

**Estimation of background processes with a  
jet misidentified as a hadronically decaying  
 $\tau$ -lepton and measurements of Simplified  
Template Cross Sections in the  $H \rightarrow \tau_{lep}\tau_{had}$   
decay channel with the ATLAS experiment at  
 $\sqrt{s} = 13 \text{ TeV}$**



Albert-Ludwigs-Universität  
Freiburg

Master thesis

for obtaining the degree of Master of Science (M.Sc.)  
for the course of studies physics

submitted by  
Fabian Becherer



## Erklärung

Hiermit versichere ich, die eingereichte Masterarbeit selbständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt zu haben. Wörtlich oder inhaltlich verwendete Quellen wurden entsprechend den anerkannten Regeln wissenschaftlichen Arbeitens (lege artis) zitiert. Ich erkläre weiterhin, dass die vorliegende Arbeit noch nicht anderweitig als Masterarbeit eingereicht wurde.

.....  
(Ort, Datum)

.....  
(Unterschrift)

Angenommen am: 01.09.2017

1. Referent Prof. Dr. M. Schumacher



## Abstract

In July 2012 both the ATLAS and the CMS experiments at CERN announced the observation of a new resonance with a combined result for the mass of  $(125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.}))$  GeV [1]. Several subsequent measurements have confirmed that the observed boson is consistent with the SM predictions of the Higgs boson. The  $H \rightarrow \tau\tau$  decay mode is the most sensitive one at the LHC to measure Yukawa couplings of the Higgs boson. This makes it an important decay channel for the ATLAS and CMS experiments in the analysis of the new dataset collected since 2015 at the LHC. This thesis presents the result of the signal strength  $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}} = 0.80 \pm 0.4$  with a cut-based analysis of the  $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$  decay mode, where one  $\tau$ -lepton decays leptonically and one hadronically, for a combined 2015 and 2016 dataset with an integrated luminosity of  $36 \text{ fb}^{-1}$  at a centre-of-mass energy of 13 TeV. The observed (expected)  $p_0$ -value is given by 0.032 (0.017) corresponding to a significance of  $1.9$  ( $2.1$ ) $\sigma$ . Events with jets misidentified as hadronically decaying  $\tau$ -leptons are one of the dominant backgrounds for this analysis. A new data-driven estimation for this background is developed, validated and applied in this thesis. Furthermore first Simplified Template Cross Sections (STXS) sensitivity studies are performed for the  $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$  decay channel and the observed (expected) cross sections for gluon fusion  $\sigma_{\text{ggF}}$  and vector boson fusion  $\sigma_{\text{VBF}}$  STXS bins of  $\sigma_{\text{ggF}} = (2.6 \pm 2.1) \text{ pb}$  ( $2.8 \text{ pb}$ ) and  $\sigma_{\text{VBF}} = (0.15 \pm 0.17) \text{ pb}$  ( $0.22 \text{ pb}$ ) are discussed.

## Zusammenfassung

Im Juli 2012 haben das ATLAS und das CMS Experiment am CERN die Entdeckung einer neuen Resonanz mit einem kombinierten Ergebnis für die Masse von  $(125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.}))$  GeV beobachtet [1]. Verschiedene anschließende Messungen haben bestätigt, dass das beobachtete Boson den Vorhersagen des Higgs Bosons des Standard Modells entspricht. Der  $H \rightarrow \tau\tau$  Zerfallskanal ist der sensitivste für die Messung der Yukawa-Kopplung des Higgs Bosons. Dies macht ihn zu einem wichtigen Zerfallskanal für das ATLAS und CMS Experiment bei der Analyse des neuen Datensatzes, welcher seit 2015 am LHC aufgenommen wird. Diese Arbeit präsentiert das Ergebnis der Signalstärke  $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}} = 0.80 \pm 0.4$  mit einer schnittbasierten Analyse des  $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$  Zerfallskanals, in welchem ein  $\tau$ -Lepton leptonic und eines hadronisch zerfällt, für einen kombinierten Datensatz aus den Jahren 2015 und 2016 mit einer integrierten Luminosität von  $36 \text{ fb}^{-1}$  und einer Schwerpunktsenergie von 13 TeV. Der beobachtete (erwartete)  $p_0$ -Wert ist 0.032 (0.017) und entspricht einer Signifikanz von  $1.9$  ( $2.1$ ) $\sigma$ . Ereignisse mit Jets, die als hadronisch zerfallendes  $\tau$ -Lepton fehl identifiziert werden, sind einer der dominanten Untergründe in dieser Analyse. Eine neue datenbasierte Bestimmung dieses Untergrunds wird entwickelt, geprüft und angewendet. Zusätzlich werden erste Studien für "Simplified Template Cross Sections" (STXS) Messungen für den  $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$  Zerfallskanal durchgeführt und die Ergebnisse für die beobachteten (erwarteten) Wirkungsquerschnitte für die Gluon-Fusion  $\sigma_{\text{ggF}}$  und Vektor-Boson-Fusion  $\sigma_{\text{VBF}}$  STXS bins von  $\sigma_{\text{ggF}} = (2.6 \pm 2.1) \text{ pb}$  ( $2.8 \text{ pb}$ ) und  $\sigma_{\text{VBF}} = (0.15 \pm 0.17) \text{ pb}$  ( $0.22 \text{ pb}$ ) diskutiert.



---

# CONTENTS

---

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Theoretical background</b>	<b>4</b>
2.1	The Standard Model . . . . .	4
2.1.1	Elementary particles . . . . .	4
2.1.2	Interactions in the Standard Model and gauge invariance . . . . .	5
2.1.3	Mass generation and Higgs mechanism . . . . .	10
2.2	The Higgs boson . . . . .	13
2.2.1	Higgs boson production at the LHC . . . . .	13
2.2.2	Higgs boson decay . . . . .	14
2.2.3	The $\tau\tau$ -decay channel . . . . .	15
<b>3</b>	<b>Experiment</b>	<b>18</b>
3.1	LHC . . . . .	18
3.2	ATLAS experiment . . . . .	20
3.2.1	Coordinate system . . . . .	20
3.2.2	Inner detector . . . . .	21
3.2.3	Calorimeters . . . . .	22
3.2.4	Muon system . . . . .	24
3.2.5	Trigger . . . . .	24
<b>4</b>	<b>Higgs boson measurements at the LHC</b>	<b>28</b>
4.1	Discovery . . . . .	28
4.2	Measurements in Run 1 . . . . .	28
4.2.1	Mass . . . . .	28
4.2.2	Signal strength . . . . .	29
4.2.3	$CP$ -properties . . . . .	30
4.3	Measurements in Run 2 . . . . .	32
4.3.1	Total cross section . . . . .	32
4.3.2	Mass . . . . .	32
4.3.3	Signal strength . . . . .	32
4.4	Simplified Template Cross Section . . . . .	33
4.4.1	Object definitions . . . . .	35
4.4.2	Splitting of production modes . . . . .	35
4.4.3	Stage 0 . . . . .	36
4.4.4	Stage 1 . . . . .	36
4.4.5	Results . . . . .	40
<b>5</b>	<b>Data set and simulated events</b>	<b>44</b>
<b>6</b>	<b>Object definitions</b>	<b>48</b>

<b>7</b>	<b>Background processes</b>	<b>54</b>
7.1	W/Z boson production in association with jets . . . . .	54
7.2	Diboson . . . . .	55
7.3	Single top quark and top quark pair production . . . . .	55
7.4	QCD multi-jet production . . . . .	56
<b>8</b>	<b>Event selection</b>	<b>58</b>
8.1	<i>Preselection</i> . . . . .	59
8.2	<i>VBF</i> region . . . . .	61
8.3	<i>Boosted</i> region . . . . .	61
8.4	Event yields . . . . .	63
<b>9</b>	<b>Signal strength measurement</b>	<b>64</b>
9.1	Background estimation . . . . .	64
9.2	Fit model . . . . .	67
9.3	Results . . . . .	67
<b>10</b>	<b>Background processes from jets misidentified as <math>\tau_{had}</math></b>	<b>70</b>
10.1	Description of the underlying method . . . . .	70
10.1.1	"Anti- $\tau$ " control region and estimation of "fake" events . . . . .	71
10.1.2	"Fake-factor" estimation . . . . .	71
10.1.3	Individual "fake-factors" . . . . .	72
10.1.4	Relative contribution . . . . .	73
10.1.5	Multi-jet contribution . . . . .	74
10.1.6	Summary of the method . . . . .	76
10.2	Application of the "fake-factor" method . . . . .	77
10.2.1	Isolation factor . . . . .	78
10.2.2	Relative contributions . . . . .	80
10.2.3	Individual "fake-factors" . . . . .	83
10.2.4	Combined "fake-factors" . . . . .	84
10.3	Further studies for the "fake-factor" method . . . . .	85
10.3.1	Use of $W + jets$ "fake-factor" for all non multi-jet events . . . . .	85
10.3.2	2015 and 2016 combined "fake-factor" . . . . .	88
10.3.3	Use of the "same-sign" isolation factor for the "opposite-sign" region . . . . .	89
10.3.4	Backgrounds in the control regions for the Isolation Factor (IF) . . . . .	91
10.3.5	"Same-sign" closure test . . . . .	92
10.3.6	Closure test for individual "fake-factors" . . . . .	94
10.4	Summary of systematic uncertainties . . . . .	97
10.4.1	Isolation factor . . . . .	97
10.4.2	Relative contribution . . . . .	99
10.4.3	Individual "fake-factor" . . . . .	99
10.4.4	Summary . . . . .	100

<b>11 Simplified Template Cross Sections</b>	<b>102</b>
11.1 STXS bin definition and expected events yields . . . . .	102
11.1.1 Stage 0 . . . . .	102
11.1.2 Stage 1 . . . . .	103
11.2 Study of expected sensitivity . . . . .	105
11.2.1 Stage 0 . . . . .	105
11.2.2 Stage 1 . . . . .	109
11.3 Results at Stage 0 . . . . .	111
<b>12 Conclusion and Outlook</b>	<b>114</b>
<b>A Appendix</b>	<b>116</b>
A.1 "Signal strength estimation" . . . . .	116
A.2 "Fake background estimation" . . . . .	117
A.3 Simplified Template Cross Sections . . . . .	121
<b>Literatur</b>	<b>126</b>



---

# 1 INTRODUCTION

---

Particle physics studies the properties and interactions of the fundamental constituents of nature. The Standard Model (SM) developed in the 1960s is the precisest description of the micro cosmos and is able to describe nearly all observations in experiments. The matter surrounding us consists of half-integer spin particles, so called fermions and their interactions are described by integer spin particles, so called bosons. The SM explains three out of the four fundamental interactions, the electromagnetic, weak and strong interaction based on relativistic quantum field theories. Only gravity, which is negligible at the energy scales considered in particles physics, is excluded.

The SM predicted several new particles, besides precisely describing the particles known at the time of its formulation. Since then, all of the predicted particles have been found experimentally, as the  $Z^0$  boson, the  $W^\pm$  boson and the top quark discovered between 1983 and 1995 [2–7].

In the years 2011 and 2012 the **L**arge **H**adron **C**ollider (LHC) [8] at **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire (CERN) near Geneva collided protons with an energy of 3.5 (2011) and 4 TeV (2012). The data taken by the **A** Toroidal **L**H**C** **A**pparatu**S** (ATLAS) and **C**ompact **M**uon **S**olenoid (CMS) collaborations were used to discover the last unobserved particle of the SM, the so called Higgs boson. The Higgs boson is predicted by the SM to describe mass terms for fermions and bosons based on the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism [9–14]. The ATLAS and CMS experiments observed in the years 2011 and 2012 the Higgs boson with a significance of  $5.9\sigma$  and  $5.0\sigma$  [15, 16] and performed a combined Higgs boson mass determination of  $m_H = (125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.}))$  GeV [1]. Several subsequent measurements, such as that of the coupling strength in different production and decay modes [17] or of the spin [18, 19] have confirmed that the observed boson is consistent with the SM predictions of the Higgs boson. The discovery of the Higgs boson at CERN lead to a Nobel Prize for Peter Higgs and Francois Englert in the year 2013 for the theoretical formulation of the underlying mechanism.

With the combined 2011 and 2012 dataset no deviations of the Higgs boson properties from SM predictions were found. Since 2015 new data is being collected by the ATLAS and CMS experiment at the LHC with a centre-of-mass energy of 13 TeV. The increased energy of the collider leads to a higher cross section for the Higgs boson production. Thus substantial improvements are expected for Higgs boson analyses like the  $H \rightarrow \tau\tau$  analysis, which with the previous dataset saw  $4.5\sigma$  evidence [20] for this process, but did not have enough statistics for the customary  $5\sigma$  observation. Furthermore several physics models beyond the SM can be probed with the additional data by using new analysis techniques like the Simplified Template Cross Sections (STXS) [21]. The STXS measure cross sections in exclusive regions of phase space for specific production modes. This reduces the dependences on theoretical uncertainties in the SM predictions and on the underlying physics model compared to the coupling strength approaches used previously [17].

The  $H \rightarrow \tau\tau$  decay mode is the most sensitive decay mode at the LHC to measure Yukawa couplings of the Higgs boson. Furthermore several methods can be used in this decay mode to measure potential  $CP$  mixing of the Higgs bosons. These properties make  $H \rightarrow \tau\tau$  an

important decay channel for the new dataset of the ATLAS and CMS experiments taken at the LHC.

This thesis will focus on the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel, where one  $\tau$ -lepton decays leptonically and one hadronically, with a combined 2015 and 2016 dataset of  $36 \text{ fb}^{-1}$  at a centre-of-mass energy  $\sqrt{s} = 13 \text{ TeV}$  collected with the ATLAS experiment.

In Chapter 2 the theoretical background for this thesis is presented, which includes the Standard Model of particle physics with a focus on the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism and the Higgs boson.

The ATLAS experiment used to collect the data for this analysis is discussed in Chapter 3 and Chapter 4 presents the current results for measurements of Higgs boson properties, like the mass, coupling strengths and  $CP$  properties, at the LHC and introduces the new measurement technique Simplified Template Cross Sections and shows first results for this technique in Run 2. The data set, the object definition, the background processes and the event selections used are described in Chapters 5-8.

This thesis measures and discusses the signal strength  $\mu = \sigma_{obs}/\sigma_{SM}$  and the significance of the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel in Chapter 9. For this a new data-driven background estimation for events with jets misidentified as hadronically decaying  $\tau$ -lepton, which is one of the dominant backgrounds in the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel, so called "fakes", is developed, validated and applied in Chapter 10. This method is based on the "fake-factor" method already used in this decay channel for Run 1 [20], but especially a new data-driven approach for calculation of the relative contribution of multi-jet processes, which contributes to the "fake" background, is developed and validated. Furthermore several assumptions and simplifications are used for the "fake" background calculation with the new combined 2015 and 2016 data-set, which will be validated by several studies and closure tests.

Furthermore the same data set and background estimations are used for sensitivity studies and measurements of the Simplified Template Cross Sections in this channel. The results of these studies and measurements are presented in Chapter 11.

---

## 2 THEORETICAL BACKGROUND

---

This chapter gives an overview of the theoretical background. First the Standard Model (SM) of particle physics is presented. Further the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism and the Higgs boson of the SM is explained. Finally the production and decay modes of the Higgs boson are discussed.

### 2.1 The Standard Model

The Standard Model of particle physics describes all known particles and their fundamental interactions with great precision. The model has successfully predicted several unknown particles, like the top quark, the gauge bosons of the weak interaction and the Higgs boson. It was possible to experimentally prove the existence of all of them. For example the  $Z^0$  boson,  $W^\pm$  boson and the top quark were discovered between 1983 and 1995 [2–7]. Out of the four known fundamental interactions three can be described by the SM. Only gravity is not included, which is negligible at the energy scales of current collider experiments. The following introduction is based on references [22–24].

#### 2.1.1 Elementary particles

The elementary particles can be grouped into two types. The first type is called fermions and includes particles with half integer spin. Particles of the second type are called bosons and carry integer spin. The fermions consist of colorless leptons ( $\nu_e, e$ ), ( $\nu_\mu, \mu$ ) and ( $\nu_\tau, \tau$ ) and color charged quarks (u,d), (c,s) and (t,b). Leptons and quarks are organised in three families or generations. The first generation of fermions includes the electron neutrino, the electron, the up and the down quark. Our surrounding matter consists of the last three of these. The second and third generation are copies of the first generation with identical quantum numbers, except for a different flavour and an increasing mass. Particles from these generations only appear in cosmic rays and collider experiments. Each generation of quarks contain one up-type quark with fractional electric charge  $q = 2/3$  and one down-type quark carrying  $q = -1/3$ . This gives six different quark flavours (up, down, charm, strange, top, bottom). The families of leptons consist of electrons  $e$ , muons  $\mu$ , and  $\tau$ -leptons with electric charge  $q = -1$  and uncharged neutrinos of the same lepton flavour: the electron neutrino  $\nu_e$ , the muon neutrino  $\nu_\mu$  and the  $\tau$ -neutrino  $\nu_\tau$ . For each particle there exists a corresponding anti-particle. The anti-particle has the same properties in spin, mass, lifetime and strength of interaction but has an opposite magnetic moment and opposite charge properties.

The fundamental interactions of the SM are the strong, the weak and the electromagnetic force, described by gauge field theories with corresponding gauge bosons with spin  $s = 1$ .

The first well understood interaction was the electromagnetic interaction. Only particles with electromagnetic charge contribute to this interaction. This means that all fermions except of neutrinos participate to it. The corresponding exchange particle is the massless photon.

The weak interaction, which among other things causes the beta decay, couples to the weak isospin  $I_w$ . Quarks and leptons have a weak isospin  $I_w = 1/2$ . This force has three massive exchange the neutral  $Z^0$  boson and the charged  $W^\pm$  bosons particles, which have a weak

isospin charge themselves. Owing to the fact that these particles have a mass in contrast to the photon, the weak interaction is weak at low energies. The exchange of a  $Z^0$  boson is called neutral weak current and couples to all fermions, while the exchange of a  $W^\pm$  boson is called charged weak current and couples only to left-handed particles and right-handed anti-particles.

The strong interaction is mediated via massless gluons and is described by the Quantum

Three generation of fermions				
	I	II	III	
mass->	2.3 MeV	1.3 GeV	173 GeV	0
charge->	2/3	2/3	2/3	0
spin->	1/2	1/2	1/2	0
name->	<b>u</b> up	<b>c</b> charm	<b>t</b> charm	<b><math>\gamma</math></b> photon
				<b>H</b> Higgs boson
	<b>d</b> down	<b>s</b> strange	<b>b</b> bottom	<b>g</b> gluon
	<b><math>\nu_e</math></b> electron-neutrino	<b><math>\nu_\mu</math></b> muon-neutrino	<b><math>\nu_\tau</math></b> tau-neutrino	<b><math>Z^0</math></b> Z boson
	<b>e</b> electron	<b><math>\mu</math></b> muon	<b><math>\tau</math></b> tau	<b><math>W^\pm</math></b> W boson
				<b>80.4 GeV</b>

Figure 2.1: All particles and their properties in the Standard Model [25].

Chromodynamics. The interaction couples to the colour charge. All quarks and the gluons themselves have a colour charge. The self-interaction ensures that the range of the interaction is limited.

The last particle of the SM is the Higgs boson, which will be discussed in Section 2.2. An overview of all particles and their properties in the SM can be found in Figure 2.1.

### 2.1.2 Interactions in the Standard Model and gauge invariance

The Standard Model uses local gauge symmetries to describe interactions. A group of global and local transformations is required under which the Lagrangian density should be invariant. These symmetries constrain the form of the underlying Lagrangian and contribute to the construction of the theories.

#### Quantum Electrodynamics

Quantum Electrodynamics (QED) describes the electromagnetic interaction with an underlying  $U(1)_Q$  symmetry. The corresponding exchange particle of this relativistic quantum field theory is the photon  $\gamma$ , which couples to the electric charge  $Q$ . The Lagrangian density  $\mathcal{L}$  of

a free fermion with mass  $m$  is given by:

$$\mathcal{L} = i\bar{\psi}\gamma_\mu\partial^\mu\psi - \bar{\psi}m\psi, \quad (2.1)$$

with the Dirac spinor  $\psi$ , the Dirac matrices  $\gamma^\mu$  and the partial derivative  $\partial_\mu = \frac{\partial}{\partial x^\mu}$  [24]. This leads to an equation of motion, which is called the Dirac equation, by applying the Euler-Lagrange equation:

$$(i\gamma^\mu\partial_\mu - m)\psi = 0. \quad (2.2)$$

For the requirement of a locally invariant gauge theory under all transformations of the  $U(1)_Q$  symmetry group, Equation 2.1 has to be invariant under the transformation of the form:

$$\psi \rightarrow e^{-iQ\alpha(x)}\psi, \quad (2.3)$$

with the electric charge  $Q$  and the local phase  $\alpha(x)$ . Applying the transformation to equation (2.1) a new term enters the Lagrangian and breaks the gauge invariance:

$$\partial_\mu\psi \rightarrow e^{i\alpha(x)}\partial_\mu\psi + ie^{i\alpha(x)}\psi\partial_\mu\alpha(x). \quad (2.4)$$

To restore local gauge invariance the derivatives have to be replaced by a covariant derivative ( $\partial_\mu \rightarrow D_\mu = \partial_\mu + ieA_\mu$ ), which corresponds to an additional coupling between the fermion and a new vector field  $A_\mu$ , which transform as:

$$A_\mu \rightarrow A_\mu + 1/e \cdot \partial_\mu\alpha. \quad (2.5)$$

The gauge invariant Lagrangian density  $\mathcal{L}$  of QED is then given by:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi. \quad (2.6)$$

Local phase invariance is ensured with the introduced vector field  $A_\mu$ . The vector field couples to fermions with the coupling strength  $e$ . The photon  $\gamma$  can be identified as this gauge field by adding a kinematic term to the Lagrangian density  $\mathcal{L}$  (Equation 2.6). Its field strength tensor  $F_{\mu\nu}$  is defined as:

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu, \quad (2.7)$$

and the Lagrangian of the QED  $\mathcal{L}_{QED}$  is defined as [24]:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \quad (2.8)$$

The introduction of any mass term of the form  $(-\frac{1}{2}m^2 A^\mu A_\mu)$  to the equation would violate the gauge invariance. Thus QED requires a massless gauge boson, which is in agreement with the experimental measurements, where upper limits of the photon mass of  $m_\gamma < 3 \cdot 10^{-27}$  eV can be set [26].

### Quantum Chromodynamics

Quantum Chromodynamics (QCD) describes the strong interaction of quarks and gluons, which couples to the color charge. Experiments have observed that quarks appear in three different flavour states: red, blue and green, or the corresponding anti-color for antiquarks [27]. Gluons carry combinations of color and anticolor. The QCD is the gauge theory of a local  $SU(3)_C$  invariance for which the four component Dirac spinor  $\psi$  is replaced by a vector of three spinors ( $\Psi = (\Psi_{red}, \Psi_{blue}, \Psi_{green})^T$ ). The dimension of the vector refers to the color degrees of freedom. The local  $SU(3)$  transformation of a free quark field  $\psi(x)$  is:

$$\psi \rightarrow \exp\left(i\frac{g_s}{2} \sum_{a=1}^8 \lambda_a \beta_a(x)\right) \psi(x) \quad (2.9)$$

with the coupling strength  $g_s$ , the eight Gell-Man matrices  $\lambda_a$  and the eight-component gauge functions  $\vec{\beta}(x)$  [28].

In contrast to QED the  $SU(3)_C$  structure is non-abelian and the gluons interact with themselves. Due to this, the field strength tensor  $G_{\mu\nu}^a$  for the gluon fields  $G_\mu^a$  ( $a=1, \dots, 8$ ) acquire an additional term:  $G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c$ , where  $f^{abc}$  refer to the structure constants of  $SU(3)_C$ . Since  $SU(3)_C$  has eight generator, eight gluon fields are defined. As for the QED a covariant derivative

$$D_\mu = \partial_\mu + ig_s \frac{\lambda_a}{2} G_\mu^a. \quad (2.10)$$

is introduced and the Lagrangian density of QCD for a single quark flavour state reads:

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - g_s(\bar{\psi}\gamma^\mu \frac{\lambda_a}{2}\psi)G_\mu^a - \frac{1}{4}G_{\mu\nu}^a G^{a,\mu\nu} \quad (2.11)$$

The self-interaction of the gluons, described by the last term of Equation 2.11, introduces a finite interaction range for QCD. The gauge invariance requires massless gluons in agreement with experimental observations. The strong force becomes weak at very short distances, leading to so-called asymptotic freedom [29, 30]. In the opposite direction, a higher spatial separation between colour-charged particles leads to an increase of the interaction potential. Thus quarks never occur as free particles, instead they are bound in mesons or baryons, so called confinement. Due to this experimentally observed particles will always be colorless or color-neutral. While mass terms for gluons break the  $SU(3)$  invariance, mass terms for quarks are allowed, given that quarks of different color have the same mass.

### Electroweak Unification

The carriers of the weak interaction are the charged  $W^\pm$  bosons and the neutral  $Z^0$  boson, with a mass of  $m_{W^\pm}=80.4$  GeV and  $m_{Z^0}=91.2$  GeV [31]. Weak interaction with and without a flavour transition for quarks are observed.

Experimental measurements observed that the weak interaction mediated by a  $W^\pm$  boson violates the parity maximally [32, 33], since the  $W^\pm$  boson couples only to left-handed particles and right-handed anti-particles. As a consequence Glashow, Salam and Weinberg

proposed a theory of electroweak interaction [34–36]. This theory is able to describe the weak and the electromagnetic force by using a  $SU(2)_{L,I_w} \times U(1)_Y$  symmetry, where  $Y$  is the hypercharge,  $I_w$  the isospin and  $L$  refers to the coupling of the weak interaction to left-handed fermions. The Gell-Mann-Nishijima relation describes the relation between the electric charge  $Q$ , the hypercharge  $Y$  and the third component of the weak isospin  $I_w^3$  [37, 38]:

$$Q = I_w^3 + \frac{Y}{2}. \quad (2.12)$$

$SU(2)_{L,I_w}$  singlets with  $I_w=0$  describe fermions in right-handed states, while the left-handed fermions are assigned to  $SU(2)_{L,I_w}$  doublets with  $I_w=1/2$  and  $I_w^3=\pm 1/2$ . An overview of fermions and their quantum numbers is shown in Figure 2.2. Recent measurements of neutrino oscillations indicate that at least two neutrinos have a non zero mass [39–41]. However this thesis will assume that neutrinos are massless. Thus only left-handed neutrinos are listed in Figure 2.2, since a right-handed massless neutrino would not interact within the SM.

The quarks appear in weak eigenstates ( $d', s', b'$ ) with  $I_3=-1/2$  as mixtures of their mass

Three generation of fermions			Quantum numbers				
	I	II	III	$I_w$	$I_w^3$	$Y$	$Q[e]$
Quarks	$u_L$	$c_L$	$t_L$	$1/2$	$1/2$	$1/3$	$2/3$
	$d'_L$	$s'_L$	$b'_L$	$1/2$	$-1/2$	$1/3$	$-1/3$
	$u_R$	$c_R$	$t_R$	$0$	$0$	$4/3$	$2/3$
	$d_R$	$s_R$	$b_R$	$0$	$0$	$-2/3$	$-1/3$
Leptons	$\nu_{e,L}$	$\nu_{\mu,L}$	$\nu_{\tau,L}$	$1/2$	$1/2$	$-1$	$0$
	$e_L$	$\mu_L$	$\tau_L$	$1/2$	$-1/2$	$-1$	$-1$
	$e_R$	$\mu_R$	$\tau_R$	$0$	$0$	$-2$	$-1$

Figure 2.2: Fermions and their quantum numbers in the Standard Model [24].

eigenstates ( $d, s, b$ ) defined by the Cabbibo-Kobayashi-Maskawa (CKM) matrix [42, 43]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.13)$$

The probability for a quark flavour transition from  $i$  to  $j$  by the interaction via a  $W^\pm$  boson is given by  $|V_{ij}|^2$ . The complex phase of the CKM matrix leads to a violation of the CP invariance [44]. The transformation of left-handed isospin doublets under  $SU(2)_{L,I_w}$  is:

$$\psi_L(x) \rightarrow \exp\left(i\frac{g}{2} \sum_{a=1}^3 \tau_a \alpha_a(x)\right) \psi_L(x), \quad (2.14)$$

with the coupling strength  $g$  and the local phase  $\alpha_a(x)$ . The generators of the  $SU(2)_{L,I_w}$  group are presented by  $\tau_a$  ( $a=1,2,3$ ) and refer to the  $(2 \times 2)$  Pauli matrices. The transformation of

left-handed isospin doublets ( $L$ ) and right-handed singlets ( $R$ ) under the  $U(1)_Y$  symmetry is given by:

$$\psi_L(x) \rightarrow \exp\left(i\frac{g'}{2}Y\beta(x)\right)\psi_L(x) \quad \text{and} \quad \psi_R(x) \rightarrow \exp\left(i\frac{g'}{2}Y\beta(x)\right)\psi_R(x), \quad (2.15)$$

here  $g'$  is a further coupling strength,  $\beta(x)$  a local phase and  $Y$  the hypercharge generator. Three vector fields  $W^a$  ( $a=1,2,3$ ) for the  $SU(2)_{L,I_w}$  group and one gauge field  $B$  for the  $U(1)_Y$  group are introduced to retain the local phase invariance. With two covariant derivatives, one for left-handed ( $D_\mu^L$ ) and one right-handed ( $D_\mu^R$ ) fermions,

$$D_\mu^L = \partial_\mu + i\frac{g}{2}\tau_a W_\mu^a + i\frac{g'}{2}YB_\mu \quad \text{and} \quad D_\mu^R = \partial_\mu + i\frac{g'}{2}YB_\mu \quad (2.16)$$

the electroweak Lagrangian density is given by:

$$\mathcal{L}_{EW} = \bar{\psi}_L i\gamma^\mu D_\mu^L \psi_L + \bar{\psi}_R i\gamma^\mu D_\mu^R \psi_R - \frac{1}{4}W_{\mu\nu}^a W^{a,\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu}. \quad (2.17)$$

The field strength tensors  $W_{\mu\nu}^a$  and  $B_{\mu\nu}$  are defined as:

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\varepsilon^{abc}W_\mu^b W_\nu^c \quad \text{and} \quad B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad (2.18)$$

where  $\varepsilon^{abc}$  defines the structure constant of the  $SU(2)$  group. Equation 2.18 enables a self-interaction of the gauge fields  $W_\mu^a$  ( $a=1,2,3$ ), while  $B_\mu$  couples only to fermions. Since  $W_\mu^3$  and  $B_\mu$  couples to neutrinos, which has no electric charge, none of these gauge fields can be identified with the physical photon field  $A_\mu$ . A linear combination of these gauge fields has to be constructed, which satisfies the requirements of  $A_\mu$ . It has to couple to left-handed and right-handed fermions with the same coupling strength and not to interact with neutrinos. Further, it has to be orthogonal to the  $Z^0$  boson. For this construction a mixing angle  $\Theta_w$  is defined:

$$\cos(\Theta_w) = \frac{g}{\sqrt{g^2 + g'^2}}. \quad (2.19)$$

The photon field  $A_\mu$  and the  $Z^0$  boson field  $Z_\mu$  can be defined orthogonal to each other by a mixing between the neutral gauge fields  $W_\mu^3$  and  $B_\mu$  of the  $SU(2)_{L,I_w} \times U(1)_Y$  group:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos(\Theta_w) & -\sin(\Theta_w) \\ \sin(\Theta_w) & \cos(\Theta_w) \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}. \quad (2.20)$$

and the charge eigenstates for the  $W^\pm$  bosons can be expressed as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (2.21)$$

The electromagnetic coupling strength  $e$  can be rewritten in terms of the coupling strengths of the  $U(1)_Y$  and  $SU(2)_{L,I_w}$  groups:

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}} = g \sin(\Theta_w) = g' \cos(\Theta_w). \quad (2.22)$$

This electroweak theory requires massless fermions and massless electroweak gauge bosons, since any mass term in Equation 2.17 would violate the gauge invariance. However experimental measurements observed gauge bosons of the electroweak interactions with nonzero mass [31] and massive fermions. The Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism explain this masses by using an additional scalar field in the context of spontaneous symmetry breaking and is described in Section 2.1.3.

### 2.1.3 Mass generation and Higgs mechanism

The Higgs mechanism was developed 1964 in [9–14]. The gauge theories described in the preceding section require invariance under local gauge transformations. The SM does not explicit by break the electroweak symmetry, instead mass terms are introduced by spontaneously breaking the symmetry via the vacuum state of a doublet of complex scalar fields with 4 degrees of freedom in the Lagrangian. This scalar field is the so called Higgs field and defined by [28]:

$$\Phi = \begin{pmatrix} \Phi_3 + i\Phi_4 \\ \Phi_1 + i\Phi_2 \end{pmatrix} \quad \Phi_i \in \mathbb{R}, Y = 1, I_w = \frac{1}{2}. \quad (2.23)$$

The corresponding Lagrangian density  $\mathcal{L}$  is:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi)(\partial^\mu \Phi) + V(\Phi). \quad (2.24)$$

To restore the invariance under the local  $SU(2)_{L,I_w} \times U(1)_Y$  transformation the derivative  $\partial_\mu$  is replaced by a covariant derivative  $D^\mu$  (Equation 2.16):

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^*(D^\mu \Phi) - V(\Phi) \text{ with} \quad (2.25)$$

$$V(\Phi) = -\mu^2 \Phi^* \Phi + \lambda (\Phi^* \Phi)^2, \quad \mu, \lambda > 0 \text{ and} \quad (2.26)$$

$$D_\mu \Phi = \left( \partial_\mu - \frac{1}{2} ig \vec{\sigma} \vec{W}_\mu - \frac{1}{2} ig' Y B_\mu \right) \Phi, \quad (2.27)$$

where  $V(\Phi)$  is the so called Higgs potential. This is the most general potential, which fullfills the requirement to be invariant under the local  $SU(2)_{L,I_w} \times U(1)_Y$  transformation and providing renormalizability. It has a local minimum for non-vanishing values of  $\Phi$  and breaks the symmetry of the  $SU(2)_{L,I_w} \times U(1)_Y$  group. The Higgs mechanism explains the masses of the  $W^\pm$  boson and  $Z^0$  boson as well as those of the fermions, while the photon remains massless. The mechanism requires a new massive scalar particle, the Higgs boson. Since the shown ground state of the Lagrangian has a  $U(1)_Y$  but no  $SU(2)_{L,I_w} \times U(1)_Y$  symmetry, the Higgs field must have non-vanishing components of weak isospin and hypercharge while it has zero charge in the second component. Thus the  $SU(2)_{L,I_w} \times U(1)_Y$  symmetry of the electroweak theory is broken to  $U(1)_Y$ .

In Figure 2.3 a simplified version of the Higgs potential, with two instead of four degree of freedom, is illustrated. This figure shows that at any fixed point in the potential minimum no rotation symmetry is given. The choice of one of these infinite points in the potential minimum refers to a spontaneous symmetry breaking. Minimizing the potential refers to a

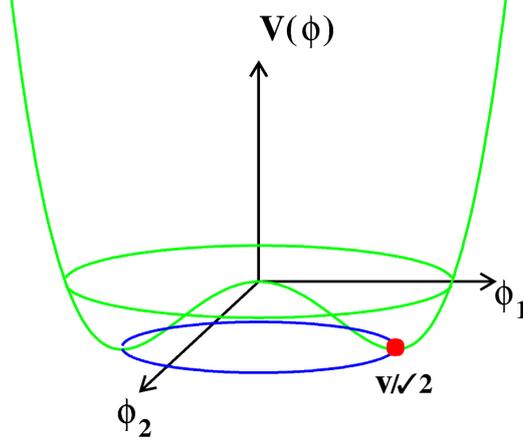


Figure 2.3: Scalar potential in the Standard Model with two degrees of freedom [45, S.12].

vacuum expectation value, which is usually defined by  $v = \mu/\sqrt{\lambda} = \Phi_3$  and  $\Phi_1 = \Phi_2 = \Phi_4 = 0$ . This choice of a particular vacuum state leads to a ground state:

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad Y = 1, I_3 = -\frac{1}{2}. \quad (2.28)$$

The vacuum expectation value  $v$  of the potential is determined by the fermi constant  $G_F$  [31]:

$$v = \left(\sqrt{2}G_F\right)^{-1/2} = 264 \text{ GeV}. \quad (2.29)$$

A parametrization of the potential  $\Phi$  around the vacuum expectation values is given by:

$$\Phi_0(x) = \frac{1}{\sqrt{2}} \exp\left(i \sum_{a=1}^3 \frac{\tau_a G_a(x)}{v}\right) \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \quad (2.30)$$

The three fields  $G_1(x), G_2(x)$  and  $G_3(x)$  refer to three Goldstone-bosons [46, 47]. These bosons are massless and scalar and an unitary gauge invariant transformation of  $\Phi_0$  ( $\Phi_0(x) \rightarrow \exp(-i \sum_{a=1}^3 \tau_a G_a(x)/v) \Phi_0(x)$ ) can eliminate them.  $H(x)$  is the Higgs boson, which is the massive excitation from the ground state. Using the Equations 2.20 and 2.21 the Lagrangian density  $\mathcal{L}$  reads:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4}\lambda H^4 \\ & + \frac{1}{2} \left(\frac{1}{2}vg\right)^2 W^{\mu,+}W_{\mu}^- + \frac{1}{2} \left(\frac{vg}{2\cos(\Theta_w)}\right)^2 Z^\mu Z_\mu \\ & + g \left(\frac{vg}{2}\right) HW^{\mu,+}W_{\mu}^- + g \left(\frac{vg}{4\cos^2(\Theta_w)}\right) HZ^\mu Z_\mu \\ & + \frac{g^2}{4}H^2W^{\mu,+}W_{\mu}^- + \left(\frac{g^2}{4\cos^2(\Theta_w)}\right) H^2Z^\mu Z_\mu + \text{const.} \end{aligned} \quad (2.31)$$

From this equation relations between different weak vector boson masses can be derived:

$$M_W = v \frac{g}{2} \quad \text{and} \quad M_Z = v \frac{\sqrt{g'^2 + g^2}}{2} = \frac{M_W}{\cos(\Theta_w S)}. \quad (2.32)$$

The mixing angle  $\Theta_w$  defines the mass ratio of the  $W^\pm$ -bosons and the  $Z^0$  boson. Further a new scalar particle with the mass  $M_H = v\sqrt{2\lambda}$  is defined. Equation 2.31 contains interaction terms of the Higgs boson  $H(x)$  with the electroweak bosons and self-interaction terms. The coupling strength of the cubic interactions of the Higgs field with the electroweak boson (HVV) are proportional to the mass of the weak vector boson. The cubic ( $H^3$ ) and the quartic ( $H^4$ ) terms describe the Higgs self-coupling. The spontaneous symmetry breaking predicts a relation between the coupling strength and gauge boson masses, which can be experimentally tested.

Further an additional coupling is introduced to describe the fermion masses. This coupling should also be invariant under  $SU(2)_{L,I_w} \times U(1)_Y$  transformations. It describes interactions between left-handed fermion  $SU(2)_{L,I_w}$ -doublets, right-handed fermion singlets and the Higgs-doublet. This so called Yukawa coupling is for the first generation of leptons given by:

$$\mathcal{L}_{Yuk}^{lep} = -g_e(\bar{\nu}_e, \bar{e})_L \Phi e_R + h.c., \quad (2.33)$$

where  $h.c.$  is the corresponding hermitian conjugate term. The equation acquires a mass for electrons, while neutrinos remain massless. For quarks an additional charge conjugated Higgs doublet

$$\Phi_C = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} \quad (2.34)$$

must be defined, to enable also interaction with  $I_w^3 = 1/2$  quarks. Thus the Lagrangian density  $\mathcal{L}_{Yuk}^{quark}$  for the first generation of quarks is:

$$\mathcal{L}_{Yuk}^{quark} = -g_d(\bar{u}, \bar{d})_L \Phi d_R + g_u(\bar{u}, \bar{d})_L \Phi_C u_R + h.c.. \quad (2.35)$$

The interactions of higher generations can be provided in the same way and the coupling constant  $g_f$  is directly proportional to the mass of the participating fermion  $M_f$ :

$$M_f = v \frac{g_f}{\sqrt{2}}. \quad (2.36)$$

The Lagrangian density of the fermion  $\mathcal{L}_{Yuk}$  is defined by:

$$\mathcal{L}_{Yuk} = -M_f \bar{f} f \left( 1 + \frac{H}{v} \right), \quad (2.37)$$

where the excitation from the ground state is considered. The final Lagrangian  $\mathcal{L}$  of the SM reads:

$$\mathcal{L} = \mathcal{L}_{Higgs} + \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{Yuk}. \quad (2.38)$$

The masses of the gauge bosons and fermions are determined by experiments with a high precision, such that only  $\lambda$  and  $M_H$  remains as free parameter. Thus with the measurement of the Higgs boson mass  $M_H$  all coupling parameters in the SM are determined.

## 2.2 The Higgs boson

### 2.2.1 Higgs boson production at the LHC

The Higgs boson can be produced in proton-proton collisions at Large Hadron Collider (LHC). The interactions of composite protons can be described at high momentum transfer in the context of hard scattering processes of quarks and gluons. The probability to find a particular parton  $i$  with a momentum fraction  $x_i$  of the total proton momentum is determined by parton distribution functions (PDF)  $f(x_i, Q^2)$  and depends on the momentum transfer  $Q^2$  in the collision. The total cross section  $\sigma_X$  for the process  $pp \rightarrow X$  can be derived by an integration over the partonic cross section  $\hat{\sigma}_{ij}$  for the interaction of parton  $i$  and  $j$  multiplied with the PDFs according to the factorization theorem [48]:

$$\sigma_X = \int \int dx_i dx_j f(x_i, Q^2) f(x_j, Q^2) \hat{\sigma}_{ij}, \quad (2.39)$$

while the partonic cross section  $\hat{\sigma}_{ij}$  is given by:

$$\hat{\sigma}_{ij} = \int \frac{|M(ij \rightarrow X)|^2}{F} dQ. \quad (2.40)$$

$M$  corresponds to the matrix element for the transition from the initial state  $ij$  to the final state  $X$ ,  $F$  gives the particle flux,  $dQ$  is the phase space factor of the given kinematics and the squared amplitude  $|M|^2$  gives the transition probability.

Different processes, which can be classified by the phenomenology and cross section, contribute to the Higgs boson production. The Feynman diagrams in Figure 2.4 illustrate the dominant production mechanisms at the LHC.

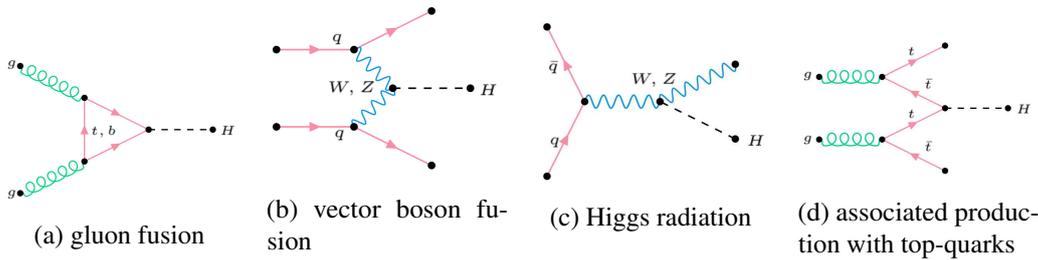


Figure 2.4: Feynman diagrams of the dominant Higgs production processes at the LHC, ordered by decreasing cross section: Gluon fusion (a), vector boson fusion (b), Higgs radiation (c) and associated production with top-quarks (d).

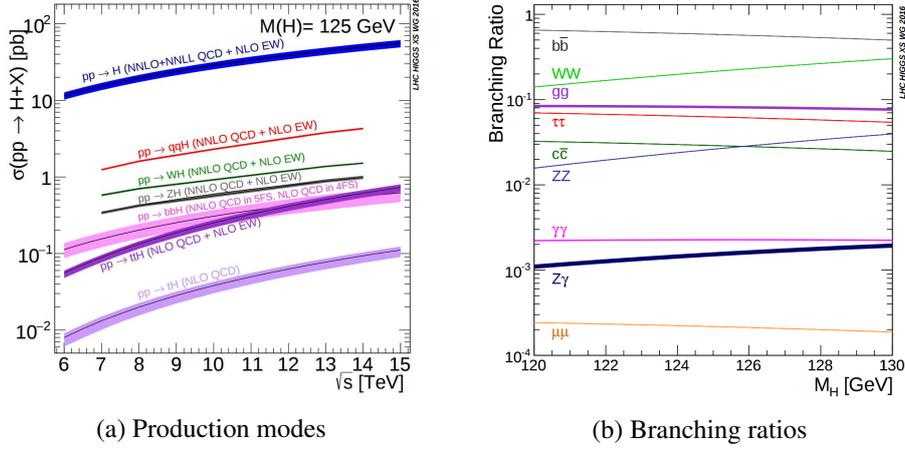


Figure 2.5: Cross sections for different production modes of the Higgs boson with a mass of 125 GeV, as a function of the centre-of-mass energy  $\sqrt{s}$  (left) and branching ratios for different Higgs boson decays as a function of the Higgs boson mass  $M_H$  (right) [49, S.148].

The gluon fusion process with a cross section of 48.6 pb at a centre-of-mass energy of 13 TeV is the dominant one for the Higgs boson production (Figure 2.4a) [49]. The Higgs boson  $H$  is produced over a quark  $q$  loop. Since the coupling strength is proportional to mass of the fermion, top  $t$  and bottom  $b$  quarks are the preferred ones for the loop.

The production process with the second highest cross section is the vector boson fusion process with a cross section of 3.8 pb at a centre-of-mass energy of 13 TeV (Figure 2.4b) [49]. Two quarks scatter in the  $t$ -channel exchanging  $W^\pm$  bosons or  $Z^0$  bosons which fuse in a Higgs boson. Characteristic for this process are the two remaining quarks in the final states with a high transverse mass  $m_{jj}$  and a high difference in the pseudorapidity  $\Delta\eta$ .

The associated production with a weak boson, so called Higgs-Strahlung, is an annihilation of a quark anti-quark pair into a weak boson, which radiates a Higgs boson (Figure 2.4c). The production cross section of this process is 2.3 pb at a centre-of-mass energy of 13 TeV [49]. The associated production with a top quark pair is an order of magnitude smaller (Figure 2.4d), due to the high invariant mass which is required to produce the heavy quarks.

The cross sections as a function of the centre-of-mass energy  $\sqrt{s}$  for these and further processes are illustrated in Figure 2.5a.

### 2.2.2 Higgs boson decay

The Higgs boson couples to massive weak vector bosons  $V$  and fermions  $f$  and its decay channels can be grouped into channels with a bosonic final state and into channels with fermions in the final state. The coupling strength to the vector bosons is proportional to the mass square of the bosons, while the coupling strength to fermions is proportional to the fermion mass.

The dominant fermion decay channels are  $H \rightarrow b\bar{b}$ ,  $H \rightarrow \tau^+\tau^-$ ,  $H \rightarrow c\bar{c}$  and  $H \rightarrow \mu^+\mu^-$  (see Figure 2.6a).

The possible decays for the bosonic case are  $H \rightarrow WW^*$ ,  $H \rightarrow ZZ^*$ ,  $H \rightarrow gg$ ,  $H \rightarrow Z\gamma^*$ ,  $H \rightarrow \gamma\gamma$ . While a direct decay into massive vector boson is possible the decay into massless gluons or photons has to use heavy quark loops or for the  $H \rightarrow \gamma\gamma$  decay over W boson loops (see Figures 2.6b-2.6d).

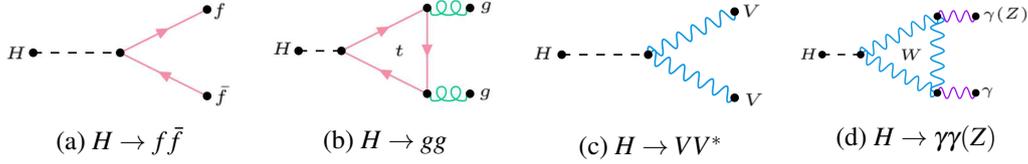


Figure 2.6: Feynman diagrams of the dominant Higgs decay channel:  $H \rightarrow f\bar{f}$  (a),  $H \rightarrow gg$  (b),  $H \rightarrow VV^*$  (c) and  $H \rightarrow \gamma\gamma(Z)$  over a W boson loop (d).

An overview of these channels and their branching ratios depending on the Higgs boson mass is given in Figure 2.5b. For a Higgs boson mass  $m_H=125$  GeV the decay into a pair of bottom quarks is the dominant one with a branching ratio of 57.7%, followed by the decay into a pair of  $W^\pm$  bosons with a branching ratio of 21.5%. The decay channel into  $\tau$ -leptons is the dominant one into leptons with a branching ratio of 6.3% [49], which is exploited in this analysis

### 2.2.3 The $\tau\tau$ -decay channel

This study focuses on the decay of a Higgs boson into a pair of  $\tau$ -leptons ( $H \rightarrow \tau^+\tau^-$ ). One challenge of this decay channel is the short lifetime ( $2.9 \cdot 10^{-13}$  s) of  $\tau$ -leptons, such that  $\tau$ -leptons are reconstructed with their decay products. There are different possibilities for the  $\tau$ -lepton decay. In Table 2.1 a summary of the important decay channels for a single  $\tau$ -lepton can be found. They can be separated into hadronic decays with a branching ratio of 65% and into leptonic decays with a branching ratio of 35%.

decay mode	$\tau^- \rightarrow$	fraction(%)
lep	$e^- \bar{\nu}_e \nu_\tau$	$17.82 \pm 0.04$
	$\mu^- \bar{\nu}_\mu \nu_\tau$	$17.39 \pm 0.04$
had	$\pi^- \nu_\tau$	$10.81 \pm 0.05$
	$\pi^- \pi^0 \nu_\tau$	$25.49 \pm 0.09$
	$\pi^- 2\pi^0 \nu_\tau$	$9.26 \pm 0.10$
	$2\pi^- \pi^+ \nu_\tau$	$8.99 \pm 0.05$

Table 2.1: Dominant decay channels of the  $\tau$ -lepton [50].

The dominant decay channel is the hadronical decay  $\tau^- \rightarrow \pi^- \pi^0 \nu_\tau$  with a branching ratio of 25.49%, followed by the two leptonic decay channels  $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$  and  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ ,

with a branching ratio of 17.82% and 17.39%. A di- $\tau$ -system has three possible decay modes. Both  $\tau$ -leptons could decay leptonically ( $\tau_{lep}\tau_{lep}$  channel) with a branching ratio of 12.4%, both could decay hadronically ( $\tau_{had}\tau_{had}$  channel) with a branching ratio of 42.6%, or one leptonically and one hadronically ( $\tau_{lep}\tau_{had}$  channel) with a branching ratio of 45.6%, on which this thesis will focus.



---

## 3 EXPERIMENT

---

### 3.1 LHC

The LHC (**L**arge **H**adron **C**ollider) [8] at CERN (**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire) near Geneva is a proton-proton (or heavy ions) accelerator, which runs since 2008. It is a two ring collider with a circumference of 27 km and is between 100 and 150 m below the groundlevel. For proton-proton collisions the beam of the accelerator contains up to 2808 bunches with  $10^{11}$  particles per bunch and a time distance of 25 ns. The proton beams are bent by 1232 superconducting dipole magnets, which are able to generate a magnetic field up to 8.3 Tesla and quadrupole magnets focus them. The provided instantaneous luminosity  $\mathcal{L}$  and centre-of-mass energy  $\sqrt{s}$  describes mainly the performance of the LHC. The instantaneous luminosity  $\mathcal{L}$  for two colliding particle bunches with an equally Gaussian shape distribution is given by [51]:

$$\mathcal{L} = \frac{N^2 \gamma_r f_{rev}}{4\pi \beta^* \epsilon_n} n_b \cdot F, \quad (3.1)$$

where  $N$  corresponds to the number of particles per bunch,  $\gamma_r$  to the relativistic Lorentz-factor,  $f_{rev}$  to the revolution frequency,  $\beta^*$  to the beta-function at the collision point,  $\epsilon_n$  to the normalized transverse beam emittance,  $n_b$  to the number of bunches per beam and  $F$  to the geometric luminosity reduction factor due to the beam crossing angle at the interaction point.

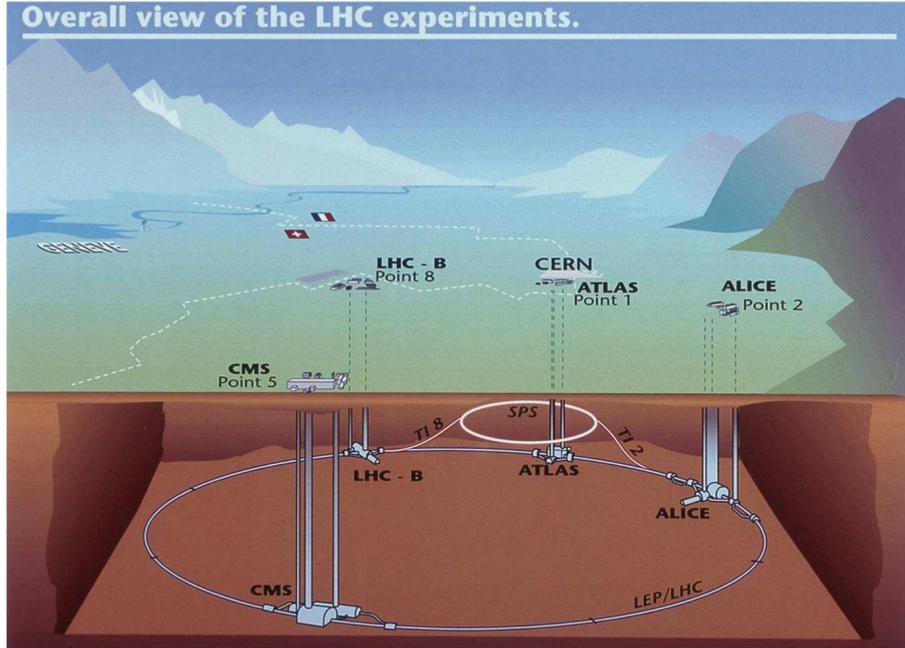


Figure 3.1: The LHC and the four major experiments [52].

With a design of an instantaneous luminosity up to  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  and a beam energy up to 7 TeV resulting in a collision centre-of-mass energies of  $\sqrt{s} = 14 \text{ TeV}$ , the LHC is the biggest and most powerful particle accelerator on earth. The LHC has four collision points. At each collision point a different detector is placed. The detectors are ALICE, LHCb and the multi-purpose detectors ATLAS and CMS. The four detector are designed for different physical questions. An overview of the LHC is given in Figure 3.1.

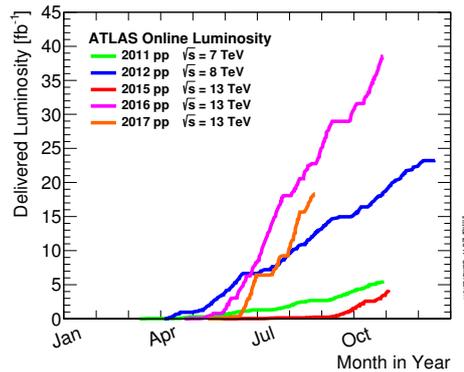


Figure 3.2: Integrated luminosity of data-taking by the ATLAS-detector split into different years until September 2017 [53].

The LHC is now in the 7th year of data-taking. The first data-taking periods were in 2011 and 2012 with centre-of-mass energies of 7 and 8 TeV. After three years of optimizations the accelerator cloud produce collisions with a centre of mass energy of 13 TeV in 2015. Table 3.1 summarize the data-taking periods. Run 1 includes the data from 2011 and 2012 with a integrated luminosity of  $\mathcal{L}_{int} = 26.4 \text{ fb}^{-1}$ . Run 2 will include the data from 2015 to 2018. In this thesis only the dataset of 2015 and 2016 will be used and defined as the Run 2 dataset. In these two years events corresponding to a integrated luminosity of  $\mathcal{L}_{int} = 39.5 \text{ fb}^{-1}$  were recorded.

Run	Year	$\sqrt{s}$ (TeV)	$\mathcal{L}_{int}$ (fb <sup>-1</sup> )
1	2011	7	5.1
1	2012	8	21.3
1	11+12	7(8)	26.4
2	2015	13	3.9
2	2016	13	35.6
2	15+16	13	39.5

Table 3.1: Overview of the LHC's data-taking until 2016.  $\sqrt{s}$  corresponds to the centre of mass energy, while  $\mathcal{L}_{int}$  refers to the integrated luminosity [53].

### 3.2 ATLAS experiment

The ATLAS (A Toroidal LHC ApparatuS) experiment is located at the LHC. The detector of the experiment is with a diameter of 25 meters and a length of 44 meters the largest particle detector ever built at a collider. It has a weight of 7000 metric tons and consists of different parts (Figure 3.3). The detector is built up in a cylindrical way, where the components from the inside to the outside are: tracking detectors, which are surrounded by superconducting magnet solenoids, electromagnetic and hadronic calorimeters and muon toroids. These components are described in the next sections and further information can be found in [54].

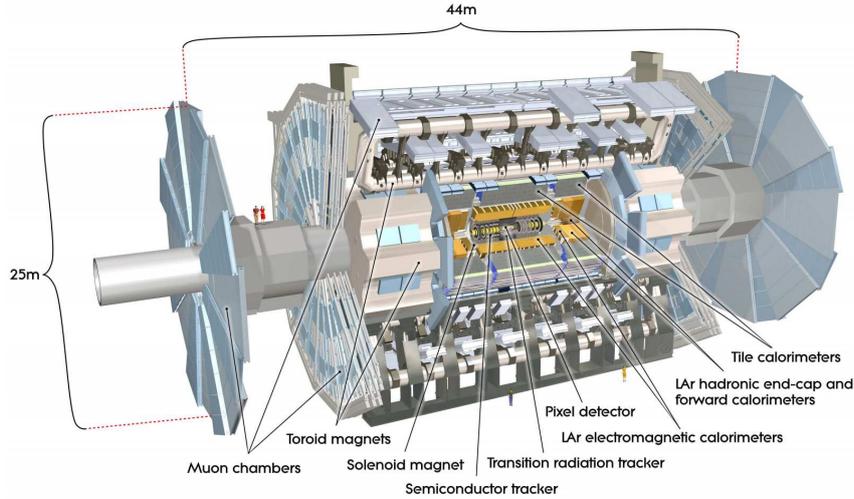


Figure 3.3: Overview of the full ATLAS detector [54, S.6].

#### 3.2.1 Coordinate system

The x-y-plane is defined transversal to the beam axis. The x-axis points to the centre of the LHC, while the y-axis is orthogonal to the x-axis and pointing upward towards the surface. The z-axis is parallel to the beam axis, so that a right-handed coordinate system results. The origin of the system is the nominal interaction point. Thus the azimuthal angle  $\Phi$  is defined in the x-y-plane. Instead of the polar angle  $\Theta$ , which describes the angle between the direction of the particle and the positive z-axis, the pseudorapidity  $\eta$  is used and given by:

$$\eta = -\ln \left( \tan \left( \frac{\Theta}{2} \right) \right). \quad (3.2)$$

The variety of soft QCD processes in a proton-proton collision induces a nearly flat distribution of the particle density in  $\eta$  [55]. For negligible masses or high energies the approximation  $\eta \approx y$ , where  $y = \frac{1}{2} \ln \left( \frac{E+p_z}{E-p_z} \right)$  is the rapidity, holds and the Lorentz invariance of  $\Delta y$  results in an invariance of  $\Delta \eta$  under a boost along the z-axis, which is unknown for two colliding particles inside the proton-proton pair.

The distance between two objects can be defined with  $\Delta\eta$  and  $\Delta\Phi$  in the corresponding coordinate system.  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2}$ , which is invariant under boosts in z-direction, describes the separation of the two objects in the  $(\eta - \Phi)$ -plane and is used in the identification as well as in the reconstruction of the objects. The particle momentum and the energy in the x-y-plane is defined by:

$$p_T = \sqrt{p_x^2 + p_y^2} \quad \text{and} \quad E_T = \sqrt{E_x^2 + E_y^2}, \quad (3.3)$$

where  $p_T$  corresponds to the transverse momentum and  $E_T$  to the transverse energy. Further information about these definitions can be found in [54, 56].

### 3.2.2 Inner detector

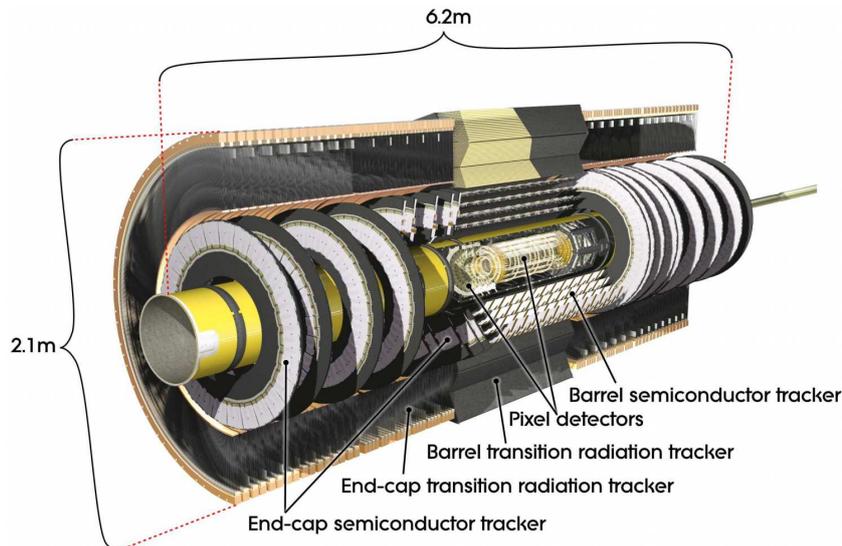


Figure 3.4: Overview of the inner detector [54, S.6].

The inner detector is the most central part of the ATLAS detector and consists of different tracking detectors (see Figure 3.4). The task of the inner detector is to provide a precise momentum measurement for charged particles, the reconstruction of the particle tracks and the reconstruction of primary and secondary vertices.

A two Tesla magnetic field, generated by the central solenoid, penetrates the inner detector. The components of the inner detector are a pixel detector and a silicon microstrip tracker (SCT) and transition radiation trackers (TRT). Pixel and SCT are especially important at small radii and cover a range up to  $|\eta| < 2.5$ , while the TRT covers a range up to  $|\eta| < 2.0$ . Both trackers are built out of a barrel region and two end-caps. In the barrel region the pixel detector and SCT are placed concentrically around the beam axis while the straw tubes of the TRT run parallel to it. The end-caps are radially mounted detectors. The inner detector has a length of 6.2 m and a radius of 2.1 m. Due to small distance between the inner detector

and the collision point, the detector has to cope with high radiation and high temperature conditions. Thus the pixel detector as well as the SCT are cooled down to  $-7^{\circ}\text{C}$ , to reduce potential damages.

The barrel and the end-cap part of the pixel detector are consist of three layers corresponding to  $250\ \mu\text{m}$  thick silicon semi-conductors with 46080 readout channels for each of the 1744 sensors. One pixel has a size between  $(R-\Phi)\times z = 50\times 400\ \mu\text{m}^2$  and  $(R-\Phi)\times z = 50\times 600\ \mu\text{m}^2$ . Thus the pixel detector reaches an intrinsic hit resolution of  $(R-\Phi)\times z = 10\times 115\ \mu\text{m}^2$ .

The SCT consists of four layers of silicon strip detector modules in the barrel region and of nine layers in the end-cap region. This ensures at least four measured points for each charged particle, which penetrates the detector. The detector has 758 active strips of 12 cm length and  $285\ \mu\text{m}$  thicknesses per strip. The SCT has a nominal hit resolution of  $(R-\Phi)\times z = 17\times 580\ \mu\text{m}^2$  and contains over 6 million readout channels.

The TRT consists of straw tubes with a 4 mm diameter filled with gas and stabilised by carbon fibres. The tubes have a length of 144 cm in the barrel region and a 37 mm length in the end-cap region. Thus only a position measurement of  $(R-\Phi) = 130\ \mu\text{m}$  is done in the barrel region and no measurements in the  $z$ -direction. The tubes enable to indicate the penetrating particle, since the transition radiation is inversely proportional to the mass of the particle. The TRT can identify electrons, which have the most transition radiation, since they are the lightest stable particles produced in the collision. An overview of the expected resolutions of the inner detector on momentum measurements is given in Table 3.2.

### 3.2.3 Calorimeters

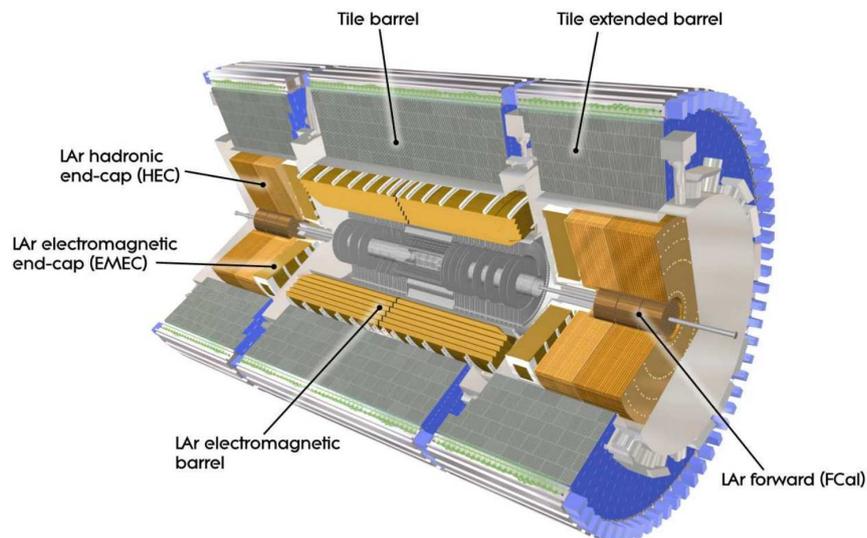


Figure 3.5: Overview of the calorimeter system [54, S.8].

### Electromagnetic calorimeter

The electromagnetic calorimeter is optimized to measure the energy of photons and electrons very precisely. It is made up of lead with a liquid argon sampling technology and has a fine granularity. The electromagnetic calorimeter is divided in a barrel part and two end-caps (see Figure 3.5). The barrel part is split into two identical half-barrels and covers a pseudorapidity range up to  $|\eta| < 1.475$  and consists of 3 layers with a granularities of  $\Delta\eta \times \Delta\Phi = 0.025/8 \times 0.025$ ,  $\Delta\eta \times \Delta\Phi = 0.025 \times 0.025$  and  $\Delta\eta \times \Delta\Phi = 0.050 \times 0.025$ . The end-caps consists of an inner and an outer wheel, which cover the range of  $1.375 < |\eta| < 3.2$ . Accordion-shaped electrodes and lead absorber plates build the calorimeter with a complete symmetry in the azimuthal angle  $\Phi$ . To compensate energy, which photons and electrons loose upstream to the calorimeter, a pre-sampler detector is added in the central region ( $|\eta| < 1.8$ ). This pre-sampler consists of an active liquid argon layer. Figure 3.6a illustrates the energy resolution of the ECAL for simulations and test measurements. Furthermore the  $\eta$ -coverage as well as the energy resolution of the electromagnetic calorimeter is given in Table 3.2.

### Hadronic calorimeter

The hadronic calorimeter is used together with other sub-detectors to measure the energy of jets and the missing transverse energy  $E_T^{miss}$ . It has a coarser granularity compared to the electromagnetic calorimeter. The barrel parts consist of a tile calorimeter ( $|\eta| < 1.0$ ) around the electromagnetic calorimeters and two extended barrels covering higher values of  $|\eta|$ :  $0.8 < |\eta| < 1.7$  (see Figure 3.5). As absorber steel plates are used while scintillating tiles are used as a active material covering a  $|\eta|$  range up to 3.2. The energy resolution of the central tile region for simulations and test beam measurements are illustrated in Figure 3.6b. Liquid argon is used as sensitive material for the hadronic end-cap (HEC) and for the forward calorimeter (FCal). While for the HEC only copper is used as an absorber, tungsten is being used for the FCal either. The  $\eta$ -coverage and the energy resolution of the different parts of the hadronic calorimeter are given in Table 3.2.

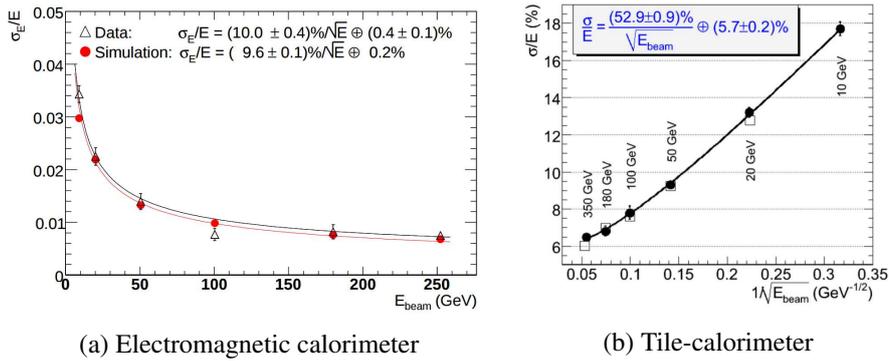


Figure 3.6: Comparison between simulations and test beam measurements of the energy resolution for the electromagnetic calorimeter (left) and the tile-calorimeter (right) of the ATLAS detector [54].

### 3.2.4 Muon system

A muon system is used to detect and measure the momentum of charged particles which transverse the inner detector and the calorimeters. Three different parts are defined in this system: the barrel region covering a absolute pseudorapidity  $|\eta|$  up to 1.4, the transition region with  $1.4 < |\eta| < 1.6$  and the end-cap region from  $|\eta| = 1.6$  up to  $|\eta| = 2.7$ . For the momentum measurements toroidal magnets are installed, which generate a field strength of 2.5 Tesla in the barrel and 3.5 Tesla in the end-cap region.

To identify the muon trajectory monitored drift tubes (MDT) are used, except for particles in large  $\eta$ -ranges, for which cathode strip chambers (CSC) are used. The CSC provide a better spatial resolution to cope with the higher signal rates in the large  $\eta$ -range.

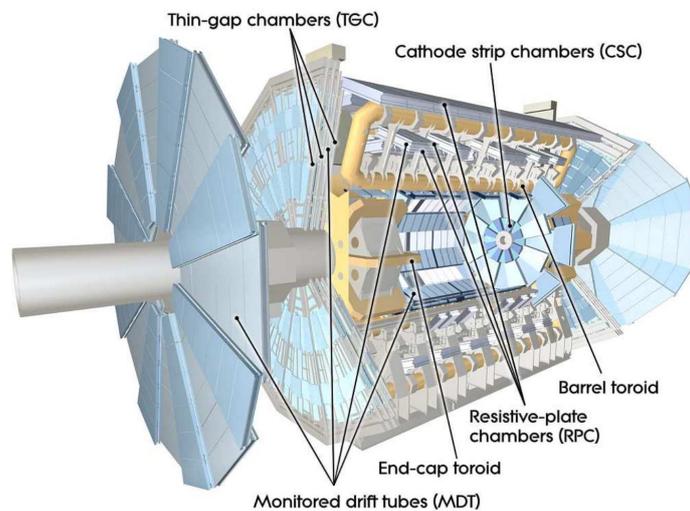


Figure 3.7: Overview of the muon system [54, S.11].

Furthermore the muon system is designed to trigger particles, which penetrate the system in the region up to  $|\eta|=2.4$ . Resistive plate chambers (RPC) in the barrel region and thin gap chambers (TGC) in the end-cap region are installed to achieve a response time of a few nanoseconds. The required resolution for the muon system is given in Table 3.2 and a schematic overview of it is illustrated in Figure 3.7.

### 3.2.5 Trigger

To reduce the LHC bunching rate of 40 MHz a two level trigger is used in Run 2 at the ATLAS experiment [57]. The trigger consists of a hardware-based first level trigger (Level-1) and a software-based high level trigger (HLT). Custom electronics are used for the Level-1 trigger to determine Regions-of-Interest (RoIs) in the detector by taking as an input coarse granularity calorimeter and muon system informations.

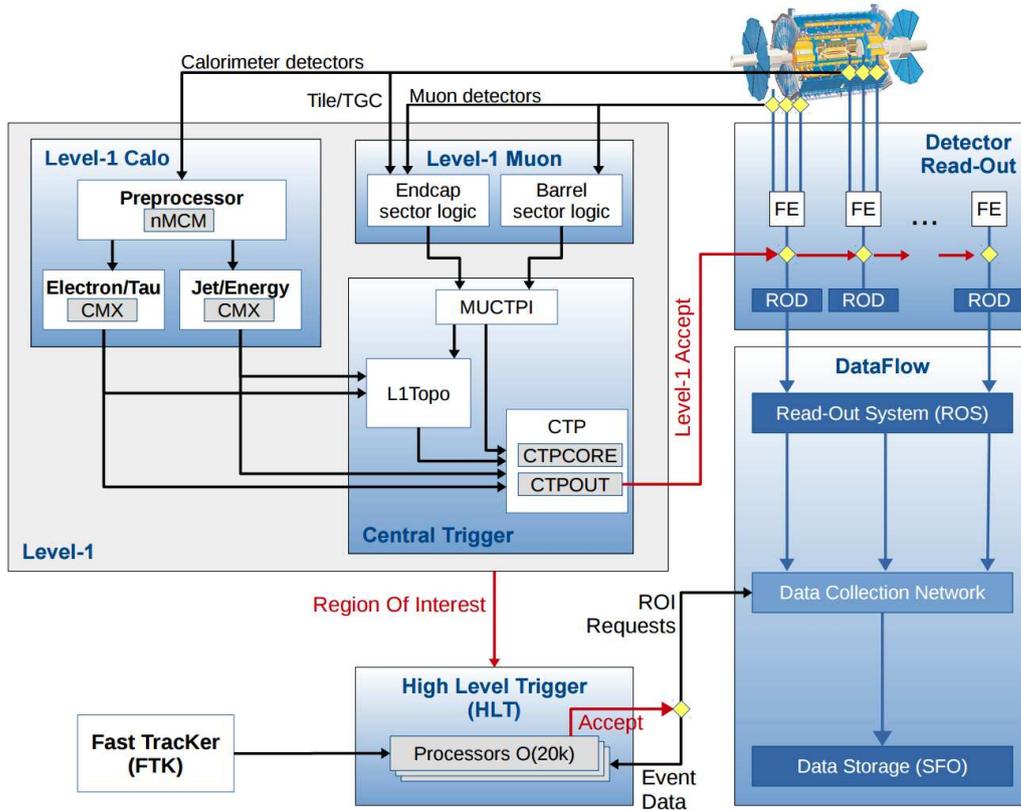
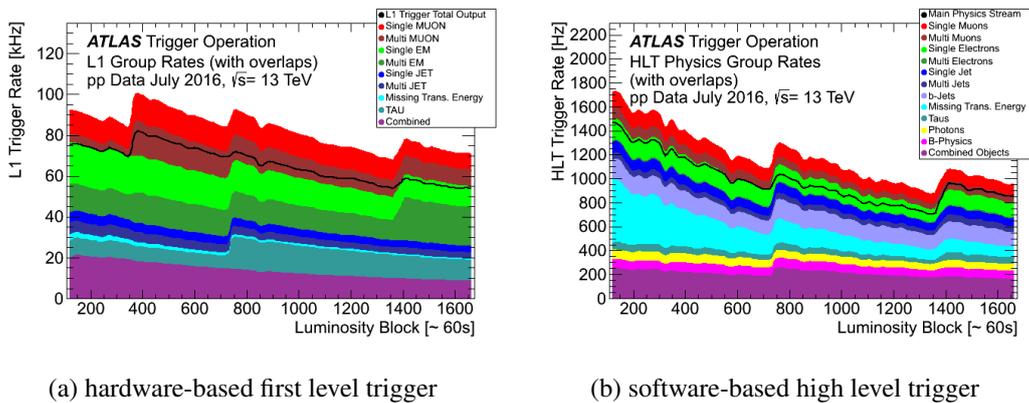


Figure 3.8: Schematic layout of the ATLAS trigger and data acquisition system used in Run 2 [57].



(a) hardware-based first level trigger

(b) software-based high level trigger

Figure 3.9: Trigger rate for the ATLAS hardware-based first level trigger (left) and the ATLAS software-based high level trigger (right) for data-taking July 2016 [58].

The Level-1 trigger reduces the event rate from 40 MHz to 100kHz, with a decision time of

2.5  $\mu$ s. These at Level-1 formed PoIs are used as an input for the HLT in which sophisticated selection algorithms are running using the full granularity detector information of the RoI and the whole event. The HLT reduces the output-rate of the Level-1 trigger of 100 kHz to approximately 1kHz on average, with a processing time of about 200 ms. A schematic overview of the ATLAS trigger and data acquisition is given in Figure 3.8. Furthermore Figure 3.9 illustrates the event rate and the distributions of so called trigger-menus for data taking in July 2016 for the Level-1 and HLT trigger.

Sub-detector	Required resolution	$\eta$ -coverage	
		Measurement	Trigger
Inner detector	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	$\pm 2.5$	
Electromagnetic calorimeter	$\sigma_E/e = 10\% / \sqrt{E} \oplus 0.7\%$	$\pm 3.2$	$\pm 2.5$
Hadronic calorimeter			
barrel and end-cap	$\sigma_E/e = 50\% / \sqrt{E} \oplus 3\%$	$\pm 3.2$	$\pm 3.2$
forward	$\sigma_E/e = 100\% / \sqrt{E} \oplus 10\%$	$3.1 <  \eta  < 4.9$	$3.1 <  \eta  < 4.9$
Muon system	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	$\pm 2.7$	$\pm 2.4$

Table 3.2: The required resolution and the pseudorapidity ( $\eta$ ) coverage of the different sub-detectors of the ATLAS experiment. Energy as well as momentum values are given in units of GeV. Further  $a \oplus b$  refers to adding  $a$  and  $b$  in quadrature:  $a \oplus b = \sqrt{a^2 + b^2}$  [54].



## 4 HIGGS BOSON MEASUREMENTS AT THE LHC

### 4.1 Discovery

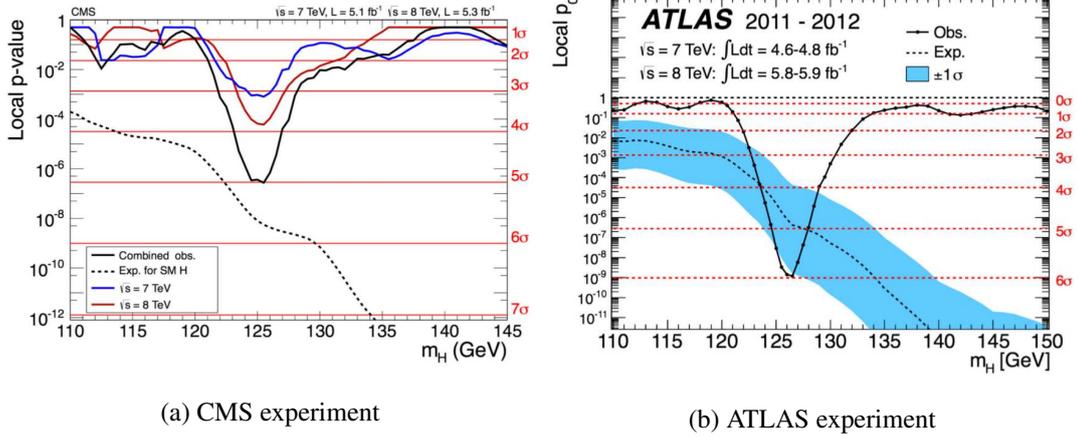


Figure 4.1:  $p_0$ -values and the corresponding significances as a functions of the Higgs boson mass  $m_H$  for the CMS experiment (left) [15] and for the ATLAS experiment (right) [16] in Run 1. The dashed lines refer to expected values and the full black lines to the observed local  $p_0$ -values.

In July 2012 both the ATLAS and the CMS experiments at CERN have announced the observation of a new resonance with a mass of 125 GeV and a significance of  $5.9 \sigma$  and  $5.0 \sigma$  [15, 16]. This corresponds to a local  $p_0$ -value of  $1.7 \cdot 10^{-9}$  and  $4.2 \cdot 10^{-6}$  (Figure 4.1), which is a criterion for the probability that the resonance is introduced by fluctuating backgrounds. A  $p_0$ -value corresponding to  $5 \sigma$  is necessary for an observation.

### 4.2 Measurements in Run 1

#### 4.2.1 Mass

Using the full dataset of Run 1 with  $25 \text{ fb}^{-1}$  the ATLAS detector was able to measure the mass of the resonance with very high precision as  $m_H = (125.36 \pm 0.37(\text{stat.}) \pm 0.15(\text{sys.})) \text{ GeV}$  [59]. This observation is consistent with the mass measurement of the CMS detector ( $m_H = (125.02 \pm 0.27(\text{stat.}) \pm 0.15(\text{sys.})) \text{ GeV}$  [60]) and the combined result of both experiments is given by [1]:

$$m_H = (125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.})) \text{ GeV}. \quad (4.1)$$

### 4.2.2 Signal strength

The experiments measured the signal strength parameter  $\mu$ , which is defined as the measured cross section normalized to SM expectation:  $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$ . This parameter is determined for different production modes and decay channels of the Higgs boson to measure potential deviations to the SM predictions. The results for the combined ATLAS and CMS measurements shown in Figure 4.2 are so far all compatible with the SM expectation  $\mu = 1$ . The current combined  $\mu$  value for all measurements is [17]:

$$\mu = 1.09 \pm 0.07(\text{stat.})_{-0.08}^{+0.09}(\text{sys.}), \quad (4.2)$$

which is consistent with the SM expectation at less than  $1\sigma$  confidence level assuming a global signal strength for all production and decay modes yields.

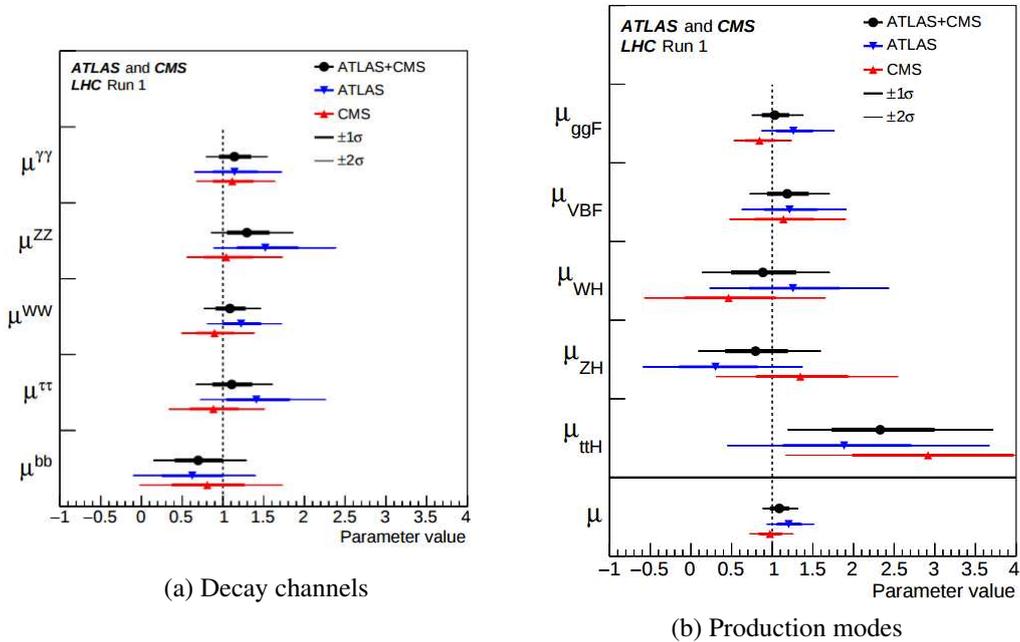


Figure 4.2: Signal strength  $\mu$  in the decay channels (left) and production modes (right) of the Higgs boson for the combined ATLAS and CMS measurements [17].

Figure 4.3a illustrates the combined ATLAS and CMS measurements of the normalized coupling strengths  $\kappa_i = g_i/g_{i,SM}$ , where  $g_i$  is the measured coupling strengths and  $g_{i,SM}$  the SM prediction. This definition includes the assumption that no beyond the SM contributions are present in loops and decays. All results are in agreement with the SM expectation of one. The reduced coupling strength scale factors  $y_{V,i} = \sqrt{\kappa_{V,i}} \frac{m_{V,i}}{v}$  for the weak bosons and  $y_{F,i} = \kappa_{F,i} \frac{m_{F,i}}{v}$  for fermions can be determined, for which the SM predicts a linearly increasing with the particle mass. In Figure 4.3b these scale factors show a good agreement as a function of the particle mass with the SM expectation.

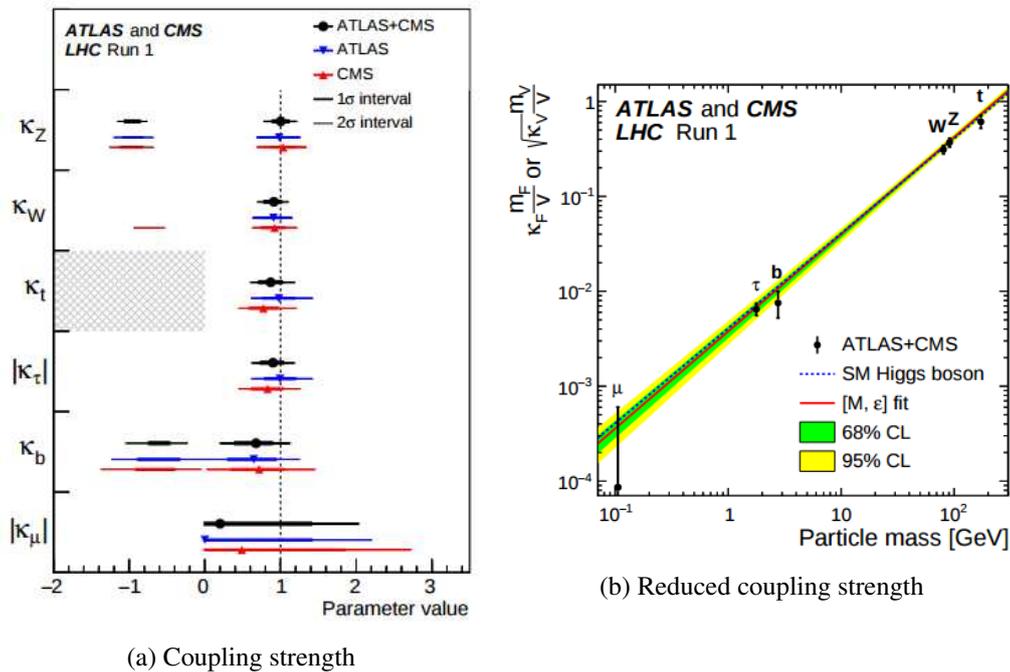


Figure 4.3: Coupling strength  $\kappa$  measurements of the Higgs boson to different particles (left) and reduced coupling strength scale factor  $y_{V,i} = \sqrt{\kappa_{V,i}} \frac{m_{V,i}}{v}$  for weak bosons and  $y_{F,i} = \kappa_{F,i} \frac{m_{F,i}}{v}$  for fermions measurements as a function of the particle mass (right), with an assumed Higgs boson mass of 125.09 GeV (Combined ATLAS and CMS measurement). The dashed lines show the Standard Model prediction [17].

### 4.2.3 CP-properties

The SM predicts a  $CP$  even Higgs boson with a spin  $J$  of zero:  $J^{CP} = 0^+$ . The  $CP$  properties of a particle describe its behaviour under a Charge conjugation and a Parity transformation. The SM predicts  $CP$  conservation in the Higgs boson production and decay, such that all fundamental interactions of a particle remain invariant under a  $CP$  transformation. For this the particle must have an eigenstate of the  $CP$  operator with even (+1) or odd (-1) eigenvalues, while a particle for which no  $CP$  eigenstate can be found would violate the  $CP$  invariance and would be direct evidence for beyond SM physics.

Run 1 results of the ATLAS and CMS experiments in the decay channels  $H \rightarrow ZZ^*$ ,  $H \rightarrow WW^*$  and  $H \rightarrow \gamma\gamma$  indicates that the Higgs boson carries  $J^{CP} = 0^+$  and all other  $J^{CP}$  values could be excluded at more than 99,9% confidence level [18, 19]. The ATLAS results for these measurements are illustrated in Figure 4.4.

However a mixing between a  $CP$  even and a  $CP$  odd eigenstate is not excluded by these measurements and its investigation is one of the major goals for the Run 2 measurements. This can be studied by using the *Optimal Observable*, which was already used in the  $H \rightarrow \tau\tau$  channel in Run 1 [61]. The  $CP$ -mixing parameter  $\tilde{d}$  of the *Optimal Observable* is directly

sensitive to the  $CP$  property of the Higgs boson coupling in the production mode. It is expected to be zero for the  $CP$ -even case of the SM, while  $CP$ -violation effects would introduce deviations from zero. The measurements are consistent with the SM expectation and exclude values of  $\tilde{d}$  over 0.05 and under -0.11 with a confidence level of 68% (see Figure 4.5).

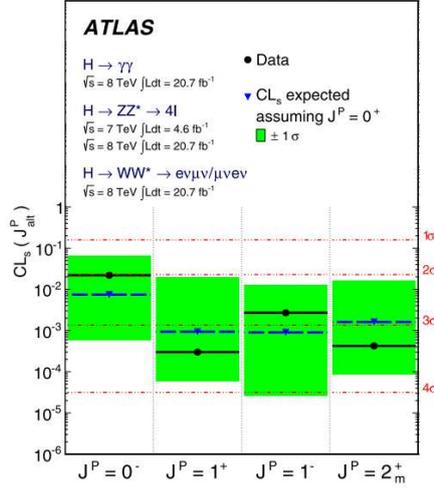


Figure 4.4: Expected (blue triangles/dashed lines) and observed (black circles/solid lines) confidence level  $CL_s$  for alternative spin-parity hypotheses assuming a  $J^P = 0^+$  signal. The green band represents the 68%  $CL_s(J^P_{\text{alt}})$  expected exclusion range for a signal with assumed  $J^P = 0^+$  [62].

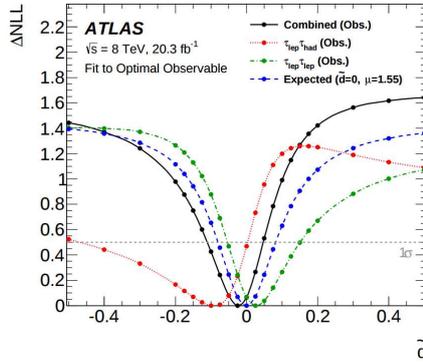


Figure 4.5: The negative log likelihood  $\Delta\text{NLL}$  as a function of the  $CP$  mixing parameter  $\tilde{d}$  for Run 1 from the Optimal Observable based ATLAS study in the  $H \rightarrow \tau\tau$  channel. The best fit value of  $\tilde{d}$  corresponds to the minimum of  $\Delta\text{NLL}$ . The grey dashed line refers to the 68% confidence level [61].

### 4.3 Measurements in Run 2

#### 4.3.1 Total cross section

In 2015 the LHC has started with the second data taking period, Run 2, at  $\sqrt{s} = 13$  TeV. The combination of the Run 1 and Run 2 measurements allows to probe the dependence of the total cross section of the Higgs boson on the centre-of-mass energy  $\sqrt{s}$ . The results for the  $H \rightarrow ZZ^* \rightarrow 4l$  and  $H \rightarrow \gamma\gamma$  decay channel of the ATLAS experiment with centre-of-mass energies of 7, 8 and 13 GeV with a data set of  $4.5 \text{ fb}^{-1}$ ,  $20.3 \text{ fb}^{-1}$ , and  $26.1 \text{ fb}^{-1}$ , respectively, are shown in Figure 4.6. The cross section increase by a factor around two from the Run 1 measurements at 7 and 8 TeV to the Run 2 measurement at 13 TeV in agreement with the SM prediction. Thus the Run 2 measurements can be used to determine further properties of the Higgs boson like the  $CP$ -properties or the coupling to bosons and fermions more precisely.

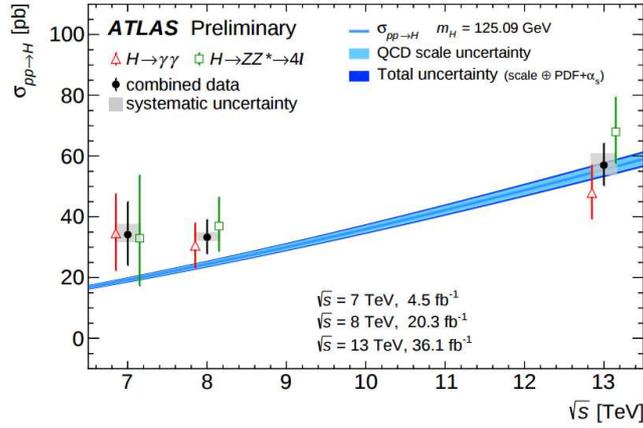


Figure 4.6: Total cross section measurements of the ATLAS detector in the  $H \rightarrow ZZ^* \rightarrow 4l$  and  $H \rightarrow \gamma\gamma$  decay channels. The blue line corresponds to SM prediction at up to  $N^3\text{LO}$  in QCD. Red refers to the  $H \rightarrow \gamma\gamma$ , green to the  $H \rightarrow ZZ^* \rightarrow 4l$  channel and black corresponds to the combined results [63].

#### 4.3.2 Mass

The Higgs mass has been determined to  $m_H = 124.98 \pm 0.19(\text{stat.}) \pm 0.21(\text{sys.}) \text{ GeV}$  by the ATLAS collaboration [64], using the full Run 2 dataset of 2015 and 2016 with an integrated luminosity of  $36 \text{ fb}^{-1}$  of the decay channels  $H \rightarrow ZZ^*$  and  $H \rightarrow \gamma\gamma$  and is consistent with the findings from Run 1.

#### 4.3.3 Signal strength

For a comparison to the Run 1 results the signal strength  $\mu$  is estimated in Run 2, by using the same definitions as in Run 1. The combined result for the  $H \rightarrow ZZ^* \rightarrow 4l$  and  $H \rightarrow \gamma\gamma$

decay channels measured with the ATLAS detector is given by [63]:

$$\mu = 1.09 \pm 0.09(\text{stat.})_{-0.05}^{+0.06}(\text{exp.})_{-0.05}^{+0.06}(\text{theo.}). \quad (4.3)$$

This result is perfectly in agreement with the SM prediction ( $\mu = 1$ ) and is more precise than the ATLAS Run 1 combined results of all decay channels.

#### 4.4 Simplified Template Cross Section

The measurement of Simplified Template Cross Sections (STXS) is a new strategy to study the Higgs boson at the LHC and is defined in [21]. The measurements of the signal strength  $\mu$  and coupling strength ( $\kappa$ ) used in Run 1 make use of several assumptions, like the kinematics predicted by the SM or an extrapolation from the measured phase space to the global phase space. These assumptions introduce theoretical uncertainties on the determined results and dependences on the underlying physics model. Furthermore the coupling strength and the signal strength are both normalized to the SM prediction ( $\kappa = g_{\text{obs}}/g_{\text{SM}}$ ,  $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$ ). The STXS technique allows in a systematic way the reduction of theory dependences, which are directly folded into the measurements. Here both dependencies, the one on theoretical uncertainties in the SM predictions and the one on the underlying physics model, are included. Cross section predictions or coupling strength predictions are not necessary for the STXS.

It provides more finely-grained measurements, which will benefit from the global combination of the measurements in all decay channels and the higher cross section of the Higgs boson in Run 2. The goal of this technique is to maximize the sensitivity, while the dependence on theory is minimized. In particular this means using a combination of all decay channels and instead of measuring the signal strength the individual cross sections of exclusive regions of phase space for specific production modes, from now on called STXS bins, are measured. The definitions of these STXS bins are motivated by minimizing the dependence on theoretical uncertainties, which has to be directly folded in for coupling strength measurements. Furthermore these bins are chosen in a way to maximize the experimental sensitivity and to isolate possible effects from physics beyond the SM.

The STXS approach allows to use advanced analysis techniques such as event categorisations or multivariate techniques.

##### STXS definitions

Figure 4.7 gives a schematic overview of the STXS. On the left-hand side the different analyses are given, which are very similar to the ones used in the coupling strength measurements. Each of these analyses define signal categories enriched with specific Higgs boson production processes and to improve the signal-to-background ratio or the invariant-mass resolution. The signal categories can be additionally optimized for the sensitivity in STXS measurements.

The centre of Figure 4.7 illustrates the part of the STXS method which is determined by a global fit of the number of events in the experimental categories that combines all decay channels and is the main result of the STXS. The results are cross sections per production mode, split into exclusive kinematic bins on truth level for each of the dominant production modes. An example for these STXS bins is the cross section of the gluon fusion production

mode of the Higgs boson with one jet in the final states and a transverse momentum of the Higgs boson greater than 200 GeV.

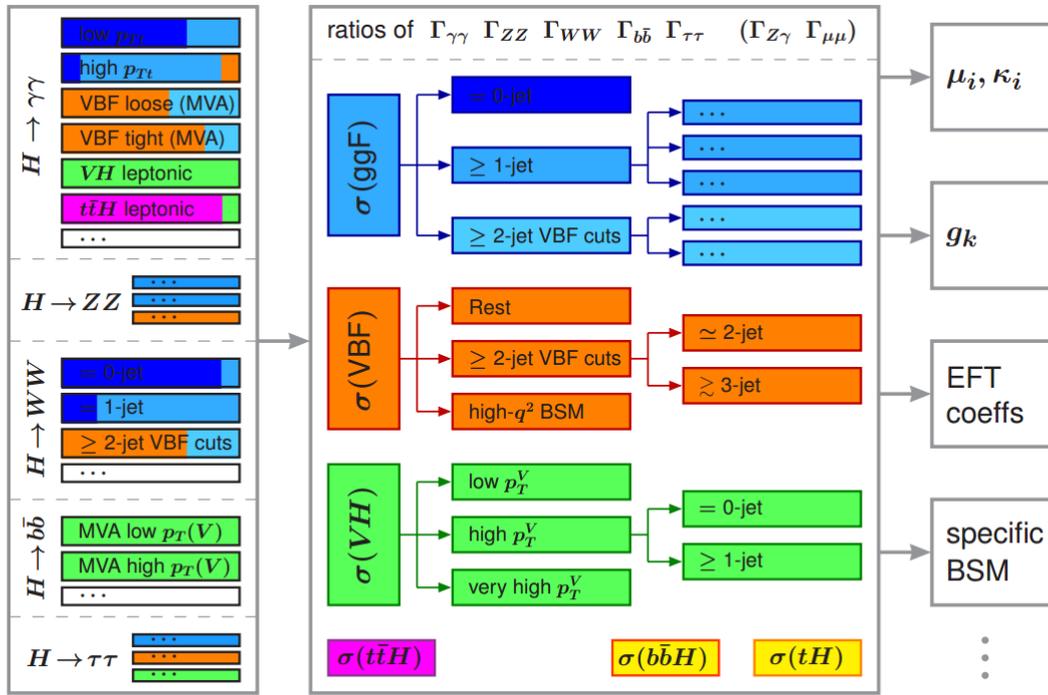


Figure 4.7: Simplified Template Cross Section framework, including the input from different analyses channel with their signal categories (left), the measured cross sections as the main result of the STXS (mid) and a list of possible interpretations based on the STXS results (right) [21].

These STXS measurements are used as input for subsequent interpretations, which are shown in Figure 4.7 on the right-hand side. Possible interpretations are Wilson coefficients in the effective field theory, tests of specific models beyond the SM, determination of signal strength modifiers or coupling scale factors  $\kappa$ , and so forth. The last two of these examples provide compatibility with earlier measurements. The experimental results should quote the full covariance matrix among the different STXS bins to minimize the theory dependence which is folded into the determination of the STXS from event categories. Instead this dependence is shifted into the interpretation step to ensure a long-term usability of the measurements. The STXS bins have some similarity to differential cross section measurements, but combine the advantages of signal strength, fiducial and differential measurements. They are complementary to full-fledged fiducial and differential measurements, but they cannot replace these. However the full fiducial or differential measurements will be only available in a subset of decay channel with the Run 2 measurements.

#### 4.4.1 Object definitions

The global fit unfolds the measured number of events in the event categories on reconstruction level in all decay channels to the STXS bins. Thus the truth final state particles need to be defined unambiguously for the measured bins and theoretical predictions from either analytic calculations or simulations. In contrast to the definitions of the fiducial cross sections, the definitions of the final state particles for the STXS are more idealized and simpler. The Higgs boson is treated as a final state particle to allow the combination of the different decay channels.

##### Higgs boson

The STXS approach requires an on-shell Higgs boson. A cut on the rapidity of Higgs boson  $y_H$  with  $|y_H| < 2.5$  is defined, since current measurements are not sensitive beyond this rapidity range.

##### Leptons

Electrons and muons from decays of signal vector bosons are required to be "dressed". This means that for example the momenta of final state radiated photons are added back to the momenta of the electron or muon.  $\tau$ -leptons are defined from the sum of their decay products. There are no restrictions on the transverse momentum or the pseudorapidity of the leptons, so that for a leptonically decaying weak vector boson the full phase space is included.

##### Jets

The STXS truth jets are defined as anti- $k_r$  jets with a jet radius of  $R=0.4$  and are built from all stable particles, including neutrinos, photons and leptons from hadron decays or produced in the shower. For a stable particle a lifetime greater than 10 ps is required, corresponding to the lifetime required in the experimental simulation for the particle to be passed to GEANT 4. Decay products of Higgs bosons are not considered, since the Higgs boson is defined as a final state particle. Leptons and neutrinos from decays of the signal weak vector boson are removed, since they are treated separately, while decay products from hadronically decaying signal weak vector bosons are included in the inputs to the truth jet finding. The truth jets are defined without a cut on their pseudorapidity, but a cut on the transverse momentum of the jets  $p_T$  is set with  $p_T > 30$  GeV, to limit the amount of phase-space extrapolation in the measurements, since the reconstruction level signal categories always include jet  $p_T$  requirements to reduce effects from pile-up and underlying events.

#### 4.4.2 Splitting of production modes

The definitions of production modes are slightly different to those introduced in Section 2.2.1. The VH production mode is explicitly defined as Higgs bosons production in association with a leptonically decaying vector boson V, while electroweak  $q\bar{q} \rightarrow H$  production comprises VBF Higgs boson production and VH production with hadronically decaying V and will be called VBF STXS bin from now on. Similarly the  $gg \rightarrow ZH$  production mode is included in the gluon fusion production. This is done due to the fact that the distinction between

$q\bar{q} \rightarrow VH$  and VBF processes and similarly between  $gg \rightarrow VH$  and gluon fusion production becomes ambiguous at higher order if the vector boson  $V$  decay hadronically.

#### 4.4.3 Stage 0

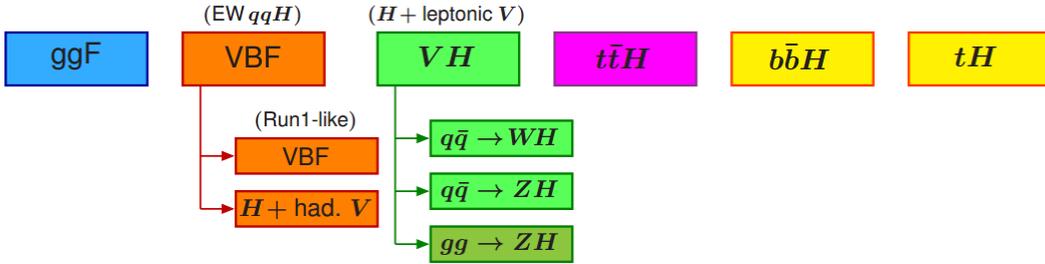


Figure 4.8: Simplified Template Cross Section Stage 0 definitions. All STXS bins in the same color correspond to the same main production mode. The VBF and VH STXS bins are further split into finer STXS bins [21].

A set of STXS bins satisfying the requirements of Section 4.4 for every analysis are difficult to define. For some analyses only a subset of the STXS bins or a sum of a set of STXS bins will be accessible. The number of sensitive STXS bins in the measurements will increase with an increasing amount of available data. The STXS is defined in several stages with an increasing number of STXS bins to account for this. The Stage 0 definition of the STXS bins is very similar to the production modes defined in Section 2.2.1. Figure 4.8 gives an overview of the different STXS bins. Each main production mode has a single inclusive STXS bin at this stage, with the associated Higgs boson production split into  $q\bar{q} \rightarrow WH$ ,  $q\bar{q} \rightarrow ZH$  and  $gg \rightarrow ZH$ . For a better comparison to the Run 1 analyses the VBF STXS bin (EW  $qqH$ ) is split into a Run1-like VBF and into a Run1-like  $V(\rightarrow jj)H$  STXS bin. For this the splitting is defined by the conventional Feynman diagrams included in the simulations. The categories are split into a Higgs boson forward category with  $|y_H| \geq 2.5$  and into a Higgs boson central category with  $|y_H| < 2.5$ .

#### 4.4.4 Stage 1

At Stage 1 the production mode STXS bins at Stage 0 are split for example into the jet multiplicity or the transverse momentum of the Higgs boson of the leading jet. At this Stage analyses are able to use a specific STXS bin merging to reach more statistics in some of the STXS bins. The merged STXS bins should have a similar acceptance, such that the individual STXS bins can still be determined in an unbiased way in a global combination of all channels.

##### Gluon fusion STXS bins at Stage 1

The Stage 0 inclusive gluon fusion STXS bin with a rapidity of the Higgs boson  $|y_H| < 2.5$

is redefined in a finer way at Stage 1. A schematic overview of these STXS bins is given in Figure 4.9.

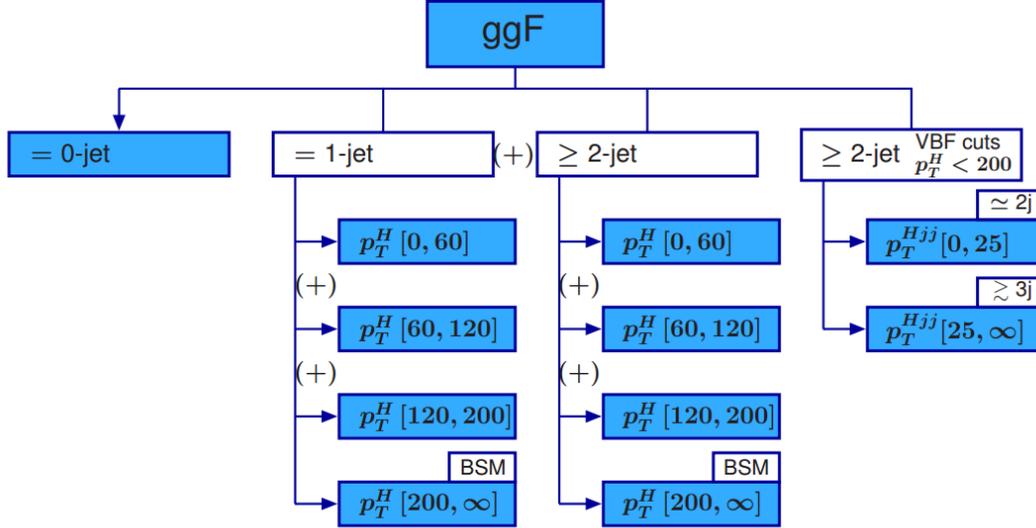


Figure 4.9: Stage 1 STXS bins for the gluon fusion production mode. The blue filled STXS bins refer to measured STXS bins, while the white ones are only given to group the measured STXS bins. STXS bins marked with a "(+)" are suggested to be merged if the statistics is too low [21].

The gluon fusion production mode is split according to the jet multiplicity  $N_j$  into STXS bins for  $N_j = 0$ ,  $N_j = 1$ ,  $N_j \geq 2$ , and  $N_j \geq 2$  with cuts for a VBF topology (defined in the same way as the corresponding STXS bin in the VBF production) and a cut on the transverse momentum of the Higgs boson  $p_T^H < 200$  GeV is required, which gives a priority to the  $p_T^H > 200$  GeV bin for  $N_j \geq 2$  without the VBF topology. These STXS bins are motivated by the use of jet bins in the experimental analyses and avoids transferring the associated theoretical uncertainties into the measurement. The separation of the VBF topology is done to measure the VBF contamination in the gluon fusion production measurements.

The VBF topology STXS bin is split into an exclusive 2-jet-like and an inclusive 3-jet-like STXS bin ( $\geq 3$  jets) by using a cut on  $p_T^{Hjj}$  at 25 GeV to prevent non-trivial theory uncertainties in the gluon fusion contribution. Here  $p_T^{Hjj}$  refers to the absolute value of the vectorial sum of the transverse momenta of the Higgs boson and the two jets with the highest transverse momentum:  $p_T^{Hjj} = |\vec{p}_T^H + \vec{p}_T^{j1} + \vec{p}_T^{j2}|$ .

The STXS bins with  $N_j = 1$  and  $N_j \geq 2$  without the VBF topology cuts are further split according to the transverse momentum of the Higgs boson  $p_T^H$ .

The low transverse momentum STXS bin with  $0 \text{ GeV} < p_T^H < 60$  GeV is the one with the highest sensitivity in all bosonic decay channels. The 60 GeV cut is chosen to give a more even split of events while it is high enough that no resummation effects are expected and that the bias from the transverse momentum of the jet should be negligible.

The STXS bin with  $60 \text{ GeV} < p_T^H < 120 \text{ GeV}$  remaining from a split into a low and a high region of the transverse momentum of the Higgs boson. Here the lower cut is high enough that the STXS bin can be treated as a hard Higgs boson plus jet system in the theoretical description.

A STXS bin with  $120 \text{ GeV} < p_T^H < 200 \text{ GeV}$  is the main contribution of the *Boosted* selection in the  $H \rightarrow \tau^+ \tau^-$  decay channel (see Chapter 8). The separate definition avoids large extrapolations for this contribution and in the case of  $N_j = 2$  it provides a substantial part of the gluon fusion contribution for the associated vector boson production with a hadronically decaying vector boson.

The STXS bin beyond the top quark mass with  $200 \text{ GeV} < p_T^H$  is defined which allow the top quark loop to be on-shell such that the top quark mass effects become relevant. This high transverse mass STXS bin ensures the usability of the heavy top expansion for the lower transverse mass STXS bins. Additionally it offers the possibility to distinguish a point-like gluon fusion vertex induced by heavier beyond the SM particles in the loop from the resolved top quark loop and is the most sensitive STXS bin to beyond the SM effects.

Several STXS bin merging approaches are suggested for the gluon fusion category at Stage 1. The three lower transverse mass STXS bins of the Higgs boson and additional or alternative the  $N_j = 1$  and  $N_j \geq 2$  STXS bins can be merged, while merging between different production modes is theoretically disfavoured. These suggestions can be used if low statistics in a single STXS bin allows no sensitivity.

### Vector boson fusion STXS bins at Stage 1

The VBF STXS bin for the rapidity of the Higgs boson  $|y_H| < 2.5$  is split into finer STXS bins at Stage 1. However at higher order the VBF production and  $VH$  production with a hadronically decaying vector boson become ambiguous. At Stage 1 the VBF category includes the  $VH$  production mode with hadronic vector boson decays and corresponds to the electroweak  $q\bar{q}H$  production.

An overview of the Stage 1 binning for the VBF category is given in Figure 4.10. The VBF STXS bin is split into two different transverse momentum regions of the leading jet  $p_T^{j_1}$  by a cut at  $p_T^{j_1} = 200 \text{ GeV}$ . While the lower  $p_T^{j_1}$  STXS bin is expected to be dominated by SM-like events, the high STXS bin should be sensitive to potential beyond the SM contributions. The cut value is chosen to reduce the contamination of SM-like events in this STXS bin.

For the lower  $p_T^{j_1}$  region a STXS sub-bin with a typical VBF topology with two or more jets is defined with a cut on the invariant mass of the two leading jets  $M_{jj}$  at  $M_{jj} > 400 \text{ GeV}$  and a cut on the difference of the pseudorapidities of the two highest- $p_T$  jets with  $\Delta\eta_{jj} = |\eta_{j_1} - \eta_{j_2}| > 2.8$ . These definitions are chosen as an intermediate compromise between the various VBF selections of the different analyses channels.

Further this STXS bin is split into an exclusive 2-jet-like STXS bin and an inclusive 3-jet-like STXS bin by a cut at  $p_T^{Hjj} = |\vec{p}_T^H + \vec{p}_T^{j_1} + \vec{p}_T^{j_2}| = 25 \text{ GeV}$  as a compromise between the different kinematic variables, which are applied in different analyses to enrich the vector boson fusion production and as a compromise between providing a good separation of gluon fusion and vector boson fusion and the selections used in the measurements.

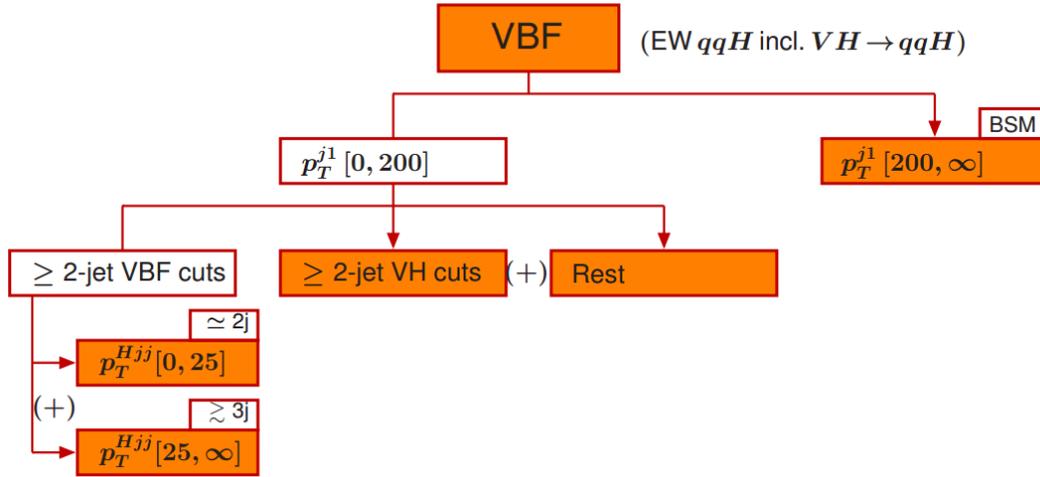


Figure 4.10: Stage 1 STXS bins for the VBF production mode. The orange filled STXS bins refer to measured STXS bins, while the white ones are only given to group the measured STXS bins. STXS bins marked with a "(+)" are suggest to be merged if the statistics is to low [21].

The events not passing the  $\geq 2$ -jet VBF cuts are divided into a VH like STXS bin and a rest STXS bin. The VH STXS bin contains events with the typical topology of associated vector boson production with a hadronically decaying vector boson, which is defined by requiring two or more jets with a di-jet mass  $M_{jj}$  between 60 GeV and 120 GeV. The rest STXS bin contains all events, which fail the requirements of the other STXS bins, including events with a jet multiplicity of zero. The rest STXS bin could be sensitive to certain beyond the SM contributions which do not follow the typical SM VBF signature of two forward jets.

### Associated vector boson production STXS bins at Stage 1

The associated vector boson production  $VH$  with a leptonic decay of the vector bosons STXS bin for the Higgs boson rapidity  $|y_H| < 2.5$  at Stage 0 is split into finer STXS bins at Stage 1 as illustrated in Figure 4.11.

The  $VH$  production is split according to the production via a quark anti-quark pair  $q\bar{q}$  or a gluon fusion  $ggF$  initial state. This split becomes ambiguous at higher order, but on the experimental side this split can be defined according to the two different simulated event samples, which are used in the analyses.

The  $q\bar{q}$  production is split into a STXS bin for  $W \rightarrow l\nu$  and one for  $Z \rightarrow ll$  plus  $Z \rightarrow \nu\bar{\nu}$ . These are further split according to the transverse momentum of the vector boson  $p_T^V$ . The cuts on  $p_T^V$  are oriented to the quantity which is used in the  $H \rightarrow b\bar{b}$  analysis, since this analysis is one of the dominant ones for the  $VH$  STXS bins.

The STXS bin with  $p_T^V < 150$  GeV is defined for events of bosonic decay channels and events from  $H \rightarrow b\bar{b}$  with  $W \rightarrow l\nu$  or  $Z \rightarrow ll$  which do not rely on the missing transverse energy  $E_T^{miss}$  triggers.

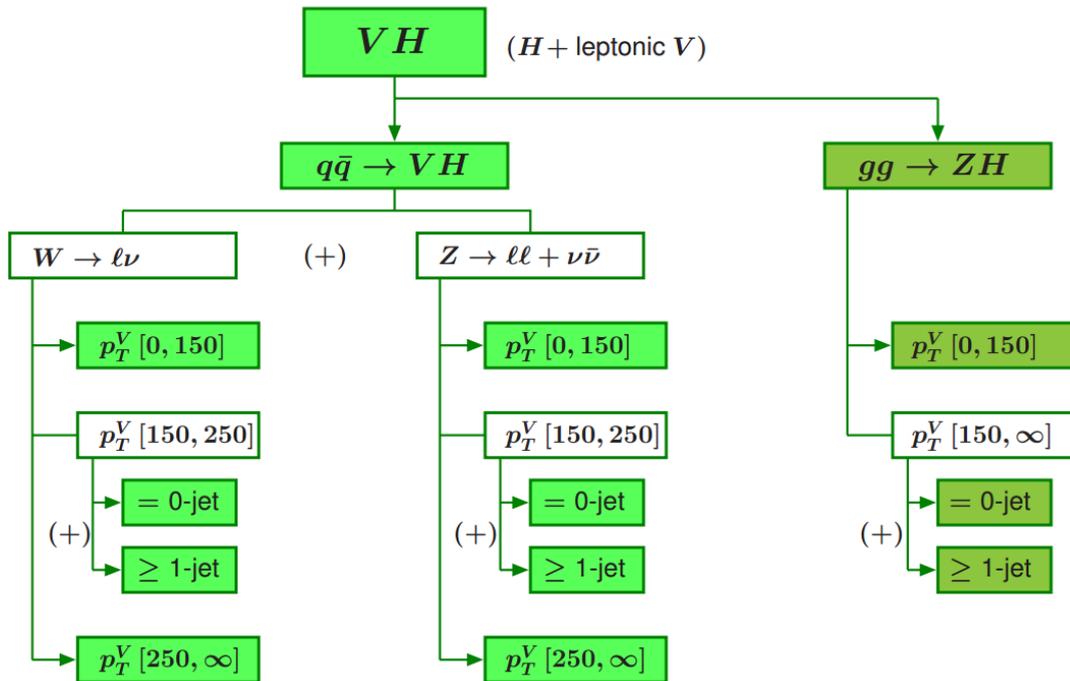


Figure 4.11: Stage 1 STXS bins for the associated vector boson production mode with a leptonic decay. The green filled STXS bins refer to measured STXS bins, while the white ones are only given to group the measured STXS bins. STXS bins marked with a "(+)" suggest to be merged if the statistics is too low [21].

The STXS bin with  $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$  has a dominant contribution of events from  $H \rightarrow b\bar{b}$  with  $Z \rightarrow \nu\bar{\nu}$  due to the high threshold of the missing transverse mass trigger and events from  $H \rightarrow b\bar{b}$  with  $W \rightarrow l\nu$  or  $Z \rightarrow \ell\ell$ . Furthermore this mid  $p_T$  STXS bin is split into a STXS bin with zero jets and one with at least one jet, to reflect the different experimental sensitivities and to avoid the corresponding theory dependence.

The STXS bin with  $250 \text{ GeV} < p_T^V$  has the highest sensitivity to beyond the SM contributions. The STXS bin for the production via a gluon-gluon initial state is split in a similar way, except for the high transverse momentum STXS bin of the vector boson, which is not separated.

#### Further production modes

For the  $t\bar{t}H$ , the  $b\bar{b}H$  and the  $tH$  STXS bins at Stage 0 no additional STXS bins are defined at Stage 1.

#### 4.4.5 Results

First preliminary results of STXS measurements have been published by the ATLAS experiment in the decay channels  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  [63]. The combined results of these

channels use a dataset with an integrated luminosity of  $36.1 \text{ fb}^{-1}$  collected in 2015 and 2016. The measured events divided into kinematic and topological categories, as summarized in Table 4.1. The categories are chosen as similar as possible to the Stage 1 STXS bins described in the previous sections.

The gluon fusion production ( $ggF$ ) is split into STXS bins of jet multiplicity, the transverse momentum of the Higgs boson and into two  $VBF$  topological STXS bins. Events initiated by a quark anti-quark pair are split into two  $VBF$  topological STXS bins, an associated vector boson production  $VH$  STXS bin, with a hadronically decaying vector boson and into a STXS bin for the remaining events. Furthermore events produced by  $VH$  processes with a leptonically decaying vector boson are split into  $WH$  and  $ZH$  processes and into STXS bins of the transverse momentum of the Higgs boson  $p_T^H$ .

$H \rightarrow \gamma\gamma$	$H \rightarrow ZZ^* \rightarrow 4l$
$t\bar{t}H + tH$ leptonic (two $tHX$ and one $t\bar{t}H$ categories)	$t\bar{t}H + tH$
$t\bar{t}H + tH$ hadronic (two $tHX$ and four BDT $t\bar{t}H$ categories)	$VH$ leptonic
$VH$ di-lepton	2-jet $VH$
$VH$ one-lepton, $p_T^{l+MET} \geq 150 \text{ GeV}$	2-jet $VBF$ , $p_T^{j_1} \geq 200 \text{ GeV}$
$VH$ one-lepton, $p_T^{l+MET} < 150 \text{ GeV}$	2-jet $VBF$ , $p_T^{j_1} < 200 \text{ GeV}$
$VH$ $E_T^{miss}, E_T^{miss} \geq 150 \text{ GeV}$	1-jet $ggF$ , $p_T^{4l} \geq 120 \text{ GeV}$
$VH$ $E_T^{miss}, E_T^{miss} < 150 \text{ GeV}$	1-jet $ggF$ , $60 \text{ GeV} < p_T^{4l} < 120 \text{ GeV}$
$VH + VBF$ , $p_T^{j_1} \geq 200 \text{ GeV}$	1-jet $ggF$ , $p_T^{4l} < 60 \text{ GeV}$
$VH$ hadronic (BDT tight and loose categories)	0-jet $ggF$
$VBF$ , $p_T^{\gamma jj} \geq 25 \text{ GeV}$ (BDT tight and loose categories)	
$VBF$ , $p_T^{\gamma jj} < 25 \text{ GeV}$ (BDT tight and loose categories)	
$ggF$ 2-jet, $p_T^{\gamma\gamma} \geq 200 \text{ GeV}$	
$ggF$ 2-jet, $120 \text{ GeV} \geq p_T^{\gamma\gamma} < 200 \text{ GeV}$	
$ggF$ 2-jet, $60 \text{ GeV} \geq p_T^{\gamma\gamma} < 120 \text{ GeV}$	
$ggF$ 2-jet, $p_T^{\gamma\gamma} < 60 \text{ GeV}$	
$ggF$ 1-jet, $p_T^{\gamma\gamma} \geq 200 \text{ GeV}$	
$ggF$ 1-jet, $120 \text{ GeV} \geq p_T^{\gamma\gamma} < 200 \text{ GeV}$	
$ggF$ 1-jet, $60 \text{ GeV} \geq p_T^{\gamma\gamma} < 120 \text{ GeV}$	
$ggF$ 1-jet, $p_T^{\gamma\gamma} < 60 \text{ GeV}$	
$ggF$ 0-jet (central and forward categories)	

Table 4.1: The signal event categories on reconstruction level of the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  analyses which enter the combined result at Stage 1 STXS measurements [63].

### Stage 0 results

For the Stage 0 results a combined fit is performed for the cross section of gluon fusion  $ggF$ , vector boson fusion  $VBF$ , associated vector boson production  $VH$  and associated top anti-top quark production  $t\bar{t}H$  for a rapidity of the Higgs boson  $y_H$  with  $|y_H| < 2.5$  and assuming the SM branching fractions. Here  $b\bar{b}H$  processes are included into  $ggF$  and  $tHX$  processes are included into  $t\bar{t}H$ . The results for this measurement are illustrated and compared to the SM predictions in Table 4.2. The cross section for  $ggF$  ( $VH$ ,  $t\bar{t}H$ ) agree with the SM prediction

at  $1\sigma$  ( $2\sigma$ ) level of confidence. The measurements of  $VBF$  results only reach a  $3\sigma$  level of confidence.

Process ( $ y_H  < 2.5$ )	Result [pb]	Uncertainty			SM prediction [pb]	
		Total	Stat.	Exp. Th.		
$ggF$	43.9	+6.2 -6.0	+5.5 -5.4	+2.7 -2.3	$\pm 1.2$	$44.5^{+2.0}_{-3.0}$
$VBF$	7.9	+2.1 -1.8	+1.7 -1.6	+0.8 -0.6	+1.0 -0.7	$3.52^{+0.08}_{-0.07}$
$VH$	0.3	+1.6 -1.4	+1.5 -1.32	$\pm 0.4$	+0.3 -0.2	$1.99^{+0.06}_{-0.05}$
$t\bar{t}H$	0.27	+0.37 -0.32	+0.36 -0.31	+0.06 -0.05	+0.05 -0.02	$0.59^{+0.03}_{-0.05}$

Table 4.2: Results for STXS Stage 0 measurements of the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  analyses [63].

### Stage 1 results

The analyses and dataset can not provide sensitivity to all Stage 1 STXS bins, such that some STXS bins were merged or neglected.

For the gluon fusion category  $ggF$  (see Section 4.4.4) the two  $VBF$  topology STXS bins are merged together with the three STXS bins with a jet multiplicity of two or more and a transverse momentum of the Higgs boson less than 200 GeV. The two beyond the SM  $ggF$  bins with a transverse momentum of the Higgs boson greater than 200 GeV are merged. For the  $VBF$  category (see Section 4.4.4) all STXS bins with a transverse momentum of the leading jet less than 200 GeV and the two beyond the SM sensitive STXS bins from the  $ggF$  and the  $VBF$  category are merged in each other. For the  $VH$  category no Stage 1 STXS bins are applied.

The remaining STXS bins of this measurement are summarized in Table 4.3 and their results are given in Figure 4.12, which show a good agreement for all STXS bins with the SM predictions.

Category	STXS bin
$ggF$	0-jet
$ggF$	1-jet, $p_T^H < 60$ GeV
$ggF$	1-jet, $60 \text{ GeV} < p_T^H < 120$ GeV
$ggF$	1-jet, $120 \text{ GeV} < p_T^H < 200$ GeV
$ggF$	$\geq 2$ -jet, $p_T^H < 200$ GeV
$VBF$	$p_T^{j_1} < 200$ GeV
$ggF+VBF$	$p_T^H > 200$ GeV or $p_T^{j_1} > 200$ GeV
$VH$	$VH$
$t\bar{t}H$	$t\bar{t}H$

Table 4.3: Measured STXS bins at Stage 1 of the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  analyses [63].

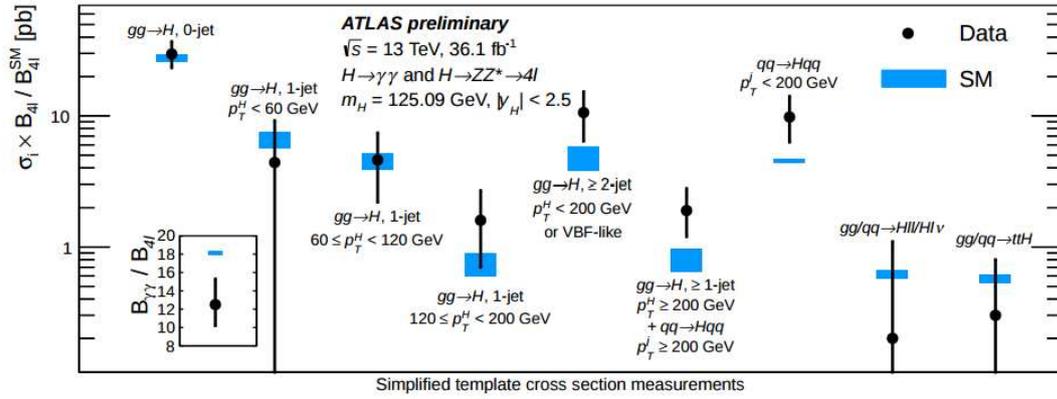


Figure 4.12: Results for the STXS Stage 1 measurement of the combined  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  decay channels [63].

---

## 5 DATA SET AND SIMULATED EVENTS

---

### Data set

The full 2015 and 2016 pp-collision data sets collected with the ATLAS detector are used for this analysis with an integrated luminosity of  $3.21 \text{ fb}^{-1}$  and  $33.26 \text{ fb}^{-1}$  [53]. For the modelling of signal and background processes simulated event samples are used, which are produced with the ATLAS simulation infrastructure [15] as a part of the ATLAS "mc15" production campaign.

### Trigger

The used trigger has changed between the 2015 and 2016 data-taking periods. For both years single lepton triggers for an electron or a muon are applied. Table 5.1 summarizes the triggers, separated into the data set years and the lepton flavour. In the trigger names "HLT" refer to the software based high level trigger and "L1" to the hardware-based first level trigger described in Section 3.2.5. The second part of the name for example "e26" or "mu26", corresponds to transverse momentum threshold on the light lepton (" $e(\mu)26$ "= $p_T^{\text{electron}(\text{muon})} > 26 \text{ GeV}$ ), which is set at the HLT, while "L1EM20VH" and "L1MU15" corresponds to transverse momentum threshold on the light lepton at L1. "lhloose", "lhmedium" and "lhtight" are likelihood identification criteria, which are set on the light lepton (see Chapter 6). The abbreviation "nod0" describes that no requirement is applied on distance of the observed tracks to the primary vertex.

The threshold on the transverse momentum on the light lepton are increased from the 2015 to 2016 data set due to the higher luminosity. The different triggers for the same object are combined by an "or" such that the object has to pass at least one of the triggers.

Data-set	Lepton flavour	Trigger
2015	Electron	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium
2016	Electron	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0
2015	Muon	HLT_mu20_iloose_L1MU15 HLT_mu40
2016	Muon	HLT_mu26_ivarmedium HLT_mu50

Table 5.1: Triggers for the 2015 and 2016 data set used in the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel.

### Signal samples

The dominant contributions to Higgs boson production from ggF and VBF are modelled with Powheg [65], interfaced with Pythia8 [66]. The CT10\_AZNLOCATEQ6L1 [67–69] tune is used for parton distribution functions. The other signal samples are produced with different generators. The ttH sample is modelled with McAtNlo [70] interfaced with Herwig [71] for

the parton shower and C6L1\_CT10 [67] for the parton distribution function tune. The VH production mode is generated with Pythia8 [66] for the parton shower and A14NNPDF23LO [72] tune is used for the parton distribution function. To model the  $\tau$ -lepton decays TAUOLA [73] is used throughout. All samples are inclusive  $H \rightarrow \tau\tau$  samples, which means that all decay modes of the  $\tau$ -leptons are considered in the event generators. The generators and their cross sections are summarized in Table 5.2.

Process	Generator	Parton distribution function	Cross Section [pb]	Order
Signal				
ggH125	Powheg + Pythia8.212	CT10 + AZNLOCTEQ6L	44.14	NNLO+NNLL QCD + NLO EW
VBF125	Powheg + Pythia8.212	CT10 + AZNLOCTEQ6L	3.78	NNLO QCD + NLO EW
WH125	Pythia8.212	A14NNPDF23LO	1.37	NNLO QCD + NLO EW
ZH125	Pythia8.212	A14NNPDF23LO	0.88	NNLO QCD + NLO EW
ttH125	aMcAtNlo + Herwigpp + EvtGen	UE5 + C6L1 + CT10	0.51	NLO QCD + NLO EW
Background				
$Z \rightarrow \mu\mu$	Sherpa 2.2.1	NNPDF30NNLO	$1.58 \cdot 10^5$	NNLO QCD
$Z \rightarrow \mu\mu$ EW	Sherpa 2.2.1	NNPDF30NNLO	22.35	NLO QCD
$Z \rightarrow ee$	Sherpa 2.2.1	NNPDF30NNLO	$1.58 \cdot 10^5$	NNLO QCD
$Z \rightarrow ee$ EW	Sherpa 2.2.1	NNPDF30NNLO	22.35	NLO
$Z \rightarrow \tau\tau$	Sherpa 2.2.1	NNPDF30NNLO	$1.60 \cdot 10^5$	NNLO QCD
$Z \rightarrow \tau\tau$ EW	Sherpa 2.2.1	NNPDF30NNLO	22.27	NLO QCD
$W \rightarrow \mu\nu$	Sherpa 2.2.1	NNPDF30NNLO	$6.98 \cdot 10^5$	NNLO QCD
$W \rightarrow \mu\nu$ EW	Sherpa 2.2.1	NNPDF30NNLO	235.51	NLO QCD
$W \rightarrow e\nu$	Sherpa 2.2.1	NNPDF30NNLO	$6.98 \cdot 10^5$	NNLO QCD
$W \rightarrow e\nu$ EW	Sherpa 2.2.1	NNPDF30NNLO	236.96	NLO QCD
$W \rightarrow \tau\nu$	Sherpa 2.2.1	NNPDF30NNLO	$7.00 \cdot 10^5$	NNLO QCD
$W \rightarrow \tau\nu$ EW	Sherpa 2.2.1	NNPDF30NNLO	236.92	NLO QCD
Diboson	Sherpa 2.2.1	CT10	$4.86 \cdot 10^3$	NLO QCD
$t\bar{t}$	Powheg + Pythia6.428	CT10 + Perugia 2012	$2.92 \cdot 10^4$	NNLO+NNLL QCD
$tW$	Powheg + Pythia6	CT10 + Perugia 2012	$2.47 \cdot 10^3$	NLO+NNLL QCD
single t (t-channel)	Powheg + Pythia6.428	CT10 + Perugia 2012	$2.51 \cdot 10^3$	NLO+NNLL QCD
single t (s-channel)	Powheg + Pythia6.428	CT10 + Perugia 2012	117.51	NLO+NNLL QCD

Table 5.2: Main signal and background samples for the  $H \rightarrow \tau_{lep} \tau_{had}$  analysis for a combined 2015 and 2016 data set, with their cross sections and the orders of the cross section calculations in perturbation theory.

### Background samples

The considered background processes are the production of vector bosons in association with jets, top quark production and diboson production.

For the simulation of  $Z + jets$  events Sherpa 2.2.1 [74] is used with the NNPDF30NNLO [75] PDF tune. To estimate modelling uncertainties alternative samples are used, which are generated with Madgraph [70] and Pythia8 [66]. The electroweak contribution of  $Z + jets$  is produced separately using Sherpa 2.2.1.

Diboson processes are simulated with Sherpa 2.2.1 for the hard scattering and the parton shower generation. For the PDF the CT10 [67] tune is used.  $t\bar{t}$  and single top quark production events are generated with Powheg [65] and Pythia6 [66] for the parton shower, with a CT10\_Perugia2012 [67, 76] tune for the PDF. An overview of all used samples, their generators and their cross sections can be found in Table 5.2.

Pile-up events produced with Pythia8 [66], overlaid corresponding to the expected pile-up profile, are used to reconstruct the simulated events. During the analysis the simulated events are re-weighted based on their pile-up, to match the observed pile-up profile in the data.

All used samples are passed through the full GEANT4 [15, 77] simulation of the ATLAS detector and have been reconstructed by the same software as used for the data. For the  $H \rightarrow \tau\tau$  decay channel DAOD\_HIGG4D2 derivations are used to produce n-tuples with the xTau-framework.



## 6 OBJECT DEFINITIONS

The topology of the  $H \rightarrow \tau\tau$  decay makes it necessary to identify and reconstruct several objects. Especially the reconstruction of hadronically decaying  $\tau$ -leptons, electrons, muons, jets and missing transverse energy  $E_T^{miss}$  is necessary. This section will describe the definitions of these different objects as used in this analysis.

### Electrons

Electron candidates are reconstructed by matching energy deposits in the electromagnetic calorimeter with tracks in the inner detector. A transverse momentum  $p_T > 15$  GeV and a pseudorapidity  $|\eta| < 2.5$  are further conditions applied to the electrons. Here the transition region between the barrel and end-cap calorimeter ( $1.37 < |\eta| < 1.51$ ) is excluded. Additionally the candidates have to pass a "medium" likelihood-based identification requirement [78] and a gradient isolation criterion [79]. The electron reconstruction and identification is measured using a *tag-and-probe technique* in  $Z \rightarrow ee$  events [79]. In this technique a tight requirement is applied to one *tag* electron, while the efficiency for the second *probe* electron in the event can be measured. The identification efficiency for the 2015 data set is given in Figure 6.1a as a function of the number of primary vertices and in Figure 6.1b the identification and reconstruction efficiency is illustrated as a function of transverse energy  $E_T$ . To correct differences in the identification and reconstruction efficiencies between data and simulated events, a scale factor is derived from the ratio between them and applied to the simulated events.

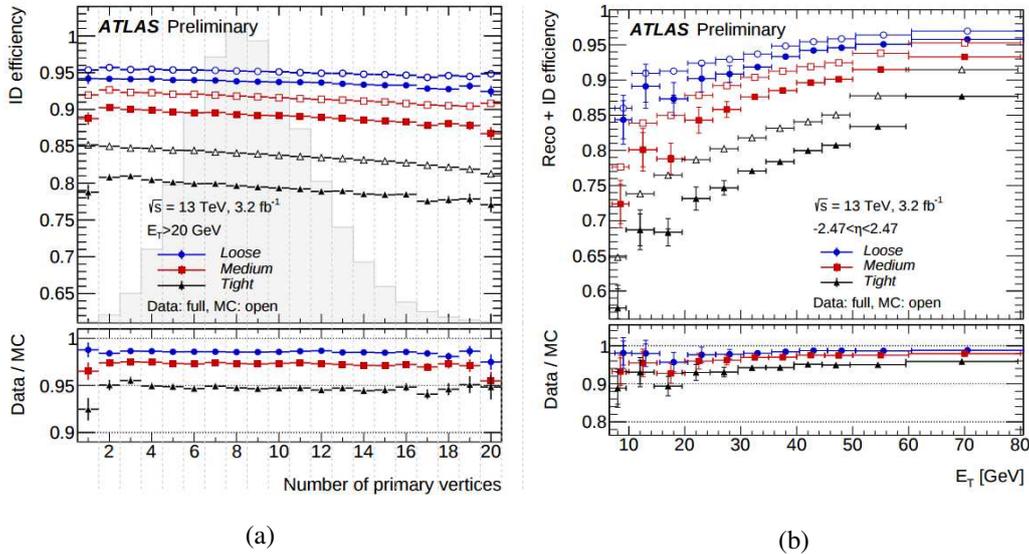


Figure 6.1: Identification efficiency for electrons as a function of the number of primary vertices (left) and identification and reconstruction efficiency for electrons as a function of the transverse energy  $E_T$  (right) with the ATLAS detector using  $Z \rightarrow ee$  events with the 2015 data set [79].

## Muons

Muon candidates are detected by identifying tracks reconstructed in the muon spectrometer, which match a reconstructed track in the inner detector. The transverse momentum  $p_T$  has to exceed 10 GeV and the absolute value of the pseudorapidity  $|\eta|$  is required to be less than 2.5. Also the muon has to satisfy a gradient isolation criterion [80] and a medium identification selection, which is based on requirements concerning the number of hits in the inner detector and the muon system [81]. The efficiency for the muon identification is measured with a *tag-and-probe technique* using  $Z \rightarrow \mu\mu$  events [80]. The efficiency for the 2015 data set is illustrated in Figure 6.2 as a function of the pseudorapidity  $\eta$ . From the ratio between data and simulation a scale factor is derived to correct identification and reconstruction efficiency deviations between them.

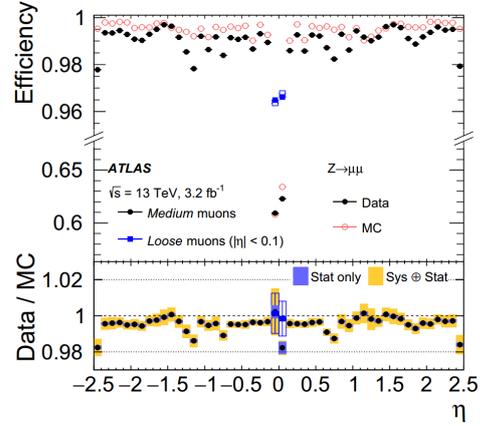
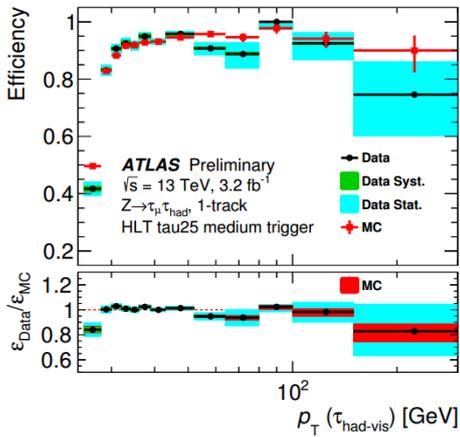
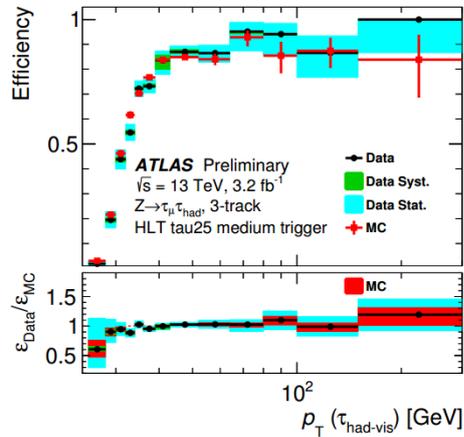


Figure 6.2: Identification efficiency for muons as a function of the pseudorapidity  $\eta$  with the ATLAS detector using  $Z \rightarrow ee$  events with a 2015 data set [80].

## Hadronic $\tau$ -lepton decays



(a)

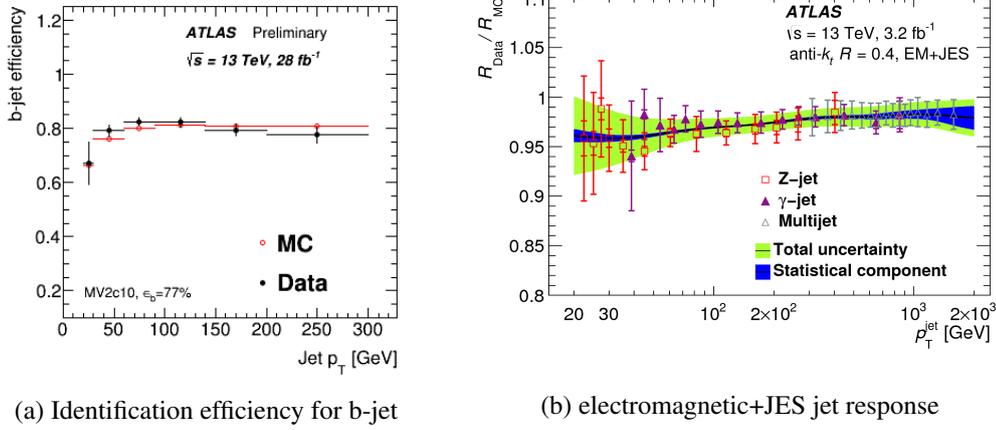


(b)

Figure 6.3: Efficiencies for signal and background subtracted data and the corresponding scale factors ( $\epsilon_{\text{Data}}/\epsilon_{\text{MC}}$ ) for the hadronically decaying  $\tau$ -lepton as a function of the transverse momentum, calculated with a 2015 data set for decays with one track (left) and three tracks (right) [82].

The seed of the hadronic  $\tau$ -lepton reconstruction are jets formed by the anti- $k_t$  algorithm applied on calibrated topological clusters with a distance parameter of  $R=0.4$  [83, 84]. The transverse momentum  $p_T$  has to be greater than 30 GeV and the absolute value of the pseudorapidity  $|\eta|$  less than 2.4. The area between the barrel and end-cap calorimeter ( $1.37 < |\eta| < 1.51$ ) is excluded and an identification selection has to fulfill a medium criteria [82], which is based on a Boosted Decision Tree. The main goal of this is to separate the true visible  $\tau$ -lepton decay products from jets, which are initiated by quarks or gluons. The efficiency of the identification of hadronically decaying  $\tau$ -lepton is measured with a *tag-and-probe technique* in  $Z \rightarrow \tau\tau$  events. The efficiencies and corresponding scale factors for the 2015 data set are illustrated in Figure 6.3.

## Jets



(a) Identification efficiency for b-jet

(b) electromagnetic+JES jet response

Figure 6.4: Identification efficiency for a b-jet as a function of the transverse momentum of the jet  $Jet p_T$  with the ATLAS detector using di-leptonic  $t\bar{t}$  events with a 2016 data set [85] (a) and ratio of the electromagnetic+JES jet response in data to that in the nominal simulated events as a function of the transverse momentum for Z-jet,  $\gamma$ -jet, and multi-jet with a 2015 data set (b)[86].

For the reconstruction of the jets the anti- $k_t$  algorithm with a distance parameter of  $R=0.4$  is used [87, 88]. To calibrate the jet energy the electromagnetic+JES scaling scheme is applied [86]. The data-simulation comparison for the electromagnetic+JES scaling scheme is shown in Figure 6.4b for a 2015 data set. Requirements on the transverse momentum and the pseudorapidity are set:  $p_T > 20$  GeV and  $|\eta| < 4.5$ . JVT is given by a Jet Vertex Tagger algorithm, which is described in [89], and suppresses jets, from pile-up events. For this a  $|JVT| > 0.64$  requirement is set on jets with  $p_T < 50$  GeV and  $|\eta| < 2.4$ . For the forward region  $|\eta| > 2.4$  a fJVT (f for forward) is defined for jets with  $p_T < 50$  GeV, which have to pass  $|fJVT| > 0.4$ . For jets which are initiated by bottom quarks additional requirements are set. *Flavour tagging* algorithms are applied to identify the flavour of hard scattered quarks causing jets. In particular to identify jets initiated by bottom quarks, so called b-jets, the

MV2c20 algorithm is implemented [90]. This algorithm uses the relatively long lifetime of b-flavoured hadrons (1.5 ps), which results in a secondary vertex with a distance of a few mm to the primary vertex. The efficiencies are calculated using  $t\bar{t}$  events with a 2016 data set and the results are given in Figure 6.4a [85].

### Missing transverse energy

The energy and momentum of the initial state partons participating in the hard scattering in a proton-proton collision is unknown along the beam axis. However four-momentum conservation can be used in the transverse plane, with the assumption of negligible transverse momenta of the initial particles. Thus the final state transverse momentum has to be zero. The missing transverse energy  $E_T^{miss}$  refers to the momentum magnitude in the transverse plane, which is necessary to achieve this requirement:

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}. \quad (6.1)$$

The dominant part of  $E_T^{miss}$  corresponds to particles which are not measured in the detector, like the weakly-interacting neutrinos. The leptonic decay of a  $\tau$ -lepton  $\tau \rightarrow l\bar{\nu}_l\nu_\tau$  provides two neutrinos in the final state, such that a significant amount of the  $\tau$ -lepton momenta in the semileptonic di- $\tau$ -lepton decay  $H \rightarrow \tau_{lep}\tau_{had}$  is carried by neutrinos. The missing energy is determined as the negative sum of all measured energy depositions in the calorimeters  $E_{x,y}^{calo}$  corrected by the reconstructed muon energies  $E_{x,y}^{muons}$ :

$$E_{x,y}^{miss} = -E_{x,y}^{calo} - E_{x,y}^{muons}. \quad (6.2)$$

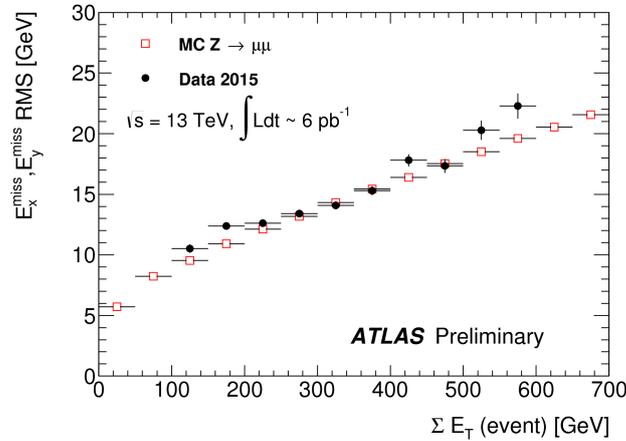


Figure 6.5: Distribution of TST  $E_x^{miss}, E_y^{miss}$  resolution as a function of the sum of the transversal energy  $\Sigma E_T$  with a 2015 data set in  $Z \rightarrow \mu\mu$  events, the black dots corresponds to data and the red squares to simulated events [91].

The energy deposition in the calorimeters includes fully reconstructed and calibrated physical objects. Since the jet reconstruction requires a cut on low the transverse momentum ( $p_T >$

20 GeV) the contribution of soft term from low  $p_T$  particles under this threshold has to be taken into account. To calculate this missing soft term the TrackSoftTerm (TST) algorithm is applied [91]. The TST uses inner-detector tracks originating from the hard-scattering vertex that are not associated to reconstructed objects and add them to the calculation. The performance for the TST is measured using  $Z \rightarrow \mu\mu$  events and is shown in Figure 6.5 for a 2015 data set [91]. The figure illustrates a good agreement between the soft term calculation from data and prediction from simulated events.

### Overlap removal

Since geometric overlap between reconstructed objects creates ambiguity in the identity of the object, an overlap removal has to be made. To decide if this is necessary for two objects, the separation  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2}$  is used. If two objects fail the  $\Delta R$  requirement described below, one object is kept, while the other is removed. For this a priority order is necessary. In this analysis the following order is used: muon, electron,  $\tau$ -lepton and jet. Different  $\Delta R$  thresholds are set for different object combinations:

1. For Muons
  - Jets in  $\Delta R = 0.4$
  - $\tau_{had}$  in  $\Delta R = 0.2$
  - Electrons in  $\Delta R = 0.2$
2. For electrons
  - Jets in  $\Delta R = 0.4$
  - $\tau_{had}$  in  $\Delta R = 0.2$
3. for  $\tau_{had}$ 
  - Jets in  $\Delta R = 0.2$

### Higgs boson candidate kinematic reconstruction

To calculate the invariant di- $\tau$ -lepton mass, the so called missing mass calculator (MMC) is used. This calculator is able to reconstruct the full event topology. The concept behind the MMC is to reconstruct the momentum of each  $\tau$ -lepton pair. For this the MMC uses all known kinematic constraints and performs a scan over the yet undetermined variables. Further information can be found in [92].



## 7 BACKGROUND PROCESSES

Background processes can be categorized into reducible and irreducible processes. Irreducible processes have the same final state signature as the signal process and are therefore difficult to suppress. Reducible backgrounds have a different final state topology than the signal process, but misidentification of physical objects and deficient reconstruction of missing transverse energy in the detector can lead to a classification as a signal process. The simulation of the background processes and the used cross-sections were discussed in Chapter 5. The estimation of their contribution to the selected event yields is described in Section 9.1.

### 7.1 W/Z boson production in association with jets

The production of Z bosons or virtual photons  $\gamma^*$  in association with jets give an important background for the  $H \rightarrow \tau_{had} \tau_{lep}$  decay channel. They can produce irreducible and reducible background events. Final states with one leptonically and one hadronically decaying  $\tau$ -lepton  $Z/\gamma^* \rightarrow \tau_{lep} \tau_{had}$  contribute as an irreducible background. Reducible backgrounds are  $Z/\gamma^* \rightarrow \tau_{had} \tau_{had}$ ,  $Z/\gamma^* \rightarrow \tau_{lep} \tau_{lep}$  and  $Z \rightarrow ll$  events. In the first processes for example the jet produced from a one hadronically decaying  $\tau$ -lepton can be misidentified as a light lepton while for the second and third process for example one light lepton has to be misidentified as a hadronically decaying  $\tau$ -lepton. Furthermore the Z boson can be produced with additional jets, which can be misidentified as  $\tau$ -leptons.

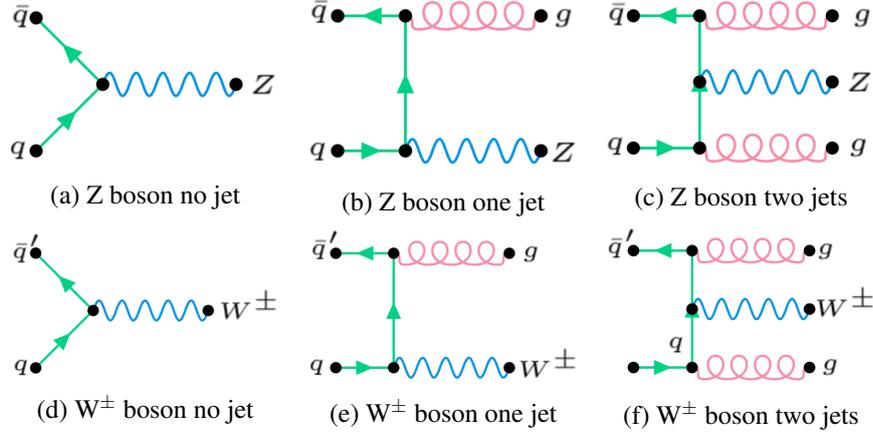


Figure 7.1: Feynman diagrams of vector boson production backgrounds with associated jets, for the Z boson (top) and for the W boson (bottom). From left to right with no associated jet, one associated jet and two associated jets.

Background contributions from  $W^\pm$  boson production in association with jets are reducible. Here mainly a jet gets misidentified as a hadronically decaying  $\tau$ -lepton, while the  $W^\pm$  boson decays leptonically.

Example Feynman diagrams for Z boson and  $W^\pm$  boson production with up to two jets are illustrated in Figure 7.1.

## 7.2 Diboson

All processes with WW-, ZZ- or WZ-diboson production contribute to this background. The W boson and Z boson can decay either leptonically or hadronically. Thus these backgrounds provide final states including leptons and jets. Example Feynman diagrams for different diboson production modes are illustrated in Figure 7.2.

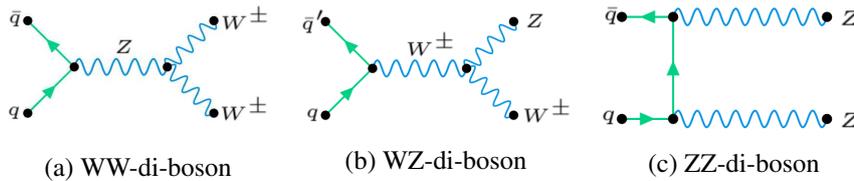


Figure 7.2: Feynman diagrams of vector boson pair production backgrounds, for example

## 7.3 Single top quark and top quark pair production

Single top quarks can be produced in the s-channel, t-channel or in association with a  $W^\pm$  boson. The Feynman diagrams for these productions modes are illustrated in Figure 7.3 (top row). Top quarks decay mostly into a W boson and a bottom quark. Thus further decays of the W boson can produce a similar final state signature as the signal process.

The dominant part of backgrounds involving top quarks are initiated by top quark pair production, with decays into bottom quarks and  $W^\pm$  bosons. Different top quark pair production modes are shown in Figure 7.3.

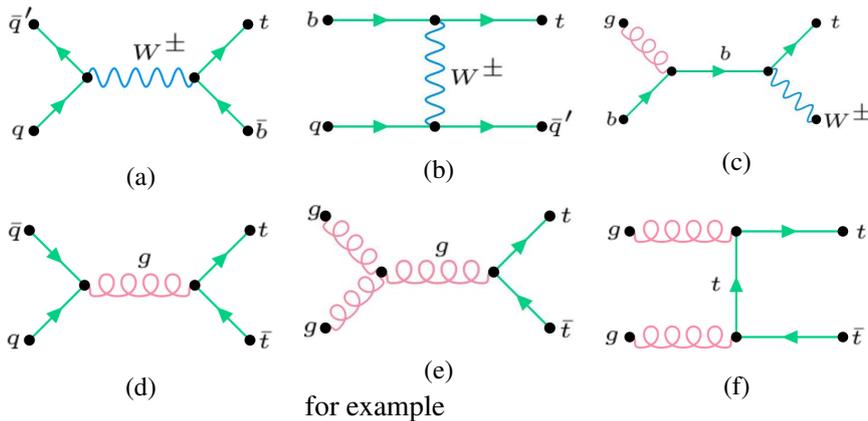


Figure 7.3: Feynman diagrams of top production backgrounds, with single top quark productions modes (top) and top quark pair productions (bottom).

#### 7.4 QCD multi-jet production

Various QCD processes involving outgoing quarks and gluons arise with large production cross sections at the LHC. In these, color-charged partons hadronize and produce multi-jet final states. These jets can be misidentified as  $\tau$ -lepton decays and thus produce a signal-like final state. Examples of Feynman diagrams for these QCD processes are given in Figure 7.4.

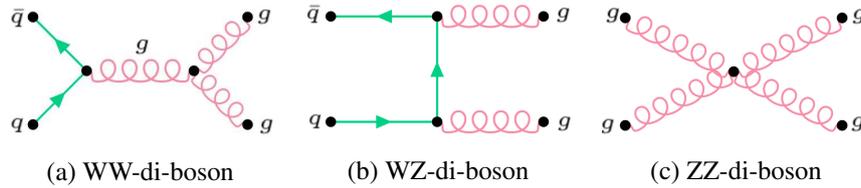


Figure 7.4: Feynman diagrams for QCD multi-jet backgrounds.



## 8 EVENT SELECTION

In this chapter the event selection for the semileptonic  $H \rightarrow \tau_{lep} \tau_{had}$  channel will be described. For the whole study a cut-based event selection is being used. A common *Preselection* is applied to select the  $\tau_{lep} \tau_{had}$  events and to suppress background processes. Furthermore two event categories are defined. The first should enhance VBF signal and is called *VBF* region. The second one is the *Boosted* region on which cuts are applied to increase the ggF signal against other processes. Furthermore both signal event categories are split into two sub-selections: *VBF tight/loose* and *Boosted high/low*. A full overview of all signal categories for example and their selection criteria can be found in Table 8.1.

Region	Requirements	Cut number
<i>Preselection</i>	Exactly one light lepton and one hadronically decaying $\tau$ -lepton	
	Light lepton medium identification and gradient isolation	(1)
	$p_T^{\tau_{had}} > 30$ GeV, $ \eta_{\tau_{had}}  < 2.4$ and $\tau_{had}$ medium identification	(2)
	Opposite charge of the light lepton and the $\tau$ -lepton	
	$ \Delta\eta(\tau_{lep}, \tau_{had})  < 1.5$	(3)
	$\Delta R(\tau_{lep}, \tau_{had}) < 2.5$	(4)
	$E_T^{miss} > 20$ GeV	(5)
	b-jet veto	(6)
	$m_T < 70$ GeV	(7)
$p_T^j > 40$ GeV	(8)	
	$0.1 < x_1 < 1.4$ and $0.1 < x_2 < 1.2$	(8)
<i>VBF</i>	$p_T^j > 30$ GeV	(9 V)
	$ \Delta\eta_{jj}  > 3$	(10 V)
	$m_{jj} > 400$ GeV	(11 V)
	$\eta_{j_1} \times \eta_{j_2} < 0$	(12 V)
<i>VBF tight</i>	Pass <i>VBF</i> requirements	
	$m_{jj} > 500$ GeV	(13 V)
	$p_T^H > 100$ GeV	(14 V)
<i>VBF loose</i>	Pass <i>VBF</i> and fail <i>VBF tight</i> requirements	(15 V)
<i>Boosted</i>	Fail <i>VBF</i> requirement	(9 B)
	$p_T^H > 100$ GeV	(10 B)
<i>Boosted high</i>	Pass <i>Boosted</i> requirements	
	$p_T^H > 140$ GeV	(11 B)
	$\Delta R(\tau_{lep}, \tau_{had}) < 1.5$	(12 B)
<i>Boosted low</i>	Pass <i>Boosted</i> and fail <i>Boosted high</i> requirements	(13 B)

Table 8.1: Signal event categories and their corresponding requirements for the  $H \rightarrow \tau_{lep} \tau_{had}$  cut-based analysis for the combined 2015 and 2016 data set.

### 8.1 Preselection

The basic requirement for the semileptonic  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel is to require exactly one light lepton (electron or muon) and one hadronically decaying  $\tau$ -lepton. Since the Higgs boson has zero charge, additionally the light lepton and the hadronically decaying  $\tau$ -lepton must have opposite electric charge. Further the light lepton and the  $\tau_{had}$ -lepton have to pass their medium identification criteria, the light lepton has to fulfill the gradient isolation criteria, the transverse momentum of the  $\tau_{had}$ -lepton has to be greater than 30 GeV and the pseudorapidity of the  $\tau_{had}$ -lepton has to fulfill  $|\eta_{\tau_{had}}| < 2.4$ . These cuts are applied first (Table 8.1 Cut 1) and they are based on the object definition in Chapter 6. Further cuts on the spatial properties of the two  $\tau$ -leptons are set:  $|\Delta\eta(\tau_{lep}, \tau_{had})| < 1.5$  (Cut 2) and  $\Delta R(\tau_{lep}, \tau_{had}) < 2.5$  (Cut 3). The distributions for these variables before the cuts are shown in Figure 8.1a and 8.1b. A cut on the missing transverse energy ( $E_T^{miss}$ ) with  $E_T^{miss} > 20$  GeV is set (Cut 4) and the distribution for this selection is illustrated in Figure 8.1c. The figures show that this cut removes the lower two bins, which are dominated by  $Z \rightarrow \tau\tau$  and  $Z \rightarrow ll$  events. To reduce background events from  $t\bar{t}$  production a b-jet veto is applied (Cut 5), which is shown in Figure 8.1d. For a two particle system the transverse mass can be calculated by:  $m_T^2 = (E_{T,1} + E_{T,2})^2 - (\vec{p}_{T,1} + \vec{p}_{T,2})^2$ , where  $E_{T,i} = m_i^2 + (\vec{p}_{T,i})^2$  is the transverse energy of the particle  $i$  and  $\vec{p}_{T,i} = (p_{x,i}, p_{y,i})$  the transverse momentum of it. To suppress the W+jets background a cut on transverse mass ( $m_T$ ) as a function of the lepton and missing transverse energy is applied:  $m_T < 70$  GeV (Cut 6 and Figure 8.1e). The transverse momentum of the leading jet  $p_T^{j_1}$  has to exceed 40 GeV, to reduce the W+jets and the  $Z \rightarrow \tau\tau$  background (Cut 7 and Figure 8.1f). The fractions of  $\tau$ -lepton momenta carried by visible decay products are calculated using the collinear approximation [93]. For the collinear mass the neutrino four momentum is reconstructed by assuming that the missing transverse energy  $E_T^{miss}$  is only caused by the neutrinos and that they are emitted in the same direction as the  $\tau$ -leptons. With this definitions the collinear mass is given by:

$$m_{coll} = \frac{m_{vis}}{\sqrt{x_1 x_2}}, \quad (8.1)$$

where  $m_{vis}$  corresponds to the mass of the visible decay products of the two  $\tau$ -leptons and  $x_1, x_2$  are the momentum fractions carried by the visible decay products:

$$\vec{p}_{vis,1(2)} = x_{1(2)} \vec{p}_{\tau,1(2)} \quad (8.2)$$

In the collinear approximation these momentum fractions can be reconstructed by using the missing transverse energy:

$$x_{1(2)} = \frac{P_{vis,2}^x P_{vis,1}^y - P_{vis,2}^y P_{vis,1}^x}{P_{vis,2}^x P_{vis,1}^y + (-)E_T^{miss,x} P_{vis,1(2)}^x - P_{vis,2}^y P_{vis,1}^x - (+)E_T^{miss,y} P_{vis,1(2)}^x} \quad (8.3)$$

To reduce irreducible backgrounds, where the missing transverse energy is not aligned in the direction of the light leptons, a cut on these fractions is set:  $0.1 < x_1 < 1.4$  and  $0.1 < x_2 < 1.2$  (Cut 8). The distribution for these variables for simulated events are given in Figure 8.1g and 8.1h.

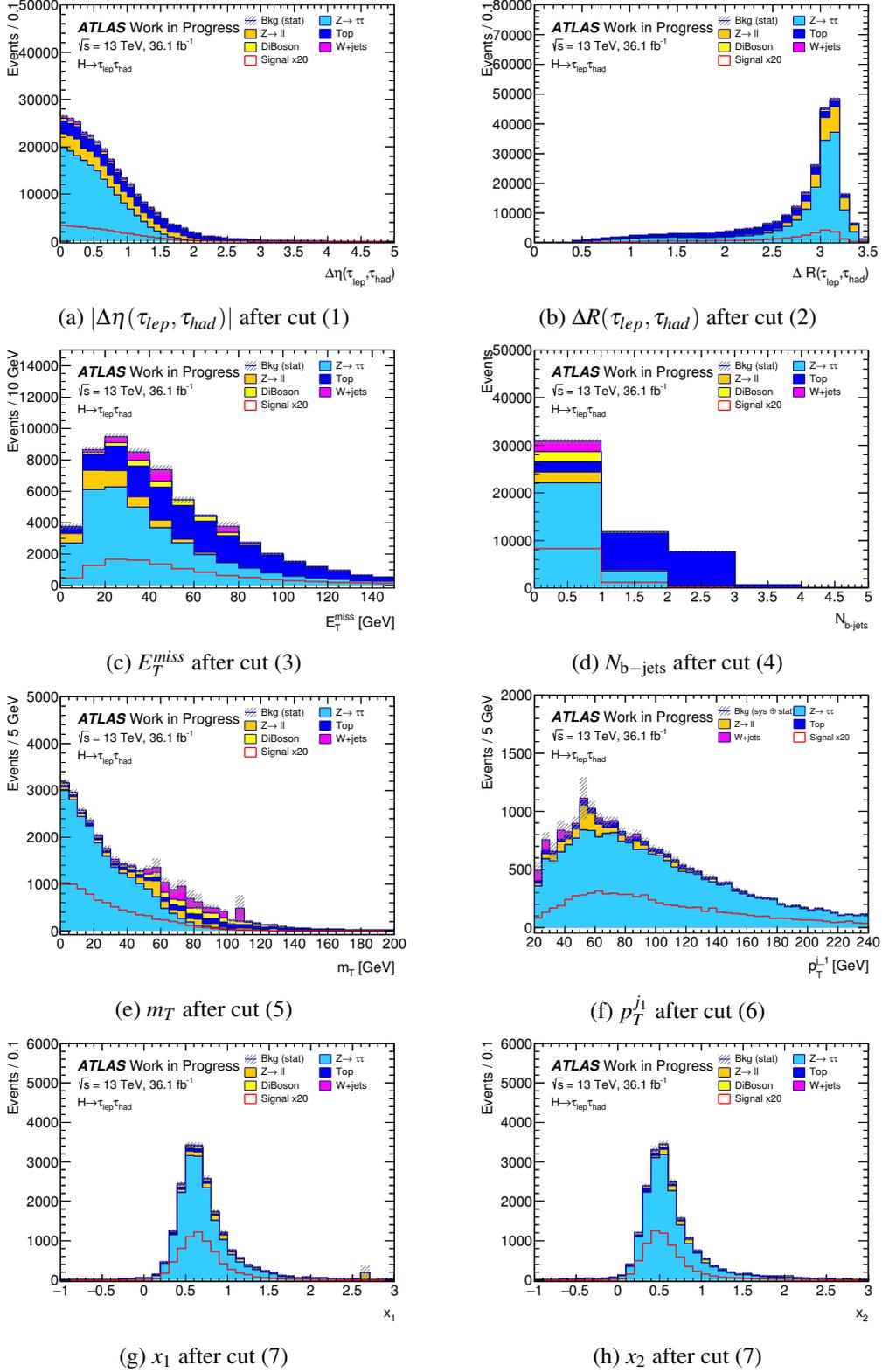


Figure 8.1: Expected number of events for background and signal processes as a function of several variables, which are used in the *Preselection*. The distributions are shown before the corresponding cut on it is applied. The colored histograms are background processes, while the combined signal processes are illustrated with the red line, multiplied by a factor of 20. The uncertainty bands on the background include only statistical uncertainties.

## 8.2 VBF region

For the *VBF* region in addition to the *Preselection*, cuts are applied to enhance the VBF signal. Since the VBF production is accompanied by two additional jets (Figure 2.4b), a requirement on a second jet with a transverse momentum  $p_T > 30$  GeV is set (Cut 9 V and Figure 8.3a). Due to the forward direction of the two jets a cut with  $|\Delta\eta_{jj}| > 3$  (Cut 10 V and Figure 8.3b) is applied. On the invariant mass  $m_{jj}$  of the two jets a cut is applied with  $m_{jj} > 400$  GeV (Cut 11 V and Figure 8.3c) and the jets are required to be in opposite hemispheres ( $\eta_{j_1} \times \eta_{j_2} < 0$ ) (Cut 12 V). The *VBF* region is split into two subregions. The *VBF tight* region is defined with two further cuts:  $m_{jj} > 500$  GeV (Cut 13 V and Figure 8.3d) and  $p_T^H > 100$  GeV (Cut 14 V and Figure 8.3e). The second region, *VBF loose*, is defined by events which fail the tight requirement but pass the *VBF* selection (Cut 15 V).

## 8.3 Boosted region

The *Boosted* region is defined to enhance the gluon fusion process against other processes. For this cuts are applied which use the property that the ggF process can radiate an additional jet (Figure 8.2). The events have to fulfill all *Preselection* requirements but fail the *VBF* requirement (Cut 9 B). Additionally a cut on the transverse momentum of the Higgs boson ( $p_T^H$ ) is set:  $p_T^H > 100$  GeV (Cut 10 B and Figure 8.3f), since Higgs bosons have a high transverse momentum through the recoil against additional jets. Further, the *Boosted* region is split into a high Higgs boson transverse momentum region and a low one. For the high transverse momentum region (*Boosted high*) the cut on the transverse momentum of the Higgs is increased to  $p_T^H > 140$  GeV (Cut 11 B and Figure 8.3g) and due to the high boost a tighter cut on  $\Delta R(\tau_{lep}, \tau_{had})$  is set with  $\Delta R(\tau_{lep}, \tau_{had}) < 1.5$  (Cut 12 B and Figure 8.3h). The (*Boosted low*) category includes all events which pass the *Boosted* but fail the *Boosted high* requirements.

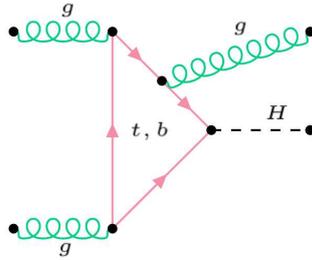


Figure 8.2: Example of a Feynman graph for a boosted ggF event.

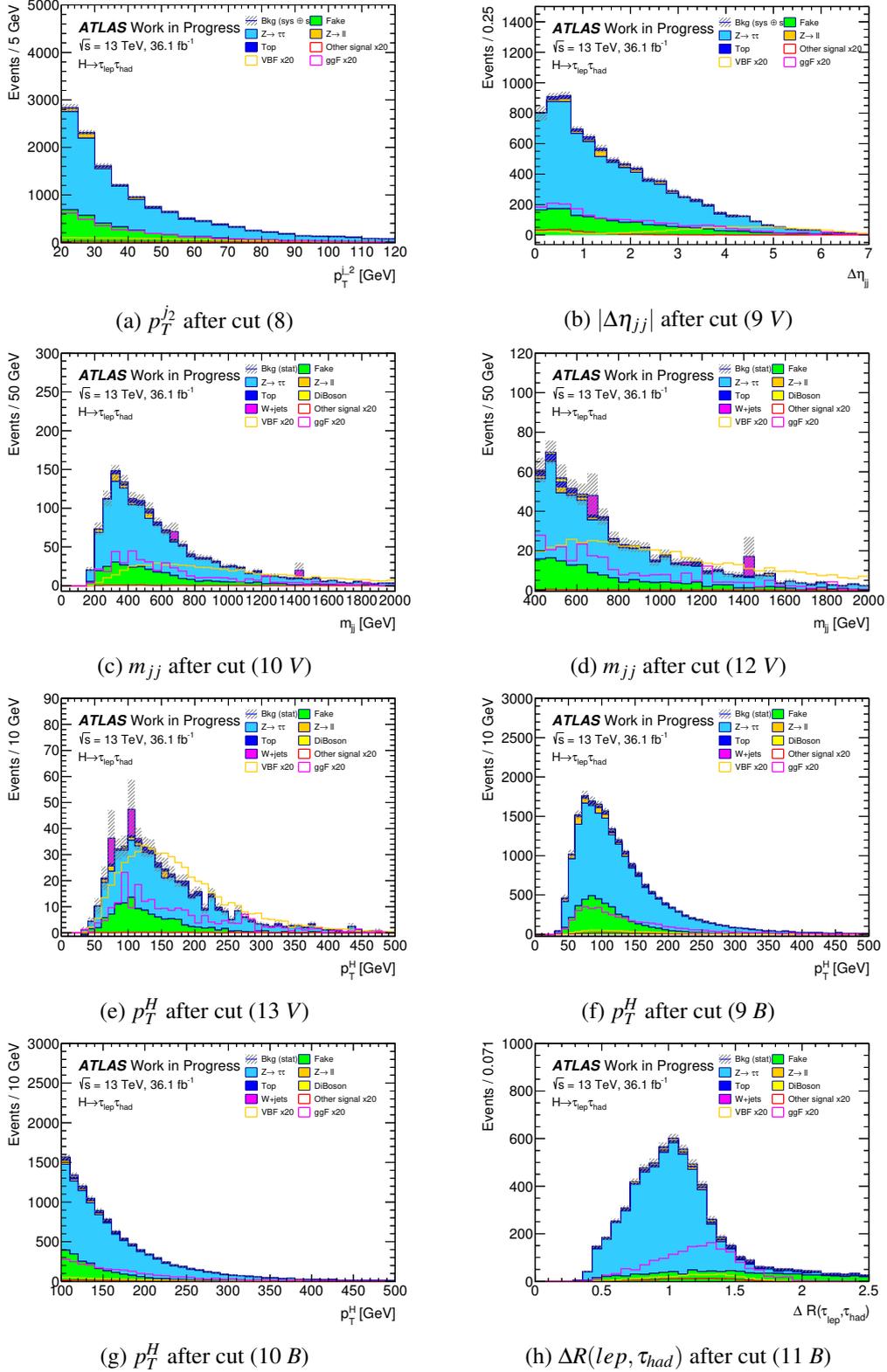


Figure 8.3: Expected number of events for background and signal processes in observables, which are used for the selection of *VBF* (a-e) or *Boosted* (f-h) event category. The distributions are shown before the corresponding cut on it is applied. The colored histograms are backgrounds processes and their uncertainty bands include only statistical uncertainties. The events of *VBF* processes (yellow), *ggF* processes (magenta) and other signal process (red) are shown by lines and are multiplied by a factor of 20.

#### 8.4 Event yields

Table 8.2 shows the event yields for data, all signal processes, the ggF and VBF processes and the combined background events at different selection levels. The last three columns of the table illustrate the proportion of all signal processes, the ggF and VBF processes. The *VBF* selection enhance as expected the VBF signal against the *Preselection*, but also the ggF contribution.

Category	Data	Signal			Background	Proportion (%) of		
		all	ggF	VBF		Signal	ggF	VBF
Preselection	20638	299±3	221±3	65±0.4	20282±157	1.45	1.02	0.30
VBF	729	43±1	14±1	30±0.3	643±18	6.27	1.92	4.12
VBF tight	445	32±1	9±1	23±0.2	384±13	7.67	2.02	5.17
VBF loose	284	12±1	5±1	7±0.1	258±12	4.44	1.75	2.46
Boosted	11615	163±2	129±3	25±0.4	11338±87	1.42	1.07	0.21
Boosted high	5606	88±2	69±2	13±0.2	5524±55	1.57	1.18	0.22
Boosted low	6009	75±2	60±2	12±0.2	5813±67	1.27	0.96	0.19

Table 8.2: Events yields of data, signal and background for the different event selections of the  $H \rightarrow \tau_{lep} \tau_{had}$  analysis with a combined 2015 and 2016 data set of  $36 \text{ fb}^{-1}$ . Only statistical uncertainties are shown.

---

## 9 SIGNAL STRENGTH MEASUREMENT

---

This chapter discuss the background estimation and modelling for the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel. Furthermore the used fit model and the results for the signal strength measurement of the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel is presented.

### 9.1 Background estimation

The processes explained in Chapter 5 can be classified into five main backgrounds. These are  $Z \rightarrow \tau\tau$ ,  $Z \rightarrow ll$ , top quark production processes, di-boson processes and "fakes". "Fakes" corresponds to events where a jet is misidentified as a hadronically decaying  $\tau$ -lepton. The "fake" background is initiated by different processes and each of these processes has a different probability to "fake" a hadronically decaying  $\tau$ -lepton. The estimation of this background is done in a data-driven way and described in Chapter 10 in detail. The remaining backgrounds are determined with simulated events with data-driven corrections for the triggering, reconstruction and identification efficiency. The expected event yields for all backgrounds in the different signal categories are given in Table 9.1. The dominant contribution in all categories are  $Z \rightarrow \tau\tau$  and "fake" events.

Category	Fake	$Z \rightarrow \tau\tau$	$Z \rightarrow ll$	Top	Di-boson	Sum
Preselection	4523±36	14682±119	557±95	243±9	277±7	20282±157
Boosted	2123±21	8640±79	216±28	161±8	197±6	11338±87
Boosted high	505±11	4793±53	76±9	49±5	101±4	5524±55
Boosted low	1618±18	3847±59	141±27	112±6	97±4	5813±67
VBF	150±6	451±16	14±4	16±2	12±1	643±18
VBF tight	80±5	277±11	8±3	11±2	9±1	384±13
VBF loose	70±4	173±12	6±2	6±1	3±1	258±12

Table 9.1: Events yields of the different backgrounds for the event selections of the  $H \rightarrow \tau_{lep}\tau_{had}$  analysis.

The relative contribution of the "fake" background to the full background is reduced from the *Preselection* (22%) to the inclusive *Boosted* event selection (19%). While the relative contribution is further reduced in the *Boosted high* event selection to 9% is the relative contribution in the *Boosted low* event selection 28%. Thus the tighter event selection reduce the reducible "fake" background. The relative contribution of the  $Z \rightarrow \tau\tau$  background is 72% in the *Preselection*. In contrast to the "fake" background the relative contribution of this background increases in the *Boosted high* event selection to 86%, while the relative contribution is 66% in the *Boosted low* event selection. The difference of the relative contributions of both backgrounds from the *Preselection* to the *VBF* event selections is smaller. In the inclusive *VBF* event selection the relative contributions are 23% and 70% for the "fake" background and the  $Z \rightarrow \tau\tau$  background.

For events with top quarks control regions are defined by inverting the bottom quark veto of

the signal categories and leaving all other requirements unchanged. These control regions are enriched in top quark initiated events.

Table 9.2 summarizes the event yields for different processes in these regions. The table shows that in each region top quark initiated events correspond to 50% up to 62% of all events. Additionally a large part of the fake background is produced by top quark events in these regions, such that the top quark initiated contribution in this region is even higher. These regions are used to obtain the normalisation of the top-quark production background in a data-driven way. They are therefore included in the final fit with the top-quark normalisation as a nuisance parameter, as discussed in the next section. This assumes that the full missmodelling in these regions is introduced through the top background, which is the dominant one in these regions.

Category	Data	Signal	Top	Fake	Other	All	Top fraction
Preselection	5001	$15 \pm 0.7$	$2632 \pm 34$	$1698 \pm 22$	$562 \pm 29$	$4906 \pm 50$	54%
Boosted	3426	$10 \pm 0.6$	$1932 \pm 29$	$1024 \pm 15$	$371 \pm 17$	$3337 \pm 37$	58%
Boosted high	643	$4 \pm 0.3$	$218 \pm 10$	$174 \pm 6$	$184 \pm 11$	$579 \pm 16$	38%
Boosted low	2783	$6 \pm 0.4$	$1714 \pm 27$	$850 \pm 14$	$187 \pm 13$	$2757 \pm 33$	62%
VBF	216	$2 \pm 0.2$	$96 \pm 6$	$45 \pm 3$	$32 \pm 9$	$175 \pm 11$	55%
VBF tight	145	$1 \pm 0.1$	$66 \pm 5$	$28 \pm 2$	$22 \pm 8$	$117 \pm 10$	57%
VBF loose	71	$1 \pm 0.1$	$29 \pm 4$	$18 \pm 2$	$10 \pm 3$	$58 \pm 5$	50%

Table 9.2: Events yields of the control regions for top quark production processes. The measured data, the expected signal events, the top quark events and the "fake" background events are shown in the first four columns, while the column other includes the remaining backgrounds and the column "all" is the sum of all predicted events. The last row shows the proportion of top quark events.

Figure 9.1 compares the data and expected event yields in  $m_{\tau\tau}^{MMC}$  for the *Boosted* categories (top) and *VBF* categories (bottom). A good agreement in all categories is observed in the background dominated low (<100 GeV) and high mass (>150 GeV) regions. Thus the "fake-factor" method and the other background estimations give a good description of the data in background dominated regions.

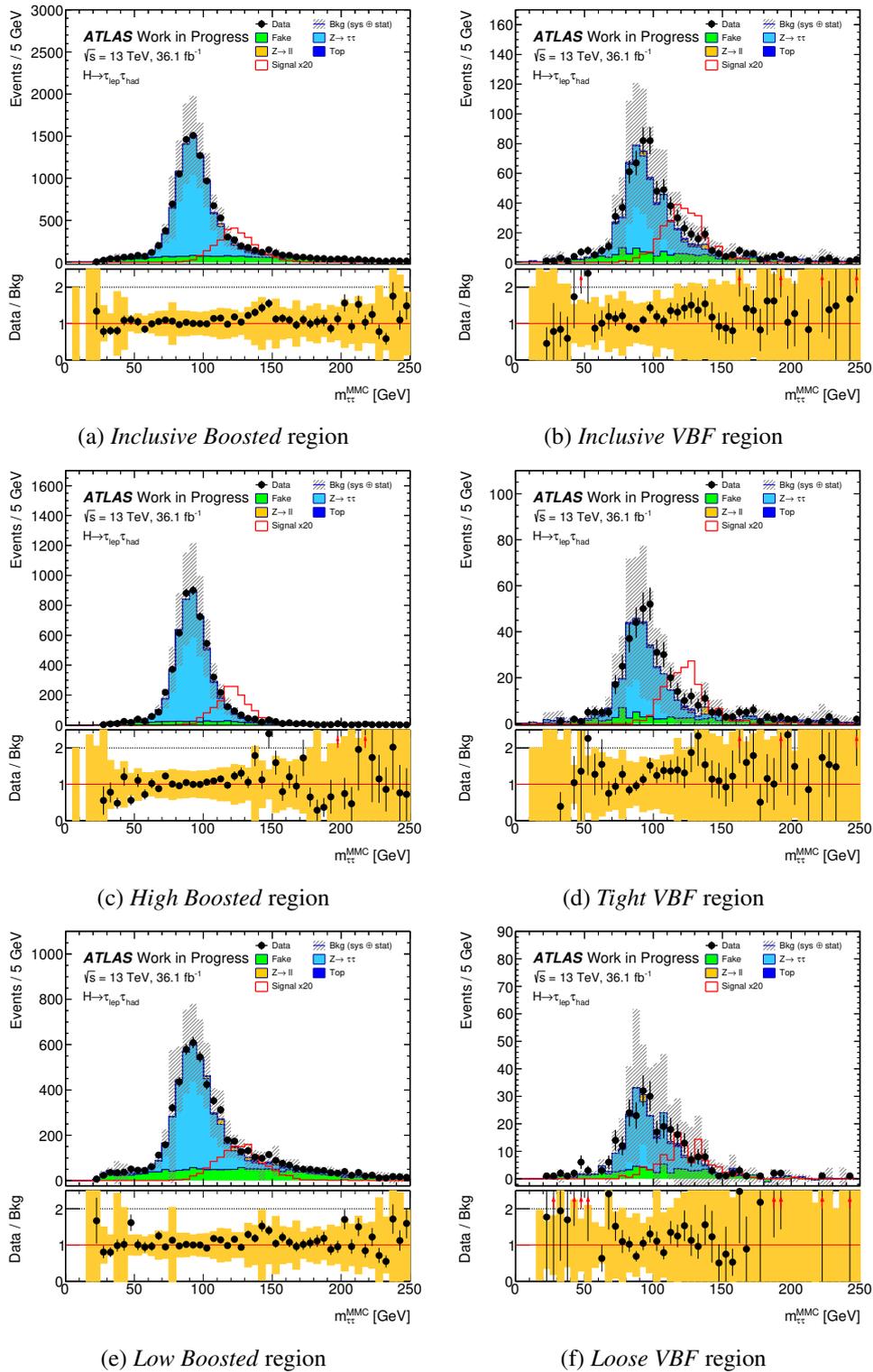


Figure 9.1: Reconstructed  $m_{\tau\tau}^{MMC}$  distributions for the *Boosted* categories (top) and the *VBF* categories (bottom). The *inclusive* signal categories, the *high/tight* and *low/loose* categories are shown from left to right. The black dots correspond to the data, while the colored histograms belongs to the different backgrounds.

## 9.2 Fit model

In this study the signal region are binned in the reconstructed di- $\tau$ -lepton mass  $m_{\tau\tau}^{MMC}$  determined by the missing mass calculator (see Section 6). The control regions are employed to constrain the event yields of top quark initiated background processes and contain a single bin. The used regions and their binning are summarized in Table 9.3. The signal strength  $\mu$

Typ	Region	Binning
Signal region	<i>Boosted high</i>	$m_{\tau\tau}^{MMC}$ binning
Signal region	<i>Boosted low</i>	$m_{\tau\tau}^{MMC}$ binning
Control region	<i>Boosted top control region</i>	single bin
Signal region	<i>VBF tight</i>	$m_{\tau\tau}^{MMC}$ binning
Signal region	<i>VBF loose</i>	$m_{\tau\tau}^{MMC}$ binning
Control region	<i>VBF top control region</i>	single bin

Table 9.3: Used regions for the final likelihood function, with their corresponding binning, in the  $H \rightarrow \tau\tau$  decay channel.

is the parameter of interest (PoI), defined as the ratio of the fitted signal cross section times the branching ratio of the  $H \rightarrow \tau\tau$  to the signal cross section times the branching ratio of the  $H \rightarrow \tau\tau$  decay predicted by the SM ( $\mu = (\sigma \cdot BR_{\tau\tau})_{obs} / (\sigma \cdot BR_{\tau\tau})_{SM}$ ). A value of zero refers to the absence of a signal and a value of  $\mu = 1$  corresponds to a signal as predicted by the SM.

A binned likelihood function is used, constructed as the product of Poisson probability terms (further information in [94]). The likelihood function is constructed by fit of the expected signal plus background to data in each signal category and control region. All backgrounds discussed in the previous section are included in the likelihood function. All backgrounds are subject to variations due to experimental and theoretical systematic uncertainties which are modelled via nuisance parameters.

## 9.3 Results

Table 9.4 show the observed results for a fit with the combined 2015 and 2016 dataset of the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel at the ATLAS experiment. The table shows the best fit value, the statistical uncertainty on data and the systematic uncertainty further split into different uncertainty sources. The best fit value for the signal strength  $\mu_{obs}$  is given by:

$$\mu_{obs} = 0.80 \pm 0.43, \quad (9.1)$$

in agreement with the SM expectation. The observed  $p_{0, obs}$ -value and the corresponding significance are given by:

$$p_{0, obs} = 0.032 (1.9\sigma). \quad (9.2)$$

To estimate the expected uncertainties on the signal strength and the expected significance a fit with an Asimov dataset is done. This data set is calculated by the sum over all simulated

background samples and simulated signal samples in the signal and control regions. The estimated expected significance has a  $p_{0, \text{exp}}$ -value of 0.017 corresponding to an expected significance of  $2.1 \sigma$ . The observed significance is lower than the expected significance. The dominant uncertainty sources shown in Table 9.4 are related to the jet and  $E_T^{\text{miss}}$  reconstruction ( $\pm 0.28$ ) and the statistical uncertainties of the dataset ( $\pm 0.22$ ), followed by the "fake" uncertainties ( $\pm 0.15$ ) and the statistical uncertainties of the used simulated samples ( $\pm 0.19$ ). The nuisance parameters and normalisation factors are given in Figure A.1 in the appendix.

Best fit values	$0.80 \pm 0.43$
Data statistic	$\pm 0.22$
Systematic unc.	$\pm 0.37$
Normalization unc.	$\pm 0.04$
Jets and $E_T^{\text{miss}}$	$\pm 0.28$
B-jets	$\pm 0.03$
Light leptons	$\pm 0.01$
$\tau$ -leptons	$\pm 0.07$
Pileup re-weighting	$\pm 0.09$
"Fake" background	$\pm 0.15$
Luminosity	$\pm 0.04$
Theory unc. on signal	$\pm 0.09$
Theory unc. on background	$\pm 0.01$
Simulation statistic	$\pm 0.19$

Table 9.4: Best fit value for the signal strength with a combined the 2015 and 2016 dataset of the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel with  $36\text{fb}^{-1}$  and its uncertainty split into several sources.

Figures 9.2a-9.2d show the  $m_{\tau\tau}^{\text{MMC}}$  distribution for the four signal event categories after the fit ("post fit"). Thus the top quark background is normalized by the control region fit and the determined nuisance parameters are applied. The colored histograms except of the red one which corresponds to the signal are the background events. The blue dashed line corresponds to the background calculation before the fit. Since only the top quark background is re-weighted with the top control regions in the fit, the difference between post-fit and pre-fit backgrounds is introduced through this re-weighting and through pulls on the nuisance parameters (see Figure A.1b). The calculated normalisation factors for the top quark background  $N_{VBF}^{\text{Top}}$  for the *VBF* region and  $N_{Boosted}^{\text{Top}}$  for the *Boosted* region are given by:

$$N_{VBF}^{\text{Top}} = 1.4 \pm 0.3 \quad \text{and} \quad N_{Boosted}^{\text{Top}} = 1.1 \pm 0.1. \quad (9.3)$$

The Figure 9.2e and 9.2f illustrate the control region post-fit yields. The figures confirm that the post fit top production background for the *VBF* region is higher than the pre fit calculated background, while the difference in the *Boosted* region is small.

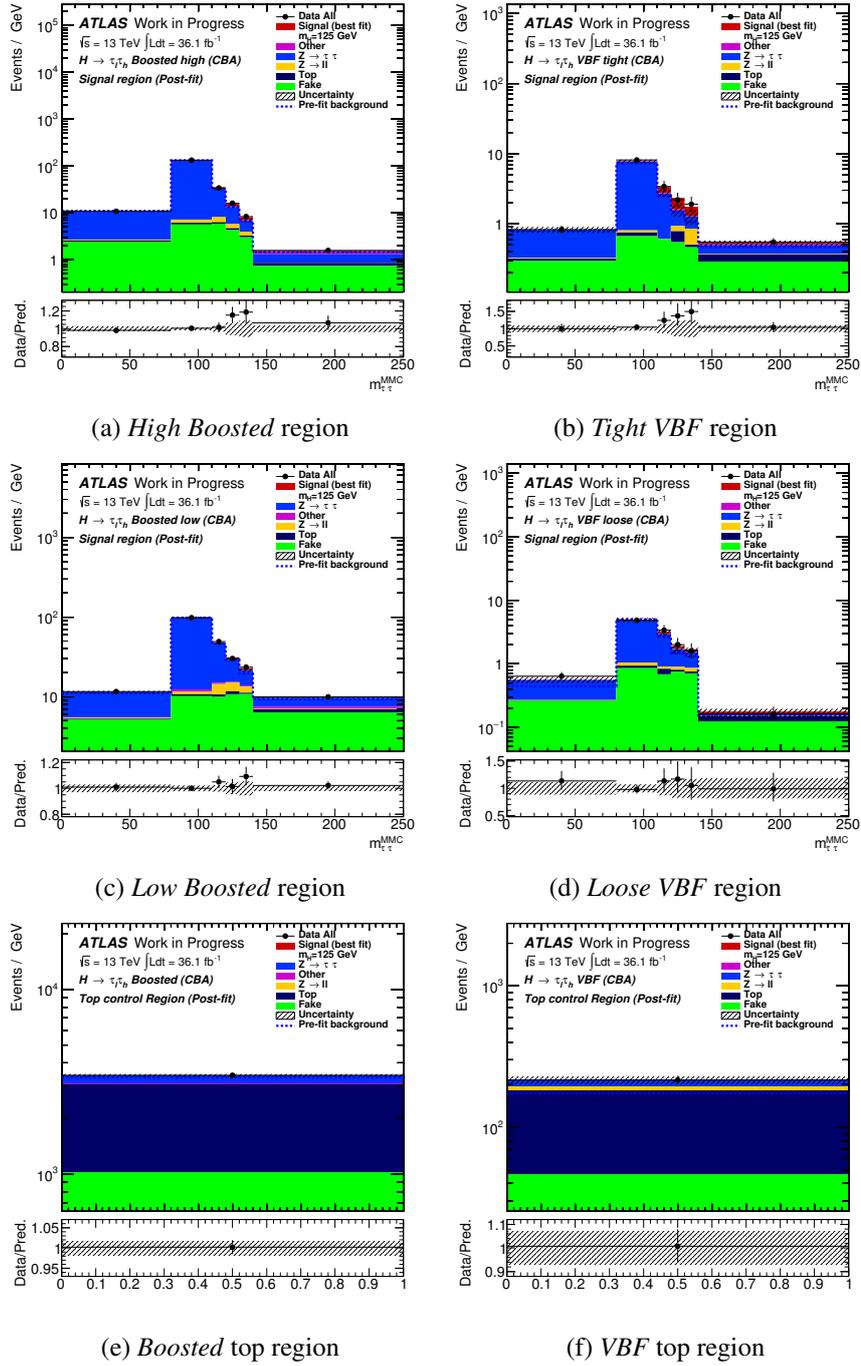


Figure 9.2: Reconstructed di- $\tau$ -lepton mass  $m_{\tau\tau}^{MMC}$  distributions after the fit for the *Boosted* categories (top) and the *VBF* categories (bottom). The signal regions are shown in the first two rows and the control regions in the last row. The black dots correspond to the data, while the colored histograms belongs to the calculated background plus signal events (red).

---

## 10 BACKGROUND PROCESSES FROM JETS MISIDENTIFIED AS $\tau_{HAD}$

---

This chapter discusses the estimation of background processes from jets misidentified as hadronically decaying  $\tau$  leptons, so called "fakes". These events are a dominant background for the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel. The "fakes" are mostly initiated by W+jets or multi-jet processes, but also by top quark production and Z+jets processes.

### 10.1 Description of the underlying method

Hadronically decaying  $\tau$ -lepton objects are identified (see Chapter 6) by using a Boosted Decision Tree (BDT). The collimation of the  $\tau$ -lepton is one of the dominant variables for this BDT. As a consequence background jets, which have a similar collimation as the jets of the  $\tau$ -lepton decay, are likely to get misidentified as a  $\tau$ -lepton. The collimation of the jet is highly depending on whether the jet is initiated by a quark or a gluon. The quark-gluon fraction is different for W+jets, multi-jet, top quark production and Z+jets processes. To take this into account the "fakes" are estimated for each process separately.

For estimating this background a data-driven method is highly desirable, since simulations are limited by statistics and systematic uncertainties. This analysis uses a so called "fake-factor" method, which is similar to the one used in Run 1 [20]. This thesis transfers the method to the new dataset, which was taken in the years 2015 and 2016, and optimizes it.

A control region, which is orthogonal to the signal region, is used for this method. This control region should be as similar as possible to the signal region, but highly dominated by "fake" events. The number and the shape of the "fake" events are calculated in this region. To transfer these into the proper signal region, a transfer factor is used, which is called "fake-factor" (FF). This transfer factor is estimated by the ratio between events with a  $\tau$ -lepton, which pass the  $\tau$ -ID medium criteria  $N_{\text{pass } \tau\text{-ID}}$  and events with a  $\tau$ -lepton, which fail the criteria  $N_{\text{fail } \tau\text{-ID}}$ :

$$\text{FF} = \frac{N_{\text{pass } \tau\text{-ID}}}{N_{\text{fail } \tau\text{-ID}}}. \quad (10.1)$$

To take the different afore-noted background contribution to "fakes" into account a individual "fake-factor"  $\text{FF}_i$  is calculated for each process  $i$ . The combined "fake-factor" FF is calculated by the sum weighted with the relative contribution  $R_i$  of the process  $i$ :

$$\text{FF} = \sum_i R_i \cdot \text{FF}_i. \quad (10.2)$$

To calculate the individual "fake-factors" further control regions are defined in which the corresponding process is enriched. Thus the method assumes that the "fake-factors" calculated in control region can be used in the signal region. The final number of "fakes" in the proper signal region is given by:

$$N_{\text{fakes}}^{\text{SR}} = N_{\text{fakes}}^{\text{anti-}\tau} \cdot \text{FF}, \quad (10.3)$$

where  $N_{\text{fakes}}^{\text{SR}}$  is the number of "fakes" in the signal region and  $N_{\text{fakes}}^{\text{anti-}\tau}$  is the number of "fakes" in the "anti- $\tau$ " control region described in the next section.

### 10.1.1 "Anti- $\tau$ " control region and estimation of "fake" events

A region called "anti- $\tau$ " region is used as the control region, in which the number of "fake" events and their shape are calculated. The "anti- $\tau$ " region is defined by a leading  $\tau$ -lepton which fails the  $\tau$ -ID medium requirement. This means, that this property is inverted compared to the signal region (see Figure 10.1a). This definition ensures a high amount of "fakes" and excludes most of the events with a true leading  $\tau$ -lepton (see Table A.1 in the appendix). At low BDT scores the quark-gluon fraction changes significantly. Since this fraction has a high influence on the "fake" rate an additional cut on the BDT score is set ( $BDT > 0.35$ ), which should bring the quark-gluon fraction in the "anti- $\tau$ " region closer to the one in the signal region without losing too much statistics. The events where the reconstructed  $\tau$ -lepton is not "faked" by a jet are subtracted by using simulated events. The number of "fakes" in the "anti- $\tau$ " control region  $N_{fakes}^{anti-\tau, SR}$  is defined by:

$$N_{fakes}^{anti-\tau, SR} = N_{data}^{anti-\tau, SR} - N_{MC, not j \rightarrow \tau}^{anti-\tau, SR}, \quad (10.4)$$

where  $N_{data}^{anti-\tau, SR}$  refers to the number of data events in the "anti- $\tau$ " region and  $N_{MC, not j \rightarrow \tau}^{anti-\tau, SR}$  to the number of simulated events in the "anti- $\tau$ " region, where the reconstructed  $\tau_{had}$ -lepton is not initiated by a jet. This definition applied to Equation 10.3 equals:

$$N_{fakes}^{SR} = (N_{data}^{anti-\tau, SR} - N_{MC, not j \rightarrow \tau}^{anti-\tau, SR}) \cdot FF. \quad (10.5)$$

Thus the "fake-factor" transfers the "fake" estimation from the anti- $\tau$  region to the  $\tau$  region (see Figure 10.1b).

### 10.1.2 "Fake-factor" estimation

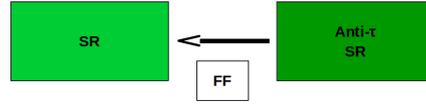
The "fake-factor" FF has to transfer the number of "fake" events from the "anti- $\tau$ " region to the signal region and has to be calculated separately for the four background processes:

$$FF = R_W \cdot FF_W + R_{top} \cdot FF_{top} + R_Z \cdot FF_Z + R_{multi-jet} \cdot FF_{multi-jet}, \quad (10.6)$$

where W refers to W+jets, top to top quark production, Z to Z+jets and multi-jet to multi-jet processes. However the contributions of top quark production and of Z+jets processes are small compared to the other backgrounds. To simplify the method the "fake-factor" of W+jets processes is used for all non-multi-jet processes, such that the relative contribution of W+jets is the sum of all non-multi-jet contributions. The underlying assumption that the difference



(a) The anti- $\tau$  region (right) is defined by inverting the medium ID criteria on the leading  $\tau$ -lepton. All further cuts are the same as for the signal region.



(b) The "fake-factor" is the transfer factor from the "anti- $\tau$ " to the  $\tau$  region.

Figure 10.1

between the "fake-factors" of the non-multi-jet processes is negligible ( $FF_W \approx FF_{top} \approx FF_Z$ ) is proven in Section 10.3.1. Thus the "fake-factor" FF is calculated in the following way:

$$FF = R_{WZt} \cdot FF_W + R_{\text{multi-jet}} \cdot FF_{\text{multi-jet}}, \text{ with} \quad (10.7)$$

$$R_{WZt} = \left( \sum_{W,Z,top} R_i \right) \quad (10.8)$$

### 10.1.3 Individual "fake-factors"

To calculate a "fake-factor" for one individual process in a data-driven way, it is essential to separate the events of this process from other backgrounds. For this control regions are defined, which enhance the contribution of the corresponding process. The control regions should be as similar as possible to the signal region and they should all be orthogonal to each other to ensure that no event is used twice. To achieve this condition one different property of the signal region is inverted for each control region.

Process	CR <sub>i</sub> = SR, except	Proportion of events (%)					
		<i>Boosted</i>			<i>VBF</i>		
		W+jets	Multi-jet	Others	W+jets	Multi-jet	Others
W+jets	Inverted $m_T$ cut ( $m_T > 70$ GeV)	84	4	12	82	3	15
Multi-jet	Inverted lepton isolation	2	97	1	2	97	1

Table 10.1: Definition and purity of the individual "anti- $\tau$ " control regions in the *Boosted* and *VBF* category.

Since multi-jet events are not expected to have true isolated leptons (see Chapter 5) the lepton isolation requirement is inverted such that the detected lepton is mostly a misidentified jet. For W+jets processes the cut on the transverse mass of the signal region is inverted, since this cut is applied on the signal region to reduce the high amount of W+jets background (see Chapter 8). The cut difference and the purity of the control regions for the *Boosted* and *VBF* signal region are summarised in Table 10.1, while the exact event yields of the control regions are given in Table A.1 in the appendix. The multi-jet processes have a proportion of 97% in both multi-jet control regions and the W+jets processes a proportion of at least 82% in the W+jets control regions.

Since the "fake-factor" should transfer the number of events with a jet misidentified as a  $\tau$ -lepton in the "anti- $\tau$ " region to the number of events with this property in the signal region, it is calculated via a fraction of events in a  $\tau$  region over events in the corresponding "anti- $\tau$ " region. This means that to calculate the  $FF_i$  of the process  $i$ , two different cuts are being used. First the cut to get the actual control region of this process and then to get the corresponding "anti- $\tau$ " region additionally the  $\tau$ -ID requirement is inverted. Thus there are six different regions included in this method (see Figure 10.2).

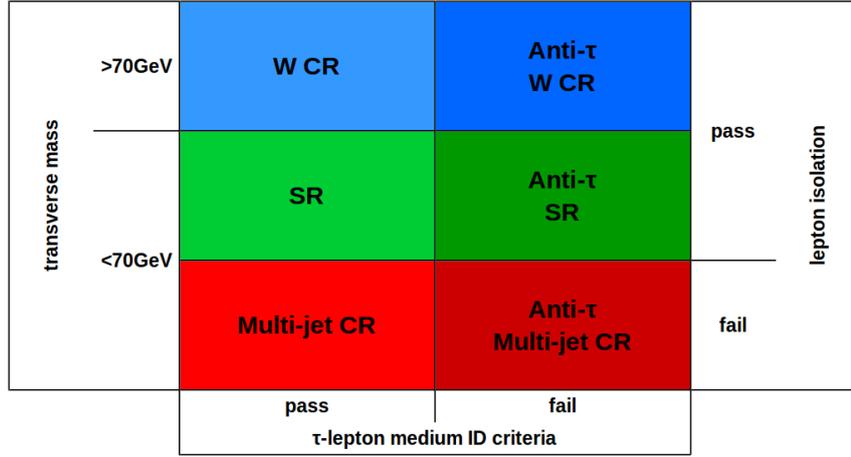


Figure 10.2: Overview of different regions, which are used in the "fake-factor" method.

These control regions are dominated by the corresponding background but not completely cleaned from other processes, such that simulated events are used to subtract the events in this region, which come from other sources. The exact formula for the individual "fake-factor" for a background process  $i$   $FF_i$  is:

$$FF_i = \frac{N_{data}^{\tau, CR_i} - N_{MC, not j \rightarrow \tau}^{\tau, CR_i} - N_{MC, other}^{\tau, CR_i}}{N_{data}^{anti-\tau, CR_i} - N_{MC, not j \rightarrow \tau}^{anti-\tau, CR_i} - N_{MC, other}^{anti-\tau, CR_i}}, \quad (10.9)$$

where  $N_{data}^{\tau, CR_i}$  are the events in the  $\tau$  control region for the process  $i$ ,  $N_{MC, not j \rightarrow \tau}^{\tau, CR_i}$  are the events in the  $\tau$  control region for this process where the  $\tau$ -signal is not initiated by a jet,  $N_{MC, other}^{\tau, CR_i}$  are events initiated by other background processes in the  $\tau$  control region for this process and  $N_{data}^{anti-\tau, CR_i}$ ,  $N_{MC, not j \rightarrow \tau}^{anti-\tau, CR_i}$ ,  $N_{MC, other}^{anti-\tau, CR_i}$  are the corresponding events in the "anti- $\tau$ " control region of the process  $i$ . Figure 10.3 illustrates how these individual "fake-factors" contribute to the whole method.

#### 10.1.4 Relative contribution

Since results of simulations are limited by statistics and systematic uncertainties, a data-driven method is preferred to determine the relative contributions. For this the relative contribution of multi-jet events  $R_{multi-jet}$  is calculated in a data-driven way. This relative contribution is then used to estimate the contributions of non-multi-jet processes  $R_{WZt}$ , using that the sum of all relative contributions must be exactly one:

$$R_{WZt} = 1 - R_{multi-jet}. \quad (10.10)$$

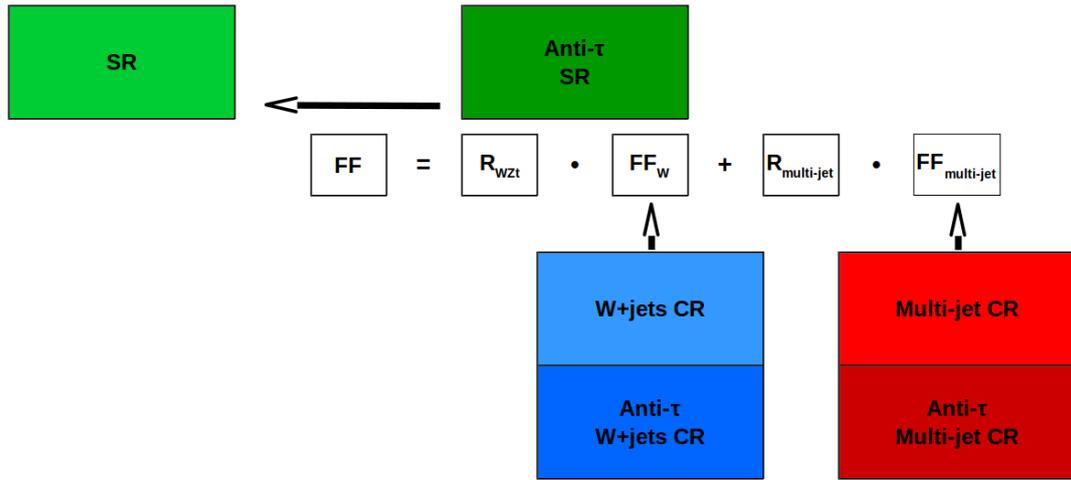


Figure 10.3: The individual "fake-factors"  $FF_W$  and  $FF_{\text{multi-jet}}$  are calculated in their respective control region and the corresponding "anti- $\tau$ " region. They are summed up with relative contributions to a combined "fake-factor"  $FF$ , which transfers the number of "fakes" from the "anti- $\tau$ " region to the signal region.

### 10.1.5 Multi-jet contribution

To calculate the relative contribution of multi-jet events in the signal region several control regions are used. The relative contribution can be calculated by:

$$R_{\text{multi-jet}} = \frac{N_{\text{multi-jet,data}}^{\text{"anti-}\tau\text{",SR}}}{N_{\text{data}}^{\text{"anti-}\tau\text{",SR}} - N_{\text{MC, not } j \rightarrow \tau}^{\text{"anti-}\tau\text{",SR}}}, \quad (10.11)$$

where  $N_{\text{multi-jet,data}}^{\text{"anti-}\tau\text{",SR}}$  is the number of events in the "anti- $\tau$ " region, which are initiated by multi-jet processes,  $N_{\text{data}}^{\text{"anti-}\tau\text{",SR}}$  the events in the "anti- $\tau$ " region,  $N_{\text{MC, not } j \rightarrow \tau}^{\text{"anti-}\tau\text{",SR}}$  events where the  $\tau$ -lepton is not initiated by a jet. The denominator can be determined directly, while a data-driven method has to be used for the numerator. For the estimation of the multi-jet events in the the "anti- $\tau$ " region the lepton isolation is used. The multi-jet control region defined by inverting the isolation criteria on the lepton, results in a region which is dominated by event with a jet misidentified as a lepton. A high amount of these events are multi-jet initiated events. Simulated events are used to subtract other backgrounds in this region. A transfer factor is applied to transfer the number of events in this multi-jet region to the "anti- $\tau$ " signal region. This transfer factor is called isolation factor "IF":

$$N_{\text{multi-jet,data}}^{\text{"anti-}\tau\text{",SR}} = \text{IF} \cdot \left( N_{\text{data}}^{\text{"anti-}\tau\text{",CR}_{\text{multi-jet}}} - N_{\text{MC, true lepton}}^{\text{"anti-}\tau\text{",CR}_{\text{multi-jet}}} \right), \quad (10.12)$$

where  $N_{\text{data}}^{\text{"anti-}\tau\text{",CR}_{\text{multi-jet}}}$  is the number of events in the "anti- $\tau$ " multi-jet control region and  $N_{\text{MC, true lepton}}^{\text{"anti-}\tau\text{",CR}_{\text{multi-jet}}}$  the number of events in this region with a true leading lepton. Figure 10.4 shows a schematic overview of the calculation of the relative contribution of multi-jet events.

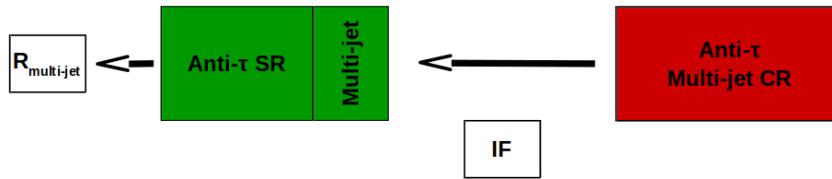


Figure 10.4: The relative contribution of multi-jet events in the "anti- $\tau$ " signal region is calculated by using the isolation factor "IF", which transfers the number of multi-jet events in the "anti- $\tau$ " multi-jet control region to the "anti- $\tau$ " signal region.

### Isolation factor

For the calculation of the transfer factor IF, which transfers the number of multi-jet events in the "anti- $\tau$ " multi-jet control region to the "anti- $\tau$ " signal region, new control regions are necessary. These control regions should have similar properties as the "anti- $\tau$ " signal region and "anti- $\tau$ " multi-jet control region.

An obvious idea is to take the definitions of these regions and invert an additional cut on both regions in the same way. Thus the new control regions are orthogonal to all other region and should have similar behaviour. For this the requirement of the signal region that the the light lepton and the hadronically decaying  $\tau$ -lepton to have an opposite charge (see Section 8.1) is inverted. Instead both of these particles are required to have the same charge sign. This "same sign" control region is used for closure tests as well (see Section 10.3.5). This implies the assumption that the *IF* in the "opposite sign" and "same sign" region are equal, for which further studies are performed and discussed in Section 10.3.3. An overview of the used control region for the isolation factor calculation is shown in Figure 10.5.

		light lepton and $\tau$ -lepton have opposite sign charge					
		true		false			
		SR	Anti- $\tau$ SR	same sign Anti- $\tau$ SR	pass	lepton isolation	
		Multi-jet CR	Anti- $\tau$ Multi-jet CR	same sign Anti- $\tau$ Multi-jet CR	fail		
		pass		fail			
		$\tau$ -lepton medium ID criteria					

Figure 10.5: The isolation factor IF is calculated in the "same sign" control region. To get a corresponding "same sign" region all cuts of the original region are kept except for the requirement on the combined charge sign of the light lepton and the hadronically decaying  $\tau$ -lepton (see Section 8.1).

The isolation factor can be calculated by the fraction of the number of multi-jet events in the "same sign" "anti- $\tau$ " signal region and the number of multi-jet events in the "same sign"

"anti- $\tau$ " multi-jet control region. To get the multi-jet events in this region both regions are cleaned from events with a true leading lepton. For this correction simulated events are being used again. This leads to the following definition of the isolation factor  $IF$ :

$$IF = \frac{N_{data}^{SS, "anti-\tau", SR} - N_{MC, true lepton}^{SS, "anti-\tau", SR}}{N_{data}^{SS, "anti-\tau", CR_{multi-jet}} - N_{MC, true lepton}^{SS, "anti-\tau", CR_{multi-jet}}}, \quad (10.13)$$

where  $N_{data}^{SS, "anti-\tau", SR}$  is the number events in the "same sign" "anti- $\tau$ " signal region,  $N_{MC, true lepton}^{SS, "anti-\tau", SR}$  the number events in this region with a true lepton and  $N_{data}^{SS, "anti-\tau", CR_{multi-jet}}$ ,  $N_{MC, true lepton}^{SS, "anti-\tau", CR_{multi-jet}}$  the corresponding numbers of events in the "same sign" "anti- $\tau$ " multi-jet control region.

### 10.1.6 Summary of the method

In the previous sections all necessary parts of the "fake-factor" method were described. This section summarises the method in a short way. Figure 10.6 shows all used regions of the method. Next to the signal region two control regions are defined. One control region for W+jets events and one for multi-jet events. For all regions "anti- $\tau$ " regions are defined. Additionally for the calculation of the isolation factor "same sign" regions for the "anti- $\tau$ " signal region and the "anti- $\tau$ " multi-jet control region are used. Figure 10.7 illustrates all steps of the method. First the isolation factor is calculated and applied to the "anti- $\tau$ " multi-jet control region to get the number of multi-jet events in the "anti- $\tau$ " signal region. With this number the relative contribution of multi-jet events and the relative contribution of W+jets events is determined. Additionally the individual "fake-factor" for W+jets and multi-jet processes are calculated. Finally they are summed up with their relative contribution and applied to the "anti- $\tau$ " signal region to get the number of "fake" events in the  $\tau$  signal region.

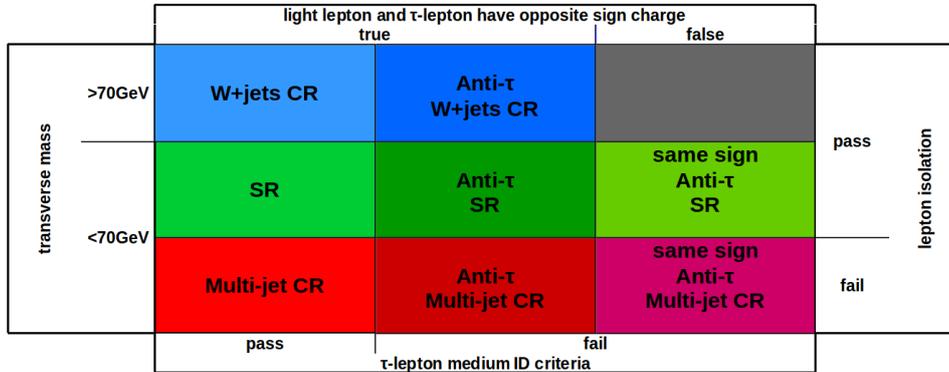


Figure 10.6: Overview of all used regions of the "fake-factor" method, with the transverse mass cut (left) which defines the W+jets control region, the isolation requirement of the leading light lepton (top), which is used to define a multi-jet control region, the  $\tau$ -lepton ID criteria (bottom), which are inverted to get the "anti- $\tau$ " control region and the requirement on the light lepton and the hadronically decaying  $\tau$ -lepton sign charge (top).

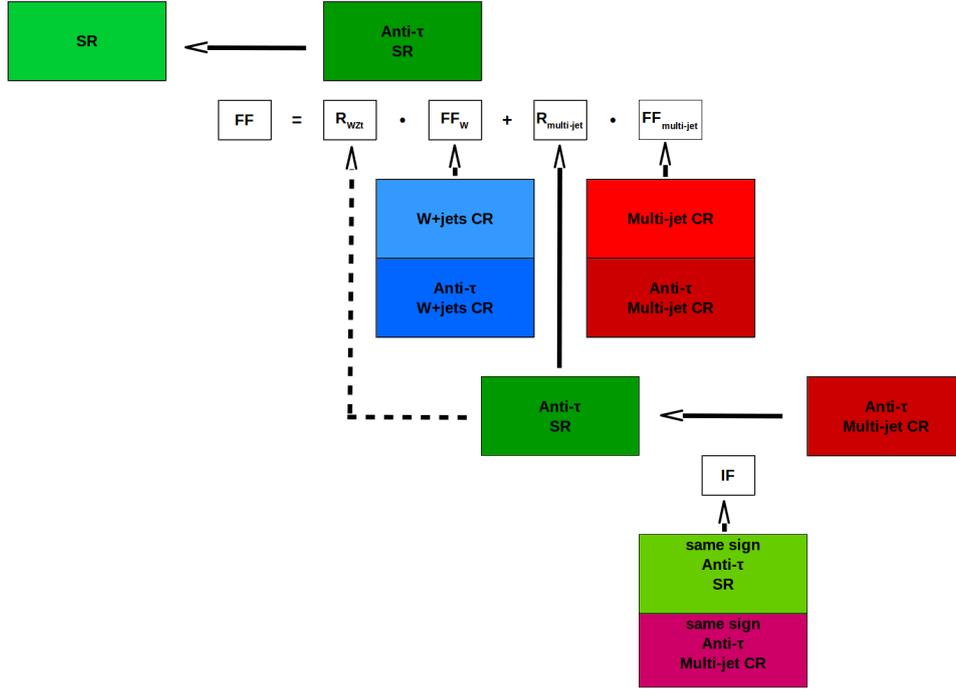


Figure 10.7: For the "fake" calculation an isolation factor IF is calculated first. This factor is applied to "anti- $\tau$ " multi-jet control regions to calculate the relative contribution of multi-jet and W+jets events in the signal region. Two individual "fake-factors" for multi-jet and W+jets processes are estimated and summed up with their relative contribution. The final "fake-factor" is then applied to the "anti- $\tau$ " signal region to get the number of "fakes" in the  $\tau$  signal region.

## 10.2 Application of the "fake-factor" method

To apply the method on the dataset of the  $H \rightarrow \tau_{had} \tau_{lep}$  analysis the different signal regions of the analysis have to be taken into account. The analysis is split into two different signal categories. These two categories have a different contribution of signal and background processes. Thus the "fake-factor" for these two regions should be calculated separately. Furthermore the isolation factor IF could also be different for these two categories. Studies have shown that the control regions for the estimation of IF have limited statistics. Due to this fact, the IF is not determined in the actual signal categories. Instead one common isolation factor is computed in at the *Preselection* level. This isolation factor is then applied on the *Boosted* and on the *VBF* categories. This assumes that the isolation factors determined at this level is the same as in the *Boosted* and the *VBF* region ( $IF_{Preselection} \approx IF_{VBF} \approx IF_{Boosted}$ ). To study this approximation several studies have been done (see Section 10.3.4). These studies have lead to additional systematic uncertainties, which are applied on the isolation factor.

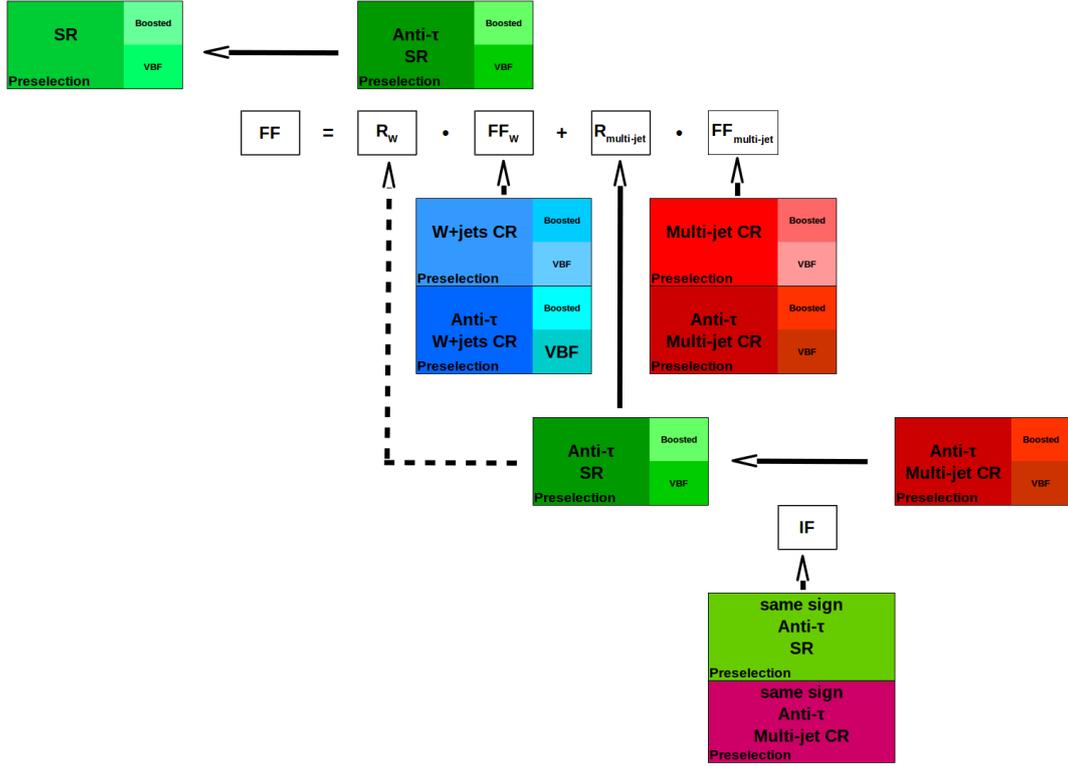


Figure 10.8: To apply the "fake-factor" method to the dataset the individual "fake-factors" and relative contribution are calculated separately for the *Preselection*, *Boosted* and *VBF* region. Since the *Boosted* and *VBF* categories are defined on top of the *Preselection*, both regions are included in the *Preselection* region (see Section 8.1). Only the isolation factor IF is always calculated in the *Preselection* due to higher statistics.

The relative contribution of the different background processes and the individual "fake-factor" are calculated separately for the event categories (*Boosted* and *VBF*). This leads to the previous method being modified in the way illustrated in Figure 10.8.

The used dataset corresponds to the year 2015 and 2016. Between these two periods of data-taking some conditions of the detector have changed, like the used triggers (see Chapter 5). This of course can also result in different "fake-factors". For this additional studies have been done, which have validated that such a split is not necessary (see Section 10.3.2).

### 10.2.1 Isolation factor

The isolation factor show dependencies on the light lepton flavour (electron or muon), on the transverse momentum of the light lepton  $p_{\text{lep}}^T$  and on the absolute value of the pseudorapidity of the light lepton  $|\eta|$ . For this three-dimensional binning a balance between fine binning and acceptable statistics has to be found. Both light lepton flavour channels are binned in three  $p_{\text{lep}}^T$  bins with  $p_{\text{lep}}^T < 30$  GeV,  $30$  GeV  $< p_{\text{lep}}^T < 40$  GeV and  $40$  GeV  $< p_{\text{lep}}^T$ . For  $|\eta|$  the

isolation factor is split into two bins at  $|\eta| = 1.5$ . For the electron channel the gap between the barrel and end-cap detector is excluded ( $1.37 < |\eta| < 1.52$ ). The isolation factor with the final binning is illustrated in Figure 10.9 and the precise values are summarized in Table 10.2.

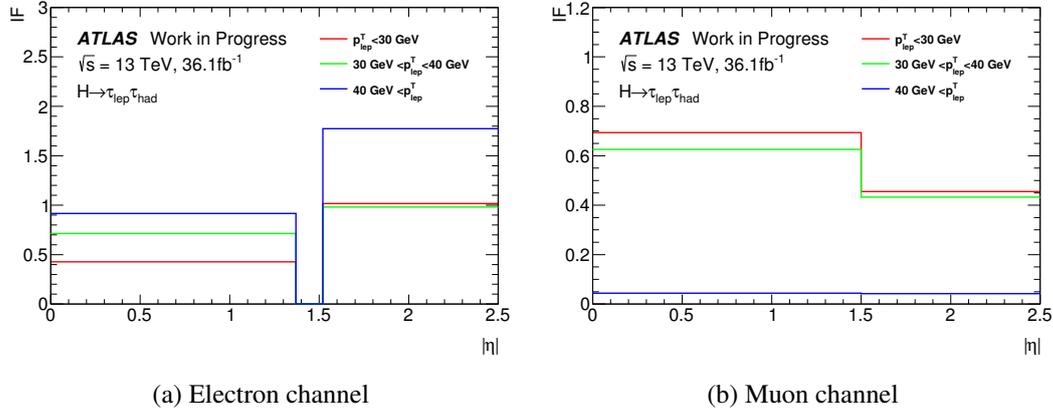


Figure 10.9: Isolation factor IF for the electron channel (left) and the muon channel (right), as a function of the transverse mass on the light lepton  $p_{lep}^T$  and the absolute value of the pseudorapidity of the light lepton  $|\eta|$ .

The electron channel shows an increase of the isolation factor in  $|\eta|$ , while in the muon channel the IF is decreasing with  $|\eta|$ . The muon channel has an isolation factor shrinking with transverse momentum of the muon, while the opposite trend is visible in the electron channel.

For the muon channel a significant difference between the first two and the third bin in the transverse momentum of the muon can be identified. For the highest transverse momentum bin the isolation factor has values around zero, while the first two bin have values between 0.4 and 0.7. This behaviour is further discussed in Section 10.3.4. Distributions and studies on the statistical and systematic uncertainties of the isolation factor can be found in Section 10.3 and 10.4.

	$p_T$ [GeV]	$p_T < 30$	$30 < p_T < 40$	$p_T > 40$
Muon	$ \eta  < 1.50$	0.694	0.626	0.044
	$ \eta  > 1.50$	0.455	0.433	0.042
Electron	$ \eta  < 1.37$	0.427	0.714	0.916
	$ \eta  > 1.52$	1.017	0.982	1.775

Table 10.2: Isolation factors IF separated into the light lepton flavour (electron, muon), the pseudorapidity of the light lepton ( $\eta$ ) and the transverse momentum of the light lepton ( $p_T$ ). The background color changes from blue to red for increasing IF values.

### 10.2.2 Relative contributions

With the results of the isolation factor IF shown in the previous section it is possible to calculate the relative contribution of the multi-jet events. The relative contributions depend on the event category (*Preselection*, *Boosted*, *VBF*), the light lepton flavour (electron, muon), the number of tracks of the hadronically decaying  $\tau$ -lepton (1-prong, 3-prong), the transverse momentum of the hadronically decaying  $\tau$ -lepton ( $p_T^\tau$ ) and the difference of azimuthal angle between hadronically decaying  $\tau$ -lepton and missing transverse energy ( $\Phi(\tau, E_T^{\text{miss}})$ ).

Figure 10.10 illustrates the obtained relative contributions for the *Boosted* category and Figure 10.11 for the *VBF* category. Both categories are binned in three  $p_T^\tau$  bins with  $p_T^\tau < 40$  GeV,  $40 \text{ GeV} < p_T^\tau < 90$  GeV and  $90 \text{ GeV} < p_T^\tau$ . Due to the higher statistics in the *Boosted* category (see Table 8.2) the relative contribution for the *Boosted* category is binned in  $\Phi(\tau, E_T^{\text{miss}})$  into  $\Phi(\tau, E_T^{\text{miss}}) < 1$ ,  $1 < \Phi(\tau, E_T^{\text{miss}}) < 1.5$ ,  $1.5 < \Phi(\tau, E_T^{\text{miss}}) < 2$ ,  $2 < \Phi(\tau, E_T^{\text{miss}})$ . Due to the lower statistic in the *VBF* category (see Table 8.2) the first three bins in  $\Phi(\tau, E_T^{\text{miss}})$  are combined to one bin. All categories are dominated by W+jets events.

All categories have a higher W+jets contribution for the hadronically decaying  $\tau$ -leptons with three tracks. The W+jets contribution in the *Boosted* categories increases with rising  $\Phi(\tau, E_T^{\text{miss}})$ , while in the *VBF* categories this effect is difficult to see due to the low statistics.

		Lepton $p_T$ [GeV]	Electron						Muon						
			$p_T < 40$		$40 < p_T < 90$		$90 < p_T$		$p_T < 40$		$40 < p_T < 90$		$90 < p_T$		
			W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet	
Boosted	1 track	$\Delta\Phi < 1.0$	0.650	0.350	0.775	0.225	0.913	0.087	0.743	0.257	0.767	0.233	0.833	0.167	
		$1.0 < \Delta\Phi < 1.5$	0.715	0.285	0.842	0.158	0.845	0.155	0.855	0.145	0.846	0.154	0.831	0.169	
		$1.5 < \Delta\Phi < 2.0$	0.747	0.253	0.782	0.218	0.818	0.182	0.866	0.134	0.883	0.117	0.872	0.128	
		$\Delta\Phi > 2.0$	0.875	0.125	0.939	0.061	0.896	0.104	0.874	0.126	0.931	0.069	0.938	0.062	
	3 track	$\Delta\Phi < 1.0$	0.893	0.107	0.929	0.071	0.976	0.024	0.890	0.110	0.915	0.085	0.949	0.051	
		$1.0 < \Delta\Phi < 1.5$	0.902	0.098	0.952	0.048	0.965	0.035	0.929	0.071	0.939	0.061	0.953	0.047	
		$1.5 < \Delta\Phi < 2.0$	0.928	0.072	0.945	0.055	0.951	0.049	0.945	0.055	0.957	0.043	0.954	0.046	
		$\Delta\Phi > 2.0$	0.962	0.038	0.985	0.015	0.972	0.028	0.960	0.040	0.968	0.032	0.977	0.023	
	VBF	1 track	$\Delta\Phi < 2.0$	0.752	0.248	0.737	0.263	0.829	0.171	0.603	0.397	0.679	0.321	0.804	0.196
			$\Delta\Phi > 2.0$	0.729	0.271	0.566	0.434	0.012	0.988	0.999	0.001	0.648	0.352	0.982	0.018
3 track		$\Delta\Phi < 2.0$	0.918	0.082	0.931	0.069	0.935	0.065	0.836	0.164	0.892	0.108	0.933	0.067	
		$\Delta\Phi > 2.0$	0.933	0.067	0.938	0.062	0.939	0.061	0.824	0.176	0.846	0.154	0.996	0.004	

Table 10.3: Relative contributions  $R_i$  separated into the number of tracks of the  $\tau$ -lepton (1,3), the event category (*Boosted*, *VBF*) and the difference of the azimuthal angle between the  $\tau$ -lepton and the missing transverse energy ( $\Delta\Phi$ ) shown on the left-hand side and the lepton flavour (electron, muon) and the transverse momentum of the  $\tau$ -lepton ( $p_T$ ) shown on the top. The background color change from blue to red for increasing  $R_i$  values.

The multi-jet contribution of the electron channel shrinks with higher values of the transverse momentum. In the muon channel this effect is not that clear. A reason for this could be the lower statistics for the muon control regions. These dependencies could be covered by statistical fluctuations in the additional binning. The precise values for the relative contributions of W+jets and multi-jet processes are given in Table 10.3.

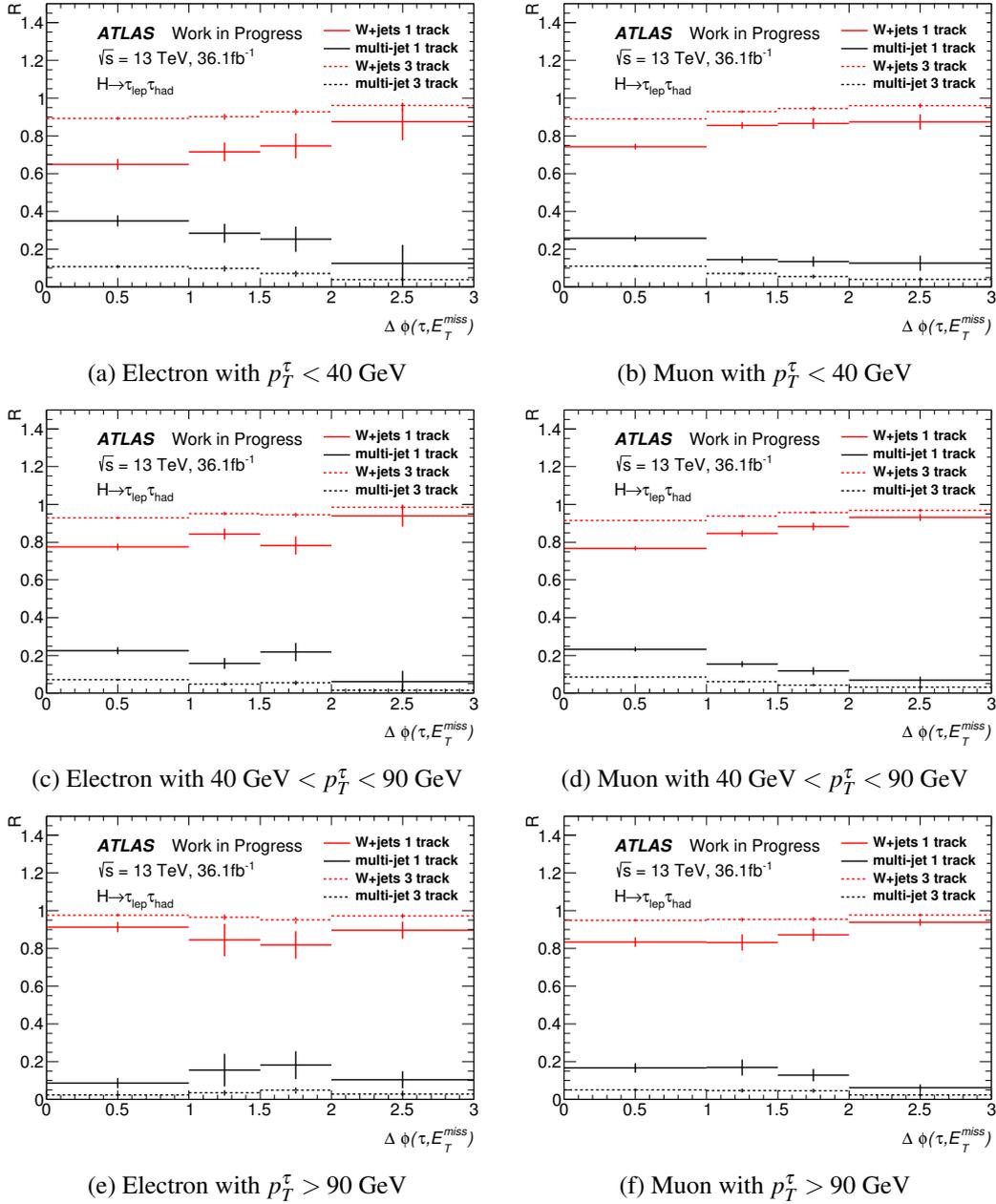


Figure 10.10: Relative contributions of the *Boosted* categories split into the electron channel (left) and muon channel (right). From the top to the bottom row different cuts on the transverse momentum of the light leptons are applied ( $p_T^\tau < 40$  GeV,  $40 \text{ GeV} < p_T^\tau < 90$  GeV,  $p_T^\tau > 90$  GeV). The dashed lines correspond to the results for hadronically decaying  $\tau$ -leptons with three tracks, while the full lines show the  $\tau$ -leptons with one track. The red line refers to the relative contribution of W+jets events, while the black one belongs to multi-jet events.

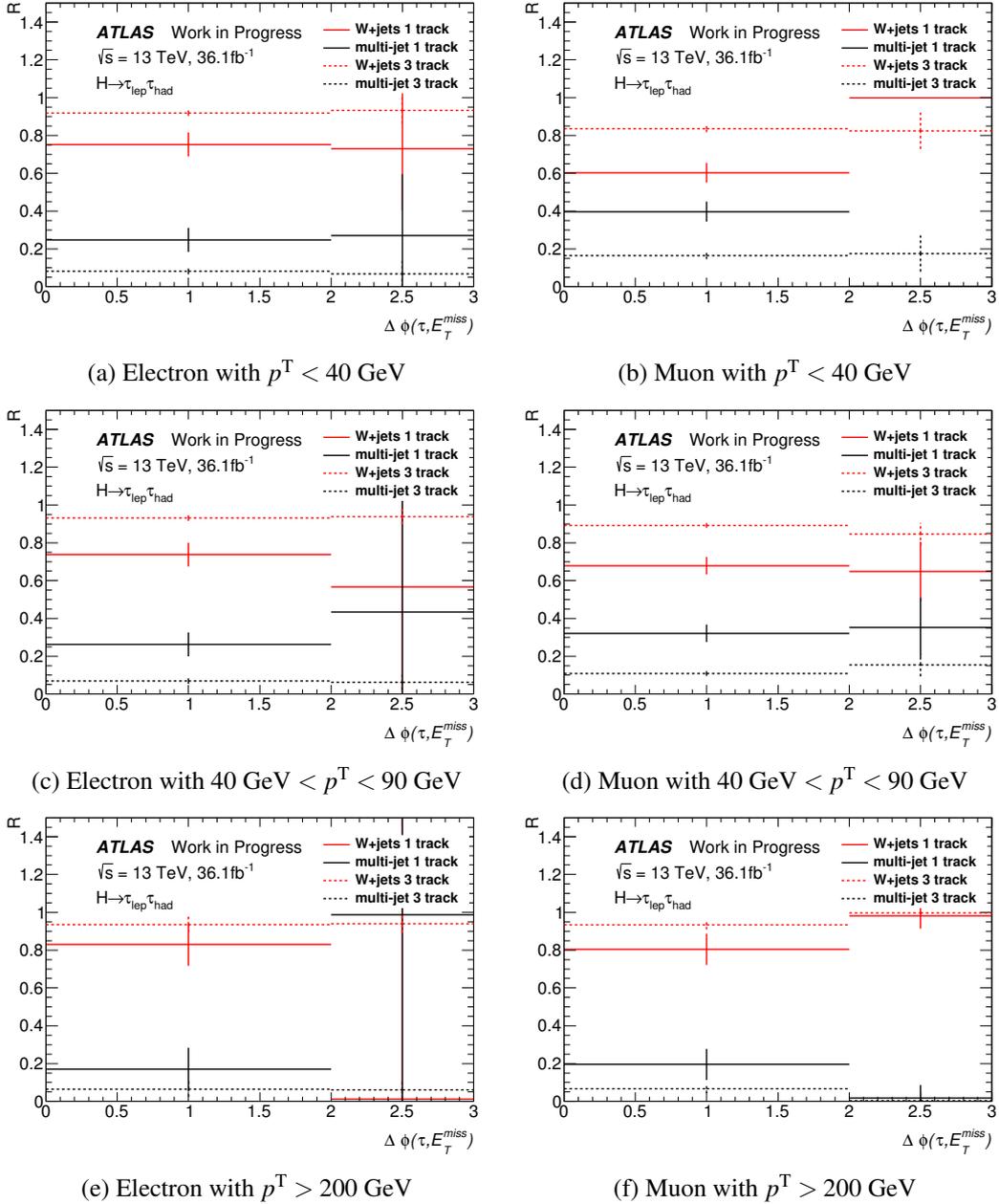


Figure 10.11: The relative contribution of the  $VBF$  category, split into the electron channel (left) and the muon channel (right). From the top to the bottom row different cuts on the transverse momentum of the light lepton are applied ( $p_T^\tau < 40 \text{ GeV}$ ,  $40 \text{ GeV} < p_T^\tau < 90 \text{ GeV}$ ,  $p_T^\tau > 90 \text{ GeV}$ ). The dashed lines correspond to the results for hadronically decaying  $\tau$ -leptons with three tracks, while the full lines show the  $\tau$ -leptons with one track. The red line refers to the relative contribution of  $W$ +jets events, while the black one belongs to multi-jet events.

### 10.2.3 Individual "fake-factors"

The result for the individual "fake-factor" depending on the transverse momentum of the hadronically decaying  $\tau$ -lepton  $p_T^\tau$  for W+jets and multi-jet processes are shown in Figure 10.12 for *Boosted* (left) and *VBF* (right). The FF of the *Boosted* and *VBF* have similar values but much lower statistics in the *VBF* categories result in higher statistical uncertainties on the FF. Higher values for the FF of hadronically decaying  $\tau$ -leptons with a track multiplicity of 1 (1-prong) than for a track multiplicity of 3 (3-prong) are observed. For the 1-prong channel the FF is slightly higher for W+jets processes, while for the 3-prong the FF is slightly higher for multi-jet processes. The precise values of the individual "fake-factor"  $FF_i$  for the *Boosted* and *VBF* category are summarized in Table 10.4.

		$p_T^\tau$ [GeV]		W+jets		Multi-jet	
		< 40	> 40	< 40	> 40	< 40	> 40
Boosted	1 track	0.252	0.161	0.178	0.165		
	3 track	0.037	0.033	0.025	0.030		
VBF	1 track	0.150	0.170	0.185	0.169		
	3 track	0.023	0.044	0.018	0.027		

Table 10.4: Individual "fake-factors"  $FF_i$  for W+jets and multi-jet processes separated into the number of tracks of the  $\tau$ -lepton (1,3), the event category (*Boosted*, *VBF*) shown on the left-hand side and the transverse momentum of the  $\tau$ -lepton ( $p_T^\tau$ ) shown on the top. The background color changes from blue to red for increasing  $FF_i$  values.

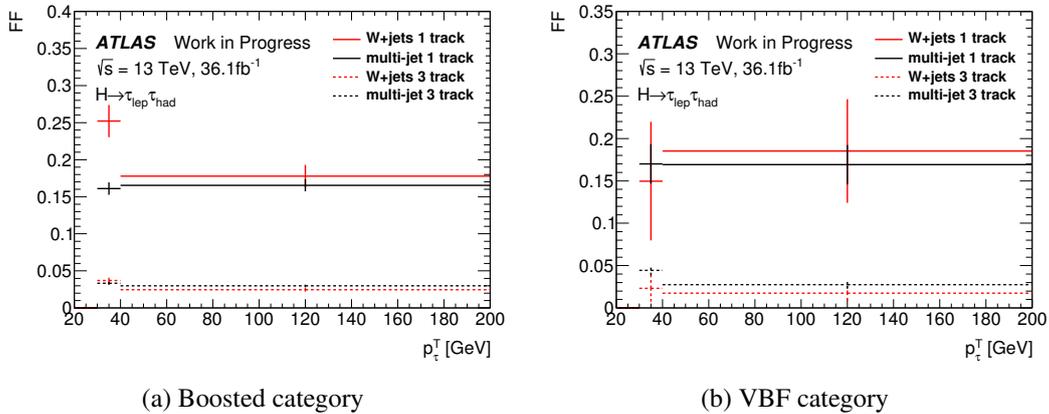


Figure 10.12: Individual "fake-factors" for the *Boosted* (left) and *VBF* (right) category. The dashed lines correspond to the results for hadronically decaying  $\tau$ -leptons with three tracks, while the full ones to  $\tau$ -leptons with one track. The red line refers to the relative contribution of W+jets events, while the black one belongs to multi-jet events.

### 10.2.4 Combined "fake-factors"

The combined "fake-factor" FF determined by:

$$FF = R_{WZt} \cdot FF_W + R_{\text{multi-jet}} \cdot FF_{\text{multi-jet}}, \quad (10.14)$$

depends on the same variables as the individual "fake-factors"  $FF_i$  and the relative contributions  $R_i$ . These are the event category (*Boosted*, *VBF*), the number of tracks of the  $\tau$ -lepton (1,3), the transverse momentum of the  $\tau$ -lepton ( $p_T$ ), the lepton flavour (electron, muon) and the difference of the azimuthal angle between the  $\tau$ -lepton and the missing transverse energy ( $\Delta\Phi$ ). The relative contributions  $R_i$  are binned in all of these variables, while the individual "fake-factors"  $FF_i$  are not binned in the lepton flavour and  $\Delta\Phi$ . Table 10.5 summarise the values for the combined "fake-factor" FF.

The combined "fake-factor" shows only a small dependence on  $\Delta\Phi$  and the lepton flavour. Since the individual "fake-factors" for W+jets and multi-jets processes (see Figure 10.12) have a similar size to each other, a variation of the relative contributions has only a small impact on the combined "fake-factor" FF. The FF is slightly higher for electrons than for muons and increases for higher  $\Delta\Phi$  values in the electron channel while the muon channel shows the opposite trend.

The number of tracks of the  $\tau$ -lepton has a large impact on the combined FF.  $\tau$ -leptons with 3 tracks have higher "fake-factor" than  $\tau$ -leptons with 1 track. Thus jets have a higher probability to get misidentified as a  $\tau$ -lepton with 3 tracks.

The FF is slightly higher for the *Boosted* category than for the *VBF* category and shows a decrease for higher  $p_T$  values in the *Boosted* category for the electron channel and the opposite trend in the muon channel. In the *VBF* category the FF shows no obvious trend for increasing  $p_T$  values.

		$p_T$ [GeV]	Electron				Muon			
			$\Delta\Phi < 1.0$	$\Delta\Phi < 1.5$	$1.5 < \Delta\Phi < 2.0$	$\Delta\Phi > 2.0$	$\Delta\Phi < 1.0$	$1.0 < \Delta\Phi < 1.5$	$1.5 < \Delta\Phi < 2.0$	$\Delta\Phi > 2.0$
1 track	Boosted	$p_T < 40$	0,220	0,226	0,229	0,241	0,200	0,176	0,181	0,169
		$40 < p_T < 90$	0,194	0,196	0,180	0,187	0,177	0,176	0,175	0,176
		$90 < p_T$	0,175	0,176	0,175	0,177	0,202	0,241	0,248	0,238
	VBF	$p_T < 40$	0,150				0,155			
		$40 < p_T < 90$	0,180				0,151			
		$90 < p_T$	0,180				0,178			
3 track	Boosted	$p_T < 40$	0,037	0,037	0,037	0,037	0,026	0,025	0,025	0,024
		$40 < p_T < 90$	0,027	0,027	0,026	0,026	0,025	0,025	0,025	0,025
		$90 < p_T$	0,025	0,025	0,025	0,025	0,034	0,036	0,037	0,036
	VBF	$p_T < 40$	0,020				0,024			
		$40 < p_T < 90$	0,020				0,019			
		$90 < p_T$	0,020				0,018			

Table 10.5: Combined "fake-factor" FF separated into the number of tracks of the  $\tau$ -lepton (1,3), the event category (*Boosted*, *VBF*) and the transverse momentum of the  $\tau$ -lepton ( $p_T$ ) shown on the left hand-side and the lepton flavour (electron, muon) and the difference of the azimuthal angle between the  $\tau$ -lepton and the missing transverse energy ( $\Delta\Phi$ ) shown on the top. The relative contributions  $R_i$  are binned in all of these variables, while the individual "fake-factors"  $FF_i$  are not binned in the variables on the top. The background color changes from blue to red for increasing FF values.

### 10.3 Further studies for the "fake-factor" method

This chapter includes several studies to justify approximations and simplifications, which are applied in the "fake-factor" method. Additionally closure tests for the validation of the "fake-factor" method will be shown. From these tests and studies some systematic uncertainties will be derived. These uncertainties will then be discussed more precisely in the next chapter.

#### 10.3.1 Use of $W + jets$ "fake-factor" for all non multi-jet events

One applied simplification in the "fake-factor" method is to use the individual  $W+jets$  "fake-factor" for all non-multi jet processes. This means that instead of calculating three different individual "fake-factors" for  $W+jets$ ,  $Z+jets$  and top quark events only the one for  $W+jets$  is determined. Instead of the relative contribution of  $W+jets$   $R_W$  the sum over all non-multi jet relative contributions of these processes  $R_{WZt}$  is used for  $FF_W$ :

$$R_{WZt} = R_W + R_Z + R_{top}, \quad (10.15)$$

This assumes that the difference between the "fake-factors" of the three processes is negligible. To prove this assumption a comparison is done between the number of "fakes" for the default "fake-factor" method and a "fake-factor" method which uses all non multi-jet processes.

In addition to the individual "fake-factors" of  $W+jets$  and multi-jet events, the individual "fake-factors" for  $Z+jets$  and top quark events have to be estimated. For this two additional control regions have to be defined. One requirement of the signal region is inverted for each region, to enrich these control regions with the corresponding process.

For the control region of top quark events instead of the b-jet veto a b-jet tag is set, since top quarks production events have mostly a bottom quark included (see Chapter 5). For the  $Z+jets$  control region two instead of one lepton are required.

Region	CR <sub>i</sub> = SR, except	Proportion of events							
		<i>Boosted</i>				<i>VBF</i>			
		W+jets	Top	Z+jets	Multi-jet	W+jets	Top	Z+jets	Multi-jet
W+jets	$m_T > 70$ GeV	0.84	0.06	0.05	0.04	0.81	0.07	0.09	0.03
Top	$N_{b-jets} > 0$	0.28	0.55	0.02	0.15	0.34	0.46	0.04	0.26
Z+jets	$N_{lep} = 2$	0.05	0.04	0.87	0.04	0.03	0.04	0.92	0.00
Multi-jet	No lepton isolation	0.02	0.01	0.00	0.97	0.02	0.00	0.01	0.97

Table 10.6: Definitions of the control regions and purities of the four different backgrounds  $W+jets$ , top quark production,  $Z+jets$  and multi-jet processes in the "anti- $\tau$ " control regions of the *Boosted* and *VBF* category.

The requirements and the purities of the different "anti- $\tau$ " control regions are summarized in Table 10.6. The  $W+jets$ ,  $Z+jets$  and multi-jet control regions have a high purity of the corresponding background process with at least 82% in both event categories. The purity of the top-quark production in the corresponding control region is with 55% in the *Boosted* category and 46% in the *VBF* category smaller than for the other processes, but still the

highest contribution in these regions. The relative contributions of the three non-multi jet processes have to be calculated individually. For this, the relative contributions of these processes are calculated with simulated events. Since the data-driven calculation of the multi-jet contribution is more confident, the results are re-weighted with the value of the data driven method:

$$R_i = R_{i, MC} \cdot \frac{1 - R_{\text{multi-jet}}}{\sum_j R_{j, MC}}, \quad (10.16)$$

where  $R_i$  is the used relative contribution of the process  $i$ ,  $R_{i, MC}$  the relative contribution of process  $i$  calculated from simulated events and  $R_{\text{multi-jet}}$  the relative contribution of multi-jet processes calculated from the data. The index  $i$  corresponds to the three non multi-jet processes W+jets, Z+jets and top quark.

		Number of tracks	Electron						Muon					
			1			3			1			3		
			Process	W+jets	Top	Z+jets	W+jets	Top	Z+jets	W+jets	Top	Z+jets	W+jets	Top
Boosted	$p_T < 40$	$\Delta\Phi < 1.0$	-0,09	0,00	-0,02	0,13	0,01	0,02	0,10	0,01	0,00	0,19	0,01	0,01
		$1.0 < \Delta\Phi < 1.5$	-0,04	-0,01	0,00	-0,02	0,00	0,00	0,32	0,04	0,01	0,16	0,03	0,00
		$1.5 < \Delta\Phi < 2.0$	-0,07	-0,01	0,00	0,00	0,00	0,00	0,12	0,02	0,01	-0,02	0,00	0,00
		$\Delta\Phi > 2.0$	0,19	0,04	0,00	-0,02	0,00	0,00	0,10	0,01	0,00	0,10	0,01	0,00
	$40 < p_T < 90$	$\Delta\Phi < 1.0$	0,12	0,01	0,01	0,12	0,01	0,02	0,02	0,00	0,00	0,15	0,01	0,01
		$1.0 < \Delta\Phi < 1.5$	0,08	0,01	0,01	0,11	0,02	0,01	0,13	0,01	0,00	0,10	0,02	0,00
		$1.5 < \Delta\Phi < 2.0$	-0,20	-0,01	-0,01	-0,04	-0,01	0,00	0,20	0,42	0,11	0,00	0,00	0,00
		$\Delta\Phi > 2.0$	-0,05	0,00	0,00	0,14	0,02	0,00	0,26	0,02	0,01	0,02	0,00	0,00
	$90 < p_T$	$\Delta\Phi < 1.0$	0,08	0,00	0,01	0,12	0,01	0,01	-0,16	-0,01	0,00	0,06	0,01	0,00
		$1.0 < \Delta\Phi < 1.5$	-0,13	-0,01	-0,01	0,07	0,00	0,00	-0,16	0,00	-0,01	0,10	0,00	0,00
		$1.5 < \Delta\Phi < 2.0$	-0,05	0,00	0,00	-0,05	0,00	0,00	0,22	0,01	0,00	0,11	0,01	0,00
		$\Delta\Phi > 2.0$	0,09	0,00	0,00	-0,03	0,00	0,00	0,37	0,01	0,01	0,04	0,00	0,00
VBF	$p_T < 40$	$\Delta\Phi < 2.0$	-0,20	-0,01	-0,04	0,37	0,03	0,02	0,04	0,00	0,00	0,10	0,01	0,01
		$\Delta\Phi > 2.0$	-0,26	0,00	-0,01	-0,04	-0,01	-0,02	0,00	0,00	0,00	-0,16	-0,02	0,00
	$40 < p_T < 90$	$\Delta\Phi < 2.0$	0,08	0,01	0,04	0,13	0,01	0,03	0,01	0,00	0,00	0,20	0,02	0,01
		$\Delta\Phi > 2.0$	-0,41	-0,01	-0,01	0,12	0,02	0,15	-0,34	-0,01	0,00	-0,14	-0,01	0,00
	$90 < p_T$	$\Delta\Phi < 2.0$	0,27	0,02	0,10	-0,02	0,00	0,00	0,23	0,00	0,01	-0,06	0,00	0,00
		$\Delta\Phi > 2.0$	-0,34	-0,07	0,00	0,14	0,01	0,02	-0,02	0,00	0,00	0,00	0,00	0,00

Table 10.7: Difference between the re-weighted relative contributions  $R_i$  and the relative contribution calculated with simulated events  $R_{i, MC}$  ( $R_i - R_{i, MC}$ ) for W+jets, top quark and Z+jets processes. The relative contribution are binned into the light lepton flavour (electron, muon) and the number of tracks of the  $\tau$ -lepton (1,3) shown on the top and into the difference of the azimuthal angle of the  $\tau$ -lepton and the missing transverse energy ( $\Delta\Phi$ ), the transverse momentum of the  $\tau$ -lepton ( $p_T$ ) and the event categories (*Boosted*, *VBF*) shown on the left-hand side.

Table 10.7 shows the difference between the relative contribution calculated with simulated events and the relative contributions re-weighted with the data-driven calculation. In most of the bins only the relative contribution of W+jets processes is re-weighted while the re-weighting of the relative contributions of top quark production and Z+jets processes is negligible. The reason for this is the high relative contribution of the W+jets processes in

most of the bins. The precise relative contributions before and after the re-weighting are illustrated in Figure A.3 in the appendix.

The individual "fake-factors"  $FF_i$  calculated for all four background processes in the *Boosted* and *VBF* category are given in Figure 10.13. The figure shows especially for the *VBF* category large statistical uncertainty on the  $FF_i$  of the *Z*+jets and the top quark production background introduced by low statistics in the corresponding control regions (see Table A.1 in the appendix).

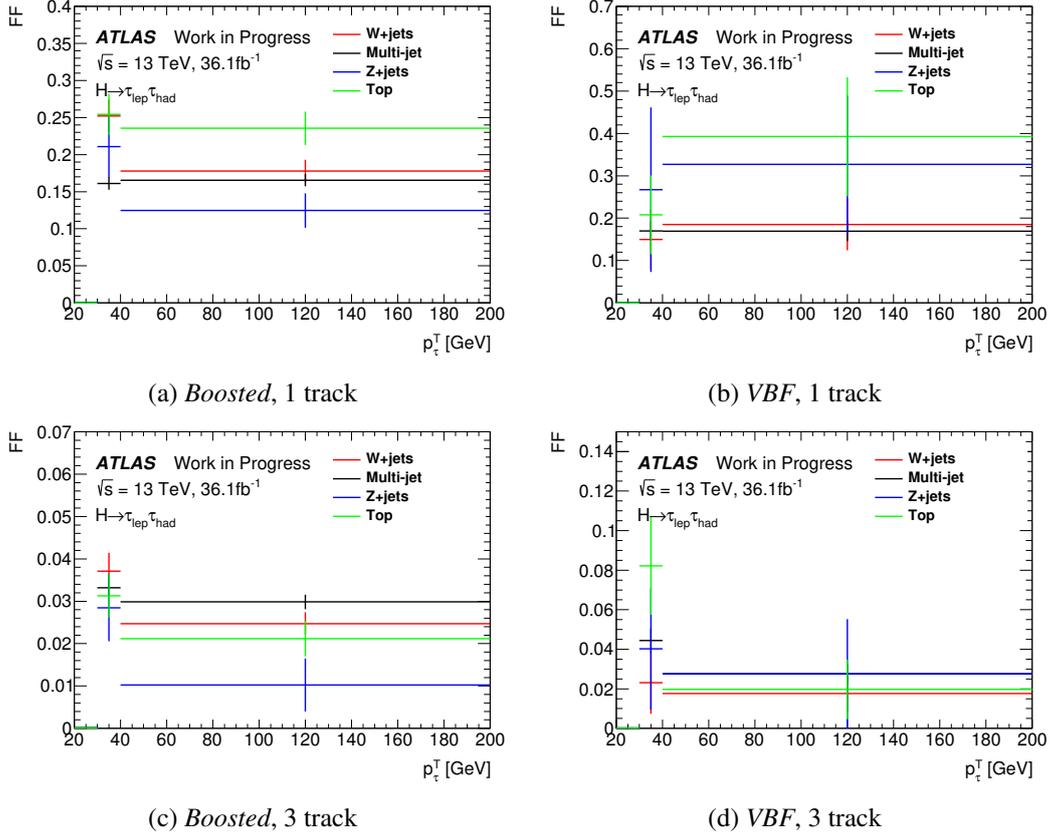


Figure 10.13: Individual "fake-factors" for the *Boosted* (left) and *VBF* (right) category, split into the number of tracks of hadronically decaying  $\tau$ -lepton (1 (top), 3 (bottom)), calculated for W+jets (red), multi-jet (black), Z+jets (blue) and top quark production processes (green).

These individual "fake-factors" and their relative contributions are used to estimate the number of "fake" events with the "fake-factor" method. This result is to be compared with the result of the default setup, which only uses the W+jets and multi-jet "fake-factor". The ratio between the number of "fakes" of the two methods as a function of the reconstructed di- $\tau$ -lepton mass  $m_{\tau\tau}^{MMC}$  is illustrated in Figure 10.14 for the *Boosted* category (left) and the *VBF* category (right). The red full lines show the result of the default method normalized to itself. Thus all values perfectly match the value one. The dashed red line refers to the full systematic and

statistical uncertainty on the default values. The systematic and statistical uncertainties are added in quadrature for this. The blue line corresponds to the "fake" events calculated with the set-up described in this section, normalized to the results of the default method. The *Boosted* category shows a deviation of less than 1% for the two methods and the *VBF* category of less than 5%. Thus in both categories this deviation is negligible in their uncertainties of around 10% for the *Boosted* category and 40% for the *VBF* category. Thus the simplification to use only the W+jets "fake-factor" can be applied without an additional uncertainty.

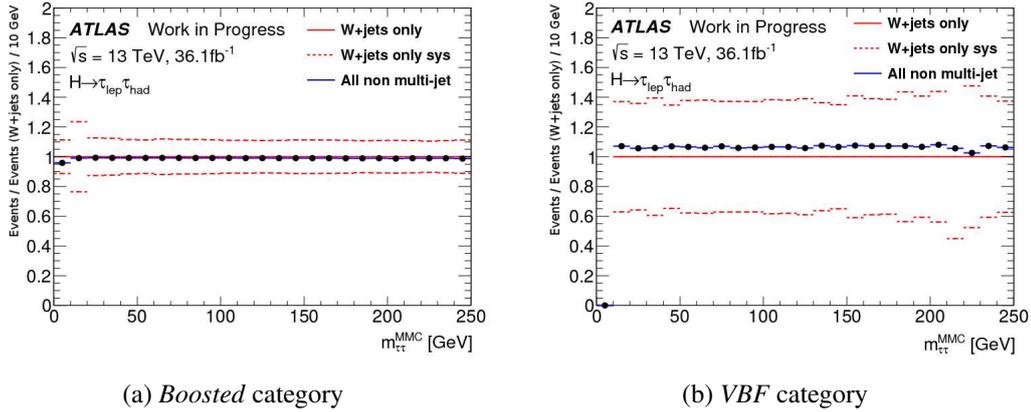


Figure 10.14: A comparison between the number of fake calculated with default "fake-factor" method (red) and the alternative version (blue) described in this section as a function of the reconstructed di- $\tau$ -lepton mass  $m_{\tau\tau}^{MMC}$  for the *Boosted* category (left) and the *VBF* category (right). For a better comparison both results are normalized to the default method. The dashed red line corresponds to the full uncertainty band, which is calculated by adding the systematic and the statistical uncertainties in quadrature.

### 10.3.2 2015 and 2016 combined "fake-factor"

The used dataset of the analysis is composed of a 2015 and a 2016 dataset. Between these two datasets settings of the ATLAS detector have changed, like the used triggers (see Chapter 5). This could have a potential influence on the "fake-factor". Since the dataset of 2015 has a small integrated luminosity ( $3.2 \text{ fb}^{-1}$ ) against the 2016 dataset ( $33.3 \text{ fb}^{-1}$ ) the difference between the two "fake-factors" could be negligible considering the big difference in the integrated luminosity. To prove this, the number of "fake" events is calculated in two different ways. Additionally to the set-up where one "fake-factor" is calculated for the combined dataset, the number of "fake" events is determined by using two individual "fake-factors" for the 2015 and 2016 dataset. The ratio between the number of "fake" events calculated with the separated datasets of 2015 and 2016 and the number of "fake" events with a combined dataset of 2015 and 2016 as a function of  $m_{\tau\tau}^{MMC}$  is illustrated in Figure 10.15 for the *Boosted* category (left) and the *VBF* category (right). In these distributions the number of "fake" events calculated with a combined dataset is shown in red and the number of "fake" events calculated with two separate "fake-factors" in blue. Both calculations are normalized to the default

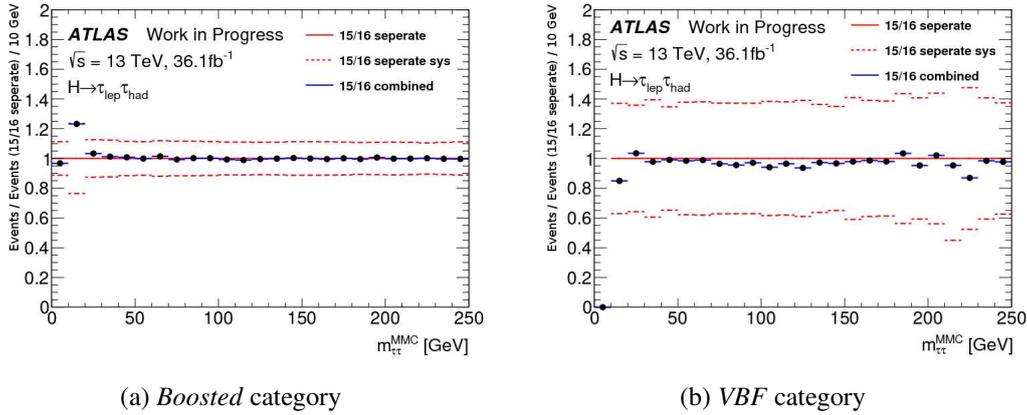


Figure 10.15: A comparison between the number of "fake" events calculated with a combined 2015 and 2016 dataset (red) and the number of "fake" events calculated separated for 2015 and 2016 (blue) as a function of  $m_{\tau\tau}^{MMC}$  for the *Boosted* category (left) and the *VBF* category (right). For a better comparison both results are normalized to the combined method. The dashed red line corresponds to the full uncertainty band of the default method.

method. The red dashed line refers the full uncertainty band for the default method, for which the systematic and the statistical uncertainty were added in quadrature. The *Boosted* category shows less than 2% deviation between the two methods, except of the three bin below 30 GeV, where a higher fluctuation is introduced due to low statistics. The *VBF* category show a higher relative deviation, up to 5%, except of two bins with a deviation of 10%-15%. However this higher deviation is covered by the high uncertainty band in the *VBF* category with  $\sim 40\%$ . Thus the simpler default method with a combined "fake-factor" for 2015 and 2016 can be used without adding an additional systematic uncertainty.

### 10.3.3 Use of the "same-sign" isolation factor for the "opposite-sign" region

The isolation factor calculated in the "anti- $\tau$ " multi-jet *Preselection* region, is applied to the "anti- $\tau$ " multi-jet *Boosted* region and to the "anti- $\tau$ " multi-jet *VBF* region (see Figure 10.8). Since the *Preselection* region includes the *Boosted* and *VBF* regions this means that the events on which the isolation factor is applied, are also used for its calculation. To solve this problem the isolation factor is instead calculated in the "same-sign" "anti- $\tau$ " multi-jet *Preselection* region, which is orthogonal to all regions in which the isolation factor is applied. This implies the assumption, that the isolation factor in the "same-sign" region and corresponding "opposite-sign" region has the same value. To prove this, the isolation factor is calculated in both regions and compared to each other.

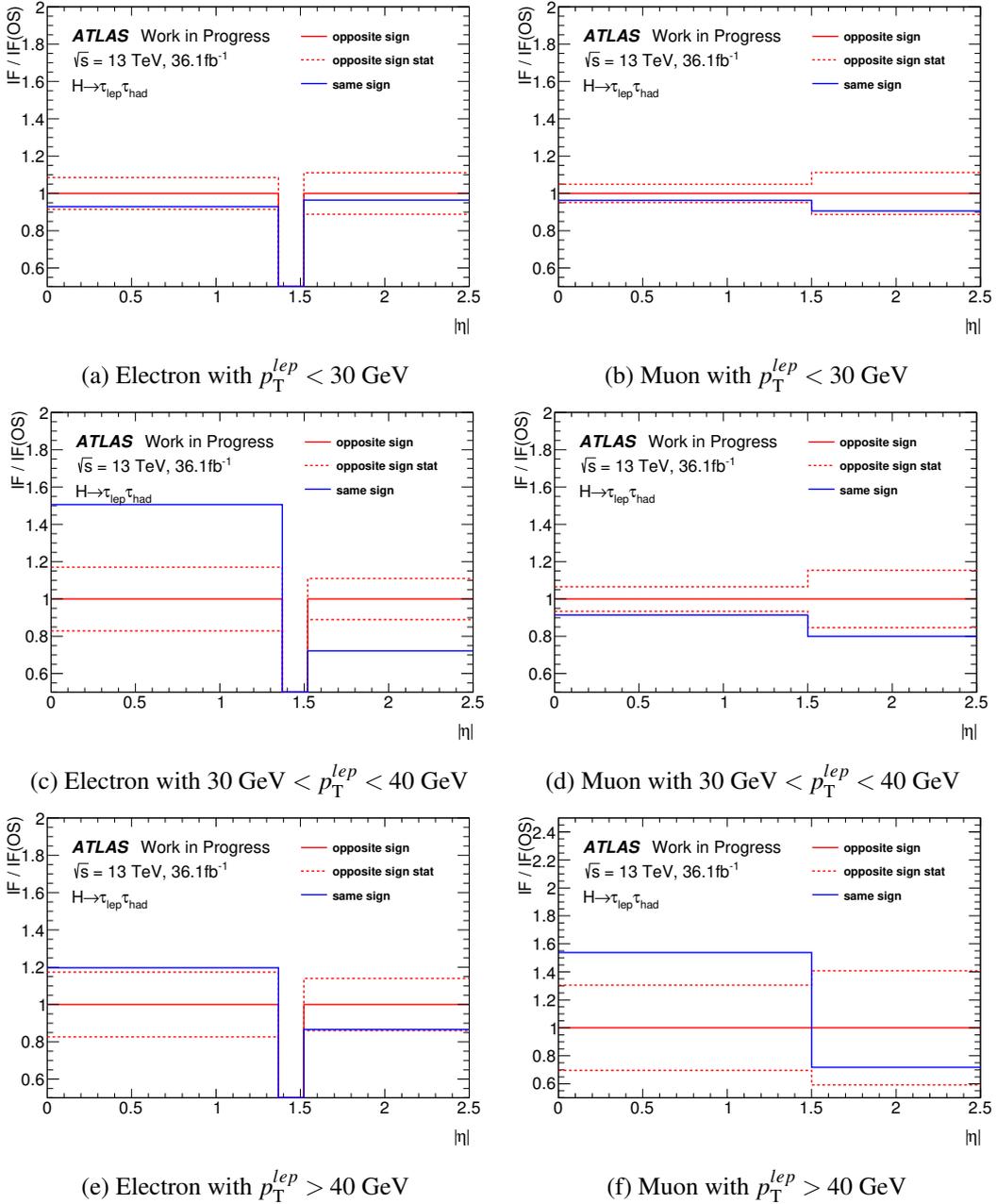


Figure 10.16: Ratio between the isolation factor calculated in the "opposite-sign" and in the "same-sign" region normalized to the "opposite-sign" values, for the electron channel (left) and muon channel (right). From top to bottom row the different cuts on the transverse momentum of the light lepton are applied. The red solid line refers to the "opposite-sign" result, while red dashed lines correspond to its statistical uncertainty. The "same-sign" calculation is illustrated in blue.

Table 10.8 shows the isolation factors calculated in the "same-sign" ("SS") and "opposite-sign" ("OS") *Preselection* region, split into the light lepton flavour (electron, muon), the transverse momentum of the light lepton  $p_T$  and the pseudorapidity of the light lepton ( $\eta$ ). Additionally the statistical uncertainties on the isolation factor of the "OS" region is given in the last column. The table illustrates that most of the isolation factors calculated in the "same-sign" region agree with their corresponding isolation factor of the "opposite-sign" region in its uncertainties.

A relative comparison of these two calculations is given in Figure 10.16. Both results are normalized to the "opposite-sign" calculation. This leads to a perfect agreement of the "opposite-sign" results (red) with one, while the "same-sign" values vary around one. The dashed red lines show the statistical uncertainties of the "opposite-sign" determination.

Figure 10.16 shows that in the most cases the difference between these two calculations is in the same magnitude as the statistical uncertainty. To cover this fact an additional systematic uncertainty is applied on the isolation factor (see Section 10.4.1). This uncertainty is calculated by the absolute difference between the "same-sign" value and corresponding value of the "opposite-sign" determination. This is done for each bin separately and added in quadrature to the other uncertainties.

$p_T$ [GeV]	electron						muon					
	$p_T < 30$		$30 < p_T < 40$		$40 < p_T$		$p_T < 30$		$30 < p_T < 40$		$40 < p_T$	
$ \eta $	$< 1.37$	$> 1.52$	$< 1.37$	$> 1.52$	$< 1.37$	$> 1.52$	$< 1.5$	$> 1.5$	$< 1.5$	$> 1.5$	$< 1.5$	$> 1.5$
"OS" IF	0.46	1.05	0.47	1.36	0.77	2.05	0.72	0.50	0.68	0.54	0.03	0.06
"SS" IF	0.43	1.02	0.71	0.98	0.92	1.77	0.69	0.46	0.63	0.43	0.04	0.04
"OS" stat.	0.04	0.12	0.08	0.15	0.13	0.29	0.04	0.06	0.05	0.08	0.01	0.02

Table 10.8: Isolation factor calculated in the "same-sign" ("SS") and "opposite-sign" ("OS") *Preselection* region, split into the light lepton flavour (electron, muon), the transverse momentum of the light lepton  $p_T$  and the pseudorapidity of the light lepton ( $\eta$ ), and the statistical uncertainty (stat.) on the "OS" isolation factor in the last row.

### 10.3.4 Backgrounds in the control regions for the Isolation Factor (IF)

This section discusses properties of the control regions, which are used for the calculation of the isolation factor. Figure 10.17 shows distributions for "same-sign" "anti- $\tau$ " signal regions split into the light lepton flavour.

In this region the numerator of the isolation factor is calculated by the number of data events subtracting events with real light leptons (see Equation 10.13). For the determination of this background with real leptons simulations are used. In the figure the data is illustrated by the black points and the background from real light leptons is shown by the colored histograms. This background has a big impact on the calculation. Especially in the high transverse momentum region for the muon channel this background is of the same order as the data. Thus the impact of a mismodelling in the simulation is high on the isolation factor. To cover potential mismodelling an additional systematic error is applied on the isolation factor. The number of events of the simulated background is varied by 35%. This value corresponds to

the average uncertainty on the main backgrounds simulations with all systematics evaluated. Since negative isolation factors have no physical meaning the isolation factor is set to zero if the variation results in a negative value for the isolation factor. The influence of this uncertainty on the isolation factor is further studied in Section 10.4.1.

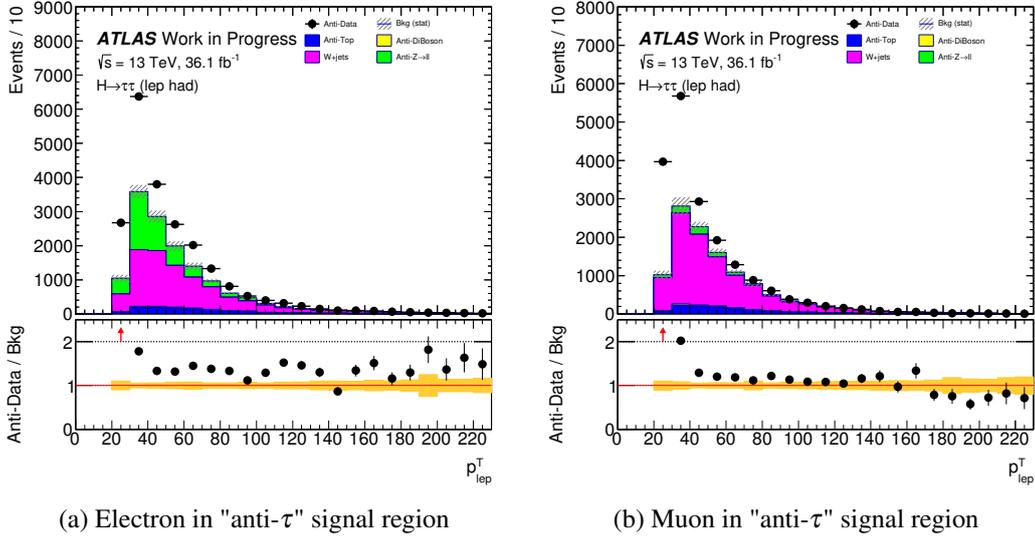


Figure 10.17: Transverse momentum of the light lepton in the "same-sign" "anti- $\tau$ " signal region, which is used for the isolation factor calculations separated into the electron (left) and muon (right) channel. Data is shown by black dots and the background by histograms.

### 10.3.5 "Same-sign" closure test

To validate the "fake-factor" method a "same-sign" closure test is used. Since the physical behaviour should be the same for the "same-sign" region as for the "opposite-sign" region the "fake-factor" method should work in both cases. The benefit of the "same-sign" region is that this region has no signal events and is dominated by "fake" events. The event yields for signal and the different backgrounds are given in Table 10.9. For this test the whole "fake-factor" calculation (see Chapter 10) is redone in the "same-sign" region.

This means that in all used control regions the requirement of the opposite charge sign for the light lepton and the hadronically decaying  $\tau$ -lepton is inverted. To keep the method consistent, the isolation factor for the "same-sign" closure test is calculated in the "opposite-sign" regions instead.

The individual "fake-factors" calculated in the "same-sign" region are illustrated in Figure 10.18. A comparison to the "fake-factors" of the "opposite-sign" calculation (see Figure 10.12) shows that the "fake-factors" for  $W$ +jets processes are larger for the "same-sign" calculation, while the ones for multi-jet processes are smaller than the ones of the "opposite-sign" calculation. The relative contribution for the "same-sign" region are shown in Figure A.2 and A.3 in the appendix and the precise values are given in Table A.2 in the appendix.

Table 10.10 summarizes the combined "fake-factors" for the "same-sign" calculation. The combined "fake-factors" are smaller compared to the ones of the "opposite-sign" (see Table 10.5). Due to the higher difference between the individual "fake-factors" in the "same-sign" region of W+jets and multi-jet processes the combined "fake-factor" has a higher dependence on the electron flavour and the difference in the azimuthal angle  $\Delta\Phi$ , in which only the relative contributions are binned.

Category	Data	Signal	Fake	$Z \rightarrow \tau\tau$	$Z \rightarrow ll$	Top	Di-boson
Boosted	1187	$1.2 \pm 0.1$	$1106 \pm 12$	$70 \pm 7$	$13 \pm 8$	$19 \pm 3$	$17 \pm 2$
Boosted high	295	$0.6 \pm 0.1$	$235 \pm 6$	$45 \pm 6$	$2 \pm 1$	$7 \pm 2$	$8 \pm 2$
Boosted low	892	$0.5 \pm 0.1$	$771 \pm 10$	$25 \pm 5$	$11 \pm 8$	$12 \pm 2$	$9 \pm 1$
VBF	102	$0.3 \pm 0.1$	$123 \pm 4$	$3 \pm 1$	$0.8 \pm 0.6$	$1.0 \pm 0.6$	$0.8 \pm 0.3$
VBF tight	49	$0.2 \pm 0.1$	$57 \pm 3$	$0.8 \pm 0.6$	$0.0 \pm 0.7$	$0.6 \pm 0.5$	$0.6 \pm 0.3$
VBF loose	53	$0.1 \pm 0.1$	$66 \pm 3$	$2 \pm 1$	$1 \pm 1$	$0.3 \pm 0.3$	$0.1 \pm 0.1$

Table 10.9: Events yields of data, signal and backgrounds in the different "same-sign" event selections.

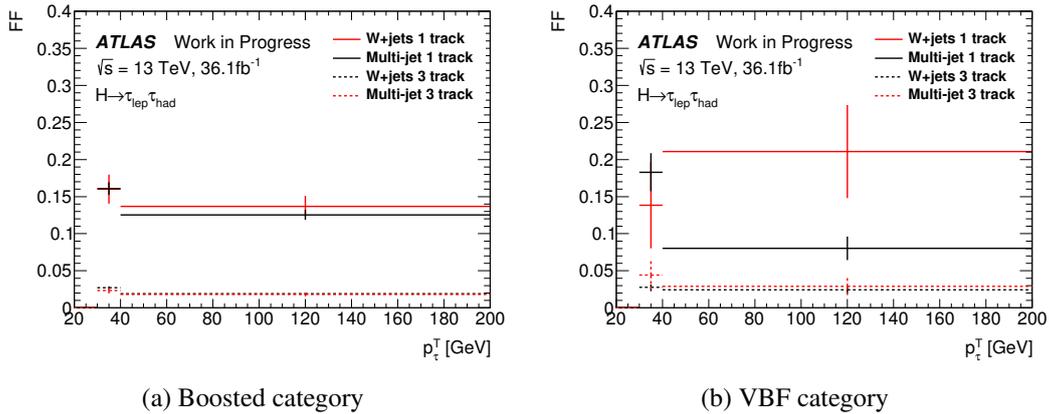


Figure 10.18: Individual "fake-factors" calculated in the "same-sign" *Boosted* (left) and "same-sign" *VBF* (right) category. The dashed lines correspond to the results for hadronically decaying  $\tau$ -leptons with three tracks and the full ones to  $\tau$ -leptons with one track. The red line refers to the relative contribution of W+jets events, while the black one belongs to multi-jet events.

Figure 10.19 illustrates the results for the inclusive and exclusive signal regions. Here the black dots correspond to the data again, while the colored histograms refer to the background. As already mentioned these regions are highly dominated by "fake" events (green). All signal categories show a good agreement between data and background in this "same-sign" closure test. Thus the closure test validates the "fake-factor" method.

		$p_T$ [GeV]	Electron				Muon			
			$\Delta\Phi < 1.0$	$\Delta\Phi < 1.5$	$1.5 < \Delta\Phi < 2.0$	$\Delta\Phi > 2.0$	$\Delta\Phi < 1.0$	$1.0 < \Delta\Phi < 1.5$	$1.5 < \Delta\Phi < 2.0$	$\Delta\Phi > 2.0$
1 track	Boosted	$p_T < 40$	0,160	0,160	0,160	0,161	0,141	0,148	0,153	0,166
		$40 < p_T < 90$	0,170	0,127	0,126	0,166	0,134	0,134	0,136	0,136
		$90 < p_T$	0,134	0,132	0,135	0,133	0,136	0,162	0,149	0,144
	VBF	$p_T < 40$	0,150			0,160	0,200			0,249
		$40 < p_T < 90$	0,180			0,200	0,180			0,211
		$90 < p_T$	0,180			0,148	0,110			0,097
3 track	Boosted	$p_T < 40$	0,027	0,027	0,027	0,027	0,019	0,020	0,020	0,020
		$40 < p_T < 90$	0,020	0,019	0,019	0,020	0,019	0,019	0,019	0,019
		$90 < p_T$	0,019	0,019	0,019	0,019	0,025	0,026	0,026	0,026
	VBF	$p_T < 40$	0,020			0,148	0,020			0,224
		$40 < p_T < 90$	0,020			0,213	0,020			0,211
		$90 < p_T$	0,020			0,189	0,026			0,125

Table 10.10: Combined "fake-factor" FF for the "same-sign" calculation separated into the number of tracks of the  $\tau$ -lepton (1,3), the event category (*Boosted*, *VBF*) and the transverse momentum of the  $\tau$ -lepton ( $p_T$ ) shown on the left hand-side and the lepton flavour (electron, muon) and the difference of the azimuthal angle between the  $\tau$ -lepton and the missing transverse energy ( $\Delta\Phi$ ) shown on the top. The relative contributions  $R_i$  are binned in all of these variables, while the individual "fake-factors"  $FF_i$  are not binned in the variables on the top. The background color changes from blue to red for increasing FF values.

### 10.3.6 Closure test for individual "fake-factors"

An assumption of the "fake-factor" method is that individual "fake-factors" in the control regions can be transferred to the signal region. To validate this assumption a further closure test is used.

For this test the "fake-factors" are calculated in the corresponding control region but instead of data simulated events of the corresponding process are used. These "fake-factors" are applied to the simulated events in the "anti- $\tau$ " signal region and compared with the actual number of simulated events in the  $\tau$  signal region. This means that, for example the W+jets background in the signal region calculated by simulation, is compared to W+jets background propagated from the "anti- $\tau$ " control region to the  $\tau$  signal region via the "fake-factor" method. Figure 10.20 illustrates the result of this closure test for the W+jets process. The black dots correspond to W+jets events from simulated events, while the green histograms refer to the calculated events via the "fake-factor" method. This test is done for the two inclusive as well as for the four exclusive signal categories. All categories show a good agreement between the "fake-factor" method and the simulation. Thus this test can validate the transfer of "fake-factors" from the control region to signal region for the W+jets process.

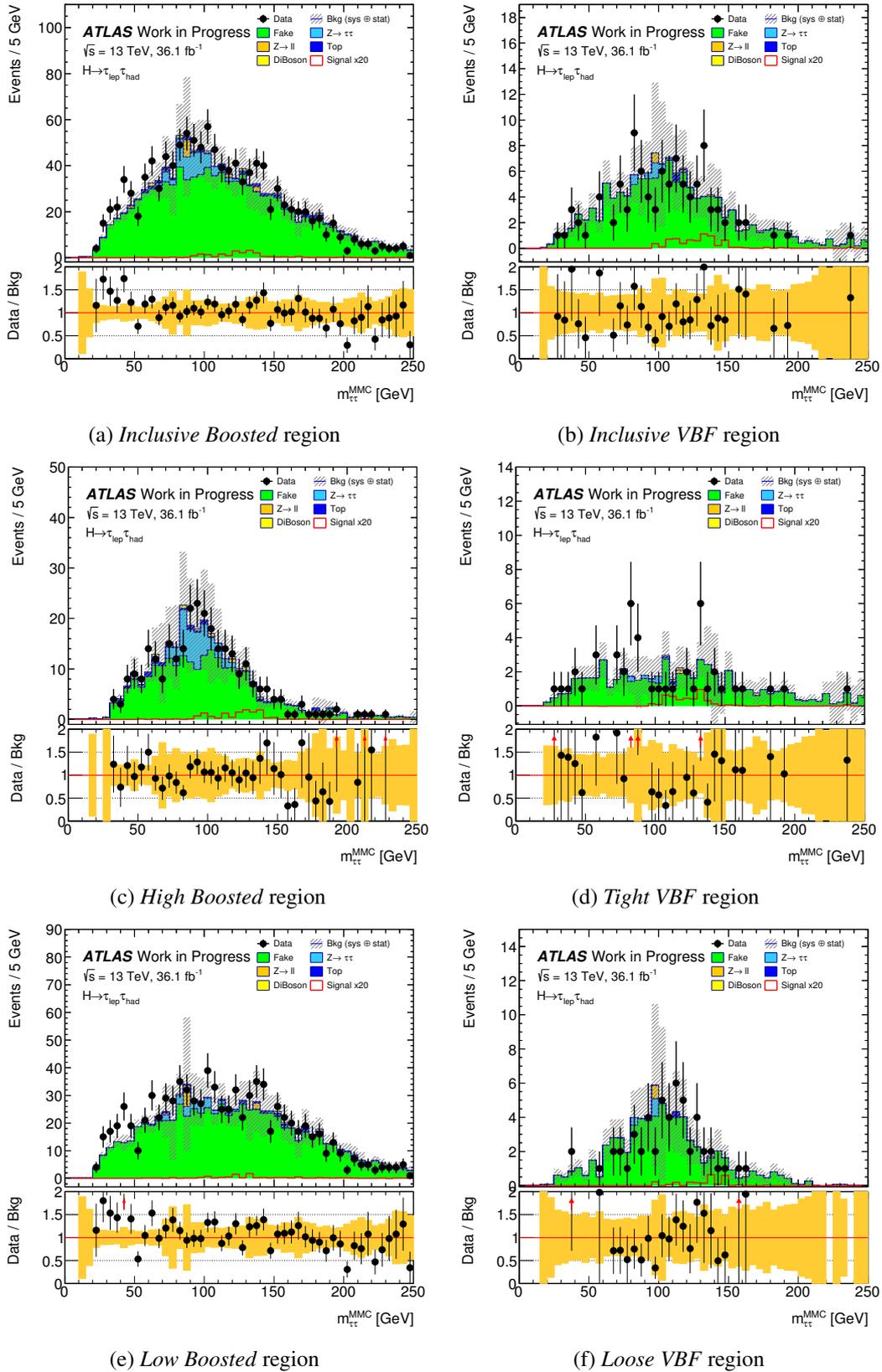


Figure 10.19: Results of the closure test as a function of  $m_{\tau\tau}^{MMC}$  for the *Boosted* (left) and *VBF* (right) "same-sign" signal regions. The first row shows the *inclusive* signal categories, the second one the *high* and the *tight* categories and the last row the *low* and *loose* categories. The data is shown by black dots and the background by the colored histograms. The background uncertainties include the statistical and systematic uncertainties added in quadrature.

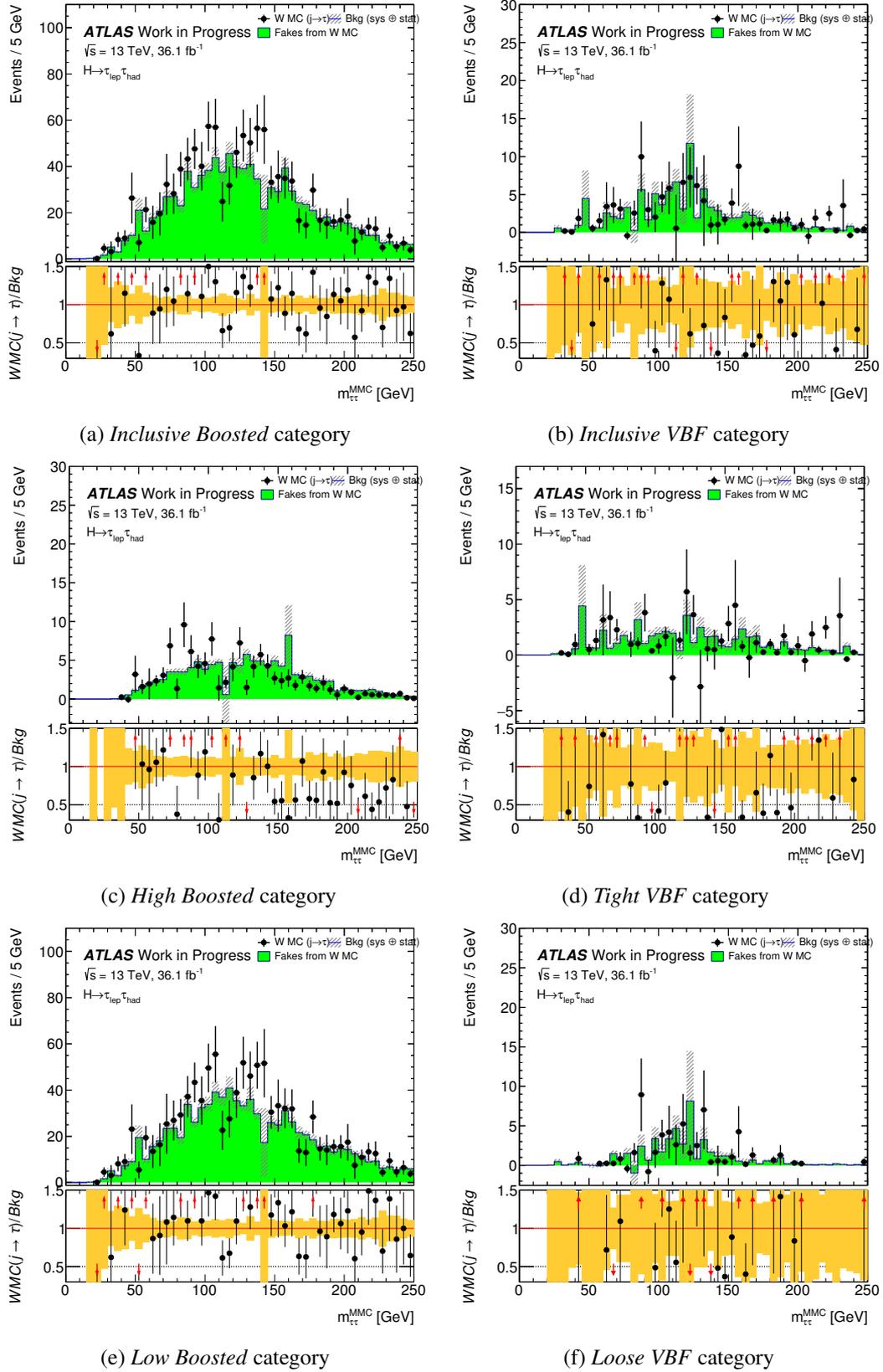


Figure 10.20: Results of the closure test for individual "fake-factors" as a function of  $m_{\tau\tau}^{MMC}$  for the *Boosted* (left) and *VBF* (right) signal regions. The first row shows the *inclusive* signal categories, the second one the *high* and the *tight* categories and the last row the *low* and *loose* categories. The data is shown by black dots and the background by the colored histograms. The background uncertainties include the statistical and systematic uncertainties added in quadrature.

## 10.4 Summary of systematic uncertainties

To give a summary of the systematic uncertainties of the estimated "fake" events all assumptions are repeated once again:

1. The individual "fake-factors" calculated in the corresponding control region can be transferred to the signal region (Section 10.3.6).
2. All non multi-jet processes can be described by using only the W+jets "fake-factor" (Section 10.3.1).
3. For the 2015 and 2016 datasets a combined "fake-factor" can be calculated (Section 10.3.2).
4. The isolation factor calculated in the "same-sign" region can be applied to the "opposite-sign" region (Section 10.3.3).

The assumptions 1-3 have shown a negligible influence on the estimated number of "fake" events so that for those no systematic uncertainties are applied and they will not be discussed in this section any further. In addition to these assumptions further sources of systematics uncertainties have to be considered:

1. The different control regions have statistical uncertainties.
2. The simulated background has a high influence on the isolation factor (Section 10.3.4).

### 10.4.1 Isolation factor

The three sources of uncertainties on the isolation factor IF are the statistical uncertainties in the control region, the difference between the isolation factor in the "same-sign" and "opposite-sign" region (Section 10.3.3) and the effect of varying the subtracted simulated events by 35% (Section 10.3.4). All uncertainties are shown in Figure 10.21.

The figure illustrates that the dominant uncertainty is the yield uncertainty on the simulated events. In most of the bins the "opposite-sign-same-sign" uncertainties ("OS-SS") and the statistical uncertainty have a similar size.

To propagate these uncertainties to a final number of "fakes" for each bin of the isolation factor a combined uncertainty  $s_{com}$  is calculated:

$$s_{com} = \sqrt{s_{stat}^2 + s_{OS-SS}^2 + s_{Yield}^2}, \quad (10.17)$$

where  $s_{stat}$  is the statistical uncertainty,  $s_{OS-SS}$  the "same-sign" "opposite-sign" uncertainty and  $s_{Yield}$  the uncertainty from the background variation. The precise values of the uncertainties and the combined uncertainty for each bin are summarized in Table 10.11.

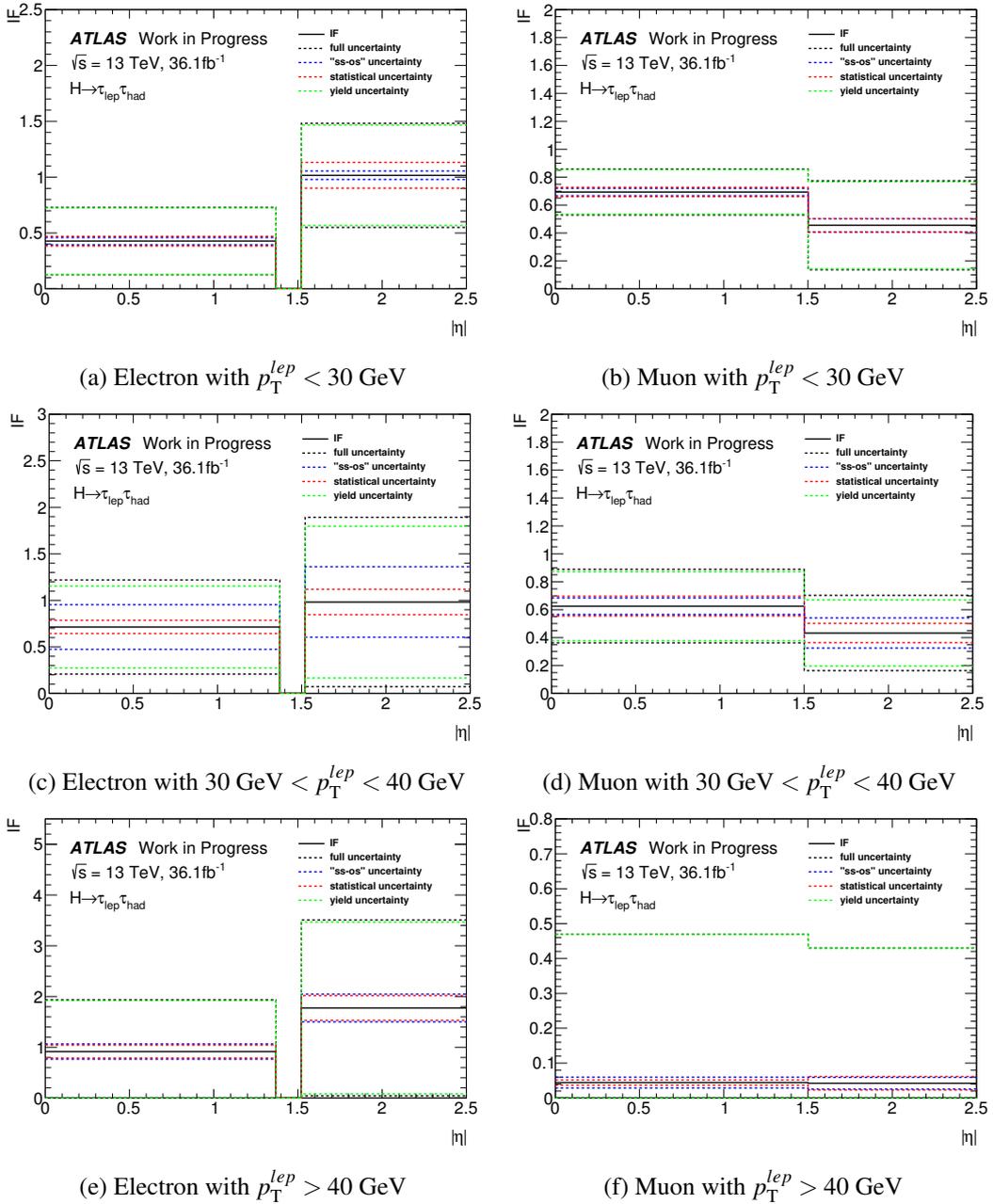


Figure 10.21: Comparison between the different uncertainties on the isolation factor. The black line corresponds to the nominal value and the dashed black line to the full uncertainties. The red line refers to the statistical uncertainties in the control region, the blue line to the uncertainty of the transfer of the "same-sign" isolation factor to the "opposite-sign" region and the green one to the uncertainty, which is obtained through the yield variation of the simulated events. The distributions are shown for the electron channel (left) and the muon channel (right) separately. The different rows shows different transverse momentum ranges for the light lepton ( $p_T^{lep} < 30$  GeV,  $30 \text{ GeV} < p_T^{lep} < 40$  GeV,  $p_T^{lep} > 40$  GeV).

Channel	$p_T^{lep}$ [GeV]	$ \eta $	Nominal	Stat.	"SS-OS"	Yield	Combined
Muon	$p_T^{lep} < 30$	$< 1.50$	0.69	0.03	0.03	0.16	0.16
		$> 1.50$	0.46	0.05	0.05	0.31	0.32
	$30 < p_T^{lep} < 40$	$< 1.50$	0.63	0.07	0.06	0.25	0.26
		$> 1.50$	0.43	0.07	0.11	0.24	0.27
	$40 < p_T^{lep}$	$< 1.50$	0.04	0.01	0.02	0.43	0.43
		$> 1.50$	0.04	0.02	0.02	0.39	0.39
Electron	$p_T^{lep} < 30$	$< 1.37$	0.43	0.04	0.03	0.30	0.30
		$> 1.52$	1.02	0.12	0.04	0.45	0.47
	$30 < p_T^{lep} < 40$	$< 1.37$	0.71	0.07	0.24	0.44	0.51
		$> 1.52$	0.98	0.14	0.38	0.82	0.91
	$40 < p_T^{lep}$	$< 1.37$	0.92	0.13	0.15	1.00	1.02
		$> 1.52$	1.77	0.24	0.27	1.69	1.73

Table 10.11: The statistical uncertainties (stat.), yield uncertainties (yield), "same-sign"- "opposite-sign" uncertainties ("SS-OS") and the combined uncertainties on the isolation factors, binned in the light lepton flavour (muon, electron), in the transverse momentum of the light lepton ( $p_T^{lep}$ ) and the pseudorapidity of the light lepton ( $\eta$ ).

#### 10.4.2 Relative contribution

For the relative contributions of multi-jet events  $R_{multi-jet}$  and W+jets events  $R_W$  the two uncertainties are the uncertainty of the used isolation factor and the statistical uncertainties in the remaining control regions, which are used for the calculation. These two uncertainties will be propagated separately to the final calculation of the "fake" events.

The statistical uncertainties of the control regions for the relative contributions are propagated to the calculation of the "fake" events. For simplicity these uncertainties treated as correlated across bins in the same category. This is a conservative assumption and significantly simplifies the uncertainty propagation.

The relative contribution of the W+jets process  $R_W$  is directly calculated from the relative contribution of the multi-jet processes  $R_{multi-jet}$  (see Equation 10.10). It follows from this linear dependency that the up variation of  $R_{multi-jet}$  leads to down variation of  $R_W$ . Additionally the statistical uncertainty of  $R_{multi-jet}$  is propagated to  $R_W$ .

#### 10.4.3 Individual "fake-factor"

For each individual "fake-factor" the uncertainty is calculated from the statistical uncertainties of the used control region through standard error propagation. For these factors no further uncertainties are applied. These uncertainties are treated as correlated across the bins in the same category.

#### 10.4.4 Summary

In Table 10.12 all uncertainties for the calculation of the number of "fake" events are summarized. The yields of both inclusive signal categories are given for the nominal case and for all different uncertainty variations. In addition the relative difference between the nominal value and the variation is given. The combined uncertainty of all three uncertainties is  $\pm 34\%$  for the *VBF* category and  $\pm 9\%$  for the *Boosted* category. The statistical uncertainties on the individual "fake-factors" have the biggest impact with  $\pm 33\%$  for the *VBF* category and  $\pm 8\%$  for the *Boosted* category. The low impact of the uncertainties of the relative contribution ( $\pm 8\%$  for the *VBF* category and  $\pm 3\%$  for the *Boosted* category) and the isolation factor ( $^{+3\%}_{-1\%}$  and  $\pm 3\%$ ) can be explained with Figure 10.12. These distributions show that the two individual "fake-factors" have similar values in the bins. Due to this fact, a variation of the relative contribution has only a small impact on the final estimated number of "fake" events. Since the uncertainty variation of the isolation factor only leads to variation of the relative contribution the influence on the number of "fake" events is also small.

	VBF	Boosted	VBF	Boosted
	Absolute yield		Rel. diff. to nom (%)	
Nominal yield	150	2123		
Individual FF stat. unc. (up)	201	2295	33	8
Individual FF stat. unc. (down)	100	1952	-33	-8
IF variation (up)	155	2067	3	-3
IF variation (down)	148	2185	-1	3
Rel. contribution stat. unc. (up)	163	2069	8	3
Rel. contribution stat. unc. (down)	138	2178	-8	-3

Table 10.12: Summary of all uncertainties on the number of "fake" events calculated with the combined 2015 and 2016 dataset with a integrated luminosity of  $36\text{fb}^{-1}$ .



## 11 SIMPLIFIED TEMPLATE CROSS SECTIONS

The first study of Simplified Template Cross Sections (STXS) in the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel is an important part of this thesis. A description of the underlying STXS method is given in Section 4.4. On the experimental side the same signal categories and objects definitions, simulated events and background estimations as in the coupling strength analysis are used for the STXS study.

### 11.1 STXS bin definition and expected events yields

A study of the event yields for the STXS bins of the Stage 0 and in particular Stage 1 is done first. The study gives an overview of the expected event yields in the different STXS bins and the contributions of the different signal event categories to STXS bins as well it discusses the binning choice, which is based on the suggestions in Section 4.4.

#### 11.1.1 Stage 0

In Table 11.1 the expected events yields in  $36 \text{ fb}^{-1}$  for the Stage 0 STXS bins are given. The STXS bins are listed from left to right, while the different signal event categories of the analysis are given from top to bottom. As expected most of the events in the signal categories contribute to the  $ggF$  bin, followed by the  $VBF$  bin and the  $VH$  bin, with a hadronically decaying vector boson.

		Stage 0								
		ggF	VBF	VH (had)	qq $\rightarrow$ WH (lep)	qq $\rightarrow$ ZH (lep)	gg $\rightarrow$ ZH (lep)	bbH	ttH	tH
Boosted	High	68.99	13.02	5.66	0.11	0.03	0.00	0.00	0.03	0.00
	Low	59.72	11.64	3.67	0.25	0.05	0.00	0.00	0.08	0.00
VBF	Tight	8.53	23.02	0.08	0.02	0.00	0.00	0.00	0.01	0.00
	Loose	5.14	6.57	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Number of events (N):		N = 0	$\geq 0.01$	$\geq 0.5$	$\geq 1$	$\geq 5$	$\geq 10$	$\geq 15$	$\geq 20$	

Table 11.1: Expected signal event yields in  $36 \text{ fb}^{-1}$  for the Stage 0 STXS bins, split into the four signal categories *Boosted* (high/low) and *VBF* (tight/loose). The background color illustrates the number of events in the corresponding. An increasing number of events change the color from blue over green and yellow to red. The accurate range of each color is given in the last row.

The  $ggF$  bin includes 142 events summed over all signal event categories. Most of the events are in the *Boosted high* category (69 events) and in the *Boosted low* category (60 events). The events of the  $ggF$  STXS bin are nearly equally distributed in these two *Boosted* categories. In both categories 79% of all signal events contribute to  $ggF$  STXS bin. Thus the *Boosted* categories are enriched in  $ggF$  events as expected. 13 events in the *Boosted high* and 12 events in the *Boosted low* category stem from the *VBF* STXS bin. This corresponds to 15% of all

events in both signal categories.

The dominant STXS bin in the *VBF* signal event categories is the *VBF* STXS bin. While 73% of all signal events in the *VBF tight* category stem from the *VBF* STXS bin, only 56% of the events belong to it in the *VBF loose* category. The tighter *VBF* category improves the *VBF* contribution further as expected. In the *VBF loose* category a dominant part of the remaining events belong to the *ggF* bin (44%).

There are no particular signal categories for *VH* STXS bins. Most of the events which correspond to the *VH* STXS bin, are in the *Boosted* categories. The contribution of *VH* processes with leptonically decaying vector boson is negligible as well as the contribution from *ttH* and *tH* processes. Furthermore no signal samples for  $gg \rightarrow ZH$  with a leptonically decaying *Z* boson, for *bbH* and *tH* processes are included in the analysis, due to the fact, that these contributions are at least one magnitude below the one for the *ttH* processes. Thus only six Stage 0 STXS bins are included into the fit. These STXS bins are the *ggF*, the *VBF*, the *VH(had)*, the  $qq \rightarrow ZH(lep)$ , the  $qq \rightarrow WH(lep)$  and the *ttH* bin.

### 11.1.2 Stage 1

For the STXS Stage 1 only STXS bins with a non-zero contribution at Stage 0 are included. In particular the  $gg \rightarrow ZH$  with a leptonically decaying *Z* boson, the *bbH* and the *tH* bin are excluded. Further no splitting for *VH* STXS bins is applied due to their low statistics at Stage 0. The expected signal events yields for the Stage 1 *ggF* STXS bins, defined in Figure 4.9, are given in Table 11.2.

Process		Stage 1										
		VBF		ggH								
Topology cuts		VBF		0		1				2		
Number of jets (Nj)		$\leq 2$	$\geq 3$	0		1				2		
Transverse momentum ( $p_T^H$ )[GeV]		< 200		$\geq 0$ < 60	$\geq 60$ < 120	$\geq 120$ < 200	$\geq 200$	$\geq 0$ < 60	$\geq 60$ < 120	$\geq 120$ < 200	$\geq 200$	
Boosted	High	0.39	0.64	0.00	0.00	0.00	14.42	14.96	0.00	0.00	13.89	24.77
	Low	0.70	0.95	0.02	0.09	16.87	17.49	0.13	0.02	9.49	13.91	0.09
VBF	Tight	2.12	2.01	0.00	0.00	0.25	0.36	0.17	0.00	0.02	0.14	3.45
	Loose	1.87	1.42	0.03	0.07	0.78	0.20	0.04	0.00	0.37	0.23	0.14
Number of events (N):				N = 0	$\geq 0.01$	$\geq 0.5$	$\geq 1$	$\geq 5$	$\geq 10$	$\geq 15$	$\geq 20$	

Table 11.2: Expected signal event yields in  $36 \text{ fb}^{-1}$  for the Stage 1 STXS bins for the gluon fusion production modes defined on the top, split into the four signal categories *Boosted* (high/low) and *VBF* (tight/loose). The background colors refer to the number of events in corresponding bin. The number of events increases from blue over green and yellow to red, defined in the last row of the table.

The contribution of the *VBF* signal categories are negligible in the STXS Stage 1 *ggF* STXS bins, except of the two STXS bins with the *VBF* topology. The *Boosted low* categories contributes mostly to the *ggF* STXS bins with a transverse Higgs boson momentum  $p_T^H$  between 60 GeV and 200 GeV, while the *Boosted high* contributes to the bins with a transverse momentum over 120 GeV. This is expected since on the experimental side a cut on the

transverse momentum of the Higgs boson  $p_T^H$  is set, with  $100 \text{ GeV} < p_T^H < 140 \text{ GeV}$  for the *Boosted low* and  $p_T^H > 140 \text{ GeV}$  for the *Boosted high* category (see Chapter 8).

The STXS bins corresponding to a transverse Higgs boson momentum less than 60 GeV or to a jet multiplicity of zero have nearly zero signal events. The STXS bin with a jet multiplicity of zero has a-priori a high acceptance, but on the experimental side all signal categories require at least one jet. However this STXS bin is not merged with further STXS bins, due to large differences in the bin acceptances.

The  $p_T^H$  STXS bins with less than 200 GeV are merged to increase the statistics. Furthermore the STXS bins with a jet multiplicity of at least one are merged, as suggest in the STXS description (see Section 4.4.4). Thus three STXS bins for the ggF without the VBF topology remains. One STXS bin with a jet multiplicity of zero, one with a jet multiplicity of at least one and  $p_T^H < 200 \text{ GeV}$  and one with a jet multiplicity of one or more and  $p_T^H \geq 200 \text{ GeV}$ .

The ggF VBF like STXS has low statistics. Summed over all signal categories the relative contribution of events in the VBF like STXS bins to all ggF events is 7%. These two STXS bins are merged into the ggF STXS bin with at least a jet multiplicity of one and  $p_T^H < 200 \text{ GeV}$ . Thus three STXS bins remain for the ggF process, which are summarized with their corresponding expected event yield in Table 11.4.

Process		Stage 1							
		VBF					qq → WH	qq → ZH	ttH
		≥ 200	< 200				leptonic		
		Number of jets (N <sub>j</sub> )	≥ 2			rest			
VBF			VH						
Topology cuts		< 25		≥ 25					
Transverse momentum ( $p_T^{\text{Hij}}$ )[GeV]									
Boosted	High	7.23	1.09	0.51	3.45	6.39	0.11	0.03	0.03
	Low	0.22	1.34	0.44	2.17	11.14	0.25	0.05	0.08
VBF	Tight	5.25	12.60	4.22	0.06	0.96	0.02	0.00	0.01
	Loose	0.08	4.53	1.05	0.00	0.96	0.00	0.00	0.00
Number of events (N):		N = 0	≥ 0.01	≥ 0.5	≥ 1	≥ 5	≥ 10	≥ 15	≥ 20

Table 11.3: Expected signal event yields in  $36 \text{ fb}^{-1}$  for the Stage 1 STXS bins for all production modes except of the ggF, split into the four signal event categories *Boosted* (high/low) and *VBF* (tight/loose). The background colors refer to the number of events in the corresponding bin. The number of events increases from blue over green and yellow to red, defined in the last row.

The event yields for the *VBF* STXS bins split into the signal event categories are given in Table 11.3, together with the event yields of the remaining non-ggF Stage 1 STXS bins. The table shows that the events are more equally distributed for these STXS bins than for the ggF STXS bins. The *VBF loose* category contributes mostly to VBF topology STXS bins and the *VBF tight* category to these and further to the beyond the SM physics STXS bin with a transverse momentum of the leading jet  $p_T^j$  over 200 GeV. The *Boosted low* category mostly contributes to the VBF rest STXS bin, while the *Boosted high* category contributes to the high transverse momentum STXS bin as well. These contributions are initiated from *VH*

production with hadronically decaying vector bosons, which are merged into the *VBF* STXS bins at Stage 1.

A merging of the *VBF* STXS bins is necessary, due to the low expected event yields. As suggested in Section 4.4.4, the two STXS bins with a *VBF* topology are merged. Further the STXS bin with a *VH* topology is merged into the rest STXS bin, since most of the events in both STXS bins are initiated by *VH* production processes. Thus three STXS bins for the *VBF* category remain for the STXS Stage 1 fit. All remaining STXS bins are summarized in Table 11.4, with their corresponding expected event yields.

Process		Stage 1								
		ggF			VBF		qq → WH	qq → ZH	ttH	
Topology cuts				VBF			leptonic			
Number of jets (N <sub>j</sub> )		0	≥ 1	≥ 2						
Transverse momentum (p <sub>T</sub> <sup>H</sup> /p <sub>T</sub> <sup>j</sup> ) [GeV]			< 200	≥ 200	< 200	≥ 200				
Boosted	High	0.00	29.30	39.68	1.60	9.84	7.23	0.11	0.03	0.03
	Low	0.02	59.48	0.22	1.78	13.31	0.22	0.25	0.05	0.08
VBF	Tight	0.00	4.90	3.62	16.83	1.03	5.25	0.02	0.00	0.01
	Loose	0.03	4.94	0.18	5.57	0.96	0.09	0.00	0.00	0.00
Number of events (N):		N = 0	≥ 0.01	≥ 0.5	≥ 1	≥ 5	≥ 10	≥ 15	≥ 20	

Table 11.4: Expected signal event yields in  $36 \text{ fb}^{-1}$  for the Stage 1 merged STXS bins, split into the four signal event categories *Boosted* (high/low) and *VBF* (tight/loose). The different colors refers to number of events in the corresponding bins, defined in the last row.

## 11.2 Study of expected sensitivity

To determine the expected sensitivity fits for the Simplified Template Cross Sections at Stage 0 and Stage 1 with different settings are performed with an Asimov dataset.

### 11.2.1 Stage 0

#### Input sample

The Stage 0 STXS are extracted from a fit to missing mass calculator  $m_{\tau\tau}^{MMC}$  distribution (see Chapter 6), with the STXS bins at Stage 0 as the parameters of interest. This distribution is directly sensitive to the signal and allows a better signal to background discrimination, since the signal is focused around the Higgs boson mass. An non-equidistant binning is chosen for this. In the range from 100 GeV to 150 GeV, which corresponds to the region around the expected Higgs boson signal of 125 GeV a fine binning of 10 GeV steps is chosen, while under and over this range a single bin is used. This binning is chosen as a compromise between a good signal to background discrimination in the  $m_{\tau\tau}^{MMC}$  distribution and reasonable statistics in each bin. The distribution for the *Boosted* categories are shown in Figure 11.1. The signal-only distributions in the first row illustrate the high contribution of ggF production mode in this region and the peak of the signal around the Higgs boson mass, while the distributions with all processes in the second row show the high background contribution

especially in the lowest and highest bin. The same behaviour is observed in Figure 11.2 for the  $VBF$  region except of a lower statistic and a domination of VBF processes in the signal distributions.

Due to the small contributions of  $Z \rightarrow ll$  and di-boson events in the signal event categories they are merged into a common background sample.

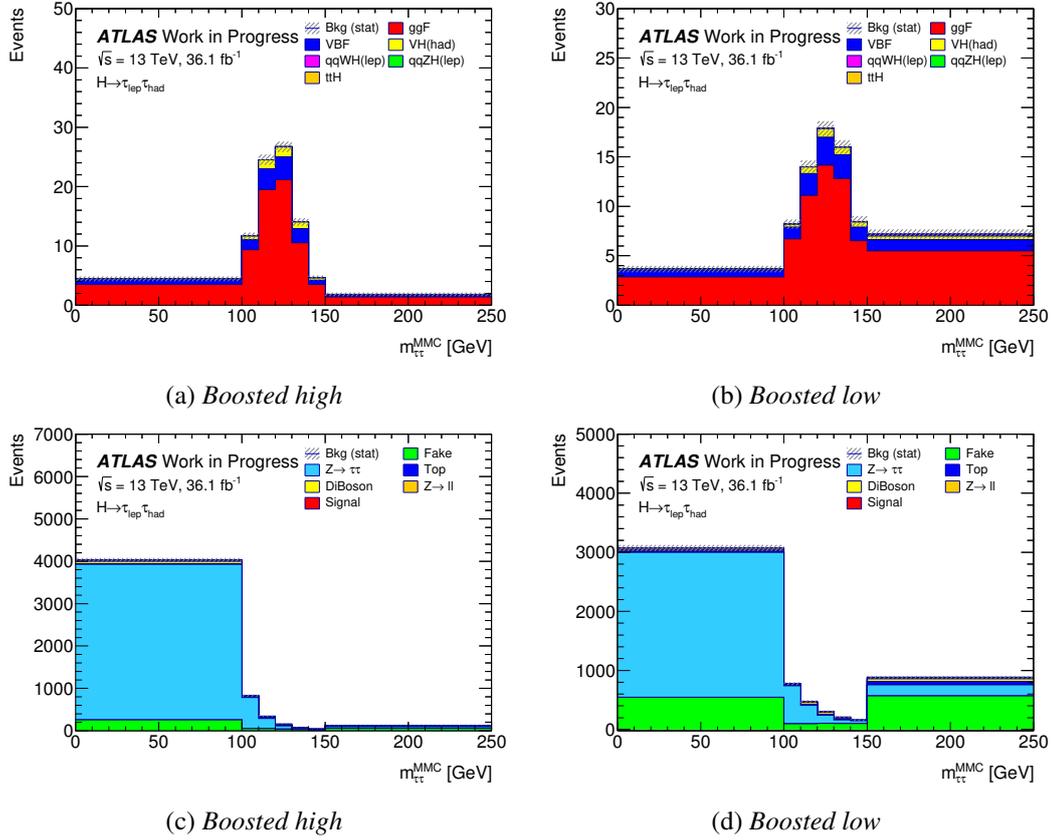


Figure 11.1: The  $m_{\tau\tau}^{MMC}$  distributions of signal processes only (top) and of all processes (bottom) in the *Boosted high* category (left) and the *Boosted low* category (right) in  $36 \text{ fb}^{-1}$ .

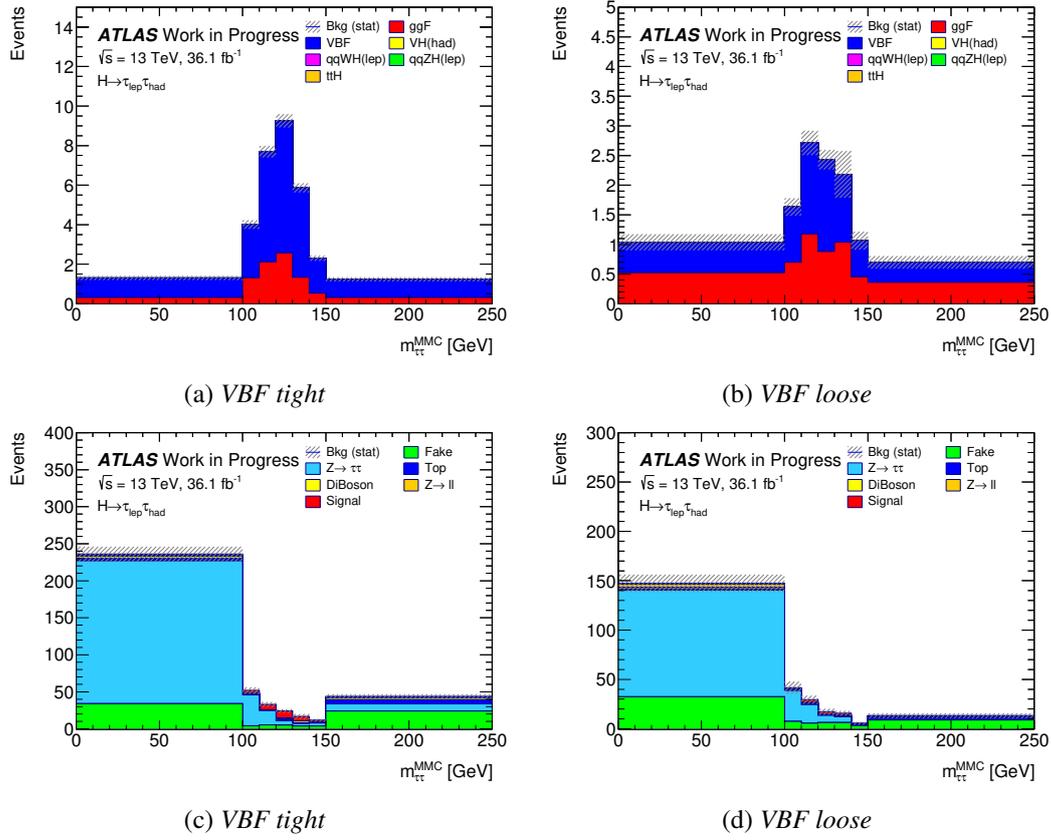


Figure 11.2: The  $m_{\tau\tau}^{MMC}$  distributions of signal processes only (top) and of all processes (bottom) in the *VBF tight* category (left) and the *VBF loose* category (right) in  $36\text{ fb}^{-1}$ .

### Parameter of interest

Table 11.5 gives an overview of the expected best fit values and their uncertainties determined with different settings. The first row (full fit) of Table 11.5 illustrates the default setup, in which the cross section for each Stage 0 STXS bin in the fit as a parameter of interest (PoI). As expected for a fit with Asimov data the values perfectly match the prediction of the SM given in the last row of the table. However the uncertainties up to 5700% point out, that no sensitivity for this full STXS Stage 0 set-up is given. The low statistics for some STXS bins introduce high uncertainties on these cross sections.

Table 11.6a gives the full correlation matrix for all six cross sections. A high correlation between the *VBF* and the *VH (had)* STXS bin and a strong anti-correlation of both to the *ggF* STXS bin exists. Since all signal processes have a similar shape in the  $m_{\tau\tau}^{MMC}$  distribution, with a peak around the Higgs boson mass of 125 GeV, the correlation depends mainly on the event yields of the processes in the different signal categories. Figure 11.1 shows that the *VBF* and *VH (had)* STXS bin has a non-negligible contribution in the *ggF* dominated *Boosted* categories and the *ggF* STXS bin has a non-negligible contribution in the *VBF* dominated

$VBF$  category, while the  $VH$  STXS bin contribution there is negligible. An increasing of the  $ggF$  yield leads to a decrease of the  $VBF$  and  $VH$  yield in the *Boosted* categories, while an increasing  $VBF$  yield in the  $VBF$  categories results only in a decrease of the  $ggF$  yield since the  $VH$  process has no contribution there. Thus the  $VBF$  and  $VH$  STXS bins are forced to a common increase or decrease, while the  $ggF$  STXS bin has the opposite trend. The behaviour of the nuisance parameters and the normalisation factors are illustrated in Figure A.4a in the appendix.

To reduce the influence of the low statistics signals two different assumptions can be used. A merging of of the low statistics processes would combined their statistics, but contradicts the idea of the STXS (see Section 4.4). Instead the low-statistics signals are fixed to the SM predictions. This contains the assumption that these cross sections have no deviations from the SM predictions. The second row (Low signal fixed) of Table 11.5 shows the results for a fit with fixed cross sections for the  $ttH$  STXS bin and the  $qq \rightarrow VH$  STXS bins, with a leptonically decaying vector boson. This reduce the uncertainties of the reaming three bins by at least 11%. To further increase the sensitivity a fit is performed with an additional fixed value for the  $VH$  (had) STXS bin, which has a relative uncertainty around 1900%. The result for this fit is shown in the third row (Only  $ggF$  and  $VBF$ ) of Table 11.5. For this fit the expected relative uncertainties of the  $ggF$  and  $VBF$  STXS bin are reduced to less then 100%.

Bin	$ggF$	$VBF$	$VH$ (had)	$ttH$	$qq \rightarrow ZH$ (lep)	$qq \rightarrow WH$ (lep)
Full fit	$2.8 \pm 9.4$	$0.22 \pm 0.31$	$0.09 \pm 1.9$	$0.03 \pm 31$	$0.01 \pm 57$	$0.03 \pm 14$
Low signal fixed	$2.8 \pm 7.6$	$0.22 \pm 0.26$	$0.09 \pm 1.7$	0.03	0.01	0.03
Only $ggF$ and $VBF$	$2.8 \pm 2.1$	$0.22 \pm 0.17$	0.09	0.03	0.01	0.03
SM prediction	2.8	0.22	0.09	0.03	0.01	0.03

Table 11.5: Comparison of the expected fit results of a global fit with six cross section as parameters of interest (Full fit) against a fit with cross sections of STXS bins with the lowest sensitivity ( $ttH$ ,  $qq \rightarrow ZH$   $qq \rightarrow WH$ ) fixed to the SM prediction (Low signal fixed) and a fit with the cross section of the  $VH$  STXS bin additionally fixed (Only  $ggF$  and  $VBF$ ). Values without an uncertainty are SM predictions.

Table 11.6b illustrates the correlation matrix for the two parameter of interest. Compared to the correlation matrix with six parameters of interest (Table 11.6a), the anti-correlation between the  $ggF$  and the  $VBF$  bin is reduced. For each STXS bin two enriched signal event categories exists, and no STXS bin without a specific signal event category remains. This lead to the assumption that the definition of specific signal event categories for each STXS bin could reduce the correlations between them. Further specific signal event categories for the low sensitivity signals increase their statistics, since most of their events do not pass the  $VBF$  or *Boosted* event category requirements. Thus a study of further signal categories should be done in the future but is beyond the scope of this thesis.

Bin	ggF	VBF	VH (had)	ttH	qq-ZH (lep)	qq-WH (lep)
ggF	1.00	-0.85	-0.95	0.05	-0.58	0.05
VBF	-0.85	1.00	0.79	0.07	0.05	-0.54
VH (had)	-0.95	0.79	1.00	-0.01	0.41	-0.44
ttH	0.05	0.07	-0.01	1.00	-0.12	-0.21
qq-ZH (lep)	-0.58	0.05	0.41	-0.12	1.00	-0.82
qq-WH (lep)	0.05	-0.54	-0.44	-0.21	-0.82	1.00

(a) 6 PoIs

Bin	ggF	VBF
ggF	1	-0.43
VBF	-0.43	1

(b) 2 PoIs

Table 11.6: Correlation matrix of the Stage 0 fit with six parameters of interest (PoI) (left) and two PoIs (right). Dark red refers to high correlation and dark blue to high anti-correlation, while the light colors refer to a small contribution.

### 11.2.2 Stage 1

The same input sample with a non-equidistant binned  $m_{\tau\tau}^{MMC}$  distribution as at Stage 0 is used at Stage 1. Further a merged low statistics background is defined, including  $Z \rightarrow ll$  and di-boson events. The STXS bins  $ttH$  and  $qq \rightarrow VH$  with a leptonically decaying vector boson, which shows no sensitivity on Stage 0, are fixed to SM prediction. Further the  $ggF$  STXS bin with a jet multiplicity  $N_j$  of zero is fixed to the SM cross section of 1.7 pb, since it has no expected event entries (see Table 11.2). The  $VH$  STXS bin with a hadronically decaying vector boson is merged to the  $VBF$  STXS bins, as suggested in Section 4.4.4.

Bin	ggF		VBF		
	Topology cut		VBF		
Number of jets ( $N_j$ )	$\geq 1$		$\geq 2$		
Transverse momentum ( $p_T^H / p_T^j$ ) [GeV]	< 200	$\geq 200$	< 200		$\geq 200$
Full fit	$1.0 \pm 7 \cdot 10^{57}$	$0.036 \pm 6$	$0.07 \pm 2 \cdot 10^{34}$	$0.2 \pm 1 \cdot 10^{55}$	$0.014 \pm 22$
Only beyond the SM STXS bins	1.0	$0.036 \pm 5$	0.07	0.2	$0.014 \pm 18$
Only $ggF$ STXS bins	$1.0 \pm 2 \cdot 10^{55}$	$0.036 \pm 4$	0.07	0.2	0.014
SM prediction	1.0	0.036	0.07	0.2	0.014

Table 11.7: Comparison of the expected fit results of Stage 1 STXS bins defined on the top with all five STXS bins as parameters of interest (PoIs) (Full fit), with only sensitive to the beyond the SM effects STXS bins as PoIs and with only  $ggF$  STXS bins as PoIs, while the remaining STXS bins are fixed to the SM cross sections.

		process		ggF		VBF		
		cuts				VBF		
		Signal Topology	jets (Nj)	$\geq 1$		$\geq 2$		
				Number of ( $p_T^{H_1}, p_T^{H_2}$ ) [GeV]	< 200	$\geq 200$	< 200	$\geq 200$
ggF	VBF	$\geq 1$	< 200	1,00	0,19	-0,14	-0,47	0,00
			$\geq 200$	0,19	1,00	0,11	-0,43	-0,36
VBF	VBF	$\geq 2$	< 200	-0,14	0,11	1,00	0,06	-0,50
			$\geq 200$	-0,47	-0,43	0,06	1,00	-0,10
			$\geq 200$	0,00	-0,36	-0,50	-0,10	1,00

(a) 5 PoIs

		process		ggF	VBF	
		cuts		$\geq 1$		
		Signal Topology	jets (Nj)	$\geq 1$		
				Number of ( $p_T^{H_1}, p_T^{H_2}$ ) [GeV]	< 200	$\geq 200$
ggF	VBF	$\geq 1$	< 200	1	-0,44	
			$\geq 200$	-0,44	1	

(b) only beyond SM bins

		process		ggF	
		cuts		$\geq 1$	
		Signal Topology	jets (Nj)	$\geq 1$	
				Number of ( $p_T^{H_1}, p_T^{H_2}$ ) [GeV]	< 200
ggF	VBF	$\geq 1$	< 200	1	-0,04
			$\geq 200$	-0,04	1

(c) only  $ggF$  bins

Table 11.8: Correlation matrix of the Stage 1 fit with five parameters of interest (PoIs) (a), only the beyond the SM sensitive bins (a) as PoIs and only the  $ggF$  bins as PoIs (c). Dark red refers to high correlation and dark blue to high anti-correlation, while the light colors refers to a small contribution.

Thus five STXS bins are remain as parameters of interest at Stage 1. One  $ggF$  STXS bins with a transverse momentum of the Higgs boson  $p_T^H$  less 200 GeV and one  $ggF$  STXS bins with  $p_T^H \geq 200$  GeV, which should be sensitive to beyond the SM physics effects. Both bins required to has a jet multiplicity of at least one. Three  $VBF$  STXS bins are used as parameters of interest at Stage 1. One with a transverse momentum of the leading jet  $p_T^{j_1}$  greater than 200 GeV sensitive to beyond the SM effects and two  $VBF$  STXS bins with  $p_T^{j_1} \leq 200$  GeV. One of these STXS bin requires  $VBF$  topology cuts, while the remaining one consists of all events which fail the requirements of the first two  $VBF$  STXS bins.

The results for Stage 1 are summarised in Table 11.7. The first row (Free fit) shows the best fit values with their uncertainties for a fit with all six remaining cross sections as parameters of interest. The uncertainties of all STXS bins are large, but the two STXS bins which are sensitive to beyond the SM physics effects has smaller relative uncertainties as the remaining STXS bins. Two further fits are produced to prove a potential decreasing of the uncertainties by setting some STXS bins to the SM predictions. One fit with all Stage 1 STXS bin fixed to SM prediction except of the two STXS bins, which should be sensitive to beyond the SM physics (see Section 4.4). This reduces the uncertainties of these two STXS bins against the first determination (see Table 11.7 second row (Only beyond the SM bins)). However both STXS bins still have a large uncertainty such that no sensitivity is expected. A third fit in which all STXS bins except of the two  $ggF$  STXS bins are fixed to the SM prediction is produced. The results for this fit is shown in the third row of the table (Only  $ggF$  STXS bins). Compared to the full fit the uncertainties are reduced but still to large to expect sensitivity in a

fit to data. The nuisance parameters and normalizations factors can be found in Figure A.5 and Table A.4 in the appendix.

Table 11.8 shows the correlation matrices for the three different fit set-ups. The matrices show an anti-correlation between the  $ggF$  and  $VBF$  STXS bins, which were already observed at Stage 0.

The study at Stage 1 shows that no sensitivity in the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel alone for the 2015 and 2016 combined dataset with the current signal event categories is expected.

### 11.3 Results at Stage 0

The gluon fusion and vector boson fusion cross sections as parameters of interest fitted to all four signal regions and two control regions for top quark produced background at STXS Stage 0 are given in Table 11.9. For the fit the combined 2015 and 2016 dataset for the  $H \rightarrow \tau_{lep}\tau_{had}$  decay channel with  $36 \text{ fb}^{-1}$  is used and the remaining STXS Stage 0 bin were fixed to the SM prediction. The best fit values for the  $ggF$  cross section  $\sigma_{ggF}$  and for the  $VBF$  cross section  $\sigma_{VBF}$  are given by:

$$\sigma_{ggF} = (2.6 \pm 2.1) \text{ pb} \quad \text{and} \quad \sigma_{VBF} = (0.15 \pm 0.17) \text{ pb}. \quad (11.1)$$

Both values agree with the predicted cross sections of the SM within their uncertainties, but are somewhat lower. This behaviour is consistent with the signal strength measurement results of  $\mu = 0.8 \pm 0.4$ , which is lower as the expected one (see Section 9.3). The relative uncertainties are 81% on the  $ggF$  cross section and 113% on the  $VBF$  cross section.

Table 11.9 illustrate the statistical uncertainty from data and the systematic uncertainty further split into the different sources. The main uncertainty sources are the data statistics, the uncertainties related to jets and the missing transverse energy  $E_T^{miss}$  and the statistical uncertainties of the simulated events.

The nuisance parameters and normalisation factors are illustrated in Figure A.6 in the appendix. The normalisation factor of the top quark production background for the *Boosted* category includes with  $1.04 \pm 0.08$  the expected value of one within its uncertainties, while the the one for the  $VBF$  category has a value of  $1.4 \pm 0.3$ . The top quark production background calculated by simulation is lower than the one measured in the control regions since both normalisation factors are larger one.

The correlation matrix of the parameters of interest is given in Table 11.10. The matrix shows an anti-correlation between the two cross sections of -0.46, as predicted by the fit with an Asimov dataset (see Table 11.6b).

Bin	$\sigma_{ggF}$ [pb]	$\sigma_{VBF}$ [pb]
SM prediction	2.8	0.22
Best fit values	$2.6 \pm 2.1$	$0.15 \pm 0.17$
Data statistic	$\pm 1.15$	$\pm 0.11$
Systematic unc.	$\pm 1.70$	$\pm 0.13$
Normalization unc.	$\pm 0.10$	$\pm 0.01$
Jets and $E_T^{miss}$	$\pm 1.11$	$\pm 0.08$
B-jets	$\pm 0.08$	$\pm 0.01$
Light leptons	$\pm 0.08$	$\pm 0.01$
$\tau$ -leptons	$\pm 0.15$	$\pm 0.02$
Pileup re-weighting	$\pm 0.16$	$\pm 0.01$
"Fake" background	$\pm 0.26$	$\pm 0.04$
Luminosity	$\pm 0.12$	$\pm 0.01$
Theory unc. on signal	$\pm 0.72$	$\pm 0.04$
Theory unc. on background	$\pm 0.02$	$\pm 0.00$
Simulation statistic	$\pm 0.86$	$\pm 0.09$

Table 11.9: Best fit values for STXS at Stage 0 with two parameter of interest ( $\sigma_{ggF}$ ,  $\sigma_{VBF}$ ) for the 2015 and 2016 dataset of the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel with  $36\text{fb}^{-1}$  and the uncertainties on the cross sections split into several sources.

Bin	ggF	VBF
ggF	1	-0.46
VBF	-0.46	1

Table 11.10: Correlation matrix of the Stage 0 fit with two parameter of interest ( $ggF$  and  $VBF$  cross section) for the combined 2015 and 2016 dataset.



---

## 12 CONCLUSION AND OUTLOOK

---

The thesis presents the cut-based signal strength measurement of the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel with a combined 2015 and 2016 dataset of an integrated luminosity of  $36 \text{ fb}^{-1}$  recorded with the ATLAS detector. For this a new data-driven method is introduced and validated for the determination of background events from jets misidentified as hadronically decaying  $\tau$ -leptons, so called "fakes". Furthermore first studies of Simplified Template Cross Section (STXS) at Stage 0 and Stage 1 are discussed.

### **Background estimation of jet misidentified as $\tau_{had}$**

The new data-driven background estimation is based on the "fake-factor" method of the  $H \rightarrow \tau_{lep} \tau_{had}$  analysis used in Run 1[20], but is modified and expanded for the new dataset. Especially a new data-driven approach is used to estimate the relative contribution of multi-jet processes, which contribute to this background. The method uses an "anti- $\tau$ " control region defined by inverting the medium ID criteria [82] of the hadronically decaying  $\tau_{had}$ -lepton to calculate the number and the shape of "fake" events (see Section 10.1.1). To transfer the number and the shape of events from the "anti- $\tau$ " control region to the signal region a transfer factor, so called "fake-factor", is used.

The "fake-factor" consists of individual "fake-factors" for W+jets, Z+jets, top quark production and multi-jet processes, which are summed up to the combined "fake-factor" weighted by their relative contributions in the signal region (see Section 10.1.2). To calculate these individual "fake-factors" a separate control region is defined for each background process. A closure test validate that these individual "fake-factors" calculated in the control region can be applied to the signal region (see Section 10.3.6). A further test proves that the assumption of using the individual "fake-factor" of W+jets processes for all non-multi-jet processes can be made without an additional systematic uncertainty (see section (10.3.1)). Since for example the used triggers have changed between 2015 and 2016 a study is done to prove that the number of "fake" events can be calculated for a combined 2015 and 2016 dataset instead of for the individual datasets (see Section 10.3.2).

The relative contributions are calculated with a new data-driven approach based on another transfer factor, the so called isolation factor (see Section 10.1.5). This isolation factor transfers the number and the shape of multi-jet events from a multi-jet control region, which is defined by inverting the gradient isolation criteria on the light lepton [79, 80], to the signal region. With the number of multi-jet events in the signal region the relative contributions of this process and of the non-multi-jet processes are determined. To avoid that the same control region is used to calculate the individual "fake-factor" of multi-jet processes and to calculate their relative contributions, the requirement of the signal region that the light lepton and the  $\tau_{had}$ -lepton must have opposite charge sign ("opposite-sign") is inverted when calculating the relative contribution. A study shows that the relative contributions calculated in this "same-sign" multi-jet control region can be applied to the "opposite-sign" region, introducing a systematic uncertainty (see Section 10.3.3) on the isolation factor to account for the small difference.

A closure test in the "same-sign" signal region validates the results of the new "fake-factor"

method (see Section 10.3.5) and the distributions in the signal regions show a good agreement between data and background in regions where no signal is expected.

### Signal strength measurement

The signal strength  $\mu = (\sigma \cdot BR_{\tau\tau})_{obs} / (\sigma \cdot BR_{\tau\tau})_{SM}$  in the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel is determined to be  $\mu = 0.8 \pm 0.4$  in agreement with the SM expectation of 1 (see Section 9.3). A  $p_0$ -value of 0.032 corresponding to  $1.9\sigma$  is found. The dominant uncertainty sources are related to the jet and missing transverse energy reconstruction ( $\pm 0.28$ ) and the statistical uncertainties of the dataset ( $\pm 0.22$ ), followed by the uncertainties in the estimation of "fake" background ( $\pm 0.15$ ) and the statistical uncertainties of the simulated samples used ( $\pm 0.19$ ).

### Simplified Template Cross Sections

The measurement of the gluon fusion STXS bin  $\sigma_{ggF}$  and vector boson fusion STXS bin  $\sigma_{VBF}$  yield  $\sigma_{ggF} = (2.6 \pm 2.1)$  pb and  $\sigma_{VBF} = (0.15 \pm 0.17)$  pb (see Section 11.3). Both results are in agreement with the SM prediction of 2.8 pb and 0.22 pb, but have large uncertainties. The dominant uncertainty sources are the data statistics ( $\pm 1.2$  pb and  $\pm 0.11$  pb), the uncertainties related to jet and the missing transverse energy reconstruction ( $\pm 1.1$  pb and  $\pm 0.08$  pb) and the statistical uncertainties of the simulated events ( $\pm 0.9$  pb and  $\pm 0.09$  pb). The best fit values for both cross sections are smaller than the SM predictions in agreement with the signal strength measurement of less than one. The sensitivity studies for the STXS Stage 1 and full Stage 0 show that no sensitivity is available in the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel with the combined 2015 and 2016 dataset and the current signal event categories.

### Outlook

The new "fake" background estimates shows a good agreement with data and can be used in analysis beyond the scope of this thesis as well. To apply this method on a new dataset investigations for the binning of individual "fake-factors", the relative contributions and the isolation factor are necessary. With a larger amount of data a calculation of the isolation factor in the signal regions should be possible. Furthermore a repetition of the closure tests to prove the assumption of the method with a new dataset is necessary.

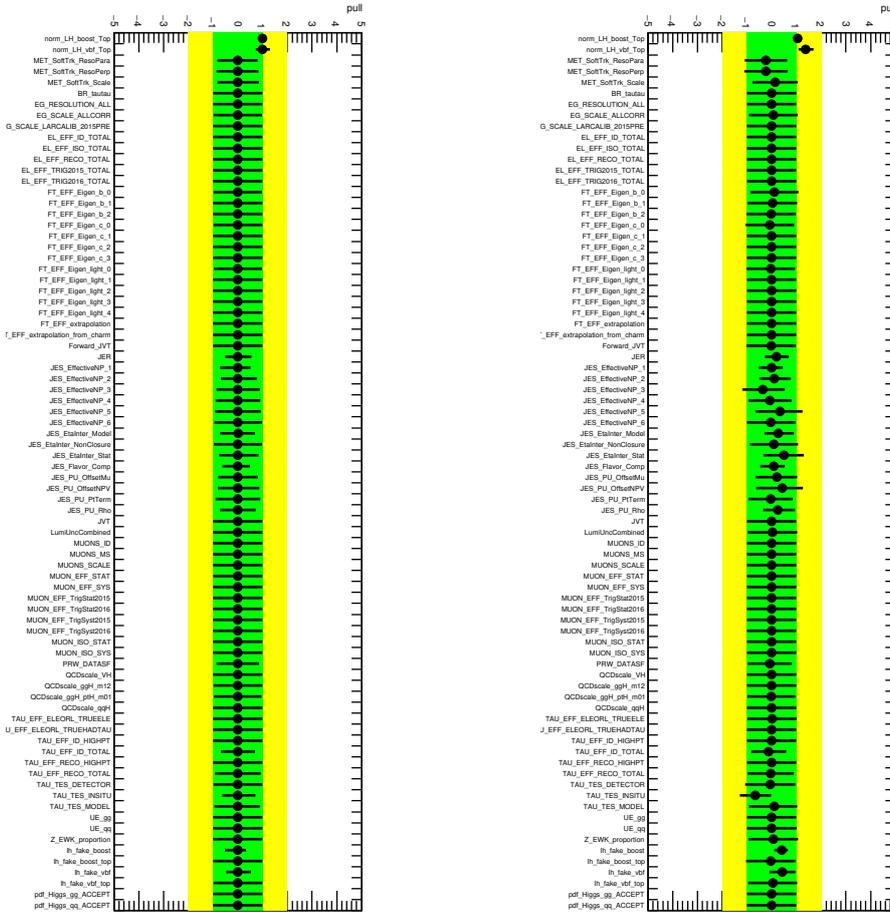
A significance of the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel with  $1.9\sigma$  is not sufficient for evidence of this decay mode of the Higgs boson with the combined 2015 and 2016 dataset of  $36 \text{ fb}^{-1}$ , but the combination of the three different  $H \rightarrow \tau\tau$  sub-decay modes which will be published beyond the scope of this thesis is expected to provide at least evidence or observation of the Yukawa coupling at the ATLAS experiment.

The Simplified Template Cross Sections studies show that no measurements at Stage 1 are possible for  $H \rightarrow \tau_{lep} \tau_{had}$  decay mode with analysis strategy and signal event categories of the combined 2015 and 2016 dataset. For a sensitive measurement the combination of the three different  $H \rightarrow \tau\tau$  sub-decay modes and an optimisation of the defined signal region would be necessary. A combination with the 2017 and 2018 datasets should be considered to increase the statistics as well. Furthermore for a complete Stage 1 measurement the implementation of further theoretical systematic uncertainties are necessary to take possible migration between different exclusive transverse momentum or jet multiplicity bins into account.

# A APPENDIX

## A.1 "Signal strength estimation"

Figure A.1 show the observed and the expected behaviour of the nuisance parameters and normalization factors. The nuisance parameters related to the jet energy resolution (JER) and jet energy scale (JES) show constraints and observed values are slightly pulled to positive values. Further the parameters related to the "fake" estimation show a similar behaviour.



(a) Expected

(b) Observed

Figure A.1: Expected (a) and observed (b) nuisance parameters and normalisation factors for the signal strength analysis in the  $H \rightarrow \tau_{had} \tau_{lep}$  decay channel with a combined 2015 and 2016 dataset of  $36 \text{ fb}^{-1}$ .

## A.2 "Fake background estimation"

Table A.1 shows the event yields of data and different background processes in the "anti- $\tau$ " signal regions and "anti- $\tau$ " control regions of the *Boosted* and *VBF* signal event category.

Category	Region	Data	$N_{\text{not } j \rightarrow \tau}$	W+jets	Z+jets	Top	Other
VBF	SR	2893	$291 \pm 51$	$1580 \pm 196$	$190 \pm 142$	$124 \pm 13$	$45 \pm 7$
	W+jets	1033	$90 \pm 19$	$769 \pm 115$	$81 \pm 27$	$65 \pm 9$	$13 \pm 3$
	Multi-jet	3725	$25 \pm 7$	$70 \pm 24$	$36 \pm 25$	$18 \pm 6$	$3 \pm 2$
	Z+jets	142	$10 \pm 3$	$5 \pm 3$	$147 \pm 32$	$7 \pm 2$	$3 \pm 1$
	Top	980	$142 \pm 16$	$280 \pm 46$	$32 \pm 8$	$384 \pm 25$	$12 \pm 3$
Boosted	SR	34569	$4711 \pm 133$	$19853 \pm 530$	$1456 \pm 131$	$1814 \pm 51$	$589 \pm 23$
	W+jets	13794	$1011 \pm 83$	$10786 \pm 286$	$661 \pm 131$	$830 \pm 33$	$2016 \pm 14$
	Multi-jet	33370	$298 \pm 33$	$729 \pm 65$	$67 \pm 19$	$267 \pm 30$	$29 \pm 5$
	Z+jets	2348	$149 \pm 20$	$103 \pm 19$	$1908 \pm 190$	$97 \pm 11$	$45 \pm 4$
	Top	16771	$2672 \pm 77$	$4007 \pm 224$	$224 \pm 28$	$7773 \pm 110$	$139 \pm 12$

Table A.1: Number of events in the "anti- $\tau$ " control regions for the W+jets, multi-jet, Z+jets and top quark production background corresponding to the *VBF* (top) and *Boosted* (bottom) category.  $N_{\text{not } j \rightarrow \tau}$  refers to events where the  $\tau$ -lepton is not initiated by a jet.

Table A.2 shows the relative contribution of W+jets and multi-jet processes to the "fake" background in the "same-sign" control region split into the signal event categories.

	Lepton	$p_T$ [GeV]	Electron						Muon					
			$p_T < 40$		$40 < p_T < 90$		$90 < p_T$		$p_T < 40$		$40 < p_T < 90$		$90 < p_T$	
			W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet	W+jets	Multi-jet
Boosted	1 track	$\Delta\Phi < 1.0$	0.548	0.452	0.709	0.291	0.764	0.236	0.663	0.337	0.648	0.352	0.798	0.202
		$1.0 < \Delta\Phi < 1.5$	0.734	0.266	0.612	0.388	0.728	0.272	0.799	0.201	0.737	0.263	0.578	0.422
		$1.5 < \Delta\Phi < 2.0$	0.888	0.112	0.791	0.209	0.926	0.074	0.870	0.130	0.846	0.154	0.870	0.130
	3 track	$\Delta\Phi > 2.0$	0.525	0.475	0.653	0.347	0.892	0.108	0.816	0.184	0.939	0.061	0.870	0.130
		$\Delta\Phi < 1.0$	0.882	0.118	0.913	0.087	0.941	0.059	0.872	0.128	0.878	0.122	0.922	0.078
		$1.0 < \Delta\Phi < 1.5$	0.912	0.088	0.902	0.098	0.951	0.049	0.926	0.074	0.903	0.097	0.894	0.106
VBF	1 track	$1.5 < \Delta\Phi < 2.0$	0.964	0.036	0.948	0.052	0.984	0.016	0.933	0.067	0.938	0.062	0.942	0.058
		$\Delta\Phi > 2.0$	0.892	0.108	0.943	0.057	0.980	0.020	0.929	0.071	0.974	0.026	0.967	0.033
	3 track	$\Delta\Phi < 2.0$	0.700	0.300	0.713	0.287	0.613	0.387	0.568	0.432	0.526	0.474	0.457	0.543
		$\Delta\Phi > 2.0$	0.510	0.490	0.522	0.478	0.999	0.001	0.703	0.297	0.658	0.342	0.637	0.363
3 track	$\Delta\Phi < 2.0$	0.909	0.091	0.917	0.083	0.902	0.098	0.820	0.180	0.820	0.180	0.889	0.111	
	$\Delta\Phi > 2.0$	0.792	0.208	0.832	0.168	0.999	0.001	0.902	0.098	0.778	0.222	0.790	0.210	

Table A.2: Relative contributions  $R_i$  of the "same-sign" region separated into the number of tracks of the  $\tau$ -lepton (1,3), the event category (*Boosted*, *VBF*) and the difference of the azimuthal angle between the  $\tau$ -lepton and the missing transverse energy ( $\Delta\Phi$ ) shown on the left-hand side and the lepton flavour (electron, muon) and the transverse momentum of the  $\tau$ -lepton ( $p_T$ ) shown on the top. The background color changes from blue to red for increasing  $R_i$  values.

Figures A.2 and A.3 shows the relative contribution for W+jets and multi-jet processes for the "same-sign" regions of the *Boosted* and *VBF* signal event categories.

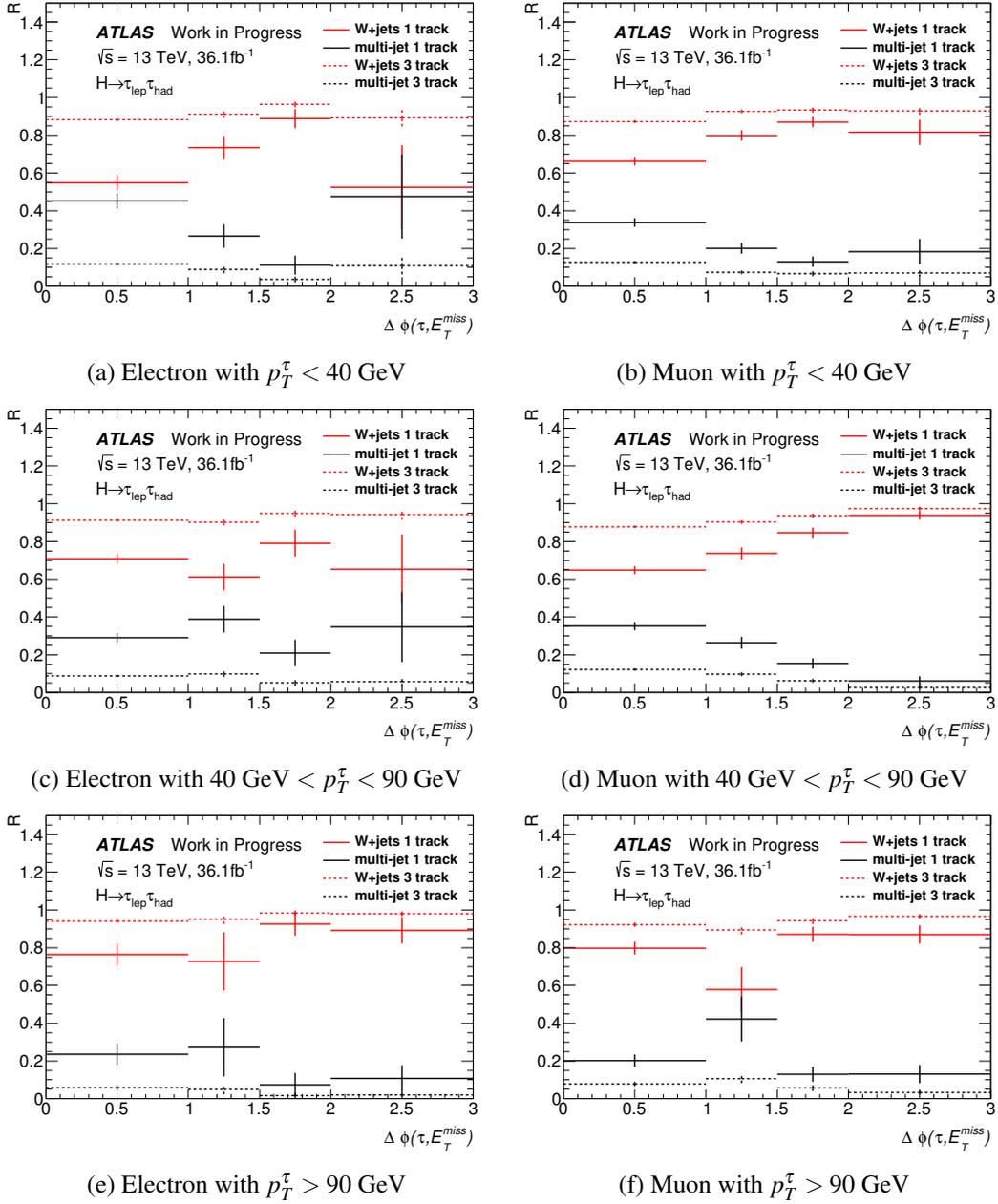


Figure A.2: The relative contribution for the "same-sign" *Boosted* categories. The left column illustrates the results for the electron channel and right one for the muon channel. From the top to the bottom row different cuts on the transverse momentum of the light leptons were applied ( $p_T^\tau < 40$  GeV,  $40 \text{ GeV} < p_T^\tau < 90$  GeV,  $p_T^\tau > 90$  GeV). The dashed lines correspond to the results for hadronically decaying  $\tau$ -leptons with three tracks, while the full lines show the  $\tau$ -leptons with one track. The red line refers to the relative contribution of W+jets events, while the black one belongs to multi-jet events.

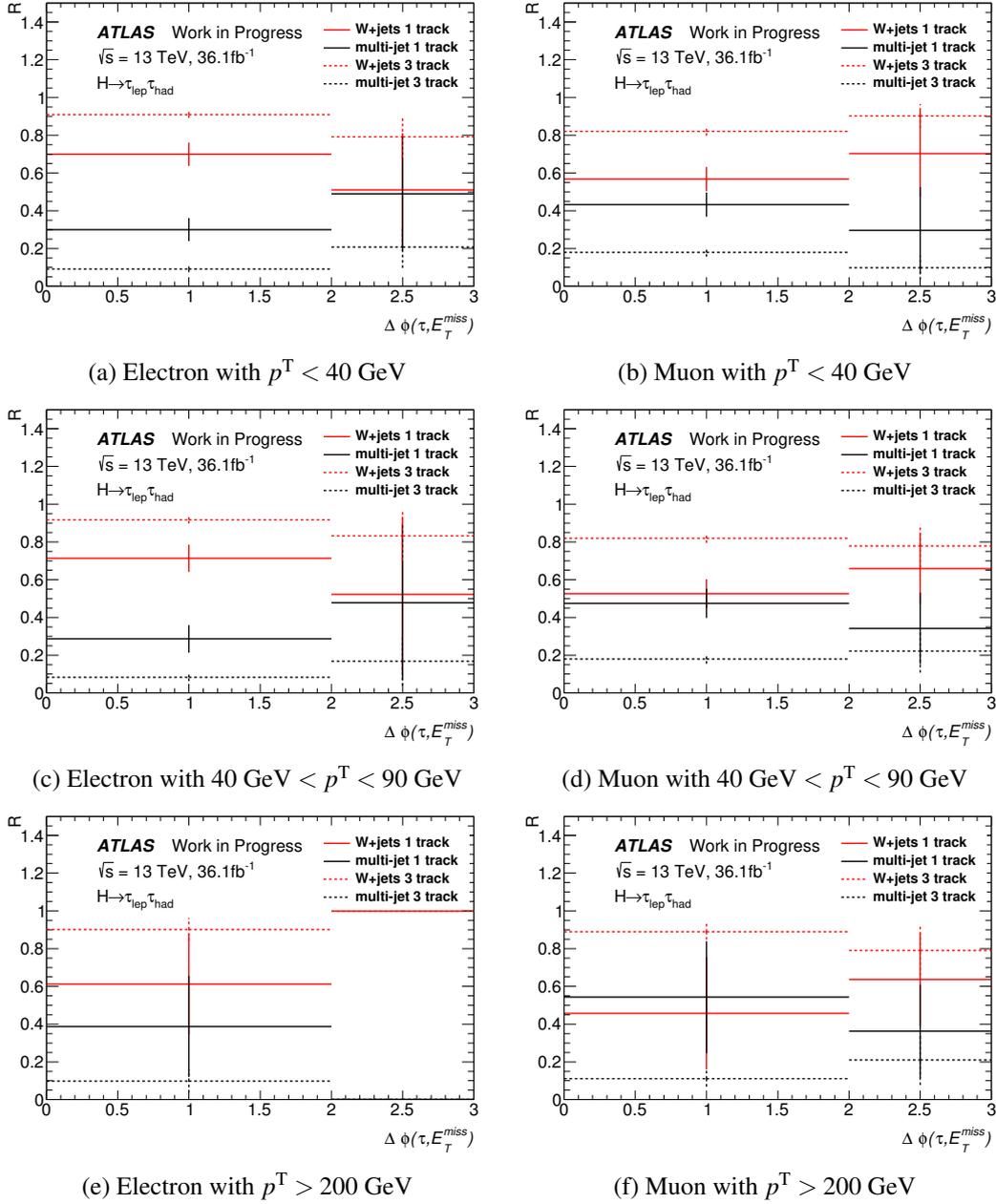


Figure A.3: The relative contribution for the "same-sign"  $VBF$  category. The first column illustrates the results for the electron channel and last one for the muon channel. From the top to the bottom row different cuts on the transverse momentum of the light lepton are applied ( $p_T^l < 40 \text{ GeV}$ ,  $40 \text{ GeV} < p_T^l < 90 \text{ GeV}$ ,  $p_T^l > 90 \text{ GeV}$ ). The dashed lines correspond to the results for hadronically decaying  $\tau$ -leptons with three tracks, while the full lines show the  $\tau$ -leptons with one track. The red line refers to the relative contribution of  $W$ +jets events, while the black one belongs to multi-jet events.

Table A.3 shows the relative contribution of W+jets, Z+jets and top quark production processes to the "fake" background with and without a re-weighting with the data-driven estimated relative contribution multi-jet processes.

$p_T$ [GeV]		Electron														
		Number of tracks		1					3							
		Process		W+jets		Top		Z+jets		Multi-jet	W+jets		Top		Z+jets	
Calculation		MC	Data	MC	Data	MC	Data	Data	MC	Data	MC	Data	MC	Data	Data	
Boosted	$p_T < 40$	$\Delta\Phi < 1.0$	0.63	0.53	0.03	0.03	0.10	0.09	0.35	0.60	0.73	0.05	0.06	0.09	0.11	0.11
		$1.0 < \Delta\Phi < 1.5$	0.64	0.60	0.09	0.08	0.03	0.03	0.28	0.77	0.75	0.12	0.12	0.04	0.04	0.10
		$1.5 < \Delta\Phi < 2.0$	0.67	0.60	0.13	0.12	0.04	0.03	0.25	0.75	0.76	0.14	0.14	0.03	0.04	0.07
		$\Delta\Phi > 2.0$	0.52	0.71	0.12	0.16	0.01	0.01	0.12	0.87	0.85	0.11	0.11	0.01	0.01	0.04
	$40 < p_T < 90$	$\Delta\Phi < 1.0$	0.54	0.66	0.03	0.04	0.06	0.07	0.22	0.63	0.76	0.05	0.06	0.09	0.11	0.07
		$1.0 < \Delta\Phi < 1.5$	0.64	0.72	0.07	0.08	0.04	0.05	0.16	0.68	0.78	0.10	0.12	0.05	0.05	0.05
		$1.5 < \Delta\Phi < 2.0$	0.90	0.71	0.06	0.05	0.03	0.03	0.22	0.82	0.78	0.12	0.12	0.06	0.05	0.05
		$\Delta\Phi > 2.0$	0.90	0.85	0.08	0.07	0.02	0.02	0.06	0.70	0.83	0.11	0.13	0.02	0.03	0.02
	$90 < p_T$	$\Delta\Phi < 1.0$	0.72	0.80	0.03	0.03	0.08	0.09	0.09	0.70	0.82	0.05	0.06	0.09	0.10	0.02
		$1.0 < \Delta\Phi < 1.5$	0.87	0.74	0.06	0.05	0.07	0.06	0.15	0.78	0.85	0.05	0.05	0.06	0.06	0.03
		$1.5 < \Delta\Phi < 2.0$	0.80	0.75	0.04	0.04	0.03	0.03	0.18	0.93	0.89	0.03	0.03	0.04	0.03	0.05
		$\Delta\Phi > 2.0$	0.75	0.84	0.03	0.03	0.02	0.02	0.10	0.94	0.91	0.02	0.02	0.04	0.04	0.03
VBF	$p_T < 40$	$\Delta\Phi < 2.0$	0.82	0.61	0.02	0.02	0.16	0.12	0.25	0.44	0.81	0.04	0.07	0.02	0.04	0.08
		$\Delta\Phi > 2.0$	0.97	0.71	0.00	0.00	0.03	0.02	0.27	0.61	0.57	0.11	0.11	0.28	0.26	0.07
	$40 < p_T < 90$	$\Delta\Phi < 2.0$	0.37	0.46	0.04	0.05	0.19	0.23	0.26	0.56	0.69	0.06	0.07	0.14	0.17	0.07
		$\Delta\Phi > 2.0$	0.96	0.54	0.03	0.01	0.02	0.01	0.43	0.27	0.38	0.05	0.07	0.34	0.49	0.06
	$90 < p_T$	$\Delta\Phi < 2.0$	0.31	0.58	0.02	0.04	0.11	0.22	0.17	0.75	0.73	0.05	0.05	0.17	0.16	0.06
		$\Delta\Phi > 2.0$	0.93	0.02	0.07	0.00	0.00	0.00	0.98	0.60	0.74	0.06	0.08	0.10	0.13	0.06

(a) Electron channel

$p_T$ [GeV]		Muon														
		Number of tracks		1					3							
		Process		W+jets		Top		Z+jets		Multi-jet	W+jets		Top		Z+jets	
Calculation		MC	Data	MC	Data	MC	Data	Data	MC	Data	MC	Data	MC	Data	Data	
Boosted	$p_T < 40$	$\Delta\Phi < 1.0$	0.57	0.67	0.04	0.05	0.02	0.03	0.26	0.60	0.79	0.04	0.06	0.03	0.04	0.11
		$1.0 < \Delta\Phi < 1.5$	0.43	0.74	0.05	0.09	0.01	0.02	0.14	0.61	0.77	0.11	0.14	0.02	0.02	0.07
		$1.5 < \Delta\Phi < 2.0$	0.60	0.72	0.09	0.11	0.03	0.04	0.13	0.83	0.82	0.12	0.11	0.02	0.02	0.06
		$\Delta\Phi > 2.0$	0.70	0.80	0.07	0.08	0.00	0.00	0.13	0.73	0.83	0.09	0.10	0.03	0.03	0.04
	$40 < p_T < 90$	$\Delta\Phi < 1.0$	0.69	0.71	0.04	0.04	0.02	0.02	0.23	0.66	0.81	0.05	0.06	0.03	0.04	0.09
		$1.0 < \Delta\Phi < 1.5$	0.64	0.77	0.05	0.06	0.02	0.02	0.15	0.68	0.78	0.11	0.13	0.03	0.03	0.06
		$1.5 < \Delta\Phi < 2.0$	0.04	0.24	0.09	0.51	0.02	0.13	0.12	0.81	0.81	0.12	0.12	0.03	0.03	0.04
		$\Delta\Phi > 2.0$	0.58	0.84	0.05	0.08	0.01	0.02	0.07	0.81	0.83	0.12	0.12	0.01	0.01	0.03
	$90 < p_T$	$\Delta\Phi < 1.0$	0.94	0.79	0.03	0.03	0.02	0.02	0.17	0.78	0.84	0.07	0.07	0.03	0.03	0.05
		$1.0 < \Delta\Phi < 1.5$	0.95	0.79	0.02	0.02	0.03	0.02	0.17	0.79	0.89	0.04	0.04	0.02	0.02	0.05
		$1.5 < \Delta\Phi < 2.0$	0.59	0.81	0.04	0.05	0.01	0.02	0.13	0.78	0.89	0.04	0.04	0.02	0.02	0.05
		$\Delta\Phi > 2.0$	0.52	0.89	0.01	0.01	0.02	0.03	0.06	0.88	0.93	0.02	0.02	0.03	0.03	0.02
VBF	$p_T < 40$	$\Delta\Phi < 2.0$	0.51	0.55	0.02	0.03	0.03	0.03	0.40	0.64	0.74	0.04	0.05	0.04	0.05	0.16
		$\Delta\Phi > 2.0$	0.88	0.88	0.00	0.00	0.12	0.12	0.00	0.88	0.73	0.11	0.09	0.01	0.01	0.18
	$40 < p_T < 90$	$\Delta\Phi < 2.0$	0.58	0.59	0.05	0.05	0.04	0.04	0.32	0.58	0.79	0.06	0.08	0.02	0.02	0.11
		$\Delta\Phi > 2.0$	0.97	0.63	0.03	0.02	0.00	0.00	0.35	0.91	0.77	0.09	0.08	0.00	0.00	0.15
	$90 < p_T$	$\Delta\Phi < 2.0$	0.54	0.76	0.01	0.02	0.02	0.02	0.20	0.90	0.84	0.07	0.07	0.02	0.02	0.07
		$\Delta\Phi > 2.0$	0.97	0.95	0.00	0.00	0.03	0.03	0.02	0.89	0.89	0.06	0.06	0.05	0.05	0.00

(b) Muon channel

Table A.3: Relative contribution of W+jets, top quark production and Z+jets event calculated from simulated events (MC) and after re-weighting with the data-driven estimation of the relative contribution of multi-jet event (Data) separated for the electron channel (a) and the muon channel (b). The contributions are determined separately for the event categories (*Boosted*, *VBF*) and binned into the number of tracks of the hadronically decaying  $\tau$ -lepton (1, 3), the transverse momentum of the  $\tau$ -lepton ( $p_T$ ) and the difference of the azimuthal angle between the  $\tau$ -lepton and the missing transverse energy ( $\Delta\Phi$ ).

### A.3 Simplified Template Cross Sections

#### Stage 0 expected

Figure A.4a illustrate the behaviour of nuisance parameters (NPs) and the normalisation factors (NFs) for the top quark production background for a fit six cross sections as a parameter of interest (PoIs) (full fit) at Stage 0. All precise values of the NFs and their corresponding uncertainties are given in the last two rows of the first column of Table A.5. The best fit values for both NFs are one, what is expected for a Asimov dataset by construction. The high uncertainty is mainly introduced through the high degree of freedom in the signal region. Due to the high uncertainties on the PoI the normalisation for top quark initiated background can be chosen in a wide range without a huge impact on the results.

The NPs have a similar behaviour as the one for the signal strength analysis (see Section 9.3). The systematic uncertainty for the jet energy resolution  $JER$  as well as for the "fakes" are constrained. The NPs with constraints of more than 3% are summarized, with their precise values in Table A.5 in the first column.

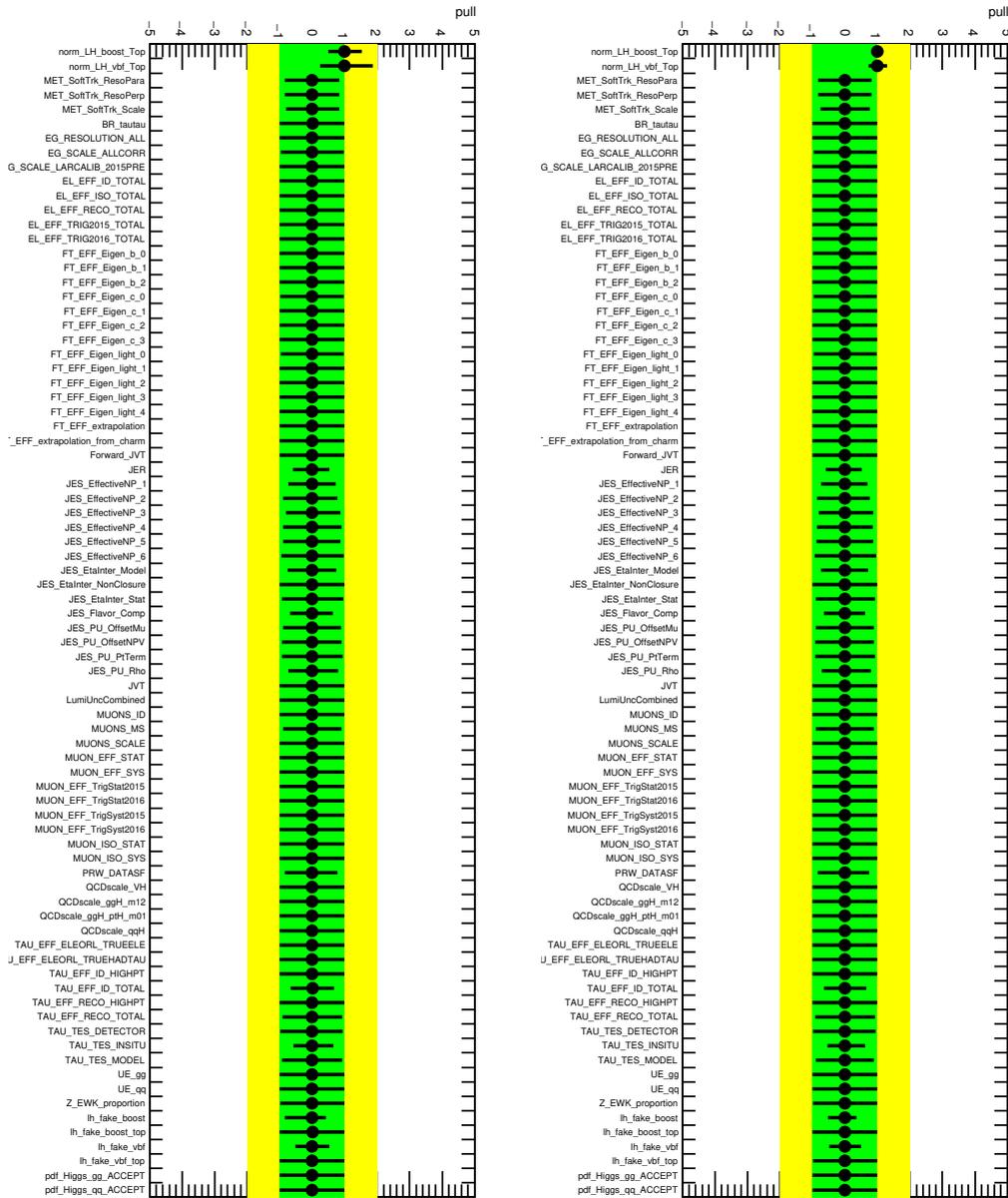
The reduction of the PoIs has only a small influence on the NP. Only the parameters, which has shown a constraint in the fit with all six PoI, has a small deviation. This deviation is illustrated in Table A.5. The first column refers to the fit with six PoIs and the second one to the fit with two PoIs. Overall the NPs have a trend to slightly smaller uncertainties for the fit with 2 PoIs. The impact on the NFs for the top quark initiated background in the last two rows is more obvious. Due to reduced degrees of freedom in the fit the uncertainties on the NFs are reduced.

#### Stage 1 expected

Figure A.5 shows the NPs and the NFs for the fit with 5 PoIs (a) and with the two beyond the SM STXS bins as PoIs (b) at Stage 1. Most of the NPs show no constraint, however the "fake" parameters are strongly constrained and the NFs have high uncertainties. The precise values for these parameters are given in Table A.4. This is a result of the large relative uncertainties on the parameters of interest. Due to the low sensitivity the fit does not need to constrain different parameters. Instead the fit constrains only the uncertainty of the "fake" parameters and introduces large uncertainties on the NFs.

	5 PoIs	only beyond SM bins	only $ggF$ bins
lh fake boost	$0^{+0.11}_{-0.15}$	$0^{+0.11}_{-0.14}$	$0^{+0.11}_{-0.15}$
lh fake vbf	$0^{+0.26}_{-0.23}$	$0^{+0.25}_{-0.22}$	$0^{+0.24}_{-0.21}$
lh fake boost top	$0^{+0.33}_{-0.32}$	$0^{+0.32}_{-0.33}$	$0^{+0.32}_{-0.33}$
lh fake vbf top	$0^{+0.39}_{-0.31}$	$0^{+0.38}_{-0.31}$	$0^{+0.38}_{-0.31}$
norm LH boost Top	$0.7^{+49}_{-0.7}$	$0.7^{+49}_{-0.7}$	$0.7^{+49}_{-0.7}$
norm LH vbf Top	$0.8^{+49}_{-0.8}$	$0.8^{+49}_{-0.8}$	$0.8^{+49}_{-0.8}$

Table A.4: Nuisance parameters of the "fake" background and normalisation factors of backgrounds initiated by top quark processes at Stage 1.



(a) 6 PoIs

(b) 2 PoIs

Figure A.4: Nuisance parameters and normalisation factors of STXS Stage 0 measurements for a fit with six parameters of interest (PoIs) (left) and a fit with two PoIs (right).

Parameter of interest	6	2
ATLAS JER	$0^{+0.52}_{-0.60}$	$0^{+0.51}_{-0.59}$
ATLAS JES EffectiveNP 1	$0^{+0.72}_{-0.74}$	$0^{+0.68}_{-0.73}$
ATLAS JES EffectiveNP 2	$0^{+0.79}_{-0.89}$	$0^{+0.76}_{-0.87}$
ATLAS JES EffectiveNP 3	$0^{+0.87}_{-0.81}$	$0^{+0.86}_{-0.80}$
ATLAS JES EffectiveNP 4	$0^{+0.90}_{-0.90}$	$0^{+0.86}_{-0.87}$
ATLAS JES EffectiveNP 5	$0^{+0.88}_{-0.90}$	$0^{+0.86}_{-0.88}$
ATLAS JES EffectiveNP 6	$0^{+0.98}_{-0.95}$	$0^{+0.97}_{-0.94}$
ATLAS JES EtaInter Model	$0^{+0.75}_{-0.75}$	$0^{+0.71}_{-0.74}$
ATLAS JES EtaInter Stat	$0^{+0.96}_{-0.93}$	$0^{+0.92}_{-0.90}$
ATLAS JES Flavor Comp	$0^{+0.63}_{-0.67}$	$0^{+0.61}_{-0.65}$
ATLAS JES PU OffsetMu	$0^{+0.88}_{-0.90}$	$0^{+0.88}_{-0.90}$
ATLAS JES PU OffsetNPV	$0^{+0.91}_{-0.92}$	$0^{+0.89}_{-0.91}$
ATLAS JES PU PtTerm	$0^{+0.93}_{-0.93}$	$0^{+0.92}_{-0.92}$
ATLAS JES PU Rho	$0^{+0.81}_{-0.74}$	$0^{+0.79}_{-0.71}$
ATLAS MET SoftTrk ResoPara	$0^{+0.83}_{-0.84}$	$0^{+0.81}_{-0.83}$
ATLAS MET SoftTrk ResoPerp	$0^{+0.84}_{-0.85}$	$0^{+0.82}_{-0.83}$
ATLAS MET SoftTrk Scale	$0^{+0.83}_{-0.80}$	$0^{+0.76}_{-0.76}$
ATLAS MUONS MS	$0^{+0.90}_{-0.89}$	$0^{+0.90}_{-0.89}$
ATLAS PRW DATASF	$0^{+0.77}_{-0.85}$	$0^{+0.75}_{-0.84}$
ATLAS TAU EFF ID TOTAL	$0^{+0.67}_{-0.67}$	$0^{+0.66}_{-0.65}$
ATLAS TAU EFF RECO TOTAL	$0^{+0.93}_{-0.92}$	$0^{+0.92}_{-0.91}$
ATLAS TAU TES DETECTOR	$0^{+0.95}_{-0.99}$	$0^{+0.95}_{-0.99}$
ATLAS TAU TES INSITU	$0^{+0.65}_{-0.57}$	$0^{+0.62}_{-0.54}$
ATLAS TAU TES MODEL	$0^{+0.91}_{-0.93}$	$0^{+0.89}_{-0.90}$
lh fake boost	$0^{+0.43}_{-0.85}$	$0^{+0.36}_{-0.53}$
lh fake vbf	$0^{+0.53}_{-0.52}$	$0^{+0.49}_{-0.49}$
lh fake boost top	$0^{+1.00}_{-1.00}$	$0^{+1.00}_{-1.00}$
lh fake vbf top	$0^{+1.00}_{-1.00}$	$0^{+0.99}_{-0.99}$
norm LH boost Top	$1.00^{+0.53}_{-0.50}$	$1.00^{+0.08}_{-0.07}$
norm LH vbf Top	$1.00^{+0.87}_{-0.76}$	$1.00^{+0.30}_{-0.27}$

Table A.5: Precise values of normalisation factors and nuisance parameters, with at least a uncertainty deviation of 3% from the expected uncertainty of one, of the STXS Stage 0 measurements.

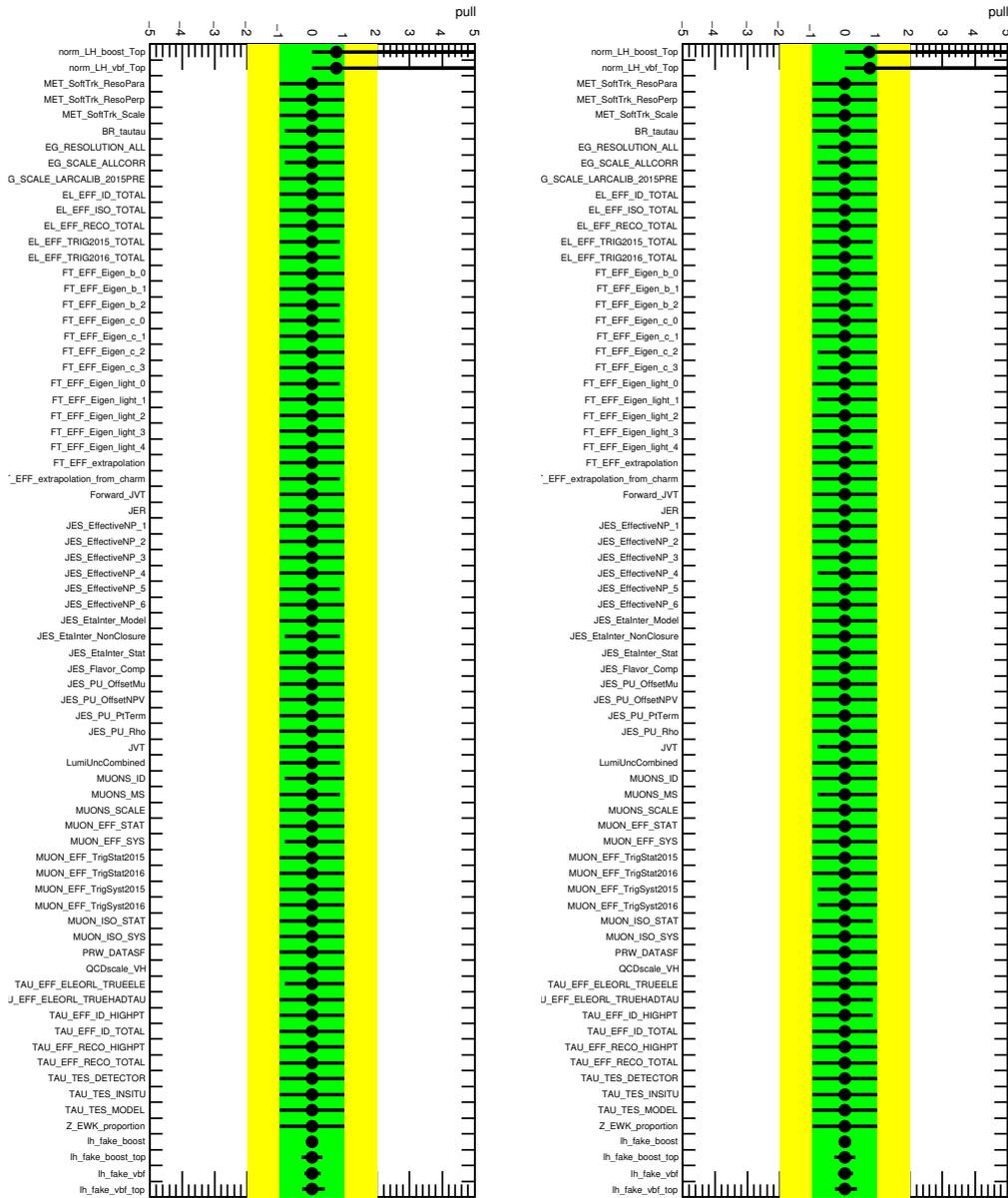


Figure A.5: Nuisance parameter and normalisation factors of STXS Stage 1 measurements for a fit with six parameter of interest (PoI) (left) and a fit with only two PoIs corresponding to the two beyond the SM bins (right).

The figure shows a constraint for the systematic uncertainties of the jet energy resolution  $JER$  as well as of the "fakes" estimation. While the jet energy resolution  $JER$  is still under study until beyond the scope of this thesis, the constraint of the "fake" systematics is introduced through the conservative uncertainty assignment in the *fake* estimation (see Section 10.4).

### Stage 0 observed

Figure A.6 shows the behaviour of the nuisance parameters and the normalisation factors of the STXS Stage 0 fit. The nuisance parameters show a similar behaviour as in the expected fit shown in Figure A.4b. The nuisance parameters related to the jet energy resolution (JER) and the jet energy scale (JES) show deviations from the expected value of zero but agree within their uncertainties with zero.

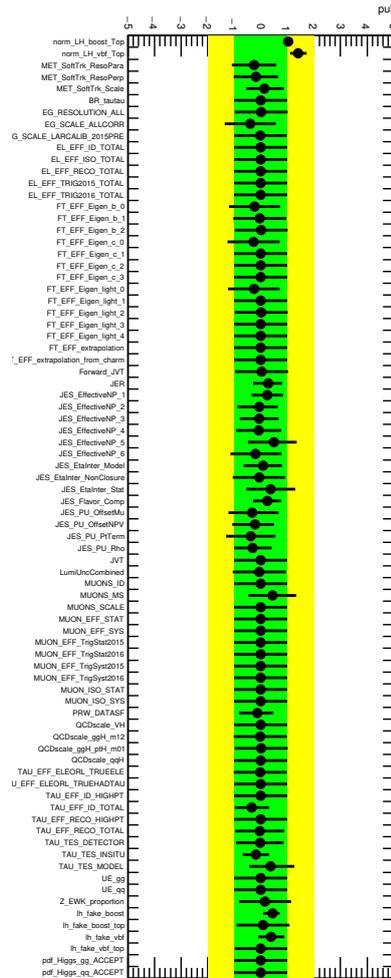


Figure A.6: Nuisance parameters and normalisation factors of STXS Stage 0 measurements for a fit with two parameters of interest ( $ggF$  and  $VBF$  cross section) for the combined 2015 and 2016 dataset of the  $H \rightarrow \tau_{lep} \tau_{had}$  decay channel measured at the ATLAS detector.

---

## REFERENCES

---

- [1] ATLAS Collaboration and CMS Collaboration. “Combined Measurement of the Higgs Boson Mass in  $pp$  Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments”. In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].
- [2] UA1 Collaboration. “Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c<sup>2</sup> at the CERN SPS Collider”. In: *Phys. Lett.* 126B (1983), p. 398. DOI: 10.1016/0370-2693(83)90188-0.
- [3] UA1 Collaboration. “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $\sqrt{s} = 540$  GeV”. In: *Phys. Lett.* 122B (1983), p. 611. DOI: 10.1016/0370-2693(83)91177-2.
- [4] CDF Collaboration. “Observation of top quark production in  $\bar{p}p$  collisions”. In: *Phys. Rev. Lett.* 74 (1995), p. 534. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002 [hep-ex].
- [5] UA2 Collaboration. “Evidence for  $Z^0 \rightarrow e^+ e^-$  at the CERN anti- $p p$  Collider”. In: *Phys. Lett.* 129B (1983), p. 130. DOI: 10.1016/0370-2693(83)90744-X.
- [6] UA2 Collaboration. “Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti- $p p$  Collider”. In: *Phys. Lett.* 122B (1983), p. 476. DOI: 10.1016/0370-2693(83)91605-2.
- [7] D0. “Observation of the top quark”. In: *Phys. Rev. Lett.* 74 (1995), p. 2632. arXiv: hep-ex/9503003 [hep-ex].
- [8] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.
- [9] Peter W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Phys. Lett.* 12 (1964), p. 125. DOI: 10.1016/0031-9163(64)91136-9.
- [10] Peter W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”. In: *Phys. Rev.* 145 (1966), p. 1156. DOI: 10.1103/PhysRev.145.1156.
- [11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (1964), p. 585. DOI: 10.1103/PhysRevLett.13.585.
- [12] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (1964), p. 321. DOI: 10.1103/PhysRevLett.13.321.
- [13] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (1964), p. 508. DOI: 10.1103/PhysRevLett.13.508.
- [14] T. W. B. Kibble. “Symmetry breaking in nonAbelian gauge theories”. In: *Phys. Rev.* 155 (1967), p. 1554. DOI: 10.1103/PhysRev.155.1554.

- [15] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett.* B716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [16] CMS Collaboration. “Observation of a new Boson at a Mass of 125 GeV With the CMS Experiment at the LHC”. In: (2015), p. 352. DOI: 10.1142/9789814623995\_0019.
- [17] ATLAS Collaboration and CMS Collaboration. “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: *ATLAS-CONF-2015-044* (2015). URL: <http://cds.cern.ch/record/2053103>.
- [18] ATLAS Collaboration. “Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector”. In: *Eur. Phys. J. C* 75.10 (2015). [Erratum: *Eur. Phys. J. C* 76, no. 3, 152 (2016)], p. 476. DOI: 10.1140/epjc/s10052-015-3685-1. arXiv: 1506.05669 [hep-ex].
- [19] CMS Collaboration. “Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV”. In: *Phys. Rev.* D92.1 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441 [hep-ex].
- [20] ATLAS Collaboration. “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”. In: *JHEP* 04 (2015), p. 117. DOI: 10.1007/JHEP04(2015)117. arXiv: 1501.04943 [hep-ex].
- [21] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].
- [22] Francis Halzen and Alan D. Martin. *Quarks and leptons: an introductory course in modern particle physics*. Wiley, 2010. ISBN: 9780471887416.
- [23] Michael E. Peskin and Dan V. Schroeder. *An Introduction To Quantum Field Theory (Frontiers in Physics)*. Westview Press, 1995. ISBN: 0201503972.
- [24] David Griffiths. *Introduction to Elementary Particles*. Wiley-VCH, 2008. ISBN: 3527406018.
- [25] D. Ebert. *Eichtheorien Grundlage Der Elementarteilchenphysik*. Wiley-VCH Verlag GmbH, 1989. ISBN: 3527278192.
- [26] G. V. Chibisov. “Astrophysical upper limits on the photon rest mass”. In: *Sov. Phys. Usp.* 19 (1976), p. 624. DOI: 10.1070/PU1976v019n07ABEH005277.
- [27] V. V. Ezhela, S. B. Lugovsky, and O. V. Zenin. “Hadronic part of the muon g-2 estimated on the  $\sigma^{2003(tot)}(e^+e^- \rightarrow \text{hadrons})$  evaluated data compilation”. In: (2003). arXiv: hep-ph/0312114 [hep-ph].
- [28] Peter Schmüser. *Feynman-Graphen und Eichtheorien für Experimentalphysiker (German Edition)*. Springer, 1994. ISBN: 3540584862.

- [29] David J. Gross and Frank Wilczek. “Ultraviolet Behavior of Nonabelian Gauge Theories”. In: *Phys. Rev. Lett.* 30 (1973), p. 1343. DOI: 10.1103/PhysRevLett.30.1343.
- [30] H. David Politzer. “Reliable Perturbative Results for Strong Interactions?” In: *Phys. Rev. Lett.* 30 (1973), p. 1346. DOI: 10.1103/PhysRevLett.30.1346.
- [31] Particle Data Group. “Review of Particle Physics”. In: *Chin. Phys.* C40.10 (2016). DOI: 10.1088/1674-1137/40/10/100001.
- [32] T. D. Lee and Chen-Ning Yang. “Question of Parity Conservation in Weak Interactions”. In: *Phys. Rev.* 104 (1956), p. 254. DOI: 10.1103/PhysRev.104.254.
- [33] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (1957), p. 1413. DOI: 10.1103/PhysRev.105.1413.
- [34] S. L. Glashow. “Partial Symmetries of Weak Interactions”. In: *Nucl. Phys.* 22 (1961), p. 579. DOI: 10.1016/0029-5582(61)90469-2.
- [35] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (1967), p. 1264. DOI: 10.1103/PhysRevLett.19.1264.
- [36] Abdus Salam and John Clive Ward. “Weak and electromagnetic interactions”. In: *Nuovo Cim.* 11 (1959), p. 568. DOI: 10.1007/BF02726525.
- [37] M. Gell-Mann. “The interpretation of the new particles as displaced charge multiplets”. In: *Nuovo Cim.* 4.S2 (1956), p. 19. DOI: 10.1007/BF02748000.
- [38] Kazuhiko Nishijima. “Charge Independence Theory of V Particles”. In: *Prog. Theor. Phys.* 13.3 (1955), p. 285. DOI: 10.1143/PTP.13.285.
- [39] Z. Maki, M. Nakagawa, and S. Sakata. “Remarks on the unified model of elementary particles”. In: (1962), p. 663. DOI: 10.1143/PTP.28.870.
- [40] K2K Collaboration. “Measurement of Neutrino Oscillation by the K2K Experiment”. In: *Phys. Rev.* D74 (2006), p. 072003. DOI: 10.1103/PhysRevD.74.072003. arXiv: hep-ex/0606032 [hep-ex].
- [41] B. Pontecorvo. “Neutrino Experiments and the Problem of Conservation of Leptonic Charge”. In: *Sov. Phys. JETP* 26 (1968), p. 984.
- [42] Makoto Kobayashi and Toshihide Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), p. 652. DOI: 10.1143/PTP.49.652.
- [43] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10 (1963), p. 531. DOI: 10.1103/PhysRevLett.10.531.
- [44] J. H. Christenson et al. “Evidence for the  $2\pi$  Decay of the  $K_2^0$  Meson”. In: *Phys. Rev. Lett.* 13 (1964), p. 138. DOI: 10.1103/PhysRevLett.13.138.
- [45] M. Schumacher. “Suche nach neutralen Higgs-Bosonen mit dem OPAL-Detektor am LEP2”. In: *Dissertation: BONN-IR-99-19* (1999).

- [46] Yoichiro Nambu. “Quasiparticles and Gauge Invariance in the Theory of Superconductivity”. In: *Phys. Rev.* 117 (1960), p. 648. DOI: 10.1103/PhysRev.117.648.
- [47] J. Goldstone. “Field Theories with Superconductor Solutions”. In: *Nuovo Cim.* 19 (1961), p. 154. DOI: 10.1007/BF02812722.
- [48] S. D. Drell and Tung-Mow Yan. “Partons and their Applications at High-Energies”. In: *Annals Phys.* 66 (1971), p. 578. DOI: 10.1016/0003-4916(71)90071-6.
- [49] C. Grojean. “Higgs Physics”. In: (2016), p. 143. DOI: 10.5170/CERN-2016-005.143. arXiv: 1708.00794 [hep-ph].
- [50] Particle Data Group. “ $\tau$  BRANCHING FRACTIONS”. In: *Particle Data Group* (2016). URL: <http://pdg.lbl.gov/2016/reviews/rpp2016-rev-tau-branching-fractions.pdf>.
- [51] Michael Hostettler and Giulia Papotti. “Luminosity Lifetime at the LHC in 2012 Proton Physics Operation”. In: 2013. URL: <http://JACoW.org/IPAC2013/papers/tupfi029.pdf>.
- [52] AC Team. “The four main LHC experiments”. 1999. URL: <https://cds.cern.ch/record/40525>.
- [53] ATLAS Collaboraion. *ATLAS public twiki*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2> (visited on 08/27/2017).
- [54] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.
- [55] ATLAS Collaboration. “ATLAS pixel detector electronics and sensors”. In: *JINST* 3 (2008), P07007. DOI: 10.1088/1748-0221/3/07/P07007.
- [56] J. Maluck. “Studien zur Optimierung und Massenrekonstruktion in der Suche nach dem Higgs-Boson des Standardmodells im Zerfall  $H \rightarrow \tau\tau \rightarrow ll+4\nu$  mit dem ATLAS-Experiment.” In: *Diplomarbeit: Albert-Ludwigs-Universität Freiburg* (2013).
- [57] A. Ruiz Martinez. “The Run-2 ATLAS Trigger System”. In: *J. Phys. Conf. Ser.* 762.1 (2016), p. 012003. DOI: 10.1088/1742-6596/762/1/012003.
- [58] ATLAS Collaboraion. *ATLAS public twiki*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults> (visited on 08/05/2017).
- [59] ATLAS Collaboration. “Measurement of the Higgs boson mass from the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  channels with the ATLAS detector using  $25 \text{ fb}^{-1}$  of  $pp$  collision data”. In: *Phys. Rev. D* 90.5 (2014), p. 052004. DOI: 10.1103/PhysRevD.90.052004. arXiv: 1406.3827 [hep-ex].
- [60] CMS Collaboration. “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”. In: *Eur. Phys. J. C* 75.5 (2015), p. 212. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].

- [61] ATLAS Collaboration. “Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector”. In: *Eur. Phys. J. C* 76.12 (2016), p. 658. DOI: 10.1140/epjc/s10052-016-4499-5. arXiv: 1602.04516 [hep-ex].
- [62] ATLAS Collaboration. “Evidence for the spin-0 nature of the Higgs boson using ATLAS data”. In: *Phys. Lett. B* 726 (2013), p. 1432. DOI: 10.1016/j.physletb.2013.08.026. arXiv: 1307.1432 [hep-ex].
- [63] ATLAS Collaboration. “Combined measurements of Higgs boson production and decay in the  $H \rightarrow ZZ^* \rightarrow 4l$  and  $H \rightarrow \gamma\gamma$  channels using  $\sqrt{s} = 13$  TeV pp collision data collected with the ATLAS experiment”. In: *ATLAS-CONF-2017-047* (2017). URL: <http://cds.cern.ch/record/2273854>.
- [64] ATLAS Collaboration. “Measurement of the Higgs boson mass in the  $H \rightarrow ZZ^* \rightarrow 4l$  and  $H \rightarrow \gamma\gamma$  channels with  $\sqrt{s}=13$ TeV  $pp$  collisions using the ATLAS detector”. In: *ATLAS-CONF-2017-046* (2017). URL: <https://cds.cern.ch/record/2273853>.
- [65] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].
- [66] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “PYTHIA 6.4 Physics and Manual”. In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].
- [67] Jun Gao et al. “CT10 next-to-next-to-leading order global analysis of QCD”. In: *Phys. Rev. D* 89.3 (2014), p. 6246. DOI: 10.1103/PhysRevD.89.033009. arXiv: 1302.6246 [hep-ph].
- [68] J. Pumplin et al. “New generation of parton distributions with uncertainties from global QCD analysis”. In: *JHEP* 07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195 [hep-ph].
- [69] Hung-Liang Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev. D* 82 (7 2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024.
- [70] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].
- [71] G. Corcella et al. “HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)”. In: *JHEP* 01 (2001), p. 010. DOI: 10.1088/1126-6708/2001/01/010. arXiv: hep-ph/0011363 [hep-ph].
- [72] ATLAS Collaboration. “ATLAS Run 1 Pythia8 tunes”. In: *ATL-PHYS-PUB-2014-021* (2014). URL: <https://cds.cern.ch/record/1966419>.

- [73] Z. Was. “TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ /... status report”. In: *Nucl. Phys. Proc. Suppl.* 98 (2001), p. 96. DOI: 10.1016/S0920-5632(01)01200-2. arXiv: hep-ph/0011305 [hep-ph].
- [74] T. Gleisberg et al. “Event generation with SHERPA 1.1”. In: *JHEP* 02 (2009), p. 4622. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].
- [75] Richard D. Ball et al. “Parton distributions for the LHC Run 2”. In: *Journal of High Energy Physics* 2015.4 (2015), p. 040. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2015)040.
- [76] Peter Zeiler Skands. “Tuning Monte Carlo Generators: The Perugia Tunes”. In: *Phys. Rev. D* 82 (2010), p. 3457. DOI: 10.1103/PhysRevD.82.074018. arXiv: 1005.3457 [hep-ph].
- [77] M. Asai. “Geant4-a simulation toolkit”. In: *Trans. Amer. Nucl. Soc.* 95 (2006), p. 757.
- [78] ATLAS Collaboration. “Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data”. In: *Eur. Phys. J. C* 74.7 (2014), p. 2941. DOI: 10.1140/epjc/s10052-014-2941-0. arXiv: 1404.2240 [hep-ex].
- [79] ATLAS Collaboration. “Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data”. In: *ATLAS-CONF-2016-024* (2016). URL: <https://cds.cern.ch/record/2157687>.
- [80] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton-proton collision data at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 76.5 (2016), p. 4120. DOI: 10.1140/epjc/s10052-016-4120-y.
- [81] ATLAS Collaboration. “Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data”. In: *The European Physical Journal C* 74.11 (2014), p. 3130. DOI: 10.1140/epjc/s10052-014-3130-x.
- [82] ATLAS Collaboration. “Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using  $pp$  collisions at  $\sqrt{s} = 13$  TeV”. In: *ATLAS-CONF-2017-029* (2017). URL: <https://cds.cern.ch/record/2261772>.
- [83] ATLAS Collaboration. “Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in  $pp$  collisions at  $\sqrt{s} = 8$  TeV”. In: *The European Physical Journal C* 75.7 (2015), p. 303. DOI: 10.1140/epjc/s10052-015-3500-z.
- [84] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *Eur. Phys. J. C* 77 (2017), p. 02934. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].
- [85] ATLAS Collaboration. *ATLAS web*. URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2016-003> (visited on 08/15/2017).

- [86] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *CERN-EP-2017-038* (2017). URL: <https://cds.cern.ch/record/2257300>.
- [87] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Anti-k(t) jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].
- [88] Matteo Cacciari and Gavin P. Salam. “Dispelling the N3 myth for the kt jet-finder”. In: *Physics Letters B* 641.1 (2006), p. 037. DOI: <http://dx.doi.org/10.1016/j.physletb.2006.08.037>.
- [89] ATLAS Collaboration. “Pile-up subtraction and suppression for jets in ATLAS”. In: *ATLAS-CONF-2013-083* (2013). URL: <http://cds.cern.ch/record/1570994>.
- [90] ATLAS Collaboration. “Expected performance of the ATLAS *b*-tagging algorithms in Run-2”. In: *ATL-PHYS-PUB-2015-022* (2015). URL: <https://cds.cern.ch/record/2037697>.
- [91] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: *ATL-PHYS-PUB-2015-027* (2015). URL: <https://cds.cern.ch/record/2037904>.
- [92] A. Elagin et al. “A New Mass Reconstruction Technique for Resonances Decaying to di-tau”. In: *Nucl. Instrum. Meth.* A654 (2011), p. 481. DOI: 10.1016/j.nima.2011.07.009. arXiv: 1012.4686 [hep-ex].
- [93] R. Keith Ellis et al. “Higgs Decay to tau+ tau-: A Possible Signature of Intermediate Mass Higgs Bosons at the SSC”. In: *Nucl. Phys.* B297 (1988), p. 221. DOI: 10.1016/0550-3213(88)90019-3.
- [94] G. Bohm and G. Zech. *Introduction to statistics and data analysis for physicists*. 2010. ISBN: 9783935702416.



---

## ACKNOWLEDGEMENTS

---

Ich möchte mich an dieser Stelle bei allen Personen bedanken, die mich während meines Masters- und Bachelorsstudiums begleitet und unterstützt haben. Einige dieser Personen möchte ich hier gerne besonders hervorheben.

Zunächst möchte ich mich bei Prof. Dr. Markus Schumacher bedanken für die Bereitstellung meines Themas und der guten Betreuung. Nicht nur während der Masterarbeit, sondern auch schon bei meiner Bachelorarbeit und bei diversen Gesprächen zwischen diesen beiden, in denen er immer wieder mein Interesse für Teilchenphysik verstärkt hat.

Einen großen Dank gilt auch meinem Betreuer Dr. Elias Coniavitis, welcher mich auch aus Genf aus hervoragend betreut hat trotz der großen Entefernung. Die Videogespräche und auch die Besuche in Genf haben immer wieder für große Fortschritte gesorgt.

Des weiteren möchte ich mich bei der ganzen Arbeitsgruppe bedanken für die angenehme Arbeitsatmosphäre. Insbesondere gilt mein Dank Dirk Sammel für die viele Zeit, die er sich immer wieder genommen hat, um mir bei alltäglichen Problemen zu helfen, sowie Dr. Ulrike Schnoor, welche mich vor allem fachlich in den letzten Monaten meiner Arbeit stark unterstützt hat.

Auch möchte ich mich bedanken bei allen Personen die immer wieder Teile meiner Arbeit Korrektur gelesen haben und mir Anregungen gegeben haben, wie Christina Schwemmbauer.

Zum Schluss möchte ich meine Familie, meinen Freunden und Mitbewohnerinnen danken. Insbesondere Monika Ellwarth ohne die ich mein Grundstudium nicht abgeschlossen hätte. Aber auch Greta Cambensi möchte ich danken, für die große Unterstützung während der gesamten Zeit und den zahlreichen notwendigen Ablenkungen, welche mir immer wieder neue Motivation gegeben haben.