

Search for CP violation in fermionic couplings of the Higgs boson generated by gluon-gluon fusion with the CMS Detector

von

Dominik Wolfschläger

Masterarbeit im Fach Physik

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften der
RWTH Aachen

im

September 2018

angefertigt am

III. Physikalischen Institut B

bei

Prof. Dr. Achim Stahl

Erstgutachter und Betreuer

Prof. Dr. Achim Stahl
III. Physikalisches Institut B
RWTH Aachen

Zweitgutachter

Priv. Doz. Oliver Pooth
III. Physikalisches Institut B
RWTH Aachen

Contents

1	The Higgs boson and CP violation in the Standard Model and beyond	1
1.1	The Standard Model	2
1.1.1	Elementary particles and interactions	3
1.1.2	The Higgs mechanism	5
1.2	CP violation	8
1.2.1	Symmetry transformations	9
1.2.2	CP violation in the Standard Model	10
1.2.3	Effective field Lagrangian	11
2	Experimental setup	15
2.1	The Large Hadron Collider	15
2.2	The CMS Detector	17
2.2.1	Detector subsystems	18
2.2.2	Trigger	20
2.3	Particle Flow	20
2.3.1	Leptons	21
2.3.2	Jets	22
2.3.3	Tau leptons	22
2.3.4	Missing transverse energy	24
3	Search for CP violation in gluon-gluon fusion events	25
3.1	Measuring CP properties in gluon-gluon fusion events	25
3.1.1	Baseline event selection	25
3.1.2	Event kinematics and CP sensitive observables	27
3.1.3	Simulated samples	29
3.1.4	Categorization	31
3.2	Background estimation	37
3.2.1	$Z \rightarrow \tau\tau/\ell\ell$	37
3.2.2	$t\bar{t}$	37
3.2.3	W+jets	39
3.2.4	QCD multijet background	47
3.2.5	VBF Higgs	55
3.2.6	Other backgrounds	56
3.3	Systematic uncertainties	57
3.3.1	Limited size of MC samples	57
3.3.2	Gluon fusion in association with jets	57
3.3.3	Jet energy correction uncertainties	57

3.4	Performance of the measurement of the CP mixing angle in gluon-gluon fusion	61
3.4.1	Signal model	61
3.4.2	Fit function	63
3.4.3	Results	65
4	Conclusion and Outlook	71
	Bibliography	73
A	Search for CP violation in gluon-gluon fusion events	79
A.1	Measuring CP violation in gluon-gluon fusion events	79
A.1.1	Categorization	79
A.2	Background estimation	83
A.2.1	Measurement of the QCD OS/SS factor	83
A.2.2	QCD control regions	86
A.3	Performance of the measurement of the gluon-gluon fusion CP mixing angle	87
A.3.1	Prefit distributions for MELA approach	87
A.3.2	Postfit distributions for MELA approach	95
A.3.3	Prefit distributions for $\Delta\phi_{jj}$	101
A.3.4	Postfit distributions for $\Delta\phi_{jj}$	109

The Higgs boson and CP violation in the Standard Model and beyond

The discovery of the Higgs boson in 2012 [1, 2] has been one of the greatest successes of the standard model (SM) of elementary particles, that is currently the most broadly accepted theory describing the fundamental constituents of matter and their interactions among each other. Although having defied unnumerous tests, it is insufficient because it cannot explain approved experimental observations such as the matter-antimatter asymmetry in the observable universe. Moreover, only three of the four known fundamental forces - the electromagnetic, strong, weak and gravitational force - can be described as a quantum theory. Currently, no verified quantum theory of gravitation exists and the SM and general relativity form two complementary descriptions of nature. There are theories which solve these problems by introducing new physics that goes beyond the description of the SM. In general physicists follow two approaches to search for physics beyond the standard model (BSM). These are direct searches for new particles and phenomena and indirect searches, where BSM physics leads to inconsistencies between measured SM parameters. Consequently, new physics would have been detected if any deviation from the prediction of the SM was observed in a precision measurement. Since the discovery of the Higgs boson, a major scientific effort has been performed to measure all the particles' properties. These include measurements of the mass m_H [3], the decay width Γ_H [4], the spin and parity quantum numbers [5, 6]. So far no deviation from the SM prediction has been found. The precise determination of the properties of the Higgs boson under parity and charge-conjugation transformations, namely the CP properties, is, however, still a major challenge in Higgs physics. Being predicted as a scalar $J^{PC} = 0^{++}$ particle, observations up to now only exclude the pure pseudoscalar scenario. A small pseudoscalar admixture to the prediction of the SM is still a valid scenario. The detection of such an admixture would prove CP violation in the Higgs sector, contributing potentially to the matter-antimatter asymmetry in the universe. This thesis presents a framework for an analysis that is dedicated to the precise measurement of the CP properties of the Higgs boson in the gluon-gluon fusion process. Beginning with a brief summary about the current understanding of elementary particles with the focus on the properties of the Higgs boson, a theoretical framework for performing a measurement of CP properties in fermionic couplings of the Higgs boson using an effective field theory (EFT) is motivated. In chapter 2 an overview about the CMS Detector and the LHC at CERN and the reconstruction of particles from recorded signals is given. Chapter 3 presents the setup, statistical tools, and quotes the result in terms of an expected sensitivity for the measurement of the CP mixing angle of the Higgs boson to top quarks.

1.1 The Standard Model

The SM describes the constituents of matter by *elementary particles* and the fundamental *electromagnetic, weak and strong forces*. Each particle is considered to be an excitation of a field with certain properties under symmetry transformations. Interactions between particles are based on the principle of *local gauge invariance* leading to bosons that mediate these interactions. The SM is mathematically formulated in the framework of relativistic quantum field theories (QFT) and is expressed through a Lagrange density $\mathcal{L}_{\text{SM}}$ that is invariant under local gauge transformations of the form $\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y$. The corresponding gauge bosons are 8 gluons with color charge resulting from the $\text{SU}(3)_C$, three weak bosons with weak isospin from $\text{SU}(2)_L$ and one boson with hypercharge from $\text{U}(1)_Y$. The importance of local gauge invariance can be inferred from t'Hooft's proof [7] that every locally gauge invariant QFT is renormalizable. Renormalizability enables the prevention of divergencies originating from energy ranges that lie beyond the scope of the theory by restricting the energy scale to the scale of the problem. Feynman diagrams, which are a graphical representation of perturbational calculus, are used to derive the matrix element that encapsulates all necessary information to calculate physics observables such as cross sections, angular distributions and decay rates.

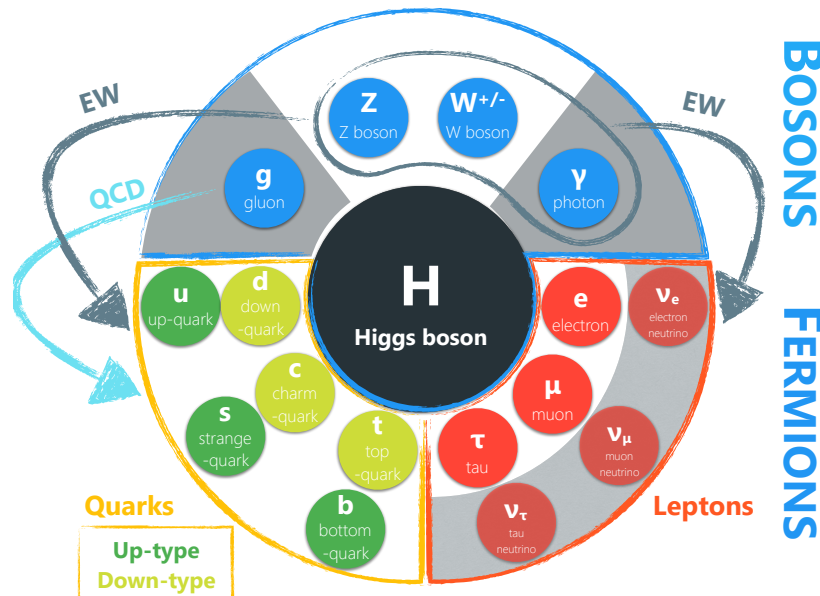


Fig. 1.1.: Particle content and description of the interactions of the SM. Fermions can be categorized into quarks and leptons. According to their similar quantum numbers, but very different masses, both types of fermions are further divided into three generations that come as left-handed doublets and right-handed singlets. The four gauge bosons are the mediators of the strong, electromagnetic and weak interactions. By means of electroweak symmetry breaking (EWSB), the Higgs boson gives rise to the masses of the weak bosons and all fermions except for the neutrinos, that are massless particles within the description of the SM and do also not form right-handed singlets. The arrows indicate that an interaction can take place between the associated bosons and fermions (EW: Electroweak interaction, QCD: Strong interaction). The shaded area behind some of the particle indicates that they do not couple directly to the Higgs boson and are as a consequence massless.

1.1.1 Elementary particles and interactions

Elementary particles are entities that do not constitute of more fundamental particles. Each particle possesses distinct quantum numbers such as mass, spin, electric charge, and color charge that precisely define its properties and possible behaviour regarding the interaction with other particles. Based on the spin, it is possible to group all particles into *fermions* with spin-1/2 and *bosons* with integer spin. The twelve fermions of the SM can be considered as the building blocks of matter. They come up as left-handed doublets and right-handed singlets in so-called generations with increasing mass but otherwise similar physical properties. *Quarks* carry color charge and are the only group of elementary particles that interacts via the strong interaction, which is modeled within the theory of quantum chromodynamics (QCD). They also have an electric charge of $+\frac{2}{3}e$ or $-\frac{1}{3}e$ for *up-type quarks* and *down-type quarks*, respectively, and carry weak isospin implying that they also take part in the electroweak interaction (EW), which is the unified theory describing the electromagnetic and weak interactions. The remaining six fermions are called *leptons*. The electron, muon and tau lepton have an electric charge and participate in the EW. The massless and chargeless neutrinos interact only weakly. Although it has been confirmed through the observation of neutrino-oscillations that they must possess very small masses [8, 9, 10], neutrinos are considered massless within the SM. Bosons are force-mediators and responsible for the interaction between fermions. Thereby, the Higgs boson plays an important role as it gives mass to the fermions and EW bosons, as discussed below. The sketch in figure 1.1 shows the particle content and the interactions of the SM.

1.1.1.1 Quantum Chromodynamics

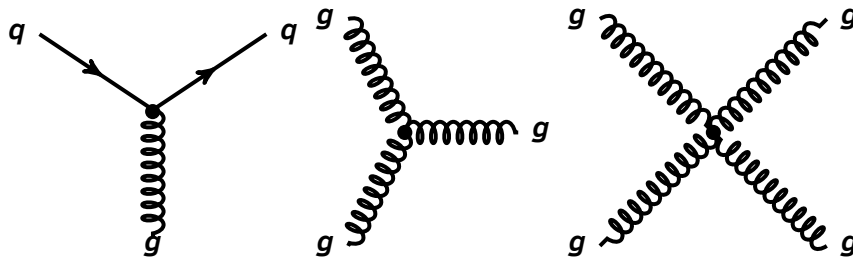


Fig. 1.2.: Standard vertices of QCD. Gluon coupling to a pair of quarks (left), three gluon-vertex (middle) and four-gluon vertex (right).

Gluons (g) are mediators of the strong force. There are eight gluon generators T^a , that are related to the 3×3 Gellmann-matrices λ^a for the $SU(3)_C$ symmetry group with the relation $T^a = \frac{1}{2}\lambda^a$, corresponding to the eight possible non-trivial combinations and superpositions of three different colors c and their associated anti-colors $\bar{c}$. As gluons possess an effective color charge $c\bar{c}$, too, they interact with themselves causing the possible standard vertices of QCD depicted in figure 1.2. The corresponding coupling structure for the QCD vertices in the Feynman diagrams is

$$\propto g_s \frac{1}{2} \lambda_{ij}^a \gamma^\mu \quad (1.1)$$

with a strong coupling constant g_s and color factor λ_{ij}^a , that is the ij -th component of the Gellmann matrix of the generator $a \in [1..8]$.

γ^μ is the vector of Dirac matrices as it appears in the Dirac equation [11].

The range of gluon interactions is restricted to the size of the color-neutral hadrons formed by the quarks. This *color confinement* is supported by the experimental observation that quarks and gluons do not propagate freely and hypothesizes that only particles with a net color charge equal to zero can propagate in free space. As a consequence, a quark or gluon being produced in a high-energy physics interaction undergoes the process of *hadronization* and produces a shower of hadrons, called *jet*.

1.1.1.2 Quantum Flavourdynamics (QFD)

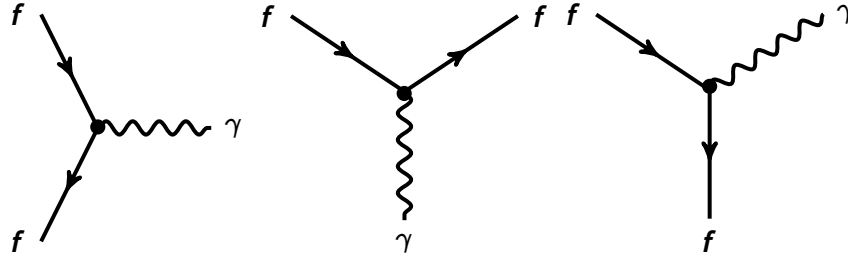


Fig. 1.3.: The three standard vertices of QED describe the annihilation of a fermion-antifermion pair into a photon (s-channel, left), the scattering of a fermion by exchange of a photon (t-channel, middle) and the emission of a photon (right).

The photon (γ) is the mediator of the electromagnetic force and couples to all particles that have an electric charge Q . Quantum electrodynamics (QED) introduces the photon field A that produces the vertices indicated in figure 1.3. The structure for the fermion current in the Feynman diagrams at an interaction vertex of QED for fermions described by the Dirac spinor u and the Dirac matrices is

$$\propto Q \bar{u} \gamma^\mu u. \quad (1.2)$$

The $W^\pm$ bosons are responsible for the flavour-changing charged current (FCCC) of the weak interaction and couple to all left-handed particles possessing weak isospin T_3 . The charged current produces vertices of the form

$$\propto g_W \frac{1}{2} \gamma^\mu (1 - \gamma^5), \quad (1.3)$$

where g_W is the weak coupling constant and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$.

The weak and the electromagnetic interactions are unified within the electroweak model of Glashow, Salam and Weinberg (GSW model) [12]. The neutral current (NC) induced by the Z boson is evoked by the third neutral generator W_3^0 of the $SU(2)_L$ gauge group to the weak isospin $T_W^{(3)}$ through mixing with a photon-like neutral field B that obeys a further $U(1)_Y$ gauge group to the weak hypercharge Y . Moreover, during the mixing the QED photon field A is retrieved naturally. The hypercharge is connected to both electric charge Q and weak isospin T by the relation

$$Y = 2 \left(Q - T_W^{(3)} \right). \quad (1.4)$$

The Feynman diagrams of the production of a W and a Z boson are shown in figure 1.4.

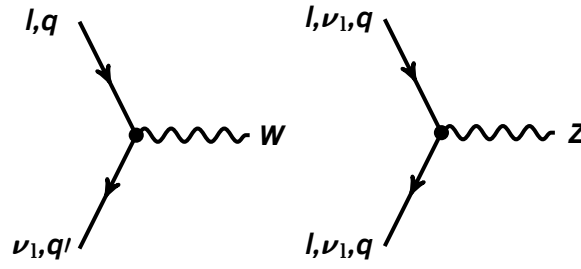


Fig. 1.4.: The standard vertices of QFD describe the production of a W boson by FCCC and the production of a Z boson. There are also t-channel diagrams which are not shown here. The fermions participating in the production of the W boson must possess left-handed helicity eigenvalues.

1.1.2 The Higgs mechanism

The Higgs mechanism is an integral part of the SM, because it ensures the self-consistency of the theory. Bosons that are introduced through a gauge symmetry are necessarily massless particles, because mass terms of the form $m^2 V^{\mu\nu} V_{\mu\nu}$ as they appear in the Klein-Gordon equation break the previously required local gauge invariance. Opposed to that claim, it was observed that the weak gauge bosons have very large masses, which are responsible for the short-rangeness of the weak interaction [13]. The Higgs mechanism proposed independently by Brout, Englert and Higgs et al. in 1964 [14, 15, 16, 17, 18] applies the principle of *spontaneous symmetry breaking* to a scalar field with a potential of the form given in figure 1.5.

In the SM two complex scalar fields are placed within a doublet of the weak isospin

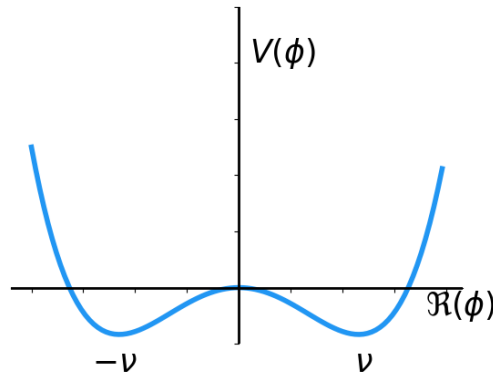


Fig. 1.5.: Projection onto the real plane of the Higgs field potential. The positions of the minima are degenerate and have a value that is different from zero. Thus, the non-zero vacuum expectation value $\nu = 246$ GeV breaks the symmetry of the Higgs field .

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \quad (1.5)$$

This gives rise to four degrees of freedom that create the masses of the two W bosons and the Z boson through interaction with the scalar field. The fourth degree belongs to the

massive boson corresponding to the scalar field. This boson is called *Higgs boson*. Without loss of generality the scalar field can be chosen to be a real constant by gauge fixing. In the unitary gauge the ground state is

$$\langle 0 | \phi | 0 \rangle = \begin{pmatrix} 0 \\ \nu \end{pmatrix} \quad (1.6)$$

with a real, positive constant ν . The Higgs mechanism creates terms of the form

$$\frac{1}{8} \nu^2 g_W^2 \left(W_\mu^{(1)} W^{(1),\mu} + W_\mu^{(2)} W^{(2),\mu} \right) + \frac{1}{8} \nu^2 \left(g_W W^{(3)\mu} - g' B^\mu \right) \left(g_W W_\mu^{(3)} - g' B_\mu \right) \quad (1.7)$$

in the Lagrangian. They have the form of mass terms, and therefore the masses of the weak gauge bosons can be read off or rather be derived within the GSW model to be

$$m_W = \frac{1}{2} g_W \nu \quad \text{and} \quad m_Z = \frac{1}{2} \nu \sqrt{g_W^2 + g'^2}. \quad (1.8)$$

This implies that by measuring the masses and coupling constants of the W bosons and the Z boson $m_{W/Z}$ and g_W/g' the vacuum expectation value can be calculated

$$\nu = 246 \text{ GeV}. \quad (1.9)$$

1.1.2.1 Higgs boson Yukawa couplings

The SM Higgs mechanism generates the masses of the W and Z bosons and solves the mass term problem for them. The different behaviour of left-handed and right-handed fermions, $\bar{f}_L$ and f_R , under $SU(2)_L$ transformation produces Dirac mass terms of the form

$$\mathcal{L}_f = -m \bar{f} f = -m (\bar{f}_L f_R + \bar{f}_R f_L) \quad (1.10)$$

that break the gauge symmetry again. Nevertheless, it can be shown that terms of the form

$$-y_f [\bar{L} \phi R + \bar{R} \phi^\dagger L] \quad (1.11)$$

and equivalent terms with the complex conjugate Higgs doublet

$$\phi_C = \begin{pmatrix} -\nu - H \\ 0 \end{pmatrix} \quad (1.12)$$

of the form

$$-y_f [\bar{L} \phi_C R + \bar{R} \phi_C^\dagger L] \quad (1.13)$$

are invariant under transformations of the complete $SU(2)_L \times U(1)_Y$ symmetry group. Thereby, L and R denote the left-handed fermion doublet and right-handed fermion singlet, respectively. For this reason, it is possible to retrieve expressions that have the form of Dirac mass terms

$$\mathcal{L}_{\text{mass}} \propto -y_f \nu (\bar{f}_L f_R + \bar{f}_R f_L) = -y_f \nu \bar{f} f. \quad (1.14)$$

The Yukawa coupling constants y_f , that were introduced above, are not predicted by the SM Higgs mechanism and must be measured. This means, y_f is proportional to the fermion mass by construction

$$y_f = \frac{m_f}{\nu}. \quad (1.15)$$

Moreover, interaction terms between the fermions and the Higgs boson appear that are proportional to the Yukawa couplings

$$\mathcal{L}_Y \propto -y_f H (\bar{f}_L f_R + \bar{f}_R f_L) = -y_f \bar{f} f H. \quad (1.16)$$

1.1.2.2 Properties of the Higgs boson

Given the mass of the Higgs boson of 125.18 GeV [19] the coupling strengths to other particles are fixed within the SM and production cross sections in proton-proton collisions and branching ratios can be predicted. The two most important production processes are shown in figure 1.6.

At the LHC (see section 2.1) the most efficient production mechanism is the gluon-gluon

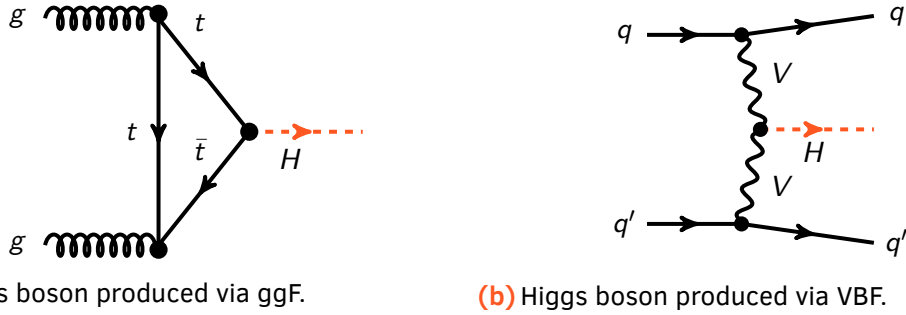


Fig. 1.6.: Feynman graphs of the two most efficient production processes of the Higgs boson.

fusion process (ggF) (see fig. 1.6a). The LHC is a proton collider which means the energy of the actual collision is determined by the amount of momentum that the two colliding partons (gluon or quark) carry inside the proton. The momentum is a fraction x of the total proton momentum p_p and defined by the parton density functions that depict the probability that a parton carries a momentum fraction x [20]. At the LHC energies the parton density function for gluons peaks at low values of x . These x values are matching to the production of a Higgs boson at 125 GeV mass. Together with the Higgs coupling to the top quark which is in the order $\mathcal{O}(1)$, this leads to the relatively high cross section of $\sigma_{gg \rightarrow H} = 43.92 \text{ pb}$ [21] for the ggF production process. Later in this thesis, ggF events with two initial-state jets are selected to study the CP properties of the Higgs boson (see fig. 1.7, page 12). These events are a next-to-next-to-leading-order (NNLO) QCD contributions and result, thus, in a cross section that is suppressed by a factor of g_s^2 . Nevertheless, the NNLO ggF process contributes with a cross section of $\sigma_{gg \rightarrow Hjj} = 3.98 \text{ pb}$, which is still a bit larger than the contribution from vector boson fusion (VBF) with a cross section of $\sigma_{VV \rightarrow H} = 3.748 \text{ pb}$. As VBF is characterized by two forward jets originating from the remnants of the two quarks that radiate off the vector bosons producing the Higgs boson, the topology of VBF is very similar to that of ggF events in association with two jets.

Tab. 1.1.: Decay modes and branching ratios (BR) of the SM Higgs boson [19].

Decay mode	BR / %	Remark
$H \rightarrow b\bar{b}$	57.8	Highest branching fraction, but large QCD background [23].
$H \rightarrow WW^*$	21.6	Off-shell produced W-boson [24].
$H \rightarrow gg$	8.6	Not detectable at LHC.
$H \rightarrow \tau^+\tau^-$	6.4	Largest direct coupling to leptons. Relatively low background compared to $b\bar{b}$ [22].
$H \rightarrow c\bar{c}$	2.9	Not discovered yet, very large QCD backgrounds.
$H \rightarrow ZZ^*$	2.7	Utilized for Higgs boson discovery. Very low background in $ZZ \rightarrow 4l$ [1, 2].
$H \rightarrow \gamma\gamma$	0.2	Utilized for Higgs boson discovery. Clear signature and low backgrounds [1, 2].

The Higgs production is decoupled from the decay because no spin information is transferred through a scalar or pseudoscalar resonance. For this reason, Higgs boson production processes can be studied with any possible decay mode. The branching ratios of the most frequent decay modes are given in table 1.1.

The decay into a pair of tau leptons $H \rightarrow \tau^+\tau^-$ is used to identify Higgs bosons in this thesis as this analysis built upon the SM search for the Higgs boson decaying into tau leptons [22]. This provides on the one hand a relatively high yield of Higgs bosons (see tab. 1.1) and on the other hand a settled framework for the selection of Higgs bosons. In the following the $H \rightarrow \tau^+\tau^-$ decay is abbreviated by $H \rightarrow \tau\tau$ keeping in mind that the tau leptons in the final state have electric charges.

1.2 CP violation

Investigating the behaviour of physics processes under symmetry transformations can provide fundamental information about the nature of the interaction. For example, physical processes are studied regarding the situation that

- the direction of time is inverted or
- the coordinate system is spatially inverted.

Invariance under these transformations gives fundamental insight into the physics of the process, because according to Emmy Noethers' theorem any symmetry leads to a conserved quantity [25, 26]. This section gives an overview about how elementary particle interactions behave under parity, charge-conjugation and time-reversel transformations and the implications of violations from these laws.

1.2.1 Symmetry transformations

The SM is based on the CPT-theorem, that implies that every QFT is invariant under the combination of

- parity transformation $\mathcal{P}$, that inverts the coordinate system with the transformation $\vec{r} \rightarrow -\vec{r}$,
- charge-conjugation $\mathcal{C}$, that inverts the sign of every additive quantum number of the particle, thus turning a particle into an anti-particle and vice-versa, and
- time-reversel $\mathcal{T}$, that inverts the sign of the time component $t \rightarrow -t$ and consequently changes the direction of time [27, 28].

The interaction terms appearing in the Lagrangian are lorentz-invariant bilinear forms that provide a distinct behaviour under these transformations. An overview about the different structures and their properties is given in table 1.2.

Tab. 1.2.: Behaviour of bilinear forms under $\mathcal{P}$, $\mathcal{C}$ and $\mathcal{T}$ transformations. The proofs are summarized in Ref. [29]. Here, $(-1)^\mu$ denotes that the transformation is +1 if $\mu = 0$ and -1 for $\mu \in [1, 2, 3]$, where μ is the μ -th component of the four-vector.

Type	Form	$\mathcal{P}$	$\mathcal{C}$	$\mathcal{T}$
Scalar	$\bar{\psi}\phi$	+1	+1	+1
Pseudoscalar	$\bar{\psi}\gamma^5\phi$	-1	+1	-1
Vector	$\bar{\psi}\gamma^\mu\phi$	$(-1)^\mu$	-1	$(-1)^\mu$
Axial vector	$\bar{\psi}\gamma^\mu\gamma^5\phi$	$-(-1)^\mu$	+1	$(-1)^\mu$
Tensor	$\bar{\psi}\sigma^{\mu\nu}\phi$	$(-1)^\mu(-1)^\nu$	-1	$-(-1)^\mu(-1)^\nu$

From table 1.2 it can be inferred that

- QED and QCD are invariant under CP-transformations because their fermion currents are vector couplings as given in sections 1.1.1.1 and 1.1.1.2.
- The charged current interaction maximally violates both parity and charge-conjugation transformations individually as experimentally proven by [30]. Mathematically, the coupling structure of the weak interaction of the form $\bar{f}\gamma^\mu - \gamma^\mu\gamma^5 f$ explains this. The consequence of this violation is that there are no couplings to right-handed neutrinos and left-handed anti-neutrinos in the SM [31].

Any quantum state that is eigenstate of the corresponding transformation has the eigenvalue ± 1 . For the CP transformation this implies, for instance,

$$\mathcal{CP} |\psi_{\text{CP}}^\pm\rangle = \pm |\psi_{\text{CP}}^\pm\rangle, \quad (1.17)$$

where +1 and -1 are the eigenvalues of the CP-even and CP-odd eigenstates, respectively. Any superposition of CP-even and CP-odd states is no longer an eigenstate of the CP transformation (see eq. (1.18)). Thus, CP violation is observed:

$$\mathcal{CP} |\psi_{\text{CP}}\rangle = \mathcal{CP} (a \cdot |\psi_{\text{CP}}^+\rangle + b \cdot |\psi_{\text{CP}}^-\rangle) = a \cdot |\psi_{\text{CP}}^+\rangle - b \cdot |\psi_{\text{CP}}^-\rangle \neq \pm |\psi_{\text{CP}}\rangle. \quad (1.18)$$

1.2.2 CP violation in the Standard Model

Summarizing the properties of the weak interaction under C and P transformations above, one would expect that the weak interaction is at least invariant under simultaneous C and P transformations. In contrast to that, evidence for CP violation was reported in neutral Kaon decays and mixing [32, 33, 34] and in B meson systems [35, 36, 37, 38, 39]. Recently, the T2K experiment reported indications that CP conservation is also not conform with the neutrino sector [40] pointing towards CP violation also there. CP violation in the SM has its source in the complex phase of the 3x3 CKM-matrix which relates the mass eigenstates of the quarks to their weak eigenstates that couple to the weak gauge bosons [30]. CP violation occurs because this complex phase is irreducible and appears, thus, in the matrix elements for box diagrams in mixing processes, for example. For a brief overview about the occurrence of CP violation in the SM the reader is pointed to Ref. [11]. A complete and detailed theoretical description, even beyond the SM, is given in Ref. [41].

1.2.2.1 CP violation and the early universe

According to the big bang theory, matter and antimatter must have been produced in equal amounts initially. This implies that all matter should have been annihilated into gamma rays when the density, and consequently the interaction rate, was still high enough in the early universe. Obviously some matter survived the annihilation phase, leading to the formation of stars and galaxies. Measurements strongly indicate that no major abundances of antimatter are present in the observable universe today. Quantitatively, this can be inferred from the matter-antimatter asymmetry $\eta_{\text{Asymmetry}}$ that is defined by the difference of baryon n_B and antibaryon $n_{\bar{B}}$ densities in relation to the cosmic microwave background photon density

$$\eta_{\text{Asymmetry}} = \frac{n_B - n_{\bar{B}}}{n_\gamma} \sim 10^{-9}. \quad (1.19)$$

In principle, the SM provides the theoretical framework required by the Sakharov conditions [42] that describe a physical configuration leading to a matter-antimatter asymmetry:

- Baryon number violation,
- CP violation and
- a deviation from thermal equilibrium.

An estimate for the total amount of CP violation in the SM can be derived using the Jarlskog determinant $\mathcal{J}$ as defined in Ref. [43]. A dimensionless quantity can be constructed by dividing the temperature present at the condensation of the Higgs field, when the masses of the quarks were generated, at a temperature $T_{\text{EWSB}} = 125 \text{ GeV}$

$$A_{\text{CP}} = \frac{\mathcal{J}}{T_{\text{EWSB}}^2} \approx 10^{-19} \ll \eta \sim 10^{-9}. \quad (1.20)$$

Comparing this roughly estimated value to the observed baryon asymmetry, it can be seen that the CP nonconservation in the SM is not large enough to explain the Baryogenesis. Therefore, the search for potentially larger sources of CP violation beyond the SM is strongly encouraged.

1.2.3 Effective field Lagrangian

In this analysis an approach utilizing an effective field theory (EFT) is followed. Thereby, the SM Lagrangian $\mathcal{L}_{\text{SM}}$ is modified with kinematically possible new terms. Often processes are introduced at a cut-off scale Λ . Models relevant for studies aiming to determine the spin and parity (J^{PC}) properties of the Higgs boson are constructed by taking the SM Lagrangian, excluding the Higgs boson, and adding new terms associated with a boson H with model-specific J^{PC} properties. An EFT provides some advantages compared to other types of BSM theory types, that are, for instance, that

- only a limited set of new parameters and combinations of parameters needs to be introduced and studied,
- the SM is naturally recovered at low energies,
- every EFT is fully renormalizable since orders of $\sqrt{s}/\Lambda$ in the expansion become small,
- if necessary, the model may be extended by adding higher order effects or by introducing further interactions and particles.

As described in Ref. [44] such a Lagrangian for a spin-0 boson has the form

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM Higgs}} + \mathcal{L}_{H(J=0)}. \quad (1.21)$$

For the analysis in this thesis a Lagrangian $\mathcal{L}_{H(J=0)}$ is used that

- allows for a mixing between scalar and pseudoscalar couplings of the boson,
- is limited to three-point interactions between H , fermions and vector bosons, meaning that only Hff and HVV vertices are considered, and
- assumes that the resonance at 125 GeV belongs to a single boson.

1.2.3.1 Determining the Higgs Yukawa coupling to the top quark

The CP properties of the Higgs boson can be fully deduced from its couplings in the production and its decay because it is a scalar or a pseudoscalar particle. There are studies dedicated to the measurement of the CP properties in couplings to vector bosons [45] and to tau leptons in the tau decay channel [46]. This analysis is dedicated to measure the CP properties in fermionic couplings in the production process which means that the Yukawa coupling to quarks needs to be investigated. At the LHC Higgs bosons are predominantly produced via ggF. As gluons are massless particles they do not couple directly to the Higgs boson. An intermediate quark loop is needed, and thus the effective coupling to gluons in the ggF process is studied to examine the Yukawa couplings of the Higgs boson. In principle, any quark flavour can run in the loop between the gluons and the Higgs boson but due to the large coupling of the top quark to the Higgs boson it becomes the dominant

contribution. In Ref. [47] it was shown that the heavy-top quark limit $m_t \rightarrow \infty$, where the top quark is much heavier than the Higgs boson, is a very good approximation for these studies. Therefore, this simplification is considered within this analysis implying that the Yukawa coupling to the top quark is directly inferred from the effective ggH coupling. Compared to the VBF production process which occurs via the t-channel, there are no visible remnants that can be used to deduce the coupling parameters. Therefore, two initial-state partons need to be radiated off as indicated in the Feynman graphs in figure 1.7. These appear as two jets in the detector. Consequently, the signature of the ggF events, that is used to study the fermionic couplings of the Higgs boson, is very similar to the typical VBF Higgs production topology with its two forward jets. For this reason, the $qq \rightarrow H$ process is considered as a background in this analysis.

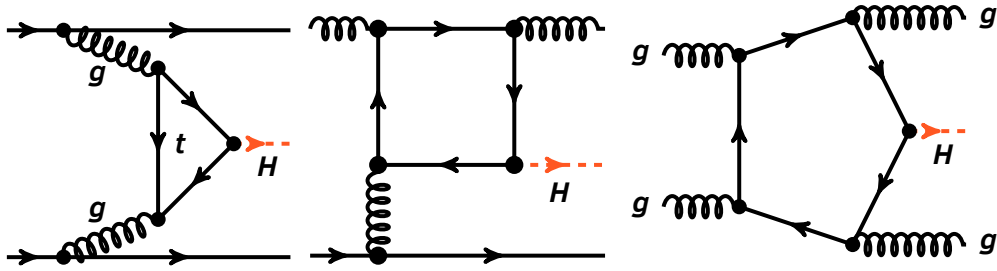


Fig. 1.7.: Feynman graphs for the production of a Higgs boson via ggF in association with two initial state radiation partons $gg \rightarrow H + 2j$.

Using these constraints, it is possible to derive the most general tensor structure of the ggF process in the heavy-top quark limit. From all triangular Feynman diagrams the tensor structure is calculated applying Furry's theorem and using Passarino-Veltman contraction [21]. This yields

$$T^{\mu\nu} = a_1 g^{\mu\nu} + a_2 (q_1 \cdot q_2 g^{\mu\nu} - q_1^\nu q_2^\mu) + a_3 \epsilon^{\mu\nu\alpha\beta} q_{1\alpha} q_{2\beta}. \quad (1.22)$$

Thereby, the q_i denote the four-momenta of the incoming gluons and the a_i are constant factors that are analytically evaluated within the heavy-top quark approximation (see eq. (1.23)). There is no direct coupling between the Higgs boson and gluons in the SM implying that $a_1 = 0$. The term proportional to a_2 causes the scalar couplings, the term proportional to a_3 evokes pseudoscalar couplings. They are directly related to the scalar and pseudoscalar Yukawa coupling strengths y_t and $\tilde{y}_t$ and are normalized here to the SM Yukawa coupling strength $y_t^{\text{SM}} = \frac{m_t}{v}$:

$$a_2 = \frac{y_t}{y_t^{\text{SM}}} \frac{g_s}{3\pi\nu} \quad a_3 = \frac{\tilde{y}_t}{y_t^{\text{SM}}} \frac{g_s}{2\pi\nu}. \quad (1.23)$$

Consequently, the part of the Lagrangian describing the couplings of the CP-odd - denoted as A - and CP-even - denoted as H - components of the Higgs boson can be written as

$$\mathcal{L}_{ggH} = \frac{y_t}{y_t^{\text{SM}}} \frac{g_s}{3\pi\nu} H G^{\mu\nu} G_{\mu\nu} + \frac{\tilde{y}_t}{y_t^{\text{SM}}} \frac{g_s}{2\pi\nu} A G^{\mu\nu} G_{\rho\sigma} \epsilon^{\rho\sigma\mu\nu}. \quad (1.24)$$

If the scalar and pseudoscalar Yukawa coupling constants, y_t and $\tilde{y}_t$, are assumed to be equal, the coupling factors can be expressed in terms of an overall coupling modifier κ and a CP mixing angle α_t , that describes the mixing in the fermionic couplings of the Higgs boson to the top quark,

$$\mathcal{L}_{ggH} = \kappa_{ggH} \cos(\alpha_t) H G^{\mu\nu} G_{\mu\nu} + \kappa_{ggH} \sin(\alpha_t) A \frac{3}{2} G^{\mu\nu} G_{\rho\sigma} \epsilon^{\rho\sigma\mu\nu}. \quad (1.25)$$

This means a pseudoscalar admixture is evoked solely under the influence of α_t . A modification of the rate of the ggF process may be due to other reasons such as heavy BSM particles running in the loop between the gluons and the Higgs boson. For $\alpha_t = 0$ the SM scenario of a scalar Higgs boson coupling to top quarks is retrieved, while $\alpha_t = \frac{\pi}{2}$ describes a purely pseudoscalar coupling. The ultimate goal of this analysis is thus the measurement of α_t . If both the scalar and pseudoscalar couplings are assumed to be at SM strength, the coupling to a pseudoscalar Higgs boson is expected to be larger by a factor $\frac{3}{2}$ compared to the scalar coupling. It is important to note, that the CP mixing angle introduced in equation (1.25) is unique for the coupling of the Higgs boson to the top quark. In general, CP mixing angles may be different for couplings to other particles. Throughout this thesis, α refers to the mixing angle as defined by equation (1.25), unless otherwise stated.

1.2.3.2 Pseudoscalar admixture in VBF

Following the same procedure as above, pseudoscalar admixtures can also be introduced for Higgs boson couplings to vector bosons. As given in Ref. [44, 48] the EFT Lagrangian for the coupling to vector bosons is

$$\mathcal{L}^{W/Z} = \left\{ \cos(\alpha_V) \kappa_{SM} \left[\frac{1}{2} g_{HZZ} Z_\mu V^\mu + g_{HWW} W_\mu^+ W^{-\mu} \right] \right. \quad (1.26)$$

$$\left. - \frac{1}{4\Lambda} [\cos(\alpha_V) \kappa_{HZZ} Z^{\mu\nu} Z_{\mu\nu} + \sin(\alpha_V) \kappa_{AVV} Z^{\mu\nu} Z_{\rho\sigma} \epsilon^{\rho\sigma\mu\nu}] \right. \quad (1.27)$$

$$\left. - \frac{1}{2\Lambda} [\cos(\alpha_V) \kappa_{HWW} W^{-\mu\nu} W_{\mu\nu}^+ + \sin(\alpha_V) \kappa_{AWW} W^{-\mu\nu} W_{\rho\sigma}^+ \epsilon^{\rho\sigma\mu\nu}] \right\} X, \quad (1.28)$$

where X is a spinless boson that is a superposition of H and A . κ_{SM} , $\kappa_{H/AZZ}$ and $\kappa_{H/AWW}$ are the corresponding coupling modifiers and α_V is the CP mixing angle for CP violation in the vector boson coupling of the Higgs boson. First of all, it is emphasized that any pseudoscalar couplings occurs only at a higher-order that is suppressed by the cut-off scale Λ . This means, any effects from CP violation in the VBF production process occurs on an perturbational order that is suppressed compared to the scalar SM-like interaction. This is not the case in the ggF process investigated in this analysis. Here, scalar and pseudoscalar couplings occur at the same order and can thus yield a sizeable pseudoscalar admixture even if α_t has small values. A framework for the characterization of a potential pseudoscalar admixture in vector boson couplings is given in Ref. [49] and followed in Ref. [45]. CP violation is modeled by a fractional coupling parameter f that describes the ratio of pseudoscalar coupling compared to the overall coupling. If $f = 0$ no pseudoscalar contribution is present and the SM scenario in the VBF process is obtained.

Experimental setup

Today, the Higgs boson and particles of the second and third generation do not appear very frequently in the universe anymore. When the temperature dropped below their rest masses, they left thermal equilibrium and decayed more or less instantly into more stable particles. They are created in high-energy processes such as the interaction of cosmic rays with the atmosphere making it difficult or impossible to study their properties in nature. Elementary particles have to be produced artificially to study their properties in a reproducible and unbiased environment. To reach energy densities comparable to the first seconds of the universe, physicists build particle colliders that accelerate charged, stable particles or nuclei to highest energies and scatter them at fixed targets or collide them with an oppositely accelerated beam of particles transforming such kinetic energy eventually into mass. Sophisticated detector systems, constructed to detect and reconstruct particles produced in the collisions, are built around the interaction point. Thus, the signature of the decay of an unstable particle eventually created in a particle collision, is recorded. In this thesis, an analysis using a data set of proton-proton collisions collected at the Large Hadron Collider (LHC) from Run 2 corresponding to an integrated luminosity of 35.9 fb^{-1} is performed. The data was recorded by the CMS Detector during the year 2016.

2.1 The Large Hadron Collider

The LHC at CERN is a particle accelerator and storage ring that is capable of accelerating protons to a maximum energy of currently 6.5 TeV and bringing them into collision at so-called interaction points. The ring has a circumference of 26.7 km and is operated in the old LEP tunnel near Geneva. It is currently the largest and most powerful particle accelerator in the world. A sketch showing the full LHC complex including relevant pre-acceleration steps is given in figure 2.1. The LHC uses superconducting radio-frequency cavities (RF cavities) operating at a frequency of 400.8 MHz to accelerate proton bunches in opposite directions. Superconducting dipole magnets with field strengths of up to 8 T force the bunches on a circular orbit [50]. The LHC as storage ring keeps the beams in circulation with a lifetime of ~ 12 h. During that time proton bunches collide every 25 ns at the four interaction points [50]. The four major experiments ALICE, ATLAS, CMS and LHCb are built around the interaction points and execute each a tailor-made physics programme. While ATLAS and CMS are general purpose experiments with a broad physics programme covering SM precision tests, Higgs sector investigations, searches for Supersymmetry (SUSY) and other exotic new physics, ALICE is examining the quark-gluon plasma in heavy-ion collisions and LHCb is specialized in measuring CP properties in B-meson decays.

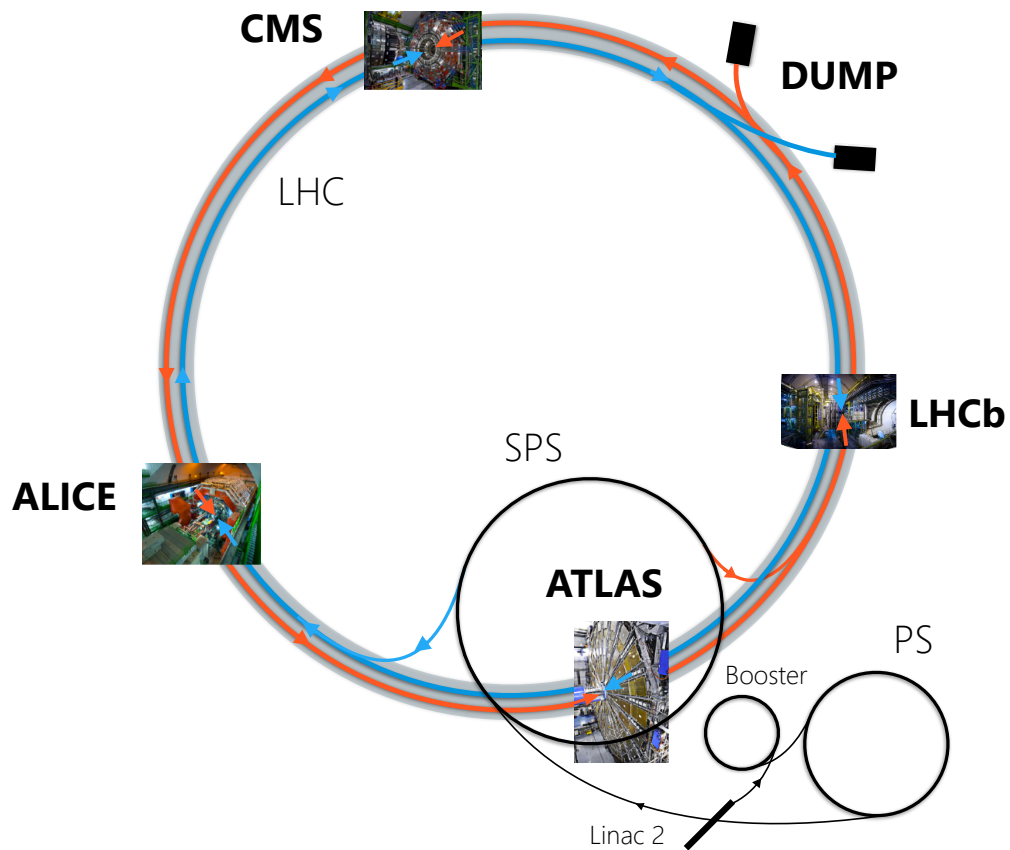


Fig. 2.1.: Schematic overview of the LHC accelerator complex condensed from Refs. [51, 52, 53, 54, 55, 56, 57]. Before the protons enter the actual LHC ring, they are pre-accelerated up to 450 GeV in the accelerator chain beginning with the LINAC 2. The particles increase their energy in the Booster and become further accelerated in the Proton Synchrotron (PS). They reach an energy of 450 GeV in the Super Proton Synchrotron (SPS). At the end of its lifetime the beam is directed into a block of carbon, the *beam dump*, to not damage the machine.

2.2 The CMS Detector

The 12.9 m long superconducting solenoid with a field strength of 3.8 T is the key feature of the *Compact Muon Solenoid* (CMS). The detector is built around the interaction point covering almost 4π of solid angle and is designed to handle the high luminosities present at the LHC to measure and reconstruct every particle produced in the 40 million bunch crossings every second. The volume within the magnet's internal diameter of 6 m is large enough to contain the three layers of silicon pixel detector¹ very near to the interaction point, the silicon strip detector, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The muon system is embedded in the iron return yoke outside of the magnet and consists of three to four layers depending on the part of the detector (see fig. 2.2). As in the following only a brief overview on the functionality of the detector components is given, the reader is pointed to Refs. [58, 59] for a more detailed description.

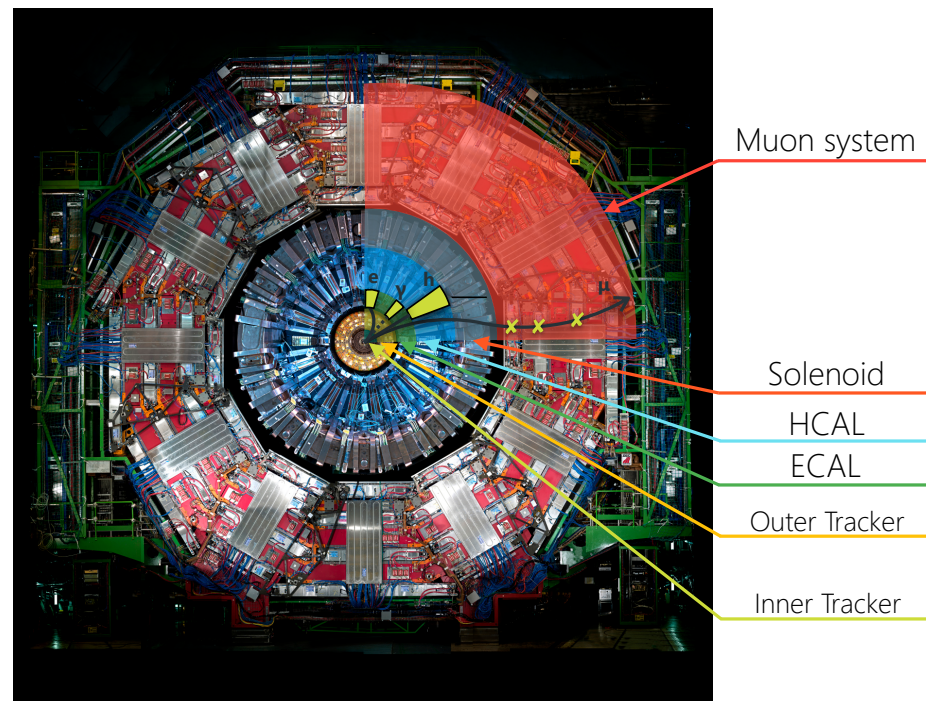


Fig. 2.2.: Cross section of the CMS detector with highlighted subdetectors. Muons are the only known particles (except for neutrinos that cannot be detected directly with CMS) that traverse the whole detector. They leave signals in the tracker, the calorimeters and the muon system. Hadrons are detected with information of the HCAL and leave signals in the tracker if they possess electric charge. The same holds for electrons that leave a record in the tracker and an energy deposit in the ECAL. Photons on the other hand do not interact with the tracker but deposit their energy in the ECAL. Adapted from Ref. [60].

¹From 2017 on, a new pixel detector is installed with a fourth layer.

2.2.1 Detector subsystems

2.2.1.1 Tracker

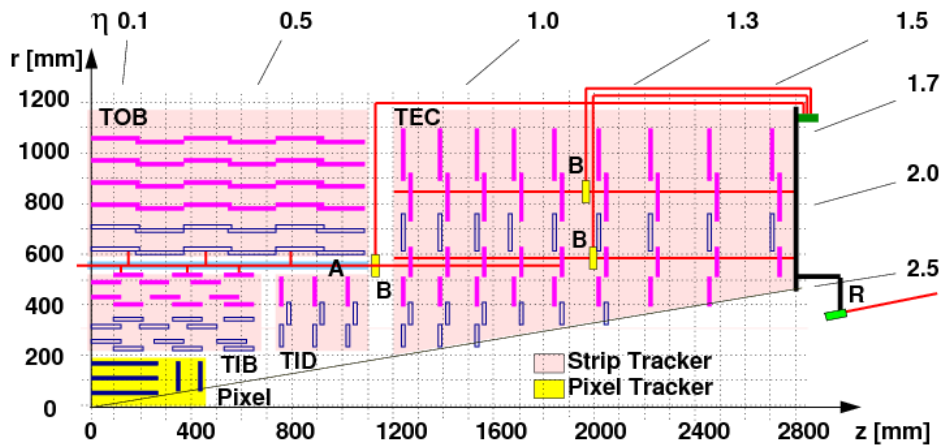


Fig. 2.3.: Layout of the cylindrical CMS tracker in the r - z -plane. The overall tracker covers pseudorapidities up to $|\eta| = 2.5$ and is made up of the pixel detector highlighted in yellow and the strip detector in rose. The pixel detector is constructed such that every particle within $|\eta| < 2.5$ traverses at least three layers of pixel modules. Tracker inner and outer barrel (TIB and TOB) provide ten layers of tracking modules accompanied by three layers of inner disk modules (TID) and up to nine layers of endcap disks (TEC). A laser alignment system (LAS - in red) is used to monitor the exact position of each module in the detector to ensure highest accuracy during the track reconstruction. Taken from Ref. [61].

The tracker is designed to measure the trajectories of charged particles. A silicon pixel detector with fine granularity is installed directly around the interaction point. The pixel detector is surrounded by the strip detector with a slightly broader granularity. The layout of the tracking system is depicted in figure 2.3. Charged particles ionize the semiconducting material when they traverse it. As a consequence electrons are elevated into the conduction band, if an electric field is applied to deplete the sensor, and induce a small electrical current. The high number of in total 66 million pixels and 9.6 million strips, that are each read out individually, leads to a channel occupancy which enables the tracking system to confidently assign tracks by combining the signals of the charged particles produced during every bunch crossing. Tracks are the key ingredient for the identification of charged particles and reconstruction of the event vertex.

2.2.1.2 Electromagnetic calorimeter

The ECAL measures the energy of electrons, positrons and photons by total absorption. It covers a region up to $|\eta| = 3.0$ and consists of 61200 homogenous lead-tungsten (PbWO_4) mono-crystals in the barrel and 7324 in the endcaps that produce scintillation light proportional to the energy of the stopped particle. The light is collected and amplified by Avalanche Photodiodes (APDs) in the barrel region and Vacuum phototriodes (VPTs) in the endcaps. A preshower device is installed in front of the endcap ECAL to detect π^0 decays.

2.2.1.3 Hadronic calorimeter

The HCAL is a sampling calorimeter surrounding the ECAL that measures the energy of the hadronic activity. It is constructed to maximize the material within the coil of the magnet in terms of interaction lengths λ . Brass plates used as absorber are interspersed by plastic scintillator tiles which act as a compact active medium. Wavelength-shifting fibres transport the scintillating light to photosensors, which are connected to the read-out electronics located outside the barrel. To ensure hermiticity additional sets of scintillator tiles are installed outside the magnet coil to catch the tail of very high-energetic hadronic showers that are not stopped by the HCAL and the magnet. The HCAL in alignment with the ECAL covers pseudorapidities up to $|\eta| = 3.0$. In addition, a hadron forward calorimeter made of steel and quartz fibres is installed in the region $3.0 < |\eta| < 5.0$ to measure the forward hadronic activity. Hadrons produce Cherenkov light in the quartz fibres that is detected by photomultipliers.

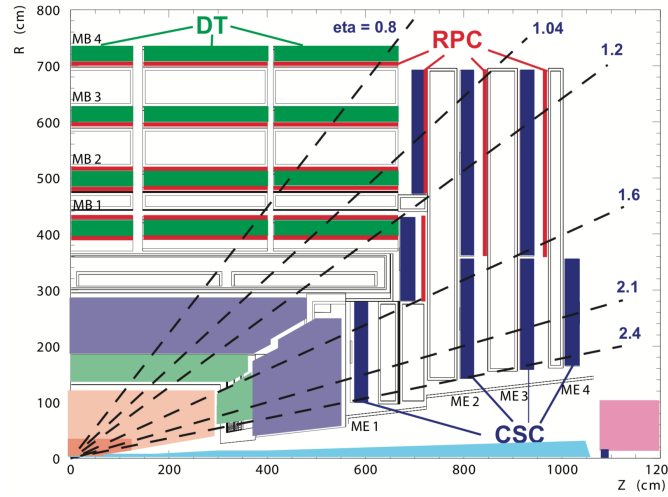


Fig. 2.4.: Schematic view of a quarter of the muon system in the r - η plane. Four layers of stacks of DTs (green) and RPCs (red) are installed around each wheel of the detector in the barrel region. RPC/CSC (red/blue) stacks are used to cover the region up to $|\eta| = 1.6$. Four layers of CSCs detect muons up to $|\eta| = 2.4$. Taken from [58].

2.2.1.4 Muon chambers

Muons are the only particles that pass through the whole inner detector and the solenoid. The muon system uses three different types of gaseous detectors to identify muons. As indicated in figure 2.4 the muon system is located outside the solenoid. The barrel covers pseudorapidities up to $|\eta| < 1.2$ and uses stacks of drift tubes (DT) and resistive plate chambers (RPC) to detect muons. The RPCs provide a very good time resolution as their response time is in the order of 1 ns and make it subsequently possible to assign the correct bunch crossing. In the endcaps the muon chambers operate in a higher magnetic field and must process higher muon rates. Therefore, six layers of cathode strip chambers (CSC) replace the DTs.

Overall, the muon system can measure muons with pseudorapidities up to $|\eta| < 2.4$. Muons leave hits in both the tracker and the muon chambers and yield an exceptionally high track and primary vertex (PV) reconstruction precision for high energetic muons. The muon system is part of the trigger system to select interesting physics events.

2.2.2 Trigger

Due to technical limitations only about 100–1000 events out of the 40 million collisions taking place in CMS every second can be written to tape and stored for high-level analyses. For this reason, the event rate must be reduced by a factor of 10^{5-6} using a sophisticated triggering system that is capable of extracting only the most relevant physics events within a couple of microseconds. CMS uses several stages of triggering. The electrical signals of all detector subsystems are stored in pipelined buffers while the *level-1* (L1) trigger, that is implemented in hardware, decides whether an event is kept or rejected. This decision is solely based on the primitive information of the calorimeters and the muon system. For this first step a computing time of $3.8\,\mu\text{s}$ is allocated per event. Events passing the L1 stage are feed into a processor farm where the high-level trigger (HLT) is applied that runs several menus of a lightweight reconstruction software. The HLT trigger is designed to reject events as early as possible. For this reason, the algorithms evaluate calorimeter and muon system data first and utilize the tracking system content and the full event information only if the event was not disregarded in the first step.

In 2016 the CMS Detector recorded proton-proton collision at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with an integrated luminosity of $L=35.9\,\text{fb}^{-1}$. Recorded data is grouped into so-called *Runs* with similar quality assessment and detector performance. 2016 data from the Runs *B-H* is used for this analysis.

2.3 Particle Flow

To perform a complex physics analysis, information provided by tracking, calorimeter and muon systems has to be combined to reconstruct the actual particles produced in the collisions. At CMS, advanced reconstruction algorithms are applied to piece together particle signatures and find vertices. An iterative tracking reconstruction algorithm described in Ref. [62] uses hits in the tracker to reconstruct trajectories and vertices of charged particles. The energy of electrons, hadrons, and photons is inferred from energy deposit clusters in the ECAL and HCAL. Tracks and clusters are interpreted by the particle-flow (PF) algorithm [63] that links the reconstructed elements according to their spatial and kinematic compatibility and tests which one of the five possible particle hypotheses describes the whole object. As a result, PF will give a list of muons, electrons, photons, neutral and charged hadrons with the associated tracks and energy deposits. These particle candidates can further be used to construct complex objects, like *jets*. The reconstruction and selection criteria as used in this analysis are outlined in the following subsections.

2.3.1 Leptons

2.3.1.1 Muons

Muons are required to be reconstructed by the PF algorithm. Thereby, muons are denoted *global* if a muon candidate based on muon chamber hits was matched to tracks in the inner tracker. *Tracker* muons are muons that were identified by assigning calorimeter and muon chamber entries to an already reconstructed inner track. In this analysis muons must pass the medium muon ID provided by the CMS collaboration [64] with the requirements that the muon

- is reconstructed as a *tracker* or *global* muon,
- has at least 49% of valid tracker hits in data for Run B-F or 80% valid hits for MC and data for Run G-H, and
- is associated with the PV of the event by $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm. Thereby, d_{xy} and d_z denote the transversal and longitudinal minimal distances between of the muon track and the PV.

Furthermore, the muon must fulfill various quality criteria based on the performance of the fit.

2.3.1.2 Electrons

Electrons are reconstructed from energy deposits in the ECAL and tracks in the inner tracker. They must pass the tight-working point of the multivariate-analysis (MVA) based electron ID corresponding to a 80% electron selection efficiency. To ensure that the electron is assigned to the PV, they must satisfy $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm.

2.3.1.3 Lepton isolation

Lepton identification is made robust by isolating the lepton from other PF candidates around the lepton candidate within a cone of size $\Delta R < 0.4$ ($\Delta R < 0.3$) for muons (electrons) according to the formula

$$I_{\ell}^{\text{rel}} = \frac{\sum p_T(h^{\pm}) + \max(\sum E_T(h^0) + \sum E_T(\gamma) - \Delta\beta \sum E_T(\text{PU}), 0)}{P_T^{\ell}}. \quad (2.1)$$

The variable compares the transverse momentum of the candidate lepton P_T^{ℓ} with the sum of the transverse momenta of charged hadrons $p_T(h^{\pm})$ and transverse energies of neutral hadrons and photons within the size of the cone. Thereby, $\Delta\beta = 0.5$ is the estimate of the ratio of neutral to charged particles originating from additional proton collisions in the same event called *pileup* (PU). Non-isolated leptons are likely to originate from heavy quark decays. In the analysis presented in the next chapter, leptons play an important role in identifying Higgs bosons. The isolation is a powerful observable to select leptons originating from the decay products of the Higgs boson.

2.3.2 Jets

Jets originate from partons, generated in EW or QCD processes, for example. They undergo hadronization and create a collimated stream of particles that is seen as a cone shaped cluster of tracks and energy deposits in the detector. As jets inherit the properties of the initial parton, it is crucial for the correct reconstruction of the jets to assign only those particles to the jet that were really produced during hadronization. For this reason, dedicated sequential jet-finding algorithms have been developed that apply distinct angular distance criteria to combine particles in the detector to jets. This analysis utilizes the *anti- k_T* algorithm [65] with a cone size of $R=0.4$ to reconstruct jets. The algorithm clusters individual particle flow candidates using a distance measure of the form

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{(\Delta y_{ij} + \Delta \phi_{ij})^2}{R^2}, \quad (2.2)$$

where $p = -1$ for the *anti- k_T* algorithm, Δy_{ij} and $\Delta \phi_{ij}$ denote the rapidity and azimuthal differences, and the k_t are the transverse momenta of the candidates. By this, particles with large transverse momenta are clustered first, making the reconstructed jets particularly resistant against soft radiation, that is not originating from the hadronization process. The jet properties are fully determined by the sum of all four-momentum vectors over all jet constituents.

To be considered in the analysis, jets are required to fulfill $|\eta| < 4.7$ and $p_T > 30$ GeV and they must pass the jet ID [66]. Moreover, they must be separated by at least

$$\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} > 0.5 \quad (2.3)$$

from each of two reconstructed tau leptons. Jets originating from b-quarks usually emerge from a secondary vertex because the b-mesons have a relatively large lifetime and travel a non-negligible distance in the detector before they decay. Such jets are tagged by the Combined Secondary Vertex (CSV) algorithm and are considered *b-tagged* if they pass the medium working point of the b-tag ID provided by the CMS collaboration and have at least $|\eta| < 2.4$ and $p_T > 20$ GeV.

2.3.3 Tau leptons

The hadron-plus-strips (HPS) algorithm [67, 68, 69] is employed to reconstruct and identify hadronically decaying tau leptons, denoted by τ_h , by their decay mode (see tab. 2.1). Jets that were reconstructed by the *anti- k_T* algorithm in the previous step are taken as inputs for the algorithm. From all electron and photon candidates with $p_T > 0.5$ GeV the one with the highest p_T seeds a π^0 candidate. Neutral pions decay almost immediately into a pair of photons. Some of the photons convert in the tracker creating electron and positron tracks. Electrons and positrons may in addition radiate bremsstrahlung when crossing tracker material. Thus, additional photon and electron candidates in a small $\phi - \eta$ window around the π^0 candidate are added. Such a cluster is called *strip*. τ_h candidates are constructed from the charged hadrons and the neutral π^0 s with $p_T > 2.5$ GeV. Due to conservation of electric charge in the hadronic tau decay, only decay modes with an odd number of

Tab. 2.1.: Tau lepton decay modes and branching fractions ($\mathcal{B}$) with intermediate resonances. Charged hadrons are denoted by $h^\pm$ and neutral hadrons by h^0 . The hadronic decay modes highlighted in light blue are reconstructed and used in this analysis. Adapted from Refs. [70] and [19].

Decay mode	Resonance	$\mathcal{B}$ / %	Selection
Leptonic decays		35.2	
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$		17.8	
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$		17.4	
Hadronic decays		64.8	
$\tau^- \rightarrow h^- \nu_\tau$		11.5	No strips
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770)$	25.9	One strip and $0.3 < m_\tau < 1.3 \sqrt{p_T/100 \text{ GeV}}$
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5	2 strips and $0.4 < m_\tau < 1.2 \sqrt{p_T/100 \text{ GeV}}$
$\tau^- \rightarrow h^- h^- h^+ \nu_\tau$	$a_1(1260)$	9.8	No strips and $0.8 < m_\tau < 1.34 \text{ GeV}$, $\Delta z < 0.4 \text{ cm}$
$\tau^- \rightarrow h^- h^- h^+ \pi^0 \nu_\tau$		4.8	
others		3.3	

charged hadrons occur. Decays with one (three) charged hadron(s) in the final state are often called *one-prong* (*three-prong*) decays. The HPS algorithm combines decays with one charged hadron with 0, 1 or 2 π^0 candidates and decay modes with three charged hadrons with no further π^0 candidate to τ_h candidates and calculates its four-momentum by summation of the four-momenta of the constituents. To select a unique τ_h candidate the criteria given in table 2.1 are applied. If a candidate can be assigned to more than one decay mode, the decay mode that maximizes the p_T of the τ_h candidate is selected. Since τ_h 's are reconstructed from jets, there is a high probability that a jets is misidentified as a tau lepton. Moreover, the charged hadrons may be misidentified as electrons and muons and vice versa. For this reason, MVA-based discriminators are utilized to reject jets and electrons taking tau-lifetime variables and isolation of the τ_h candidate into account. Muons are vetoed using a cut-based discriminant. The discriminator against other jets are often referred to as the *Tau ID*.

Leptonically decaying taus are treated as if they were prompt leptons. Although tau leptons possess a relatively large lifetime and should be travelling some distance in the detector, it is rather challenging to identify leptons originating from tau decays. The determination of the impact parameter with respect to the PV is not very precise. Moreover, the large missing transverse energy (see section 2.3.4) coming from the two neutrinos in the leptonic decay is more or less balanced in the case that the decaying Higgs boson is not strongly boosted in the laboratory frame, since the taus are produced more or less back-to-back in the lab frame, too.

2.3.4 Missing transverse energy

The missing transverse energy of the event is defined as the negative vectorial sum of the transverse momenta of all particles that are reconstructed by the PF algorithm in the event [71]

$$\vec{E}_T^{\text{miss}} = - \sum_i \vec{p}_{T,i}. \quad (2.4)$$

Search for CP violation in gluon-gluon fusion events

The performance of the measurement of the CP mixing angle in couplings of the Higgs boson to top quarks is presented in this chapter. First, kinematic observables that are sensitive to scalar and pseudoscalar couplings of the Higgs boson are introduced. Based on these observables categories are defined that yield an enhanced fraction of ggF events and lead to a better sensitivity of the discriminating variables. A new data-driven background estimation technique is presented to estimate the contribution of multijet and W+jets backgrounds. The measurement of the CP mixing angle α based on a pseudo-dataset is discussed in the last section of this chapter.

3.1 Measuring CP properties in gluon-gluon fusion events

3.1.1 Baseline event selection

This analysis is developed on top of the SM search for the decay of the Higgs boson into a pair of tau leptons [22]. Consequently, Higgs boson events are tagged exploiting the same baseline event selection. They are identified in four of the in total six combinatorically possible decay channels of the tau leptons. These are the $e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels. The ee and $\mu\mu$ channels are disregarded because they have both a small branching fraction and suffer from an overwhelming background of $Z \rightarrow ee/\mu\mu$ events. The notation used above refers to events with leptonically decaying tau leptons by the final state lepton that triggers the event. For instance, τ_e is denoted e and τ_μ is called μ . The lepton and tau candidates are triggered with the same HLT menus as used in Ref. [22]. The requirements of the trigger menus and the criteria on the isolation are summarized in table 3.1. Only events with exactly two leptons are considered to make the channels exclusive among each other. This means events with additional loosely identified and isolated leptons are rejected. Moreover, leptons must have opposite-sign electric charges as they possibly originate from the electrically neutral Higgs boson. Ditau pairs are constructed using the pair matching algorithm from Ref. [22]. Last but not least, requirements on the compatibility with the PV of the event are imposed. Therefore, leptons must satisfy $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm. In the semileptonic decay channels, $e\tau_h$ and $\mu\tau_h$, W+jets events become a major source of background. These are rejected by requiring the transverse mass to be

$$m_T = \sqrt{2p_T^\ell E_T^{\text{miss}} (1 - \cos(\Delta\phi))} < 50 \text{ GeV}, \quad (3.1)$$

with the transverse momentum of the lepton p_T^ℓ and the angle $\Delta\phi$ between the lepton and the missing transverse energy E_T^{miss} . The choice of $m_T < 50$ GeV was investigated in [22] and found to provide a close-to-optimal signal acceptance and background rejection. The distribution of m_T for the 0 -jet category (see section 3.1.4) in the semileptonic channels are shown in figure 3.1.

Tab. 3.1.: Minimum selection criteria for the lepton pairs. The requirements are slightly tighter than actually imposed in the HLT menus to reduce the impact originating from the turn-on curve of the trigger. In the $e\mu$ and $\mu\tau_h$ channels several triggers are used. The setups used for these different triggers are highlighted accordingly. Adapted from Ref. [22].

Channel	Selection criteria		
	p_T/GeV	η	Isolation
$e\mu$	$p_T^e > 13$	$ \eta^e < 2.5$	$I^e < 0.15$
	$p_T^\mu > 24$	$ \eta^\mu < 2.4$	
	$p_T^e > 24$	$ \eta^e < 2.5$	$I^\mu < 0.2$
	$p_T^\mu > 15$	$ \eta^\mu < 2.4$	
$e\tau_h$	$p_T^e > 26$	$ \eta^e < 2.1$	$I^e < 0.1$
	$p_T^{\tau_h} > 30$	$ \eta^{\tau_h} < 2.3$	MVA τ_h ID
$\mu\tau_h$	$p_T^\mu > 23$	$ \eta^\mu < 2.1$	$I^\mu < 0.15$
	$p_T^{\tau_h} > 30$	$ \eta^{\tau_h} < 2.3$	
	$20 < p_T^\mu < 23$	$ \eta^\mu < 2.1$	MVA τ_h ID
	$p_T^{\tau_h} > 15$	$ \eta^{\tau_h} < 2.3$	
$\tau_h\tau_h$	$p_T^{\tau_h} > 50 \text{ \& } 40$	$ \eta^{\tau_h} < 2.1$	MVA τ_h ID

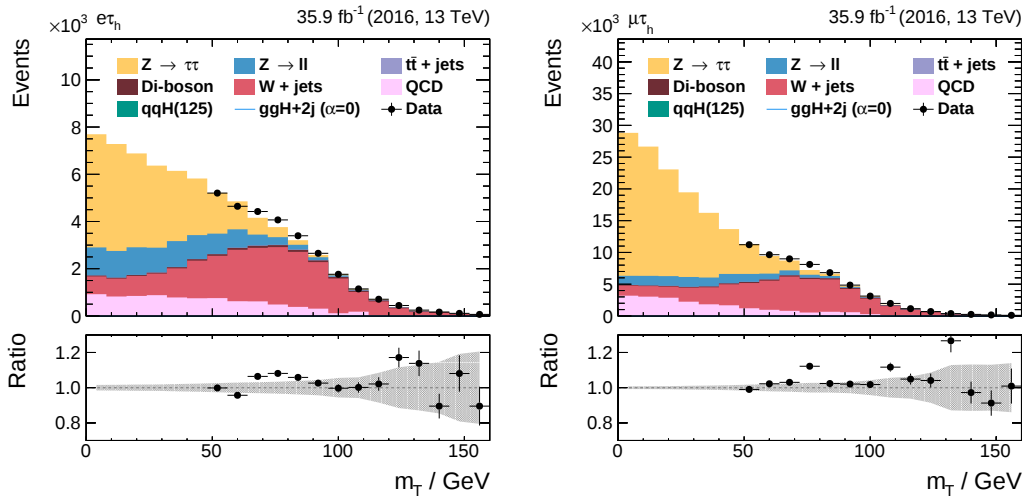


Fig. 3.1.: The transverse mass expected background distribution in the 0 -jet category estimated with the techniques introduced in section 3.2 and data blinded for $m_T < 50$ GeV are shown for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

3.1.2 Event kinematics and CP sensitive observables

The different tensor structures in the effective Lagrangian of the scalar and pseudoscalar couplings between the Higgs boson and the gluon fields lead to different kinematic distributions of the observable objects in the fermionic couplings in $gg \rightarrow H + 2j \rightarrow \tau^+ \tau^- + 2j$ events. The production via ggF and decay into tau leptons of a Higgs boson at rest is shown in figure 3.2. The outgoing particles produced in the event carry information about the quantum numbers and couplings because these quantities must be conserved at each interaction vertex. Generally, information about spin and parity is encoded in angular distributions. Consequently, the five angles $\Omega = \{\Delta\phi_{jj}, \theta_1, \theta_2, \theta^*, \phi_2\}$ in the sketch in figure 3.2 must contain all this information. The figure shows the kinematic situation of $gg \rightarrow H + 2j \rightarrow \tau^+ \tau^- + 2j$ events in the rest frame of the Higgs boson. The two initial-state jets span up two production planes together with the incoming partons, whose directions are aligned with the directions of the colliding protons. Furthermore, the intersection of the two production planes and the flight direction of the tau leptons span up an additional decay plane. As the tau leptons are produced back-to-back in the rest frame of the Higgs boson, they do not build a linearly independent set of two vectors necessary to span up a plane by their own.

It was found that the signed azimuthal difference between the two initial-state jets $\Delta\phi_{jj}$ defined by

$$\Delta\phi_{jj} = \phi_{j_1} - \phi_{j_2}, \quad \text{with } \eta_{j_1} > 0 \text{ and } \eta_{j_2} < 0 \quad (3.2)$$

already contains much of the total information and is consequently particularly sensitive to the dissimilarities of scalar and pseudoscalar couplings [72, 73]. $\Delta\phi_{jj}$ is a signed angle, that is uniquely defined by subtracting the azimuthal angle of the backward jet from the forward jet. As indicated in figure 3.2 the angle represents the azimuthal separation of the



Fig. 3.2.: Production (blue and orange) and decay (grey) planes in the rest frame of a Higgs boson produced via gluon-gluon fusion and decaying into a pair of tau leptons. The incoming partons and the initial-state produced gluons and quarks span up two production planes with azimuthal angular difference $\Delta\phi_{jj}$. A further decay plane is defined by the intersection of the two production planes and the direction of the two tau leptons.

two production planes spanned by the incoming and outgoing partons creating the two gluons that produce the Higgs boson. The shape of $\Delta\phi_{jj}$ for three different CP hypotheses is

shown in figure 3.3. Without any further selection criteria, all hypotheses have very similar shapes. Using an additional cut of $m_{jj} > 300$ GeV as motivated in Ref. [74] the shape differences become more pronounced. For scalar couplings the two initial-state jets are emitted parallel or anti-parallel, whereas jets originating from a pseudoscalar interaction are more likely to be orthogonal to each other in the azimuthal plane.

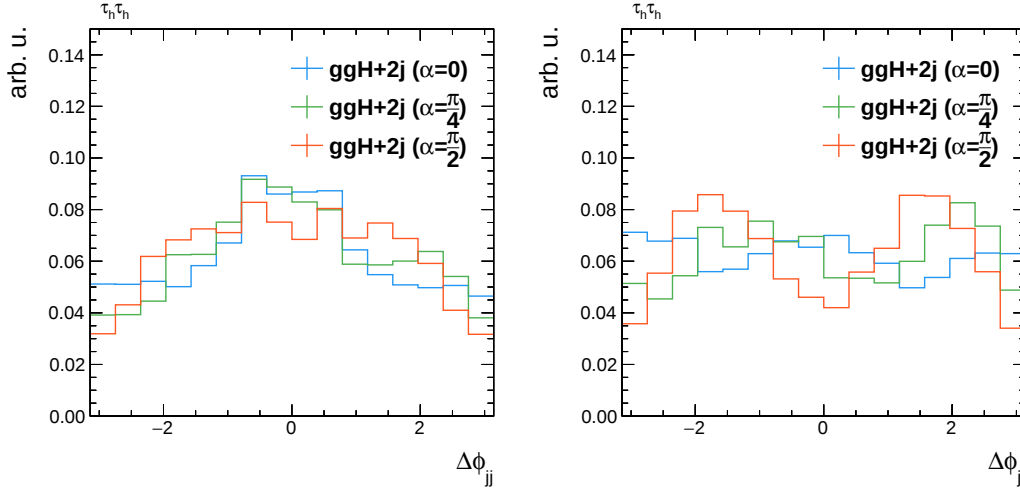


Fig. 3.3.: Comparison of the normalized $\Delta\phi_{jj}$ distributions of three CP hypotheses in the $\tau_h\tau_h$ channel. Simulated events provided by the three samples generated by the JHU Monte Carlo event generator were utilized for this study. The left figure shows the shapes without applying any further criteria on top of the event selection requirements from section 3.1.1. For the right plot events are required to have an $m_{jj} > 300$ GeV yielding a better separability.

3.1.2.1 Matrix Element Likelihood Analysis (MELA)

The matrix element $\mathcal{M}$ calculated by Monte Carlo generators is used to simulate the distributions of kinematic observables. It encaptures all information about a quantum physical process and vice versa. This fact is exploited in the Matrix Element Likelihood Approach (MELA) that aims to project the full Higgs boson production kinematics of multiple observables into a set of only a few kinematic discriminants [75, 76, 77, 78]. This approach was already applied during the search for the Higgs boson decaying into a pair of Z bosons in Run 1 [79]. So far only one of the five angles present in figure 3.2 was examined. By means of MELA, the full information of all five angles can be exploited projected into a 1D-variable. MELA is seeded with the Lorentz four-vector of the Higgs boson as reconstructed by the *SV-Fit* algorithm [80] and the jet four-vectors clustered by the *anti- k_t* algorithm and calculates the likelihood $\mathcal{P}$ of the kinematics to be compatible with the Matrix element of a given CP scenario on an event-by-event basis. These likelihoods are used further to build ratios that are optimal variables in terms of the Neyman-Pearson-Lemma [81] to distinguish between two different hypotheses. To be able to distinguish between scalar, pseudoscalar and CP

violating couplings of Higgs bosons produced via gluon-gluon fusion two observables are needed:

$$D_{0-} = \frac{\mathcal{P}_{0+}}{\mathcal{P}_{0+} + \mathcal{P}_{0-}} \quad (3.3)$$

$$D_{\text{CP}} = \frac{\mathcal{P}_{\text{CP violation}}}{\mathcal{P}_{0+} + \mathcal{P}_{0-}} \quad (3.4)$$

D_{0-} is used to discriminate between the scalar and pseudoscalar hypotheses and is constructed from the likelihoods $\mathcal{P}_{0+}$ and $\mathcal{P}_{0-}$ to be compatible with the scalar ($J^P = 0^+$) and pseudoscalar ($J^P = 0^-$) matrix elements and yields a score between 0 and 1. It is expected that D_{0-} peaks at 1 for a scalar Higgs boson, where pseudoscalar events should yield lower values on average. D_{CP} investigates a possible CP violating term using the likelihood $\mathcal{P}_{\text{CP violation}}$. It projects the event information into a score between -1 and $+1$ and is expected to possess an asymmetric shape if CP violation is present. The distributions of both discriminators for three CP scenarios are shown in figure 3.4.

Here again the shapes of the three different CP hypotheses without applying any selection cuts are hard to distinguish. For this reason, the $m_{jj} > 300$ GeV criterion is applied again leading to more pronounced shapes, too. Furthermore, it is observed that all distributions peak around 0.5 implying that the event kinematics for different CP scenarios are very similar. Therefore, the likelihoods to belong to a certain CP scenario are rather equal. Nevertheless, it can be seen that the peaks for scalar and pseudoscalar couplings lie at different values (blue and orange histograms in the upper row, right plot) and that the intermediate scenario causes an asymmetry in the D_{CP} distribution (green histogram right plot, bottom row).

3.1.3 Simulated samples

This analysis is based on the data set collected during Run 2 at the LHC in 2016 corresponding to an integrated luminosity of 35.9 fb^{-1} . Simulated samples for all backgrounds and SM Higgs bosons decaying into a pair of tau leptons were taken from the analysis for the search for $H \rightarrow \tau\tau$ events in Ref. [22]. The signal samples were generated with the POWHEG generator and contain simulated Higgs bosons created by all production processes. This means, there are only few events with two initial-state jets provided, which make these POWHEG samples not extraordinary suited for the usage in this analysis, because two jets are needed to calculate the CP sensitive observables. For this reason, three inclusive dijet samples for Higgs bosons produced via gluon-gluon fusion and decaying into tau leptons with CP mixing angles $\alpha = 0$, $\alpha = \frac{\pi}{4}$, and $\alpha = \frac{\pi}{2}$ were generated with the JHU generator [75, 76, 77, 78]. PYTHIA 8 [82] is used for modeling the decay of the Higgs boson into tau leptons, parton-showering, and hadronization. JHU uses only matrix elements that were calculated at leading order (LO). By comparing the kinematic distributions of the scalar samples with a sample generated by the POWHEG event generator at next-to-leading order (NLO), it was found that the emission of two jets is well-described [45]. The samples are listed in table 3.2 together with the SM cross sections of the processes. JHU assumes equal cross sections for the scalar and pseudoscalar scenarios. In the case of the intermediate scenario twice the cross section is yielded which can be inferred from equation (3.25) if

the normalization are equal to one as realized by the JHU generator. Command to produce the shapes for DCPStar

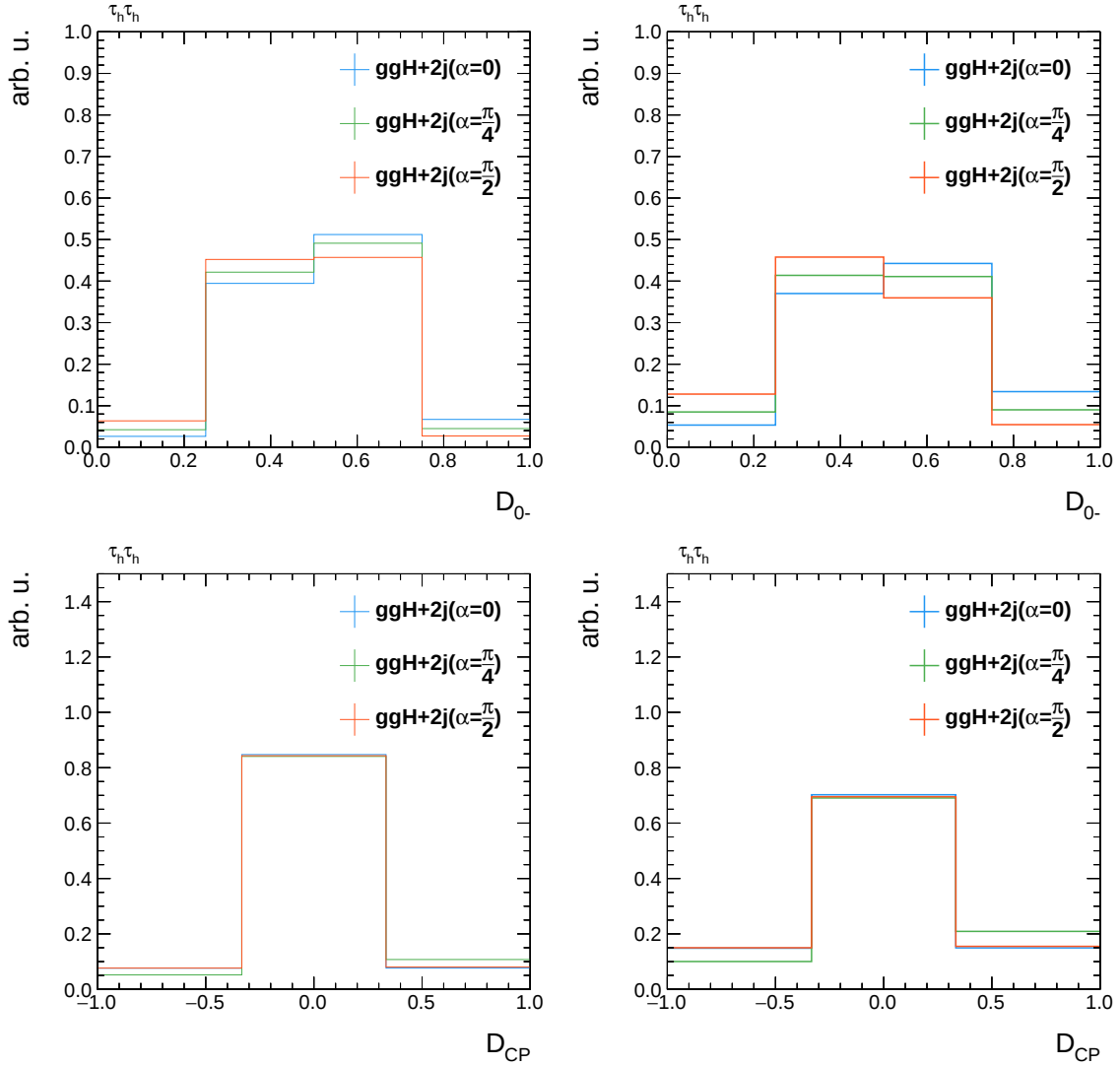


Fig. 3.4.: Comparison of the normalized D_{0-} (upper row) and D_{CP} (bottom row) distributions for three CP scenarios in the $\tau_h\tau_h$ channel. Simulated events generated by the JHU Monte Carlo event generator where utilized to predict the shapes. The left plots shows the distribution without any further criteria on top of those defined in section 3.1.1 are used. For the right plot events are required to satisfy $m_{jj} > 300$ GeV yielding a better separation between the different hypotheses. With the m_{jj} criterion the peak of the different CP scenarios in D_{0-} (right plot, upper row) lies in neighboring bins around 0.5. This means the different CP scenarios have very similar shapes. Three bins are used in D_{CP} to put emphasis on events that were well reconstructed and lie therefore in the tails of the distribution. The CP violating scenario shows here an asymmetry, whereas the pure states cannot be distinguished (right plot, bottom row).

Tab. 3.2.: Gluon-gluon fusion in association with two jets and VBF samples generated by the JHU generator used in this analysis. Events are generated with an increased cross section and later normalized to SM cross section [19].

Name	Events	σ_{SM}/pb
GluGluH2JetsToTauTau_M125_13TeV_CPmixing_sm_JHU	10^6	0.399
GluGluH2JetsToTauTau_M125_13TeV_CPmixing_maxmix_JHU	10^6	0.789
GluGluH2JetsToTauTau_M125_13TeV_CPmixing_pseudoscalar_JHU	10^6	0.386
VBFHiggsOPM_M-125_13TeV-JHUGenV6	10^6	2.671
VBFHiggsOMf05ph0_M-125_13TeV-JHUGenV6	10^6	0.474
VBFHiggsOMM_M-125_13TeV-JHUGenV6	10^6	0.237

3.1.4 Categorization

Previously, it has been found that a simple requirement on m_{jj} leads to better distinguishability between the three signal samples. This motivates that there are regions of phase space that provide a better separation between these samples than the overall separation in an inclusive event selection. Moreover, the selection of events is not pure in Higgs boson events. The baseline selection comes with a large contamination of background events which have signatures that mimic the signal in the detector. In this section, additional quality requirements are introduced to separate signal from background. Another goal here is to increase the fraction of selected Higgs bosons generated by ggF compared to VBF produced Higgs bosons in the signal categories to be sensitive to fermionic couplings. In this analysis events, that pass the baseline selection, are split into four categories to maximize the sensitivity to measure the CP properties in the fermionic couplings. For this purpose, the categories defined for [22] are adapted and optimized.

3.1.4.1 0-jet and boosted categories

Events with no jets fall into the *0-jet* category. All events that have either one jet or a dijet system without an invariant dijet mass greater than $m_{jj} = 300 \text{ GeV}$ fall into the *boosted* category. Since signal events are required to have at least two jets, these two categories

Tab. 3.3.: Selection cuts and observables of the *0-jet* and *boosted* categories used within this analysis.

Channel	0-jet	boosted
Selection cuts		
$e\mu, e\tau_h, \mu\tau_h$	No jet & No b-jet	Not in 0-jet & not in dijet
$\tau_h\tau_h$	No jet	
Observables		
All	$m_{\tau\tau}$	$m_{\tau\tau} \times p_{\text{T}}^{\tau\tau}$

serve as background containers, that help to control the contamination of backgrounds and nuisance parameters in the signal categories during the final fit. The requirements for these two categories are summarized in table 3.4. Compared to Ref. [22], a b-tag veto is added to the $e\mu$, $e\tau_h$ and $\mu\tau_h$ channels to further reduce the contamination from top-anti-top quark production, $t\bar{t}$. Moreover, events in the *0-jet* category are no longer discriminated in bins of the reconstructed tau decay mode but remain to be examined using the reconstructed invariant ditau mass $m_{\tau\tau}$. This has the advantage that the five associated decay-mode specific systematic uncertainties listed in table 3.5 become redundant.

$m_{\tau\tau}$ is the full invariant ditau mass as reconstructed by the *SVFit* algorithm. As shown in figure 3.5 the *SVFit* mass is capable of separating the Z and the Higgs boson resonances and is thus a powerful discriminant to reject Drell-Yan (DY) background events. For the *boosted* category a two-dimensional discriminator is chosen in bins of $m_{\tau\tau}$ and the p_T of the Higgs boson, that is constructed from the reconstructed four-momenta of the two taus and the missing transverse energy

$$\vec{p}_T^{\tau\tau} = \vec{p}_T(\tau_1) + \vec{p}_T(\tau_2) + \vec{E}_T^{\text{miss}}. \quad (3.5)$$

For the ggF signal in the *0-jet* and *boosted* categories POWHEG samples are utilized because

Tab. 3.4.: Selection cuts and observables of the *0-jet* and *boosted* categories used within this analysis.

Channel	0-jet	boosted
Selection cuts		
$e\mu, e\tau_h, \mu\tau_h$	No jet & No b-jet	Not in 0-jet & not in dijet
$\tau_h\tau_h$	No jet	
Observables		
All	$m_{\tau\tau}$	$m_{\tau\tau} \times p_{\text{T}}^{\tau\tau}$

the JHU samples are inclusive dijet samples and therefore not usable for these categories. An overlap with dijet events with $m_{jj} < 300$ GeV is prevented by using only events from the POWHEG samples in the *boosted* category. The modeling of the backgrounds and ggF signal is estimated with the methods described in section 3.2. They are shown in figure 3.5 for the $\tau_h\tau_h$ channel.

Tab. 3.5.: Decay-mode specific systematic uncertainties that are removed from this analysis because the decay mode of the tau leptons is no longer used as an observable in the *0-jet* category.

Name
CMS_tauDMReco_1prong_13TeV
CMS_tauDMReco_1prong1pizero_13TeV
CMS_tauDMReco_3prong_13TeV

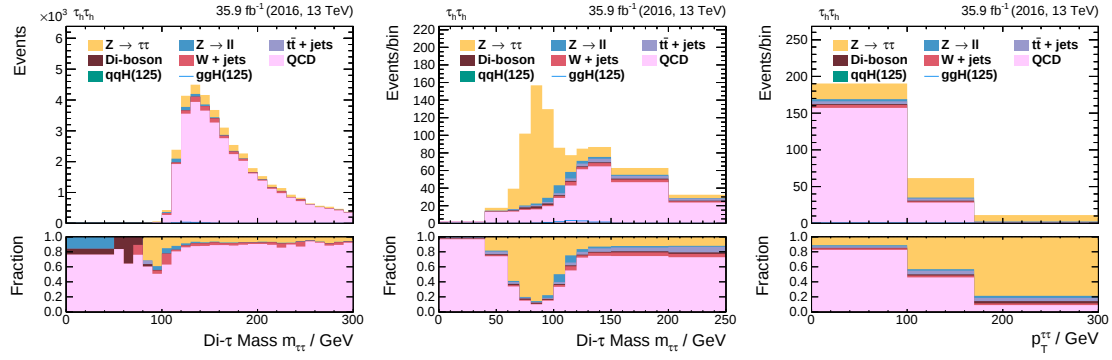


Fig. 3.5.: Estimated background and ggF signal distributions of observables in the *0-jet* category (left) and *boosted* category (middle, right) for the $\tau_h\tau_h$ channel. Background and signal are modeled from simulation. From the middle plot it can be inferred that the invariant ditau mass is capable of separating Higgs bosons from Drell-Yan background, as both the Z and the Higgs boson peaks are visible and separated. The subplot shows the relative fraction of each background in each bin. In the $\tau_h\tau_h$ channel is the dominant background are multijet events. More plots can be found in the appendix (see fig. A.1, page 80).

3.1.4.2 Dijet signal categories

Events containing two or more jets are further divided into a *dijet boosted* and a *dijet lowboost* category according to the transverse momentum of the Higgs boson $p_T^{\tau\tau}$. To be part of the *dijet boosted* category $p_T^{\tau\tau} > 200$ GeV must be satisfied. Otherwise events fall in the *lowboost* category. As shown for the $\tau_h\tau_h$ channel in figure 3.5, this criterion is particularly suited to separate signal from background. Furthermore, the two jets in the dijet categories are required to have $m_{jj} > 300$ GeV as discussed in section 3.1.2 and section 3.1.2. The distribution of m_{jj} for dijet events is shown in figure 3.6. In the final fit, three-dimensional discriminators in bins of $m_{\tau\tau}$, D_{0-} , and D_{CP} are constructed. The distributions of the individual observables and the background modeling are shown in figure 3.7. The categorization criteria are summarized in table 3.6. For the three observables the binnings

- $m_{\tau\tau} \in [0; 80; 100; 115; 130; 150; \infty]$ GeV,
- $D_{0-} \in [0.0; 0.25; 0.5; 0.75; 1.0]$, and
- $D_{CP} \in [-1.0; -0.4; 0.4; 1.0]$

were found to provide a reasonable compromise between sensitivity and bin population. In principle two bins in D_{CP} are sufficient to measure an asymmetry pointing to the presence of a CP violating term, but since it was observed that the event kinematics of different CP scenarios are rather similar, three bins were chosen. The idea is that events with a clear signature, and therefore a high likelihood, found in the tails of the distribution get a larger impact on the analysis sensitivity. The binning in $m_{\tau\tau}$ was chosen to find regions in the phase space with improved signal-to-background ratios close to the Higgs boson mass of 125 GeV.

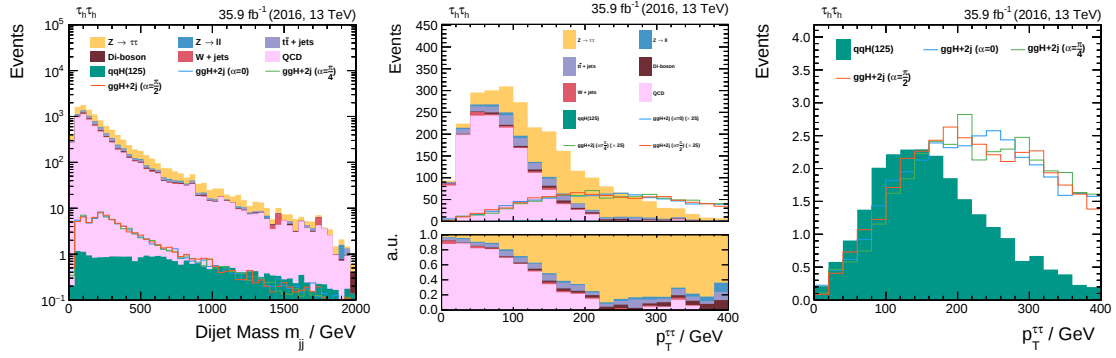


Fig. 3.6.: Expected signal and background modeling of the invariant dijet mass m_{jj} (left) and the ditau transverse momentum $\bar{p}_T^{\tau\tau}$ (middle and right) for the $\tau_h\tau_h$ channel are shown for events with two jets or more. The cross section of all three CP scenarios are normalized to the SM ggF cross section. The signal distributions in the middle plot of ditau transverse momentum is scaled by a factor of 25 for a better visibility. The right plots shows the same selection as applied in the middle plot focusing on the distributions of Higgs boson events. Above $\bar{p}_T^{\tau\tau} = 200$ GeV the fraction of ggF events compared to VBF events increases. More plots are shown figure A.2.

Tab. 3.6.: Selection criteria and observables used in the two dijet categories.

Channel	dijet lowboost	dijet boosted
Selection cuts		
$e\mu$	$>1 \text{ jet}, m_{jj} > 300 \text{ GeV}, p_T^{\tau\tau} < 200 \text{ GeV}, D_\zeta < -10 \text{ GeV}, b\text{-tag veto}$	$>1 \text{ jets}, m_{jj} > 300 \text{ GeV}, p_T^{\tau\tau} > 200 \text{ GeV}, D_\zeta < -10 \text{ GeV}, b\text{-tag veto}$
$e\tau_h$	$>1 \text{ jets}, m_{jj} > 300 \text{ GeV}$	$>1 \text{ jets \& } m_{jj} > 300 \text{ GeV}$
$\mu\tau_h$	$p_T^{\tau\tau} < 200 \text{ GeV}, b\text{-tag veto}$	$p_T^{\tau\tau} > 200 \text{ GeV}, b\text{-tag veto}$
$\tau_h\tau_h$	$>1 \text{ jets}, m_{jj} > 300 \text{ GeV}, p_T^{\tau\tau} < 200 \text{ GeV}$	$>1 \text{ jets}, m_{jj} > 300 \text{ GeV}, p_T^{\tau\tau} > 200 \text{ GeV}$
Observables		
All	$m_{\tau\tau} \times D_{CP} \times D_{0-}$	

For cross-checks the analysis is also performed using $\Delta\phi_{jj}$ as a discriminating variable. In that approach a two-dimensional discriminator in bins of $m_{\tau\tau}$ and $\Delta\phi_{jj}$ is constructed. The binning used in this case is

- $m_{\tau\tau} \in [0; 80; 100; 115; 130; 150; \infty] \text{ GeV}$ and
- $\Delta\phi_{jj} \in [-3.2; -2.7; -2.1; -1.6; -1.1; -0.5; 0.0; 0.5; 1.1; 1.6; 2.1; 2.7; 3.2]$.

As a qualitative test for the potential of the two CP sensitive observables to distinguish between different CP hypotheses the signal efficiency of a scalar Higgs boson coupling (0^{++}) is plotted versus the signal efficiency of a pseudoscalar Higgs boson coupling (0^{-+}). This so-called *ROC curve* is shown in figure 3.8 for the *dijet boosted* category in the $\tau_h\tau_h$ channel. The curve proves the statement made before that the approach with the optimal observables provides more potential than using $\Delta\phi_{jj}$ as a single observable. This means, there is indeed more spin information encapsulated in the remaining four angles that are fully taken into account by the matrix element likelihood approach.

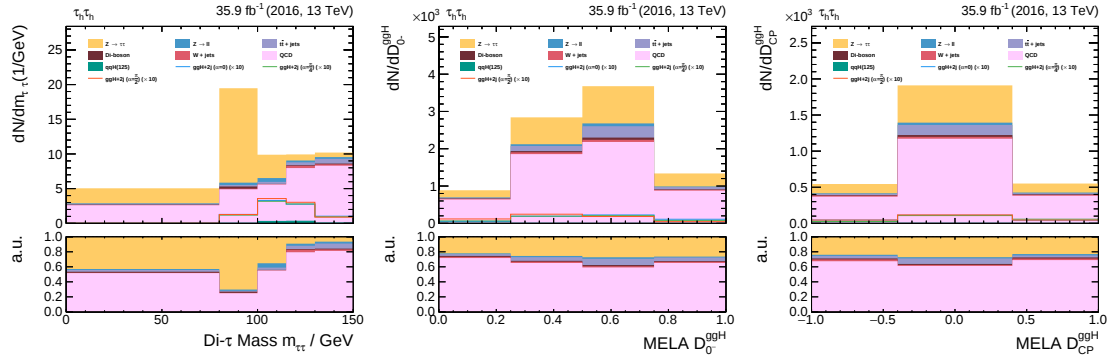


Fig. 3.7.: Background and ggH signal distributions for a scalar (blue), a CP violating (green) and a pseudoscalar (orange) CP scenario of the observables in the dijet lowboost category for the $\tau_h\tau_h$ channel. The cross section of the different CP scenarios are normalized to the SM ggF cross section. All histograms are normalized by the bin width and the ggF signal scaled by a factor 10 to improve the visibility. More plots can be found in figure A.3.

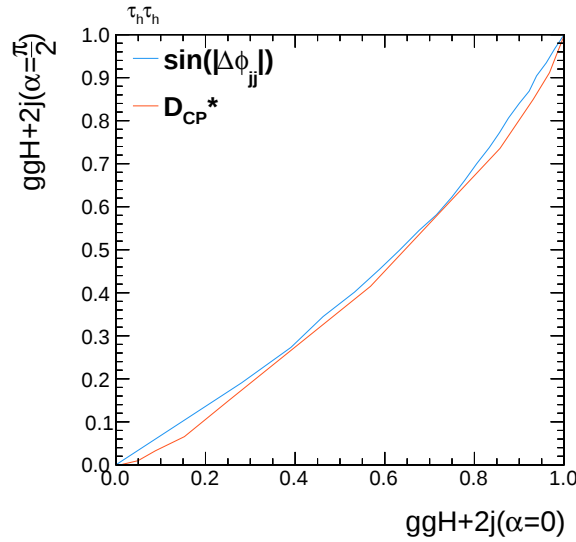


Fig. 3.8.: Comparison of ability to separate scalar and pseudoscalar fermionic couplings in the $\tau_h\tau_h$ channel in the *dijet boosted* category. It can be seen, that the MELA approach has more potential to distinguish between scalar and pseudoscalar hypotheses.

3.1.4.3 Purity and background contamination

The fraction of ggF and VBF events per channel and category is calculated from the histograms corresponding to the one-dimensional, two-dimensional and three-dimensional discriminants that are passed to the final fit. The contribution of VBF was split from the others backgrounds to cross-check whether indeed more ggF events than VBF events are selected. As shown in table 3.7 the fraction of ggF is largest in the *dijet boosted* category. Although VBF events tend to fall into the *dijet boosted* category, too, they only superceed ggF events in the *dijet lowboost* category. This means that the *lowboost* category is exceptionally capable of controlling the VBF background contamination in the final fit.

Albeit the fraction of signal events does not exceed 6% in any category, regions of phase space (especially bins of $m_{\tau\tau}$) can be found that show a larger signal-to-background ratio.

Tab. 3.7.: Fraction in percent of ggF signal events (always first number given per cell), VBF background events (second number) and other background events (last number). The fractions are calculated from the histograms entering the final fit (Prefit). The categorization successfully provides the largest fraction of ggF events in the *dijet boosted* category, where the fraction also superceeds the VBF contamination.

Fraction [%] of ggF, VBF, Bkg events per channel and category				
Channel/category	0-jet	boosted	dijet lowboost	dijet boosted
$e\mu$	0.29, 0.00, 99.71	0.51, 0.11, 99.38	0.85, 1.13, 98.01	4.74, 1.65, 93.61
$e\tau_h$	0.47, 0.01, 99.53	0.67, 0.14, 99.20	0.81, 1.06, 98.13	4.66, 1.76, 93.58
$\mu\tau_h$	0.44, 0.01, 99.56	0.67, 0.13, 99.20	0.82, 1.15, 98.03	4.27, 1.76, 93.97
$\tau_h\tau_h$	0.54, 0.01, 99.45	0.88, 0.20, 98.92	0.97, 1.22, 97.81	5.55, 2.11, 92.34

These regions are produced by the combined binning of the three (two) discriminating observables used in the matrix element ($\Delta\phi_{jj}$) approach.

3.2 Background estimation

3.2.1 $Z \rightarrow \tau\tau/\ell\ell$

As Z bosons may decay into a pair of tau leptons ($Z \rightarrow \tau\tau$), they are an important source of background in this analysis. The contribution is estimated using simulated events and is split into three groups according to the matching of the tau lepton candidate to the generator objects, that are the actual particles as they were produced by the Monte Carlo event generator.

- ZTT: τ_h 's are matched to the generated hadronically decaying tau lepton in the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels and electrons and muons to the generated leptonically decaying tau leptons in the $e\mu$ final state.
- ZL: τ_h 's are matched to a generated electron or muon and
- ZJ: τ_h 's otherwise.

The latter two contributions are denoted $Z \rightarrow \ell\ell$ within this analysis.

3.2.2 $t\bar{t}$

The contribution of events with a top-anti-top quark pair, that predominantly decay into a bottom quark and a W boson, is the largest background in the $e\mu$ channel. The shape of the distribution is fully taken from simulated events. A first estimate of the normalization is taken from simulation, but later dynamically adjusted in the final fit by including a $t\bar{t}$ control region. Again, the contribution is split according to the matching of the tau lepton candidate to the generated object.

- TTT: τ_h matches generated hadronically decaying tau leptons in the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels,
- TTJ: otherwise.

To reduce its contribution a b-tag veto is applied in all channels, except for $\tau_h\tau_h$ where the $t\bar{t}$ contamination is negligible. As additional requirement, events must satisfy $D_\zeta < -10$ GeV as defined in Refs. [22, 70] in the $e\mu$ channel. D_ζ describes the collinearity of the visible and invisible decay products of the two tau leptons. If the decay products originate from the decay of a resonance such as a Z boson, D_ζ yields values close to zero. As shown in figure 3.9 for the inclusive distribution in the $e\mu$ channel, this variable is a powerful discriminant to reject $t\bar{t}$ events.

3.2.2.1 $t\bar{t}$ control region

The initial normalization is derived taking into account the theory cross section, reconstruction and identification efficiencies, and the integrated luminosity. Later, a dedicated

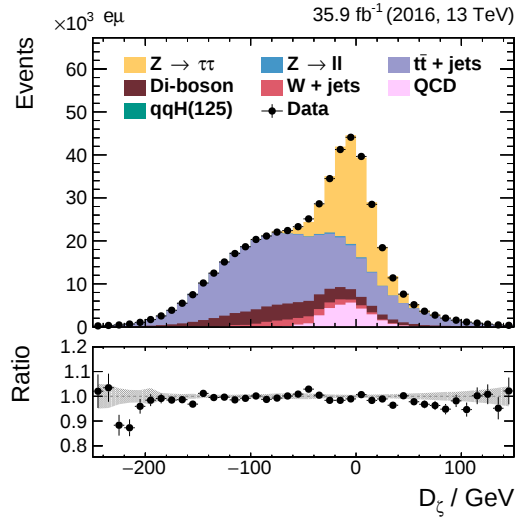


Fig. 3.9.: The inclusive background distribution and data of D_ζ neglecting the $m_T < 60$ GeV baseline criterion and the b-tag veto for the $e\mu$ channel is shown. Requiring $D_\zeta < -10$ GeV the $t\bar{t}$ contribution is significantly reduced.

$t\bar{t}$ control region is added to the final fit to adjust the $t\bar{t}$ normalization dynamically. The control region (CR) is constructed in the $e\mu$ channel by requiring

$$D_\zeta < -50 \text{ GeV and } m_{\text{vis}} > 90 \text{ GeV} \quad (3.6)$$

and keeping all other selection criteria as defined in section 3.1.4 fixed. The visible mass criterion is applied to reduce the small contamination of $Z \rightarrow \tau\tau$ events leading to a purity of 82% $t\bar{t}$ events in the control region. The event content of the control region is shown in figure 3.10.

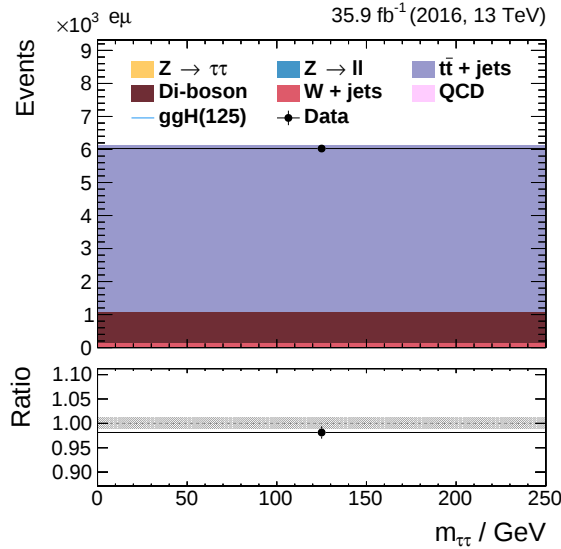


Fig. 3.10.: Expected number of background events and data in the $t\bar{t}$ control region as given to the final fit.

3.2.3 W+jets

$W+jets$ events are a major background of $H \rightarrow \tau\tau$ events. Especially in the semileptonic channels $W+jets$ becomes large if the W boson decays leptonically and one jet in the event is identified as a hadronically decaying tau lepton. The estimation of $W+jets$ and QCD multijet backgrounds distinguishes from the estimation of the other backgrounds as Monte Carlo generated events in that case do not provide a trustworthy prediction of the shape and yield of these backgrounds leading to large uncertainties. Therefore, a new data-driven estimation technique is applied that provides a simultaneous estimation of both backgrounds in the semileptonic channels. Throughout this thesis QCD multijet background is abbreviated by QCD .

3.2.3.1 $e\mu$ and $\tau_h\tau_h$ channels

In the $e\mu$ and $\tau_h\tau_h$ channels shape and yield are predicted from simulated $W+jets$ events.

3.2.3.2 $e\tau_h$ and $\mu\tau_h$ channels

For the $e\tau_h$ and $\mu\tau_h$ channels the $W+jets$ contribution is determined using a high transverse mass control region (high m_T CR) with $m_T > 70$ GeV. This CR is chosen because it profits from a high purity of $W+jets$ events. In order to estimate the $W+jets$ background in this CR, all other sources of backgrounds are subtracted from data. Except for QCD , which is also estimated using a data-driven technique, all background shapes and yields are taken from simulation here. As a consequence, the estimation of $W+jets$ and QCD backgrounds are entangled with each other and only the sum of both yields can be derived using a single CR. Splitting the high m_T CR according to the electric charge the final state leptons possess, the sum of yields of $W+jets$ and QCD in both sub-CRs can be calculated. Thereby, SS denotes events where the leptons have equal electric charges and OS denotes events with oppositely charged leptons. The yields in both subregions fulfill the following set of equations:

$$N_W^{SS,CR} + N_{QCD}^{SS,CR} = N_{data}^{SS,CR} - N_{other}^{SS,CR} \quad (3.7)$$

$$N_W^{OS,CR} + N_{QCD}^{OS,CR} = N_{data}^{OS,CR} - N_{other}^{OS,CR} \quad (3.8)$$

Solving them for $N_W^{SS,CR}$, the $W+jets$ yield in the SS high m_T region can be determined without a first estimate of the QCD background. First, the OS yields of QCD and $W+jets$ is expressed by the OS/SS extrapolation factors $R_{QCD}^{OS/SS}$ and $R_W^{OS/SS}$:

$$\begin{cases} N_{QCD}^{SS,CR} &= N_{data}^{SS,CR} - N_{other}^{SS,CR} - N_W^{SS,CR} \\ R_W^{OS/SS} \cdot N_W^{SS,CR} &= N_{data}^{OS,CR} - N_{other}^{OS,CR} - R_{QCD}^{OS/SS} \cdot N_{QCD}^{SS,CR} \end{cases}$$

$N_{QCD}^{SS,CR}$ can then be substituted by the upper equation

$$\Rightarrow R_W^{OS/SS} \cdot N_W^{SS,CR} = N_{data}^{OS,CR} - N_{other}^{OS,CR} - R_{QCD}^{OS/SS} \cdot (N_{data}^{SS,CR} - N_{other}^{SS,CR} - N_W^{SS,CR})$$

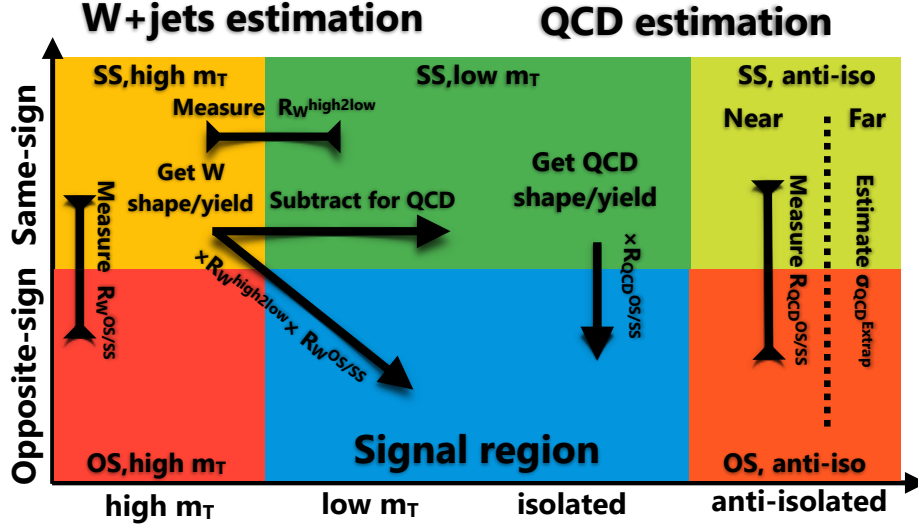


Fig. 3.11.: Procedure of the *simultaneous equation method* - a data-driven estimation of $W+jets$ and QCD backgrounds in the semileptonic ditau decay channels. The shape and yield of the $W+jets$ background is taken from a $W+jets$ enriched sideband. This sideband is realized by selecting events with a transverse mass $m_T > 70$ GeV (high m_T). QCD is estimated from a sideband where the charged hadron and the lepton from the tau lepton decays have the same electric charge. Usually an estimate of $W+jets$ is needed to derive the QCD estimate in the sideband or vice-versa. Here, the high m_T sideband is split again into same-sign and opposite-sign events (yellow and red boxes). By means of this, the $W+jets$ yield can be derived in a same-sign control region and is calculated without a first estimate of the multijet contribution using formula (3.9). For this, an opposite-sign to same-sign extrapolation factor $R_W^{OS/SS}$ must be measured (yellow and red boxes). Using further high-to-low (high2low) extrapolation factors (yellow and green boxes) the $W+jets$ yield can be estimated in every region. As a consequence, the $W+jets$ contribution, that must be subtracted from data in a same-sign control region to get the shape and preliminary normalization for QCD events, is taken from the data-driven estimate in the yellow box. Last but not least, a QCD opposite-sign to same-sign extrapolation factor is measured in an additional sideband with inverted isolation of the light lepton (light-green and orange boxes). With all these factors the contribution of both $W+jets$ and QCD can be estimated in the signal region (blue box).

After that $N_W^{SS,CR}$ is extracted

$$\Rightarrow R_W^{OS/SS} \cdot N_W^{SS,CR} - R_{QCD}^{OS/SS} N_W^{SS,CR} = N_{data}^{OS,CR} - N_{other}^{OS,CR} - R_{QCD}^{OS/SS} \cdot (N_{data}^{SS,CR} - N_{other}^{SS,CR})$$

giving an expression that is independent of the QCD yield in the high m_T SS region:

$$\Rightarrow N_W^{SS,CR} = \frac{1}{R_W^{OS/SS} - R_{QCD}^{OS/SS}} \left(N_{data}^{OS,CR} - N_{other}^{OS,CR} - R_{QCD}^{OS/SS} \cdot (N_{data}^{SS,CR} - N_{other}^{SS,CR}) \right). \quad (3.9)$$

This approach is called *simultaneous equation method* and its major advantage compared to the approach used in [22] is that no $W+jets$ Monte Carlo is needed to calculate the yield of QCD in the control region to later determine the shape and yield of $W+jets$ there. This implies that the yield of $W+jets$ is obtained naturally by this technique and no QCD Monte Carlo sample is needed at all. The technique is visualized in the sketch in figure 3.11.

The $W+jets$ yield in the signal region is finally extrapolated utilizing

$$N_W^{SR} = R_W^{SS,high2low} \cdot R_W^{OS/SS} \cdot N_W^{SS,CR}. \quad (3.10)$$

Summarizing the procedure above, three extrapolation factors need to be measured:

1. A $W+jets$ OS/SS extrapolation factor $R_W^{OS/SS}$ that is measured using simulated $W+jets$ events. For the *0-jet* and *boosted* categories, this factor is measured in events that satisfy the criterion $m_T > 70$ GeV. For the two dijet categories, an inclusive selection is utilized by omitting the requirement that events must have a transverse mass smaller than 50 GeV.
2. A QCD OS/SS extrapolation factor $R_{QCD}^{OS/SS}$ that is measured from data in a sideband with anti-isolated leptons (see section 3.2.4.2 and the light-green and orange sideband in figure 3.11).
3. A $W+jets$ high transverse mass ($m_T > 70$ GeV) to low transverse mass ($m_T < 50$ GeV) extrapolation factor $R_W^{OS/SS,high2low}$ that is measured using Monte Carlo simulated $W+jets$ events and is calculated in SS and OS events, respectively.

In the following subsections the results for the measurement of the different $W+jets$ extrapolation factors are discussed. The procedure of measuring the QCD OS/SS factor is described in section 3.2.4.2.

$W+jets$ extrapolation factors

Simulated $W+jets$ events are utilized to measure all three $W+jets$ extrapolation factors. Figure 3.12 (left) shows the measured OS/SS factors in each category for both channels. Only statistical uncertainties are shown here. The OS/SS ratios for both m_T control regions were found to not always agree within their uncertainties. Therefore, it was decided that $R_W^{OS/SS}$ cannot be measured in an inclusive m_T distribution for all categories and the high m_T control region is used to measure the factor only in the *0-jet* and the *boosted* categories. In the dijet categories the statistics become so low that the factors agree within their uncertainties and $R_W^{OS/SS}$ can be measured without distinguishing between different transverse mass regions. As a consequence, factors measured in the inclusive selection profit from a larger bin population. The trend clearly indicates, that the OS/SS factor depend on the number of jets in the events implying that the $W+jets$ extrapolation factors must be measured for each signal category individually. The high-to-low transverse mass extrapolation factors are measured individually for OS and SS events, because they should be as proximate as possible to the region where they are applied, and are shown in figure 3.12 in the right canvas. The exact scale factors with their associated statistical uncertainties of the measurements for both $e\tau_h$ and $\mu\tau_h$ channels for all signal categories are summarized in tables 3.8 and 3.9, respectively.

$W+jets$ systematic uncertainties

Systematic uncertainties covering the extrapolation from the high m_T CR to the signal region are calculated using the Ersatz method [83]. In addition to the high m_T to low m_T

Tab. 3.8.: Measured W +jets extrapolation factors and associated statistical uncertainties in the 0 -jet and *boosted* categories. $R_W^{OS/SS}$ is measured in the high m_T region for these two categories. The relative uncertainty of the measurement is given in the brackets.

Category	Channel	$R_W^{OS/SS, m_T > 70 \text{ GeV}}$	$R_W^{SS, high2low}$	$R_W^{OS, high2low}$
0-jet	$e\tau_h$	5.14 ± 0.08 (1.4%)	0.284 ± 0.003 (1.3%)	0.283 ± 0.008 (2.8%)
	$\mu\tau_h$	5.39 ± 0.06 (1.0%)	0.301 ± 0.003 (0.9%)	0.300 ± 0.006 (2.0%)
boosted	$e\tau_h$	3.84 ± 0.07 (1.8%)	0.282 ± 0.005 (1.8%)	0.269 ± 0.010 (3.6%)
	$\mu\tau_h$	3.39 ± 0.04 (1.2%)	0.295 ± 0.003 (1.3%)	0.313 ± 0.007 (2.2%)

Tab. 3.9.: Measured W +jets extrapolation factors and associated statistical uncertainties in the *dijet lowboost* and *dijet boosted* categories. The full region of m_T events, omitting the baseline $m_T < 50 \text{ GeV}$ criterion, is taken to measure $R_W^{OS/SS}$. Relative uncertainties of the measurement are stated in the brackets.

Category	Channel	$R_W^{OS/SS}$	$R_W^{OS, high2low}$	$R_W^{SS, high2low}$
2-jet lowboost	$e\tau_h$	2.70 ± 0.20 (7.6%)	0.291 ± 0.024 (8.2%)	0.185 ± 0.024 (13.2%)
	$\mu\tau_h$	2.70 ± 0.15 (5.4%)	0.300 ± 0.017 (5.9%)	0.209 ± 0.018 (8.8%)
2-jet boosted	$e\tau_h$	2.96 ± 0.73 (25%)	0.376 ± 0.089 (23%)	0.249 ± 0.098 (39%)
	$\mu\tau_h$	2.59 ± 0.50 (19%)	0.463 ± 0.084 (18%)	0.250 ± 0.068 (28%)

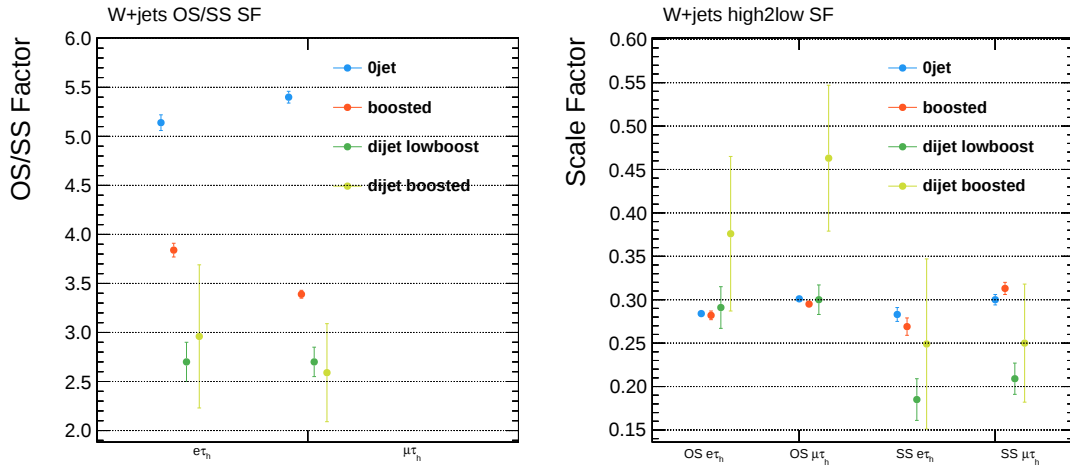


Fig. 3.12.: W +jets OS/SS extrapolation factors (left) and high m_T to low m_T scale factors (right) measured from simulated events. For the OS/SS factors a clear dependence on the number of jets is noted as the factors strongly depend on the number of jets of the different categories.

uncertainty, there is a low p_T to high p_T extrapolation uncertainty in the *dijet boosted* category that takes the higher Higgs boson transverse momentum regime in this region into account.

As the W +jets OS/SS factor is estimated from simulated events a potential mismodeling of these events is accounted for by systematic uncertainties on this factor. These uncertainties are estimated in a CR with inverted isolation of the hadronically decaying tau lepton. Other background contributions are subtracted from data in this sideband separately

Tab. 3.10.: Systematic uncertainties arising from the $W+jets$ background estimation method. Extrapolation uncertainties are derived using the *Ersatz method* from [83] in the *0-jet* and *boosted* categories. Uncertainties for the *dijet lowboost* and *dijet boosted* categories are measured in an inclusive selection for both categories simultaneously. An additional low p_T to high p_T extrapolation uncertainty is used in the *dijet boosted* category. $W+jets$ OS/SS uncertainties were measured by comparing simulated events with data in a sideband with a relaxed selection of hadronically decaying tau leptons. As these uncertainties affect only the rate of the $W+jets$ process, they are assigned as log-normal (lnN) uncertainties.

Name	Channel	Type	Value
Ersatz method			
WHighMTtoLowMT_0jet_CHANNEL_13TeV	$e\tau_h, \mu\tau_h$	lnN	3.3%
WHighMTtoLowMT_boosted_CHANNEL_13TeV	$e\tau_h, \mu\tau_h$	lnN	6.7%
WHighMTtoLowMT_dijet_CHANNEL_13TeV	$e\tau_h, \mu\tau_h$	lnN	18.2%
WlowPTtoHighPT_dijet_boosted_CHANNEL_13TeV	$e\tau_h, \mu\tau_h$	lnN	27.9%
Data-MC comparison			
WOSSS_syst_0jet_et_13TeV	$e\tau_h$	lnN	0.2%
WOSSS_syst_boosted_et_13TeV	$e\tau_h$	lnN	2.9%
WOSSS_syst_dijet_et_13TeV	$e\tau_h$	lnN	13.1%
WOSSS_syst_0jet_mt_13TeV	$\mu\tau_h$	lnN	1.2%
WOSSS_syst_boosted_mt_13TeV	$\mu\tau_h$	lnN	4.9%
WOSSS_syst_dijet_mt_13TeV	$\mu\tau_h$	lnN	8.6%
WOSSS_stat_0jet_et_13TeV	$e\tau_h$	lnN	3.5%
WOSSS_stat_boosted_et_13TeV	$e\tau_h$	lnN	2.6%
WOSSS_stat_dijet_et_13TeV	$e\tau_h$	lnN	8.2%
WOSSS_stat_0jet_mt_13TeV	$\mu\tau_h$	lnN	2.6%
WOSSS_stat_boosted_mt_13TeV	$\mu\tau_h$	lnN	2.0%
WOSSS_stat_dijet_mt_13TeV	$\mu\tau_h$	lnN	6.6%

for OS and SS events. The OS/SS factor is measured from this data-driven $W+jets$ estimate and then compared to the MC estimate from before. The difference $|R_W^{OS/SS, \text{data}} - R_W^{OS/SS, \text{MC}}|$ is used as systematic uncertainty and is measured in the *0-jet*, *boosted* and an inclusive *dijet* category. These uncertainties were measured in [84] and are summarized in table 3.10. For the dijet categories a single uncertainty was measured.

Control plots

The results of the background estimation for the transverse mass variable are shown for all categories in figures 3.13-3.16. The distribution of data are blinded in the signal region up to $m_T < 50$ GeV and only statistical uncertainties are taken into account. After the background estimation, an overall good agreement between data and simulation in the high m_T region is observed.

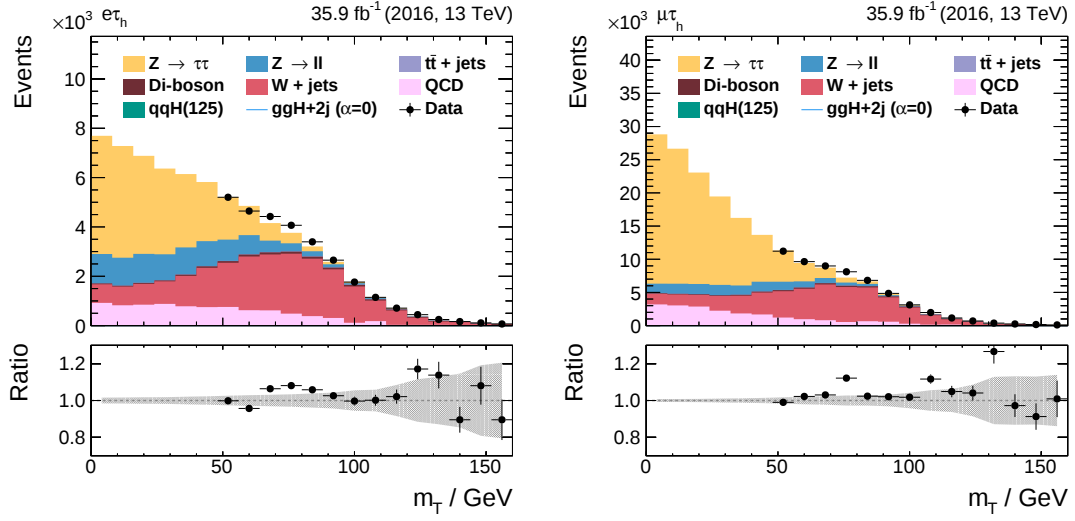


Fig. 3.13.: Blinded m_T distribution in the 0-jet category for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channel. Backgrounds were estimated with the techniques introduced in section 3.2.

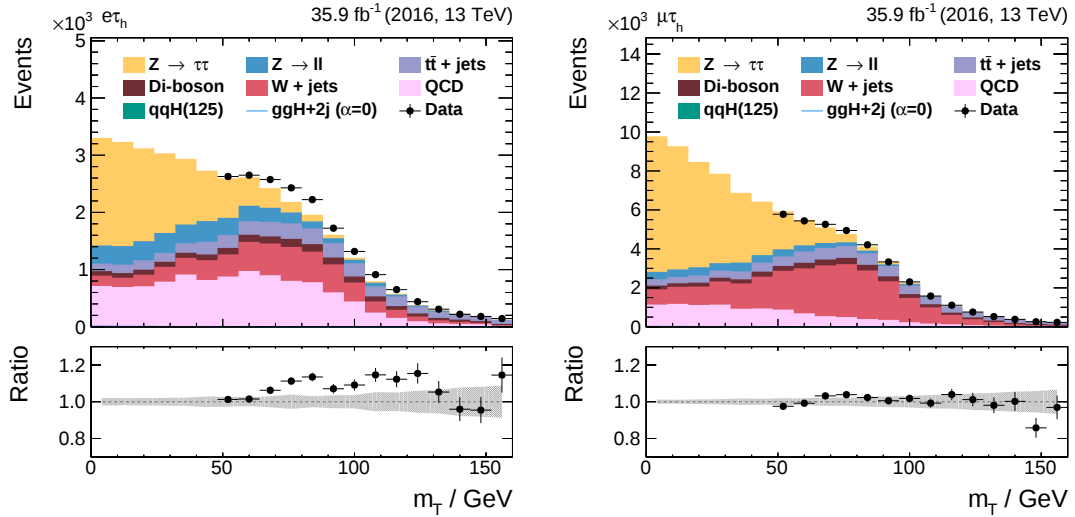


Fig. 3.14.: Blinded m_T distribution in the boosted category for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channel. Backgrounds were estimated with the techniques introduced in section 3.2.

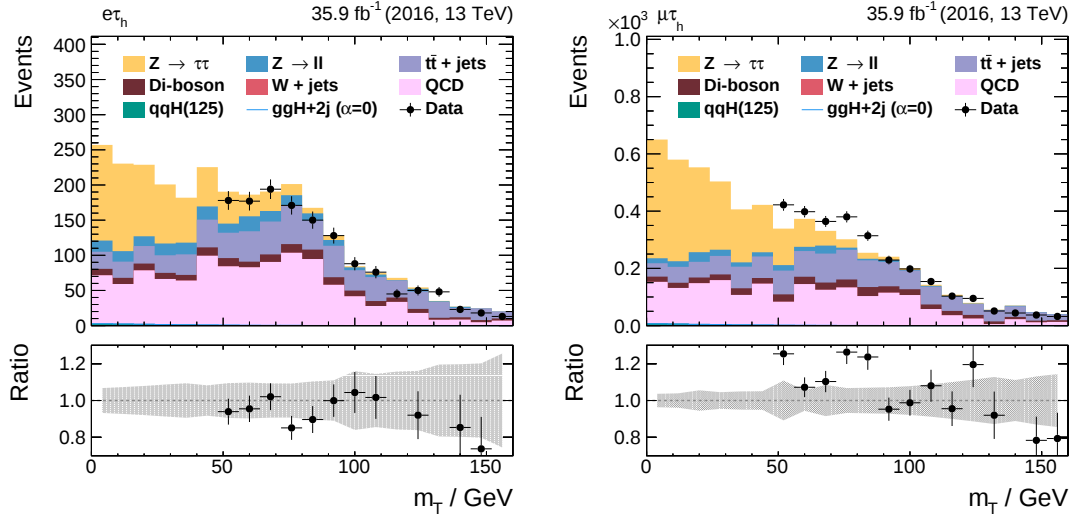


Fig. 3.15.: Blinded m_T distribution in the *dijet lowboost* category for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channel. Backgrounds were estimated with the techniques introduced in section 3.2.

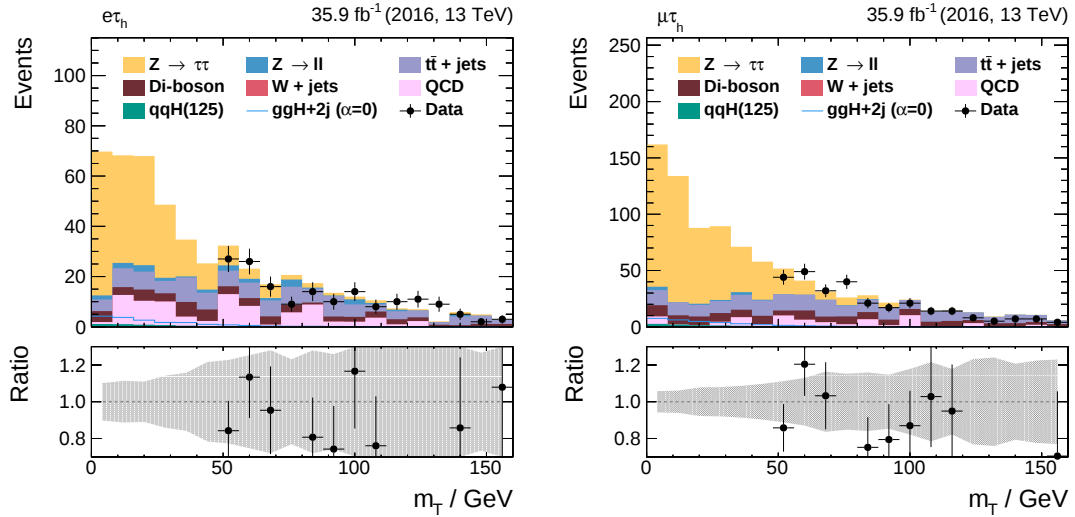


Fig. 3.16.: Blinded m_T distribution in the *dijet boosted* category for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channel. Backgrounds were estimated with the techniques introduced in section 3.2.

W+jets control regions

One OS high m_T and one SS high m_T control region are added to the final fit in the *0-jet* and *boosted* categories, respectively. As no *W+jets* events remain in the SS high m_T CR of the two dijet category, no dijet control regions are added to the final fit. This means in total eight *W+jets* control regions are used with this setup. The expected background contents and event numbers in these control regions are shown in figures 3.17 and 3.18 for the high m_T and SS high m_T region, respectively.

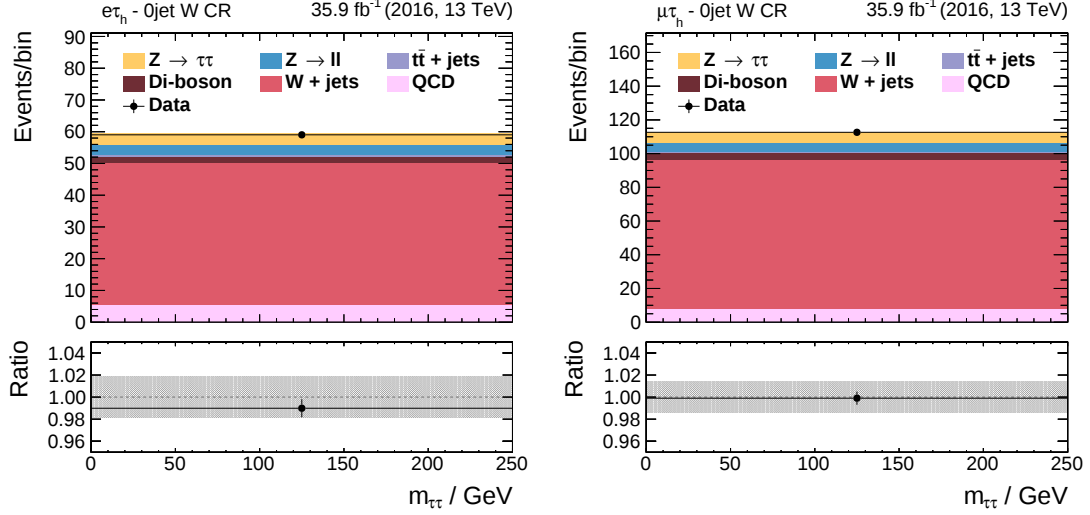


Fig. 3.17.: Control region expected background normalization and data as given into the final fit for the *W+jets* OS high m_T control region in the *0-jet* category for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

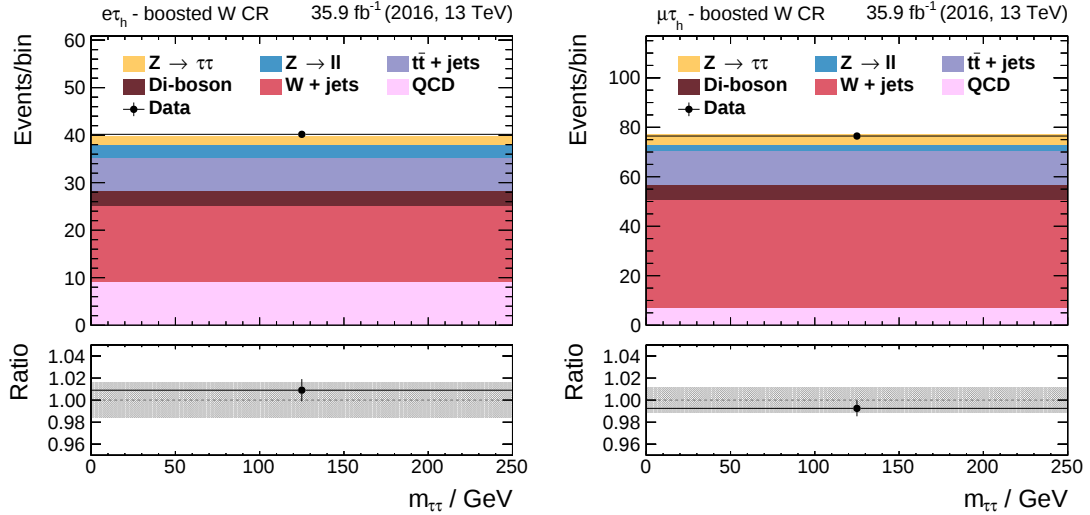


Fig. 3.18.: Control region expected background normalization and data as given into the final fit for the *W+jets* OS high m_T control region in the *boosted* category for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channel.

3.2.4 QCD multijet background

The contribution of QCD changes with for different decay channels as jets are likely to mimick hadronically decaying taus. As hadronization is a complicated process, data-driven estimation techniques are utilized in all channels.

3.2.4.1 $e\mu$ channel

In the $e\mu$ channel the QCD background is estimated in a sideband where the electric charges of electron and muon have the same sign. As the QCD background is small in this channel, a customized background estimation has not been performed. This means, the global OS/SS extrapolation factor measured for [22] to be $R_{\text{QCD}}^{\text{OS/SS}} = 2.22$ is used in all categories. The SS QCD shape and yield is calculated by subtracting all other sources of backgrounds from SS data. The extrapolation factor is applied as a scale factor to get the normalization in the signal region.

3.2.4.2 $e\tau_h$ and $\mu\tau_h$ channels

The QCD background is estimated from events where the τ_h and the light lepton (e, μ) have SS electric charges. The sideband is chosen such that it keeps all other selection criteria, that are applied in the signal region, fixed. To estimate the shape and the yield in the SS region, the small amounts of other backgrounds are subtracted from data. Except for $W+jets$, that is modeled by the data-driven estimate obtained in section 3.2.3.2 with

$$N_{\text{W}}^{\text{SS,SR}} = R_{\text{W}}^{\text{SS, high2low}} \cdot N_{\text{W}}^{\text{SS,CR}}, \quad (3.11)$$

all other background contributions are taken from simulation. In order to obtain the final QCD contribution

$$N_{\text{QCD}}^{\text{OS,SR}} = R_{\text{QCD}}^{\text{OS/SS}} \cdot N_{\text{QCD}}^{\text{SS,SR}} \quad (3.12)$$

is used. Here, $R_{\text{QCD}}^{\text{OS/SS}}$ is the QCD OS/SS extrapolation factor.

QCD SS/OS extrapolation factor

Studies of QCD Monte Carlo samples indicate that different fractions of SS and OS events in QCD backgrounds may be expected. For this reason, the ratio of SS to OS events must be measured to get a reliable QCD multijet background estimate in the signal region. $R_{\text{QCD}}^{\text{OS/SS}}$ is determined in two sidebands orthogonal to the signal region with inverted isolation of the light lepton. The factor measured in the *near* sideband ($0.1 < I_{\text{rel}}^e < 0.17$ in $e\tau_h$ and $0.15 < I_{\text{rel}}^\mu < 0.25$ in $\mu\tau_h$) is applied to SS events to estimate the final yield. This *near* sideband ensures proximity to the signal region, but possesses a rather low purity in QCD events of roughly 30-50%. The measurement in the *far* sideband ($0.17 < I_{\text{rel}}^e < 0.5$ in $e\tau_h$ and $0.25 < I_{\text{rel}}^\mu < 0.5$ in $\mu\tau_h$) is used to cross-check the measurement in the *near* sideband and to estimate the systematic uncertainty due to the assumption that the measurement of the ratio in the *near* sideband can be extrapolated to the signal region. Moreover, this region of phase space provides a significantly larger fraction of QCD events in the order of

80-90%. Since the fraction may strongly depend on the selection, $R_{\text{QCD}}^{\text{OS/SS}}$ is measured for each signal category individually. In order to increase the event population in the two dijet categories, an inclusive selection of dijet events, omitting the $p_T^{\tau\tau}$ and the m_{jj} criteria, is chosen and the working point of the tau-ID is relaxed to *very loose*. An exception of this procedure is made for the *dijet lowboost* category in the $\mu\tau_h$ channel because the inclusive factor does not provide a reasonable data to MC agreement. There the same procedure as for the *0-jet* and *boosted* categories is applied. A summary of the exact selection criteria is given in table 3.11.

Tab. 3.11.: Selection criteria applied in the anti-isolated sidebands used to measure $R_{\text{QCD}}^{\text{OS/SS}}$. The factor is measured in the *0-jet*, *boosted* and an inclusive *dijet* category for the two dijet signal categories. An exception is made for the *dijet lowboost* signal category in the $\mu\tau_h$ channel, in which the inclusive scale factor did not provide a good agreement between data and MC.

Name	Category	Near	Far
0-jet	No jet, b-tag veto		
boosted	(!0-jet, !dijet) (One jet, b-tag veto)	$0.1 < I_{\text{rel}}^e < 0.17$ in $e\tau_h$	$0.17 < I_{\text{rel}}^e < 0.5$ in $e\tau_h$
dijet	≥ 1 jet, b-tag veto, τ_h WP VLoose Isolation		
$\mu\tau_h$ lowboost	≥ 1 jet, $p_T^{\tau\tau} < 200$ GeV, b-tag veto, $m_{jj} > 300$ GeV	$0.15 < I_{\text{rel}}^\mu < 0.25$ in $\mu\tau_h$	$0.25 < I_{\text{rel}}^\mu < 0.5$ in $\mu\tau_h$

The extrapolation factor is determined performing a binned maximum-likelihood fit to the visible mass distribution in OS events treating QCD as signal process. As a first estimate for QCD in the respective OS sideband, $N_{\text{QCD}}^{\text{OS,prefit}}$, the QCD contribution from SS events is taken which is calculated subtracting all sources of simulated backgrounds (in this sideband also *W+jets*) from SS data

$$N_{\text{QCD}}^{\text{OS,prefit}} = N_{\text{QCD}}^{\text{SS}} = N_{\text{data}}^{\text{SS}} - \sum_{i \in \{ZTT, ZLL, TT, VV, W, qqH\}} N_i^{\text{SS}}. \quad (3.13)$$

This is equivalent to choosing $R_{\text{QCD}}^{\text{OS/SS}} = 1$. During the fit all backgrounds are allowed to float within their uncertainties. Therefore, normalization uncertainties for all backgrounds that are subtracted from data; an uncertainty on the luminosity of 2.5%; trigger efficiency uncertainties for tau leptons (3%), muons (2%) and electrons (2%); the $e \rightarrow \tau$ misidentification rate uncertainty of 12%; the $\mu \rightarrow \tau$ misidentification rate uncertainty of 25% and the decay-mode specific tau energy scale shape uncertainty are taken into account.

The QCD signal strength modifier r_{QCD} after performing the fit in the *near* sideband is subsequently taken as the QCD OS/SS factor. This factor is read off from the profile likelihood function for r_{QCD} as exemplary shown for the *boosted* category in the $\mu\tau_h$ channel for both *near* and *far* sidebands in figure 3.19. More fit result are found in the appendix in section A.2.1.

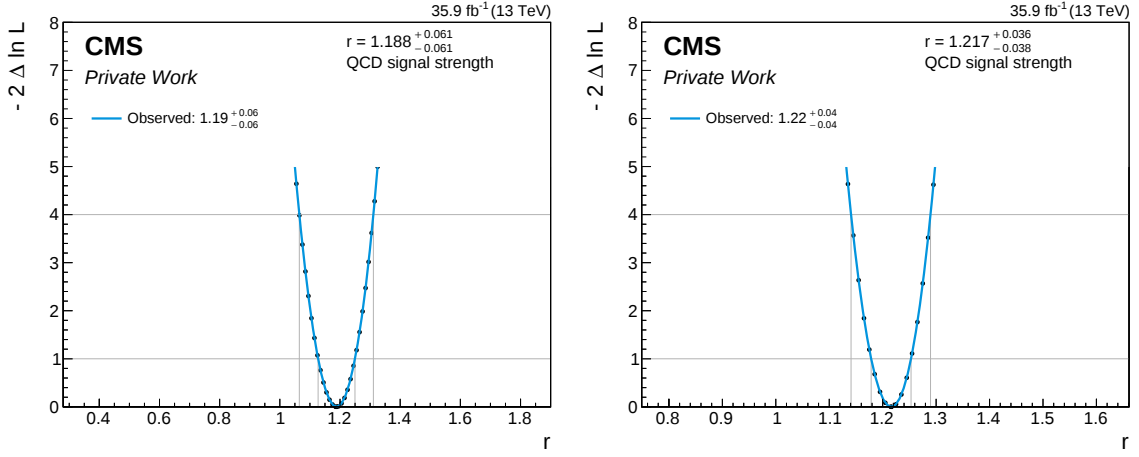


Fig. 3.19.: Results of the maximum likelihood scan to find the signal strength of the QCD multijet background in the $\mu\tau_h$ channel in the *boosted* category. The signal strength parameter from the near sideband (left figure) $r_{\text{QCD}}^{\text{near}} = 1.19 \pm 0.6$ is taken as the QCD OS/SS factor. As statistical uncertainty the larger value of the two given thresholds is chosen. $r_{\text{QCD}}^{\text{near}}$ is compared to the value $r_{\text{QCD}}^{\text{far}} = 1.22 \pm 0.04$ obtained in the *far* sideband (right plot). As they agree within their uncertainties the statistical uncertainties of the far sideband is assigned as systematic extrapolation uncertainty.

The uncertainty given by the fit is added as a systematic uncertainty. Furthermore, an uncertainty $\sigma_{\text{QCD-Extrap}}$, covering the extrapolation from the *near* sideband to the signal region, is estimated comparing the measured factors in the *near* and the *far* sideband and added to the fit uncertainty of the *near* measurement. If both factors agree within their uncertainties, the extrapolation uncertainty is covered by the fit uncertainty in the *far* sideband

$$\sigma_{\text{QCD-Extrap}} = \sigma_{\text{far}}. \quad (3.14)$$

If they do not agree, the difference between the two measured factors

$$\sigma_{\text{QCD-Extrap}} = \left| R_{\text{QCD, near}}^{\text{OS/SS}} - R_{\text{QCD, far}}^{\text{OS/SS}} \right| \quad (3.15)$$

is taken. The quadratic sum $\sigma_{\text{QCD}}^{\text{tot}}$ of the near fit uncertainty and the extrapolation uncertainty is given as a single systematic uncertainty to the final fit. The results of the QCD OS/SS factor measurements are stated in table 3.12. Examples for the outcome of the fits in the *near* and *far* sidebands are given in figure 3.19. The expected background modeling and data before (Prefit) and after the likelihood fit (Postfit) for each category in the *near* sideband is shown in figures 3.21 and 3.22. Additionally, the distributions in the $\mu\tau_h$ channel in the *dijet lowboost* category can be found in figure 3.20.

Tab. 3.12.: Measured QCD OS/SS extrapolation factors and systematic uncertainties in the *0-jet*, *boosted* and *dijet* categories. Furthermore, an individual factor is measured for the *dijet lowboost* category in the $\mu\tau_h$ channel, because the inclusively measured factor does not provide a good data to MC agreement. The best-fit signal strength modifier value and the uncertainty read off from the profile likelihood function are given and the relative uncertainty stated in brackets. The total uncertainty is calculated from the quadratic sum of the uncertainties from the *near* fit and the extrapolation uncertainty given by the *far* fit uncertainty or the difference of the *near* and *far* measurements.

Category	Channel	$r_{\text{QCD}}^{\text{near}} (\pm \text{rel.unc}[\%])$	$r_{\text{QCD}}^{\text{far}} (\pm \text{rel.unc}[\%])$	$\sigma_{\text{QCD}}^{\text{tot}} [\%]$
0-jet	$e\tau_h$	1.20 ± 0.08 (7.6)	0.98 ± 0.07 (7.1)	14.8
	$\mu\tau_h$	1.15 ± 0.04 (3.4)	1.25 ± 0.02 (1.6)	9.3
boosted	$e\tau_h$	1.76 ± 0.22 (12.5)	1.11 ± 0.12 (11)	38
	$\mu\tau_h$	1.19 ± 0.06 (5.0)	1.18 ± 0.04 (3.3)	6
dijet (inclusive)	$e\tau_h$	2.00 ± 0.27 (13.5)	1.02 ± 0.09 (8.8)	47
	$\mu\tau_h$	1.30 ± 0.06 (4.6)	1.13 ± 0.06 (5.3)	11
dijet lowboost	$\mu\tau_h$	1.60 ± 0.31 (19.3)	1.42 ± 0.20 (14)	23

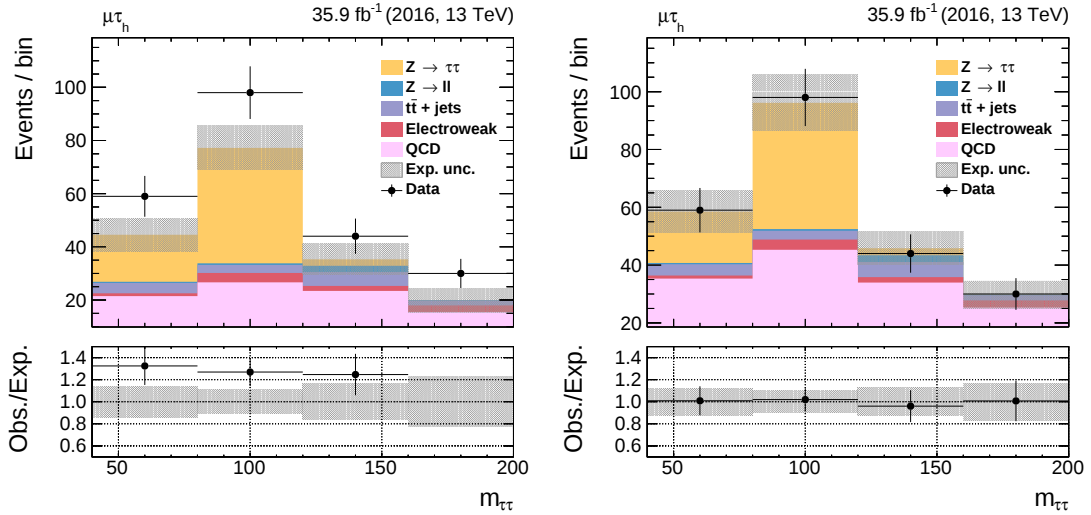


Fig. 3.20.: Prefit and postfit distributions in the $\mu\tau_h$ channel and the *dijet lowboost* category for the *near* sideband.

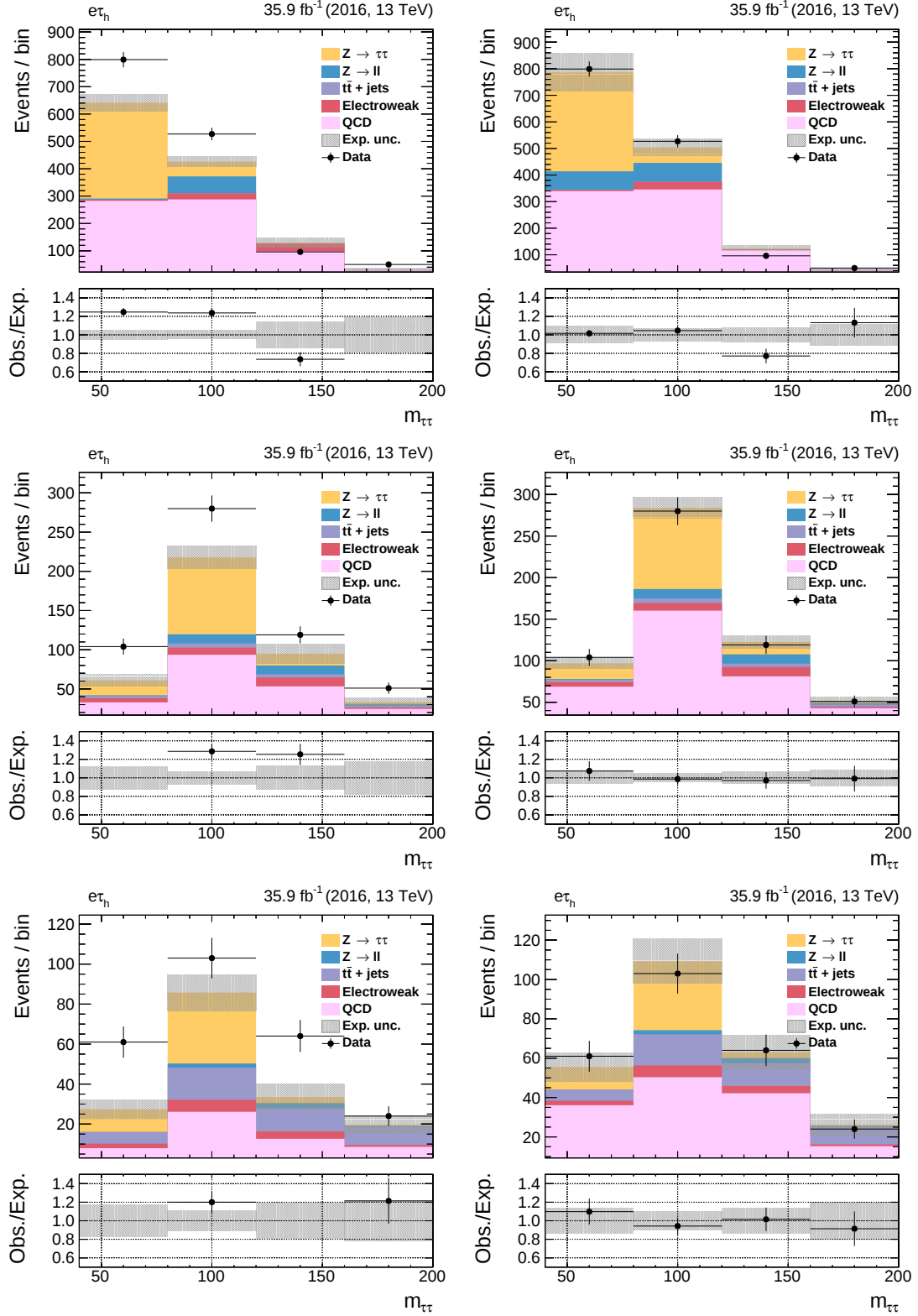


Fig. 3.21.: Prefit (left) and Postfit (right) distributions in the $e\tau_h$ channel for the *near* sideband for the *0-jet* (upper row), *boosted* (middle row) and *dijet* (bottom row) categories. During the fit, the QCD rate is pulled by the fit to find an optimal normalization of the background to describe data.

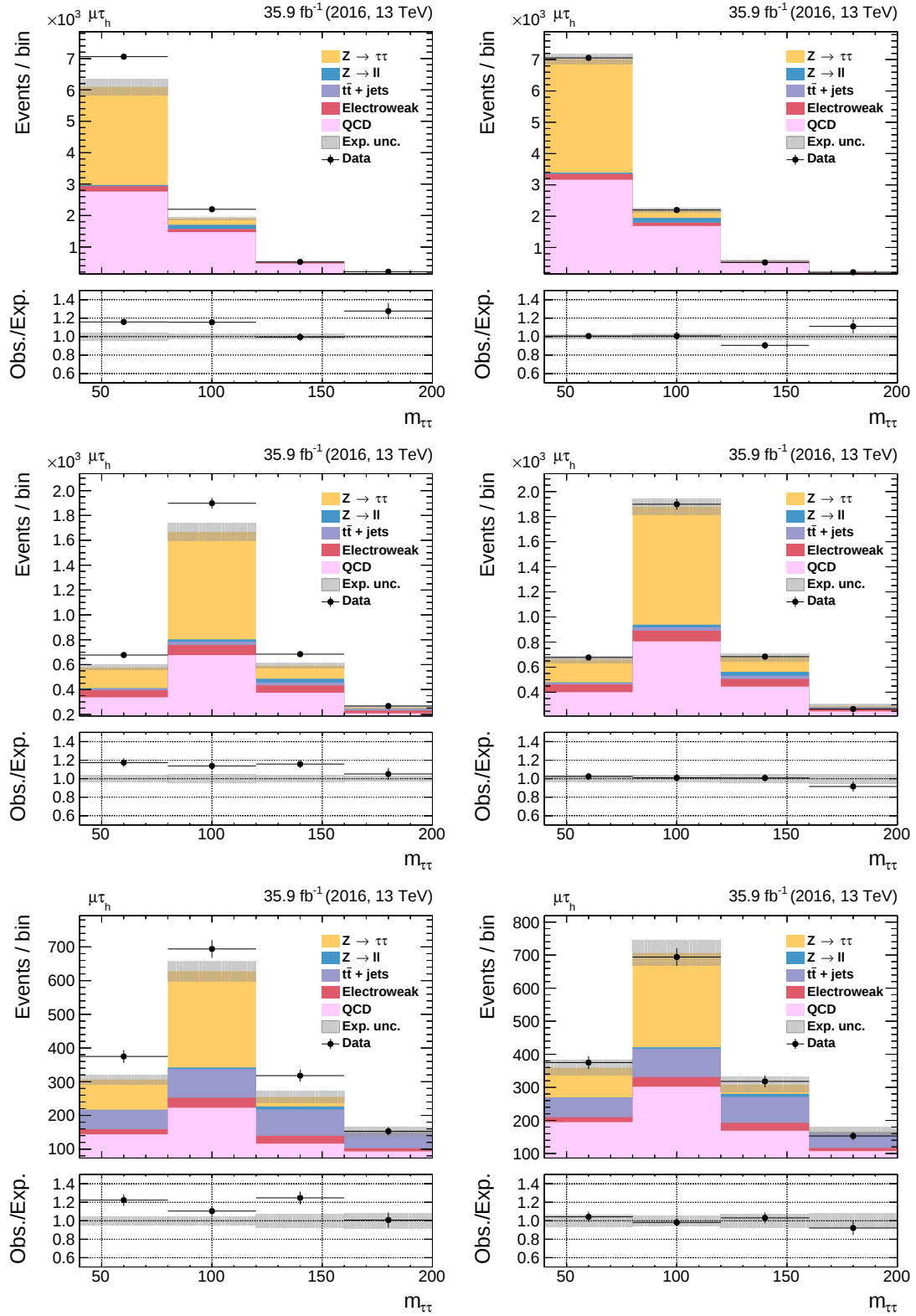


Fig. 3.22.: Prefit (left) and postfit (right) distributions in the $\mu\tau_h$ channel for the *near* sideband for the *0-jet* (upper row), *boosted* (middle row) and *dijet* (bottom row) categories. During the fit, the QCD rate is pulled by the fit to find an optimal normalization of the background to describe data.

QCD control regions

For all signal categories one QCD control region is included to the final fit to constrain the QCD yield in the $e\tau_h$ and $\mu\tau_h$ channels. These control regions require the charges of the tau-plus-lepton pair to have the same sign. All other cuts defining the selection in the signal region are kept fixed. Examples of these control regions are shown in figure 3.23. More plots can be found in the appendix in figure A.10.

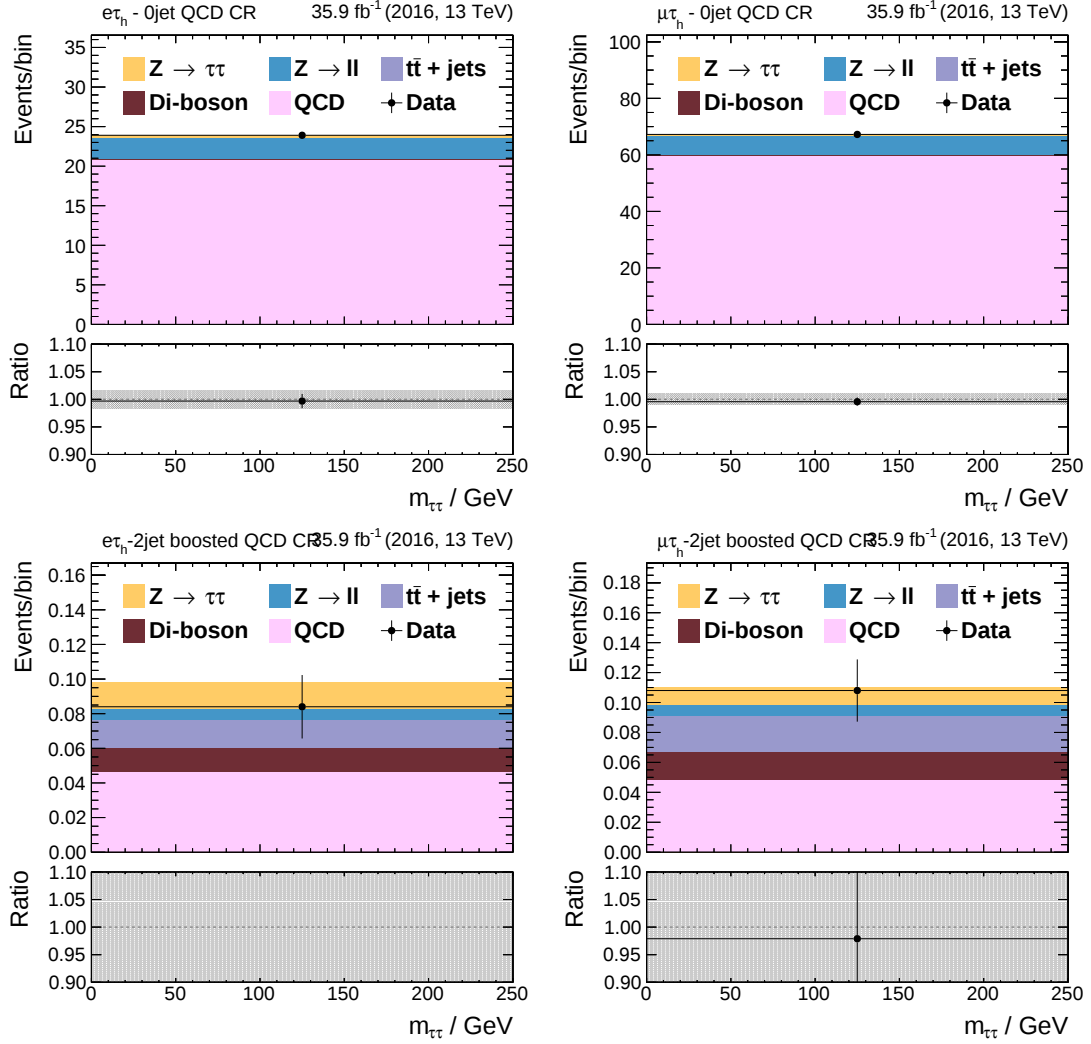


Fig. 3.23.: Background modeling and data in the QCD control region in the 0 -jet category (upper row) and the $dijet$ boosted category (bottom row) for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channel.

3.2.4.3 $\tau_h \tau_h$ channel

In the $\tau_h \tau_h$ channel QCD becomes a dominant background because quark or gluon jets are likely to be misidentified as hadronically decaying tau leptons.

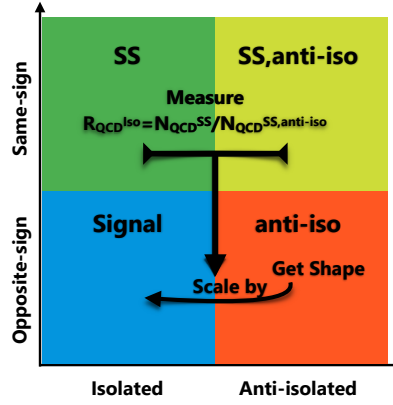


Fig. 3.24.: Estimation procedure of the QCD background in the $\tau_h \tau_h$ channel. The control region in anti-isolated events is used to extract the shape of QCD from data. The anti-isolated to isolated extrapolation factor $R_{\text{QCD}}^{\text{iso}}$ is measured in same-sign events. The factor scales the yield of the anti-isolated QCD estimate to the signal region giving the final estimate.

The multijet background in this channel is estimated from data in an anti-isolated control region as proposed in [22]. For this, the working point (WP) of the isolation for one of the two hadronically decaying tau leptons is relaxed to pass the *medium* WP of the MVA-based tau ID, which describes the tau lepton candidate isolation, while simultaneously the isolation of the other τ_h is inverted, meaning that only tau leptons that do not pass the *tight* WP are selected. In any case, the loose working point must be passed by both tau leptons. The condition for the control region is

$$I_{\tau_h,1}! = \text{tight WP} \quad \& \quad I_{\tau_h,1} = \text{loose WP} \quad \& \quad I_{\tau_h,2} = \text{medium WP} \quad \text{or} \quad (3.16)$$

$$I_{\tau_h,2}! = \text{tight WP} \quad \& \quad I_{\tau_h,2} = \text{loose WP} \quad \& \quad I_{\tau_h,1} = \text{medium WP}. \quad (3.17)$$

By inverting the tight WP of one of the τ_h 's it is ensured that no overlap exists with tau leptons that enter the signal regions. The relaxation of the other τ_h leads more events in the sideband regions, giving statistically more robust results. Furthermore, another sideband is constructed by inverting the requirement that the two tau leptons must have oppositely signed electric charges (see figure 3.24). Backgrounds taken from simulation are subtracted from data in both isolated and anti-isolated sideband regions to obtain the number of multijet events. This SS sideband is used to calculate the ratio of isolated to anti-isolated multijet events

$$R_{\text{QCD}}^{\text{iso}} = \frac{N_{\text{QCD}}^{\text{SS}}}{N_{\text{QCD}}^{\text{SS,anti-iso}}} = \frac{N_{\text{data}}^{\text{SS}} - \sum_i N_{\text{bkg},i}^{\text{SS}}}{N_{\text{data}}^{\text{SS,anti-iso}} - \sum_i N_{\text{bkg},i}^{\text{SS,anti-iso}}}, \quad (3.18)$$

Tab. 3.13.: Scale factors and systematic uncertainties arising from the estimation of the multijet background in the $\tau_h\tau_h$ channel. From [85].

Category	Scale factor	Type	rel. uncertainty
0-jet	0.26	lnN	2%
boosted	0.23	lnN	4%
dijet lowboost	0.22	lnN	8%
dijet boosted	0.20	lnN	48%

yielding the scale factors in table 3.13. The naming convention follows the scheme that is introduced in (see figure 3.24). The scale factor is applied to the QCD estimate in the anti-isolated control region. The result is the final QCD yield in the signal region.

As a same-sign QCD control region is added to the final fit, statistical uncertainties are covered by the bin-by-bin uncertainties in that region. Nevertheless, the choice of the anti-isolated region might be biased. For this reason, a systematic uncertainty is added to account for the extrapolation from the semi-isolated region (in which only one tau is anti-isolated) to the signal region. This uncertainty is derived by repeating the estimation technique described above using a fully-isolated sideband, in which both taus are anti-isolated, to estimate the QCD contribution in the semi-isolated region $N_{QCD}^{Semi-anti,estimated}$ in a data-driven manner. This means both tau leptons must now fulfill the criterion

$$I_{\tau_h,1!} = \text{tight WP} \quad \& \quad I_{\tau_h,1} = \text{loose WP} \quad \text{and} \quad (3.19)$$

$$I_{\tau_h,2!} = \text{tight WP} \quad \& \quad I_{\tau_h,2} = \text{loose WP}. \quad (3.20)$$

The difference between $N_{QCD}^{Semi-anti,estimated}$ and the number of events as calculated for equation (3.18) is consequently taken as the systematic uncertainty. This procedure was followed by [85] yielding the values in table 3.13.

QCD control regions

A QCD enriched control region for each signal category is added to the final fit to adjust the normalization of the QCD background dynamically. The expected number of background events and data is shown for all categories in figure 3.25. The *0-jet* and *boosted* category provides contain 95% QCD events, while the *dijet lowboost* category still shows a fraction of 85%. The *dijet boosted* category only contains 20% expected QCD events.

3.2.5 VBF Higgs

Higgs bosons produced via vector boson fusion (VBF) are an important background in this analysis as their two quark-initiated forward jets are likely to be misidentified as the signature of the two jets expected for the ggH signal. Dedicated JHU samples are used to derive the shape and yield of this background (see tab. 3.2). As it must be accounted for the possibility of CP violation also in this vector boson coupling of the Higgs boson,

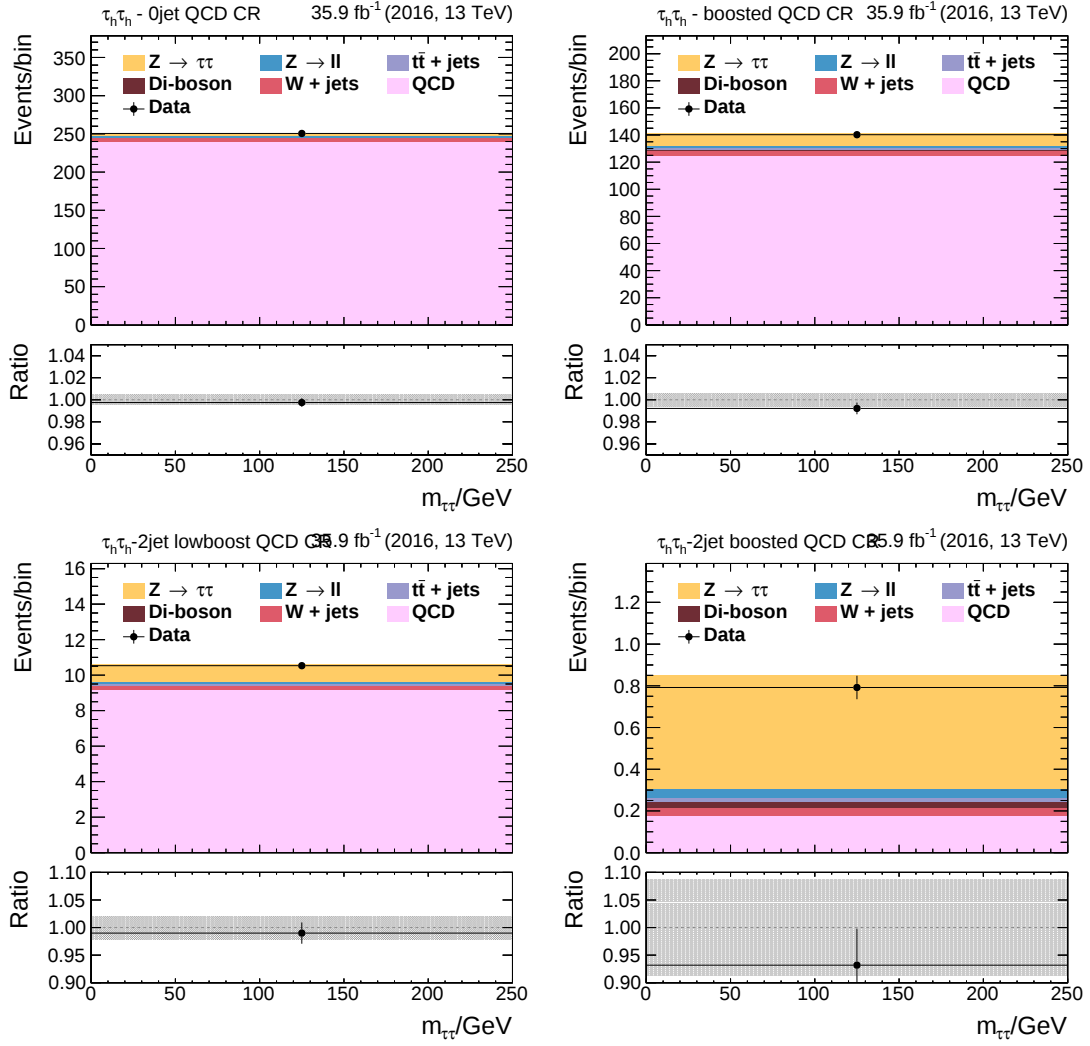


Fig. 3.25.: Background modeling and data of the QCD control regions in the $\tau_h \tau_h$ final state. The *0-jet* (upper left), *boosted* (upper right) and *dijet lowboost* (bottom left) control regions reach purities in the QCD background above 85%. The *dijet boosted* category (bottom right) only a purity of 20%.

the shape and normalization of the $qq \rightarrow H$ background is modeled by the physics model introduced in section 3.4.1.

3.2.6 Other backgrounds

There are other smaller sources of background such as events with two vector bosons or events with single top quarks and top-quarks in association with a W boson. The shapes and yields of these backgrounds are fully determined by simulation. As for the $Z \rightarrow \tau\tau$ background they are categorized according to the match of the reconstructed τ_h and the generated object. This means there are two contributions

- VVT: τ_h is matched to the generated hadronic tau in $e\tau_h$, $\mu\tau_h$ or $\tau_h\tau_h$.
- VVJ: otherwise.

3.3 Systematic uncertainties

In addition to the systematic uncertainties arising from the background estimation, uncertainties on normalizations, reconstruction and identification efficiencies and theoretical ambiguities are applied. There are in general two types of systematic uncertainties. *Shape* uncertainties affect the yield in every bin of the investigated histograms individually and thus alters the shape, and thereby to some extent also the normalization, of them. Whereas, *log-normal* (lnN) uncertainties modify the overall rate of a process. The majority of systematic uncertainties included in this analysis was taken from the search of the Higgs boson decaying into tau leptons [22] as this analysis is based on it. These uncertainties are summarized in table 3.16 and 3.17. Additional analysis-specific systematic uncertainties and modifications of the existing ones are discussed in this section.

3.3.1 Limited size of MC samples

As the simulated Monte Carlo event samples only have a finite number of events, the number of events from simulated samples in every bin is influenced by statistical fluctuations. To incorporate these statistical uncertainties in the final fit, bin-by-bin uncertainties are added as additional nuisance parameters [86, 87]. These nuisances depend only on the bin content and allows the processes to float during the final fit to take these fluctuations into account.

3.3.2 Gluon fusion in association with jets

As this analysis is performed in bins with different number of jets, so-called jet bin acceptance uncertainties must be considered as theory uncertainties on the Higgs boson in association with jets cross section. In addition to an uncertainty on the total yield in each jet bin that can be calculated with fixed-order perturbative calculations, migration uncertainties between the bins must be taken into account. These cancel each other if the inclusive jet bin is investigated. Uncertainties on this jet bin acceptance for the Higgs boson signal in association with 0, 1 or more jets are applied according to the Stewart-Tackmann method in [21, 88]. As listed in table 3.14 these uncertainties are applied in different categories according to the possible number of jets in these categories.

3.3.3 Jet energy correction uncertainties

The energy of jets is corrected for various detector effects and the resulting systematic uncertainties are provided by the CMS collaboration as a function of the pseudorapidity and transverse momentum of the jet. There are in total 27 individual, uncorrelated sources of jet energy correction uncertainties (JEC uncertainties) [89]. Every of these sources modifies the four-momentum of the jets and has consequently an impact on the shape of the final observables as calculated by MELA. It is technically very challenging, and more important very time-consuming, to propagate the impact of every single JEC uncertainty through the

Tab. 3.14.: Jet bin acceptance systematic uncertainties added as theory uncertainties on the $gg \rightarrow H$ process according to the Stewart-Tackmann method.

Source	Type	Channel	Processes	Value
STXSmig01 - 0-jet	lnN	All	$gg \rightarrow H$	4.1%
STXSmig01 - boosted	lnN	All	$gg \rightarrow H$	7.9%
STXSmig01 - dijet	lnN	All	$gg \rightarrow H$	3.9%
STXSmig12 - boosted	lnN	All	$gg \rightarrow H$	6.8%
STXSmig12 - dijet	lnN	All	$gg \rightarrow H$	16.1%
STXSVBF2j - dijet	lnN	All	$gg \rightarrow H$	20%

Tab. 3.15.: List of JEC uncertainty sources assigned to the three pseudorapidity regions used in this analysis to group these systematic uncertainties.

η	JEC sources
0-3	SinglePionECAL, SinglePionHCAL, AbsoluteFlavMap, AbsoluteMPFBias, AbsoluteScale, AbsoluteStat, Fragmentation, FlavorQCD, TimePtEta, PileUpDataMC, RelativeFSR, RelativeStatFSR, PileUpPtRef
3-5	PileUpPtEC1, PileUpPtEC2, PileUpPtBB, RelativeJEREC1, RelativeJEREC2, RelativePtEC1, RelativePtEC2, RelativeStatEC, RelativePtBB
0-5	RelativeStatHF, RelativePtHF, PileUpPtHF, RelativeJERHF

entire analysis. For this reason usually the total JEC uncertainty calculated by the quadratic sum of the individual sources is applied. Despite that, it was found in Ref. [22] that the final fit over-constrained the total JEC uncertainty, and it was therefore decided to use the 27 individual sources instead. For this analysis it makes sense to seek for a good comprise in terms of physical modeling of the JEC uncertainties and computing time. Therefore, JEC uncertainties groupings are formed (see tab. 3.15) whereby the individual uncertainties build mutual sets according to the region of the detector where they are applied. It is possible to divide the detector into regions of pseudorapidity in this case. The only single uncertainty that forms a group by itself is the *RelativeBal* uncertainty that was found to be unexplicably large. This uncertainty was introduced in Run 2 and covers the difference between the fits of MPF and p_T balance methods for deriving the JECs [90]. Thus, it affects jets in the whole detector. A comparison of the individual, total and grouped uncertainties was not implemented by means of studying the impact of the nuisances on the parameter of interest. But the effect was studied by the CMS collaboration and shown in Ref. [91] with an analysis setup which is reasonably close to the setup used here. Therefore, the groupings are expected to be less constrained in the final fit than the total JEC uncertainty, too.

Tab. 3.16.: List of systematics uncertainties. All uncertainties were introduced in the SM search of the decay of the Higgs boson in a pair of tau leptons and discussed there. In this table * denotes that all channel or all processes taken from simulation are affected by the corresponding uncertainty.

Source	Type	Channel	Processes	Value
Efficiencies, Luminosity and triggers				
Luminosity	lnN	*	*	2.5%
$\epsilon_{\text{Trigger}}$	lnN	$e\mu, e\tau_h, \mu\tau_h$	*	2%
		$\tau_h\tau_h$	*	10%
ϵ_e	lnN	$e\mu, e\tau_h$	*	2%
ϵ_μ	lnN	$e\mu, \mu\tau_h$	*	2%
		$e\mu, \mu\tau_h$ - correlated	*	4.5%
		$e\mu, \mu\tau_h$ - channel-specific	*	2%
		$\tau_h\tau_h$ - correlated	ZTT,VVT,TTT,EWKZ	9%
ϵ_τ	lnN	$\tau_h\tau_h$ - channel-specific	ZTT,VVT,TTT,EWKZ	4%
		$\tau_h\tau_h$ - correlated	TTJ,ZJ,VVJ,W	6%
		$\tau_h\tau_h$ - channel-specific	TTJ,ZJ,VVJ,W	2%
$\epsilon_{\text{b-tag}}$	shape	$e\mu, e\tau_h, \mu\tau_h$	*	
b-tag mis. id. rate	shape	$e\mu, e\tau_h, \mu\tau_h$	*	
Energy scale (ES)				
τ ES - e	shape	$e\mu$	*	
τ ES - τ_h - 1prong	shape	$e\tau_h, \mu\tau_h, \tau_h\tau_h$	ZTT,TTT,TT,VV,VVT,EWKZ	
τ ES - τ_h - 1prong1pi0	shape	$e\tau_h, \mu\tau_h, \tau_h\tau_h$	ZTT,TTT,TT,VV,VVT,EWKZ	
τ ES - τ_h - 3prong	shape	$e\tau_h, \mu\tau_h, \tau_h\tau_h$	ZTT,TTT,TT,VV,VVT,EWKZ	
Jet ES - $0 < \eta < 3$	shape	*	*	
Jet ES - $0 < \eta < 5$	shape	*	*	
Jet ES - $3 < \eta < 5$	shape	*	*	
Jet ES - RelativeBal	shape	*	*	
MET ES - unclustered	shape	*	*	
MET ES - clustered	shape	*	*	
Normalization				
VV xsec	lnN	*	VV,VVT,VVJ	5%
ZJ xsec	lnN	*	ZTT,ZJ,ZL,ZLL	4%
EWKZ xsec	lnN	*	EWKZ	4%
Reweighting				
DY reweighting	shape	*	ZTT,ZJ,ZL,ZLL	
$t\bar{t}$ reweighting	shape	*	TTJ,TTT,TT,	

Tab. 3.17.: List of systematics uncertainties (continuation).

Source	Type	Channel	Processes	Value
Misidentification rates				
ZL - 1prong	shape	$e\tau_h, \mu\tau_h$	ZL	
ZL - 1prong1pi0	shape	$e\tau_h, \mu\tau_h$	ZL	
$\mu \rightarrow \tau_h$ mis.id.	lnN	$\mu\tau_h$	ZL	7%
$e \rightarrow \tau_h$ mis.id.	lnN	$e\tau_h$	ZL	10%
	lnN	$\tau_h\tau_h$	W	20%
jet $\rightarrow \tau_h$ mis.id.	lnN	$e\tau_h, \mu\tau_h$	VVJ, TTJ, ZJ	20%
	shape	$e\tau_h, \mu\tau_h$	W	
Theoretical uncertainties				
gg Scale	shape	*	gg $\rightarrow$ H	
Finite quark mass	shape	*	gg $\rightarrow$ H	
Parton showering - 0-jet	lnN	*	gg $\rightarrow$ H	1.5%
Parton showering - boosted	lnN	*	gg $\rightarrow$ H	5.5%
Parton showering - dijet	lnN	*	gg $\rightarrow$ H	3%
Theory uncertainty	lnN	*	$H \rightarrow \tau\tau$	1.7%
Pile Up	lnN	*	$H \rightarrow \tau\tau$	0.99%
Pile Up - g_s	lnN	*	$H \rightarrow \tau\tau$	0.62%
Theory uncertainty	lnN	*	$H \rightarrow WW$	0.99%
Pile Up	lnN	*	$H \rightarrow WW$	0.99%
Pile Up - g_s	lnN	*	$H \rightarrow WW$	0.66%
	lnN	*	gg $\rightarrow$ H	3.9%
	lnN	*	qq $\rightarrow$ H	0.4%
QCD Scale	lnN	*	WH	0.7%
	lnN	*	ZH	3.8%
	lnN	*	gg $\rightarrow$ H	3.2%
	lnN	*	qq $\rightarrow$ H	2.1%
Higgs PDF	lnN	*	WH	1.9%
	lnN	*	ZH	1.6%

3.4 Performance of the measurement of the CP mixing angle in gluon-gluon fusion

The performance of the analysis to distinguish different CP scenarios in the coupling of the Higgs boson to gluons is quantitatively evaluated by a measurement of the CP mixing angle α in a MC simulation. For this purpose a pseudo-data set is constructed from the background-plus-SM-signal hypothesis for a scalar SM Higgs boson. To perform the measurement a likelihood function is constructed and maximized for various signal+background scenarios that depend on α . The scenario that shows the best compatability of the background-plus-signal hypothesis and the pseudo-data set is, thus, equivalent to a measurement of the actual mixing angle realized in (pseudo-)data.

3.4.1 Signal model

As seen before, a different CP scenario predicts a different shape and yield of the signal templates. Since the production of dedicated Monte Carlo samples for a larger number of CP scenarios consumes a lot of CPU time and is thus very expensive, a model is developed that allows for the reweighting of three existing templates to provide shape and yield of any CP scenario as a function of the mixing angle α . In this analysis, these templates are constructed using the three $gg \rightarrow H + 2j$ JHU samples ((see tab. 3.2)).

The differential cross section of any CP hypothesis is proportional to the matrix element squared $|M|^2$. This matrix element can be calculated from the scalar and the pseudoscalar hypotheses and has the form

$$|M(\alpha)|^2 = \left| aM_{\alpha=0} + bM_{\alpha=\frac{\pi}{2}} \right|^2 = a^2 |M_{\alpha=0}|^2 + b^2 |M_{\alpha=\frac{\pi}{2}}|^2 + 2abM_{\text{interference}}^2 \quad (3.21)$$

with the matrix elements $M^{\alpha=0}$ and $M^{\alpha=\frac{\pi}{2}}$ for the scalar and pseudoscalar scenarios. The mixing parameters a and b , that alter the admixture of the pseudoscalar matrix element, are functions of the mixing angle α . As a consequence, a matrix element $M_{\text{interference}}$ for the interference of the two scenarios is build from the product of this two matrix elements. The form of this element is non-trivial and must be calculated by or given to the event generator for any set of mixing parameters. The cross section is proportional to M^2 meaning that it holds

$$\frac{d\sigma_{\text{tot}}}{dx} = a^2 \frac{d\sigma_{\alpha=0}}{dx} + b^2 \frac{d\sigma_{\alpha=\frac{\pi}{2}}}{dx} + 2ab \frac{d\sigma_{\text{interference}}}{dx}. \quad (3.22)$$

As a consequence, the differential cross section for a given CP scenario with fixed angle $\tilde{\alpha}$ and non-vanishing mixing parameters $\tilde{a} \neq 0$ and $\tilde{b} \neq 0$ is

$$\frac{d\sigma_{\text{tot}}(\tilde{\alpha})}{dx} = \tilde{a}^2 \frac{d\sigma_{\alpha=0}}{dx} + \tilde{b}^2 \frac{d\sigma_{\alpha=\frac{\pi}{2}}}{dx} + 2\tilde{a}\tilde{b} \frac{d\sigma_{\text{interference}}}{dx}. \quad (3.23)$$

The interference term can be substituted in equation (3.22) by

$$\frac{d\sigma_{\text{interference}}(\tilde{a}, \tilde{b})}{dx} = \frac{1}{2\tilde{a}\tilde{b}} \frac{d\sigma_{\text{tot}}(\tilde{\alpha})}{dx} - \frac{\tilde{a}}{\tilde{b}} \frac{d\sigma_{\alpha=0}}{dx} - \frac{\tilde{b}}{\tilde{a}} \frac{d\sigma_{\alpha=\frac{\pi}{2}}}{dx}. \quad (3.24)$$

Tab. 3.18.: Mixing parameters used for the signal model. The fractions of the cross sections for the process $gg \rightarrow H + 2j \rightarrow \tau\tau + 2j$ are $\sigma_{ggH+2j,SM} = 0.399064$ pb and $\sigma_{ggH+2,PS} = 0.393961$ fb. The fractions of the cross sections for the VBF process are $\sigma_{qqH,SM} = 2.6707$ pb and $\sigma_{qqH,PS} = 0.2371$ fb.

Generator	Process	$\tilde{a}$	$\tilde{b}$	a	b
JHU	ggH	1	1.0062	$\sqrt{\mu_F} \cos(\alpha)$	$\sqrt{\mu_F} \sin(\alpha) \sqrt{\frac{\sigma_{ggH,SM}}{\sigma_{ggH,PS}}}$
MadGraph		$\cos(\pi/4)$	$\sin(\pi/4)$		
JHU	VBF	1	0.297979	$\sqrt{(1- f)} \mu_V$	$\text{Sgn}(f) \sqrt{ f } \mu_V \frac{\sigma_{VBF,SM}}{\sigma_{VBF,PS}}$

This means, any scenario for a given CP mixing angle α with corresponding parameters $a = a(\alpha)$ and $b = b(\alpha)$ is determined by the function

$$\frac{d\sigma_{\text{tot}}}{dx} = \left(a^2 - ab \frac{\tilde{a}}{\tilde{b}} \right) \frac{d\sigma_{\alpha=0}}{dx} + \left(b^2 - ab \frac{\tilde{b}}{\tilde{a}} \right) \frac{d\sigma_{\alpha=\frac{\pi}{2}}}{dx} + \frac{ab}{\tilde{a}\tilde{b}} \frac{d\sigma_{\text{tot}}(\tilde{\alpha})}{dx}. \quad (3.25)$$

The fixed mixing parameters $\tilde{a}$ and $\tilde{b}$ depend on the setup used to produce the Monte Carlo samples and are summarized for the JHU and POWHEG generators in table 3.18. In this case the intermediate sample with a fixed mixing angle of $\tilde{\alpha} = \frac{\pi}{4}$ provides the shape for $\frac{d\sigma_{\text{tot}}(\tilde{\alpha})}{dx}$. The JHU samples used in this analysis were produced such that the cross sections for the scalar and pseudoscalar terms are roughly equal. Therefore, the normalizations are close to one. Last but not least, it is noted that the samples providing the templates give the shape of the CP sensitive observables $x \in [D_{CP}, D_{0-}, \Delta\phi_{jj}]$ in terms of the number of events per bin $\frac{dN_{\text{tot}}}{dx}$, which are proportional to the cross section by the integrated luminosity L of the event sample

$$\frac{dN_{\text{tot}}}{dx} = L \cdot \frac{d\sigma_{\text{tot}}}{dx}. \quad (3.26)$$

The occurrence of CP violation in the coupling of the Higgs boson to vector bosons is not completely ruled out yet. As pseudoscalar couplings to vector bosons can only occur at a higher order of perturbation theory compared to the scalar couplings, they are suppressed by at least g_W^2 , where g_W is the weak coupling constant. The effect of CP violation in the VBF background is subsequently expected to cause a rather small impact in this analysis. To account for the possibility the physics model allows to float the shape of the VBF background, too. The CP scenarios are provided by the dedicated JHU samples given in table 3.2. The shape is modeled by the parameter f (see section 1.2.3.2) as given in table 3.18, whereas the overall rate is modeled by the signal strength modifier μ_V . Here, unless otherwise stated, the VBF shape will be fixed to the SM-value $f = 0$. As a pure shape analysis is performed, the larger pseudoscalar coupling is not taken into account by this physics model. However, for cross-checks it might be interesting to fix the coupling strength to the SM value to make a more model-dependent measurement. In that case, the gluon-gluon fusion Higgs boson production rate is parametrized by

$$\sigma(\alpha) = \cos^2 \alpha + \left(\frac{3}{2} \right)^2 \sin^2 \alpha \quad (3.27)$$

and taken in account accordingly.

As a conclusion, signal and background are modeled by four independent parameters of interest (POI) where α and f determine the CP scenario in fermionic and vector boson couplings, respectively, and are sensitive to the underlying tensor structure of the interaction. μ_F and μ_V are the coupling strengths that modify the overall rates of the two production processes. These four POI are denoted by the vector $\vec{\zeta} = \{\alpha, f, \mu_F, \mu_V\}$ in the following.

3.4.2 Fit function

To measure the mixing angle α the signal-plus-background expectation as a function of α must be compared to data. This is done by constructing a binned likelihood function. To get a quantitative estimate of the analysis sensitivity, a pseudo-data set replaces the number of observed events in the following.

3.4.2.1 Pseudo-data set

The pseudo-data set represents the SM scenario with $\alpha = 0$ and is constructed by adding the number of expected background events as estimated with the methods described in section 3.2 and the number of events expected from a scalar Higgs boson hypothesis:

$$N_i^{\text{pseudo}} = B_i(f = 0, \mu_V = 1) + S_i(\alpha = 0, \mu_F = 1). \quad (3.28)$$

The number of signal and background events are added in every bin i specified by the categories and binnings in section 3.1.4. The VBF background and ggF signal is thereby modeled as described in section 3.4.1.

3.4.2.2 Likelihood function

To find the CP scenario that is realized in data the maximum likelihood method is used. The binned likelihood function is defined through the number of observed, background and signal events in each bin i and the probability density functions of each of the j nuisance parameters θ_j , here denoted by the vector of nuisances $\vec{\theta}$. Here, the likelihood function has the form

$$\mathcal{L}(\vec{D}|\vec{\zeta}, \vec{\theta}) = \prod_i \frac{(S_i(\alpha, \mu_F) + B_i(f, \mu_V))^{D_i^{(\text{pseudo})}}}{D_i^{(\text{pseudo})}!} e^{-(S_i(\alpha, \mu_F) + B_i(f, \mu_V))} \prod_j p(\theta_j|\tilde{\theta}_j). \quad (3.29)$$

The first term is constructed from the product over all bins specified in section 3.1.4 of Poisson probabilities to measure D_i events, while expecting S_i signal and B_i background events in the i -th bin. As defined in the signal model the S_i can be expressed through a function of α and a fermionic signal strength parameter μ_F . The background B_i also depends on the CP scenario occurring in the VBF production mode of the Higgs boson.

Accordingly, the number of expected background events in each bin is modeled by the sum of VBF events and the other CP-independent SM processes through

$$B_i(\mu_V, f) = B_i^{\text{VBF}}(\mu_V, f) + B_i^{\text{others}}. \quad (3.30)$$

This means, the likelihood function depends on the full vector of POIs $\vec{\zeta}$.

The second term of $\mathcal{L}$ models the impact of systematic uncertainties and is the product over the probability density functions of the nuisance parameters $\vec{\theta}$ defined in section 3.3. In the fit the likelihood function is maximized with respect to all POIs and all nuisance parameters. Consequently, the value of α that provides the best description of the taken data is the result of the fit. Here, the combine tool [92], a software package based on RooStats [93] and RooFit [94], is used to perform the fit. The likelihood function must be maximized for various mixing angles and tested against a certain null-hypothesis. In particle physics this null-hypothesis is usually the SM prediction which is represented by the pseudo-data set in this case. The test statistic used here is the profile likelihood ratio

$$q = -2 \ln \frac{\mathcal{L}(\vec{D}^{(\text{Pseudo})} | \vec{\zeta}, \hat{\vec{\theta}}_{\vec{\zeta}})}{\mathcal{L}(\vec{D}^{(\text{Pseudo})} | \hat{\vec{\zeta}}, \hat{\vec{\theta}})}, \quad (3.31)$$

with $\hat{\vec{\zeta}} = (\alpha = 0, f = 0, \mu_F = 1, \mu_V = 1)$, which is the configuration of the pseudo-data set that subsequently provides the global maximum of the likelihood function together with the corresponding best-fit values of the nuisance parameters $\hat{\vec{\theta}}$. It can be inferred that the denominator subsequently sets the minimum of q . For the numerator various signal and background hypotheses described by $\vec{\zeta}$ can be evaluated with corresponding best-fit nuisance parameters $\hat{\vec{\theta}}_{\vec{\zeta}}$ that maximize $\mathcal{L}$ for a given $\vec{\zeta}$. q becomes minimal when the tested CP hypothesis matches the scenario realized in (pseudo-)data because the likelihood ratio becomes equal to one then. Consequently, q is zero and minimal then. Thus, the measurement of α is performed by finding the CP hypothesis that minimizes the test statistic. The profile likelihood ratio facilitates the determination of the uncertainty of a profiled parameter. With this form the 1σ and 2σ intervals can be determined by evaluating q at $q = 1.0$ and $q = 4.0$, respectively.

3.4.2.3 Postfit distributions

Except for the 0-jet category, two-dimensional and three-dimensional discriminators are used. Consequently, three-dimensional histogram must be fitted to data. In this analysis, these histograms are first unrolled to one-dimensional histograms to be passed to the combine tool that performs the maximum likelihood fit. The fit is performed on the combination of the histograms corresponding to the categories in each channel and the control regions specified in the background estimation. The unrolled postfit distributions of the input histograms in the $\tau_h \tau_h$ channel are shown in figures 3.26-3.27 with pseudo-data and the expected background and signal distributions after the fit. The input histograms before the fit that can be found in the appendix in section A.3.1 and a perfect agreement between data and the signal-plus-background hypothesis is found in all distributions there. The

reason for this is, that the observed events depict the pseudo-data set that was constructed from the background and signal model.

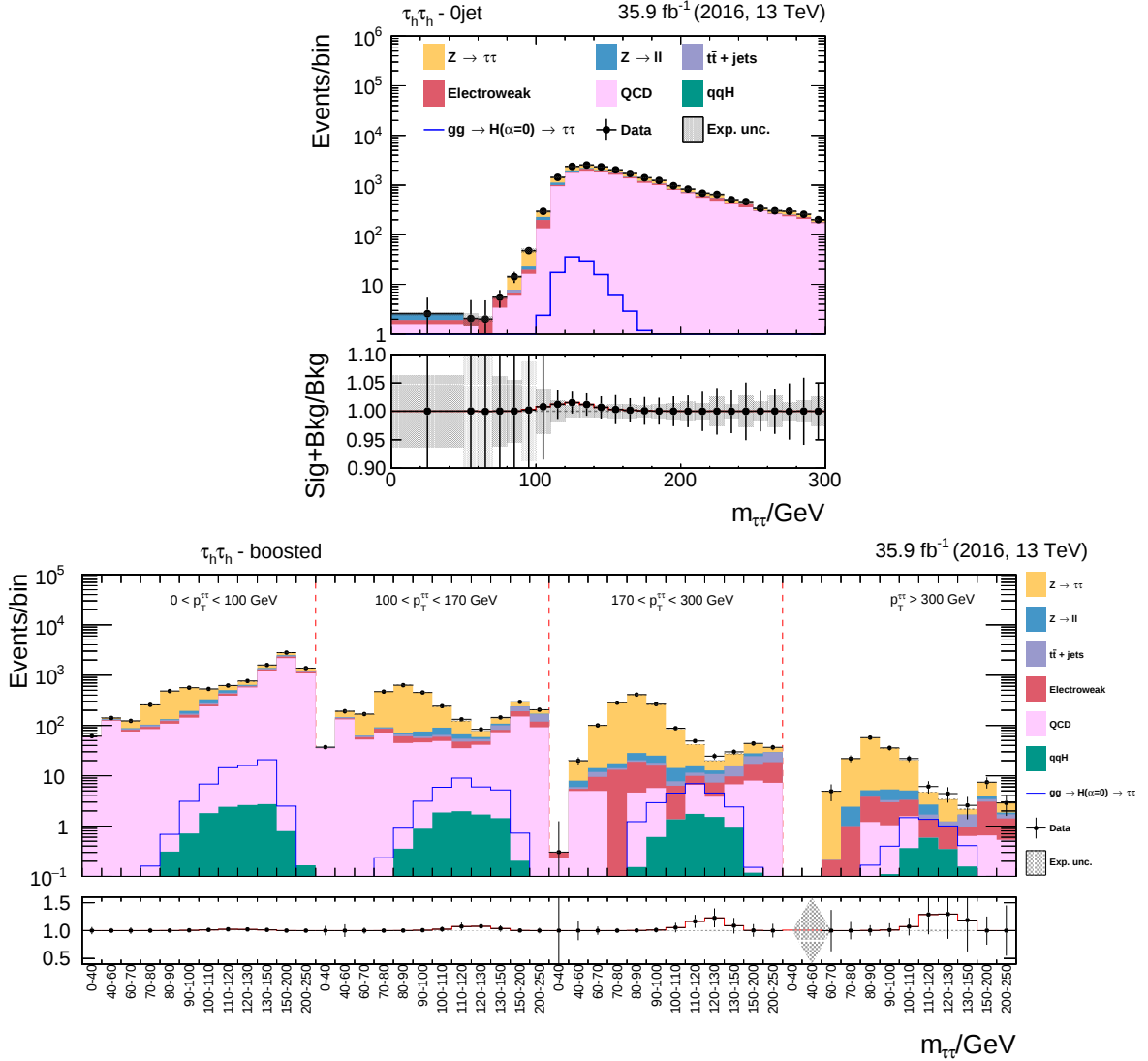


Fig. 3.26.: Postfit distributions of the expected background and signal model and pseudo-data with signal-plus-background-to-background ratio in the *O-jet* (upper row) and *boosted* (bottom row) categories for the $\tau_h \tau_h$ channel. For the *boosted* category the unrolled 2D histogram is shown. First, the distribution is unrolled in bins of $p_T^{\tau\tau}$ and then along $m_{\tau\tau}$. The borders between bins of $p_T^{\tau\tau}$ are indicated through the red dashed lines.

3.4.3 Results

Figure 3.28 shows the expected performance of the measurement of the CP mixing angle α using the kinematic observables D_{CP} and D_{0-} . To perform a pure shape analysis, the signal strength parameters μ_F and μ_V are left floating in the fit. The VBF background is assumed to be SM-like by fixing $f = 0$ in the fit. Including all systematic uncertainties and control regions, the analysis is able to distinguish between a pure scalar and pseudoscalar fermionic coupling at 1.34σ (right plot in figure 3.28). The settings of the pseudo-data set

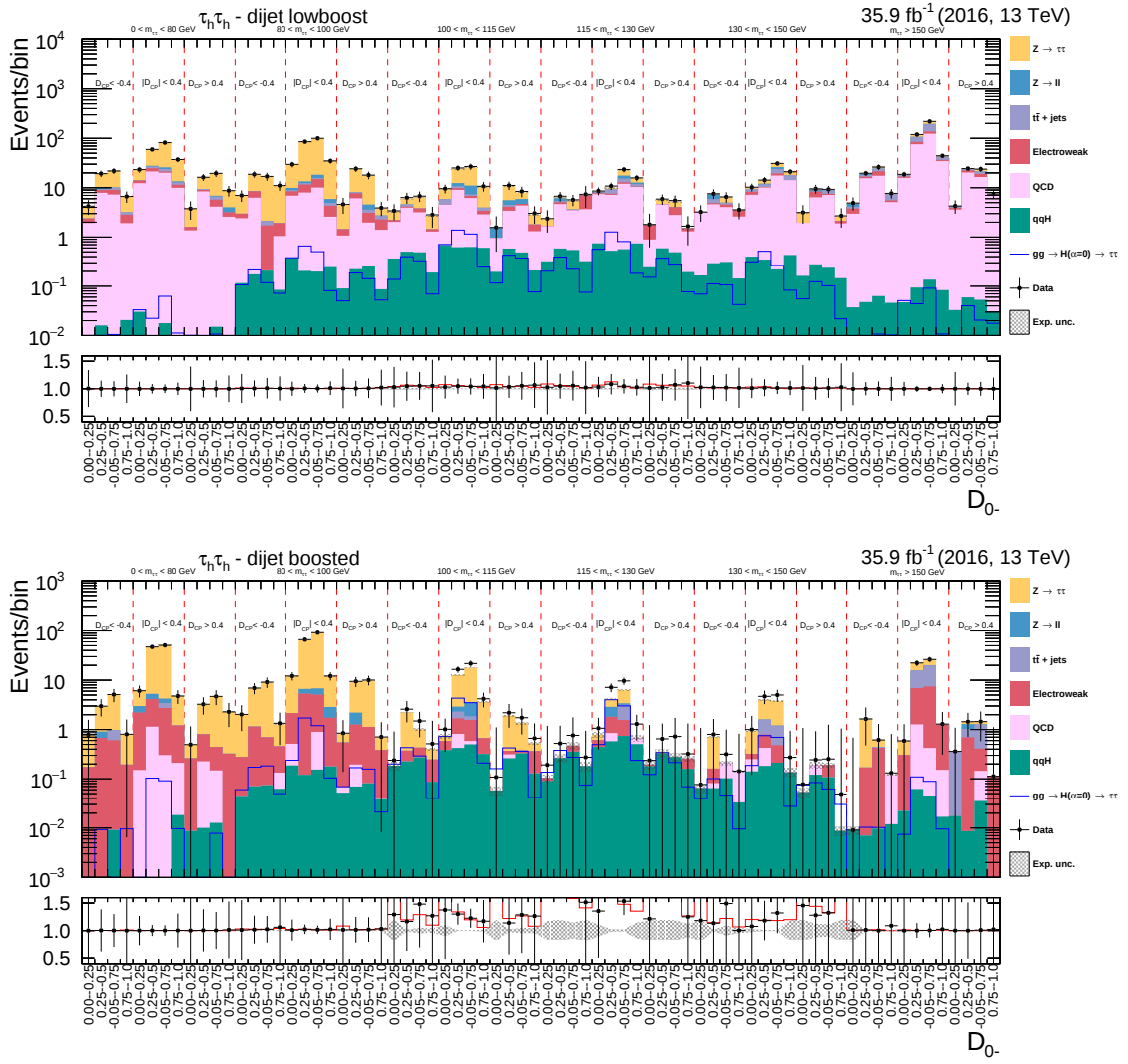


Fig. 3.27.: Postfit distributions of the expected background and signal model and pseudo-data with signal+background to background ratio in the *dijet lowboost* (upper row) and *dijet boosted* (bottom row) categories for the $\tau_h \tau_h$ channel. The 3D histogram is unrolled first along $m_{\tau\tau}$, then along D_{CP} and finally along D_{0-} . The borders between neighboring bins of D_{CP} are indicated by the red dashed lines. Bins of $m_{\tau\tau}$ are indicated by on top of the frame and belong always to three bins of D_{CP} . In the bottom plot it is possible to see that the signal-to-background ratio is optimal in bins of $m_{\tau\tau}$ that are close to the Higgs boson mass of 125.18 GeV.

are found by measuring α with an accuracy of 41° , which is read off the 1σ band. Examining the performance in the individual decay channels (middle plot in figure 3.28) one observes the $\tau_h \tau_h$ channel to provide the best performance to discriminate between different CP scenarios. As expected the $e\mu$ final state yields the worst performance because it has the smallest signal-to-background ratio (see fig. A.40, page 116). In order to see whether the fit is capable of reconstructing the imposed signal strength of the pseudo-data set, μ_F was profiled with the same setup leaving α floating in the fit. As can be seen in figure 3.29, the fit is able to find the correct signal strength of $\mu_F = 1$ in all channels. Here, the measurement comes with an uncertainty of 0.5 for the optimal discriminators.

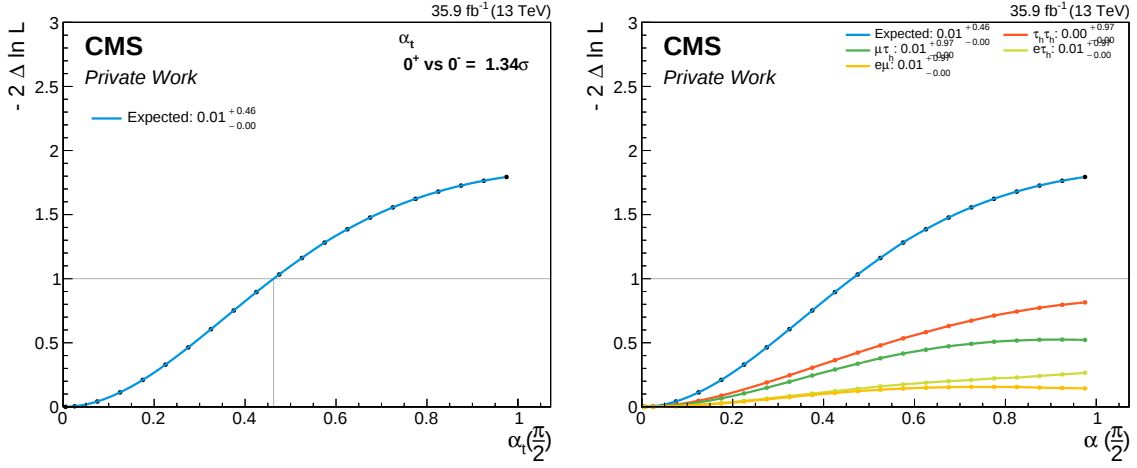


Fig. 3.28.: Measurement of the CP mixing angle α in the fermionic couplings of the Higgs boson to the top quark inferred from the effective gluon coupling under the assumption of a scalar Higgs boson and SM-like VBF production ($f = 0$). The analysis is able to distinguish a scalar from a pseudoscalar CP scenario using the MELA approach at 1.34σ . The right plot shows the measurement in each individual decay channel. The $\tau_h\tau_h$ channel is the most significant channel followed by the $\mu\tau_h$ channel.

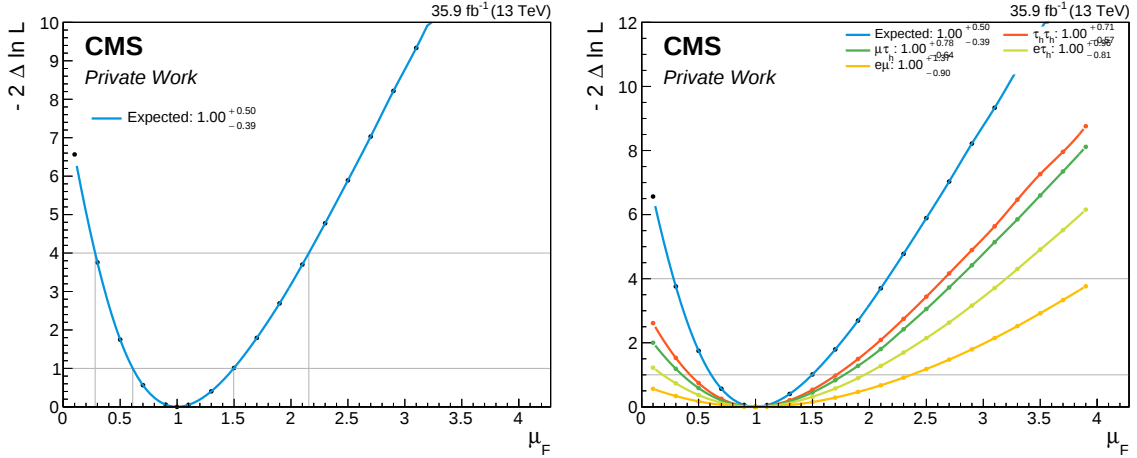


Fig. 3.29.: Measurement of the gluon-gluon fusion signal strength utilizing a signal model assuming a scalar Higgs boson and SM-like VBF production ($f = 0$). The analysis is able to find μ_F at the value set by the pseudo-data set with an accuracy of 0.5. The right plot shows the measurement in each individual decay channel. Again the $\tau_h\tau_h$ channel yields the tightest constraint followed by the $\mu\tau_h$ channel. In the left plot the measurement is performed for $\Delta\phi_{jj}$ showing that it gives the same result but with a looser constraint only.

3.4.3.1 Comparison to the measurement using $\Delta\phi_{jj}$

α was also profiled using $\Delta\phi_{jj}$ as discriminator instead of the combination of D_{CP} and D_{0-} . The profile is shown in the left plot in figure 3.30. Using $\Delta\phi_{jj}$ as the only discriminant, a less accurate measurement of $\alpha = 0.0^{+45}_{-0}^\circ$ is obtained because the information from four of the five angles in the event kinematics is missing. Moreover, this setup is only capable of distinguishing scalar and pseudoscalar couplings at 1.25σ . Also here a measurement of the signal strength was performed providing also a less accurate measurement, as depicted in the right plot in figure 3.30. The fit uncertainty here yields 0.59.

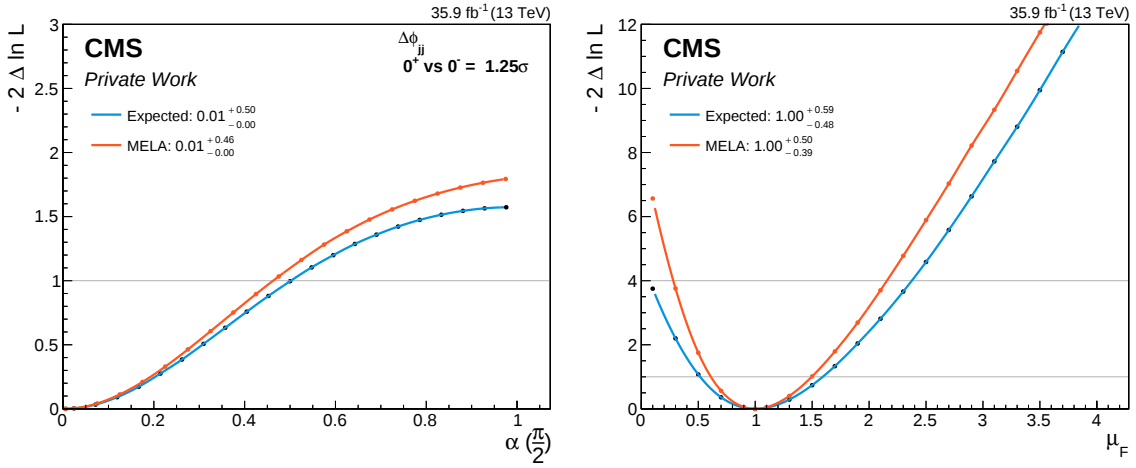


Fig. 3.30.: Measurement of the CP mixing angle α (left) and the signal strength μ_F (right) in the fermionic coupling of the Higgs boson to the top quark inferred from the effective gluon coupling under the assumption of a scalar Higgs boson and SM-like VBF production ($f = 0$). As shown in the left figure, the analysis is able to distinguish a scalar from a pseudoscalar CP scenario using approach utilizing $\Delta\phi_{jj}$ (blue) at 1.25σ . A lower performance is obtained than with the full angular information taken into account by the MELA approach (orange). In the left plot the measurement is performed for $\Delta\phi_{jj}$ showing that it gives the same result but with a looser constraint only. The right plot shows the measurement of the signal strength modifier μ_F and also shows a larger fit uncertainty.

3.4.3.2 Impact of CP violation in VBF and SM rate constraint

To study whether the CP state realized in the VBF background has any impact on the performance of the analysis, α is profiled leaving the VBF CP parameter f floating. The result is given in the left plot in figure 3.31 showing that the performance is not affected by letting the fit find a best-fit value for f as no change in the fit result is found compared to figure 3.28. This implies that the analysis setup is currently not sensitive to the possible higher-order effects from the VBF CP scenario. The reason for this are both that the yield is smaller compared to ggF events in the signal categories and that the effect of a pseudoscalar admixture does not enter at tree level but is suppressed by a loop. As a consequence, fixing $f = 0$ during the measurement of α is a justified approximation.

With the signal model introduced above, the rate of the processes and their shapes, that provides the information about the pseudoscalar admixture, can be altered independently from each other. Therefore, the shape analysis performed in this thesis gives the most unbiased measurement of α . Any effect that also modifies the rate of the ggF process can be due to other physics scenarios such as heavy BSM fermions running in the loop between the gluons and the Higgs boson instead of a top quark and lighter SM fermions. For this reason, a measurement of α constraining the ggF rate to the known SM value of $\mu_F = 1$ comes with the model-dependent assumption that there are no other BSM effects present in the coupling. The impact of fixing $\mu_F = 1$ is examined in the right plot in figure 3.31. The model-dependent measurement of α leads to a significantly better precision of 24.3° . Using the constraint the analysis setup can distinguish between a scalar and a pseudoscalar Higgs boson fermionic coupling at 3.65σ . The reason for this is the increased coupling

constant of the pseudoscalar coupling in equation (3.27) expected by the signal model in this case.

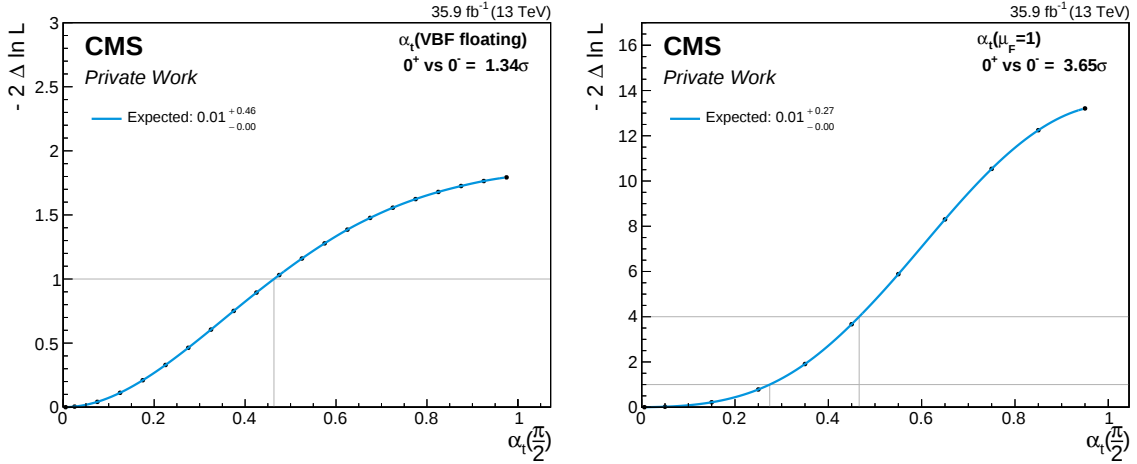


Fig. 3.31.: Measurement of α leaving the VBF CP scenario floating in the fit (left) and using the SM constraint on the ggF signal strength (right). No change in the accuracy of the measurement is observed compared to the nominal measurement from figure 3.28 when f is left floating in the fit. This justifies setting $f = 0$. The rate constraint leads to a better discrimination of scalar and pseudoscalar scenarios due to the expected larger coupling strength of the pseudoscalar coupling.

3.4.3.3 Uncertainty breakdown

The analysis is performed taking both statistical and systematic uncertainties into account. To find out the contribution of both to the final fit uncertainty and uncertainty breakdown is performed. First, the likelihood scan with the setup used for figure 3.28 is performed and the result with the total uncertainty σ_{tot} stored. The likelihood profile is repeated taking into account only statistical uncertainties by fixing the nuisance parameters to the values obtained from the stored global fit from figure 3.28 this time. Thus, the result only incorporates statistical uncertainties σ_{stat} as only those can float in the fit. Using quadratic error summation the systematic uncertainty can be inferred from the the two previous measurements by

$$\sigma_{\text{syst}} = \sqrt{\sigma_{\text{tot}}^2 - \sigma_{\text{stat}}^2}. \quad (3.32)$$

Both fit results and the resulting uncertainty split into a statistical and systematic component are shown in figure 3.32. The total uncertainty of made of a statistical component of 31.5° and a systematic uncertainty of 27° . The measurement of α currently shows a larger statistical component of 0.35 with the integrated luminosity given by the 2016 data set implying that the performance of the measurement is expected to improve when combined with the data collected in 2017 and 2018. This encourages a combination of the data sets at the end of Run 2.

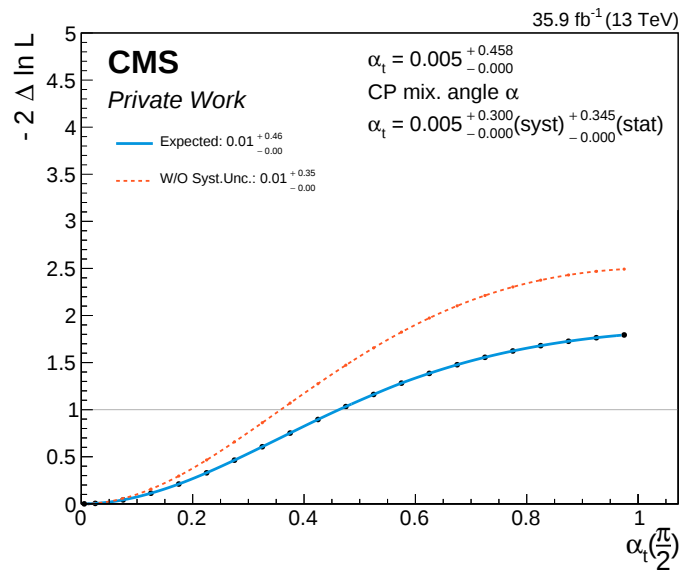


Fig. 3.32.: Final measurement of α stating the systematic and statistical composition of the total uncertainty on α . With the given dataset the uncertainty of the measurement shows a larger statistical component and is thus expected to be improved with more data.

Conclusion and Outlook

The performance of the measurement of the CP mixing angle α in the fermionic coupling of the Higgs boson to the top quark was presented in this thesis exploiting the effective coupling of the Higgs boson to gluons. Therefore, the analysis of the search for the decay of the Higgs boson into a pair of tau leptons [22] was adapted to investigate the gluon-gluon fusion production process of the Higgs boson. A categorization was developed leading to a larger number of selected Higgs bosons produced via gluon-gluon fusion in the signal categories compared to VBF produced Higgs bosons. In the same step, it was taken care to achieve a close-to-optimal signal-to-background ratio. Furthermore, the background estimation was tuned by implementing the data-driven *simultaneous equation method* in the semileptonic decay channels of the tau lepton pair. With the cut-based categorization of $gg \rightarrow H + 2j$ events used in this approach, a measurement, based on a pseudo-data set constructed from the background and the SM Higgs boson prediction, was performed for two sets of CP sensitive observables - D_{0-} , D_{CP} and $\Delta\phi_{jj}$.

For the final fit systematic uncertainties covering biases in the Higgs boson plus jets cross sections were added and groups of jet energy correction uncertainties were formed in addition to the uncertainties taken from Ref. [22]. Using the azimuthal angular difference $\Delta\phi_{jj}$ between the two initial-state jets, α is measured with a precision of 45° . A scalar and a pseudoscalar coupling can be separated at 1.25σ . Exploiting the full kinematic information in a matrix element likelihood approach with optimal observables a better performance is obtained with a precision of 41° . Here, a separation between scalar and pseudoscalar scenarios of 1.36σ is achieved.

During the evaluation of the analysis setup, the impact of a potential CP violating contribution in the vector boson coupling of the Higgs boson was studied. It was proven that CP violation in VBF does not affect the measurement of the fermionic coupling. Moreover, the model-dependent performance of the analysis was investigated. For this, the signal strength of the fermionic coupling was set to the SM prediction $\mu_F = 1$. It is noted that the analysis is then capable to discriminate scalar and pseudoscalar hypotheses by 3.65σ with an accuracy of 24° . This measurement comes with the assumption that no other BSM effects modify the rate of the effective interaction between the gluons and the Higgs boson.

In order to improve the performance quoted here several further step could be executed. The cut-based approach developed for this thesis only provides a close-to-optimal sensitivity. In the latest CMS publications of the observation of the Higgs boson decaying into bottom quarks [23] and in the associated production with top quarks [95] modern machine learning techniques proved to be ideal choices for an optimal event categorization. A similar approach is currently developed for this analysis, too. On the technical side, the background estimation can be further improved implementing new techniques such as the fake-factor method and using embedded $Z \rightarrow \tau\tau$ samples [96]. Last but not least, it has

been shown that the precision of the measurement is limited by statistical fluctuations. Thus, a combination with the full Run 2 data set is expected to improve the performance. The analysis presented in this thesis provides a simple setup that is sensitive to the CP information in fermionic couplings of the Higgs boson produced via gluon-gluon fusion. The approach looks promising as it already now can measure the CP mixing angle to an accuracy of roughly 41° . As soon as, the analysis may be unblinded, it will provide the first measurement of this CP property in data with the CMS experiment.

Bibliography

- [1] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [2] Serguei Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].
- [3] ATLAS and CMS Collaborations. “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, 191803 (2015) (Mar. 26, 2015). DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589v1 [hep-ex].
- [4] ATLAS Collaboration. “Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector”. In: *Eur. Phys. J. C* (2015) 75:335 (Mar. 3, 2015). DOI: 10.1140/epjc/s10052-015-3542-2. arXiv: 1503.01060v4 [hep-ex].
- [5] 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. D* 92, 012004 (2015) (Nov. 13, 2014). DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441v2 [hep-ex].
- [6] 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 (2015) 476 (June 18, 2015). DOI: 10.1140/epjc/s10052-015-3685-1. arXiv: 1506.05669v2 [hep-ex].
- [7] Gerard 't Hooft and M. J. G. Veltman. “Regularization and Renormalization of Gauge Fields”. In: *Nucl. Phys. B* 44 (1972), pp. 189–213. DOI: 10.1016/0550-3213(72)90279-9.
- [8] Y. Fukuda, T. Hayakawa, E. Ichihara, et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.
- [9] Q. R. Ahmad et al. “Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory”. In: *Phys. Rev. Lett.* 87 (2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301. arXiv: nucl-ex/0106015 [nucl-ex].
- [10] Q. R. Ahmad, R. C. Allen, T. C. Andersen, et al. “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”. In: *Phys. Rev. Lett.* 89 (1 2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301.
- [11] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013.

- [12] Hans Volker Klapdor and Klaus Grotz. “Die Glashow-Weinberg-Salam Theorie der elektroschwachen Wechselwirkung”. In: *Die schwache Wechselwirkung in Kern-, Teilchen- und Astrophysik: Eine Einführung*. Wiesbaden: Vieweg+Teubner Verlag, 1989, pp. 196–238. DOI: 10.1007/978-3-322-84833-8_6.
- [13] G. Arnison, A. Astbury, B. Aubert, et al. “Experimental observation of isolated large transverse energy electrons with associated missing energy at $s=540$ GeV”. In: *Physics Letters B* 122.1 (1983), pp. 103–116. DOI: [https://doi.org/10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [14] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.
- [15] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.
- [16] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.
- [17] P. W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Physics Letters* 12 (Sept. 1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.
- [18] Peter W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”. In: *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.
- [19] C. Patrignani et al. “Review of Particle Physics”. In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.
- [20] J. D. Bjorken. “Asymptotic Sum Rules at Infinite Momentum”. In: *Phys. Rev.* 179 (5 1969), pp. 1547–1553. DOI: 10.1103/PhysRev.179.1547.
- [21] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka (Eds.) “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties”. In: *CERN-2013-004* (CERN, Geneva, 2013). DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].
- [22] Albert M Sirunyan et al. “Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector”. In: *Phys. Lett. B* 779 (2018), pp. 283–316. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex].
- [23] CMS Collaboration. “Observation of Higgs boson decay to bottom quarks”. In: (Aug. 24, 2018). arXiv: 1808.08242v1 [hep-ex].
- [24] Georges Aad et al. “Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector”. In: *Phys. Rev. D* 92.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.92.012006. arXiv: 1412.2641 [hep-ex].
- [25] E. Noether. “Invarianten beliebiger Differentialausdrücke.” German. In: *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* 1918 (1918), pp. 37–44.
- [26] Emmy Noether. “Invariante Variationsprobleme.” German. In: *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* 1918 (1918), pp. 235–257.
- [27] Gerhart Lüders. “On the equivalence of invariance under time reversal and under particle-antiparticle conjugation for relativistic field theories”. In: (Jan. 1954).
- [28] Wolfgang Pauli. “Exclusion Principle, Lorentz Group and Reflection of Space-Time and Charge”. In: *Wolfgang Pauli: Das Gewissen der Physik*. Ed. by Charles P. Enz and Karl v. Meyenn. Wiesbaden: Vieweg+Teubner Verlag, 1988, pp. 459–479. DOI: 10.1007/978-3-322-90270-2_41.
- [29] Paolo Nason. *Lecture notes in Computer Assisted Diagnosis*. 2011.

- [30] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.
- [31] M. Goldhaber, L. Grodzins, and A. W. Sunyar. “Helicity of Neutrinos”. In: *Phys. Rev.* 109 (3 1958), pp. 1015–1017. DOI: 10.1103/PhysRev.109.1015.
- [32] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. “Evidence for the 2π Decay of the K^0_2 Meson”. In: *Phys. Rev. Lett.* 13 (1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138.
- [33] H Blümer. “Measurement of direct CP-violation with the NA48 experiment at the CERN SPS”. In: (1999).
- [34] B Gorini. “New result on direct CP violation from NA48”. In: (2001).
- [35] B. Aubert, D. Boutigny, J.-M. Gaillard, et al. “Observation of CP Violation in the B^0 Meson System”. In: *Phys. Rev. Lett.* 87 (9 2001), p. 091801. DOI: 10.1103/PhysRevLett.87.091801.
- [36] Bernard Aubert et al. “Measurement of branching ratios and CP asymmetries in $B^- \rightarrow D_{CP}^0 K^-$ decays”. In: *High energy physics. Proceedings, 31st International Conference, ICHEP 2002, Amsterdam, Netherlands, July 25-31, 2002*. 2002. arXiv: hep-ex/0207087 [hep-ex].
- [37] S. K. Swain, T. E. Browder, K. Abe, et al. “Measurement of branching fraction ratios and CP asymmetries in $B^\pm \rightarrow D_{CP} K^\pm$ ”. In: *Phys. Rev. D* 68 (5 2003), p. 051101. DOI: 10.1103/PhysRevD.68.051101.
- [38] B. Aubert, R. Barate, D. Boutigny, et al. “Measurements of Branching Fractions and CP -Violating Asymmetries in $B^0 \rightarrow \rho^\pm h^\mp$ Decays”. In: *Phys. Rev. Lett.* 91 (20 2003), p. 201802. DOI: 10.1103/PhysRevLett.91.201802.
- [39] B. Aubert, D. Boutigny, J.-M. Gaillard, et al. “Measurements of Branching Fractions and CP -Violating Asymmetries in $B^0 \rightarrow \pi^+ \pi^-, K^+ \pi^-, K^+ K^-$ Decays”. In: *Phys. Rev. Lett.* 89 (28 2002), p. 281802. DOI: 10.1103/PhysRevLett.89.281802.
- [40] K. Abe et al. “Combined Analysis of Neutrino and Antineutrino Oscillations at T2K”. In: *Phys. Rev. Lett.* 118.15 (2017), p. 151801. DOI: 10.1103/PhysRevLett.118.151801. arXiv: 1701.00432 [hep-ex].
- [41] Gustavo C. Branco, Luis Lavoura, and Joao P. Silva. “CP Violation”. In: *Int. Ser. Monogr. Phys.* 103 (1999), pp. 1–536.
- [42] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967). [Usp. Fiz. Nauk161,no.5,61(1991)], pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.
- [43] C. Jarlskog. “Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation”. In: *Phys. Rev. Lett.* 55 (10 1985), pp. 1039–1042. DOI: 10.1103/PhysRevLett.55.1039.
- [44] P.and Demartin F.and Frederix R.and Frixione S.and Maltoni F.and Mandal M. K.and Mathews P.and Mawatari K.and Ravindran V.and Seth S.and Torrielli P.and Zaro M. Artoisenet P.and de Aquino. “A framework for Higgs characterisation”. In: *Journal of High Energy Physics* 2013.11 (2013), p. 43. DOI: 10.1007/JHEP11(2013)043.
- [45] CMS Collaboration. *CP measurements in VBF Htautau*. CMS Analysis Note CMS-HIG-17-034. 2017.

- [46] Claudia Pistone. “Study of the Higgs boson CP properties in the tautau decay channel with the CMS experiment”. PhD thesis. RWTH Aachen University, 2018.
- [47] V. Del Duca, W. Kilgore, C. Oleari, C. Schmidt, and D. Zeppenfeld. “Higgs + 2 jets via gluon fusion”. In: *Phys. Rev. Lett.* 87 (2001), p. 122001. DOI: 10.1103/PhysRevLett.87.122001. arXiv: hep-ph/0105129 [hep-ph].
- [48] Federico Demartin, Fabio Maltoni, Kentarou Mawatari, Ben Page, and Marco Zaro. “Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction”. In: *Eur. Phys. J. C* 74.9 (2014), p. 3065. DOI: 10.1140/epjc/s10052-014-3065-2. arXiv: 1407.5089 [hep-ph].
- [49] Ian Anderson, Sara Bolognesi, Fabrizio Caola, et al. “Constraining anomalous HVV interactions at proton and lepton colliders”. In: *Phys. Rev. D* 89, 035007 (2014) (Sept. 18, 2013). DOI: 10.1103/PhysRevD.89.035007. arXiv: 1309.4819v3 [hep-ph].
- [50] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (2008), S08001.
- [51] *CERN Accelerators*. <http://ps-div.web.cern.ch/ps-div/PSComplex/accelerators.pdf>. Accessed: 2018-08-02.
- [52] *LHC layout. Schema general du LHC*. <https://cds.cern.ch/record/841573>. Accessed: 2018-08-02.
- [53] *The CERN Accelerator complex*. <https://cds.cern.ch/record/2197559>. Accessed: 2018-08-02.
- [54] *CERN Open Days 2013, Point 5 - CMS: CMS Experiment*. <https://cds.cern.ch/record/1603466?ln=de>. Accessed: 2018-08-02.
- [55] *Side A of the ATLAS cavern*. <https://cds.cern.ch/record/1973064/>. Accessed: 2018-08-02.
- [56] *Point 2: ALICE Experiment*. <https://cds.cern.ch/record/1551026/>. Accessed: 2018-08-02.
- [57] *View of the LHCb cavern on 20 May 2009*. <http://cdsweb.cern.ch/record/1178706>. Accessed: 2018-08-02.
- [58] Albert de Roeck, U Gasparini, D Marlow, et al. *CMS Physics Technical Design Report Volume I: Detector Performance and Software*. Ed. by D Acosta. Technical Design Report CMS. Geneva: CERN, 2006.
- [59] S Chatrchyan, G Hmayakyan, A M Khachatryan V and Sirunyan, et al. “The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment”. In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08004. 361 p.
- [60] *CMS Experiment*. <http://cds.cern.ch/record>. Accessed: 2018-08-02.
- [61] S Chatrchyan, V Khachatryan, W Sirunyan A M and Adam, et al. “Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays”. In: *JINST* 5. arXiv:0910.2505 (2009), T03009. 41 p.
- [62] The CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (2014), P10009.
- [63] A.M. Sirunyan, A. Tumasyan, W. Adam, et al. “Particle-flow reconstruction and global event description with the CMS detector”. In: *Journal of Instrumentation* 12.10 (2017), P10003.
- [64] CMS Collaboration. *Baseline Muon Selections for Run-II*. Accessed: 2018-09-05.

- [65] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- k_t jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (2008), p. 063.
- [66] CMS Collaboration. *Performance of the Particle-Flow jet identification criteria using proton-proton collisions at 13 TeV for Run2016 data*. CMS Analysis Note CMS-HIG-17-074. 2017.
- [67] CMS Collaboration. “Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS”. In: *JINST 11 (2016) P01019* (Oct. 26, 2015). DOI: 10.1088/1748-0221/11/01/P01019. arXiv: 1510.07488v2 [physics.ins-det].
- [68] “Performance of tau identification with 2016 data at $\sqrt{s} = 13$ TeV”. In: (2017).
- [69] “Tau-Id performance with full 2016 dataset using $Z \rightarrow \tau_\mu \tau_h$ events”. In: (2017).
- [70] Alexander Rafael Nehrkorn. “Search for the standard model Higgs boson decaying into pairs of tau leptons with the CMS Experiment”. Veröffentlicht auf dem Publikationsserver der RWTH Aachen University; Dissertation, RWTH Aachen University, 2017. Dissertation. Aachen: RWTH Aachen University, 2017, 1 Online-Ressource (xi, 132 Seiten) : Illustrationen, Diagramme. DOI: 10.18154/RWTH-2017-09574.
- [71] The CMS collaboration. “Performance of the CMS missing transverse momentum reconstruction in pp data at 8TeV”. In: *Journal of Instrumentation* 10.02 (2015), P02006.
- [72] V. Hankele, G. Klamke, D. Zeppenfeld, and T. Figy. “Anomalous Higgs boson couplings in vector boson fusion at the CERN LHC”. In: *Phys. Rev. D* 74 (2006), p. 095001. DOI: 10.1103/PhysRevD.74.095001. arXiv: hep-ph/0609075 [hep-ph].
- [73] Tilman Plehn, David L. Rainwater, and Dieter Zeppenfeld. “Determining the Structure of Higgs Couplings at the LHC”. In: *Phys. Rev. Lett.* 88 (2002), p. 051801. DOI: 10.1103/PhysRevLett.88.051801. arXiv: hep-ph/0105325 [hep-ph].
- [74] Matthew J. Dolan, Philip Harris, Martin Jankowiak, and Michael Spannowsky. “Constraining CP-violating Higgs Sectors at the LHC using gluon fusion”. In: *Phys. Rev. D* 90, 073008 (2014) (June 12, 2014). DOI: 10.1103/PhysRevD.90.073008. arXiv: 1406.3322v1 [hep-ph].
- [75] Ian Anderson, Sara Bolognesi, Fabrizio Caola, et al. “Constraining anomalous HVV interactions at proton and lepton colliders”. In: *Phys. Rev. D* 89, 035007 (2014) (Sept. 18, 2013). DOI: 10.1103/PhysRevD.89.035007. arXiv: 1309.4819v3 [hep-ph].
- [76] Yanyan Gao, Andrei V. Gritsan, Zijin Guo, et al. “Spin determination of single-produced resonances at hadron colliders”. In: *Phys.Rev.D* 81:075022,2010 (Jan. 19, 2010). DOI: 10.1103/PhysRevD.81.075022. arXiv: 1001.3396v2 [hep-ph].
- [77] Sara Bolognesi, Yanyan Gao, Andrei V. Gritsan, et al. “On the spin and parity of a single-produced resonance at the LHC”. In: *Phys.Rev.D* 86:095031,2012 (Aug. 20, 2012). DOI: 10.1103/PhysRevD.86.095031. arXiv: 1208.4018v3 [hep-ph].
- [78] Andrei V. Gritsan, Raoul Roentsch, Markus Schulze, and Meng Xiao. “Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques”. In: *Phys. Rev. D* 94, 055023 (2016) (June 9, 2016). DOI: 10.1103/PhysRevD.94.055023. arXiv: 1606.03107v3 [hep-ph].
- [79] The CMS collaboration, S. Chatrchyan, V. Khachatryan, et al. “Search for a Higgs boson in the decay channel $H \rightarrow ZZ(*) \rightarrow q\bar{q} + \text{in pp collisions at } \sqrt{s} = 7 \text{ TeV}$ ”. In: *Journal of High Energy Physics* 2012.4 (2012), p. 36. DOI: 10.1007/JHEP04(2012)036.

- [80] L. Bianchini, B. Calpas, J. Conway, et al. “Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of tau leptons using matrix element techniques”. In: (Mar. 18, 2016). DOI: 10.1016/j.nima.2017.05.001. arXiv: 1603.05910v3 [hep-ex].
- [81] and. “IX. On the problem of the most efficient tests of statistical hypotheses”. In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009. eprint: <http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf>.
- [82] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput.Phys.Commun.* 178:852-867,2008 (Oct. 20, 2007). DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820v1 [hep-ph].
- [83] CMS Collaboration. “Search for additional neutral MSSM Higgs bosons in the final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: (Mar. 17, 2018). arXiv: 1803.06553v1 [hep-ex].
- [84] Daniel Winterbottom. Private communication. May 17, 2018.
- [85] Daniel Winterbottom. Private communication. Sept. 4, 2018.
- [86] Roger Barlow and Christine Beeston. “Fitting using finite Monte Carlo samples”. In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. DOI: [https://doi.org/10.1016/0010-4655\(93\)90005-W](https://doi.org/10.1016/0010-4655(93)90005-W).
- [87] J. S. Conway. “Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra”. In: (Mar. 2, 2011). arXiv: 1103.0354v1 [physics.data-an].
- [88] Iain W. Stewart and Frank J. Tackmann. “Theory uncertainties for Higgs mass and other searches using jet bins”. In: *Phys. Rev. D* 85 (3 2012), p. 034011. DOI: 10.1103/PhysRevD.85.034011.
- [89] Vardan Khachatryan et al. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].
- [90] CMS Collaboration. *JEC Uncertainty TWIKI Page*. Accessed: 2018-09-05.
- [91] Daniel Winterbottom. *RegionalJES*. Status report, HTauTau meeting, CERN. 2018.
- [92] CMS Collaboration. *TWIKI - Jet energy scale uncertainty sources*. Sept. 25, 2018.
- [93] Lorenzo Moneta, Kevin Belasco, Kyle Cranmer, et al. “The RooStats Project”. In: *PoS ACAT2010*. arXiv:1009.1003 (2010). Comments: 11 pages, 3 figures, ACAT2010 Conference Proceedings, p. 057.
- [94] Wouter Verkerke and David P. Kirkby. “The RooFit toolkit for data modeling”. In: *eConf C0303241* (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].
- [95] Albert M Sirunyan, Armen Tumasyan, Federico Adam Wolfgang Ambrogio, et al. “Observation of $t\bar{t}H$ production”. In: *Phys. Rev. Lett.* 120.CMS-HIG-17-035. CMS-HIG-17-035-003 (2018). Replaced with the published version. Added the journal reference and the DOI. All the figures and tables, including additional supplementary figures and tables, can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-17-035> (CMS Public Pages), 231801. 17 p.
- [96] Janek Bechtel. “Cross-check of the CMS search for additional MSSM Higgs bosons in the di- τ final state using $\mu \rightarrow \tau$ embedded events”. Karlsruhe Institute of Technology, Masterarbeit, 2016. MS. Karlsruhe Institute of Technology, 2017.

Search for CP violation in gluon-gluon fusion events

A.1 Measuring CP violation in gluon-gluon fusion events

A.1.1 Categorization

A.1.1.1 0-jet and boosted categories

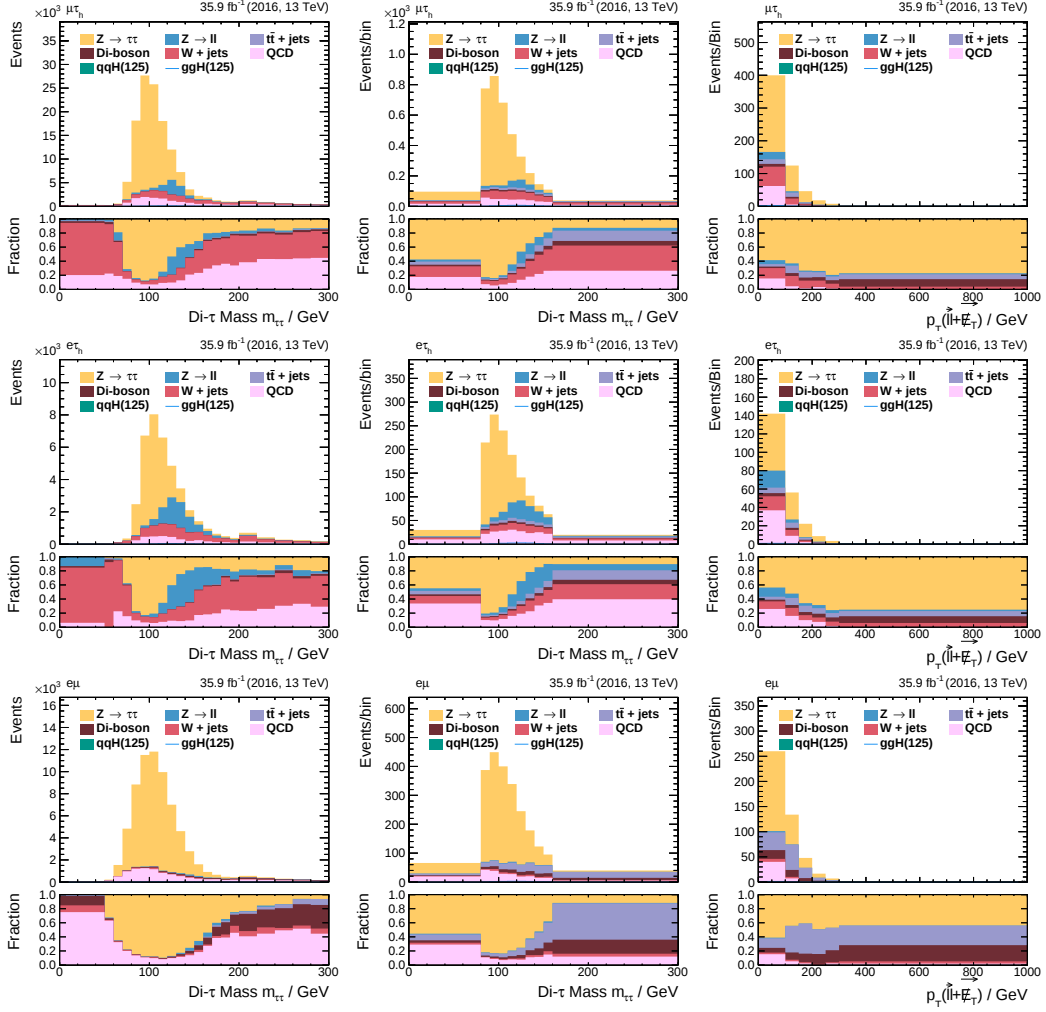


Fig. A.1.: Background and expected qqH and ggH signal distributions of observables in the 0-jet category (left) and *boosted* category (middle, right) for the $\mu\tau_h$ (upper row), $e\tau_h$ (middle row) and $e\mu$ (bottom row) channels. The subplot shows the relative fraction of each background in each bin.

A.1.1.2 Dijet signal categories

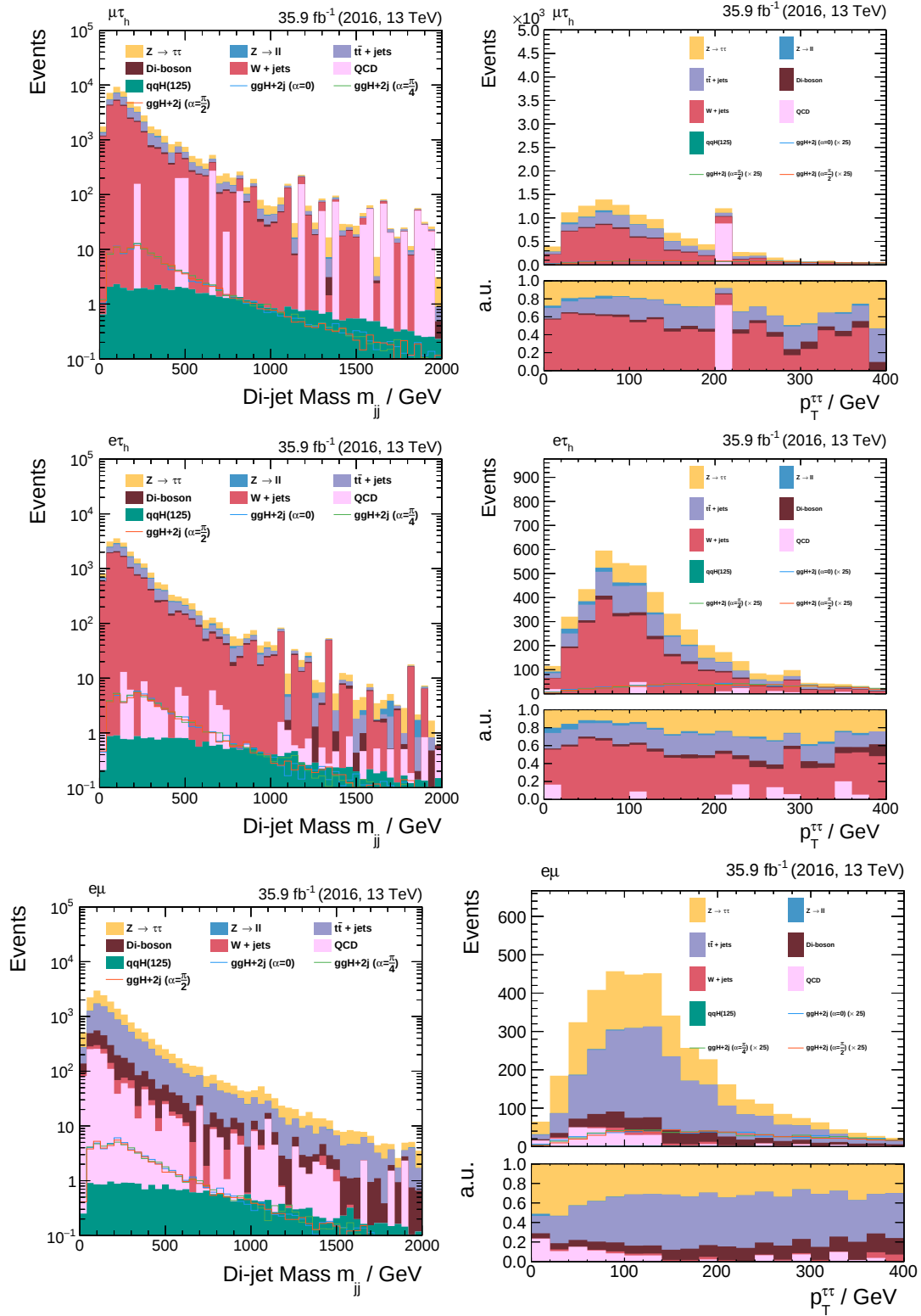


Fig. A.2.: Distributions of the invariant dijet mass m_{jj} (left) and the reconstructed Higgs transverse momentum $p_T^{\tau\tau}$ (middle) for the $\mu\tau_h$ (upper row), $e\tau_h$ (middle row) and $e\mu$ (bottom row) channels are shown for events with 2 jets or more.

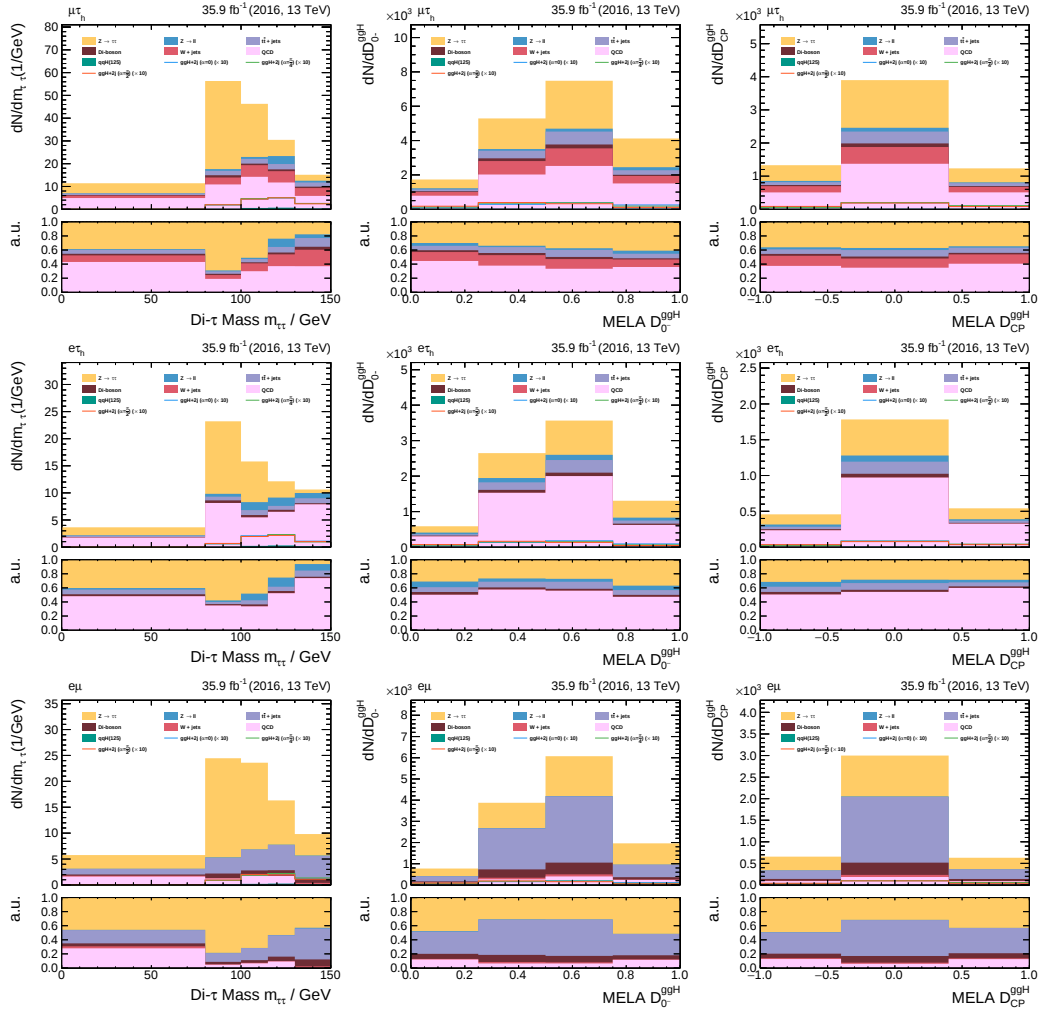


Fig. A.3.: Background and ggH signal distributions for three scalar, maxmix and pseudoscalar CP scenarios of the observables in the 2-jet lowboost category for the $\mu\tau_h$ (upper row), $e\tau_h$ (middle row) and $e\mu$ (bottom row) channels. All histograms are normalized by the bin width and the ggH signal scaled by a factor 10.

A.2 Background estimation

A.2.1 Measurement of the QCD OS/SS factor

A.2.1.1 $e\tau_h$ channel

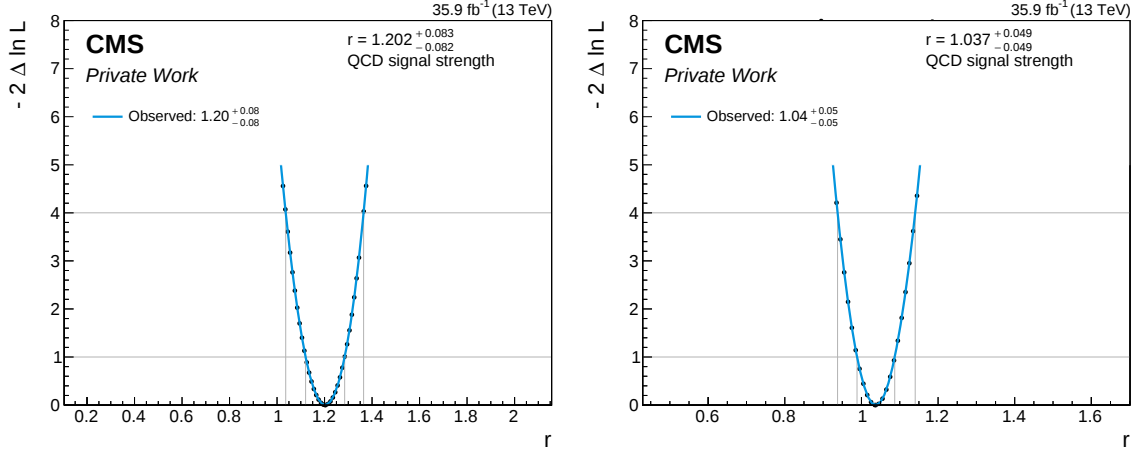


Fig. A.4.: Results of the maximum likelihood scan to find the signal strength of the QCD multijet background in the $e\tau_h$ channel in the *boosted* category. The signal strength parameter from the near sideband (left figure) $r_{\text{QCD}}^{\text{near}} = 1.20 \pm 0.08$ is taken as the QCD OS/SS factor. As statistical uncertainty the larger value of the two given thresholds is chosen. $r_{\text{QCD}}^{\text{near}}$ is compared to the value $r_{\text{QCD}}^{\text{far}} = 1.04 \pm 0.05$ obtained in the *far* sideband (right plot). As they agree within their uncertainties the statistical uncertainties of the far sideband is assigned as systematic extrapolation uncertainty.

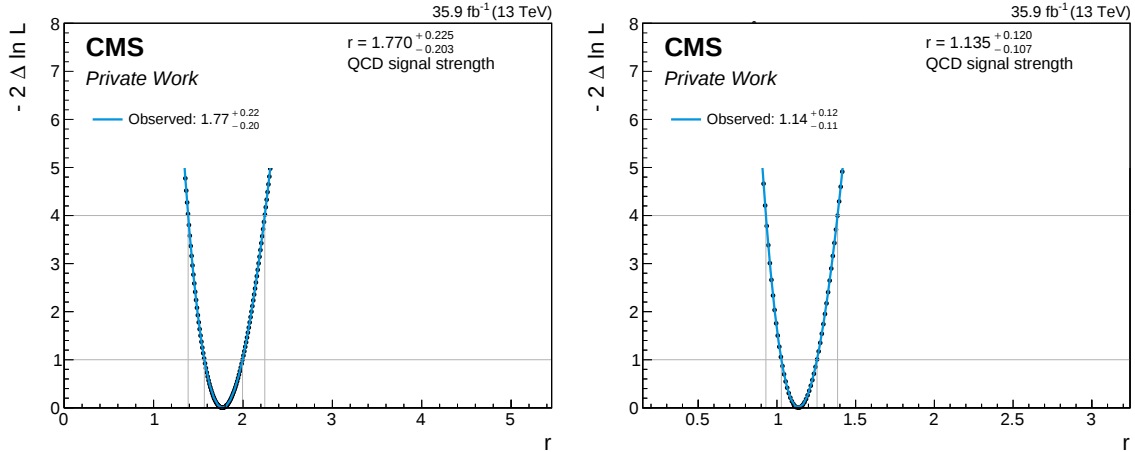


Fig. A.5.: Results of the maximum likelihood scan to find the signal strength of the QCD multijet background in the $e\tau_h$ channel in the *boosted* category. The signal strength parameter from the near sideband (left figure) $r_{\text{QCD}}^{\text{near}} = 1.77 \pm 0.23$ is taken as the QCD OS/SS factor. As statistical uncertainty the larger value of the two given thresholds is chosen. $r_{\text{QCD}}^{\text{near}}$ is compared to the value $r_{\text{QCD}}^{\text{far}} = 1.13 \pm 0.12$ obtained in the *far* sideband (right plot). As they agree within their uncertainties the statistical uncertainties of the far sideband is assigned as systematic extrapolation uncertainty.

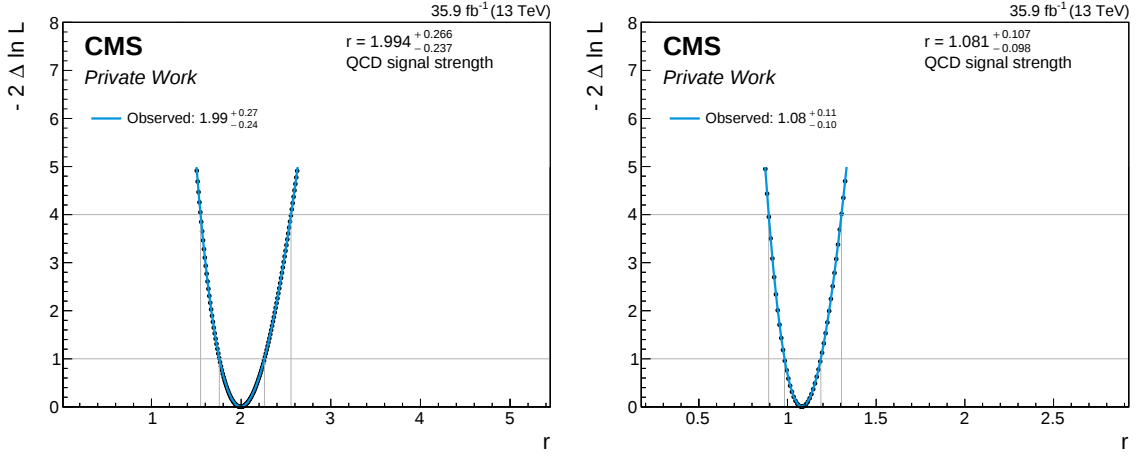


Fig. A.6.: Results of the maximum likelihood scan to find the signal strength of the QCD multijet background in the $e\tau_h$ channel in the *boosted* category. The signal strength parameter from the near sideband (left figure) $r_{\text{QCD}}^{\text{near}} = 2.00 \pm 0.27$ is taken as the QCD OS/SS factor. As statistical uncertainty the larger value of the two given thresholds is chosen. $r_{\text{QCD}}^{\text{near}}$ is compared to the value $r_{\text{QCD}}^{\text{far}} = 1.08 \pm 0.1$ obtained in the *far* sideband (right plot). As they agree within their uncertainties the statistical uncertainties of the far sideband is assigned as systematic extrapolation uncertainty.

A.2.1.2 $\mu\tau_h$ channel

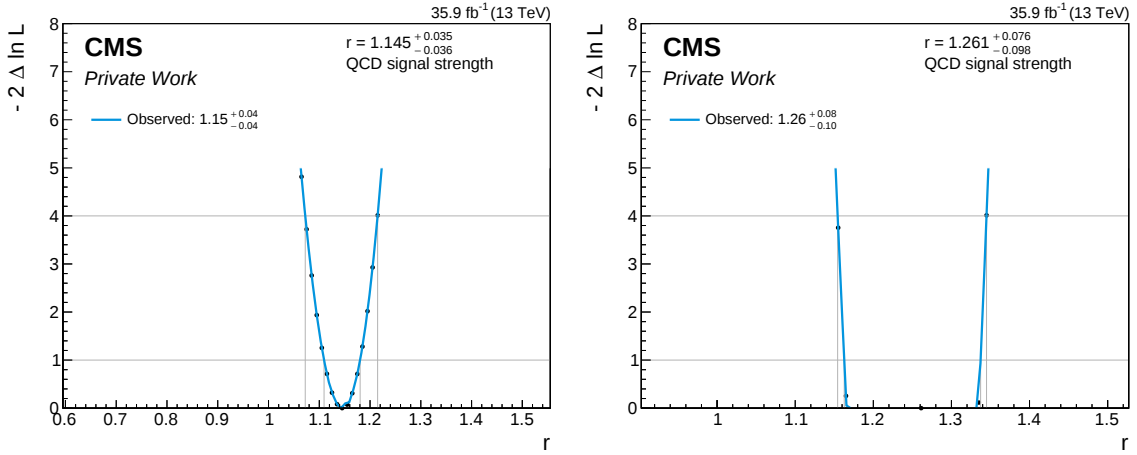


Fig. A.7.: Results of the maximum likelihood scan to find the signal strength of the QCD multijet background in the $\mu\tau_h$ channel in the *boosted* category. The signal strength parameter from the near sideband (left figure) $r_{\text{QCD}}^{\text{near}} = 1.19 \pm 0.06$ is taken as the QCD OS/SS factor. As statistical uncertainty the larger value of the two given thresholds is chosen. $r_{\text{QCD}}^{\text{near}}$ is compared to the value $r_{\text{QCD}}^{\text{far}} = 1.26 \pm 0.10$ obtained in the *far* sideband (right plot). As they agree within their uncertainties the statistical uncertainties of the far sideband is assigned as systematic extrapolation uncertainty.

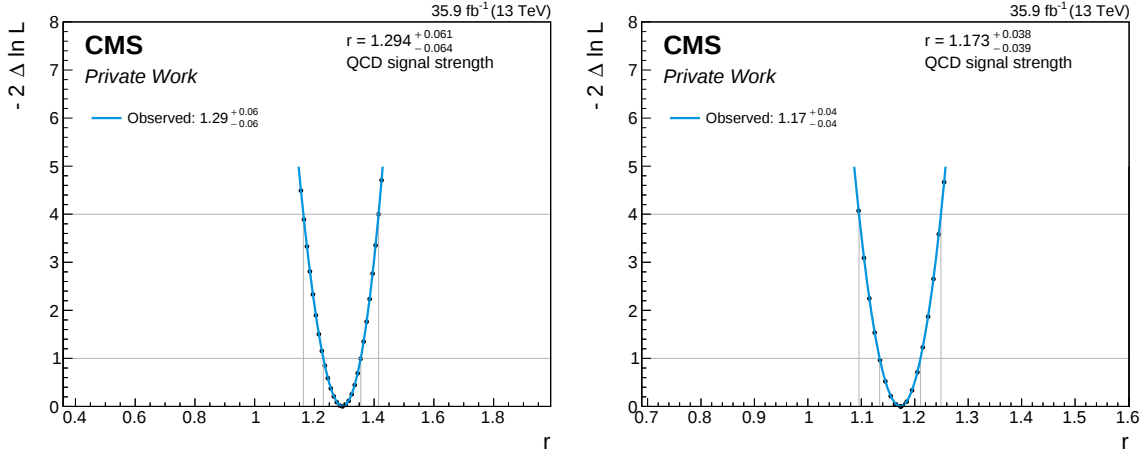


Fig. A.8.: Results of the maximum likelihood scan to find the signal strength of the QCD multijet background in the $\mu\tau_h$ channel in the *boosted* category. The signal strength parameter from the near sideband (left figure) $r_{\text{QCD}}^{\text{near}} = 1.29 \pm 0.6$ is taken as the QCD OS/SS factor. As statistical uncertainty the larger value of the two given thresholds is chosen. $r_{\text{QCD}}^{\text{near}}$ is compared to the value $r_{\text{QCD}}^{\text{far}} = 1.17 \pm 0.04$ obtained in the *far* sideband (right plot). As they agree within their uncertainties the statistical uncertainties of the far sideband is assigned as systematic extrapolation uncertainty.

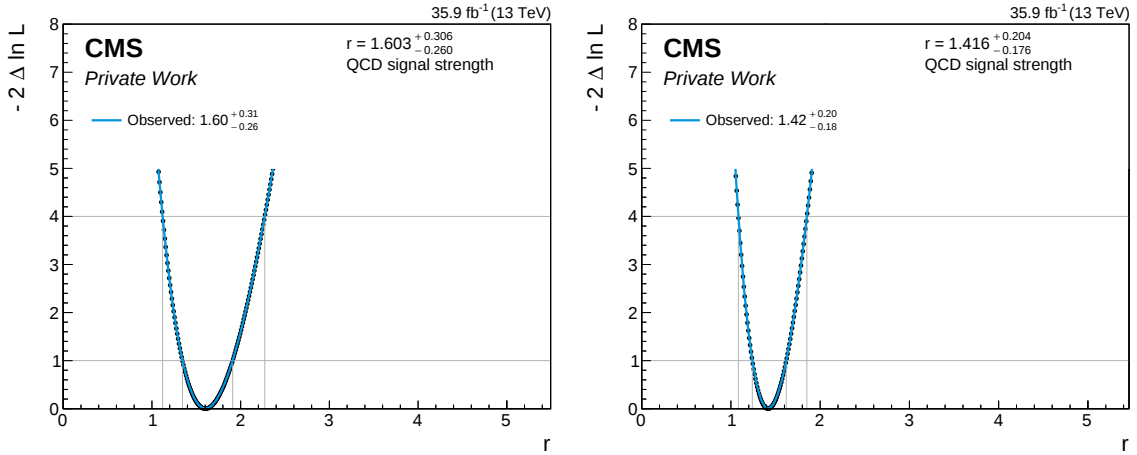


Fig. A.9.: Results of the maximum likelihood scan to find the signal strength of the QCD multijet background in the $\mu\tau_h$ channel in the *boosted* category. The signal strength parameter from the near sideband (left figure) $r_{\text{QCD}}^{\text{near}} = 1.60 \pm 0.3$ is taken as the QCD OS/SS factor. As statistical uncertainty the larger value of the two given thresholds is chosen. $r_{\text{QCD}}^{\text{near}}$ is compared to the value $r_{\text{QCD}}^{\text{far}} = 1.42 \pm 0.2$ obtained in the *far* sideband (right plot). As they agree within their uncertainties the statistical uncertainties of the far sideband is assigned as systematic extrapolation uncertainty.

A.2.2 QCD control regions

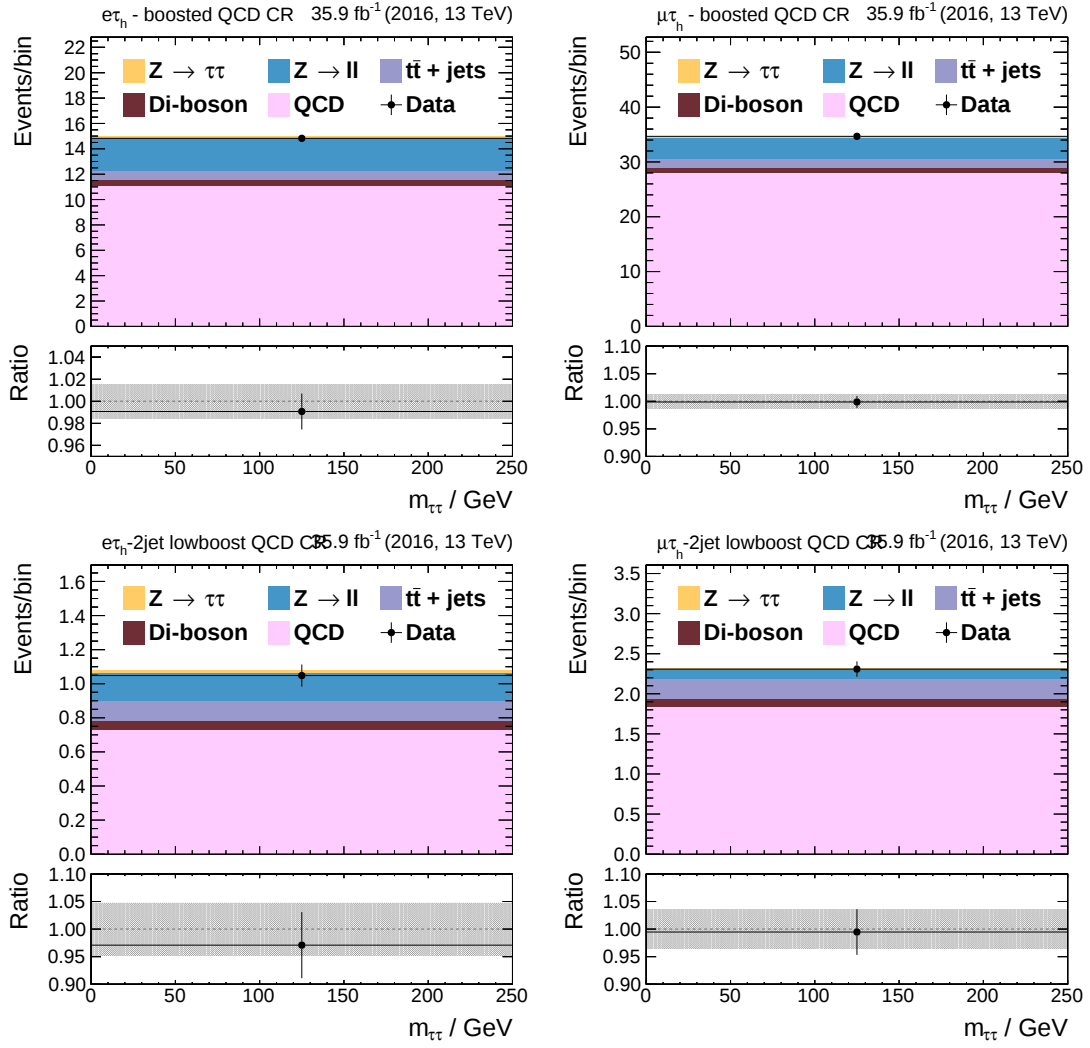


Fig. A.10.: Background modeling and data in the QCD control region in the *boosted* category (upper row) and the *dijet lowboost* category (bottom row) for the $e\tau_h$ (left) and $\mu\tau_h$ (right) channel.

A.3 Performance of the measurement of the gluon-gluon fusion CP mixing angle

A.3.1 Prefit distributions for MELA approach

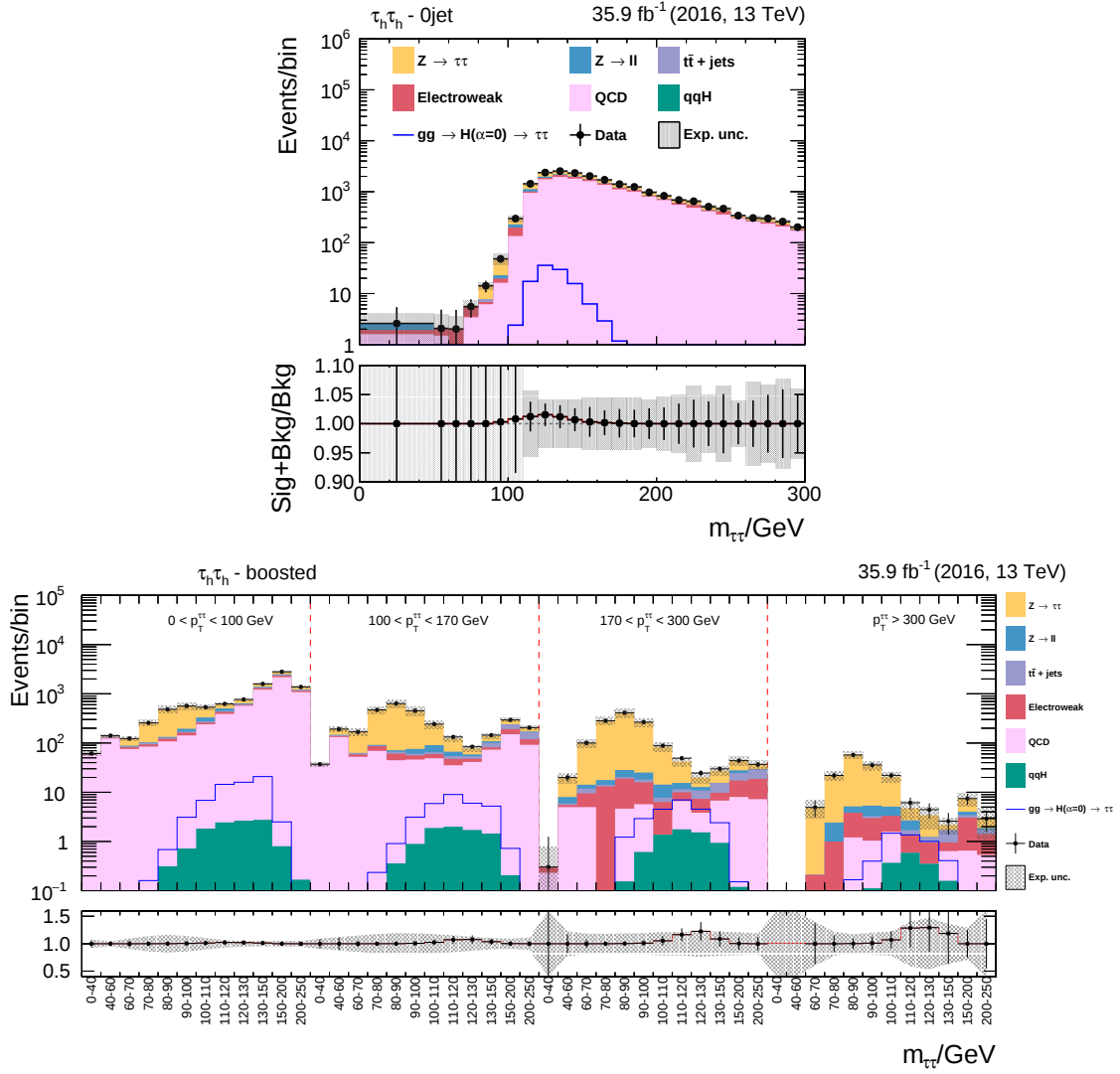


Fig. A.11.: Prefit distributions of the 0-jet and boosted categories in the $\tau_h \tau_h$ channel using the MELA approach.

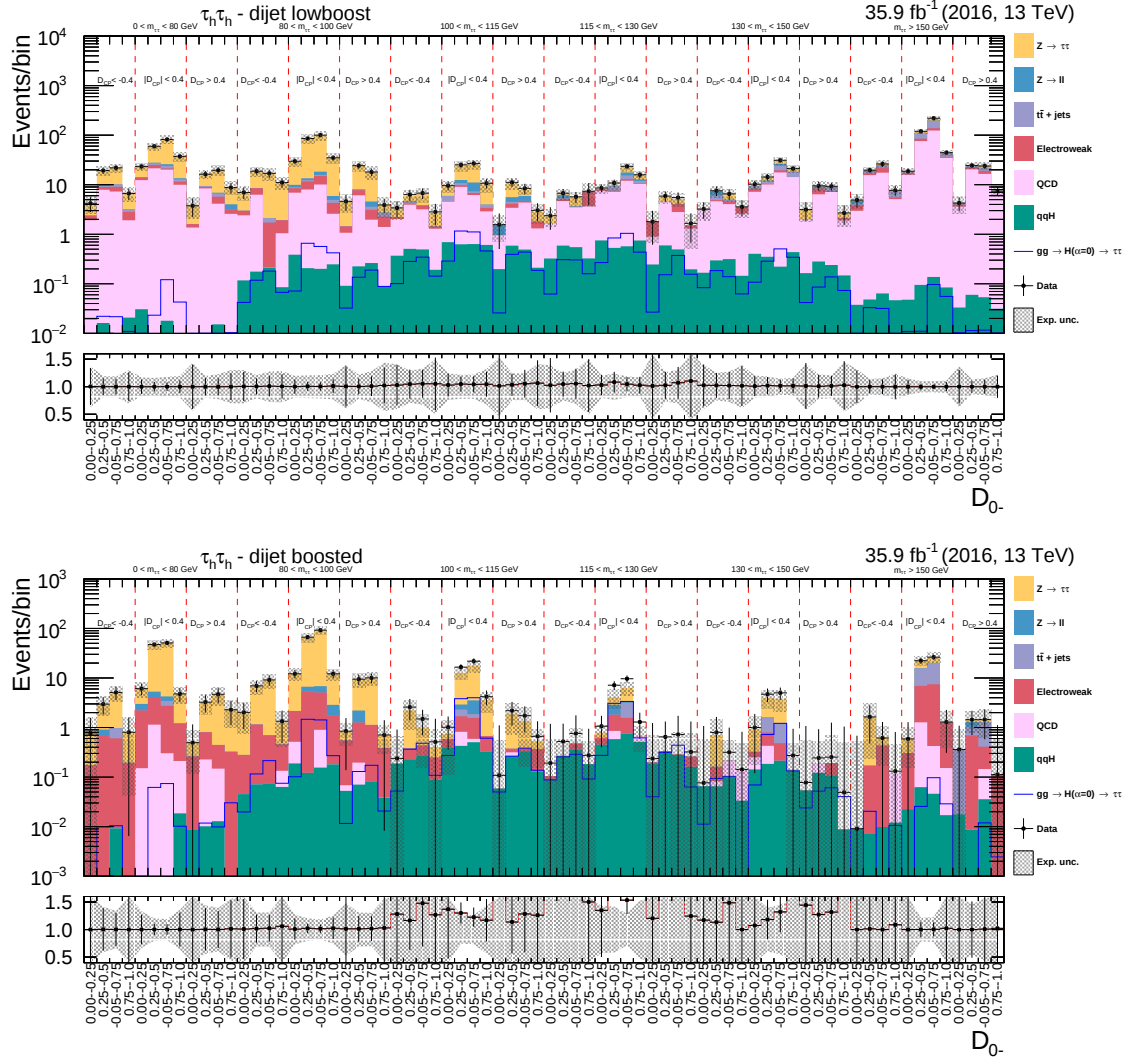


Fig. A.12.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $\tau_h \tau_h$ channel using the MELA approach.

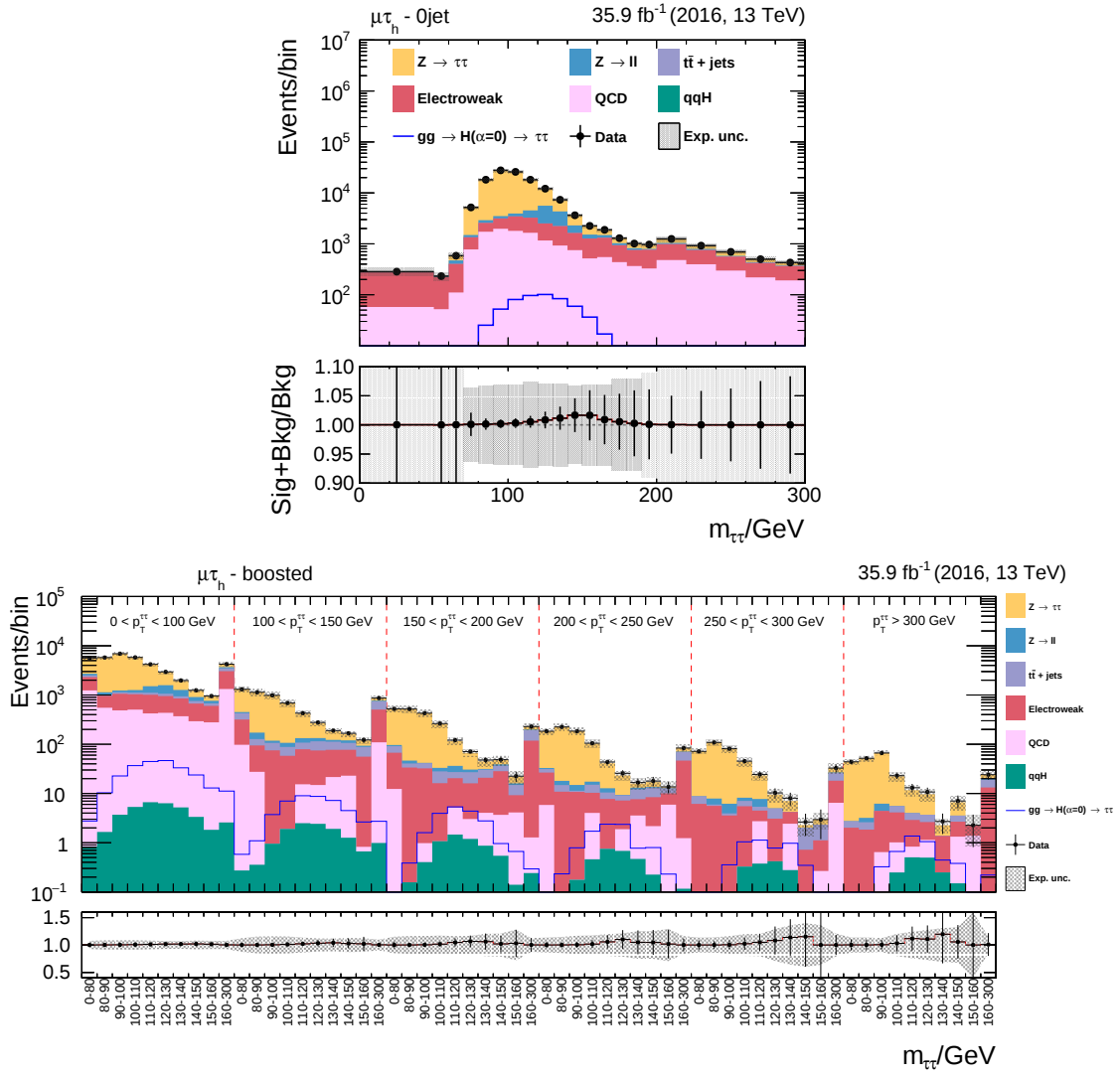


Fig. A.13.: Prefit distributions of the *0-jet* and *boosted* categories in the $\mu\tau_h$ channel using the MELA approach.

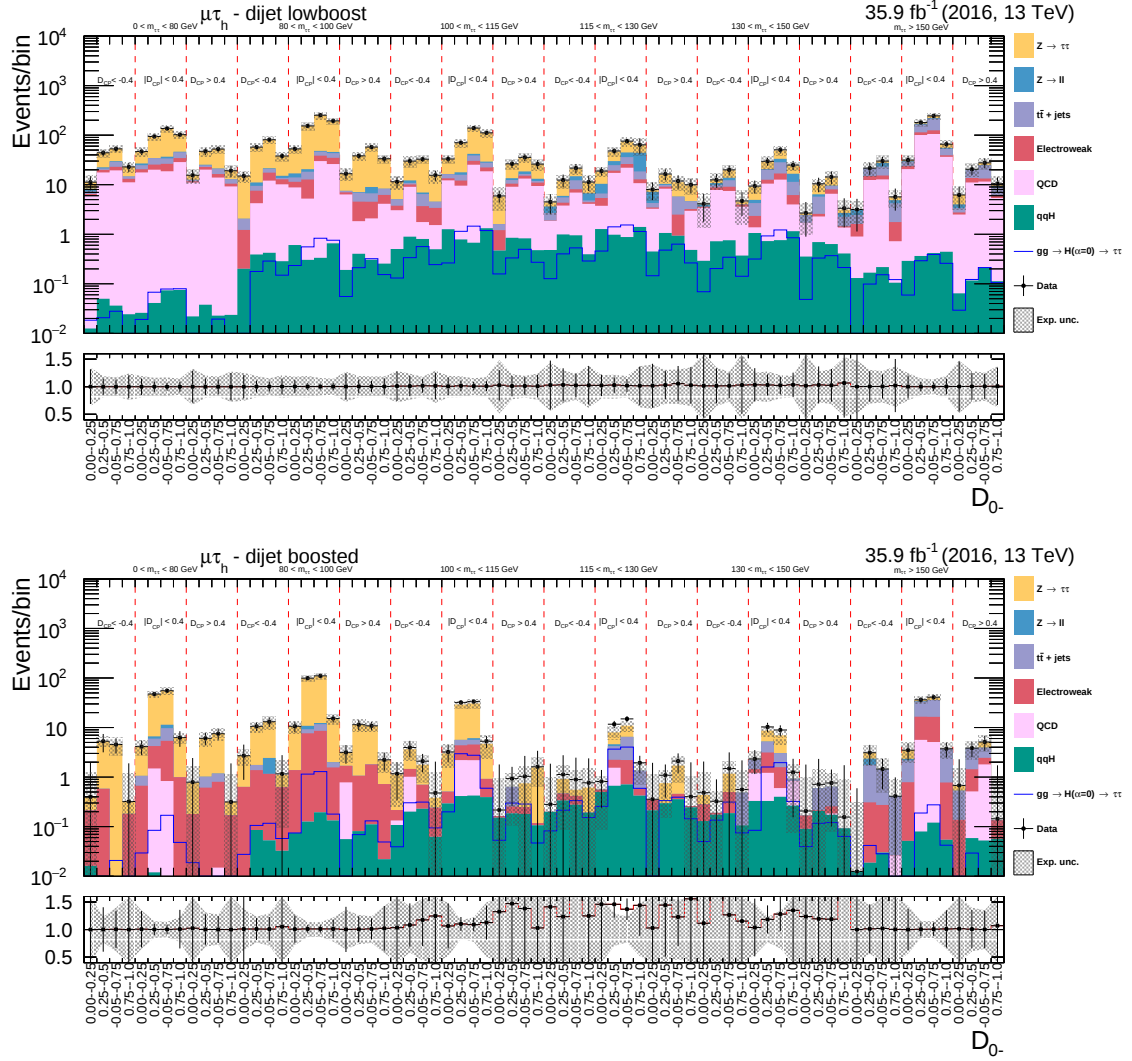


Fig. A.14.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $\mu\tau_h$ channel using the MELA approach.

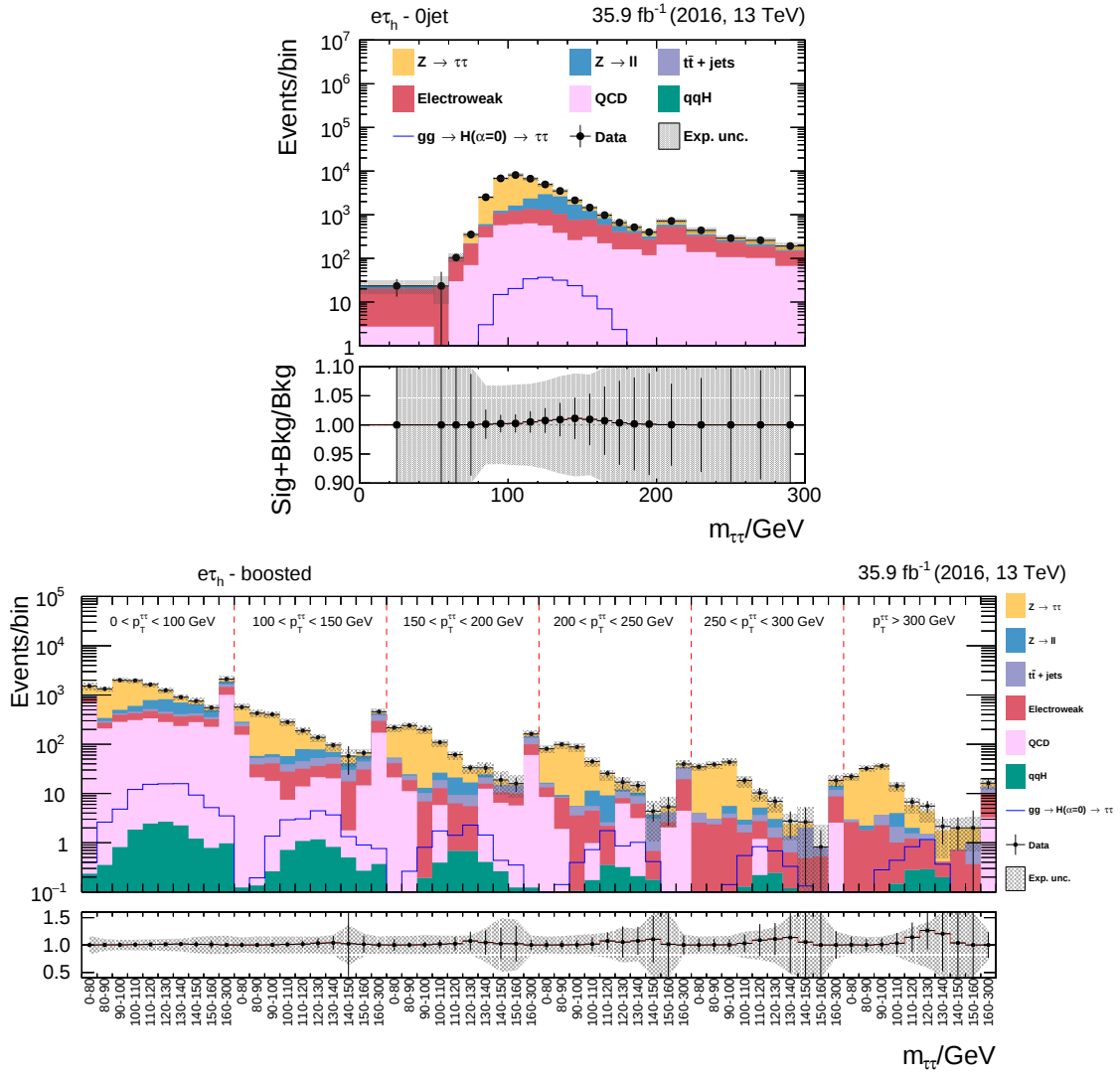


Fig. A.15.: Prefit distributions of the *0-jet* and *boosted* categories in the $e\tau_h$ channel using the MELA approach.

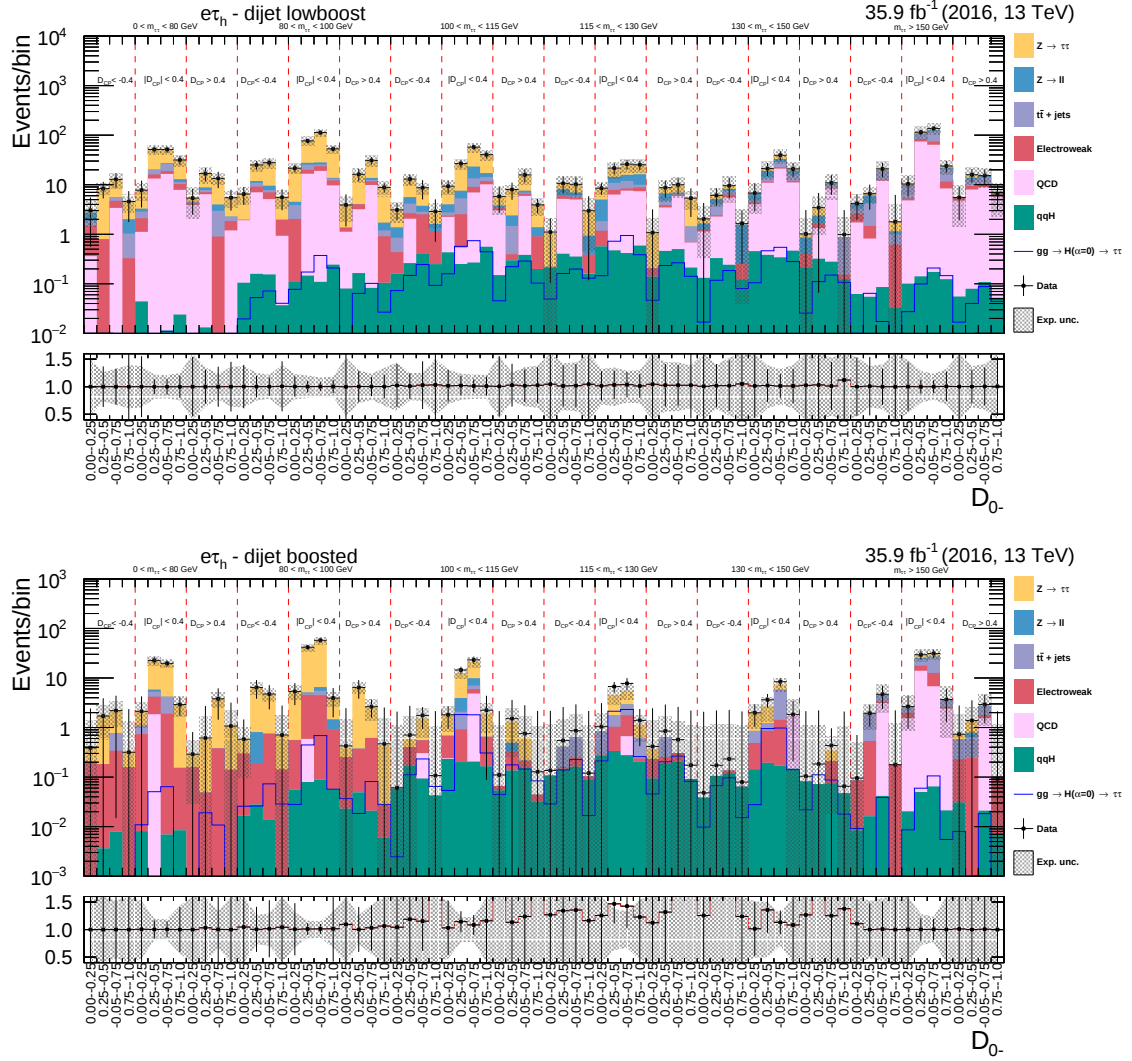


Fig. A.16.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\tau_h$ channel using the MELA approach.

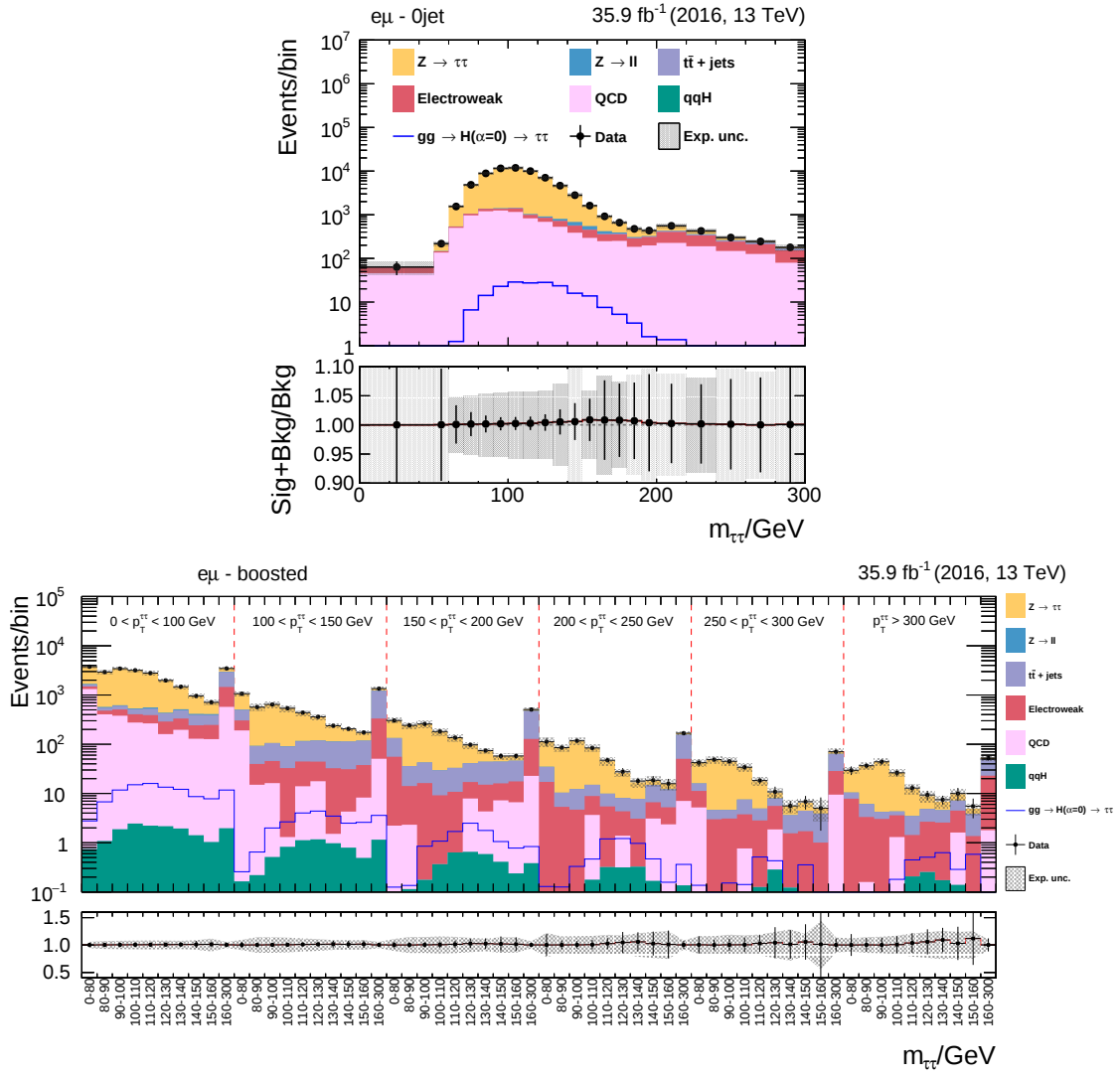


Fig. A.17.: Prefit distributions of the 0-jet and boosted categories in the $e\mu$ channel using the MELA approach.

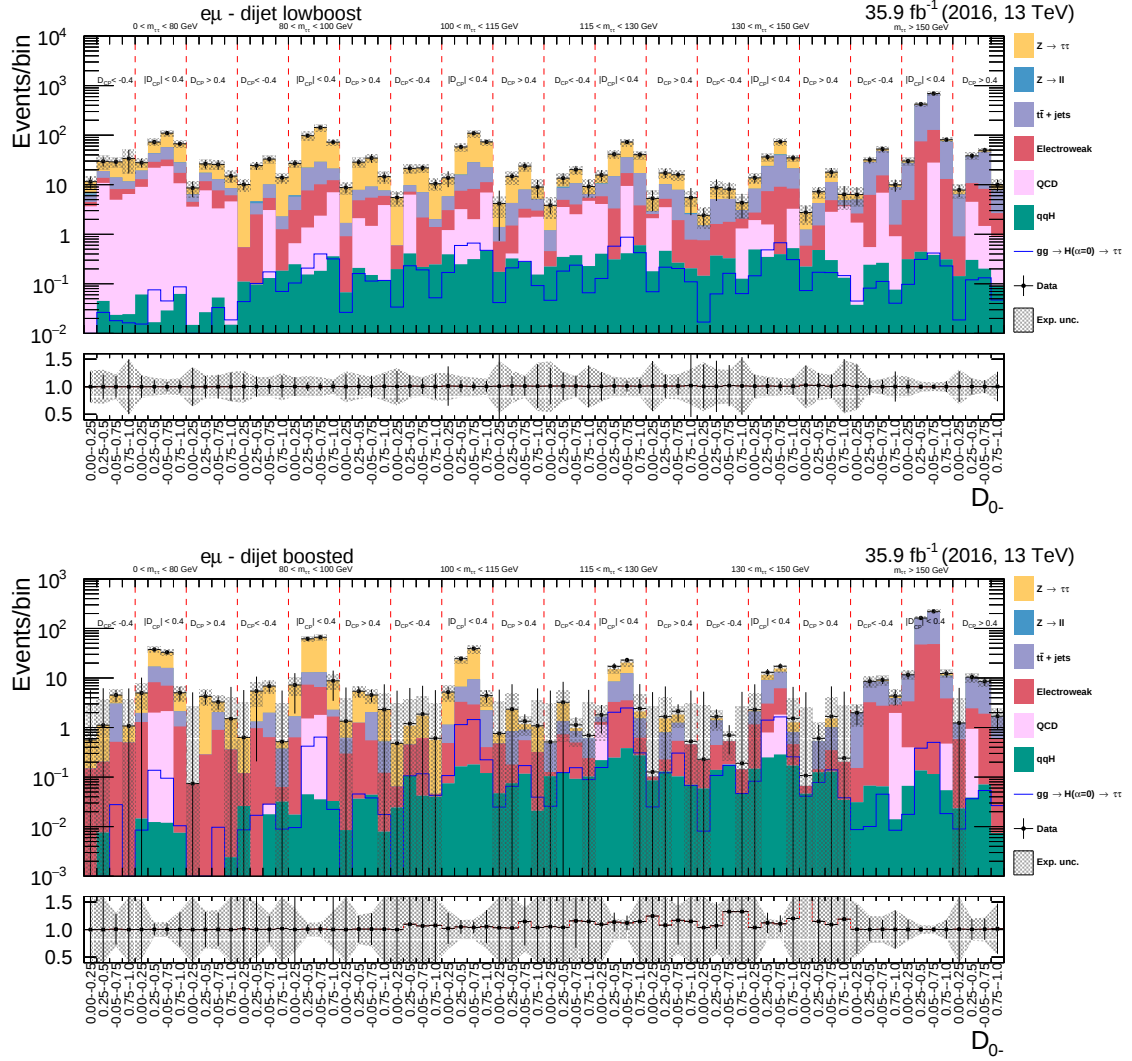


Fig. A.18.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\mu$ channel using the MELA approach.

A.3.2 Postfit distributions for MELA approach

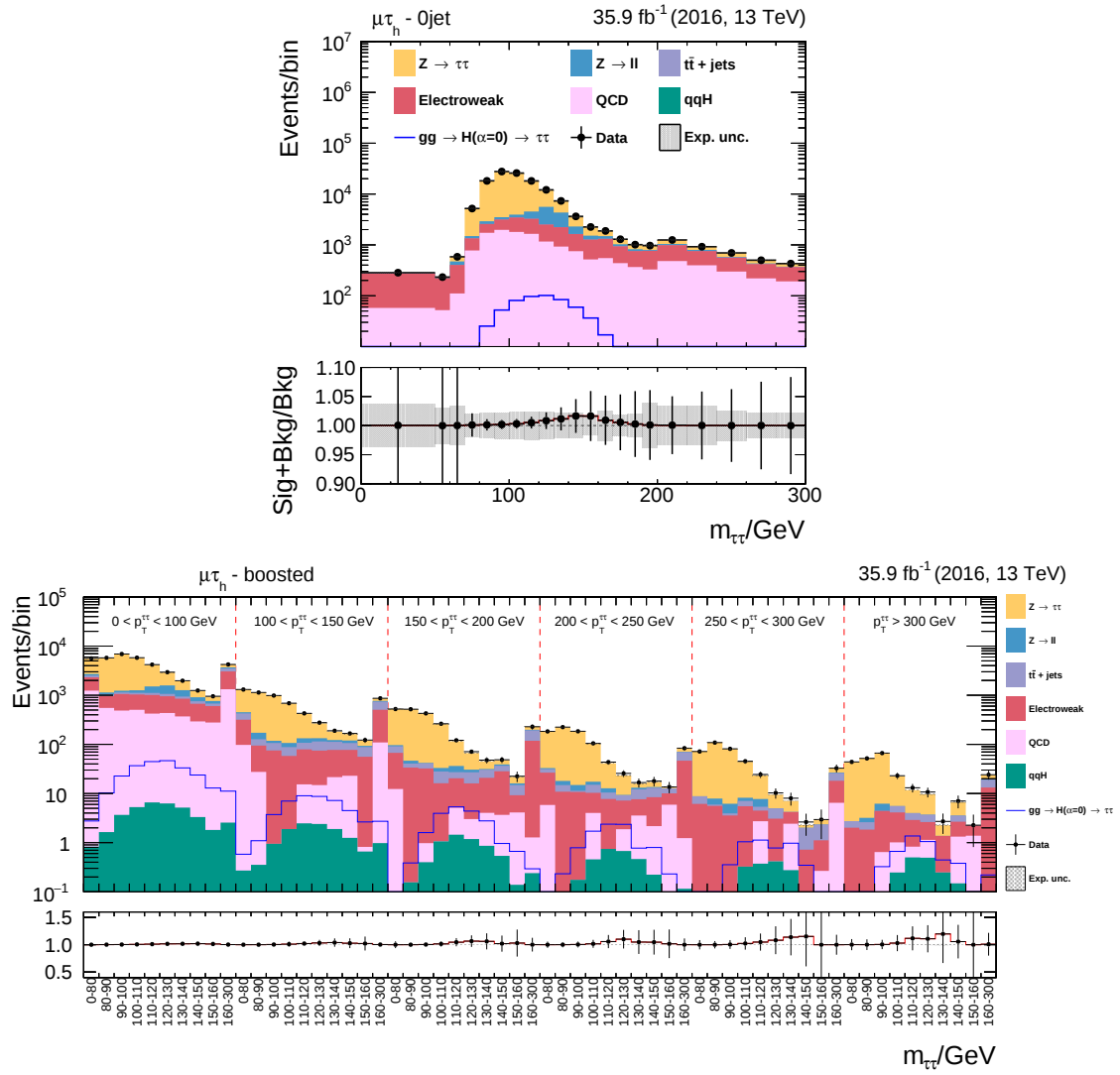


Fig. A.19.: Postfit distributions of the *0-jet* and *boosted* categories in the $\mu\tau_h$ channel using the MELA approach.

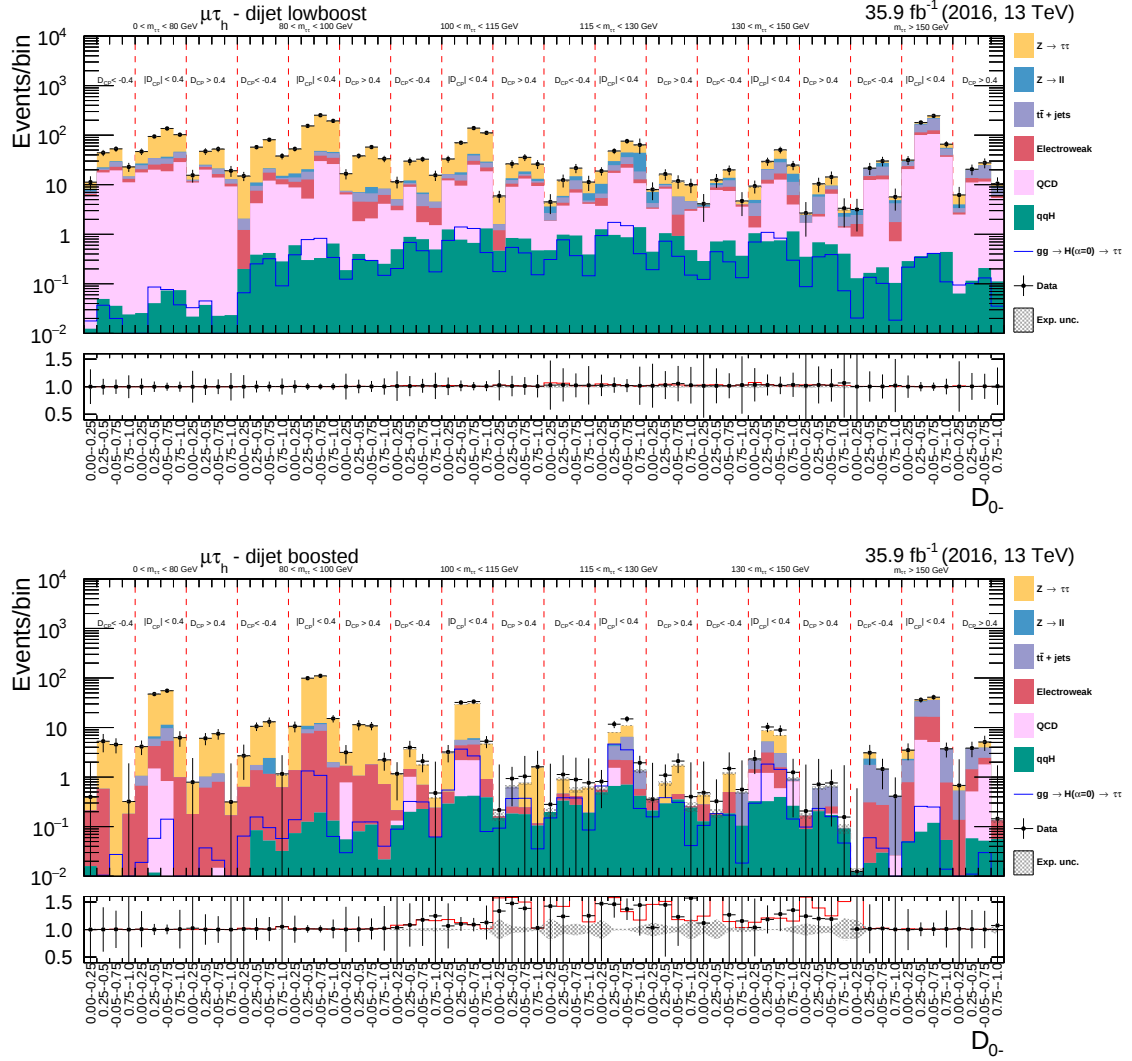


Fig. A.20.: Postfit distributions of the *dijet lowboost* and *dijet boosted* categories in the $\mu\tau_h$ channel using the MELA approach.

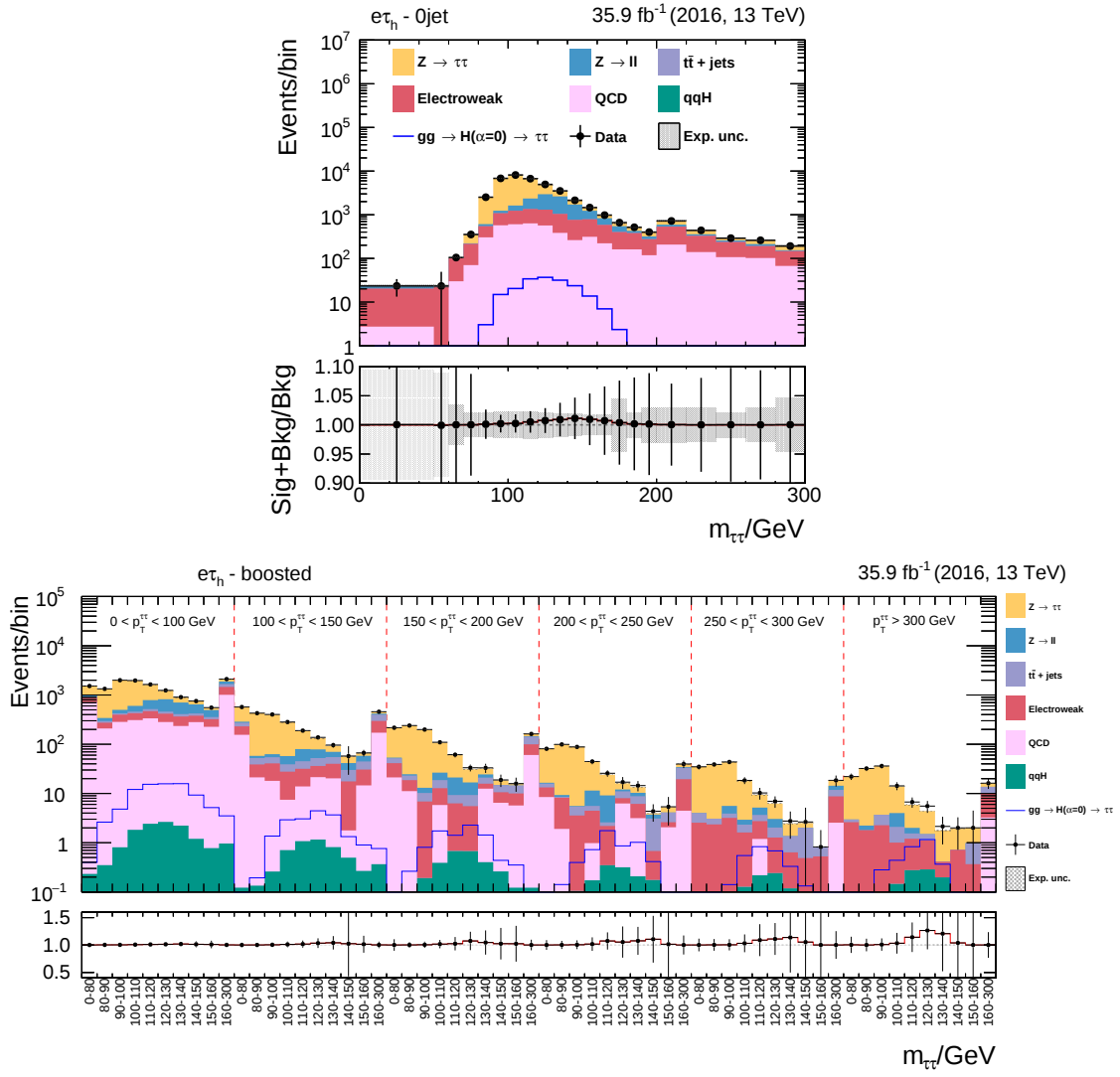


Fig. A.21.: Postfit distributions of the *0-jet* and *boosted* categories in the $e\tau_h$ channel using the MELA approach.

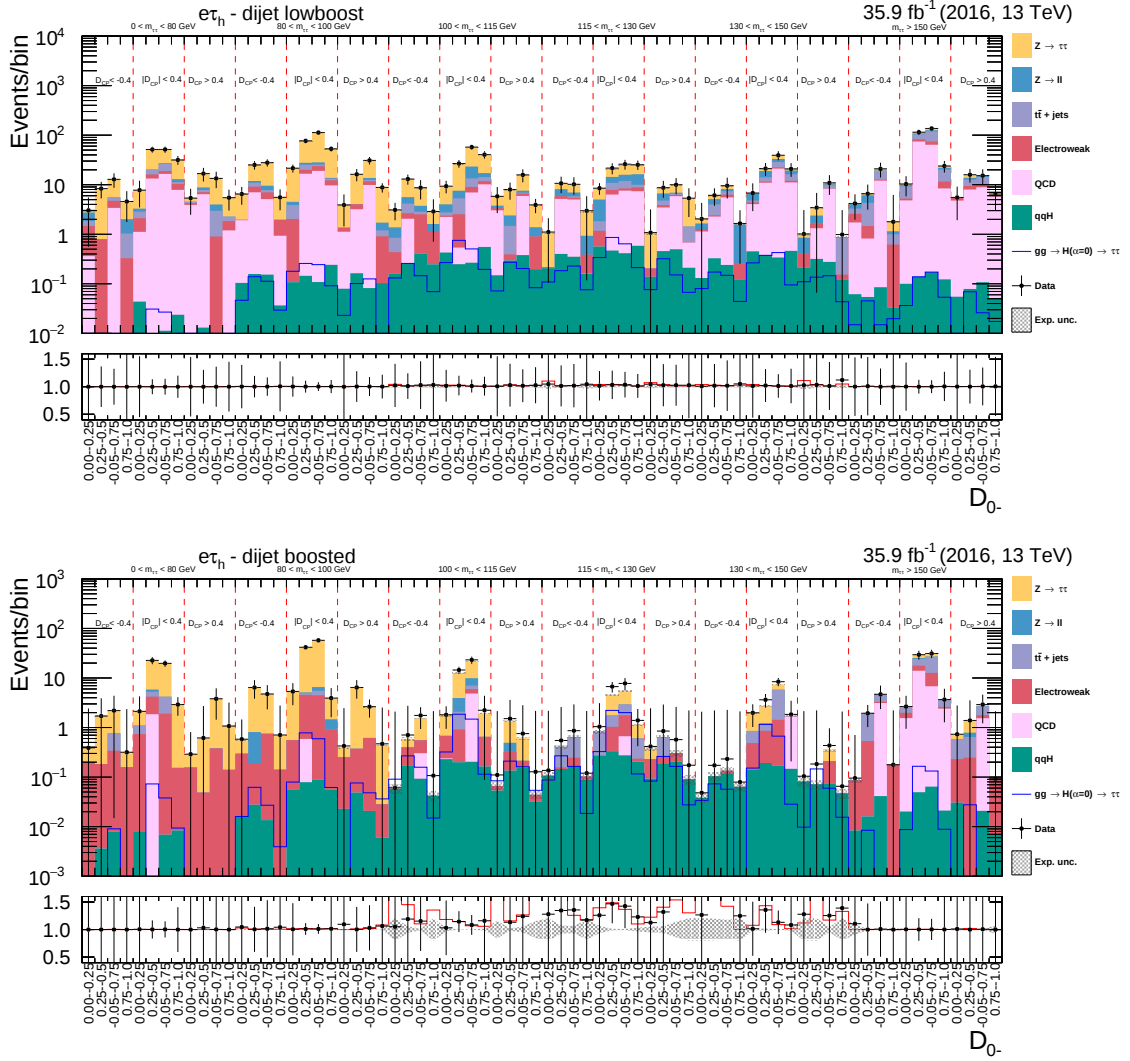


Fig. A.22.: Postfit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\tau_h$ channel using the MELA approach.

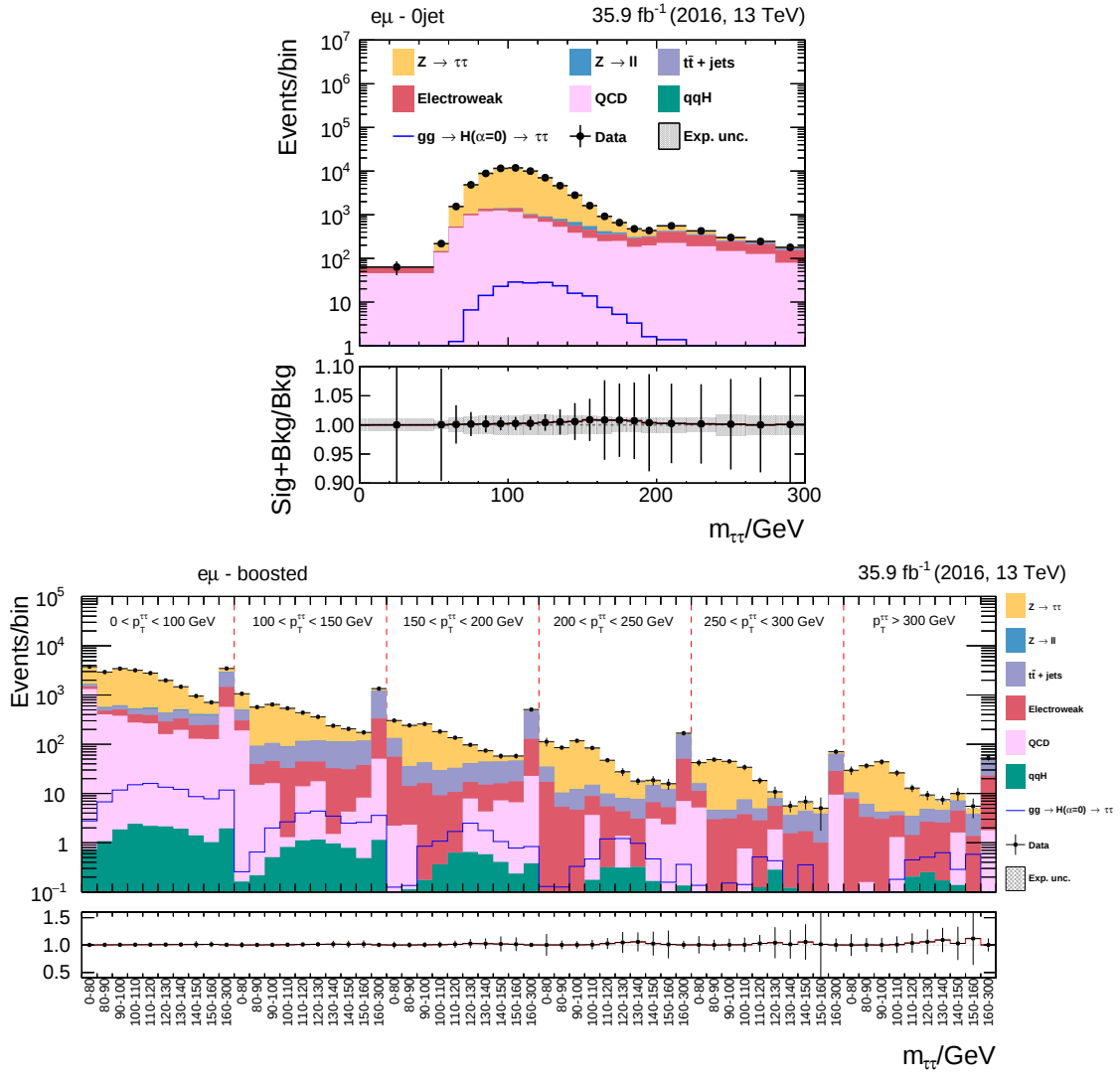


Fig. A.23.: Postfit distributions of the 0 -jet and $boosted$ categories in the $e\mu$ channel using the MELA approach.

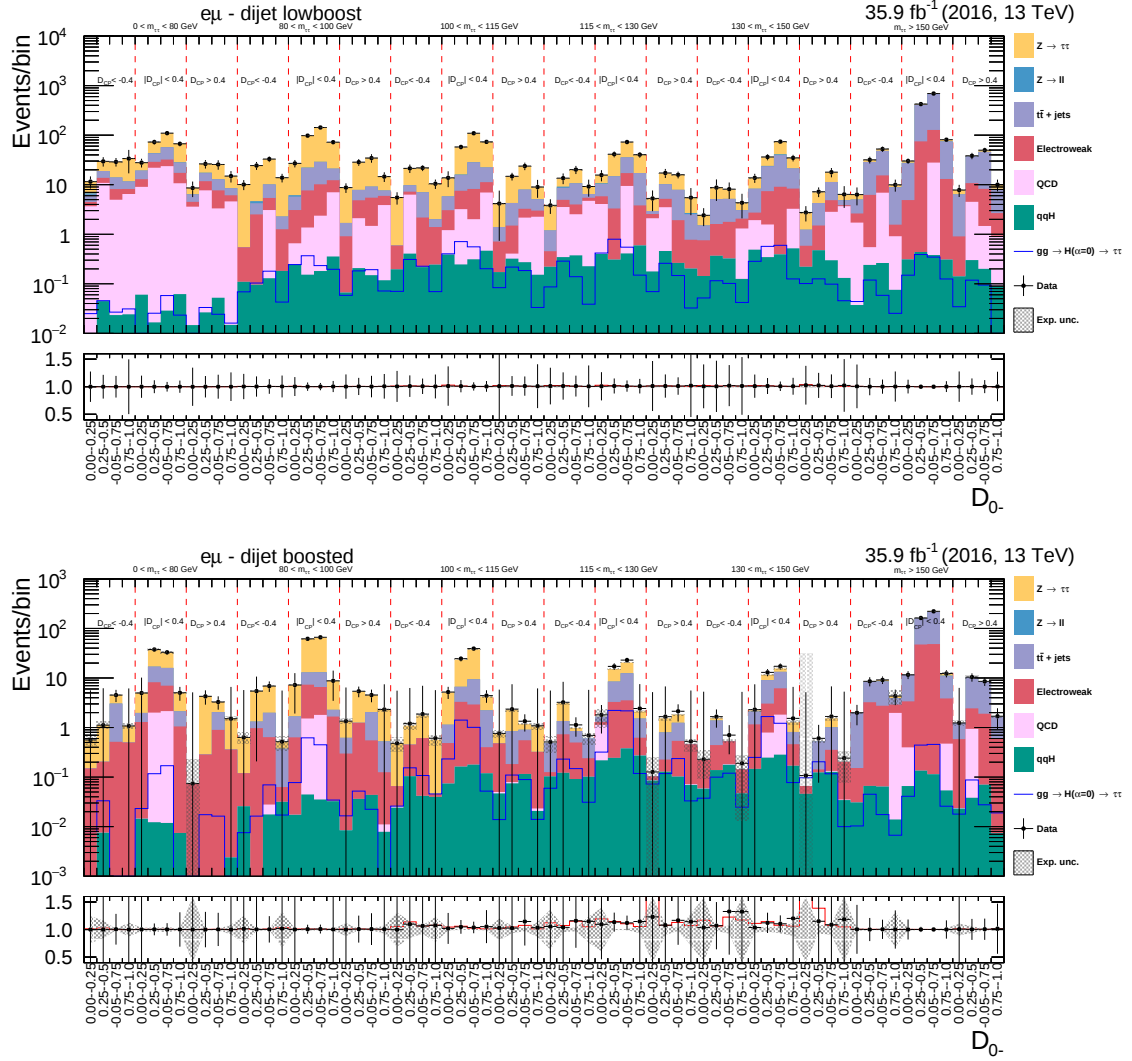


Fig. A.24.: Postfit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\mu$ channel using the MELA approach.

A.3.3 Prefit distributions for $\Delta\phi_{jj}$

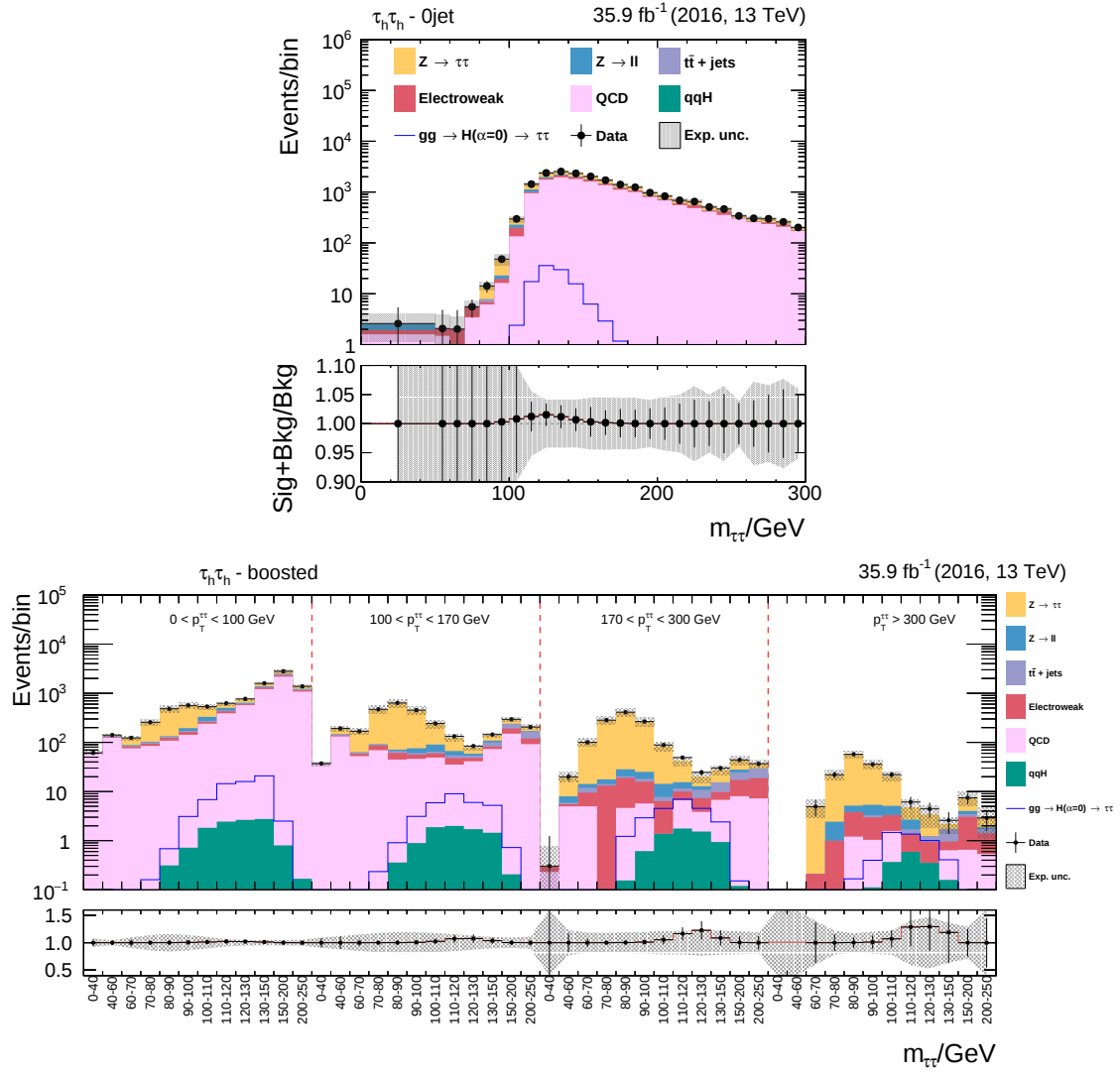


Fig. A.25.: Prefit distributions of the *0-jet* and *boosted* categories in the $\tau_h \tau_h$ channel using $\Delta\phi_{jj}$.

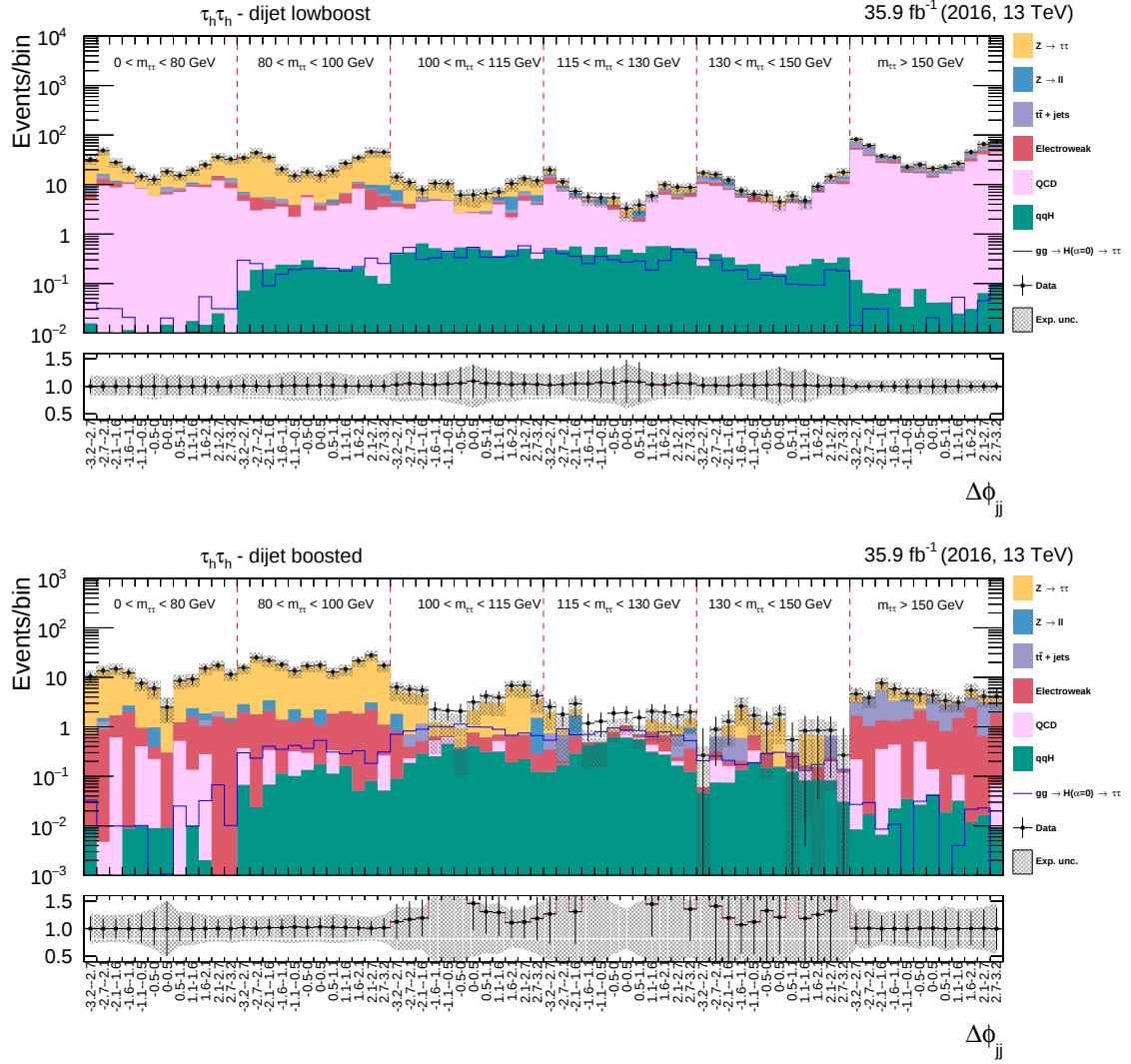


Fig. A.26.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $\tau_h \tau_h$ channel using $\Delta\phi_{jj}$.

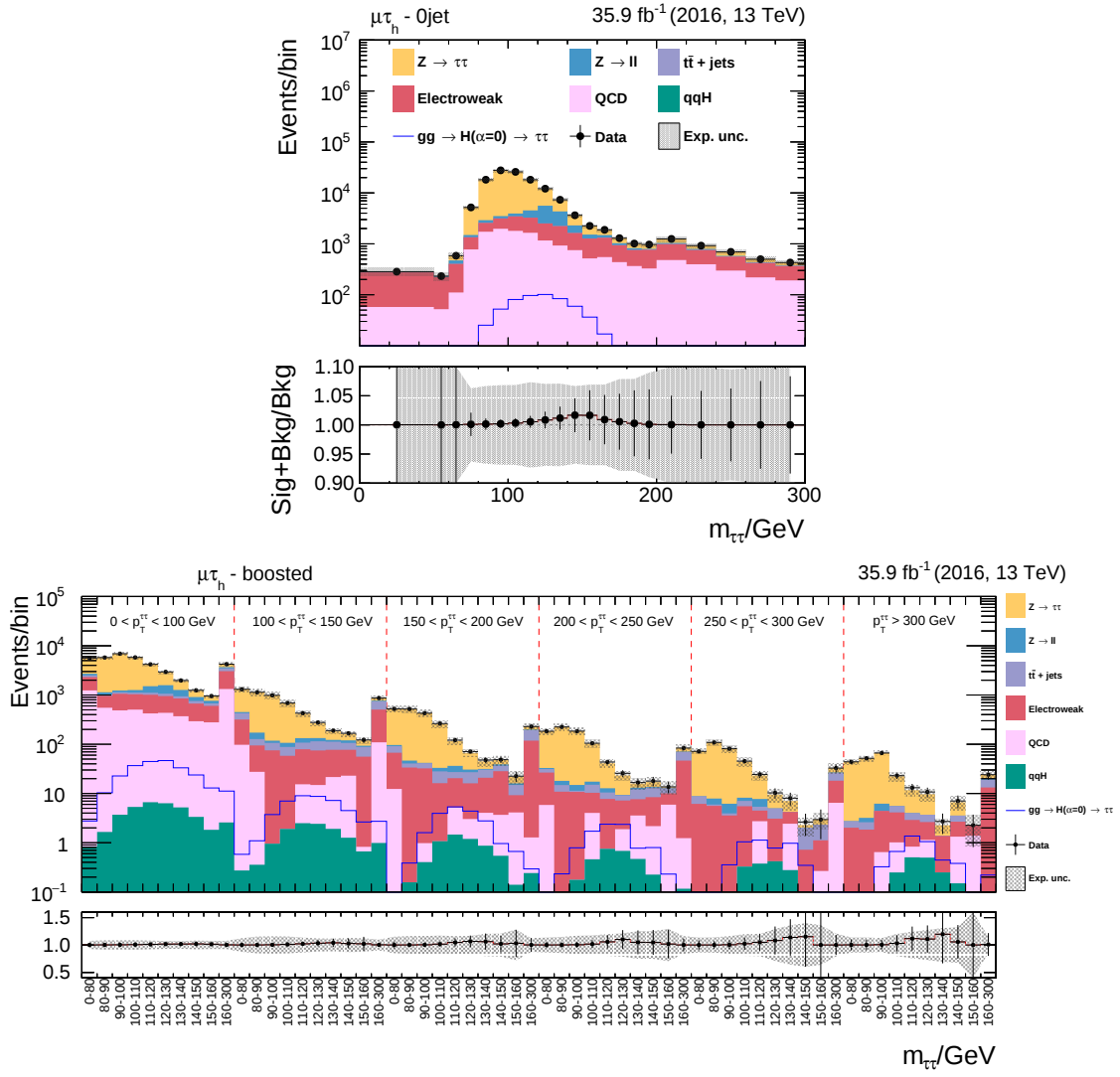


Fig. A.27.: Prefit distributions of the *0-jet* and *boosted* categories in the $\mu\tau_h$ channel using $\Delta\phi_{jj}$.

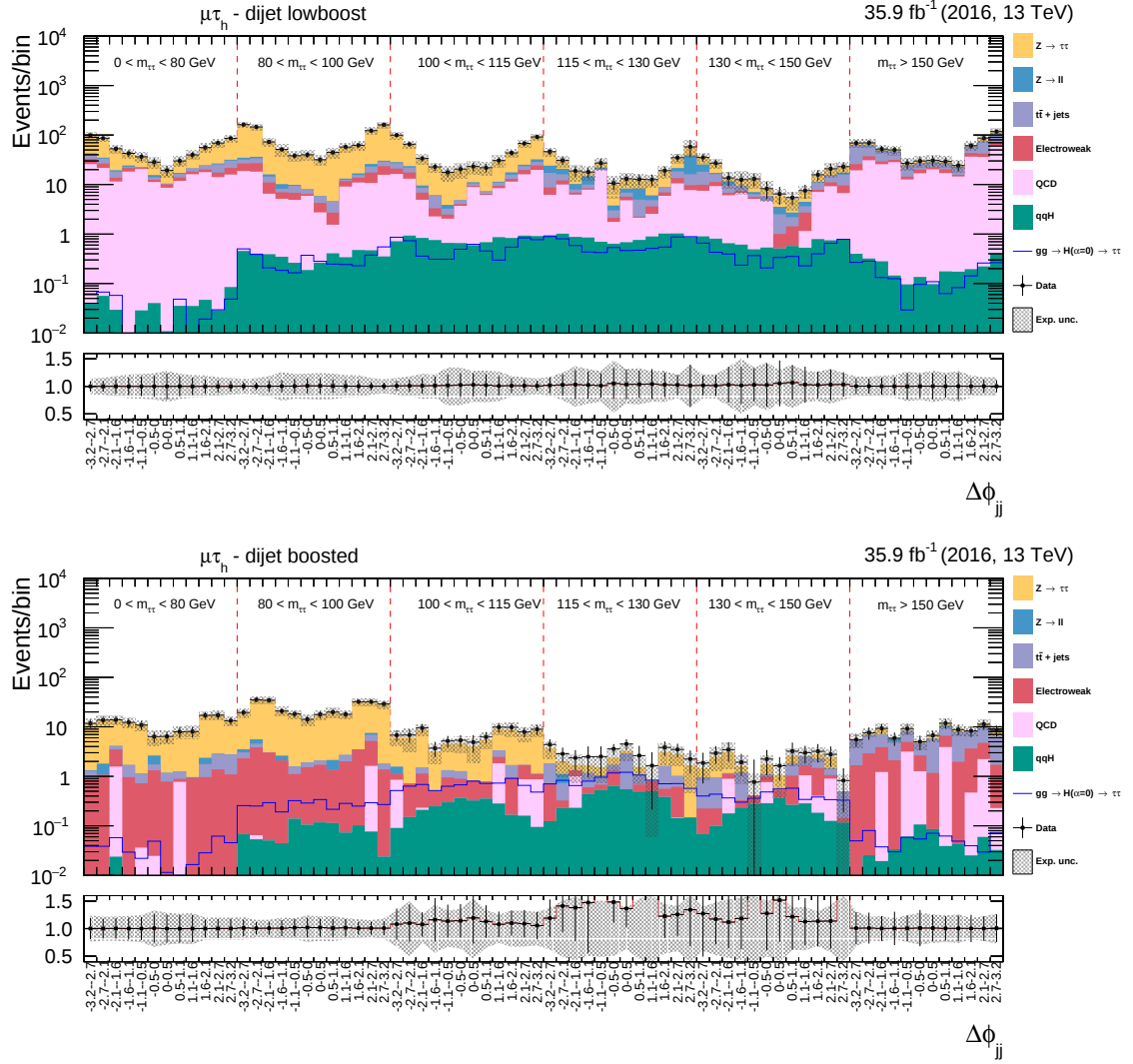


Fig. A.28.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $\mu\tau_h$ channel using $\Delta\phi_{jj}$.

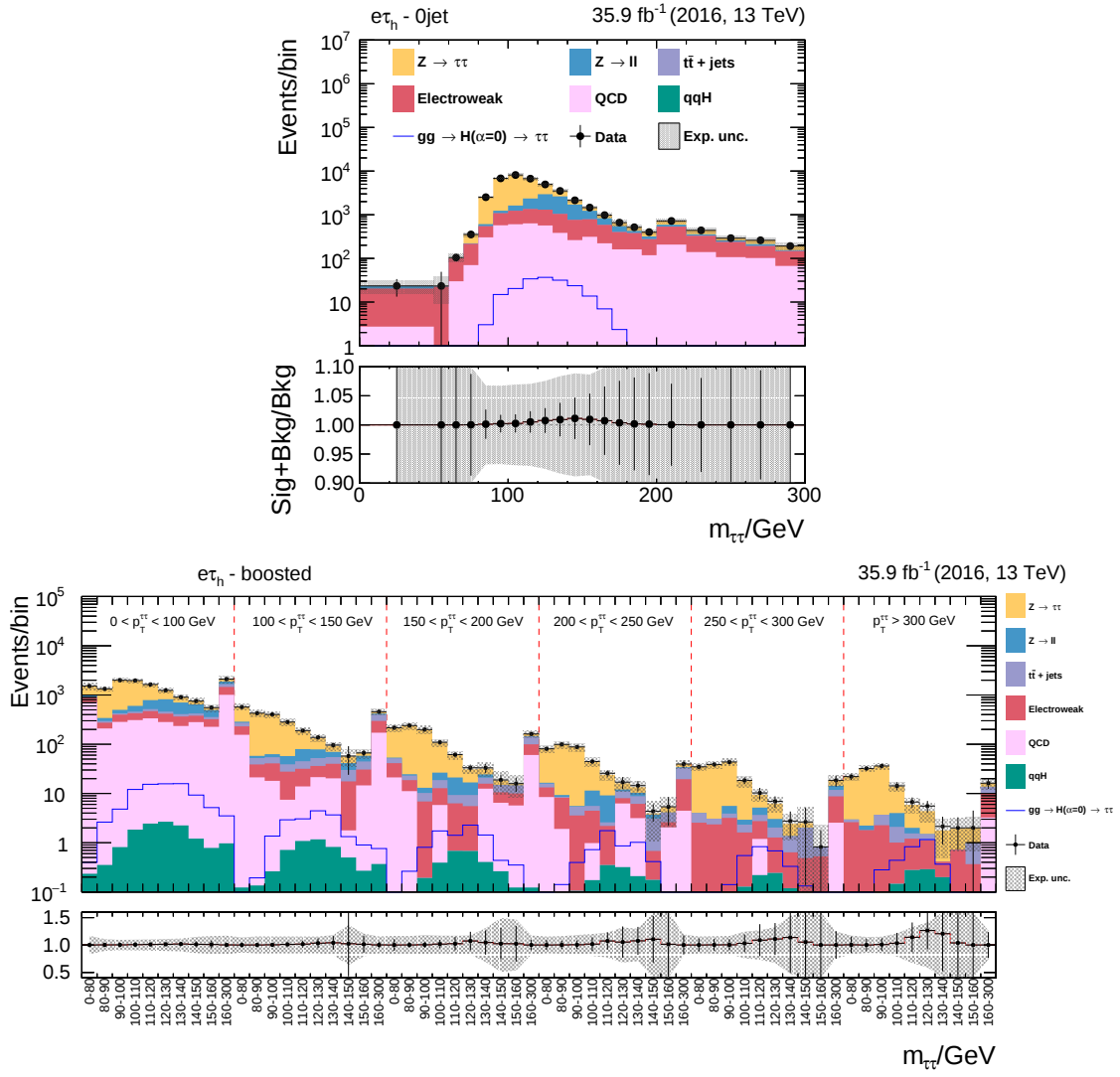


Fig. A.29.: Prefit distributions of the *0-jet* and *boosted* categories in the $e\tau_h$ channel using $\Delta\phi_{jj}$.

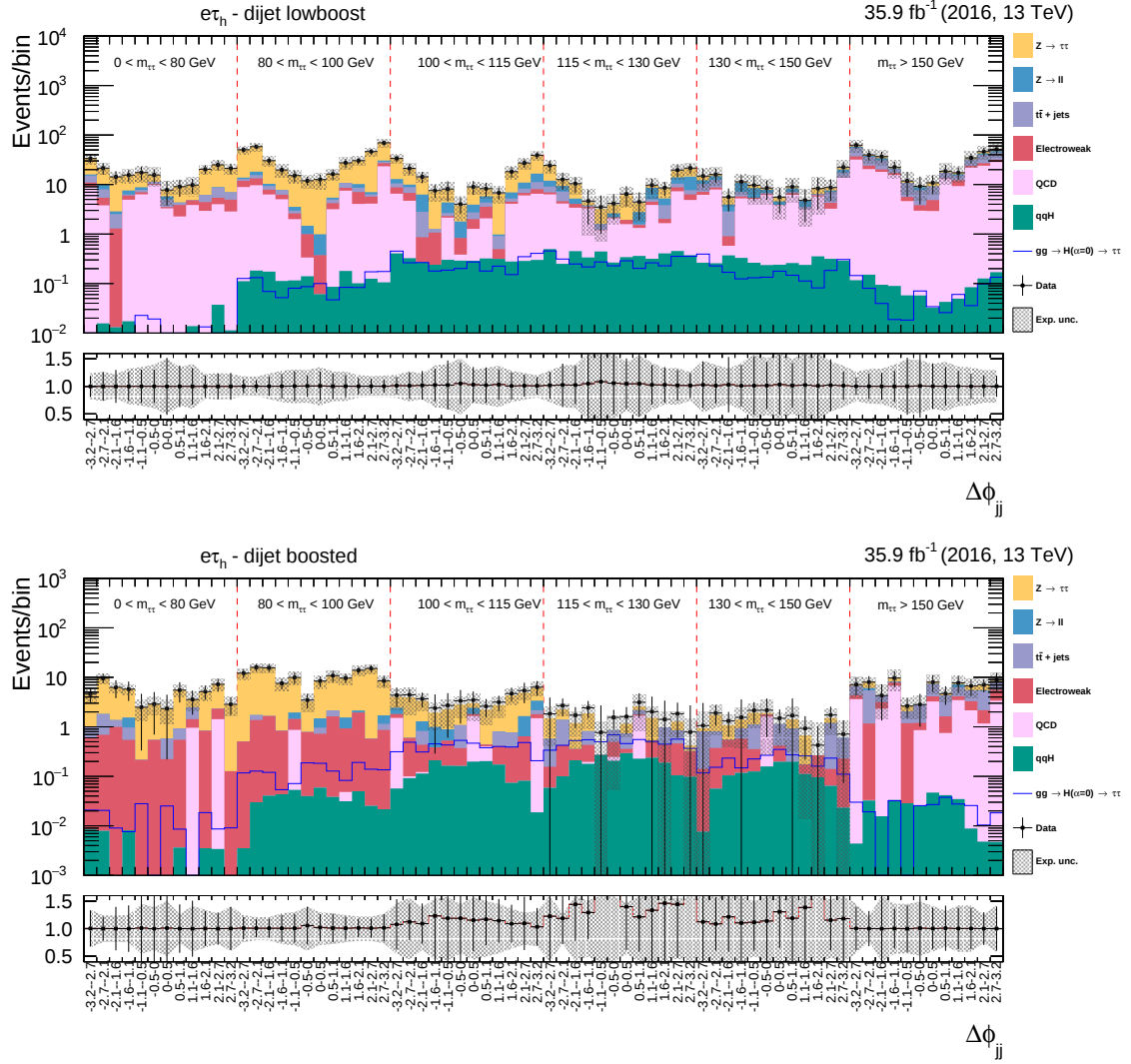


Fig. A.30.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\tau_h$ channel using $\Delta\phi_{jj}$.

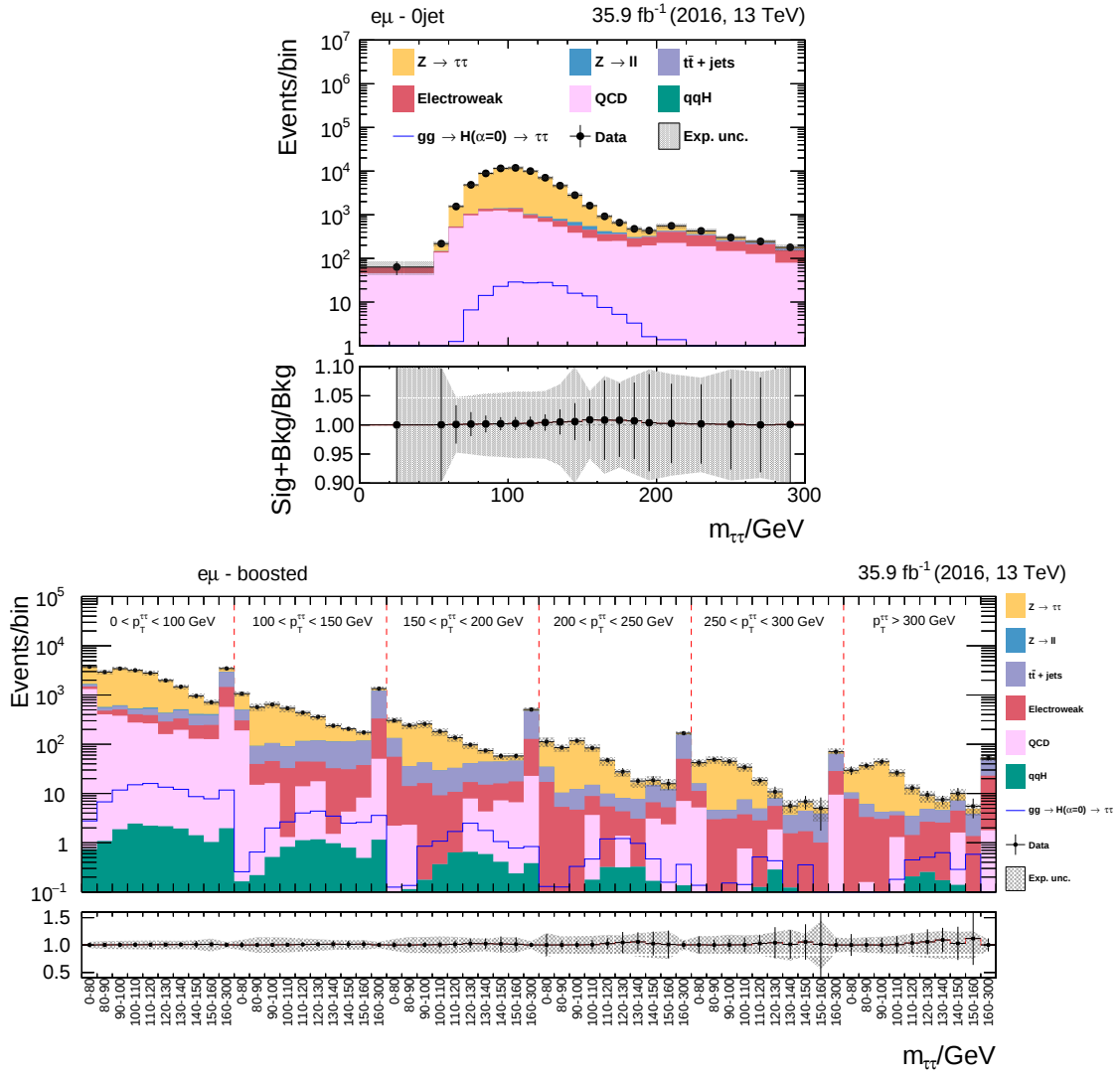


Fig. A.31.: Prefit distributions of the *0-jet* and *boosted* categories in the $e\mu$ channel using $\Delta\phi_{jj}$.

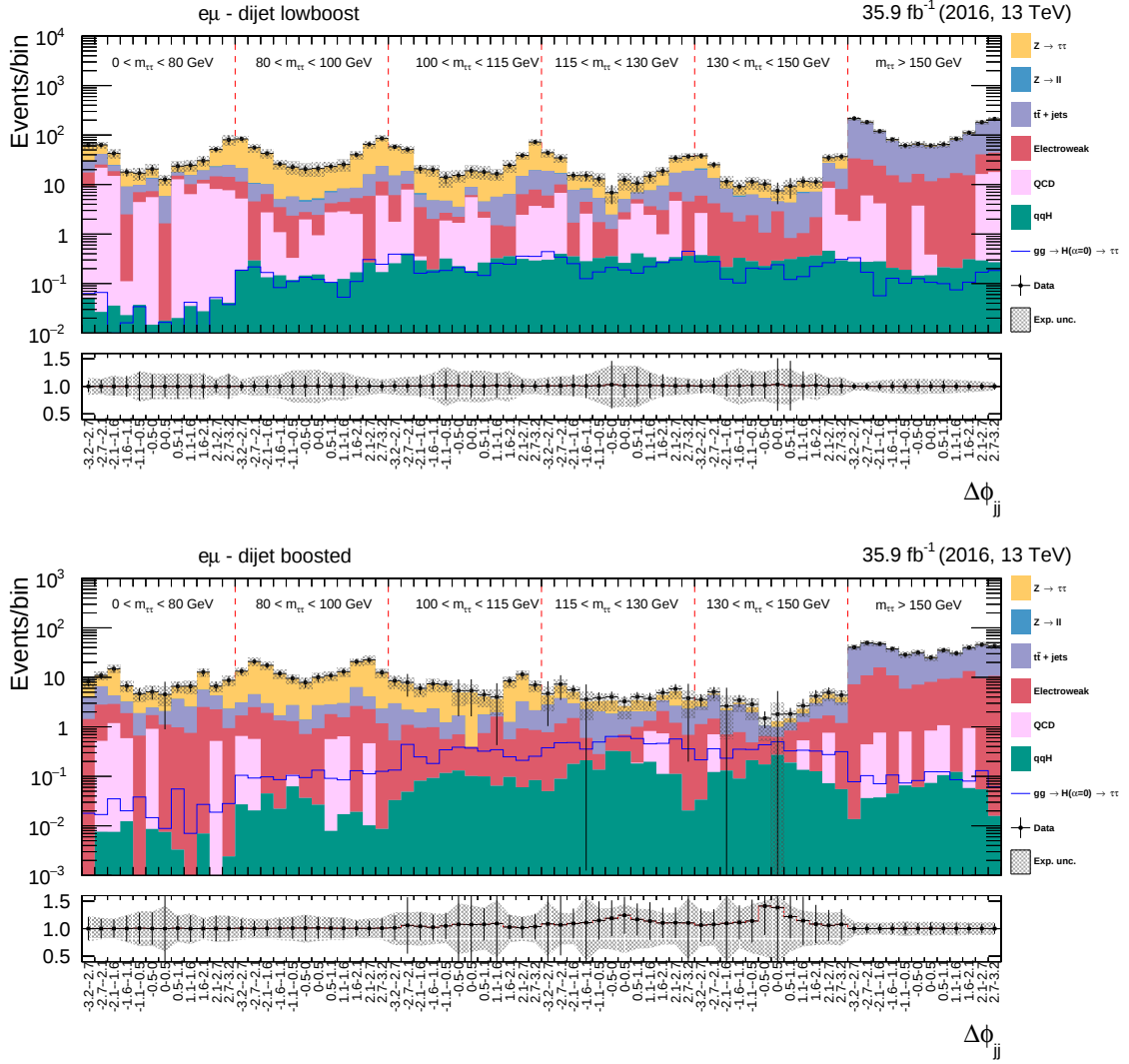


Fig. A.32.: Prefit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\mu$ channel using $\Delta\phi_{jj}$.

A.3.4 Postfit distributions for $\Delta\phi_{jj}$

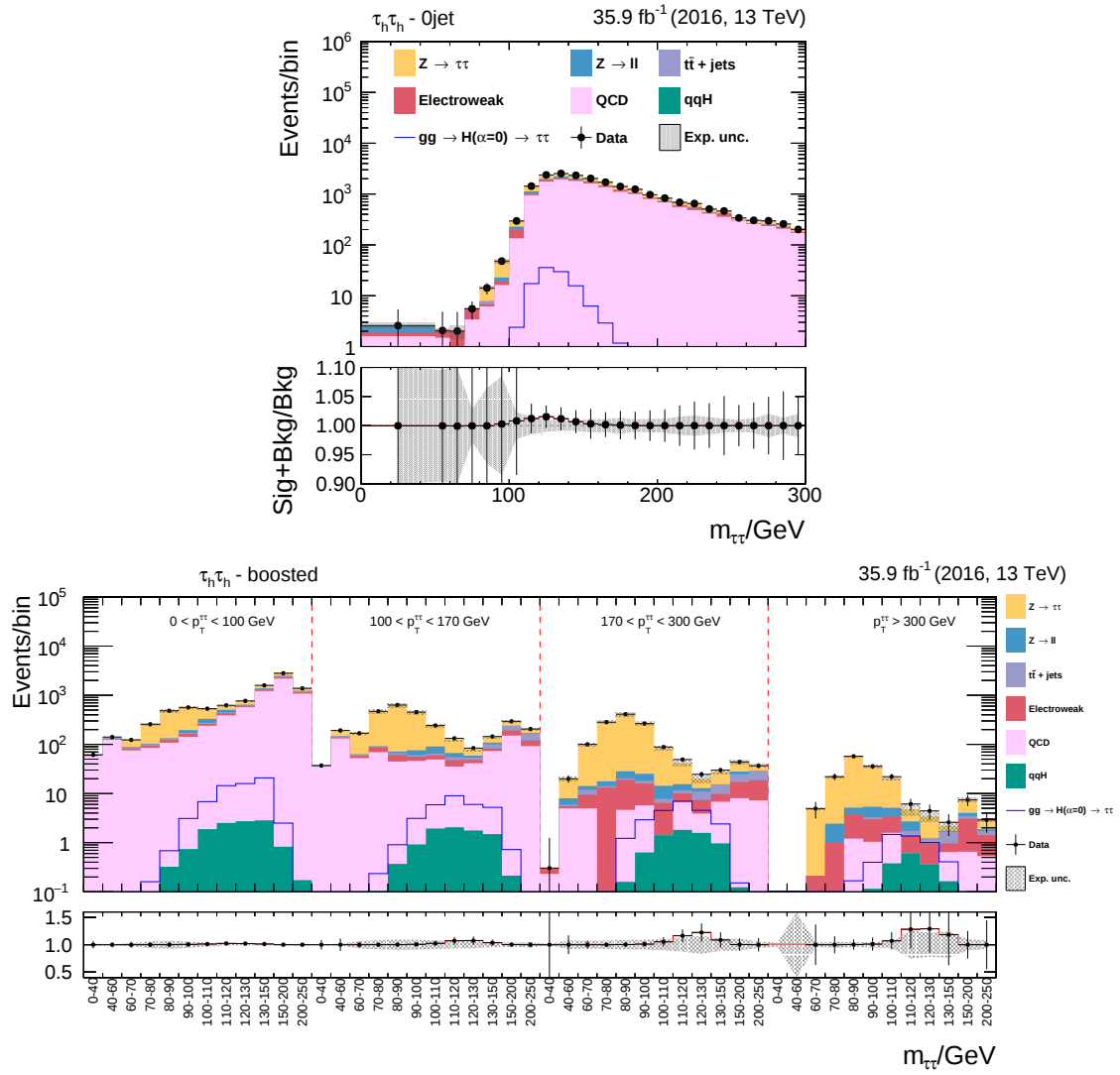


Fig. A.33.: Postfit distributions of the *0-jet* and *boosted* categories in the $\tau_h\tau_h$ channel using $\Delta\phi_{jj}$.

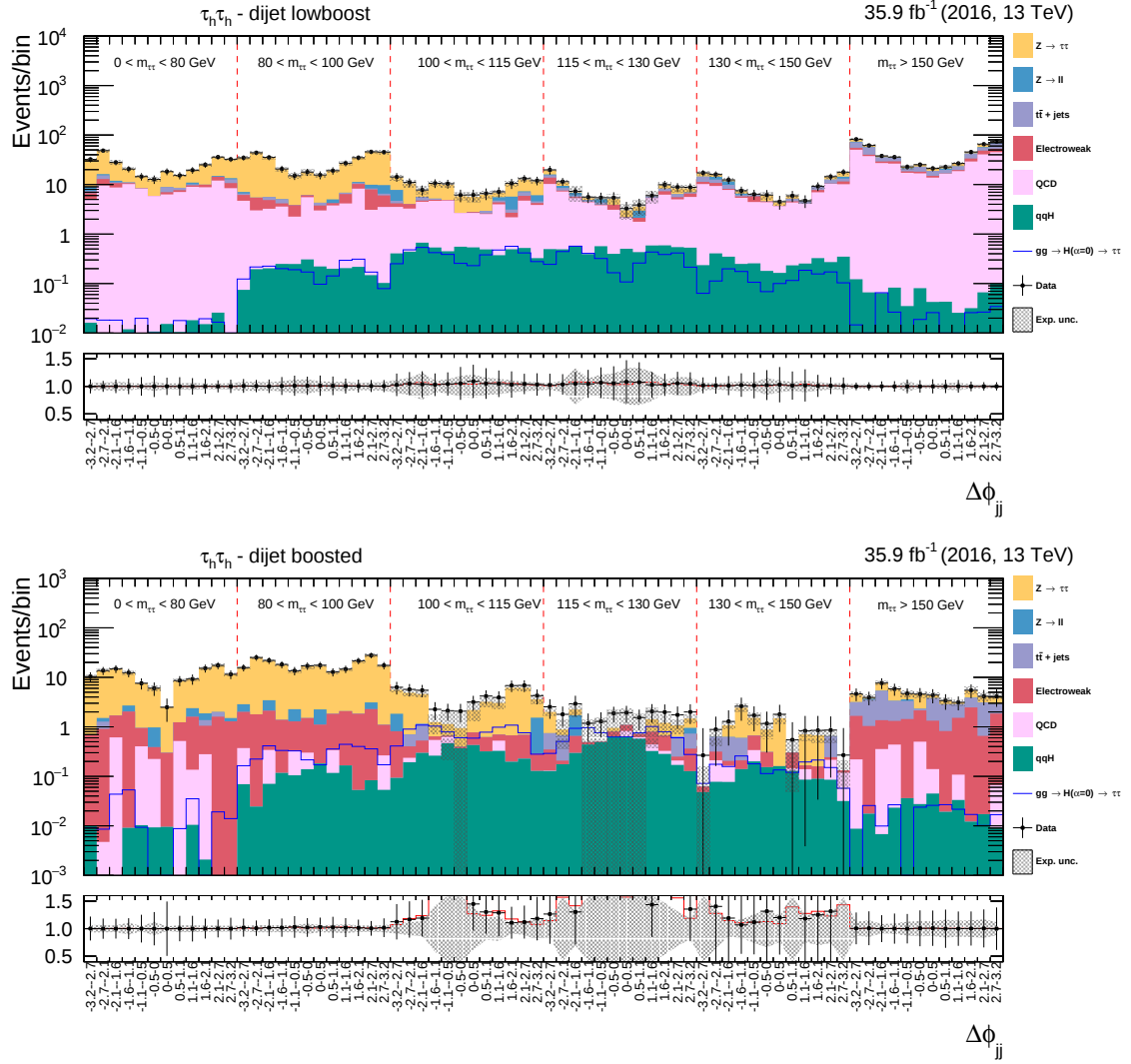


Fig. A.34.: Postfit distributions of the *dijet lowboost* and *dijet boosted* categories in the $\tau_h \tau_h$ channel using $\Delta\phi_{jj}$.

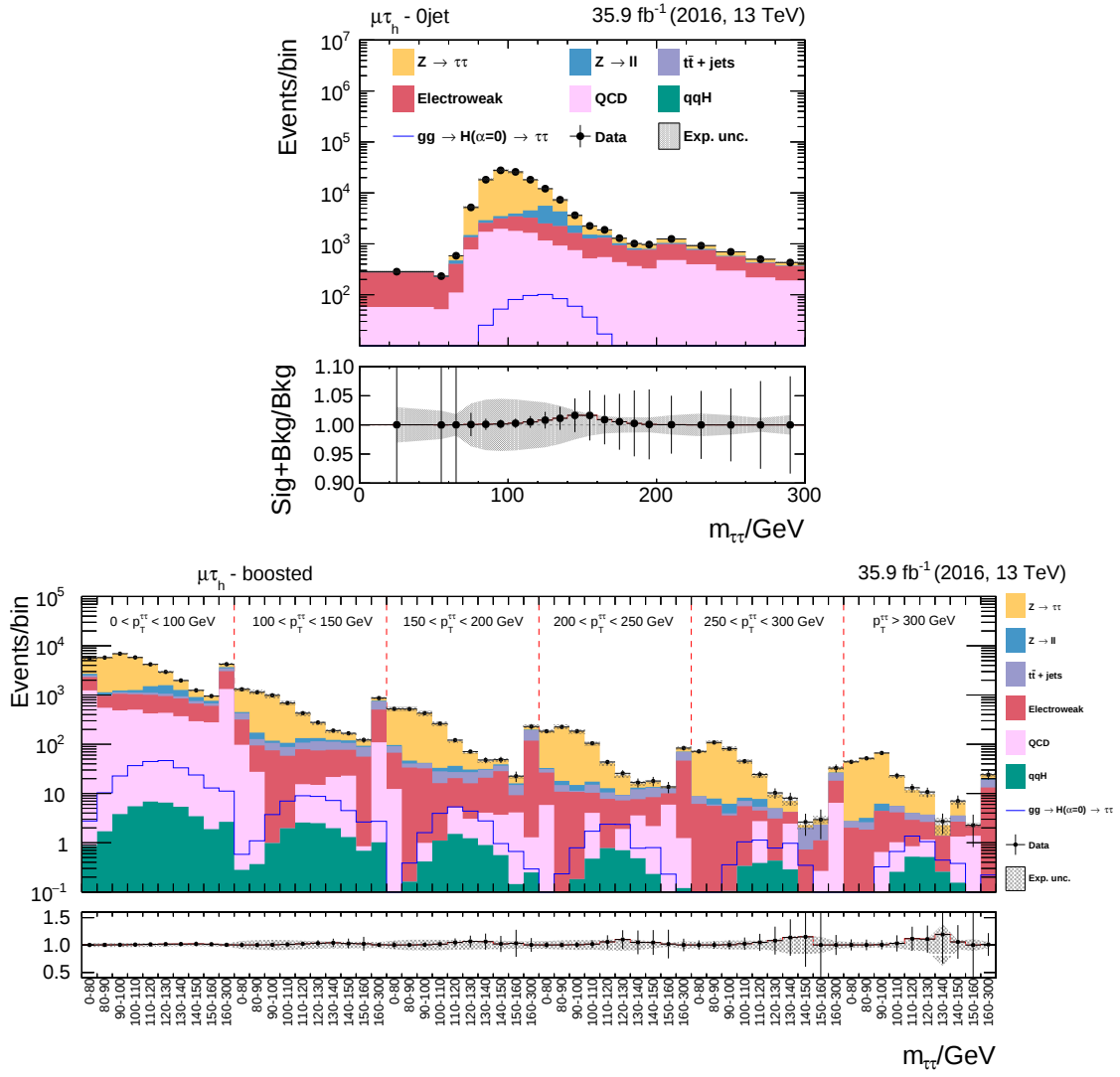


Fig. A.35.: Postfit distributions of the *0-jet* and *boosted* categories in the $\mu\tau_h$ channel using $\Delta\phi_{jj}$.

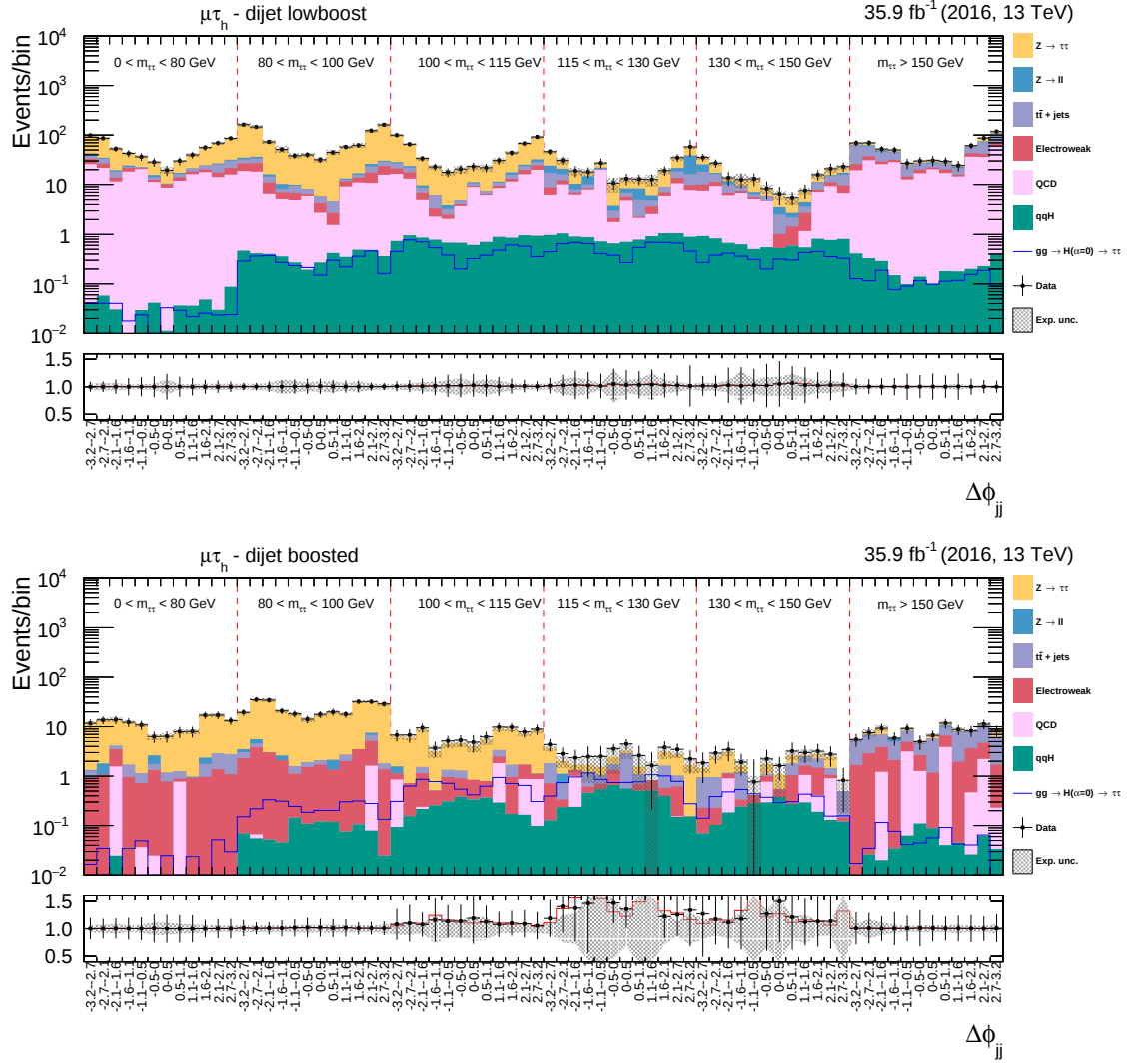


Fig. A.36.: Postfit distributions of the *dijet lowboost* and *dijet boosted* categories in the $\mu\tau_h$ channel using $\Delta\phi_{jj}$.

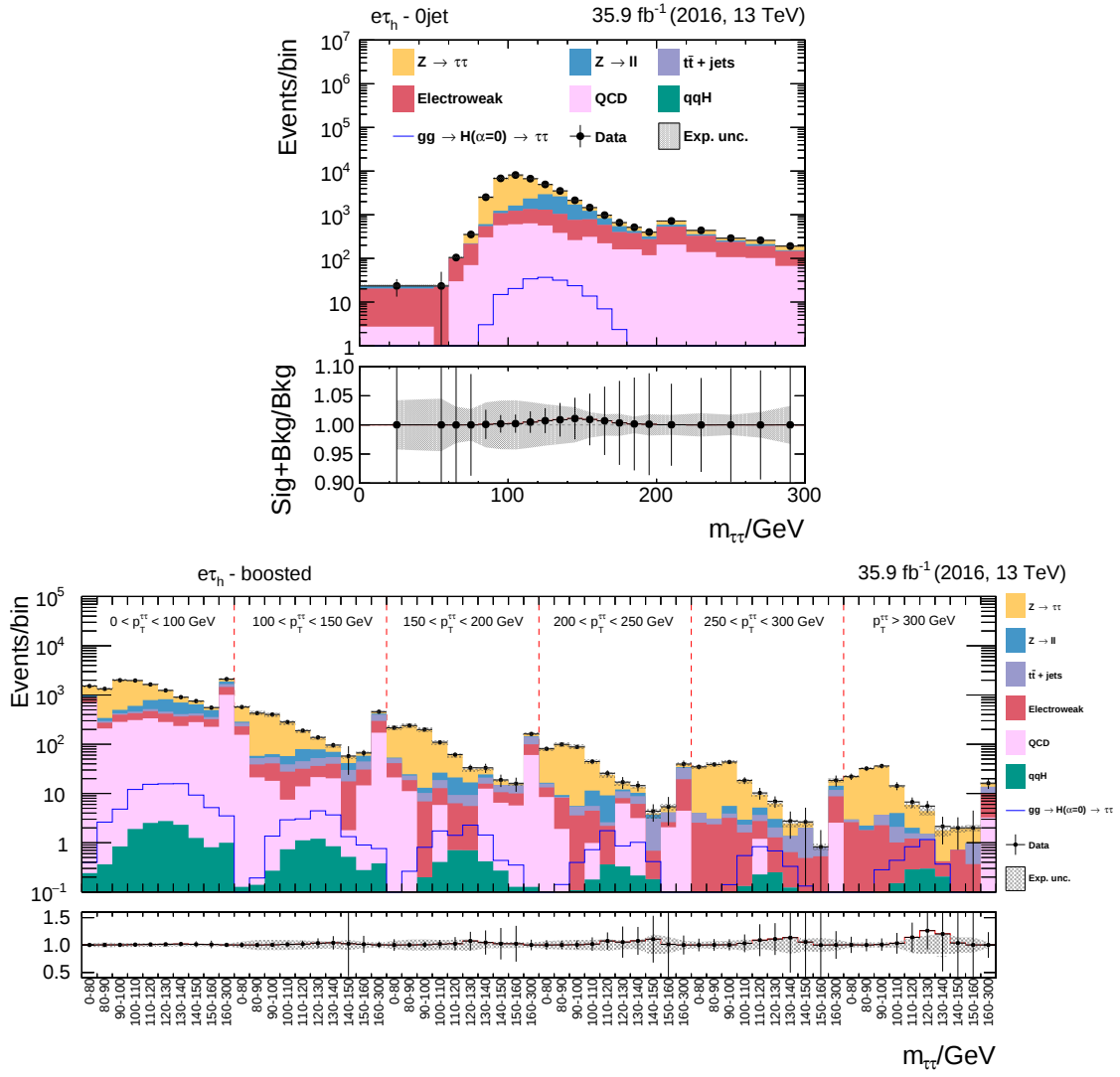


Fig. A.37.: Postfit distributions of the *0-jet* and *boosted* categories in the $e\tau_h$ channel using $\Delta\phi_{jj}$.

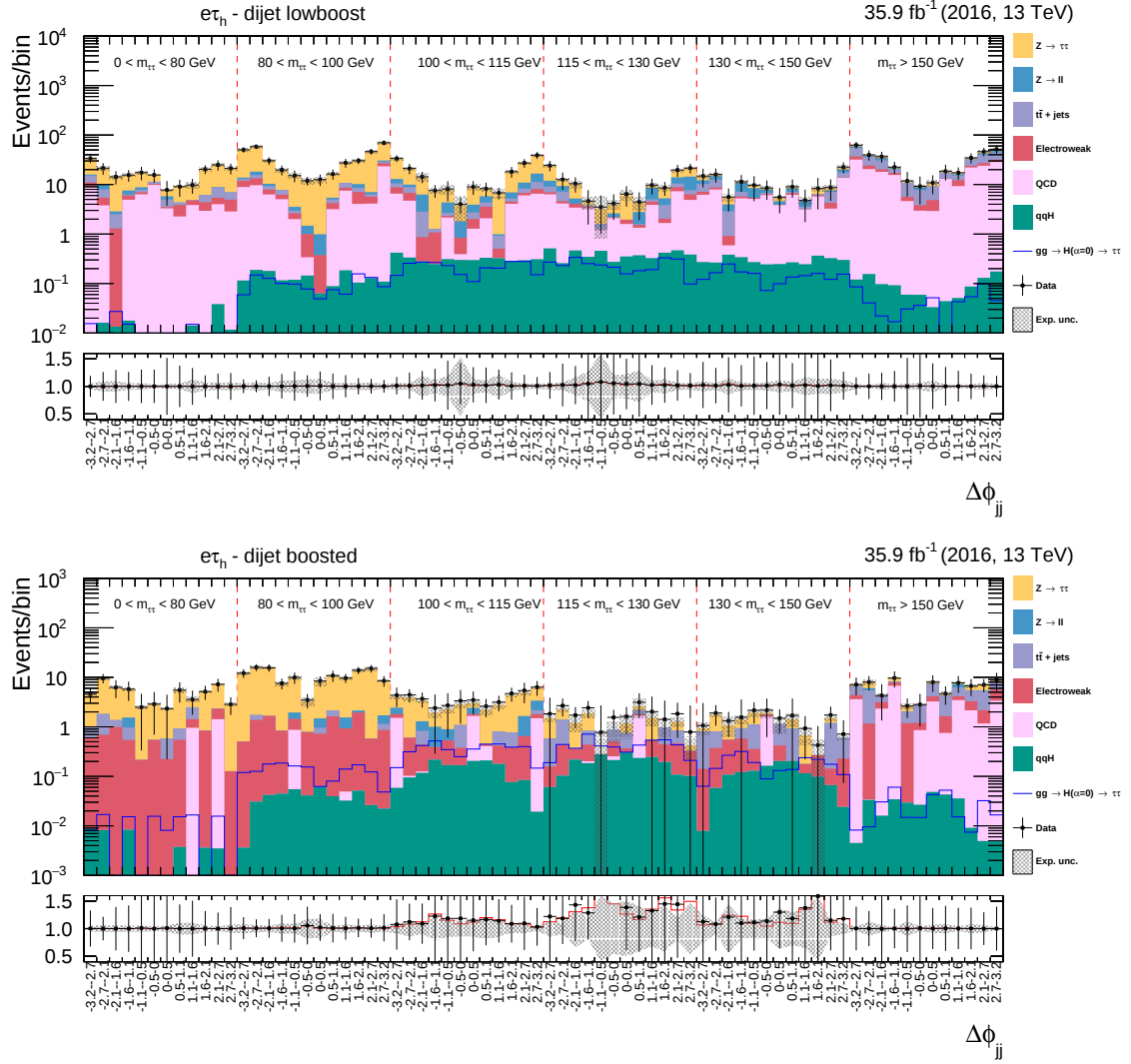


Fig. A.38.: Postfit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\tau_h$ channel using $\Delta\phi_{jj}$.

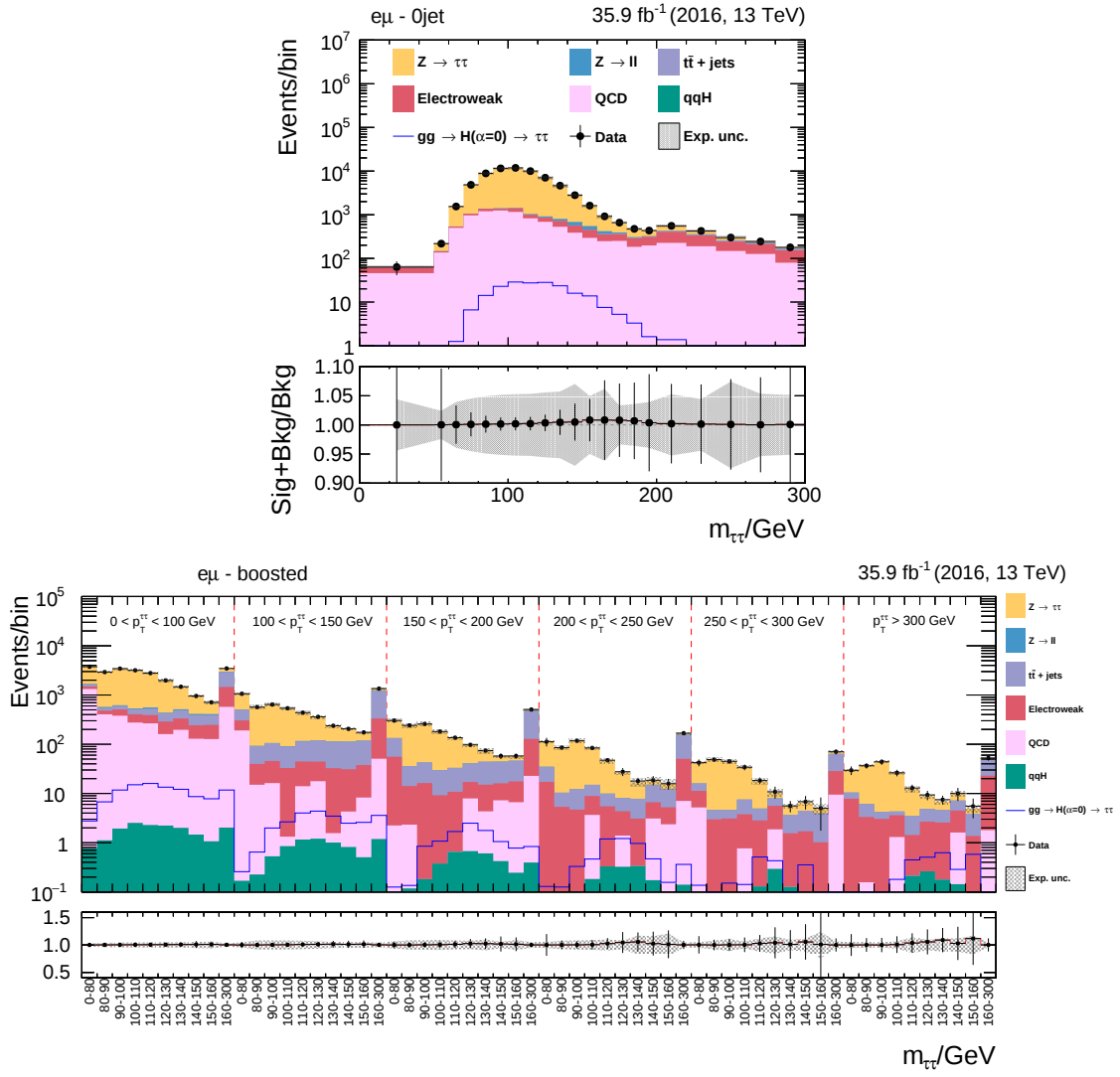


Fig. A.39.: Postfit distributions of the *0-jet* and *boosted* categories in the $e\mu$ channel using $\Delta\phi_{jj}$.

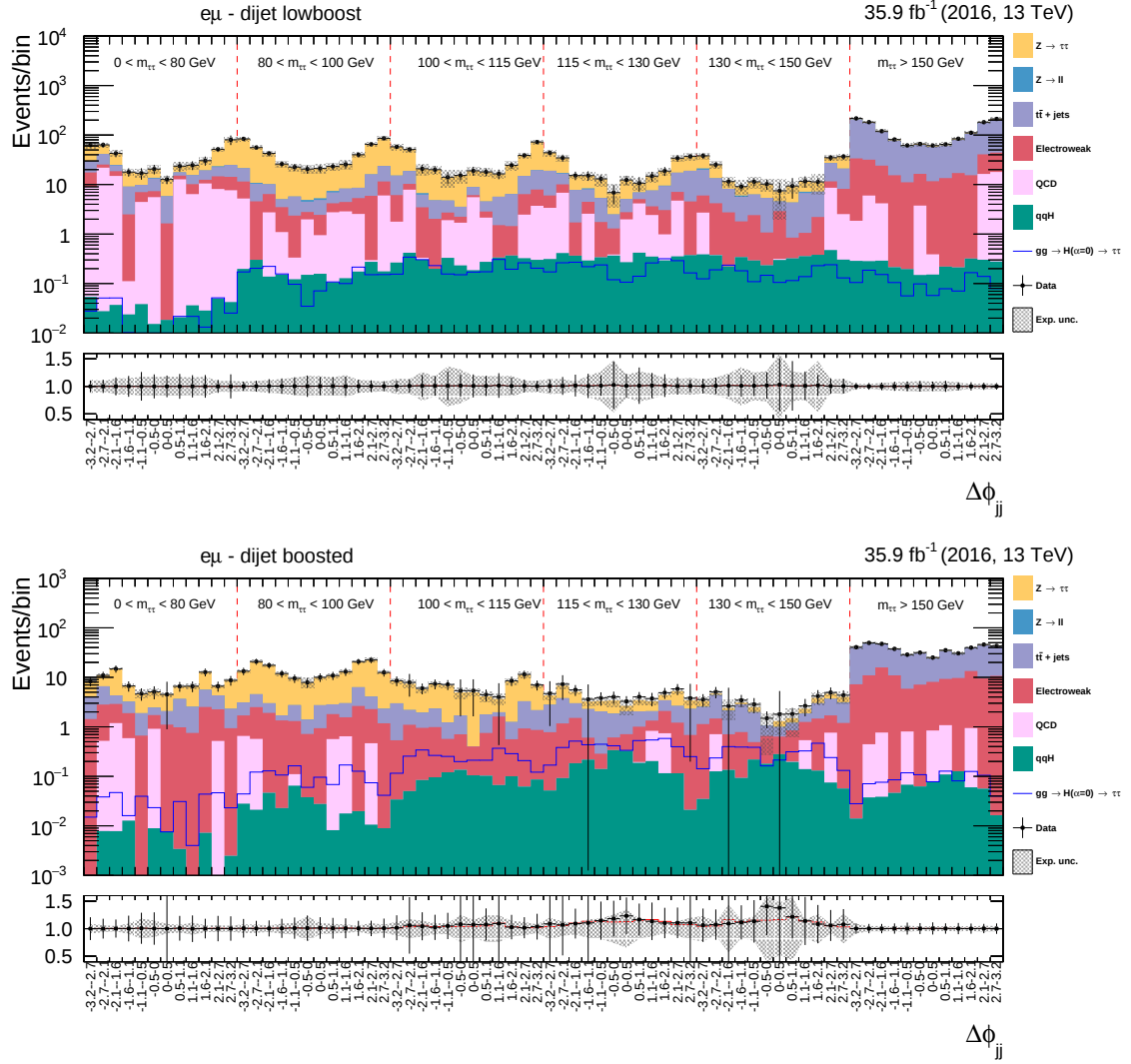


Fig. A.40.: Postfit distributions of the *dijet lowboost* and *dijet boosted* categories in the $e\mu$ channel using $\Delta\phi_{jj}$.

Eidesstattliche Versicherung Statutory Declaration in Lieu of an Oath

Wolfschläger, Dominik

Name, Vorname/Last Name, First Name

331848

Matrikelnummer (freiwillige Angabe)

Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende ~~Arbeit/Bachelorarbeit/~~
Masterarbeit* mit dem Titel

I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

Search for CP violation in fermionic couplings of the Higgs boson
generated by gluon-gluon fusion with the CMS Detector

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Aachen,

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:

Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

Para. 156 StGB (German Criminal Code): False Statutory Declarations

Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Strafflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.

Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence

(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.

(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

I have read and understood the above official notification:

Aachen,

Ort, Datum/City, Date

Unterschrift/Signature

