ETP-KA/2023-3

Karlsruher Institut für Technologie

Measurement of the top-Yukawa coupling in top quark associated Higgs boson production and blueprint for a novel top quark decay width measurement in Wb scattering

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)
von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.Sc. Marco Link

Tag der mündlichen Prüfung: **28. Juli 2023**

Referent: **Prof. Dr. Thomas Müller**
Korreferent: **Prof. Dr. Günter Quast**

Contents

1 Introduction

With the discovery of the Higgs boson in 2012 [1, 2], the last missing piece of the Standard Model of particle physics was found, yet some questions still remain unanswered: Currently the Standard Model does not include dark matter candidates to explain cosmological observations. A description of gravity compatible with the Standard Model still has to be found and the observation of neutrino oscillations [3] contradicts the massless neutrinos of the Standard Model. In search for deviations from the Standard Model, precise measurements of particle properties are performed.

This thesis probes the Standard Model through the measurement of top quark properties in order to find hidden traces of new physics. The measurement of the coupling strength of top quarks with the Higgs boson and the top quark decay width is of great interest, since the top quark as the heaviest particle of the Standard Model has a special place in many theories extending the Standard Model. Due to its large mass, it couples strongly with the Higgs boson and the short lifetime suppresses the formation of hadrons. By precisely measuring properties of the top quark, small deviations from the Standard Model can be found or constraints on new theories can be determined. With the Large Hadron Collider, a large amount of top quarks are produced and the Yukawa coupling between top quarks and the Higgs boson can be measured precisely.

The observation of top quark pair production in association with a Higgs boson ($t\bar{t}H$) [4] paved the path for the direct measurement of the top-Higgs coupling, described in terms of a coupling modifier κ_t relative to the Standard Model. In parallel, the difficult search for single top quark production in association with a Higgs boson (tH) [5] was performed. Both analyses used the data accumulated by the CMS experiment at the LHC up to 2016 and applied machine learning methods to increase their sensitivity. Due to overlap in the selected events, a combination of the results obtained by the two analyses is not possible. However, such a combination would be beneficial, because tH production is sensitive to the sign of the top-Higgs coupling modifier κ_t relative to the gauge boson coupling modifier κ_V but lacks general sensitivity due to the small production cross section compared to that of $t\bar{t}H$ production. Until now, a measurement of $\mathcal{CP}$ mixing of the Higgs boson, that could be a hint to extended physics in the Higgs sector, has not been performed in the $H \to b\bar{b}$ decay channel at the CMS experiment.

In this thesis, the top-Higgs coupling modifier κ_t is measured in tH and $t\bar{t}H$ production with the Higgs boson decaying into a bottom quark pair, both in a Standard Model and non-Standard Model coupling scenario. The used data accounts for a total of $138\,\mathrm{fb}^{-1}$ collected by the CMS experiment during the Run 2 data taking period between 2016 and 2018. The measurement in the non-Standard Model scenario introduces mixing with a pseudoscalar Higgs boson, predicted by many Higgs-doublet models. This measurement is then combined with the CMS results in the $H \to ZZ/\gamma\gamma$ [6] and multilepton Higgs boson decay channel [7], yielding world-leading constraints on $\mathcal{CP}$ mixing of the Higgs boson. Additionally upper limits on tH production are derived. These measurements are based on a new (t)tH analysis, a combination of the previously separate tH and $t\bar{t}H$ analyses improved with state-of-the-art developments in the reconstruction and classification of high energy particle collisions.

The current world average of the top quark decay width is based on three measurements, performed by the D0 [8], CMS [9] and ATLAS Collaboration [10]. The methods used in these measurements require model-dependent assumptions on the top quark branching ratio $\mathrm{BR}_{t\to Wb}$ or are severely limited by systematical uncertainties. For the measurement of the top quark width described in this thesis, a novel approach [11] is adopted. The direct measurement of the top quark width in Wb scattering ought to reduce systematic uncertainties drastically while remaining fully model-independent. A blueprint on how such a first measurement can be performed with the CMS experiment is given. Due to the small systematical uncertainties, this approach is going to remain relevant in the future with the currently ongoing Run 3 data taking period and the HL-LHC era afterward.

This thesis is structured as follows: In chapter 2, the theoretical foundation and the experimental setup is described. The subsequent chapter 3 contains a description of the (t)tH analysis, used to measure the top-Higgs coupling and for the search of tH production. The general structure of the measurement of the top quark decay width is given in chapter 4. Chapter 5 gives some insight into the currently ongoing development of b-tagging in Run 3, which is important for the acquisition of data to study top quarks and the Higgs boson in more detail. The summary in chapter 6 completes the thesis.

2 Foundations

2.1 The Standard Model

The Standard Model of particle physics (SM) is a theory describing three fundamental forces acting between elementary particles. Four spin-1 gauge bosons mediate the forces between twelve spin-$\frac{1}{2}$ fermions and among themselves. The fermions are divided into three leptons families (e, μ and τ) and three generations of quark pairs. Each lepton family is a pair of an electrically charged lepton ℓ^- and the corresponding uncharged neutrino ν_ℓ. Each of the three quark generations consists of an up-type and a down-type quark. Up-type quarks (u, c, t) carry an electrical charge of $+\frac{2}{3}\,e$ and down-type quarks (d, s, b) carry an electrical charge of $-\frac{1}{3}\,e$. For each fermion there is also an antifermion with identical quantum numbers but opposite electrical charge.

There are three fundamental forces described by the SM: The electromagnetic force is described by quantum electrodynamics (QED). Its mediator particle is the photon γ, coupling to all electrically charged particles. The coupling constant is $\alpha \approx \frac{1}{137}$ in the asymptotic limit of zero energy and increases for higher energy scales.

The weak force is mediated by a neutral Z boson, coupling to all fermions, and two oppositely charged W bosons, coupling to weak isospin doublets consisting of the left-handed component of a charged lepton ℓ and the corresponding neutrino ν_ℓ or a pair of up-type and down-type quark. Since in the SM W bosons couple only to left-handed fermions and right-handed antifermions, the charged current of the weak interaction is maximally parity violating. The mass eigenstates of quarks differ from their flavor eigenstates which are coupling to the W bosons. By convention the mass eigenstate q is set equal to the flavor eigenstate q′ for up-type quarks and the flavor eigenstates of the down-type quarks are described by a rotation with the Cabibbo–Kobayashi–Maskawa matrix V_{CKM} [12]:

$$\underbrace{\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix}}_{\text{flavor}} = \underbrace{\begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix}}_{=V_{\mathrm{CKM}}} \cdot \underbrace{\begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix}}_{\text{mass}} \tag{2.1}$$

The matrix elements $V_{\mathrm{q}\tilde{\mathrm{q}}}$ are determined experimentally. In the Wolfenstein parametrization [13] the matrix is approximated as:

$$V_{\mathrm{CKM}} = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right) \tag{2.2}$$

with the observed parameters $\lambda = 0.224\,53 \pm 0.000\,44$, $A = 0.836 \pm 0.015$, $\rho = 0.122\,^{+0.018}_{-0.017}$ and $\eta = 0.355\,^{+0.012}_{-0.011}$. In this prescription, it is clearly visible that transitions within a quark generation are preferred and that mixing between the first and second generation is stronger than mixing between the second and third generation. Mixing between first and third generation is suppressed even stronger.

Electromagnetic and weak force can be unified into the electroweak force. An effect of this is the coupling between the photon, W and Z boson. By electroweak symmetry breaking, described in section 2.1.1, the massless W and Z gauge bosons receive their mass.

The strong force acts between color-charged particles, e.g. quarks, and is mediated by gluons g, which carry color charge themselves. There are three colors: red, green and blue, as well as their corresponding anticolors. With the gluons carrying color charge, they can couple with themselves, resulting in tri- and quad-gluon vertices. The coupling strength of quantum chromodynamics (QCD) α_{s} increases with lower energy scales. The consequence is, that only color-neutral particles are observed, which is referred to as color confinement. Quarks form color-neutral bound states as mesons, a quark-antiquark pair with opposite color-charges, and baryons, three (anti)quarks with three different (anti)colors. For high energy scales the coupling α_{s} becomes small and quarks behave like free particles. This is called asymptotic freedom.

An overview over the particles of the SM and their properties is given in figure 2.1. In figure 2.2 the vertices of Feynman diagrams [14] included in the SM, excluding those containing a Higgs boson, are shown.

2.1.1 The Brout-Englert-Higgs mechanism and the Higgs Boson

In theory gauge bosons are massless. The observation of massive W and Z bosons contradict this prediction. To explain massive gauge bosons in the SM a new scalar field $\Phi = (\Phi^+, \Phi^0)^\intercal$ is introduced in the Brout-Englert-Higgs mechanism [13] with the Lagrangian $\mathcal{L}_{\mathrm{Higgs}}$:

$$\mathcal{L}_{\mathrm{Higgs}} = (D_\nu \Phi)^\dagger (D^\nu \Phi) - V(\Phi) \tag{2.3}$$

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda |\Phi^\dagger \Phi|^2 \tag{2.4}$$

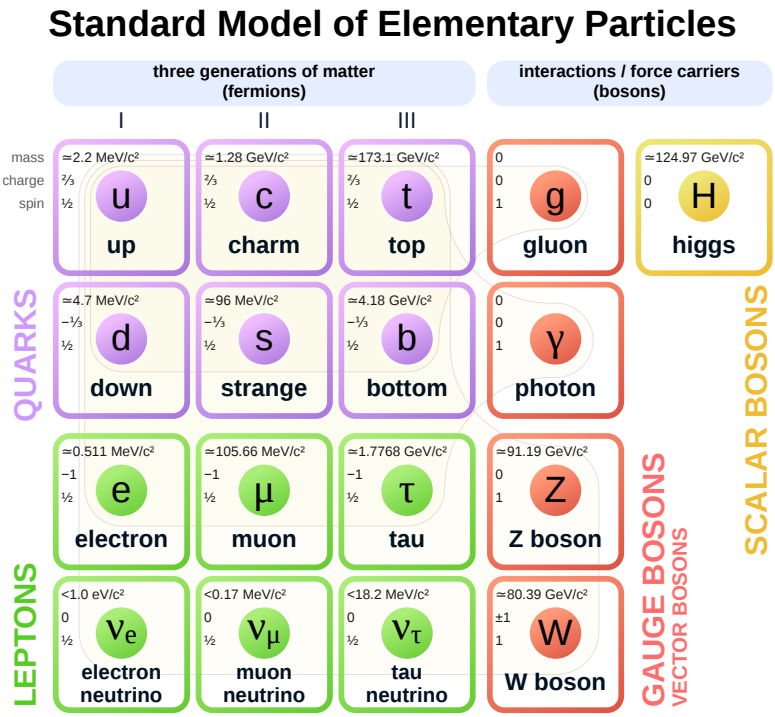


Figure 2.1: The elementary particles in the Standard Model [15] grouped by the type of particle. The numbers in each box are from top to bottom the mass, electrical charge in units of elementary charges and spin of the particle. The three left columns are the three fermion families/quark generations with $s = \frac{1}{2}$, with the quarks on the upper two rows in purple and the leptons on the lower two rows in green. The forth column in red contains the gauge bosons with spin $s = 1$, which are exchanged in the three fundamental interactions. The fifth column completes the SM with the Higgs boson.

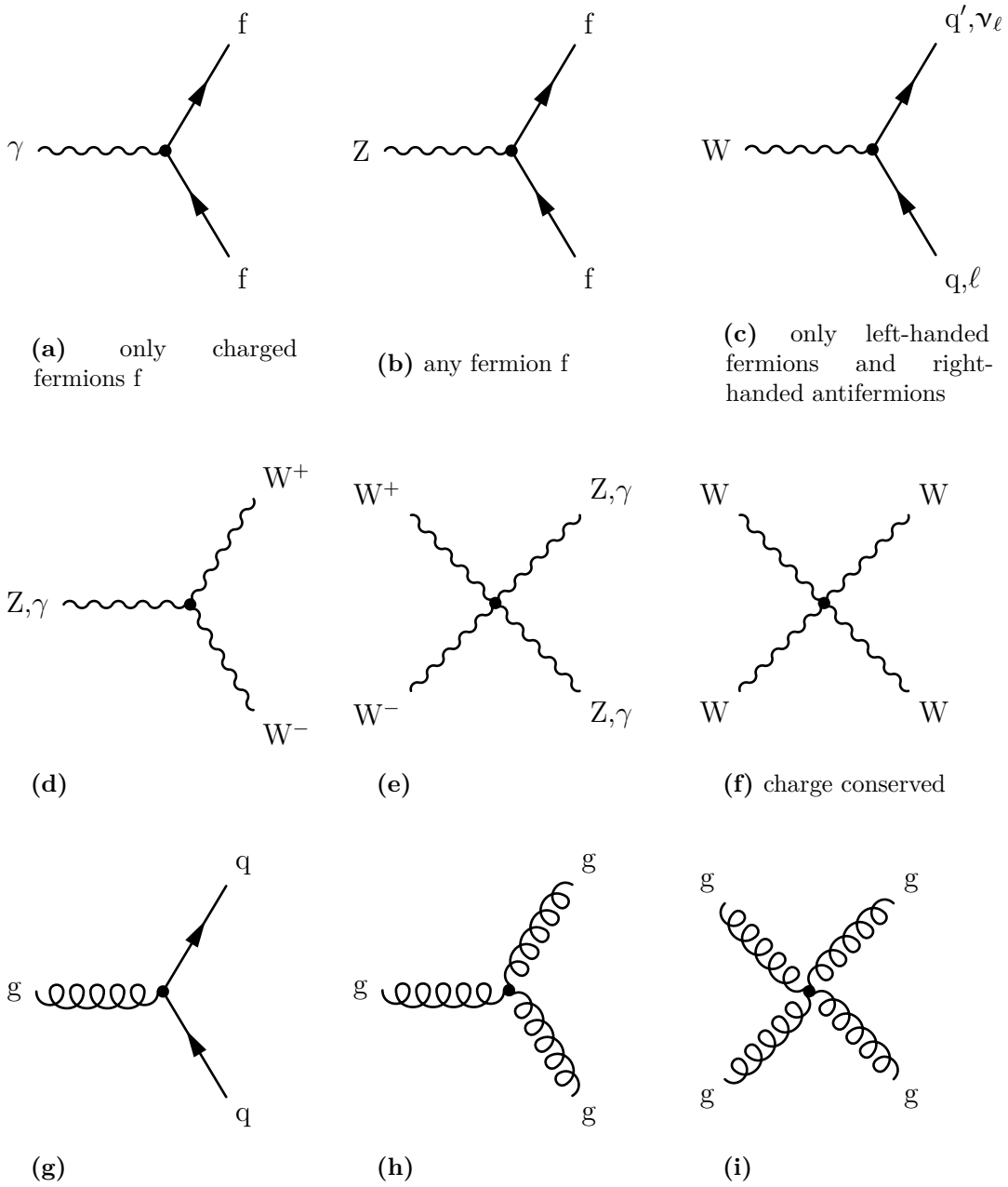


Figure 2.2: Standard Model of particle physics vertices for interactions mediated by gauge bosons. The first row are couplings of the electroweak force with fermions. The second row shows the couplings between gauge bosons of the electroweak force. The third row shows the couplings of quantum chromodynamics (QCD).

Through electroweak symmetry breaking masses are introduced to the W and Z boson. Symmetry breaking happens, when the ground state has fewer symmetries than the excited state. This is the case for $\mu^2 < 0$.

The Higgs field is constrained by massless photons to

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \tag{2.5}$$

with the vacuum expectation value $v = \sqrt{\frac{-\mu^2}{\lambda}} \approx 246\,\text{GeV}$ and an excitation of the field h describing a new particle, the Higgs boson H.

Fermion masses can be introduced through a coupling with the Higgs field h, called Yukawa coupling [16]:

$$\mathcal{L}_{\text{Yukawa}} = -\underbrace{\frac{y_\text{f}\, v}{\sqrt{2}}}_{m_\text{f}} \bar{\text{f}}\text{f} - \underbrace{\frac{y_\text{f}}{\sqrt{2}}}_{g_\text{Hff}} h\bar{\text{f}}\text{f} \tag{2.6}$$

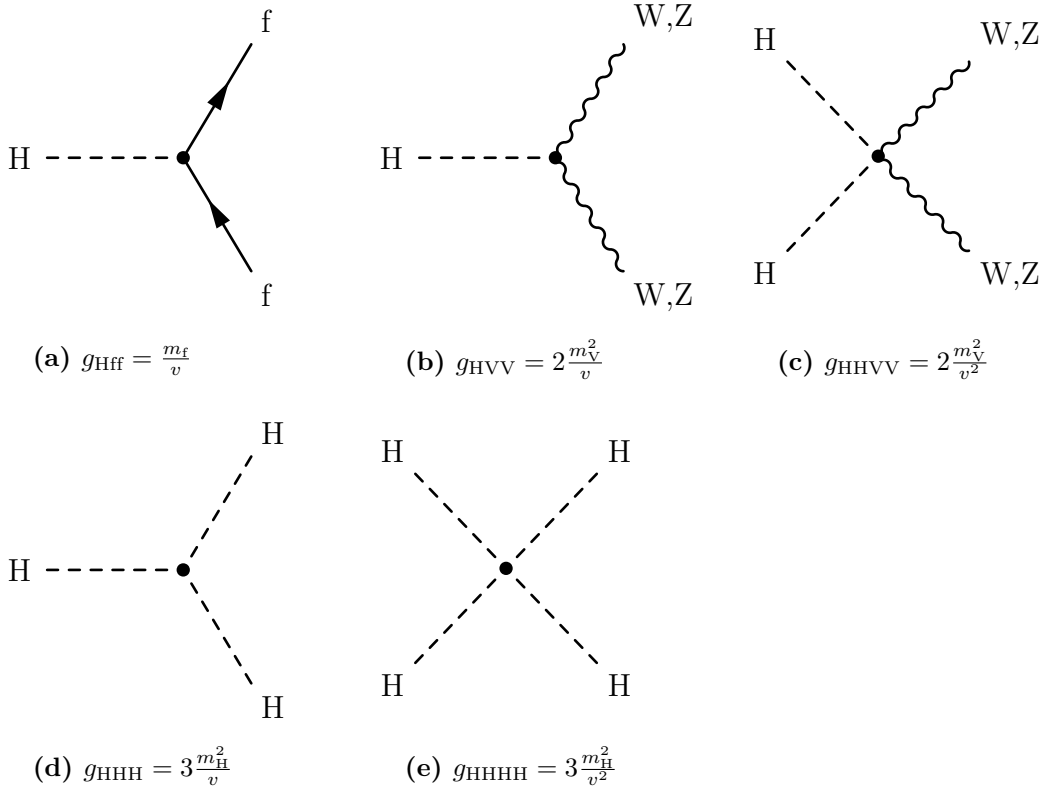
The coupling strength of the Higgs field h to fermions f is then given by $g_\text{Hff} = \frac{y_\text{f}}{\sqrt{2}} = \frac{m_\text{f}}{v}$. In figure 2.3 all interactions of the Higgs boson including the self-couplings and the coupling strengths are shown. A summary of the Higgs coupling measurements performed at the CMS experiment is given in figure 2.4.

At the Large Hadron Collider (LHC), Higgs bosons are mainly produced in gluon-gluon fusion (ggF) via a quark loop dominated by top quarks and vector boson fusion (VBF). They can also be radiated off by a W or Z boson or produced in association with top quarks. The dominant decay of the Higgs boson [17] with a mass of $m_\text{H} = (125.18 \pm 0.16)\,\text{GeV}$ is into a pair of bottom quarks $\text{b}\bar{\text{b}}$, followed by the decay into two W bosons. Higgs boson decays into four leptons and into photons produce a distinctive signature in the detector.

2.1.2 The Top quark

With a mass of $m_\text{t} = (173.0 \pm 0.4)\,\text{GeV}$, the top quark t is the heaviest particle in the SM. Its mean lifetime of $\tau \approx 5 \times 10^{-25}\,\text{s}$ [13] is one order of magnitude smaller than the timescale of hadronization. As a result the top quark does not live long enough to form bound states. Due to the suppressed mixing of the third generation ($V_\text{tb} \approx 1$), top quarks decay nearly exclusively into a bottom quark and a W boson. The decay is classified by the subsequent decay of the W boson into leptons or quarks.

The production of top quark pairs is a common process at the LHC. The cross



(a) $g_{\mathrm{Hff}} = \frac{m_{\mathrm{f}}}{v}$ **(b)** $g_{\mathrm{HVV}} = 2\frac{m_{\mathrm{V}}^2}{v}$ **(c)** $g_{\mathrm{HHVV}} = 2\frac{m_{\mathrm{V}}^2}{v^2}$

(d) $g_{\mathrm{HHH}} = 3\frac{m_{\mathrm{H}}^2}{v}$ **(e)** $g_{\mathrm{HHHH}} = 3\frac{m_{\mathrm{H}}^2}{v^2}$

Figure 2.3: Standard Model vertices and coupling strengths for the Higgs coupling to massive fermions, vector bosons (W, Z) and Higgs self-interaction.

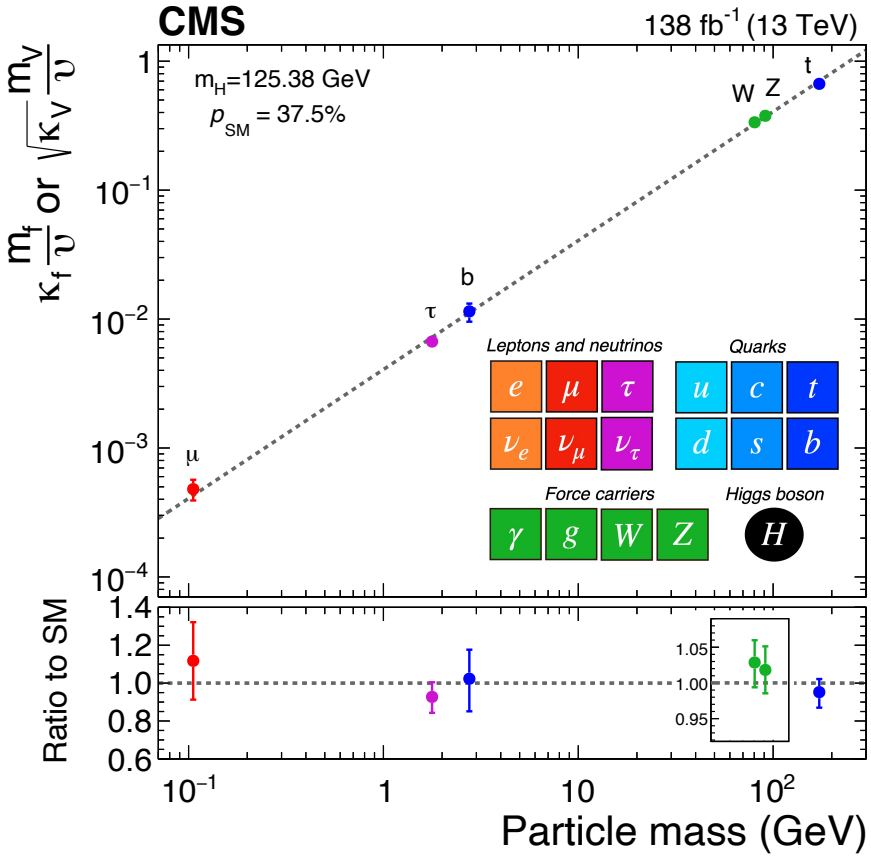


Figure 2.4: Summary of Higgs coupling measurements reported by the CMS experiment until 2022 [18]. The coupling strength to vector bosons, second and third generation fermions is shown in different colors. The measured couplings agree with the SM prediction over more than three orders of magnitude in particle mass.

section is:

$$\sigma_{\mathrm{t\bar{t}}} = (831.76\,^{+19.77}_{-29.20}\,(\mathrm{scale}) \pm 35.06\,(\mathrm{PDF}{+}\alpha_\mathrm{s})\,^{+23.18}_{-22.45}\,(\mathrm{mass}))\,\mathrm{pb} \tag{2.7}$$

for a top quark mass of 172.5 GeV at next-to-next-to-leading order (NNLO) accuracy and a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$. Some exemplary production modes are shown in the upper row of figure 2.5. Top quark pair production is categorized by the decay of the two W bosons into dileptonic, semileptonic and full hadronic decays.

A less common mode of producing top quarks is single top quark production, because it relies on a weak interaction to produce a top and bottom quark or convert a bottom quark to a top quark. The lower row of figure 2.5 depicts the different channels of single top quark production. The s and t channel are named after the Mandelstam variables [19], where s can be interpreted as the square of the center-of-mass energy and t as the square of the four-momentum transfer. s can be used to describe the resonant production in s channel Feynman diagrams, while t is used to describe a scattering process. In tW-associated production the top quark is produced by radiating off a W boson from a bottom quark. The cross sections at next-to-leading order (NLO) [20] accuracy are:

$$\sigma_{s\text{ channel}} = (10.32\,^{+0.40}_{-0.36})\,\mathrm{pb} \tag{2.8}$$

$$\sigma_{t\text{ channel}} = (216.99\,^{+9.04}_{-7.71})\,\mathrm{pb} \tag{2.9}$$

$$\sigma_{\mathrm{tW}} = (71.7\,^{+1.80}_{-1.80}\,(\mathrm{scale})\,^{+3.40}_{-3.40}\,(\mathrm{PDF}))\,\mathrm{pb} \tag{2.10}$$

at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ and a top quark mass of 172.5 GeV.

Due to the large mass, top quarks couple strongly with Higgs bosons. This makes them interesting for studying Higgs bosons or probing new theories at high energies. Since the top quark does not form bound states, its spin is conserved in the decay products and polarization studies can be conducted.

2.1.3 Physics beyond the Standard Model

In 2012 the last particle of the SM, the Higgs boson was discovered by the ATLAS and CMS experiment [1, 2]. Still many physical phenomena are not explained by the SM. For example neutrinos are massless in the SM, which is proved to be wrong by the observation of neutrino oscillations [3]. Gravity is also not included and becomes important before reaching the Planck scale $\Lambda_\mathrm{P} = 1.22 \times 10^{19}\,\mathrm{GeV}$, where effects of gravity have to be considered. The SM does not contain a dark matter candidate to sufficiently explain the large fraction of dark matter in the universe [21] and the $\mathcal{CP}$ violation in the SM is not large enough to explain the shortage of antimatter observed in the universe.

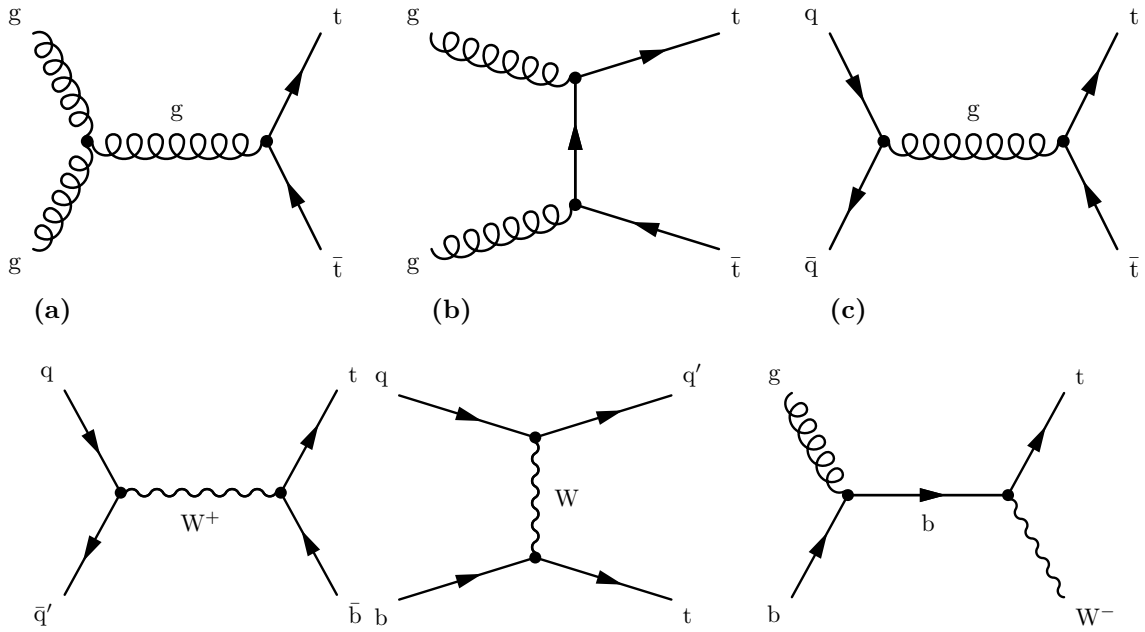


Figure 2.5: Top quark pair production from two gluons (a, b), a quark-antiquark pair (c) and single top quark production in different channels (d-f).

Currently the SM is probed on a range of different fronts. High precision measurements search for deviations from SM predictions. The nature of neutrinos is investigated by different neutrino experiments. Dedicated dark matter experiments try to directly observe dark matter particles in large volumes of matter. However until now, no sufficient direct evidence of physics beyond the Standard Model (BSM) has been observed.

2.2 The Large Hadron Collider

The LHC [22] located at the European Organization for Nuclear Research (CERN) complex in Geneva is a hadron collider for protons and heavy ions. The beam pipe is built in the 26.7 km long ring tunnel of its predecessor, the Large Electron Positron Collider (LEP). In the Run 2 measurement campaign from 2015 to 2018 protons were accelerated to an energy of 6.5 TeV in the two opposing beam pipes and collided with a center-of-mass-energy of $\sqrt{s} = 13\,\text{TeV}$. For the Run 3 data taking period, the center-of-mass-energy for proton-proton collisions was further increased to $\sqrt{s} = 13.6\,\text{TeV}$.

The LHC is a storage ring and the last accelerator in a chain of accelerators of the CERN accelerator complex, shown in figure 2.6. Each of the two beam pipes has eight radio frequency cavities to accelerate particles. 1232 superconducting dipole magnets are utilized to keep the beam inside the beam pipe. For focusing, 392 quadrupole magnets and higher order magnets are used.

An important characteristic parameter of a collider experiment is the luminosity. The instantaneous luminosity L is the number of collisions per time normalized to the cross section:

$$L = \frac{1}{\sigma_{\mathrm{pp}}} \frac{\mathrm{d}N}{\mathrm{d}t} \tag{2.11}$$

with the proton-proton cross section σ_{pp} and number of collisions N. In 2022 the instantaneous luminosity at the LHC surpassed $2.5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [23].

By integrating over time, one gets the integrated luminosity L_{int} which can be used to describe the total number of collisions:

$$L_{\mathrm{int}} = \int L(t)\,\mathrm{d}t \tag{2.12}$$

Four larger experiments are located at four beam crossing points of the LHC: A Large Ion Collider Experiment (ALICE), A Toroidal LHC ApparatuS (ATLAS), Compact Muon Solenoid (CMS) and Large Hadron Collider beauty (LHCb). ATLAS and CMS are general-purpose detectors used for a wide variety of analyses. ALICE is used to investigate quark-gluon-plasma in heavy ion collisions. At LHCb flavor physics is studied using B-mesons produced in forward direction.

2.3 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) detector is an almost 4π spatial angle general-purpose detector located at one of the four collision points of the LHC. It is built out of four subdetector systems cylinder symmetrical around the interaction point, divided into a barrel part and two endcap sections. The barrel part houses the superconducting solenoid producing a strong magnetic field and providing structural support. A drawing of the detector is shown in figure 2.7.

Polar coordinates are used to describe the trajectories of particles inside the detector. The origin is the nominal collision point with the x axis pointing inwards to the center of the LHC and the y axis pointing upwards to the sky. The z axis points in beam direction so that the right-handed coordinate system is completed. When

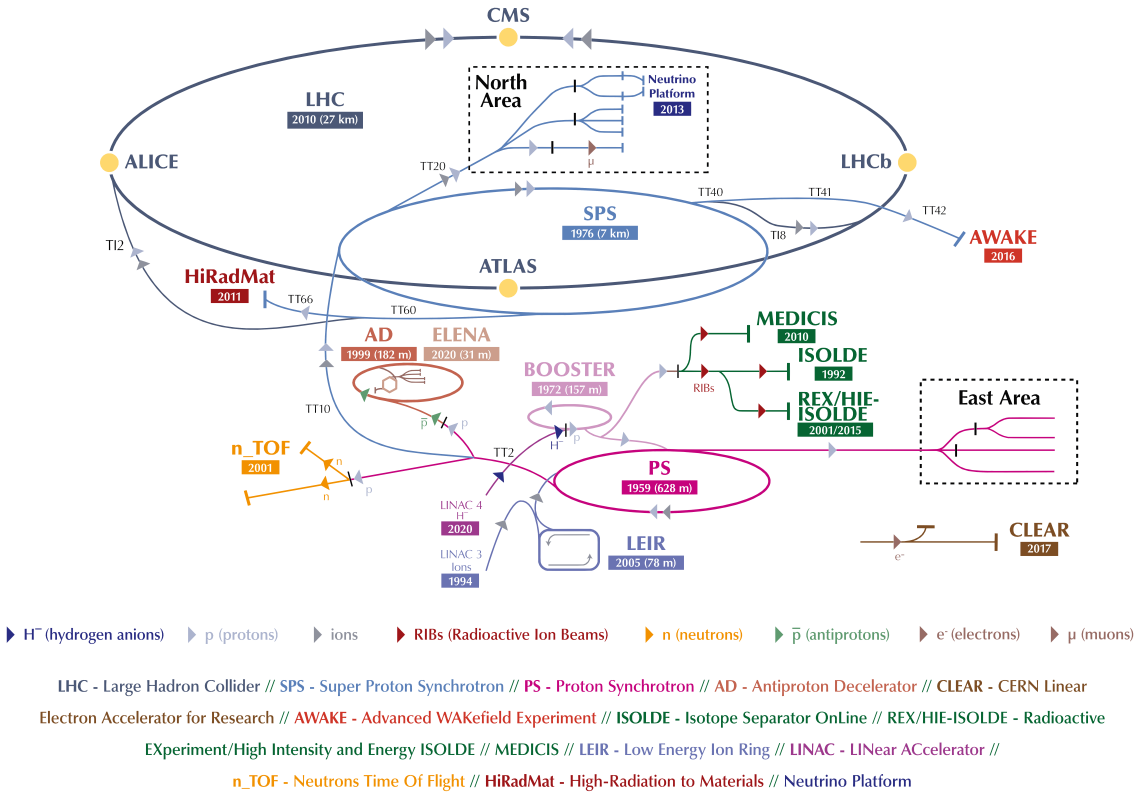


Figure 2.6: Schematic view of the CERN accelerator complex [24] in January 2022 with a chain of accelerators used to fill the Large Hadron Collider (LHC). The accelerator chain for protons is LINAC4 → PSB (Booster) → PS → SPS → LHC.

transformed into polar coordinates, the azimuthal angel φ describes the angle towards the x axis in the $x-y$ plane. The polar angle Θ, which is the angle towards the z axis, is often replaced by the pseudorapidity η, defined as:

$$\eta = -\ln\left(\tan\frac{\Theta}{2}\right) \tag{2.13}$$

For massless particles or light particles with high energies $p \gg m$ the pseudorapidity is equivalent to the rapidity y, which is defined as:

$$y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z} \tag{2.14}$$

Distances between two objects i and j in the detector are often expressed in terms of ΔR:

$$\Delta R = \sqrt{(\eta_i-\eta_j)^2+(\varphi_i-\varphi_j)^2} \tag{2.15}$$

ΔR and differences in rapidity Δy are Lorentz invariant under boosts along the beam axis. In this coordinate system, the four-momentum of a particle can be fully described by its energy E, azimuthal angle φ, pseudorapidity η and transverse momentum p_T. The transverse momentum p_T is the momentum in the $x-y$ plane and can be directly measured for charged particles in the tracker, described in section 2.3.1.

Due to the nature of protons, the momentum of the colliding constituents of the protons, the partons, before collision is unknown. But because the partons collide heads on along the z axis, the transverse momentum p_T before the collision is close to zero. By summing up the transverse momenta of all observed particles after collision, it is possible to measure the missing transverse momentum $p_\mathrm{T}^\mathrm{miss}$, caused by particles leaving the detector undetected.

2.3.1 The tracker

Tracks of charged particles are observed by the innermost detector of CMS, an all silicon tracker [26–28]. It is split into the pixel, inner tracker and outer tracker. The modules cover the η range of $|\eta| < 2.5$. Due to the magnetic field, tracks of charged particles are curved which allows measuring the charge polarity and transverse momentum p_T with the curvature of the tracks.

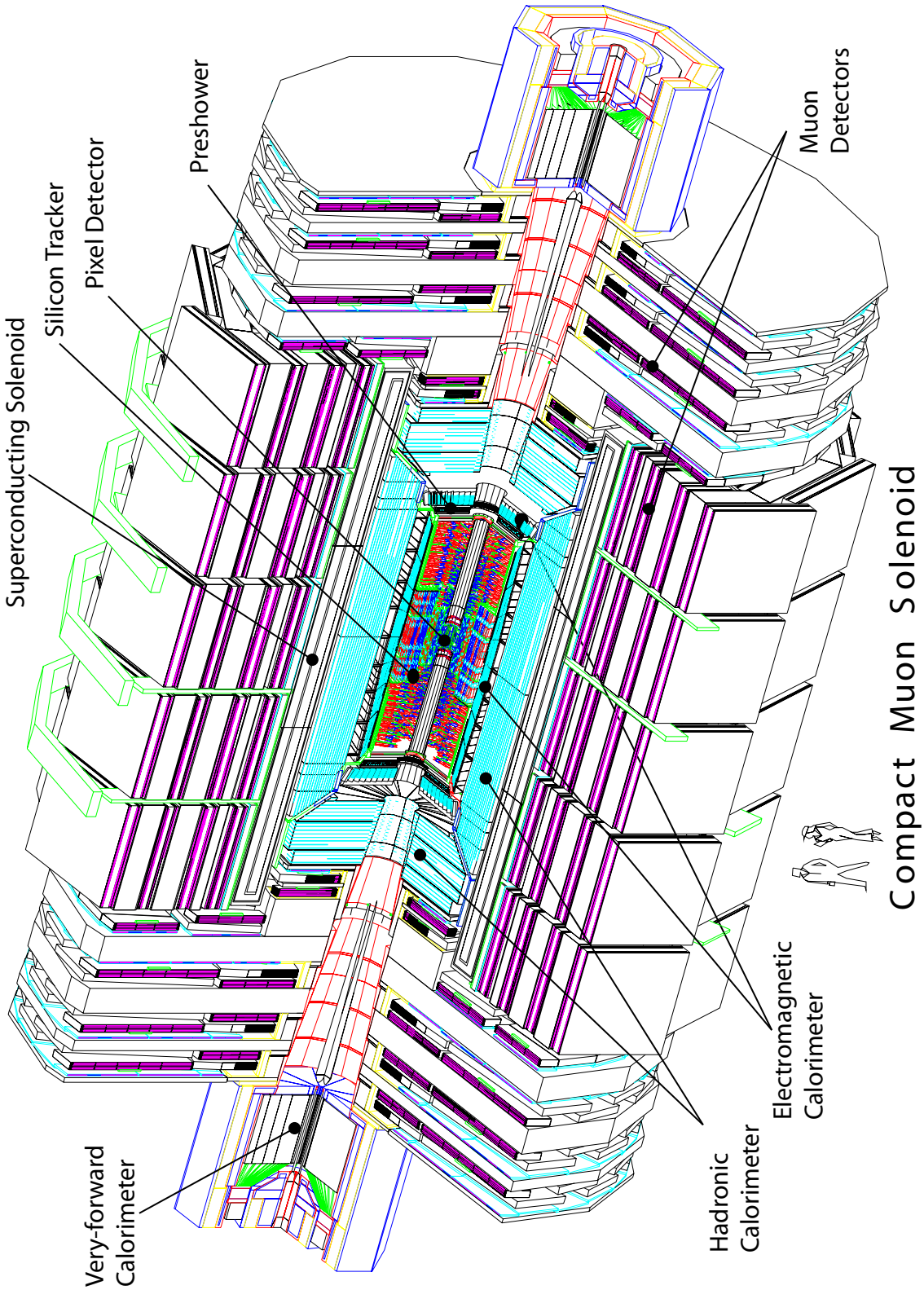


Figure 2.7: Drawing of CMS detector [25] with annotated subdetectors. The beam spot in the center surrounded by the tracker, calorimeters, solenoid and muon system.

The p_{T} resolution of the CMS tracker [29] for isolated tracks in the central part is given by:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = \frac{0.000\,15}{\mathrm{GeV}} \cdot p_{\mathrm{T}} \oplus 0.005 \tag{2.16}$$

The pixel tracker, which is segmented into 100 µm × 150 µm pixels, is used to reconstruct primary and secondary vertices and allows measuring the instantaneous luminosity. In addition, it is used to seed the track reconstruction algorithms, which utilize the full tracker and in case of muons also the muon system to reconstruct tracks and measure the transverse momentum.

2.3.2 The calorimeters

Two calorimeters are used to measure the energies of particles in the CMS detector. The Electromagnetic Calorimeter (ECAL) [30, 31] is built as a homogeneous calorimeter out of 82 728 $PbWO_4$ crystals. It is used to measure the energy of electromagnetic showers caused by electrons/positrons and photons up to $|\eta| < 3.0$. The crystals act as absorber and scintillator. With a length of around 25 times the radiation length, the full electromagnetic shower is contained inside the ECAL.

The Hadron Calorimeter (HCAL) [32, 33] is built as sampling calorimeter with the brass acting as absorber and the plastic in between used as scintillator. In the barrel part brass is used as absorber while in the endcaps it is substituted with steel. Like the ECAL it covers the range up to $|\eta| < 3.0$ with extended coverage by the forward calorimeter up to $|\eta| < 5$. At $\eta = 0$ the HCAL is only 5.82 nuclear interaction lengths λ deep, which means that the hadronic showers produced by hadrons are not always fully contained inside the calorimeter. Because of this, the HCAL in the central part is supplemented by a tail catcher after the solenoid to detect hadronic leakage into the muon system.

The energy resolution of the ECAL [30] and HCAL [32] measured in test beams at $\eta = 0$ is:

$$\frac{\sigma_{\mathrm{ECAL}}}{E} = \frac{0.027}{\sqrt{E/\mathrm{GeV}}} \oplus \frac{0.210}{E/\mathrm{GeV}} \oplus 0.0055 \tag{2.17}$$

$$\frac{\sigma_{\mathrm{HCAL}}}{E} = \frac{0.70}{\sqrt{E/\mathrm{GeV}}} \oplus 0.095 \tag{2.18}$$

2.3.3 The solenoid

Around the tracker and calorimeters is the superconducting solenoid [34, 35] producing a magnetic field of up to 3.8 T inside the coil. For the titanium alloy to reach superconductivity it has to be cooled with superfluid helium. The field outside the coil is guided by an iron return yoke, which also houses the muon system. An advantage with the coil around the calorimeter is that particles do not have to pass through the massive coil, causing undetected energy loss, before energy measurement, but the calorimeter is limited by the size of the solenoid.

2.3.4 The muon system

Three different systems together make the muon system [36, 37]. In the barrel region drift tubes (DTs) and resistive plate chambers (RPCs) are placed between the iron yoke, while in the endcaps cathode strip chambers (CSCs) are used instead of DTs. The RPCs with faster readout are used in the first trigger stage to detect the occurrence of muons. For the reconstruction all three systems are used together with the tracker system.

2.3.5 The trigger system

To select interesting events from the collisions happening every 25 ns, a two-stage trigger system [38–40] is employed at the CMS experiment. In the Level 1 trigger (L1) system, the information from the calorimeter and RPCs from the muon system is used to reject common events with low energies and without muons. When a L1 is fired, the subdetectors are readout and a fast online event reconstruction is performed. Only if a subsequent High Level Trigger (HLT) is fired based on the online reconstruction, the event is written to storage for further analysis. Figure 2.8 shows the stages of the CMS trigger system.

2.4 Object reconstruction

2.4.1 Particle identification and reconstruction

Particles in the detector are identified and reconstructed by combining the signals from the different subdetectors. Electrons and positrons are reconstructed from the

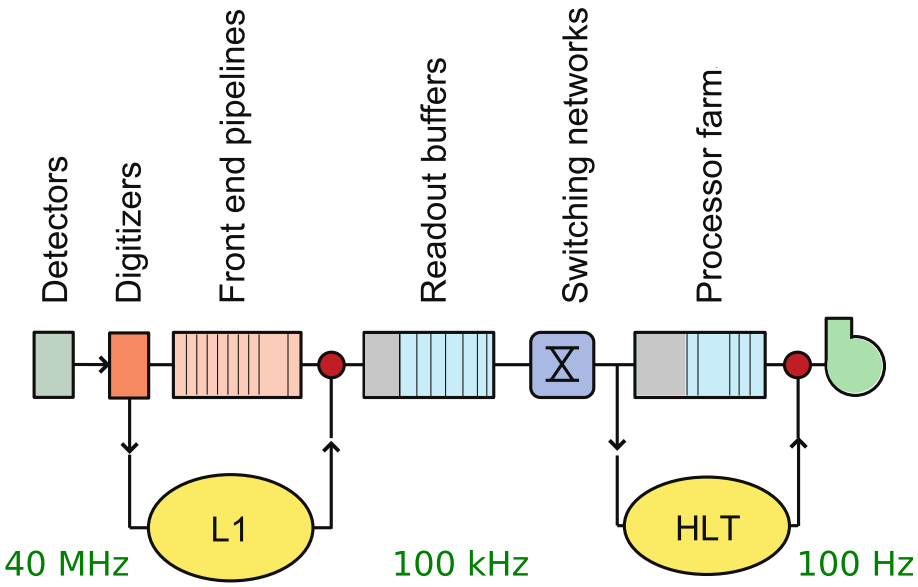


Figure 2.8: The two-stage trigger system of the CMS detector adapted from Ref. [41]. The graphic is annotated with the event rate in each stage.

tracks in the tracker and entries in the ECAL. Since photons do not produce tracks in the tracker, they are reconstructed from ECAL hits without associated tracks. Hadrons are reconstructed from the entries in the ECAL, HCAL and in case of a charged hadron, tracks in the tracker. Muon tracks are reconstructed from the muon system and the matched tracks from the tracker. Neutrinos leave the detector without interaction and can only be observed as missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$.

In the CMS experiment the energy measurement from the calorimeters and p_{T} measurement of the tracker is combined with a method called Particle Flow (PF)[42]. The p_{T} resolution of the tracker is higher for particles with lower energies while the resolution of the calorimeters increases with higher energies. By combining both, the lower resolution of the HCAL can be mitigated by the higher resolution of the tracker.

2.4.2 Jets

After production, color charged quark and gluons undergo a chain of processes starting with parton shower and hadronization, followed up by the decay of the hadrons and showering in the detector. The resulting signature in the detector is a cone of tracks and hits in the calorimeters. For event interpretation these are clustered into objects called jets using a cluster algorithm. Certain requirements have to be fulfilled by the algorithm to result in jets that can be well described by

theory. The algorithm needs to be infrared and collinear safe, which means it is robust against the soft and collinear radiation of quarks and gluons respectively.

At the CMS experiment the sequential, infrared and collinear safe anti-kt algorithm [43] is used for jet clustering. It belongs to a group of similar defined cluster algorithms with these properties. Two distance measures are defined:

$$d_{ij} = \min\left({p_{\mathrm{T}}}_i^{2n}, {p_{\mathrm{T}}}_j^{2n}\right) \frac{\Delta R_{ij}^2}{R^2} \tag{2.19}$$

$$d_{iB} = {p_{\mathrm{T}}}_i^{2n} \tag{2.20}$$

with $n = -1$ for anti-kt, distance between objects i and j ΔR_{ij} and a parameter R which corresponds to the radius of the clustered jets. d_{ij} is the distance measure between two objects i and j, while d_{iB} is the distance measure between object i and the beam. At the CMS experiment two different jet sizes R are used: $R = 0.4$ for normal jets (AK4) and $R = 0.8$ for wider jets (AK8).

First the distances of all objects i to the beam d_{iB} and distances between all pairs of objects i and j d_{ij} are calculated. If the smallest distance is between two objects i, j, they are merged by adding their four-momenta. This is repeated until the closest distance is the distance of an object i to the beam pipe d_{iB}. The remaining merged objects are the clustered jets.

With the anti-kt algorithm ($n = -1$) objects with higher transverse momenta p_{T} are clustered first. Two variations of this algorithm exist: The inclusive kt algorithm with $n = 1$, which clusters objects with lower p_{T} first, and the Cambridge/Aachen algorithm with $n = 0$, which does not take the transverse momenta into account and clusters objects solely depending on their distance in the detector.

2.4.3 b-tagging

Many interesting processes including Higgs bosons or top quarks produce bottom quarks. This raises the need to distinguish jets originating from bottom quarks against those produced by charm quarks or lighter partons (u, d, s and g). Luckily the jets originating from bottom quarks have some properties different from other jets: The B mesons formed by bottom quarks have a lifetime long enough to travel a distance of 1 mm to 2 mm from the primary interaction before they decay and are reconstructed as secondary vertices from displaced tracks. Due to the large mass of the B mesons it carries a significant fraction of the jets energy, called hard fragmentation.

At the CMS experiment, heavy flavor tagging on AK4 jets among others is performed using a tagger called DeepJet [44]. It is based on a neural network trained to classify

jets into b, c and light jet classes using kinematic, substructure and secondary vertex properties of the jet. The DeepJet classification value for b-tagging is calculated as the sum of the neural network output for different b classes. Three different working points, loose, medium and tight, are set as cuts on the DeepJet classification value with 10 %, 1 % and 0.1 % misidentification probability of light jets as b jets, respectively.

2.5 Simulated data

Particle collisions are simulated using the Monte Carlo (MC) method. First the matrix element of the hard scattering process is calculated. Using the matrix element $M_{i_1,i_2\to f}$ describing the transition from initial partons i_1 and i_2 to final state f, the cross section can be calculated:

$$\sigma_{\mathrm{pp}\to f} = \sum_{i_1,i_2} \int_{x_1,x_2=0}^{1} f_{i_1}(x_1) f_{i_2}(x_2) \cdot |M_{i_1,i_2\to f}|^2 \mathrm{d}x_1 \mathrm{d}x_2 \tag{2.21}$$

with the parton distribution function $f_i(x)$ for a parton i carrying the fraction of proton momentum x. The matrix element can be calculated with tools such as MadGraph5 [45] or Herwig [46]. The parton distribution functions are derived experimentally.

In the next step the hadron shower is simulated. It includes the simulation of parton shower and hadronization followed by decay of hadrons. Underlying event, consisting of beam remnants, multi parton interaction, initial state and final state radiation, is also calculated matching the expected conditions in the detector. The particles are then fed into a Geant4 [47] detector simulation which simulates all interaction of particles with matter and the following readout electronics. From here on the simulated data is processed the same way as real data.

2.6 Statistical methods

2.6.1 Maximum likelihood estimation

To estimate the model parameters $\vec{a}$ so they optimally describe the data $\vec{x}_i$, the likelihood function L is maximized. The likelihood L is defined as:

$$L = \prod_{i}^{N} f\left(\vec{x}_i|\vec{a}\right) \tag{2.22}$$

with the probability density function $f(\vec{x}_i|\vec{a})$ with model parameters $\vec{a}$ at the data point $\vec{x}_i$.

Often the negative log likelihood (NLL) $-2\ln L$ is minimized, because most optimization algorithms are optimized for minima and with the logarithm the product in the likelihood formula becomes a simple sum:

$$-2\ln L = -2\sum_{i}^{N} \ln f\left(\vec{x}_i|\vec{a}\right) \tag{2.23}$$

The factor 2 is selected so that for a normal distribution the likelihood values are equivalent to the χ^2-estimation.

In a histogram where each bin follows a poisson distribution $f\left(k|\lambda\right) = \frac{\lambda^k \exp(-\lambda)}{k!}$, the binned likelihood is used:

$$L = \prod_{i}^{N_{\text{bins}}} \frac{\lambda^{n_i} \exp(-\lambda_i)}{n_i!} \tag{2.24}$$

with the number of bins N_{bins}, the observed number of events n_i and expected events λ_i in bin i.

2.6.2 Exclusion limits

Exclusion limits [48] are estimated by modeling the number of events as sum of expected background events b_i and expected signal events s_i scaled with the signal strength μ. A test statistic q_μ is defined as likelihood ratio of the binned likelihood from equation (2.24):

$$q_\mu = -2\ln \frac{L(\mu, \hat{\Theta}_\mu)}{L(\hat{\mu}, \hat{\Theta}_{\hat{\mu}})} \tag{2.25}$$

with $\hat{\mu}$ maximizing the likelihood L globally and the nuisance parameters $\hat{\Theta}_\mu$ maximizing the likelihood for a specific μ.

The test statistic q_μ is subject to statistical fluctuations in the data and p-values can be calculated:

$$p_{\text{s+b}} = \int_{q_{\mu_\text{c}}}^{\infty} g(q'_\mu|\mu = \mu_\text{c})\,\mathrm{d}q'_\mu \tag{2.26}$$

$$p_{\text{b}} = \int_{-\infty}^{q_{\mu_\text{c}}} g(q'_\mu|\mu = 0)\,\mathrm{d}q'_\mu \tag{2.27}$$

with a cut μ_c, the test statistic q_{μ_c} of the cut and the probability distribution of the test statistic $g(q_\mu)$ determined by repeating the calculation many times on fluctuating data obtained by MC simulation.

p_{s+b} describes the probability to observe a q_μ that is larger than q_{μ_c} for a given μ_c. This can be interpreted as the probability for a down fluctuation of the signal to appear background-like. To get the result for a specific significance α, μ_c is varied until $p_{s+b} = \alpha$. In that case signal strengths larger than μ_c can be excluded with a confidence level (CL) of $1 - \alpha$, because the probability for down fluctuation of the signal to appear background-like and thus falsely rejecting the signal hypothesis is α.

Because of problems with this interpretation in cases of small number of expected background events, where minor fluctuations of the data can lead to large significances, the fluctuations of the background are taken into account in the more conservative CL_s-approach. A definition of the confidence level is obtained as:

$$CL_s = \frac{p_{s+b}}{1 - p_b} \tag{2.28}$$

2.6.3 Machine learning

In particle physics often different machine learning techniques are used. They are applied in the classification of events or reconstruction of quantities. The most commonly used ones are boosted decision trees (BDTs) and neural networks. These models have numerous tunable parameters that are optimized by minimizing a cost function on a test dataset. Because of the high complexity of such models, they have to be validated and the performance thoroughly studied.

BDTs consist of many binary decision trees, which act as weak learners. A single binary decision tree is defined by nodes with cuts on variables, an example is depicted in figure 2.9. The selection of the variable x_i to cut on and the cut value c_i in each node are trainable parameters. Key figures of BDTs are the number of binary trees and their depth.

Neural networks are graphs built out of layers of neurons and interconnections between them. Each neuron receives weighted inputs that are summed up with an additional bias term and evaluated with an activation function. The output is then fed into the next layer or in case of the last layer, is interpreted as the network response. Activation functions are necessary to get non-linear responses. Otherwise

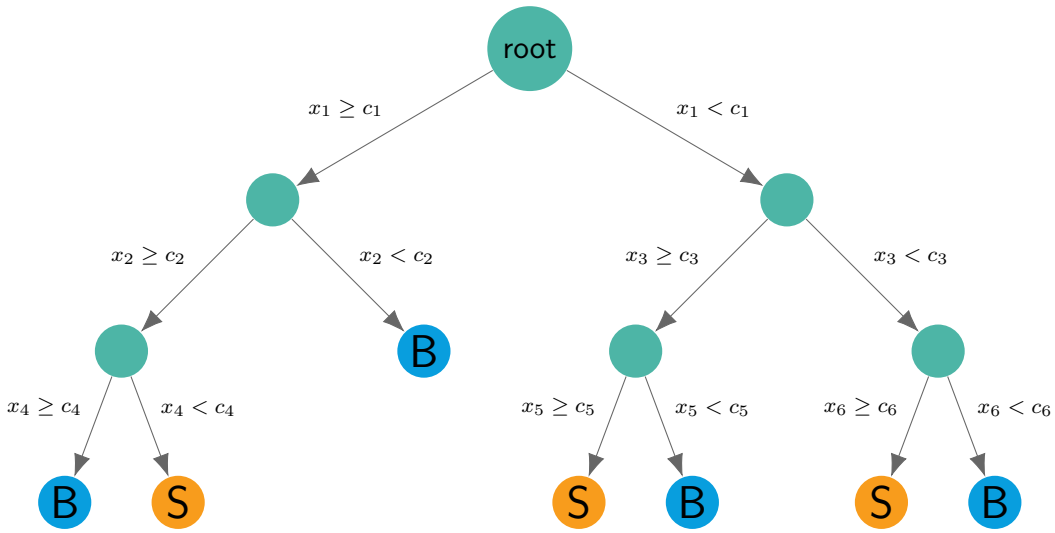


Figure 2.9: Binary decision tree with three layers depth. At each node a cut c_i on a variable x_i is applied to split the dataset. Output nodes are labeled with signal (S) or background (B) depending on the majority of the content.

commonly used activation functions are ReLu and sigmoid, defined as:

$$f_{\text{ReLU}}(x) = \max(0,\, x) \qquad \mathbb{R} \to [0,\, \infty) \tag{2.29}$$

$$f_{\text{sigmoid}}(x) = \frac{1}{1 + \exp(-x)} \qquad \mathbb{R} \to (0,\, 1) \tag{2.30}$$

The trainable parameters of neural networks are the weights of the interconnections and the value of the bias in each node. In figure 2.10 a neural network with four layers is shown.

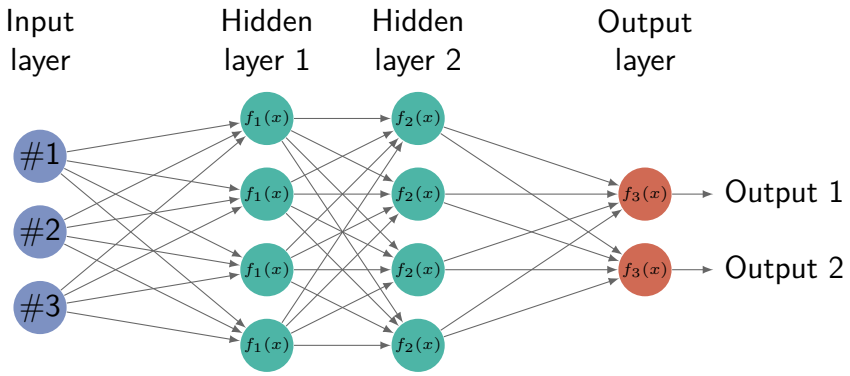


Figure 2.10: Neural network with four fully connected layers and two output nodes The value x put into the activation function $f_j(x)$ of node j is the weighted sum of all input values x_i and the bias term b: $x = \sum_i w_i x_i + b$.

3 The search for single top quark production in association with a Higgs boson and measurement of the top-Higgs coupling

The observation of top quark pair production in association with a Higgs boson ($t\bar{t}H$) in 2018 [4, 49] opened the way for a direct measurement of the coupling between top quarks and the Higgs boson. A separate analysis searched for the production of a single top quark in association with a Higgs boson (tH) [5]. For the full dataset of $138\,\mathrm{fb}^{-1}$ collected at the CMS experiment in Run 2 between 2016 and 2018, the previously separate analyses are combined. The targeted Higgs boson decay channel is the decay into a bottom quark pair ($H \to b\bar{b}$). By bringing together both analyses, not only is overlap avoided, but also the best methods from both analyses can be utilized.

With the analysis as foundation, a number of measurements can be performed: In addition to the inclusive $t\bar{t}H$ cross section measurement, the $t\bar{t}H$ cross section is measured binned in simplified template cross sections (STXS). The search for tH production is complemented by a measurement of the coupling between the top quark and the Higgs boson, both for a SM coupling scenario and for a mixture with a BSM pseudoscalar Higgs boson.

In the end, the measurement in the $H \to b\bar{b}$ decay channel is combined with the measurement in the $H \to ZZ/\gamma\gamma$ [6] and Higgs boson to multilepton [7] decay channels.

3.1 Signal processes

The signal processes of the analysis are the associated production of a Higgs boson with a single top quark or a top quark pair. In $t\bar{t}H$ production, the Higgs boson couples to a top quark produced in a $t\bar{t}$ process, as shown in figure 3.1. For tH

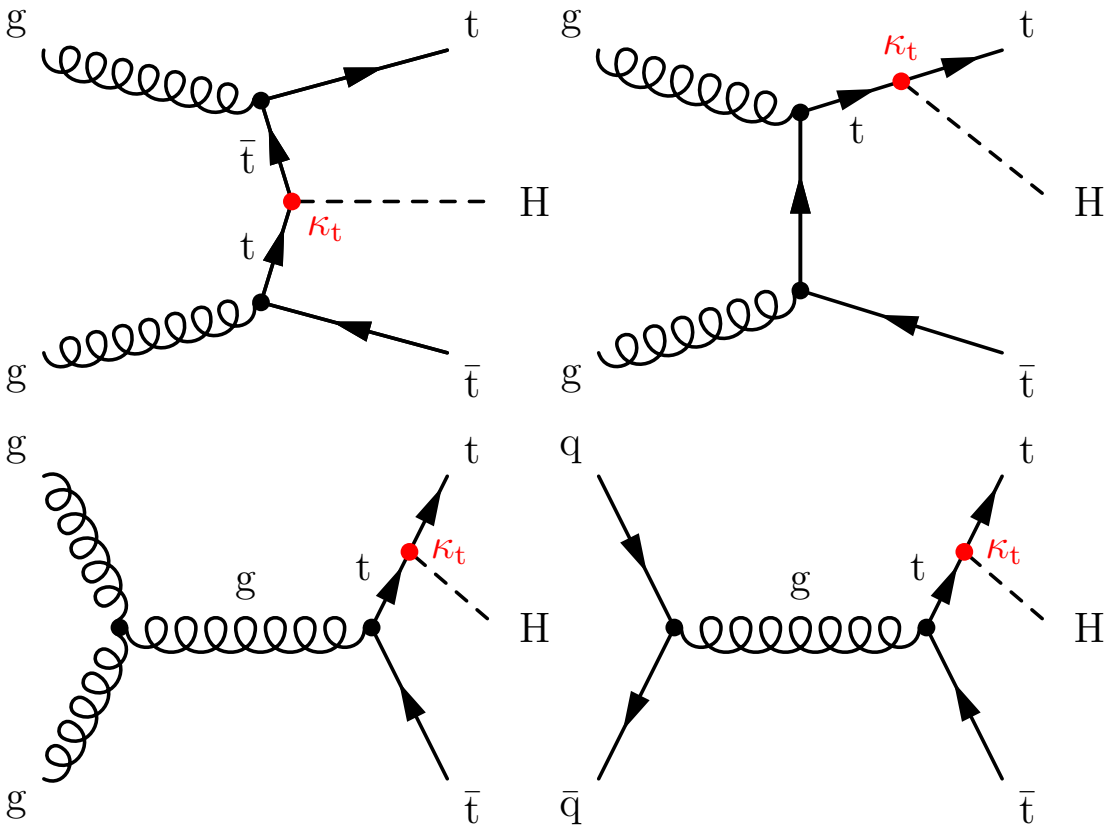
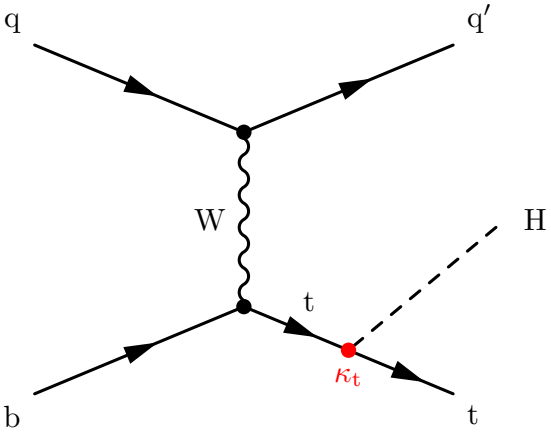


Figure 3.1: Feynman diagrams for the associated production of a Higgs boson with top quark pairs. The Higgs boson couples strongly to the top quark or antiquark, due to the large mass of the top quark. Each process contains the coupling κ_t exactly once.

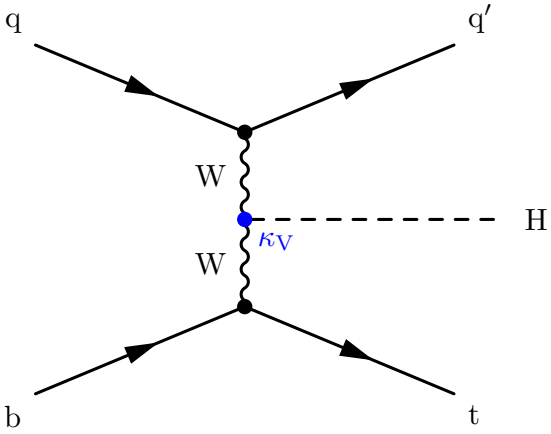
production, there exist two main production channels, shown in figure 3.2: Firstly tH production in association with a light quark (tHq), which is equivalent to single top t channel production with a Higgs boson radiated off either by the top quark or the W boson. Secondly tH production in association with a W boson (tHW) is equivalent to single top production in association with a W boson, where the Higgs boson is again radiated off by either the top quark or the W boson.

The dependency of the $t\bar{t}H$ production cross section on the top-Higgs coupling is quadratic while the dependence of the tH production cross section includes additional terms due to interference between the Feynman diagrams where the Higgs boson is radiated off by the top quark and the W boson.

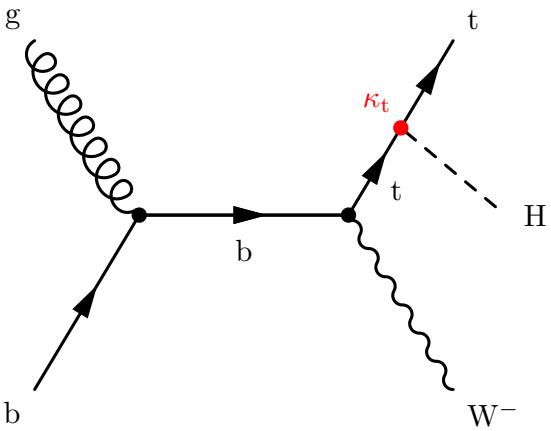
The dependencies of the production cross sections on the Higgs boson coupling can



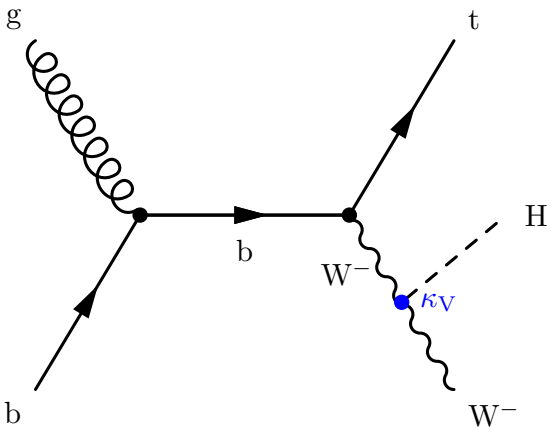
(a) tHq production with the Higgs boson coupling to the top quark



(b) tHq production with the Higgs boson coupling to the W boson



(c) tHW production with the Higgs boson coupling to the top quark



(d) tHW production with the Higgs boson coupling to the W boson

Figure 3.2: Feynman diagrams for the associated production of a Higgs boson with a single top quark. tHq production on the top (a-b) and tHW production at the bottom (c-d). The Higgs boson couples either to the top quark with κ_t (left) or the W boson with κ_V (right).

be expressed in the kappa-framework [50]:

$$\sigma_{\mathrm{t\bar{t}H}} = \kappa_{\mathrm{t}}{}^2 \cdot \sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}} \tag{3.1}$$

$$\sigma_{\mathrm{tHq}} = \left(2.63 \cdot \kappa_{\mathrm{t}}{}^2 + 3.58 \cdot \kappa_{\mathrm{V}}{}^2 - 5.21 \cdot \kappa_{\mathrm{t}}\kappa_{\mathrm{V}}\right) \sigma^{\mathrm{SM}}_{\mathrm{tHq}} \tag{3.2}$$

$$\sigma_{\mathrm{tHW}} = \left(2.91 \cdot \kappa_{\mathrm{t}}{}^2 + 2.40 \cdot \kappa_{\mathrm{V}}{}^2 - 4.22 \cdot \kappa_{\mathrm{t}}\kappa_{\mathrm{V}}\right) \sigma^{\mathrm{SM}}_{\mathrm{tHW}} \tag{3.3}$$

with the coupling modifiers κ_{t} and κ_{V} defined relative to the SM expectation:

$$\kappa_{\mathrm{V}} = \frac{g_{\mathrm{HVV}}}{g^{\mathrm{SM}}_{\mathrm{HVV}}} \tag{3.4}$$

$$\kappa_{\mathrm{t}} = \frac{y_{\mathrm{t}}}{y^{\mathrm{SM}}_{\mathrm{t}}} \tag{3.5}$$

using the Higgs boson coupling strength to gauge bosons g_{HVV} and the Yukawa coupling strength y_{t} introduced in section 2.1.1.

In case of the SM, both coupling modifiers are equal to one ($\kappa_{\mathrm{t}} = \kappa_{\mathrm{V}} = 1$). Due to the quadratic dependency of the $\mathrm{t\bar{t}H}$ production cross section on κ_{t}, there is no sensitivity for the sign of the top-Higgs coupling in $\mathrm{t\bar{t}H}$ production. tH production on the other hand is sensitive to the relative sign of κ_{t} to κ_{V} through the interference term, allowing to distinguish the SM from the inverse top coupling (ITC) ($\kappa_{\mathrm{t}} = -1, \kappa_{\mathrm{V}} = 1$) scenario.

The SM production cross sections at NLO QCD accuracy [51] are:

$$\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}} = (507.1\,^{+29.4}_{-46.7}\,(\mathrm{scale}) \pm 18.3\,(\mathrm{PDF}{+}\alpha_{\mathrm{s}}))\,\mathrm{fb} \tag{3.6}$$

$$\sigma^{\mathrm{SM}}_{\mathrm{tHq}} = (74.3\,^{+4.8}_{-11.1}\,(\mathrm{scale}) \pm 2.7\,(\mathrm{PDF}{+}\alpha_{\mathrm{s}}))\,\mathrm{fb} \tag{3.7}$$

$$\sigma^{\mathrm{SM}}_{\mathrm{tHW}} = (15.2\,^{+0.7}_{-1.0}\,(\mathrm{scale}) \pm 1.0\,(\mathrm{PDF}{+}\alpha_{\mathrm{s}}))\,\mathrm{fb} \tag{3.8}$$

for a top quark mass of $m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$, a Higgs boson mass of $m_{\mathrm{H}} = 125.0\,\mathrm{GeV}$ and a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$. In figure 3.3, the individual cross sections of tH and $\mathrm{t\bar{t}H}$ production are shown in dependence of the strength of the top-Higgs coupling κ_{t} with gauge-boson-Higgs-boson coupling fixed to the SM value of $\kappa_{\mathrm{V}} = 1$.

The branching ratio for a decay of a Higgs boson into a pair of bottom quarks [52], which is targeted by this analysis, is predicted as $\mathrm{BR}(\mathrm{H} \to \mathrm{b\bar{b}}) = 0.577 \pm 0.019$ for a Higgs boson mass of $m_{\mathrm{H}} = 125.0\,\mathrm{GeV}$.

For tH and $\mathrm{t\bar{t}H}$ production, the signal strengths μ_{tH} and $\mu_{\mathrm{t\bar{t}H}}$ are defined relative to the SM cross sections as:

$$\mu_{\mathrm{t\bar{t}H}} = \frac{\sigma_{\mathrm{t\bar{t}H}}}{\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}}} \tag{3.9}$$

$$\mu_{\mathrm{tH}} = \frac{\sigma_{\mathrm{tH}}}{\sigma^{\mathrm{SM}}_{\mathrm{tH}}} \tag{3.10}$$

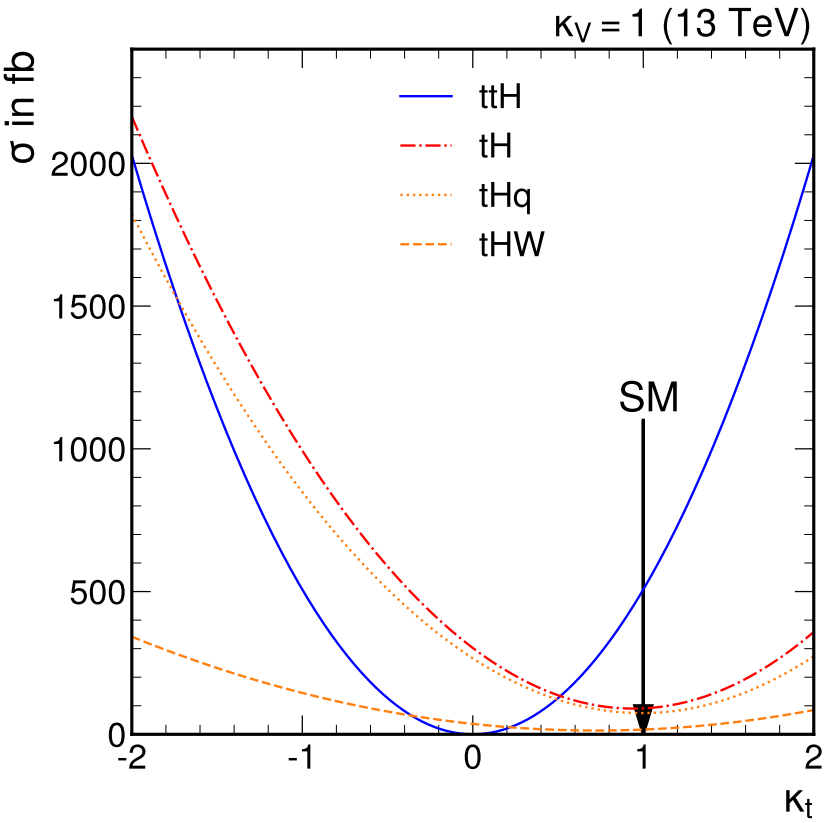


Figure 3.3: Dependence of the signal process cross sections on the top-Higgs coupling κ_t for a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with κ_V fixed to the SM value $\kappa_V = 1$. The tH production cross section is calculated as the sum of the tHq and tHW production cross sections. A vertical black vertical line marks the SM value of $\kappa_t = 1$.

Table 3.1: MC simulated datasets for the signal processes. datasets marked with * are only used to increase number of events in training and are not used in the final fit.

channel	dataset name
$t\bar{t}H$ (H $\rightarrow b\bar{b}$)	`ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8`
$t\bar{t}H$ (H $\rightarrow$non-$b\bar{b}$)	`ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8`
$t\bar{t}H$ (H $\rightarrow b\bar{b}$, SL)*	`ttHTobb_ttToSemiLep_M125_TuneCP5_13TeV-powheg-pythia8`
$t\bar{t}H$ (H $\rightarrow b\bar{b}$, DL)*	`ttHTobb_ttTo2L2Nu_M125_TuneCP5_13TeV-powheg-pythia8`
$t\bar{t}H$ (incluisive)	`ttH_4f_ctcvcp_TuneCP5_13TeV_madgraph_pythia8`
tHq (2016)	`THQ_ctcvcp_HIncl_M125_TuneCP5_13TeV-madgraph-pythia8`
tHq (2017)	`THQ_ctcvcp_4f_Hincl_13TeV_madgraph_pythia8`
tHW (2016)	`THW_ctcvcp_HIncl_M125_TuneCP5`
tHW (2017)	`THW_ctcvcp_5f_Hincl_13TeV_madgraph_pythia8`

Two different types of simulated MC datasets are used for the $t\bar{t}H$ process: NLO POWHEG datasets, split into Higgs boson decay into $b\bar{b}$ and others, and inclusive NLO MadGraph5 datasets. Two additional POWHEG$t\bar{t}H$ datasets with semi- or dileptonic $t\bar{t}$ decays are used to increase the size of the dataset in the training of the classification neural network, but are not used in the final fit. For tHq and tHW MadGraph5 datasets are used. The datasets are scaled to the NLO SM cross sections listed in equation (3.6). In table 3.1 the names of the datasets for each of the signal processes are listed. The MadGraph5 datasets contain special weights to reweight the cross sections and shapes into different coupling scenarios.

3.2 Background processes

The main background of the analysis is the production of a top quark pair with additional jets ($t\bar{t}$ + jets). The production diagrams are similar to those of the $t\bar{t}H$ signal process, shown in figure 3.1, but with the Higgs boson replaced by gluon radiation producing a quark pair. Feynman diagrams for the production of a top quark pair with an additional bottom quark pair ($t\bar{t} + b\bar{b}$) are shown in figure 3.4. In this case, the final-state particles are identical to those of the $t\bar{t}H$ signal process.

Depending on the flavor of the additional quark pair, the $t\bar{t}$ + jets background is split into different categories: If the additional jets are light flavored, the process is called $t\bar{t}$ + lf and if the additional jets are induced by a charm quark pair $c\bar{c}$, it is denoted as $t\bar{t} + c\bar{c}$. The production with bottom quark pairs is differentiated depending on how the bottom quarks are resolved in the detector: If they are both clustered into two separate jets it is denoted as $t\bar{t} + b\bar{b}$, but if one jet from a bottom quark is lost or if

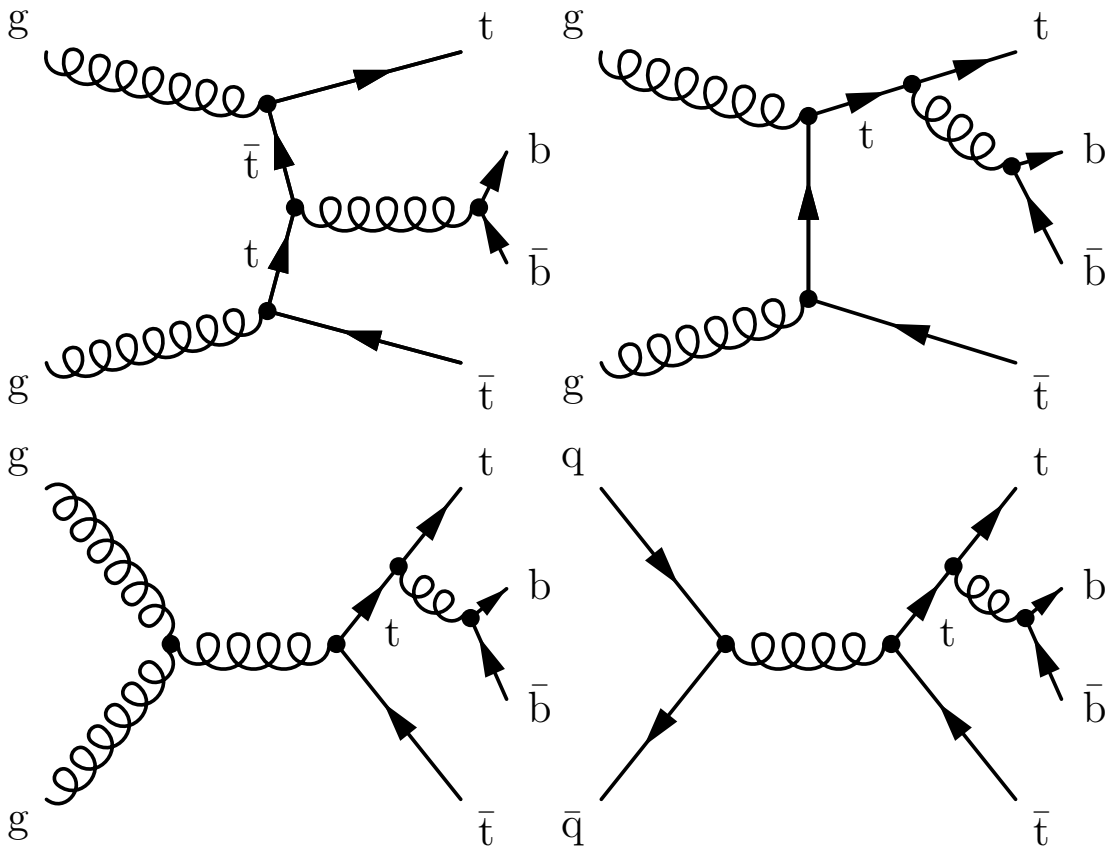


Figure 3.4: Feynman diagrams for the production of top quark pairs with an additional bottom quark pair. The additional bottom quark pair $b\bar{b}$ is produced by a gluon radiated off by a top quark.

the quark pair is clustered into one jet it is called $t\bar{t} + b$ or $t\bar{t} + 2b$ respectively.

Another important background process is the single top quark production. The production is similar to tH production without the Higgs boson. Again, gluon radiation can make it hard to distinguish it from the signal processes due to identical final-state particles.

Minor background contributions stem from processes containing gauge bosons, for example top quark pair production in association with a gauge boson $t\bar{t} + W$ and $t\bar{t} + Z$. The diboson production in the form of WW, WZ or ZZ can produce signatures similar to the signal due to the massive gauge bosons. The non-reducible QCD background also needs to be considered, particularly in the full hadronic channel, where it is determined with a data-driven method.

3.3 Coupling with pseudoscalar Higgs boson

Two-Higgs-doublet models predict the existence of additional Higgs boson including a pseudoscalar Higgs boson A. This additional pseudoscalar Higgs boson is introduced into the Lagrangian $\mathcal{L}$ describing the coupling with the top quark [53]:

$$\mathcal{L} = -\frac{y_t}{\sqrt{2}} \bar{t} \left(\kappa_t + i\gamma_5 \tilde{\kappa}_t\right) t X_0 \tag{3.11}$$

with the coupling modifier $\tilde{\kappa}_t$ for the coupling strength of the pseudoscalar Higgs boson and X_0 as mixture between the scalar SM Higgs boson H and the pseudoscalar Higgs boson A.

The coupling of the top quark with the two Higgs bosons can be described in polar coordinates with an absolute coupling strength κ'_t and a mixing angle α:

$$\kappa_t = \kappa'_t \cdot \cos\alpha \tag{3.12}$$
$$\tilde{\kappa}_t = \kappa'_t \cdot \sin\alpha \tag{3.13}$$

with $\kappa'_t = 1$ and $\alpha = 0$ in the SM case.

3.4 Analysis strategy

3.4.1 Event selection

Events are split into three channels by the number of leptons: the full hadronic (FH), single lepton (SL) and dilepton (DL) channel. The event selection applied in each channel is listed in table 3.2. The selection criteria are chosen to avoid overlap between the channels and with other $t\bar{t}H$ analyses, allowing for a combination of the results.

In each channel, triggers are chosen to select events as close to the selection as possible. For the FH channel this means that triggers selecting events with multiple jets are chosen. The triggers of the SL and DL channels rely on the presence of an electron or muon. Table 3.3 lists the triggers used in the SL channel of analysis. In table A.1 and table A.2, the triggers used in the FH and DL channel are shown. Trigger scale factors are used to correct for inefficiencies in the data acquisition caused by the triggers. In table 3.4, the used datasets collected by the CMS experiment during the three years of Run 2 are listed. To select events with good conditions, for example where the detector was fully functional, luminosity mask listed in table 3.5 are applied.

3.4.2 Event reconstruction

Different methods are used to interpret events and reconstruct final-state particles. The matrix element method [55], a likelihood ratio is derived from the matrix element to distinguish between the $t\bar{t}H$ signal process and the $t\bar{t} + b\bar{b}$ background, is used in all channels of the analysis.

Additionally in the SL channel, a BDTs-based jet assignment [56, 57] is used under the tHq, tHW, $t\bar{t}H$ and $t\bar{t}$ hypotheses to reconstruct the four-momenta of the final-state particles. The missing transverse momentum and the lepton are used to reconstruct a W boson. To get the missing momentum in beam direction p_z^{miss}, it is reconstructed by setting the reconstructed W boson mass to the literature value of $m_{\text{W}} = 80.4\,\text{GeV}$ [58]. In cases with two solutions from the quadratic equation, the smallest absolute value of p_z^{miss} is selected. The event is reconstructed with all possible jet assignments under a certain hypothesis and the best assignment is then determined by evaluating BDTs. These BDTs are trained on correct and wrong jet assignments to select the best assignment based on particle masses, kinematic quantities and flavor tagging values. This approach yields a better reconstruction efficiency than a simple χ^2 approach using the particle masses of reconstructed top

Table 3.2: Event selection [54] applied in the three analysis channels: full hadronic (FH), single lepton (SL) and dilepton (DL). The minimum p_T of the electron/muon in the SL channel is given for the individual years.

	FH channel	SL channel	DL channel
Number of leptons	0	1	2
Sign and flavor of leptons	—	$e^{\pm}$, $\mu^{\pm}$	e^+e^-, $\mu^{\pm}e^{\mp}$, $\mu^+\mu^-$
Min. p_T of p_T-leading electron (GeV)	—	29/30/30	25
Min. p_T of p_T-leading muon (GeV)	—	26/29/26	25
Min. p_T of additional leptons (GeV)	—	—	15
Max. p_T of additional leptons (GeV)	15	15	—
Max. $\|\eta\|$ of leptons	2.4	2.4	2.4
Min. number of jets	6	4	2
Min. p_T of jets (GeV)	40	30	30
Min. p_T of 6th hardest jet (GeV)	40	—	—
Max. $\|\eta\|$ of jets	2.4	2.4	2.4
Min. number of b-tagged jets	2	3	1
Min. p_T^{miss} (GeV)	—	20	40
Min. H_T (GeV)	500	—	—
Min. $m_{ee/\mu\mu}$ (GeV)	—	—	20
$m_{ee/\mu\mu}$ (GeV)	—	—	< 76, > 106
m_{qq} (GeV)	> 30, < 250	—	—

Table 3.3: Triggers used in the SL channel of the (t)tH analysis

HLT path	Run era
electron triggers	
`Ele27_WPTight_Gsf`	2016 B–H
`Ele28_eta2p1_WPTight_Gsf_HT150`	2017 B–F
`Ele32_WPTight_Gsf_L1DoubleEG` AND `ele32DoubleL1ToSingleL1Flag`	2017 B–F
`Ele28_eta2p1_WPTight_Gsf_HT150`	2018 A–D
`Ele32_WPTight_Gsf`	2018 A–D
muon triggers	
`IsoMu24`	2016 B–H
`IsoTkMu24`	2016 B–H
`IsoMu27`	2017 B–F
`IsoMu24`	2018 A–D

Table 3.4: Names of the datasets used by the $t\bar{t}H$ analysis, collected by the CMS experiment in the three years of Run 2. In 2018 the SingleElectron and DoubleEG datasets were merged into the EGamma dataset.

dataset	2016 ($35.9\,\mathrm{fb}^{-1}$)	2017 ($41.5\,\mathrm{fb}^{-1}$)	2018 ($59.7\,\mathrm{fb}^{-1}$)
JetHT	X	X	X
MET	X	X	X
BTagCSV		X	
SingleMuon	X	X	X
DoubleMuon	X	X	X
MuonEG	X	X	X
SingleElectron	X	X	
DoubleEG	X	X	
EGamma			X

Table 3.5: Luminosity mask applied to select events where the detector was fully functional. The listed luminosity is the remaining luminosity after the mask is applied.

year	luminosity	list of good lumi sections
2016	$35.9\,\mathrm{fb}^{-1}$	`Cert_271036-284044_13TeV_ReReco_07Aug2017_Collisions16_JSON.txt`
2017	$41.5\,\mathrm{fb}^{-1}$	`Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON.txt`
2018	$59.7\,\mathrm{fb}^{-1}$	`Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt`
total	$\mathbf{138\,fb^{-1}}$	

quarks, W and Higgs bosons.

3.4.3 Event classification

The event classification is performed with neural networks. Channels are divided into regions with additional cuts on the number of jets and number of b-tags. In each region a neural network is trained to classify events. In the FH channel, neural networks with a single output node are used to discriminate between signal and background. The networks are also applied in the control region (CR), which are used to determine the QCD background. In the SL and DL channels, neural networks with multiple output nodes for signal and background processes are used. A likelihood ratio is calculated from the output of the $t\bar{t}H$ category $O(t\bar{t}H)$ and the output from a background category $O(t\bar{t} + b(b)$ or $O(t\bar{t} + B)$:

$$\mathrm{LLR}_{\mathrm{SL}} = \frac{O(t\bar{t}H)}{O(t\bar{t}H) + O(t\bar{t} + b(b))} \tag{3.14}$$

$$\mathrm{LLR}_{\mathrm{DL}} = \frac{O(t\bar{t}H)}{O(t\bar{t}H) + O(t\bar{t} + B)} \tag{3.15}$$

The category with the largest output value is the category assigned to an event. Either the shape of the output or the yield is used as input to the fit.

In figure 3.5, an overview over the regions in each channel and the output categories used in the fit are given. Figure 3.6 shows the distributions of neural network output for all channels and regions in 2017. The signal is expected towards higher output values.

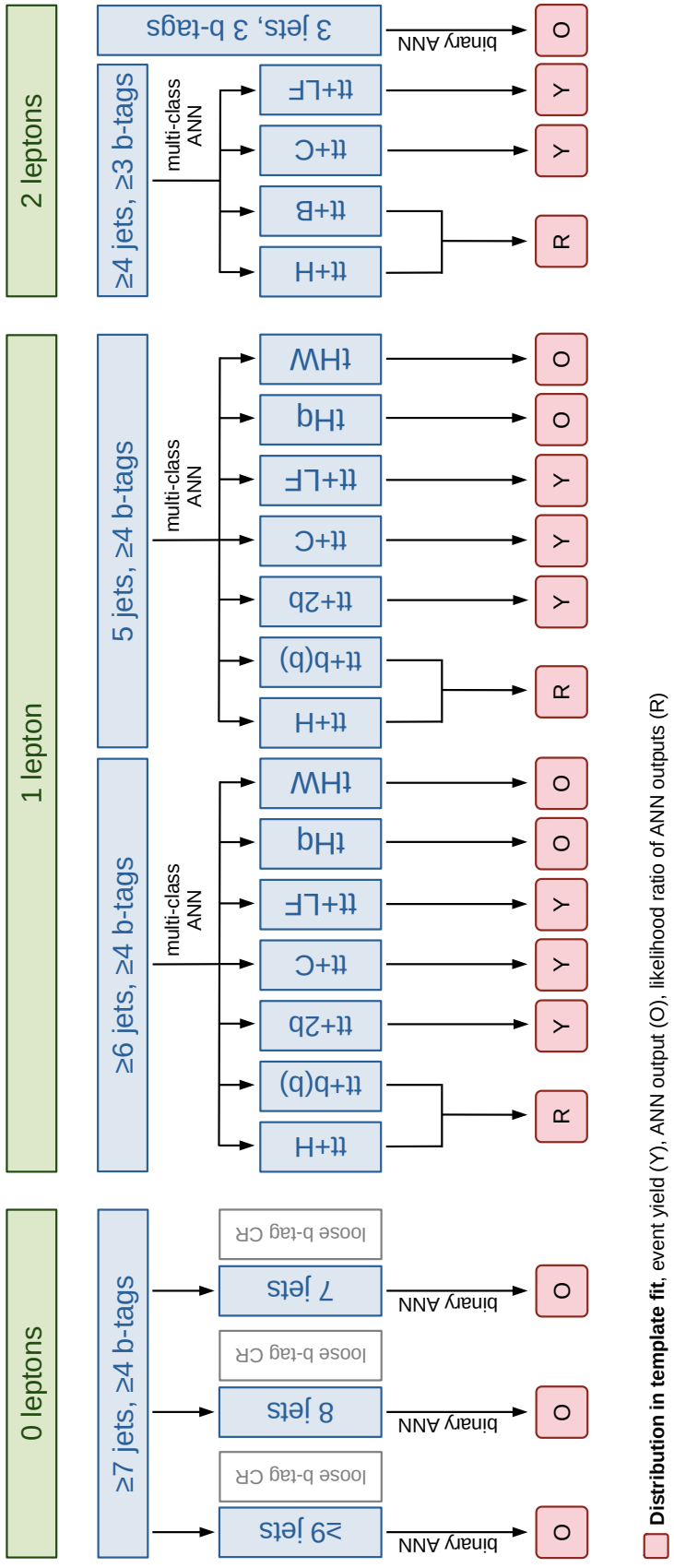


Figure 3.5: Categories of the $t\bar{t}H$ analysis [59] divided in the three channels (green) and regions (blue). Regions are defined by cuts on the number of jets and number of b-tags. Classification into signal and background is performed by neural networks. The input for the fit is either the yield of a node (Y), the shape (O) or a likelihood ratio of two nodes (R).

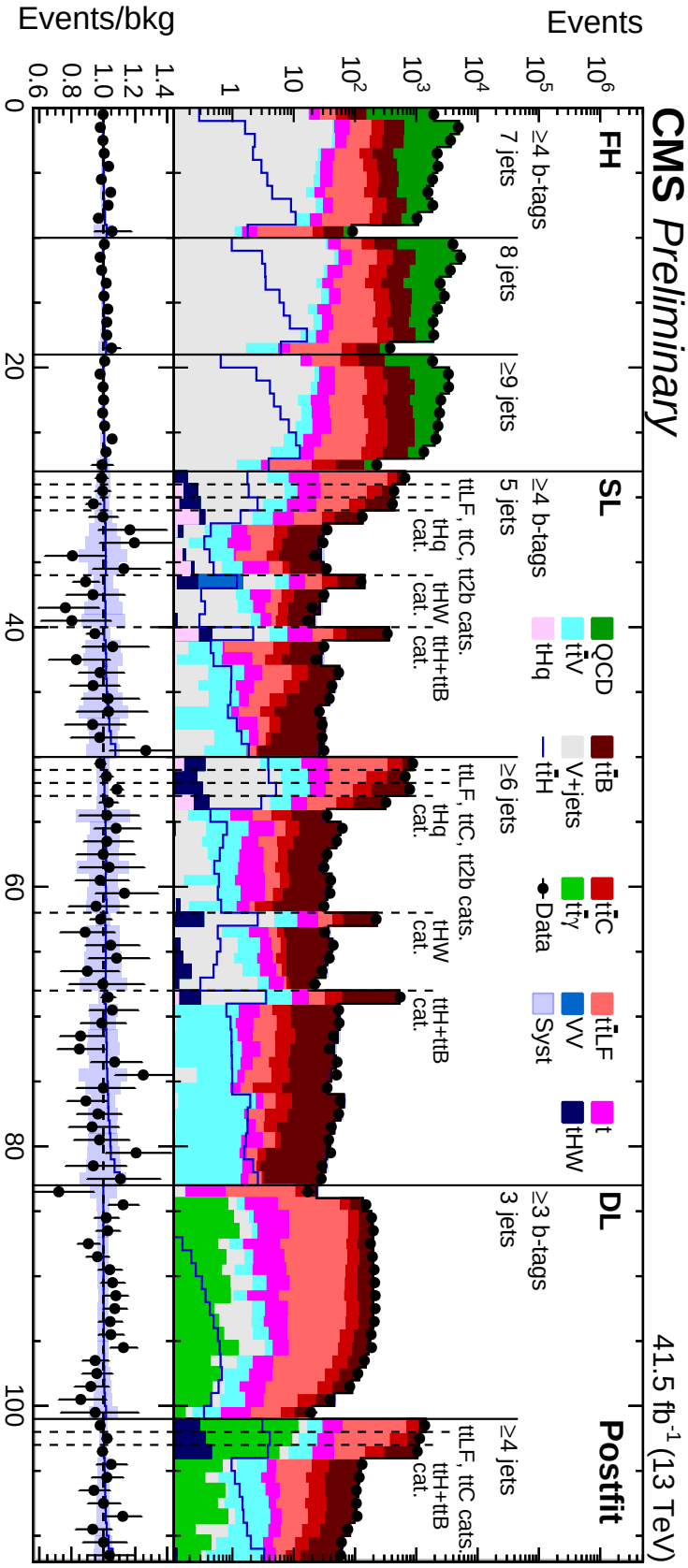


Figure 3.6: Postfit neural network output distributions with all channels and regions for 2017 [59].

3.4.4 Uncertainties

Experimental measurements are subject to statistical and systematic uncertainties. While the size of statistical uncertainties can be decreased by collecting more data, systematic uncertainties are an intrinsic property of a theory or experimental setup. The uncertainties can change the normalization or the shape of a distribution all together. These effects of shape uncertainties are estimated by up and down variation of an uncertainty while rate uncertainties are modeled by a log-normal distribution. The variations are determined by repeating the analysis either using special event weights or a special dataset. The uncertainties considered in the (t)tH ($H \to b\bar{b}$) analysis are listed below, grouped into theoretical and experimental uncertainties.

Theoretical uncertainties

PDF (shape+rate, rate) Shape uncertainties on the different production cross sections from uncertainties in the parton density functions (PDFs). Effects of the QCD scale α_s uncertainty in the calculation of PDFs are taken into account by an additional rate nuisance parameter for each of the tHq, tHW, $t\bar{t}H$, $t\bar{t}$ and single top quark production processes.

factorization and renormalization scale (shape+rate) Uncertainties in the selection of the factorization and renormalization scale, determined by varying the scale with factors of 0.5 and 2 while keeping the other fixed at the nominal scale.

ISR and FSR (shape+rate) Uncertainties of initial-state radiation (ISR) and final-state radiation (FSR) in $t\bar{t}$ production processes. For each of the $t\bar{t}+lf$, $t\bar{t}+c\bar{c}$, $t\bar{t}+b\bar{b}$ and $t\bar{t}H$ production processes an ISR and FSR uncertainty is introduced.

matrix element-parton shower matching (rate) Uncertainty from matching between matrix element generator and parton shower for the $t\bar{t}+lf$, $t\bar{t}+c\bar{c}$ and $t\bar{t}+b\bar{b}$ production processes.

underlying event (rate) Effect of underlying event on the cross section, fully correlated across all three years.

collinear gluon splitting (shape+rate) Uncertainty for the $t\bar{t}+2b$ process correlated over all three years.

background normalization (rate) The normalization of the MC simulated datasets for $t\bar{t}$ processes with additional bottom quarks $t\bar{t}+B$ and with additional charm quarks $t\bar{t}+C$ is scaled using two separate rate parameters, correlated over all

years and channels. The fitted values for these normalization parameters are given together with the best-fit values in section 3.5.3 and section 3.6.4.

Experimental uncertainties

luminosity (rate) For each year's luminosity measurement, an uncertainty is assigned and partial correlated between the years following the minimal correlations scheme [54]:

Year	uncertainty in %		
	2016	2017	2018
Uncorrelated 2016	1.0	—	—
Uncorrelated 2017	—	2.0	—
Uncorrelated 2018	—	—	1.5
Correlated 2017,2018	—	0.6	0.2
Correlated all years	0.6	0.9	2.0

pileup (shape+rate) Uncertainty in the number of pileup interactions, fully correlated across all three years.

L1 prefiring correction (shape+rate) Uncertainties on the L1 pre-firing corrections for 2016 and 2017. The L1 pre-firing happened due to a gradual timing shift in part of the ECAL. This caused the L1-trigger to associate the ECAL energy in that region to the previous bunch crossing and thus possibly vetoing the actual events, because triggering on two consecutive bunch crossings is forbidden.

trigger efficiency (shape+rate) Uncertainties on the trigger scale factors used to correct inefficiencies of the individual triggers, fully uncorrelated between the three years.

lepton efficiency (shape+rate) Uncertainties on the electron and muon reconstruction, identification and isolation scale factors, fully uncorrelated between the three years.

jet energy resolution (JER) (shape+rate) Uncertainty to account for differences in the shape of the jet energy between simulation and data, fully uncorrelated between the three years.

jet energy scale (JES) (shape+rate) Uncertainties introduced by the correction of the jet energy scale split with a set of eleven uncertainty sources in each year and over all three years.

b-tagging (shape+rate) Uncertainties from the b-tag scale factors split into sources for light, charm and heavy flavored jets and uncertainties from statistical limitations in the datasets used to obtain the scale factors. 8 different sources are considered for each year with partial correlation between 2017 and 2018.

QCD background estimation (shape+rate) Uncertainty from the loose b-tag requirement in the data-driven QCD background estimation, separate for each of the three years.

statistical MC & binning uncertainties (shape+rate) Statistical uncertainty from the limited statistics of the MC simulated dataset. The bin-by-bin level uncertainties are directly handled in COMBINE [60] with the "`* autoMCStats 10 0 1`" option [61], following the Barlow Beeston approach [62].

3.5 Search for single top quark production in association with a Higgs boson

3.5.1 Fit setup

The outputs of the classification for signal and background datasets are stored in COMBINE [60] datacards. These datacards contain the nominal and systematically varied distributions of the classification output with the information about the correlation of systematic uncertainties between different years and channels. In the calculation of upper limits, the production of tH is considered as signal while $t\bar{t}H$ production is considered as background and its cross section is fixed to the SM value. In a separate fit, the tH and $t\bar{t}H$ signal strengths are fitted simultaneously. As in the inclusive $t\bar{t}H$ measurement, the POWHEG MC dataset is used for the $t\bar{t}H$ process.

The background normalizations for the $t\bar{t}$ + B, $t\bar{t}$ + C and g $\rightarrow$ $b\bar{b}$ splitting backgrounds are determined in the fit by adding free floating scale factors to the corresponding datasets. Following the recommended procedure, the analysis strategy is performed blindly using Asimov toy data.

3.5.2 Goodness-of-fit tests and importance of nuisances

Different goodness-of-fit-tests are performed to check the compatibility of the fitted model with data. The statistical tests performed are the saturated goodness-of-fit

Table 3.6: The p-values from the different goodness-of-fit tests performed for the tH measurement. Tests are performed for the full analysis and parts of it.

	p-value		
	saturated	KS	AD
2016	0.91	0.56	0.78
2017	0.91	0.99	0.62
2018	0.74	0.22	0.18
DL	1.00	0.94	0.97
SL	0.95	0.80	0.71
Combined	0.90	0.85	0.69

test [63], which is an improved version of the χ^2 test, the Kolmogorov-Smirnov test [64] and the Anderson-Darling test [65]. The p-values of these tests for splits of years and channels as well as the full analysis are listed in table 3.6. The observed p-values show a good agreement of the fit result with data. In figure 3.7, the distributions of the saturated goodness-of-fit test statistic for 500 pseudo experiments are shown. The distributions of the goodness-of-fit tests for fits of parts of the analysis and other test methods are depicted in figure A.1, figure A.2 and figure A.3. As expected, the test statistics follow a χ^2 distribution and no problem with the fit can be observed.

The importance of nuisance parameters is determined by fitting with the up and down variation and evaluating the effect on the central fit value, in this case μ_{tH}. Constraints indicate how much a nuisance parameter is constrained by the fit compared to the expected pre-fit value. In figure 3.8, the constraints and importance of the most important nuisance parameters in the tH fit are shown. Figures for all nuisance parameters are given in appendix A.3. The nuisance parameters are not overly constrained and the impact of up and down variations on the fit values are not one-sided.

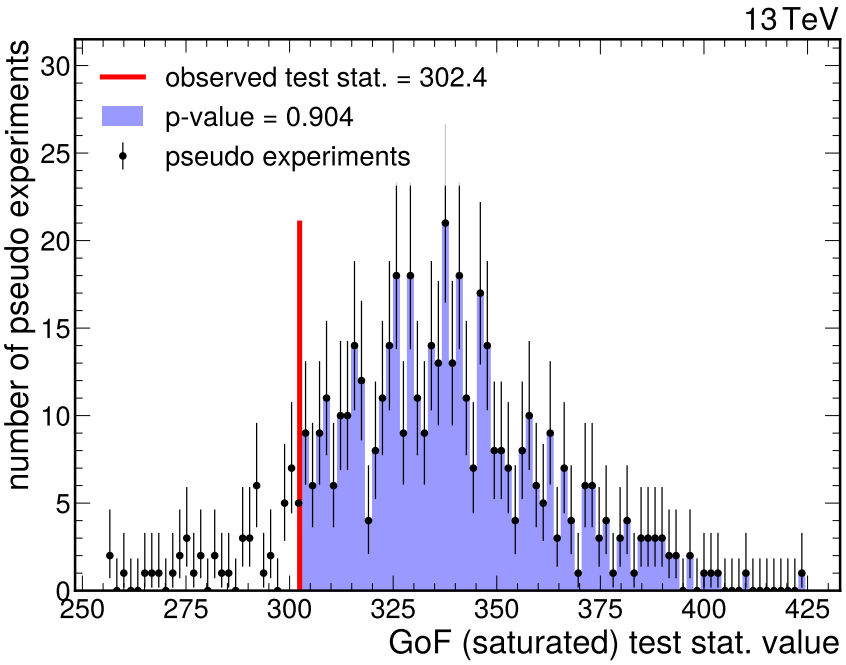


Figure 3.7: Distribution of the saturated goodness-of-fit test statistic for 500 pseudo experiments fitting tH signal for the full analysis.

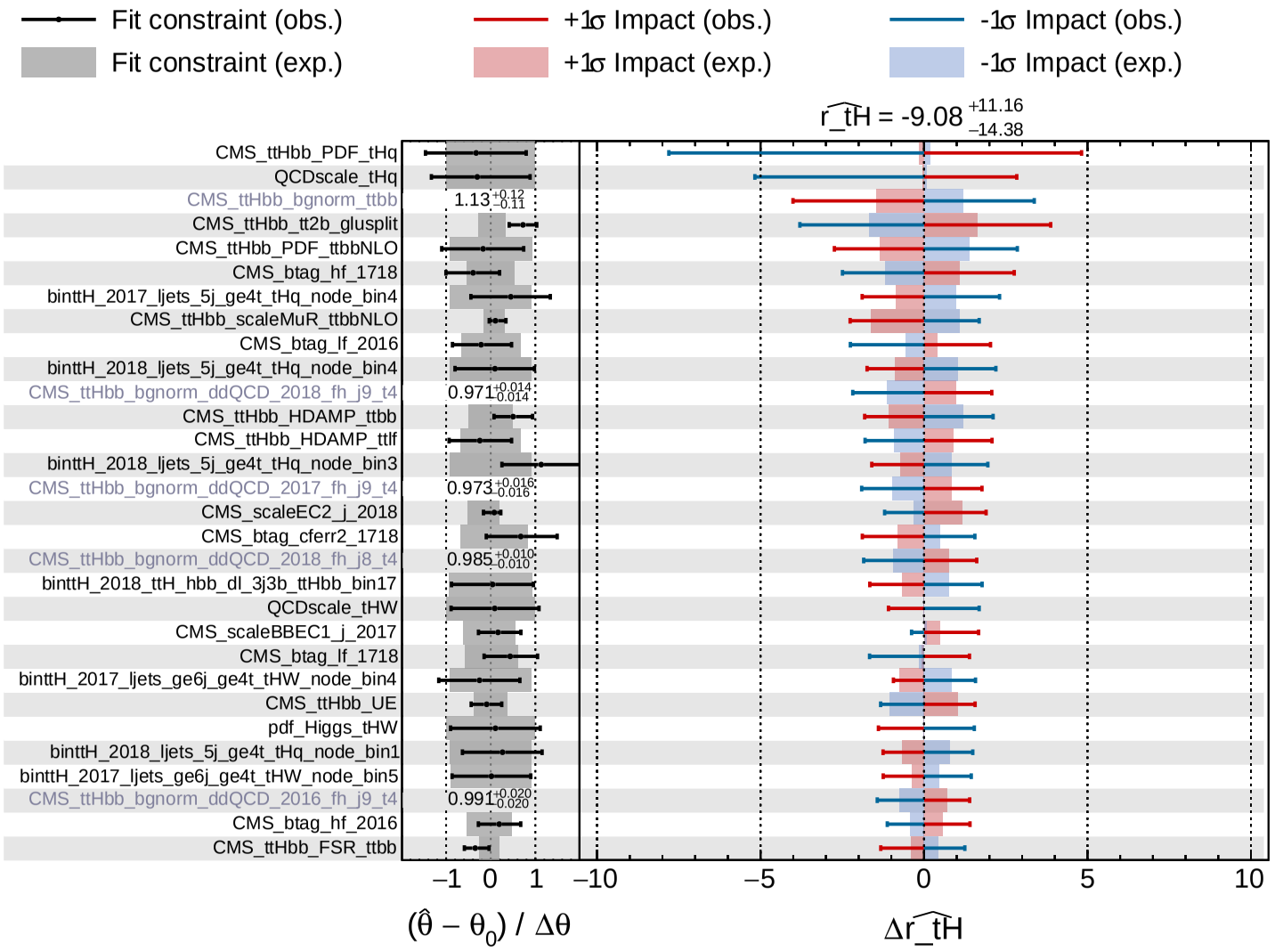


Figure 3.8: Expected and observed constraints and importance of the thirty highest ranking nuisance parameters in the fit for the tH measurement.

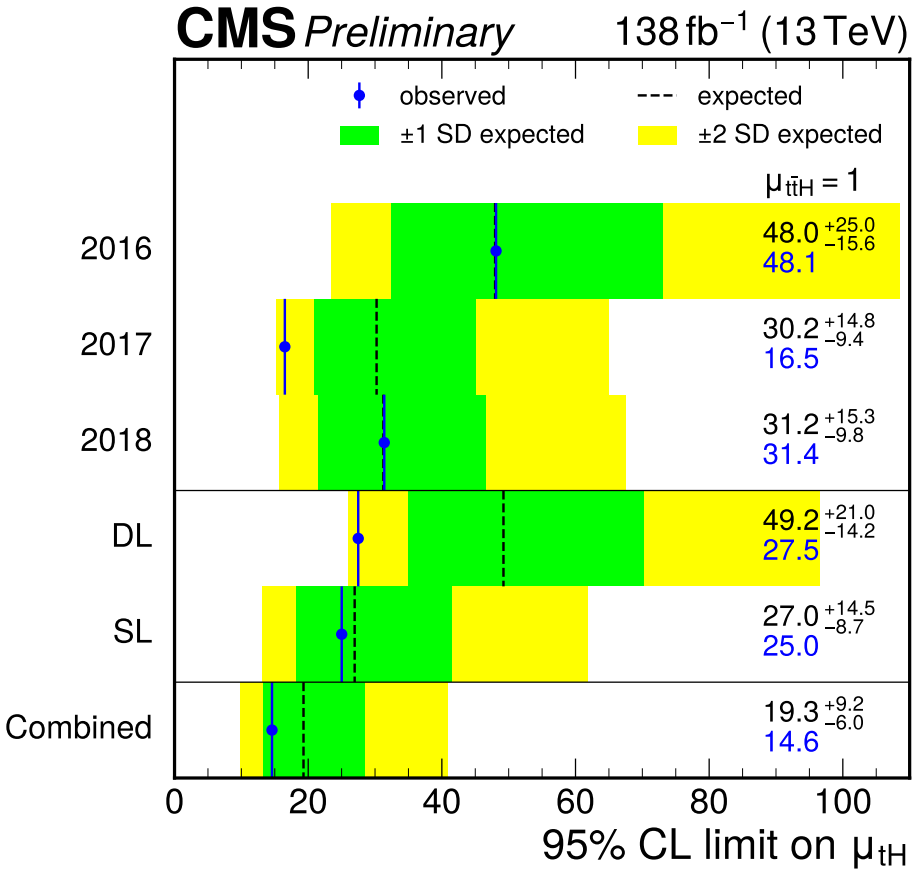


Figure 3.9: Expected and observed upper limits on the tH signal strength with 95 % CL. Limits are calculated for the full analysis and split into individual channels and years. The $t\bar{t}H$ signal strength is fixed to the SM expectation $\mu_{t\bar{t}H} = 1$. Published by the CMS collaboration in reference [59].

3.5.3 Results

In figure 3.9, the expected and observed upper limits on the tH signal strength μ_{tH} at 95 % CL are shown for the full analysis and split into individual channels and years. For the full analysis an upper limit of 14.6 ($19.3^{+9.2}_{-6.0}$) on the tH signal strength μ_{tH} is observed (expected) with 95 % CL.

In table 3.7, the best fit values for the tH signal strength and background normalization parameters are listed. The fitted values for the background normalization parameters are within their boundaries and compatible with the values fitted in the inclusive $t\bar{t}H$ measurement with fixed $\mu_{tH} = 1$.

Table 3.7: Best fit values of the tH signal strength μ_{tH} and the background normalization parameters for the full analysis and splits of years/channels.

	μ_{tH}		$\mathrm{t\bar{t}}+\mathrm{B}$		$\mathrm{t\bar{t}}+\mathrm{C}$		$g \to \mathrm{b\bar{b}}$ splitting	
2016	2	$^{+22}_{-23}$	1.00	$^{+0.17}_{-0.14}$	0.84	$^{+0.29}_{-0.27}$	1.16	$^{+0.54}_{-0.54}$
2017	-28	$^{+18}_{-19}$	1.17	$^{+0.19}_{-0.15}$	1.02	$^{+0.30}_{-0.34}$	0.84	$^{+0.53}_{-0.45}$
2018	2	$^{+15}_{-17}$	1.25	$^{+0.18}_{-0.15}$	1.41	$^{+0.28}_{-0.25}$	0.22	$^{+0.41}_{-0.37}$
DL	-45	$^{+24}_{-25}$	1.21	$^{+0.22}_{-0.18}$	1.49	$^{+0.43}_{-0.35}$	0.62	$^{+0.62}_{-0.62}$
SL	-2	$^{+13}_{-48}$	1.20	$^{+0.15}_{-0.14}$	1.36	$^{+0.54}_{-0.42}$	1.03	$^{+0.39}_{-0.39}$
Combined	-9	$^{+11}_{-14}$	1.13	$^{+0.12}_{-0.11}$	1.11	$^{+0.20}_{-0.19}$	0.72	$^{+0.31}_{-0.31}$

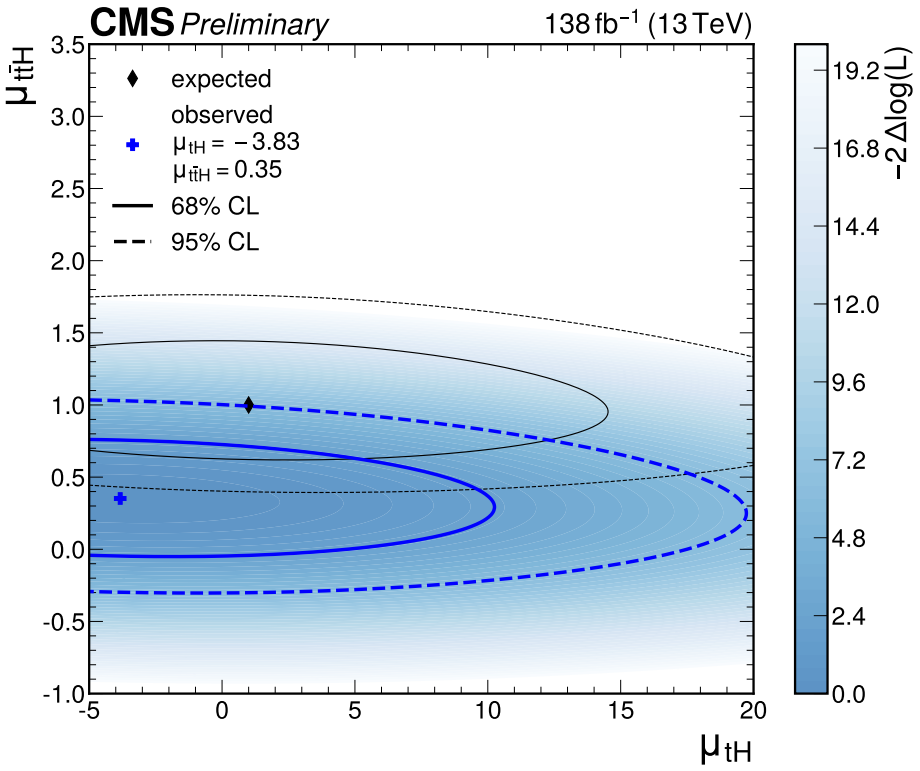


Figure 3.10: Asimov expected (top) and observed likelihood scan (bottom) for the simultaneous fit of μ_{tH} and $\mu_{\mathrm{t\bar{t}H}}$. Published by the CMS collaboration in reference [59].

3.6 Measurement of the Higgs boson coupling to top quarks

3.6.1 Fit model setup

To test different coupling hypotheses, signal templates are built from the distributions in different coupling scenarios. The $t\bar{t}H$ signal template is built from the SM template **A** (κ_t =1, $\tilde{\kappa}_t$ =0) and a template **B** with pure pseudoscalar Higgs boson coupling (κ_t =0, $\tilde{\kappa}_t$ =1):

$$c_{t\bar{t}H}(\kappa_t, \tilde{\kappa}_t) = \mathbf{A} \cdot {\kappa_t}^2 + \mathbf{B} \cdot {\tilde{\kappa}_t}^2 \tag{3.16}$$

Due to the more complex coupling dependencies of tHq and tHW from the interference, described in equation (3.1), more templates are required:

$$\begin{aligned} c_{tH}(\kappa_t, \kappa_V, \tilde{\kappa}_t) = \mathbf{A} \cdot \kappa_t(\kappa_t - \kappa_V) + \mathbf{B} \cdot {\tilde{\kappa}_t}^2 \\ + \mathbf{C} \cdot \kappa_V(\kappa_V - \kappa_t) + (\mathbf{D} + \mathbf{A} + \mathbf{C}) \cdot \kappa_t\kappa_V \end{aligned} \tag{3.17}$$

with the SM template $\mathbf{D} + \mathbf{A} + \mathbf{C}$ (κ_t =1, $\tilde{\kappa}_t$ =0, κ_V =1) and additional templates **A** (κ_t =1, $\tilde{\kappa}_t$ =0, κ_V =0), **B** (κ_t =0, $\tilde{\kappa}_t$ =1, κ_V =0) and **C** (κ_t =0, $\tilde{\kappa}_t$ =0, κ_V =1).

The templates are obtained by applying weights for a specific coupling scenario to events when filling the distribution. MadGraph5 [45] is used to generate the weights for the different coupling scenarios. For the tH datasets, the templates **A**, **B** and **C** are not included in the original dataset and have to be calculated separately. This is done by rerunning MadGraph5 with the generator information from the original dataset. The calculated weights are matched to the original events using unique ids.

The complete fit model [66] contains additional coupling modifiers for the Higgs coupling to gluons, photons, muons and charm quarks. In the fit, all other coupling modifiers are fixed to the SM value of one. Changes in the Higgs boson decay are modeled by scaling the branching ratios with the coupling modifiers and applying a normalization so that the Higgs boson decay width is fixed to the SM expectation.

The coupling between the top quark and the Higgs boson can also be expressed in polar coordinates with absolute coupling strength κ'_t and mixing angle $\cos\alpha$. By expressing the coupling modifiers κ_t and $\tilde{\kappa}_t$ in terms of κ'_t and $\cos\alpha$, the original coupling model can be used after the transformation:

$$\kappa_t = \kappa'_t \cdot \cos\alpha \tag{3.18}$$

$$\tilde{\kappa}_t = \kappa'_t \cdot \sin\alpha = \kappa'_t \cdot \sqrt{1 - (\cos\alpha)^2} \tag{3.19}$$

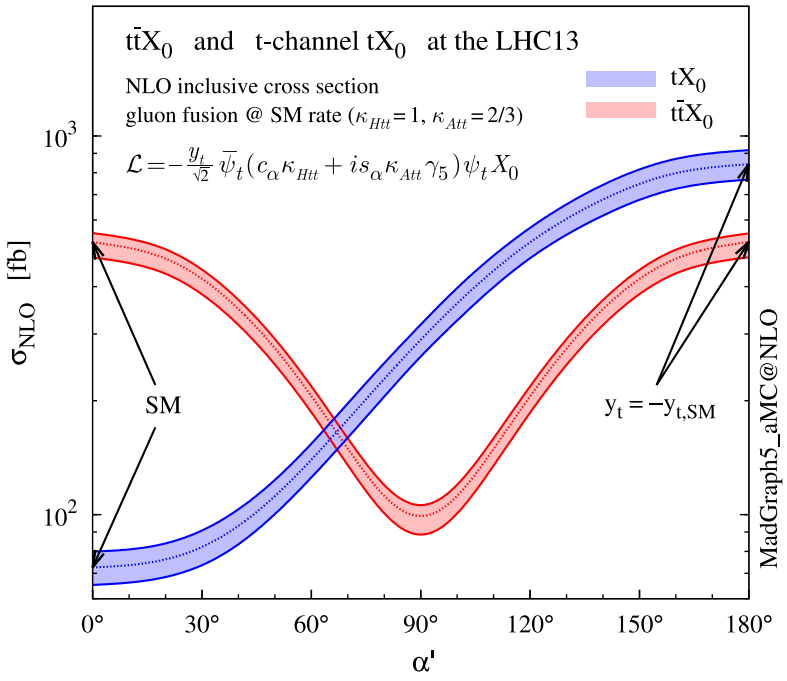


Figure 3.11: Cross sections at NLO [53] for tH (blue) and $t\bar{t}H$ (red) for different effective $\mathcal{CP}$ mixing angles $\cos\alpha'$. The $t\bar{t}H$ cross section is symmetrical to $\cos\alpha' = 90°$, thus only sensitive to absolute values of $\cos\alpha'$.

3.6.2 Fit model validation

The original model used to generate the non-SM weights in the MadGraph5 signal datasets is constrained by the Higgs boson production through gluon-gluon fusion at SM rate. As shown in figure 3.11, this results in an effective mixing angle α' defined by:

$$\kappa_t = 1 \cdot \cos\alpha' \tag{3.20}$$

$$\tilde{\kappa}_t = \frac{2}{3} \cdot \sin\alpha' \tag{3.21}$$

The weights included in the MadGraph5 datasets are produced for 21 different values of $\cos\alpha'$ between -1 and 1 in increments of 0.1.

To check the agreement between this model and the results obtained with the more general and less constrained coupling model, described in section 3.6.1, several tests were performed. First the required weights need to be produced and matched to the events. Then as a cross check, weights for already present coupling scenarios are produced and compared with the weights included in the datasets. Only tiny numerical differences between old and new weights are observed.

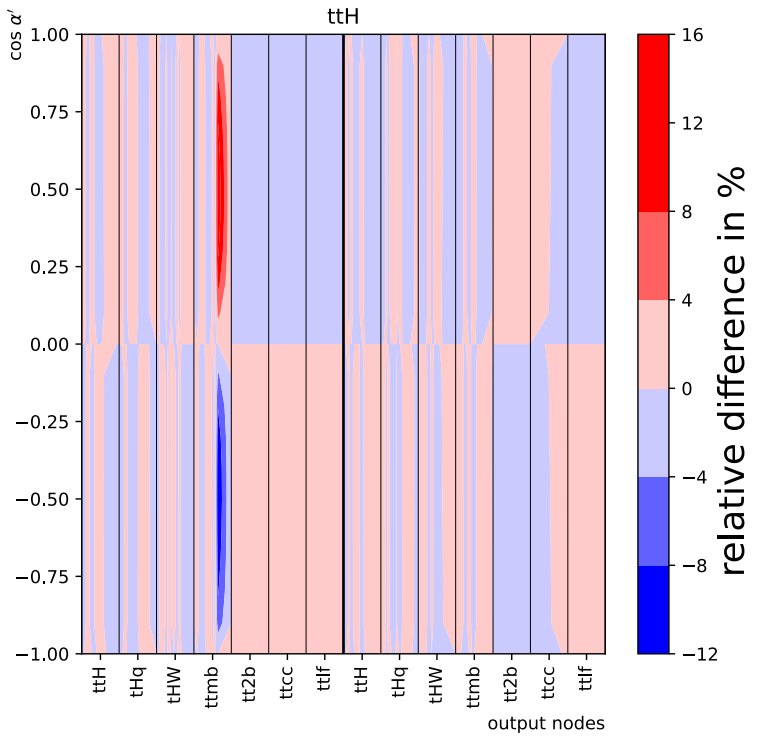


Figure 3.12: Difference in the distributions obtained with the template scaling relative to the distributions with the original weights for the $t\bar{t}H$ process. The classification neural network output is taken from two regions of the SL channel.

Lastly the definition of the templates is verified by producing the required templates and scale them to the old coupling scenarios of $\cos\alpha'$ between -1 and 1, as defined in equation (3.16) and equation (3.17). In figure 3.12 and figure 3.13 the classification output distributions are compared between old and new model.

For the $t\bar{t}H$ signal process, shown in figure 3.12, the largest deviation is observed in the shape of one of the background categories, because this category is dominated by background, the effect can be neglected. The shapes in the other distributions diverge less than 4 %.

For the tHq and tHW signal processes, shown in figure 3.13, there are also larger deviations in the background bins observed. Again, this effect can be neglected because the category is dominated by background. In addition to that, the tHq process shows deviations in the shape of the tHW category with less than 7.5 % and the tHW process deviates in the shape of the tHq category with less than 12 %. Since this difference only affects the category of the other tH signal process and tH contribution is expected to be much smaller than $t\bar{t}H$, this is considered acceptable.

Lastly, distributions and expected fit results are produced with the fit model for

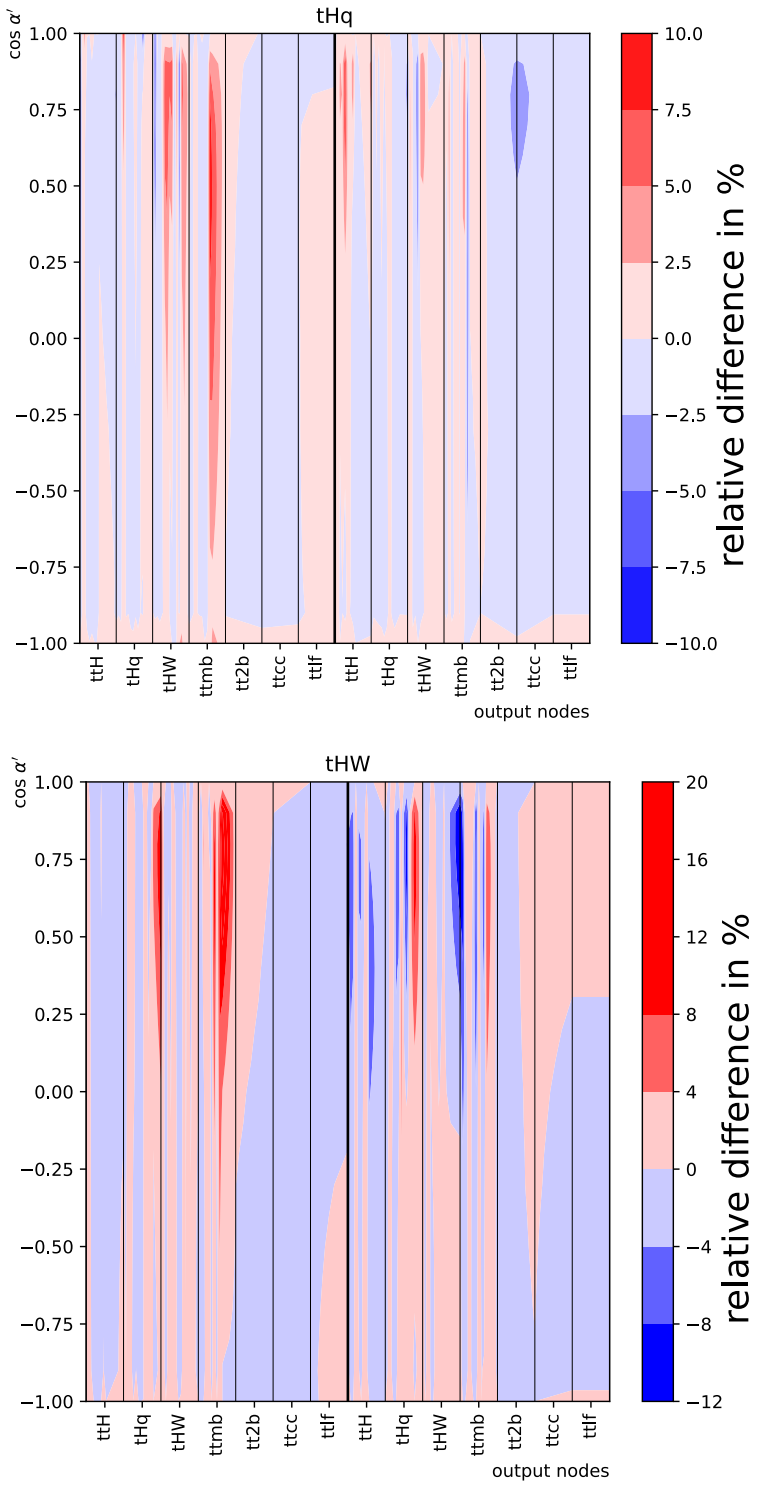


Figure 3.13: Difference in the distributions obtained with the template scaling relative to the distributions with the original weights for the tHq (top) and tHW (bottom) processes. The classification neural network output is taken from two regions of the SL channel.

Table 3.8: The p-values from the different goodness-of-fit tests performed in the SM coupling scenario. Tests are performed for the full analysis and parts of it.

	p-value		
	saturated	KS	AD
2016	0.95	0.53	0.72
2017	0.90	0.98	0.67
2018	0.82	0.13	0.16
DL	0.98	0.96	0.98
SL	0.98	0.79	0.68
Combined	0.79	0.89	0.64

$\cos\alpha'$ between -1 and 1 to compare with the original weights. In figure 3.14, a comparison of output distributions and expected fit results between the original and the new fit model is shown. The shapes of the distributions for $\kappa_t = 1$ and $\tilde{\kappa}_t = \frac{2}{3}$ agree within statistical fluctuations. The 16.5 % increase in normalization is caused by a scale applied to the $H \to b\bar{b}$ branching ratio in the fit model and not present in the final version of the model. This also effects the expected fit result, where a higher sensitivity is expected for $\cos\alpha' < 0$ due to the increase of signal.

Fits are performed in two coupling scenarios: In the SM coupling scenario the parameter of interest are κ_t and κ_V with fixed $\tilde{\kappa}_t = 0$. In addition to a two-dimensional likelihood scan in κ_t and κ_V, a likelihood scan in κ_t with fixed $\kappa_V = 1$ is performed. With the $\mathcal{CP}$ coupling scenario, the κ_t and $\tilde{\kappa}_t$ parameters are fitted with fixed $\kappa_V = 1$. In addition to the two-dimensional likelihood scan, a fit in polar coordinates over $\cos\alpha$ with κ'_t profiled is performed.

3.6.3 Goodness-of-fit tests

The results of the goodness-of-fit tests for the fit in the SM and $\mathcal{CP}$ coupling scenario are listed in table 3.8 and table 3.9 respectively. In figure 3.15, the distributions of the test statistic from the saturated goodness-of-fit tests are shown. A good agreement of the fit with data is seen in both cases.

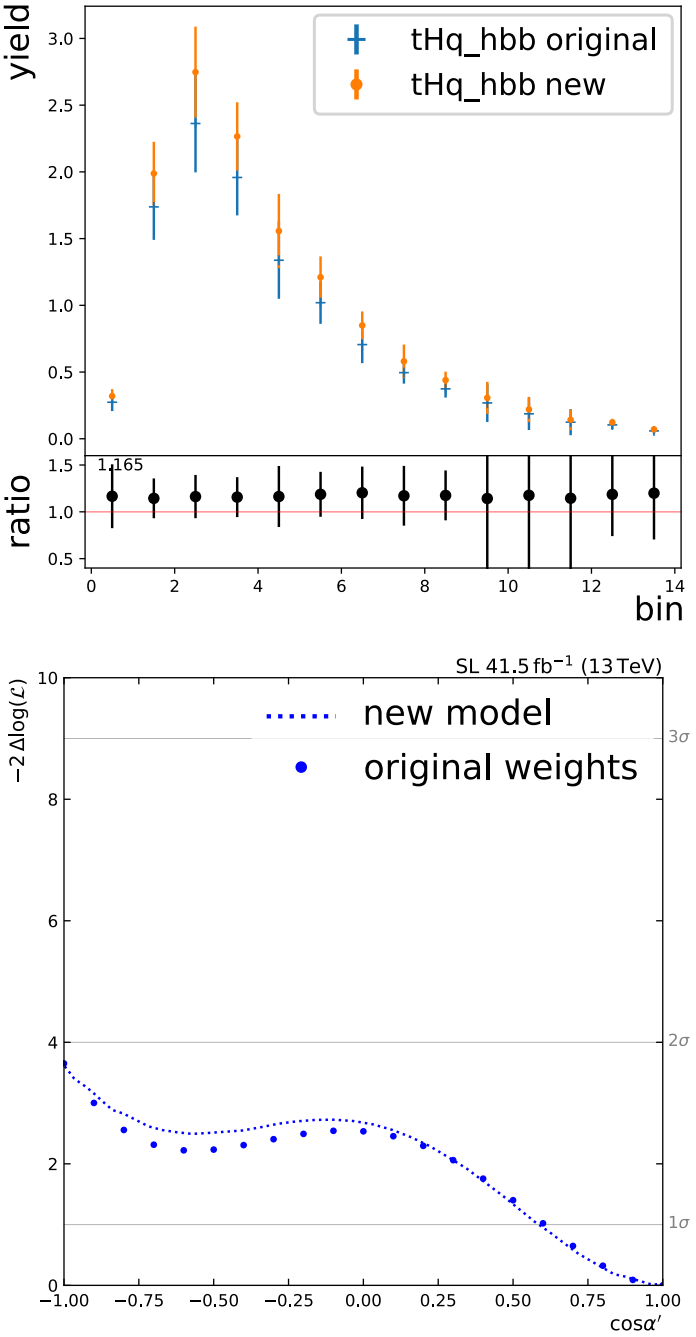
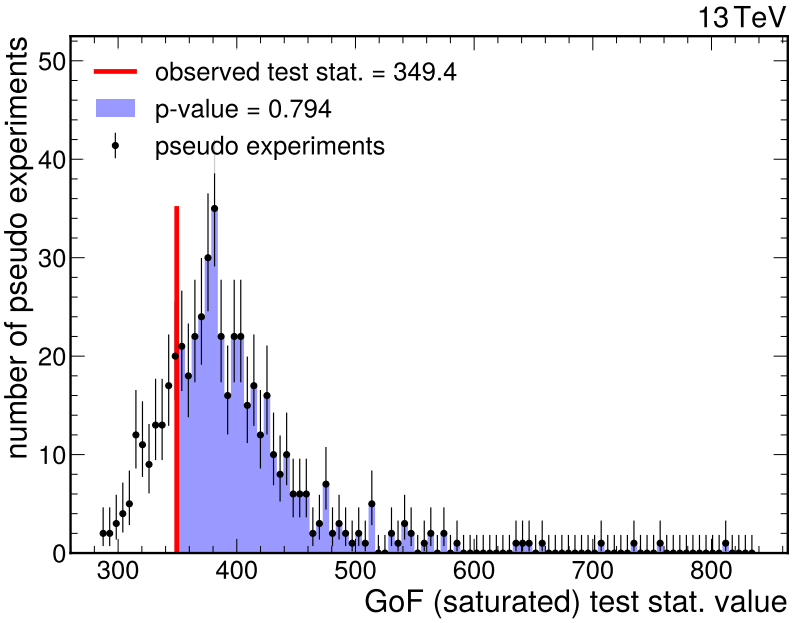
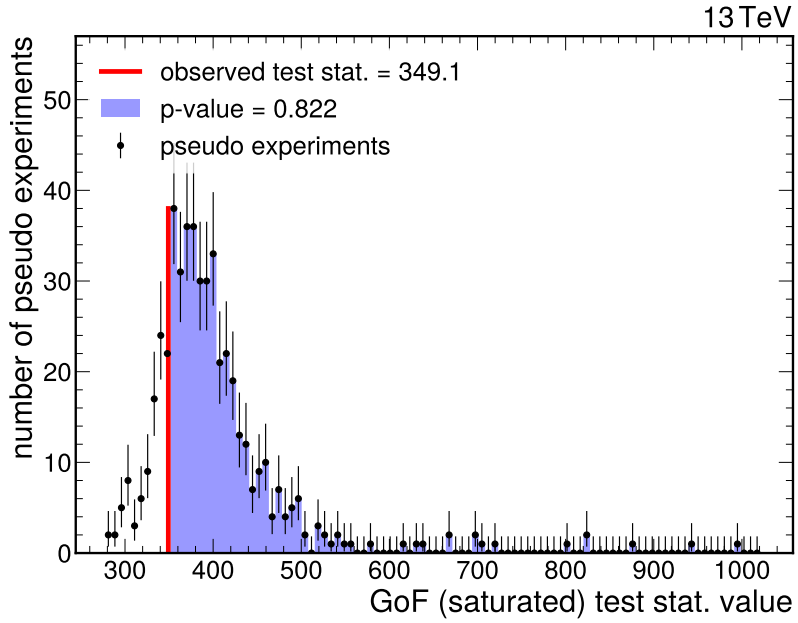


Figure 3.14: Comparison of the tHq distribution in a SL signal node with the original weights and the new model with $\kappa_t = 1$ and $\tilde{\kappa}_t = \frac{2}{3}$ (top) and comparison of the expected fit results with the original weights and the new fit model for different effective mixing angles $\cos\alpha'$ (bottom). The distribution with the new fit model has a 16.5 % higher normalization.



(a) SM coupling scenario



(b) $\mathcal{CP}$ coupling scenario

Figure 3.15: Distribution of the saturated goodness-of-fit test statistic for 500 pseudo experiments fitting in the SM (a) and $\mathcal{CP}$ coupling scenario (b).

Table 3.9: The p-values from the different goodness-of-fit tests performed in the $\mathcal{CP}$ coupling scenario. Tests are performed for the full analysis and parts of it.

	p-value		
	saturated	KS	AD
2016	0.96	0.64	0.74
2017	0.87	0.98	0.52
2018	0.84	0.11	0.21
DL	0.98	0.97	0.98
SL	0.98	0.82	0.70
Combined	0.82	0.86	0.66

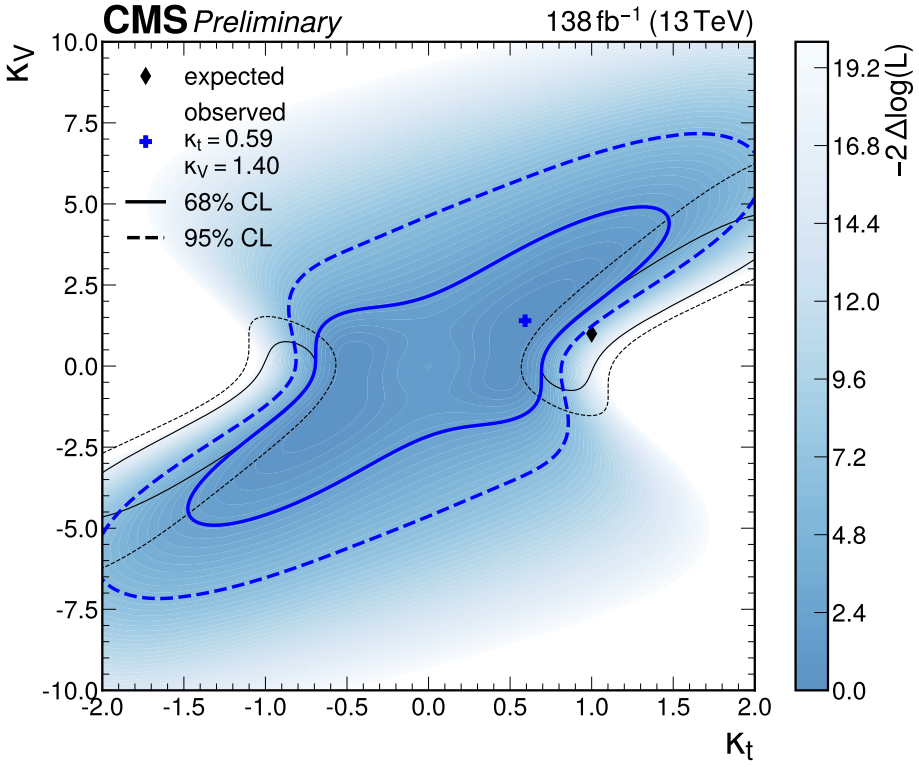


Figure 3.16: Asimov expected and observed 2D likelihood scan in the SM coupling scenario. Published by the CMS collaboration in reference [59].

3.6.4 Results

In figure 3.16 the expected and observed 2D likelihood in κ_t and κ_V with fixed $\tilde{\kappa}_t = 0$ are shown. The expected and observed 1D likelihood for κ_t with profiled κ_V is depicted in figure 3.17. The observed coupling strength is lower than the SM value and compatible with the lower $t\bar{t}H$ signal strength observed in the inclusive measurement. The best fit results for the coupling modifiers κ_t/κ_V and background normalizations are given in table 3.10.

In figure 3.18, the expected and observed likelihood scan in κ_t and $\tilde{\kappa}_t$ with fixed $\kappa_V = 1$ are shown. The observed sensitivity to $\mathcal{CP}$ mixing is lower than expected. This is simulated in figure 3.19, where the coupling strength for the expectation is reduced to $\kappa_t = 0.5$. The resulting shape is similar to the observation and the degradation of the contour to a circle explains the reduced sensitivity to $\mathcal{CP}$ mixing states. The best fit results for the coupling modifiers $\kappa_t/\tilde{\kappa}_t$ and background normalizations are given in table 3.11. The fit results for the $\mathcal{CP}$ mixing angle $\cos\alpha$ with κ'_t profiled are given in figure 3.20 and table 3.12.

Figure 3.21 contains the likelihood observed by the ATLAS experiment in the $H \to b\bar{b}$

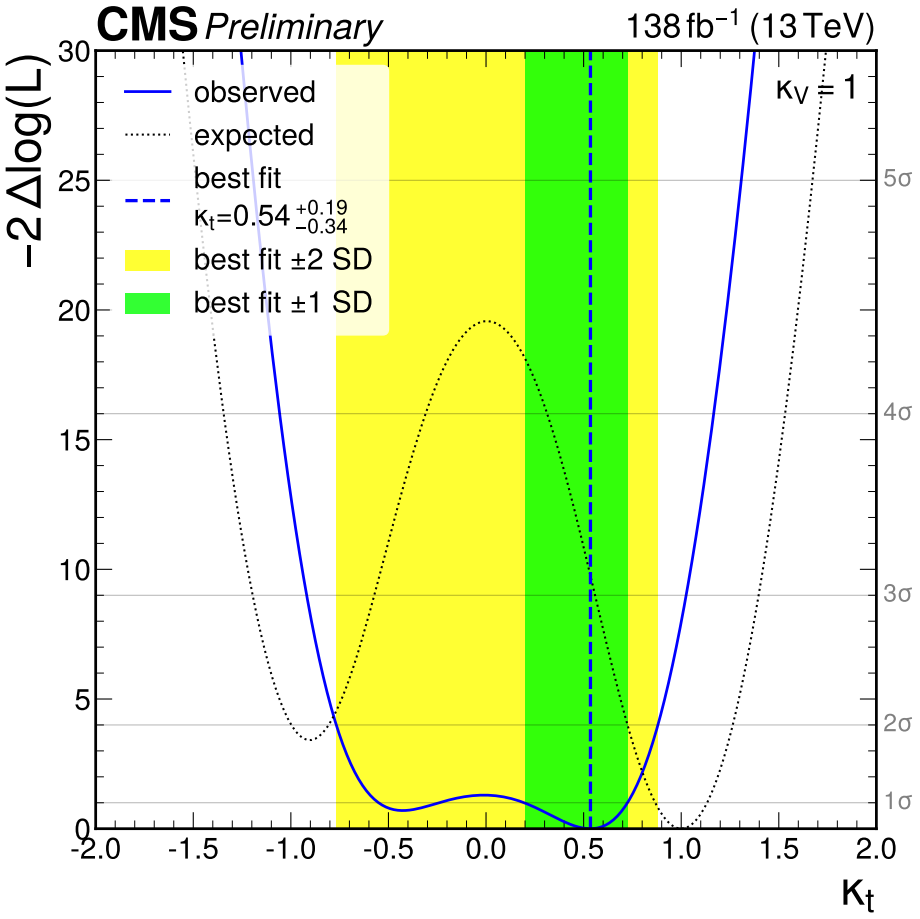


Figure 3.17: Expected and observed likelihood scan depending on κ_t with κ_V fixed to the SM. Published by the CMS collaboration in reference [59].

Table 3.10: Best fit values of κ_t, κ_V and the background normalization parameters for the full analysis and splits of years/channels.

	κ_t		κ_V		$t\bar{t}+B$		$t\bar{t}+C$		$g \to b\bar{b}$ splitting	
2016	1.73	$^{+0.27}_{-1.30}$	5.92	$^{+4.08}_{-8.63}$	1.03	$^{+0.18}_{-0.15}$	0.73	$^{+0.34}_{-0.36}$	1.01	$^{+0.53}_{-0.53}$
2017	0.53	$^{+0.48}_{-1.53}$	0.51	$^{+1.73}_{-2.75}$	1.17	$^{+0.17}_{-0.15}$	0.96	$^{+0.30}_{-0.27}$	0.83	$^{+0.51}_{-0.46}$
2018	0.17	$^{+0.54}_{-0.85}$	−2.38	$^{+1.62}_{-2.56}$	1.33	$^{+0.18}_{-0.16}$	1.33	$^{+0.27}_{-0.24}$	−0.05	$^{+0.39}_{-0.31}$
DL	−0.00	$^{+0.48}_{-0.48}$	−0.00	$^{+2.58}_{-2.58}$	1.30	$^{+0.20}_{-0.18}$	1.41	$^{+0.42}_{-0.33}$	0.12	$^{+0.63}_{-0.48}$
SL	0.99	$^{+0.94}_{-0.53}$	2.02	$^{+3.05}_{-3.46}$	1.21	$^{+0.15}_{-0.14}$	1.38	$^{+0.54}_{-0.42}$	0.92	$^{+0.40}_{-0.40}$
Combined	0.59	$^{+0.54}_{-0.39}$	1.40	$^{+2.17}_{-2.53}$	1.19	$^{+0.13}_{-0.11}$	1.07	$^{+0.20}_{-0.19}$	0.51	$^{+0.33}_{-0.33}$

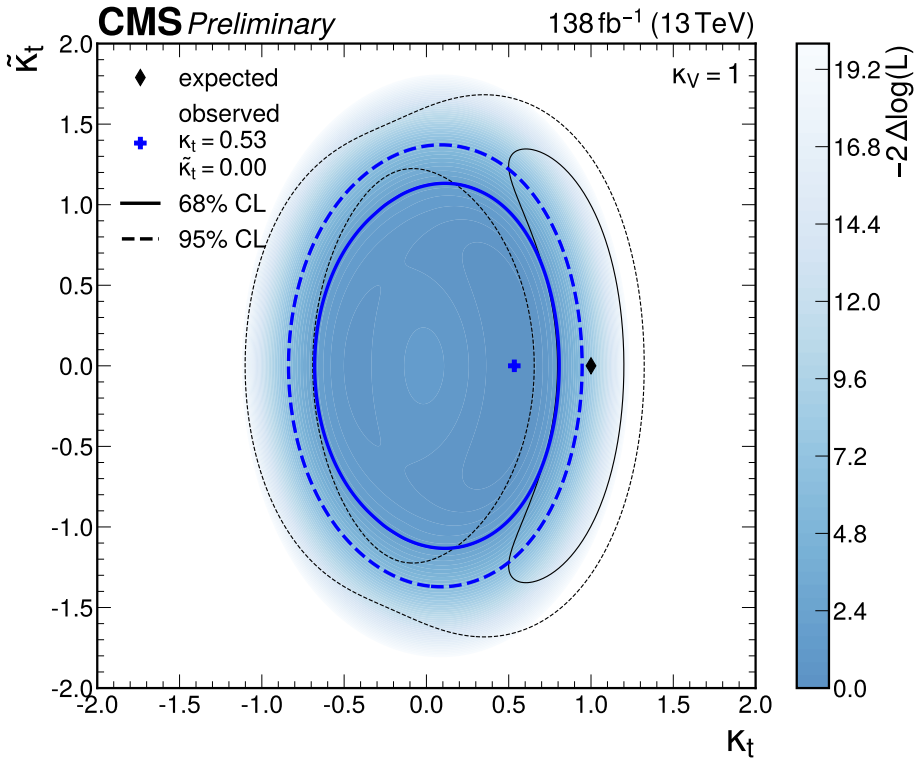


Figure 3.18: Asimov expected and observed 2D likelihood scan in the $\mathcal{CP}$ coupling scenario. Published by the CMS collaboration in reference [59].

decay channel [67] for comparison. Both prefer a SM-like scalar Higgs boson without excluding $\mathcal{CP}$ mixing. Other Higgs boson decay channels, like $H \to ZZ/\gamma\gamma$ [6] and Higgs to multilepton [7], already exclude the pure pseudoscalar case ($\kappa_t = 0, \tilde{\kappa}_t = 1$) with more than 95 %commands CL.

Table 3.11: Best fit values of κ_t, $\tilde{\kappa}_t$ and the background normalization parameters for the full analysis and splits of years/channels.

	κ_t		$\tilde{\kappa}_t$		$t\bar{t}+B$		$t\bar{t}+C$		$g \to b\bar{b}$ splitting	
2016	0.58	$^{+0.40}_{-1.52}$	−0.80	$^{+2.33}_{-0.73}$	1.03	$^{+0.18}_{-0.15}$	0.83	$^{+0.30}_{-0.27}$	1.02	$^{+0.56}_{-0.56}$
2017	0.62	$^{+0.24}_{-0.39}$	0.00	$^{+0.92}_{-0.92}$	1.16	$^{+0.17}_{-0.15}$	0.96	$^{+0.31}_{-0.27}$	0.86	$^{+0.51}_{-0.45}$
2018	−0.32	$^{+0.74}_{-0.35}$	−0.00	$^{+0.84}_{-0.84}$	1.33	$^{+0.19}_{-0.17}$	1.39	$^{+0.27}_{-0.24}$	−0.03	$^{+0.41}_{-0.33}$
DL	0.07	$^{+0.45}_{-0.46}$	0.00	$^{+0.67}_{-0.67}$	1.29	$^{+0.20}_{-0.19}$	1.41	$^{+0.42}_{-0.33}$	0.14	$^{+0.64}_{-0.49}$
SL	0.77	$^{+0.19}_{-0.33}$	0.00	$^{+0.91}_{-0.92}$	1.21	$^{+0.15}_{-0.14}$	1.38	$^{+0.54}_{-0.42}$	0.92	$^{+0.40}_{-0.40}$
Combined	0.53	$^{+0.19}_{-1.08}$	−0.00	$^{+0.95}_{-0.95}$	1.20	$^{+0.13}_{-0.12}$	1.07	$^{+0.20}_{-0.19}$	0.49	$^{+0.32}_{-0.32}$

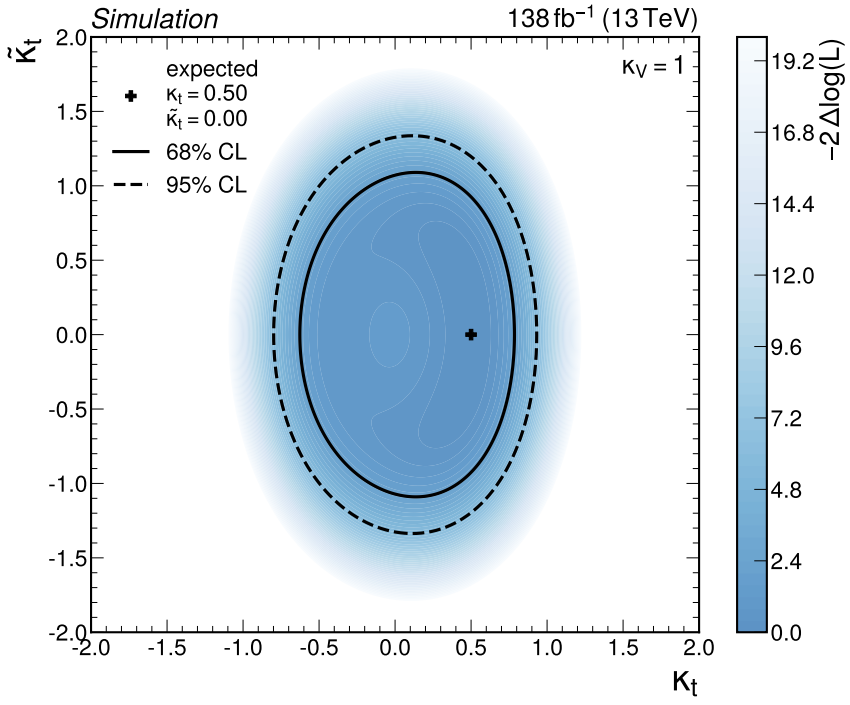


Figure 3.19: Asimov expected 2D likelihood scan in the $\mathcal{CP}$ coupling scenario with reduced coupling $\kappa_t = 0.5$ and $\tilde{\kappa}_t = 0$.

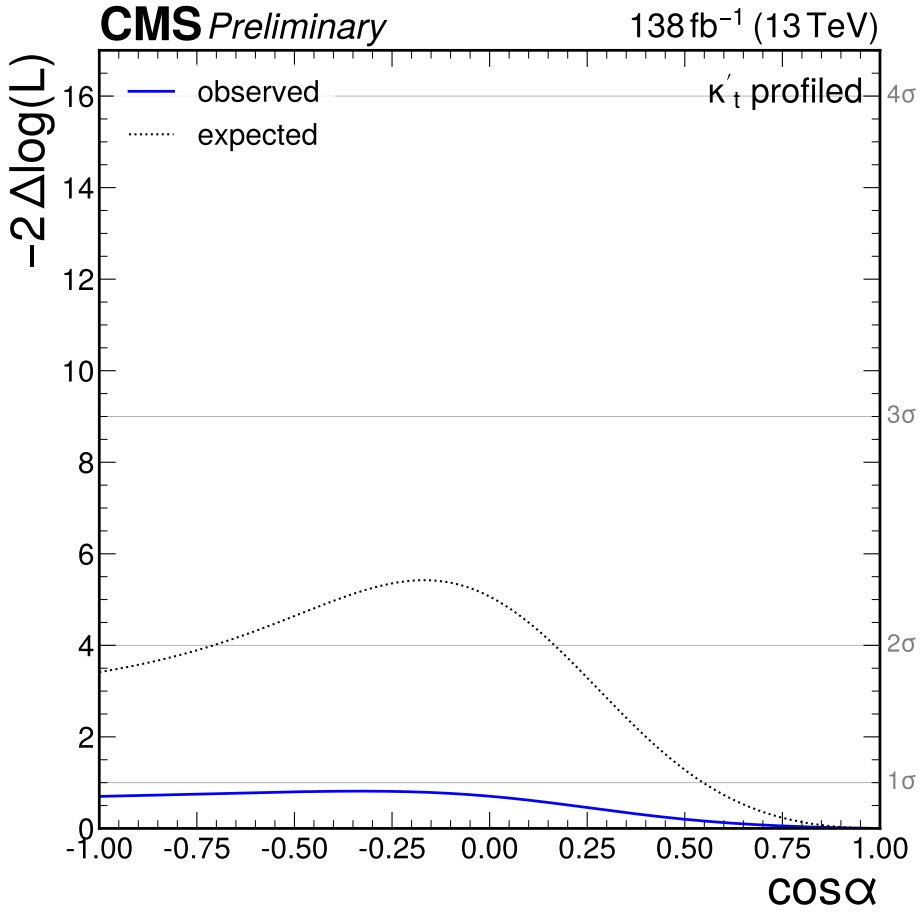


Figure 3.20: Asimov expected and observed likelihood scan for the mixing angle $\cos\alpha$ with profiled κ'_t. Published by the CMS collaboration in reference [59].

Table 3.12: Best fit values of $\cos\alpha$ and the background normalization parameters for the full analysis and splits of years/channels.

	$\cos\alpha$		κ'_t		$t\bar{t}+B$		$t\bar{t}+C$		$g \to b\bar{b}$ splitting	
2016	0.58	$^{+0.41}_{-1.58}$	0.98	$^{+0.55}_{-0.57}$	1.03	$^{+0.18}_{-0.15}$	0.83	$^{+0.29}_{-0.27}$	1.02	$^{+0.56}_{-0.56}$
2017	1.00	$^{+0.00}_{-0.65}$	0.62	$^{+0.41}_{-0.39}$	1.16	$^{+0.17}_{-0.15}$	0.96	$^{+0.31}_{-0.27}$	0.86	$^{+0.51}_{-0.45}$
2018	−1.00	$^{+2.00}_{-0.00}$	0.27	$^{+0.55}_{-0.27}$	1.34	$^{+0.18}_{-0.17}$	1.32	$^{+0.27}_{-0.24}$	−0.06	$^{+0.41}_{-0.32}$
DL	1.00	$^{+0.00}_{-2.00}$	0.07	$^{+0.61}_{-0.07}$	1.29	$^{+0.20}_{-0.19}$	1.41	$^{+0.41}_{-0.34}$	0.13	$^{+0.64}_{-0.48}$
SL	1.00	$^{+0.00}_{-2.00}$	0.76	$^{+9.24}_{-0.76}$	1.21	$^{+3.79}_{-6.21}$	1.38	$^{+3.62}_{-6.38}$	0.92	$^{+6.08}_{-7.92}$
Combined	1.00	$^{+0.00}_{-2.00}$	0.53	$^{+0.45}_{-0.34}$	1.20	$^{+0.13}_{-0.12}$	1.07	$^{+0.20}_{-0.19}$	0.50	$^{+0.32}_{-0.32}$

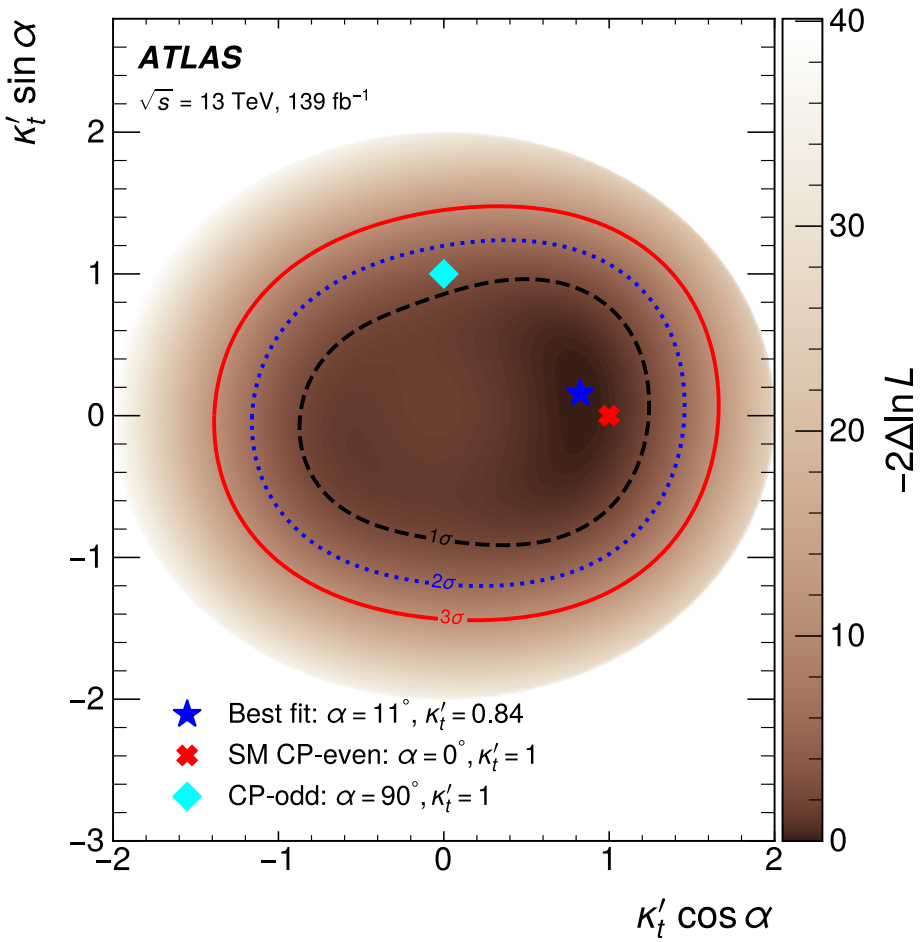


Figure 3.21: 2D likelihood scan in the $\mathcal{CP}$ coupling scenario observed by the ATLAS experiment [67] in the H → b$\bar{\text{b}}$ decay channel.

Table 3.13: The p-values from the different goodness-of-fit tests performed in the SM and $\mathcal{CP}$ coupling measurement for the combination of Higgs boson decay channels.

	p-value		
	saturated	KS	AD
SM coupling scenario	0.41	0.31	0.98
$\mathcal{CP}$ coupling scenario	0.40	0.44	—

3.7 Combination with other Higgs boson decay channels

The measurement in the $H \to b\bar{b}$ decay channel is combined with the $t\bar{t}H$ measurements performed in the $H \to ZZ/\gamma\gamma$ [6] and multilepton channel [7]. Both analyses provide COMBINE datacards compatible with the fit model described in section 3.6.1. Naming of nuisance parameters follows the recommendations by the CMS Collaboration so that the same nuisance parameters are correlated between analyses. There is no overlap in the selected events between the analyses so no special treatment is required in the combination.

In figure 3.22 the expected results from the combination in the $\mathcal{CP}$ coupling scenario is shown. The goodness-of-fit test results are given in table 3.13 and figure 3.23. Although the p-values are lower than before, they still are in an acceptable range.

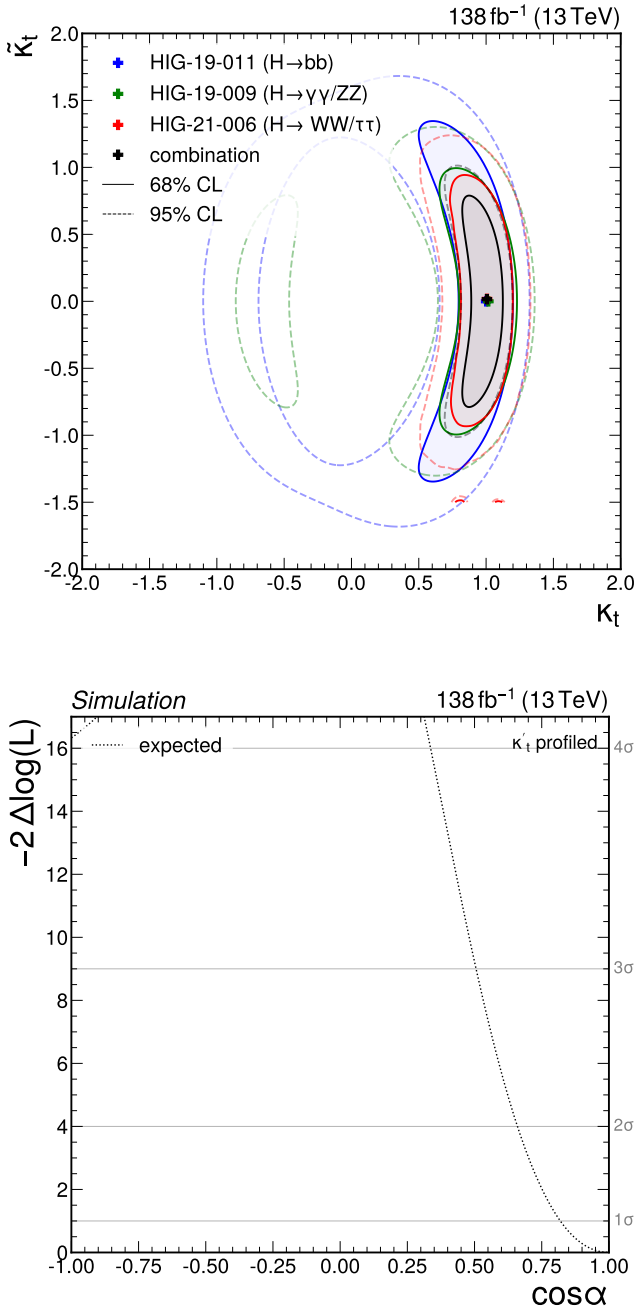
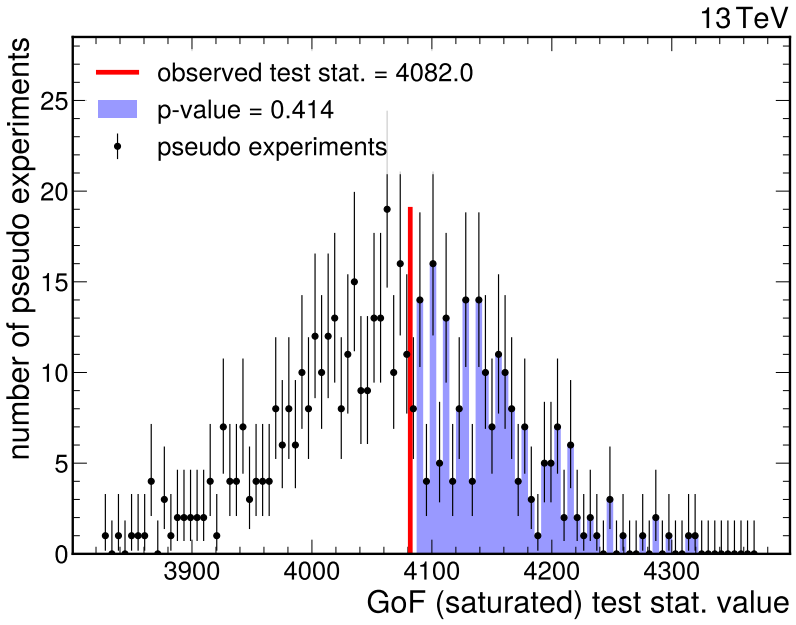
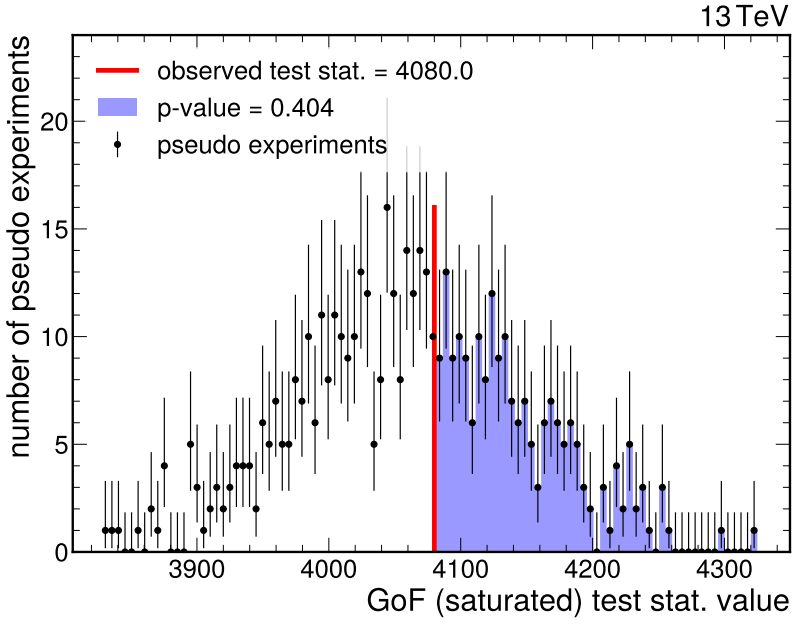


Figure 3.22: Asimov expectation for the combination of the $H \to b\bar{b}$, $H \to ZZ/\gamma\gamma$ and multilepton Higgs boson decay channels. The 2D likelihood scan in the $\mathcal{CP}$ coupling scenario for the combined results and individual analyses (top) and likelihood scan in $\cos\alpha$ with κ_t' profiled (bottom) are shown for a fixed $\kappa_V = 1$. Since the previous results are not available in the $\cos\alpha$ parametrization, only the combined expectation is given.



(a) SM coupling scenario



(b) $\mathcal{CP}$ coupling scenario

Figure 3.23: Distribution of the saturated goodness-of-fit test statistic for 500 pseudo experiments fitting in the SM (a) and $\mathcal{CP}$ coupling scenario (b) for the combination of Higgs boson decay channels. The distribution is fitted with a χ^2 distribution and the p-value of the goodness-of-fit test is shown.

4 Measurement of the top quark decay width in Wb scattering

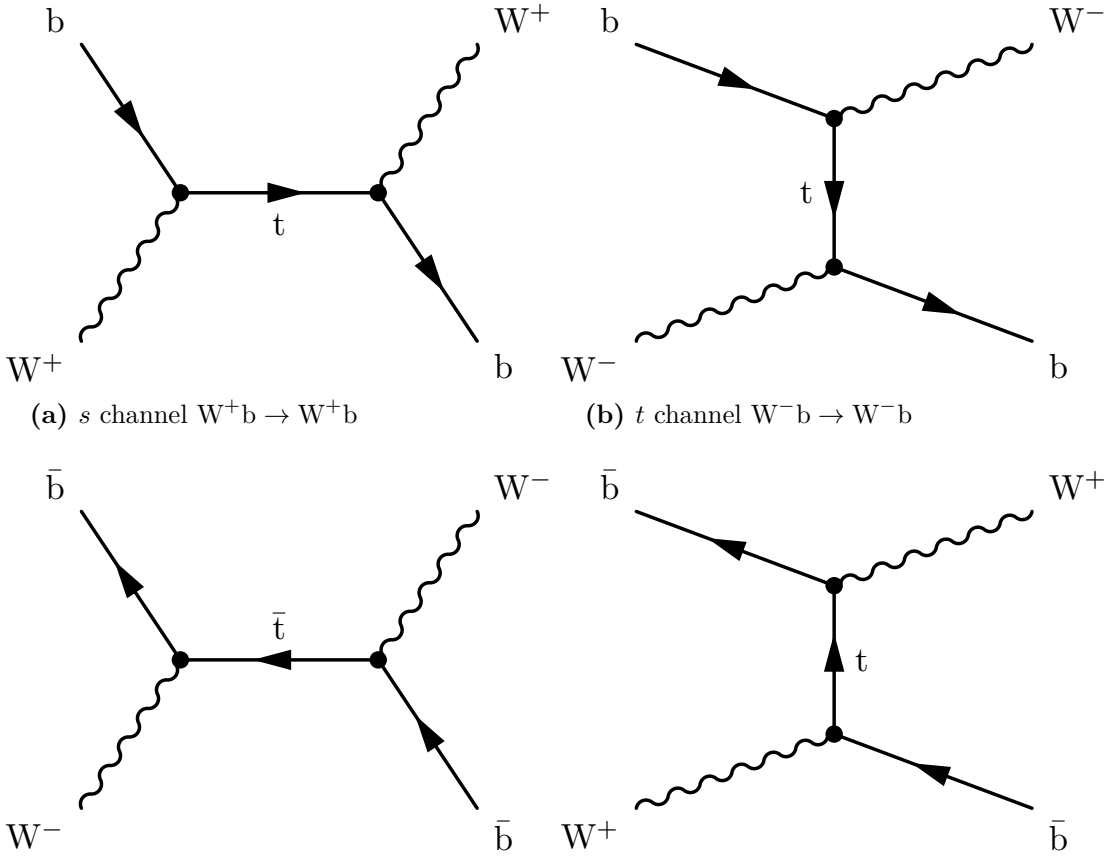
The current world average for the top quark decay width of $\Gamma_t = 1.42^{+0.19}_{-0.15}$ [58] is the combination of three independent measurements performed by the D0, CMS and ATLAS Collaboration. The measurements by the D0 [8] and CMS Collaboration [9] are obtained from the single top quark production cross section in combination with model-dependent assumptions on the branching ratio $BR_{t\to Wb}$, while the ATLAS measurement [10] follows a more direct approach by performing a template fit on kinematic variables like the b-jet lepton mass. Even though the direct measurement performed by ATLAS requires no assumption on the branching ratio, it is already systematically limited by experimental and theoretical uncertainties.

With the increasing luminosity of Run 3 and the HL-LHC, a new direct approach to measure the top quark decay width comes within in reach: The direct measurement in Wb scattering [11]. Based on this approach, an analysis using the full $138\,\mathrm{fb}^{-1}$ of data collected during Run 2 by the CMS experiment is developed. This approach will remain relevant for the upcoming Run 3 and the following HL-LHC era due the advantage of small systematic uncertainties.

4.1 Direct measurement approach

The measurement of the top quark decay width is performed in the Wb scattering process. It can be divided into two major contributions: s (resonant) and t channel (non-resonant) process. In each channel there exist two Feynman diagrams, one for each W boson charge as depicted in figure 4.1. The s and t channels can be distinguished by the charge of the W boson and of the b quark. In the s channel process, the W boson and b quark charges have opposite signs while for the t channel they have the same sign.

Since W bosons and b quarks in the 4-flavor scheme are not constituents of protons, they need to be produced by additional interactions. The W boson is radiated off a light quark while the b quark is created by gluon splitting. This leads to



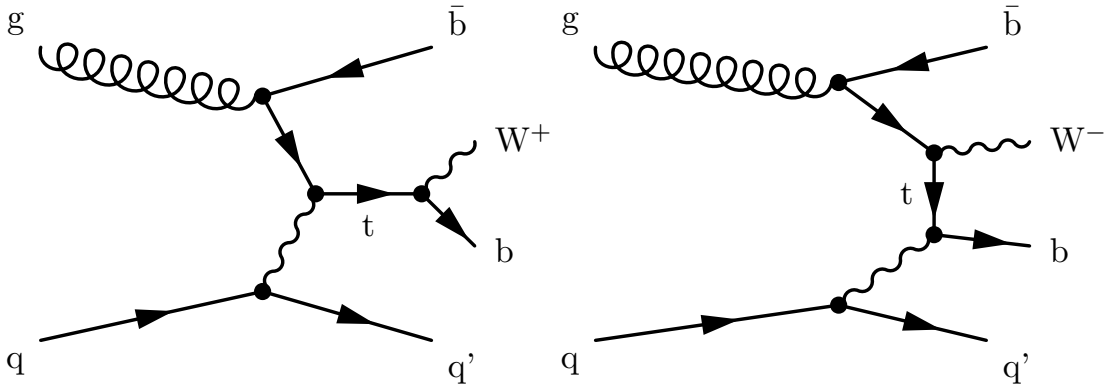
(a) s channel $W^+b \to W^+b$ **(b)** t channel $W^-b \to W^-b$

(c) s channel $W^-\bar{b} \to W^-\bar{b}$ **(d)** t channel $W^+\bar{b} \to W^+\bar{b}$

Figure 4.1: Feynman diagrams for s and t channel Wb processes. For each channel, both the $Wb \to Wb$ and the $W\bar{b} \to W\bar{b}$ diagrams are shown.

two additional final-state particles, a light quark q' and a b quark, as depicted in figure 4.2. A dedicated signal MC dataset is simulated using MadGraph5. Event weights are produced to reweight to different combinations of top quark masses from 171.5 GeV to 173.5 GeV and top quark decay widths from 0.5 GeV to 10 GeV.

Looking at the mass reconstructed from the W boson and b quark in MC simulation of Wb scattering for s and t channel, depicted in figure 4.3, differences between both channels become apparent: While t channel production has only mild dependence on the reconstructed mass, the s channel distribution shows a resonant peak around the top quark mass. Due to this resonant production of a top quark in s channel process, the cross section has a strong dependency on the top quark decay width. This is expected for a s channel process and can be exploited to precisely measure the top quark decay width by using the t channel to reduce the effect of systematic uncertainties.



(a) s channel **(b)** t channel

Figure 4.2: Feynman diagrams for Wb → Wb scattering in the 4-flavor scheme, producing an additional b and light quark.

The generator production cross section for Wb scattering and a subsequent decay of the W boson into leptons for a top quark mass of $m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$ and a top quark decay width of $\Gamma_{\mathrm{t}} = 1.322\,\mathrm{GeV}$ in leading order (LO) QCD is:

$$\sigma_{\text{Wb scattering}} \times \mathrm{BR}_{\mathrm{W}\to\ell\nu} = 64.55\,\mathrm{pb} \tag{4.1}$$

In figure 4.4, the dependence of the Wb cross section on the top quark mass and decay width is shown. While the Wb cross section has only a small dependency on the top quark mass, the Wb cross section diverges towards smaller top quark decay widths. The majority of the cross section stems from the resonant s channel process.

4.2 Event selection

Events with exactly one electron or one muon are selected. The electron needs at least a transverse momentum of 29 GeV, $|\eta| < 2.4$ and satisfy an electron identification criterion to be considered. Further quality cuts are applied to the displacement of the electron from the primary vertex: $d_{xy} < 0.05\,\mathrm{cm}$ and $d_z < 0.1\,\mathrm{cm}$ for $|\eta| < 1.479$ and $d_{xy} < 0.1\,\mathrm{cm}$ and $d_z < 0.2\,\mathrm{cm}$ else. The event is rejected if there is a second electron with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ in the central region of $|\eta| < 2.4$. Muons with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are considered in the central part if they fulfill the tight muon identification and isolation criteria. Events with a second muon with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ in the central part are rejected. In table 4.1, the trigger paths of the analysis, selecting events containing electrons or muons, are listed.

Jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 4.8$ are considered. In a ΔR-cleaning, jets within $\Delta R < 0.4$ of the electron or muon are removed. Jets in the central part are categorized

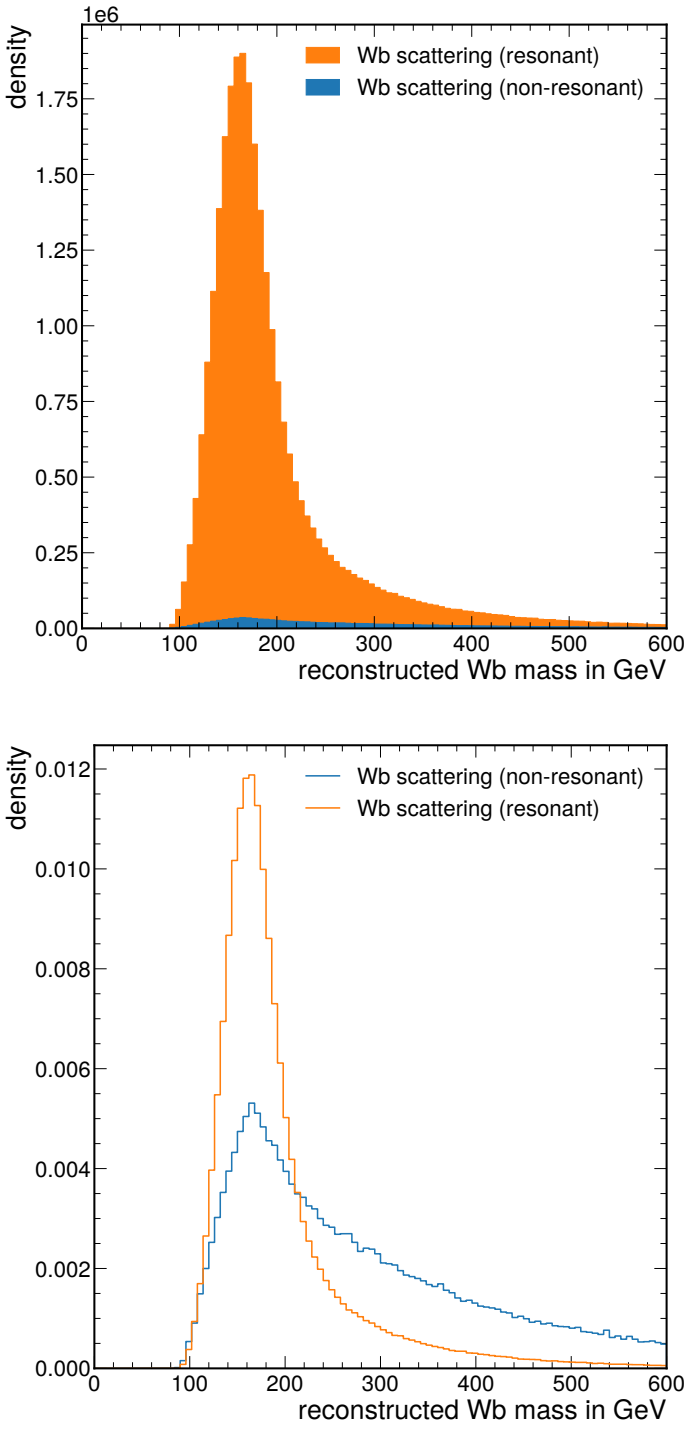


Figure 4.3: Differences in the stacked (top) and normalized number of events (bottom) distributions between s (on-shell) and t channel (off-shell) in the reconstructed Wb mass.

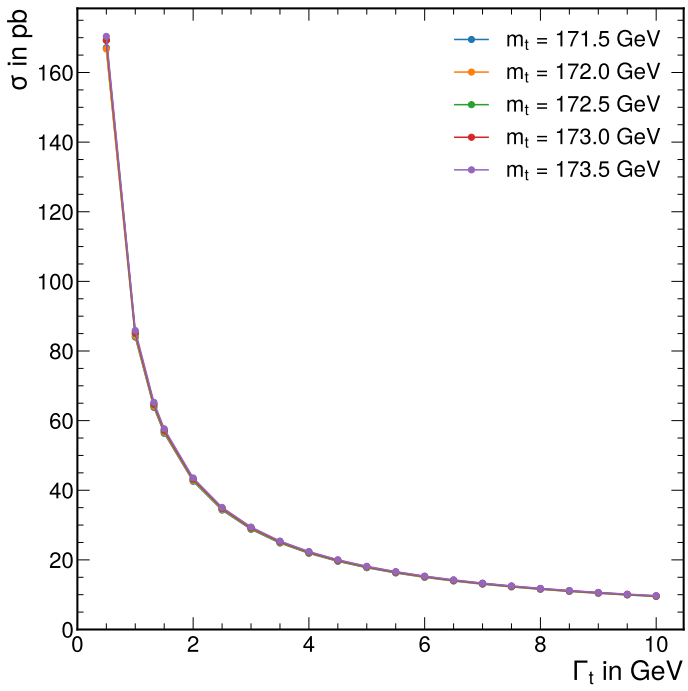


Figure 4.4: Dependence of the Wb cross section on the top quark mass and decay width. A strong dependency on the top quark decay width Γ_t is visible while the curves for different top quark masses show only minor differences.

Table 4.1: HLT trigger paths used in the Wb analysis. The triggers collect events with at least one electron or muon passing certain thresholds.

path	**year**
Ele27_WPTight_Gsf	2016
Ele32_WPTight_Gsf_L1DoubleEG	2017
Ele32_WPTight_Gsf	2018
IsoMu24	2016, 2018
IsoTkMu24	2016
IsoMu27	2017

Table 4.2: Event selection defining the signal and control regions of the Wb → Wb analysis. Separate regions are defined for the lepton being an electron or muon.

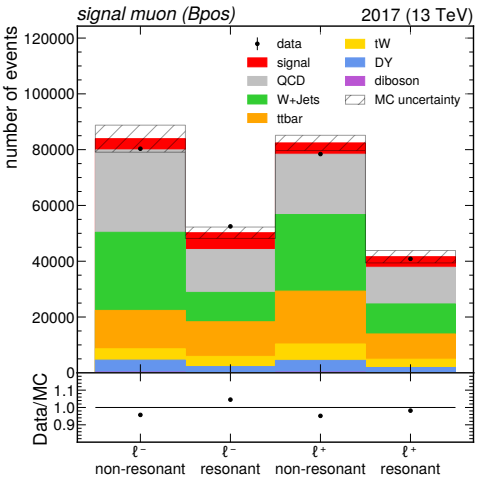
region	number of jets	b-tags	electrons	muons
W + jets control region (electron)	2	0	1	0
W + jets control region (muon)	2	0	0	1
signal region (electron)	2	1	1	0
signal region (muon)	2	1	0	1
$t\bar{t}$ control region (1 tag, electron)	3	1	1	0
$t\bar{t}$ control region (1 tag, muon)	3	1	0	1
$t\bar{t}$ control region (2 tag, electron)	3	2	1	0
$t\bar{t}$ control region (2 tag, muon)	3	2	0	1

into tagged and untagged jets using a tight working point with the DeepJet b-tagging algorithm. Depending on the total number of jets, b-tags and flavor of the lepton, events are categorized in regions. Apart from the signal region, control regions enriched with W + jets, QCD multijets or $t\bar{t}$ events are defined. In table 4.2 the selection criteria for the signal and control regions are listed.

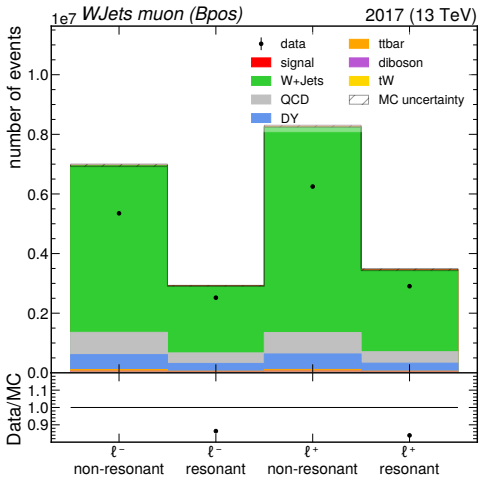
Further enrichment with signal events can be achieved using machine learning and is going to be performed with BDTs.

4.3 Background processes

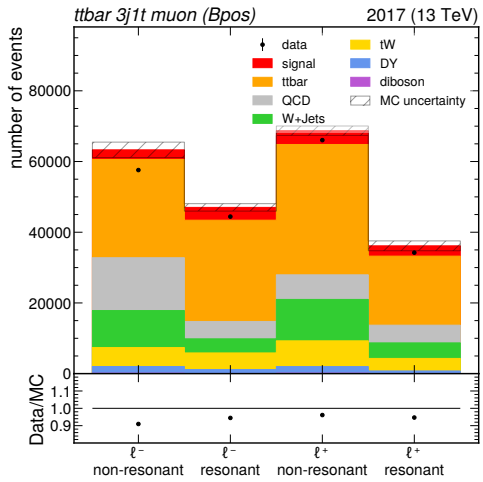
The largest background in the signal region is from non-reducible QCD multijet production, which is not well modeled in simulation and thus derived from data in a dedicated control region. The second largest background is the production of W bosons and additional jets (W+jets), which gives a similar signal as the Wb scattering processes due to the presence of a W boson. Due to the enrichment of the signal region with b-tagged jets, a large background contribution is expected from $t\bar{t}$ and tW production. For $t\bar{t}$ events, the normalization is determined in a dedicated control region. A smaller background contribution is expected from Drell-Yan processes and diboson production. In figure 4.5, the expected signal and background contributions for the B^+ charge tagger category, introduced in section 4.6, are shown together with the observed number of events. The distributions for all categories and regions can be found in appendix B.



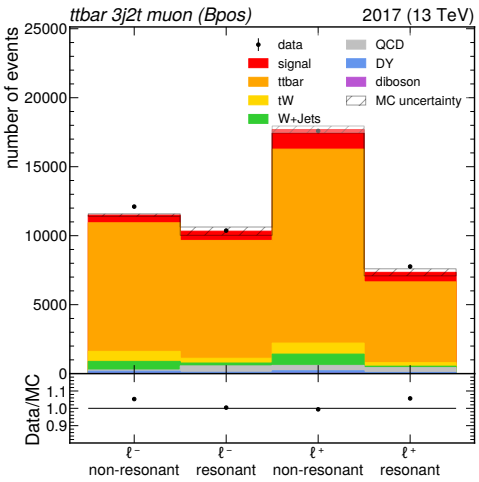
(a) signal muon region



(b) W + jets muon region



(c) $t\bar{t}$ (1 tag) muon region



(d) $t\bar{t}$ (2 tag) muon region

Figure 4.5: Pre-fit distributions for the B^+ category charge tagger in the four muon regions for 2017. The order of the background is by size.

4.4 Event reconstruction

The same method as used in the (t)tH analysis, described in section 3.4.2, is used to reconstruct a W boson from the lepton ℓ and missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$. Again the solution with the smallest absolute value of p_z^{miss} is used when solving for the literature value of the W boson mass.

The four-vectors of the b-tagged jet with the highest transverse momentum and the reconstructed W boson are used to reconstruct the Wb mass m_{Wb}. In case of no b-tagged jet, the jet with the highest transverse momentum is used instead.

4.5 b charge tagging

Differentiating between Wb and W$\bar{\mathrm{b}}$ scattering requires the ability to predict the charge of the b quark. Since b quarks are color charged, they form hadrons and are reconstructed as jets in the detector. While flavor tagging is a common technique employed at the CMS detector, there currently exists no method to determine the charge of a jet. Inspired by the DeepJet algorithm, a b charge tagger is developed to predict the hadron contained inside a b-tagged jet [68]. One output category for each B hadron is defined: B^+, B_0, $\bar{\mathrm{B}}_0$ or B^-. An in-situ calibration of the b charge tagger is going to be performed in the t$\bar{\mathrm{t}}$ control regions.

In figure 4.6, receiver operating characteristic (ROC) curve and distribution of the sum of the B_0 and B^+ output nodes are shown. The performance of the tagger is higher than that of a simple charged sum approach, where all charges of the particles in the jet are summed up. Unsurprisingly it is visibly harder to differentiate between neutral B mesons (B_0 and $\bar{\mathrm{B}}_0$) than between charged B mesons (B^+ and B^-). The sum of the B_0 and B^+ output can be used to discriminate $\bar{\mathrm{b}}$ quark against b, as both B mesons contain a $\bar{\mathrm{b}}$ quark. A large overlap in the distributions for the neutral B mesons can be observed, explaining the difficulties to correctly classify them.

At the point of equal b and $\bar{\mathrm{b}}$ identification efficiency, the developed b charge tagger is able to correctly classify the charge of the b or $\bar{\mathrm{b}}$ quark in 65.9 % of the cases. This performance is comparable to the performance of the b charge tagger developed by the ATLAS experiment [69].

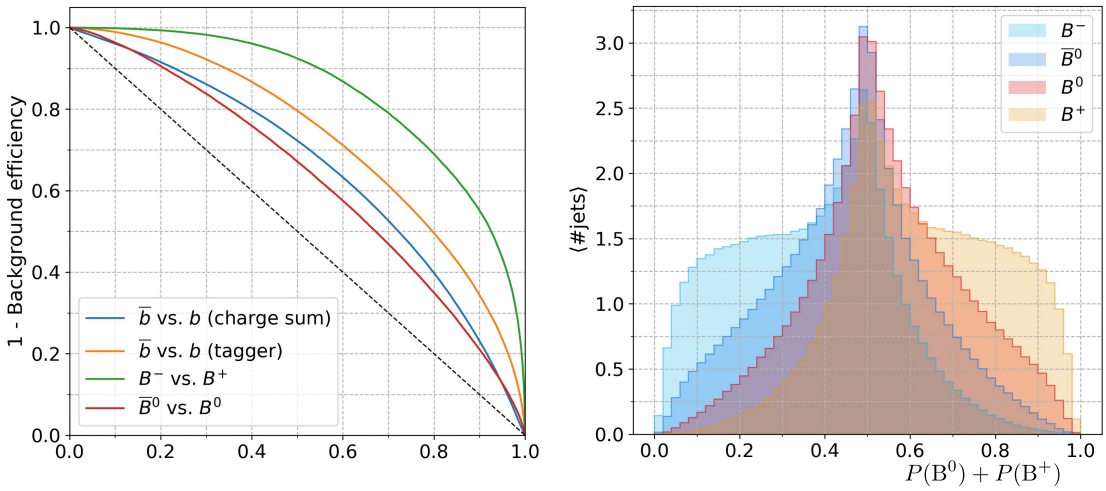


Figure 4.6: ROC curve of the b charge tagger compared to a simple charge sum approach (left) and b charge tagger output for $\bar{\mathrm{b}}$ classification (right) [68]. The ROC curve of the tagger is split into curves for charged (B^+ and B^-) and neutral B mesons (B_0 and $\bar{\mathrm{B}}_0$). It is much harder to correctly predict the charge of the b in neutral mesons than in charged ones. This is also visible when looking at the output for $\bar{\mathrm{b}}$ classification, where the distributions for neutral mesons cannot easily be separated and significantly overlap. Compared to calculating a sum over all charged particles in a jet, the tagger shows better performance.

4.6 Event categories and fit model

The events are categorized based on the highest b charge tagging output into B^+, B_0, $\bar{B}_0$ or B^-, charge of the lepton, which is identical to the lepton charge, and resonant or non-resonant scattering. The distinction into resonant or non-resonant is realized with a $\pm 25\,\text{GeV}$ window around a top quark mass of $m_t = 172.5\,\text{GeV}$ on the reconstructed Wb mass m_{Wb}. The content of these 16 categories per region is evaluated in a fit over all regions in different top quark mass and width scenarios.

A COMBINE [60] model is used to scale the signal contribution into different top quark mass and width scenarios. The model works by linear interpolating between the four closest signal templates out of the 105 scenarios simulated in the signal dataset. This makes it possible to fit any top quark mass and width scenario withing the boundaries of $171.5\,\text{GeV} \leq m_t \leq 173.5\,\text{GeV}$ and $0.5\,\text{GeV} \leq \Gamma_t \leq 10\,\text{GeV}$.

4.7 Expected result

In figure 4.7, the likelihood scan in top quark mass and width for an expectation of $m_t = 172.5\,\text{GeV}$ and $\Gamma_t = 1.322\,\text{GeV}$ is shown. This preliminary result is obtained in the signal and $t\bar{t}$ regions with two uncertainties implemented: a 2 % luminosity rate uncertainty and the statistical bin-by-bin MC uncertainties, as described in section 3.4.4. Instead of using the data-driven QCD estimation, the QCD background is currently derived from MC simulation with an additional normalization scale of 0.6 to match the observed data points by eye. The b charge tagger calibration is also not yet implemented. Even though the final result is going to differ, the fit model can be tested and issues can be solved early. In terms of sensitivity, the expected result shows a feature also expected from the final result: Low sensitivity to the top width Γ_t due to statistical limitations and only minor dependence on the top quark mass. Once the proposed analysis is fully implemented, it can serve as a base for future analyses performed with data from the currently ongoing Run 3 and following HL-LHC, where through the increase in acquired data the full potential of this approach can be utilized.

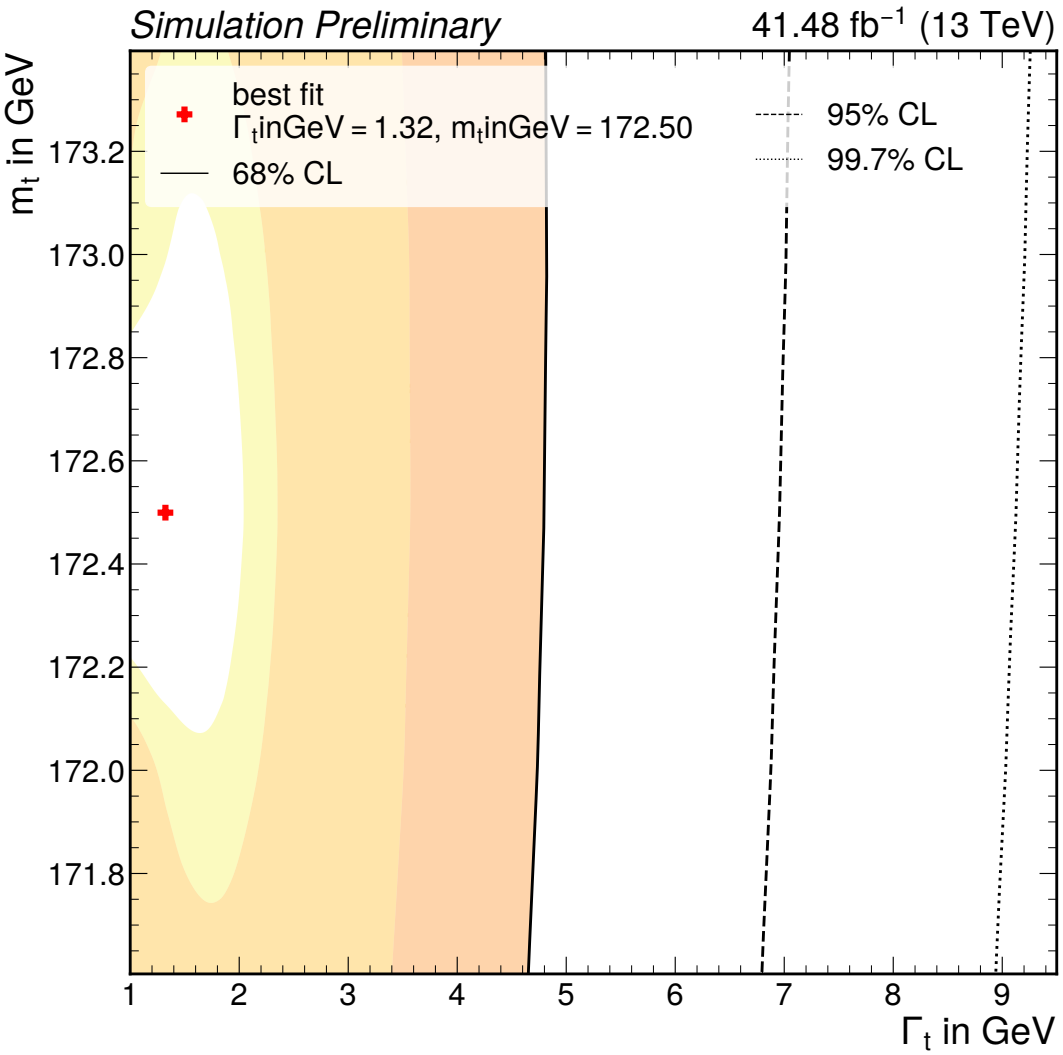


Figure 4.7: Asimov expected likelihood scan for the measurement of the top quark decay width using the signal and $t\bar{t}$ control regions of the Wb analysis. The fit is performed with a small set of uncertainties and due to change when the full set of uncertainties gets implemented. As expectation the values of $m_t = 172.5\,\text{GeV}$ and $\Gamma_t = 1.322\,\text{GeV}$ are used.

5 Study of the b-tagging performance at trigger level in 2022

With the start of the Run 3 data taking period at the LHC in 2022, the High Level Triggers (HLTs) level at the CMS experiment switched to an improved method to trigger on heavy flavor jets. The previously used DeepCSV flavor tagging algorithm [70] is replaced with the DeepJet flavor tagging algorithm [44], already used for offline flavor tagging during Run 2. Both are neural network based algorithms utilizing jet kinematics, substructure and secondary vertices. While DeepCSV is just a simple fully connected neural network, DeepJet uses convolutional layers and recurrent neural networks together with additional low-level features to increase performance. With DeepJet a lower light flavor misidentification rate is expected for a given b jet identification efficiency [71]. The 2022 data taking period also marks the first time a dedicated training of the tagging neural networks is performed with online reconstructed jets. This enables a further increase in tagging efficiency by accounting for the differences between the online and offline reconstruction.

The working point is the tagging output value above which jets are tagged as b jet. Using the working point, the b jet identification efficiency and light flavor misidentification rate are defined as:

$$\text{b jet identification efficiency} = \frac{N_{\text{tag \& b jet}}}{N_{\text{b jet}}} \tag{5.1}$$

$$\text{light flavor jet misidentification rate} = \frac{N_{\text{tag \& light jet}}}{N_{\text{light jet}}} \tag{5.2}$$

For offline b-tagging, three working points are defined by a light flavor jet misidentification rate of 0.1, 0.01 and 0.001, called loose, medium and tight respectively. For the b-tagging application at HLT level, the definition of general working points is not sensible, because the tagging is applied after a trigger specific selection and the approved trigger rate should be utilized completely. The selection of the working point is then tuned together with other selection criteria applied in the trigger.

To estimate trigger specific working points when switching to the new flavor tagging algorithm, the response of the DeepCSV and DeepJet algorithms is evaluated on

the same QCD enriched MC simulated dataset. The simulation is tuned for Run 3 conditions assuming a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$. In figure 5.1, the ROC curve as well as the light flavor jet misidentification rate and b jet identification efficiency against the working point are shown for the DeepCSV and DeepJet algorithms at HLT level. From these curves, a starting point for the tuning of the working points can be selected by the developer of trigger paths.

To study the pileup dependency and confirm the functionality of the b-tagging trigger sequences in 2022 collisions, triggers utilizing DeepJet b-tagging are compared to reference triggers. The reference triggers are identical to the examined triggers, but without b-tagging selection criterion. Due to the missing selection criterion, the rate of the reference triggers is higher. To mitigate the higher rate, these triggers are prescaled, meaning only a fraction of events selected by the reference trigger is actually saved.

Events from a dataset enriched in muons and leptons, collected by the CMS experiment in 2022, are used for the study. Further cleaning is performed using the recommended luminosity masks [72]. An event selection is used to enrich the phase space with heavy flavor jets. The event selection targets top quark pair production with one electron and one muon from the decay of the two W bosons:

- at least **2 jets** with $p_\text{T} > 30\,\text{GeV}$ and $|\eta| < 2.5$
- exactly **1 electron** with $p_\text{T} > 10\,\text{GeV}$, $|\eta| < 2.5$, $|d_z| < 0.2$ and $|d_{xy}| < 0.1$
- exactly **1 muon** with $p_\text{T} > 10\,\text{GeV}$, $|\eta| < 2.5$, $|d_z| < 0.2$ and $|d_{xy}| < 0.1$
- electron and muon have **opposite charges**
- dilepton mass $m_{\text{e}\mu} > 20\,\text{GeV}$

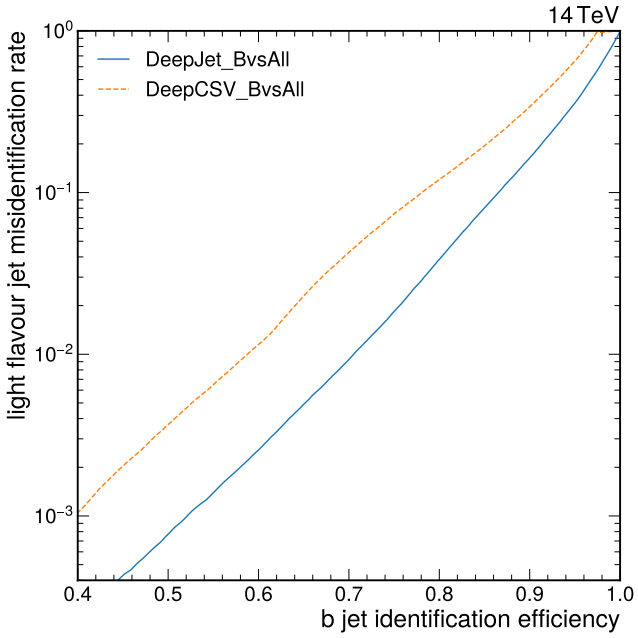
These selections allow a comparison with results obtained by analysis groups, which use the same selection criteria.

Furthermore, only events where the reference trigger fired are considered. An efficiency is defined as the fraction of events where the trigger fired:

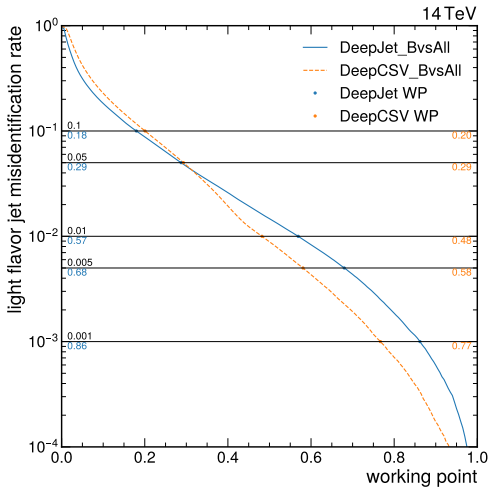
$$\epsilon = \frac{N_{\text{selection \& reference \& trigger}}}{N_{\text{selection \& reference}}} \tag{5.3}$$

This efficiency describes the fraction of events passing the b-tagging selection criterion of the trigger.

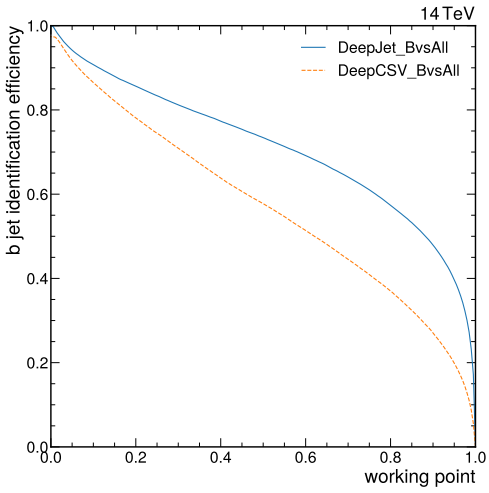
A variable that can be used to study pileup is the number of primary vertices. Unfortunately it is also an input to the DeepJet tagging algorithm. To rule out



(a) Receiver operating characteristic



(b) light flavor jet misidentification rate



(c) b jet identification efficiency

Figure 5.1: Performance of DeepJet and DeepCSV at HLT level evaluated on a QCD enriched MC simulated dataset. The ROC curve shows the performance advantage of DeepJet over DeepCSV (a). The light flavor jet misidentification rate (b) and b jet identification efficiency (c) in relation to the working point is given. For different values of the light flavor jet misidentification rate, the working point for both flavor tagging algorithm is marked.

bias, the average instantaneous luminosity is assigned to events based on their luminosity section. For studying the pileup dependence of the tagging sequence, events are binned in the number of offline reconstructed primary vertices and average instantaneous luminosity. In figure 5.2, the results of this study are shown for two different triggers. Only minimal dependency on pileup is observed.

The tagging sequence turn-on curves are studied by binning the efficiency in the offline DeepJet b-tag value. At the time of the study, the updated training of the offline DeepJet algorithm for Run 3 was still in development and the Run 2 DeepJet model was used. The turn-on is visible in the n-highest offline DeepJet b-tagging value, with the n corresponding to the number of b-tags selected at trigger level. Binned in the m-highest offline DeepJet b-tagging value with $0 < m < n$, the efficiency is flat. In figure 5.3, the turn-on curves for three different triggers, requiring one, two or three b-tags, are shown. The turn-on curves show the preference of the tagging sequences to select events with higher offline DeepJet b-tagging values, as intended. Overall the new DeepJet tagging algorithm a HLT-level works as expected and can continue to do so with the increasing pileup in future data taking conditions.

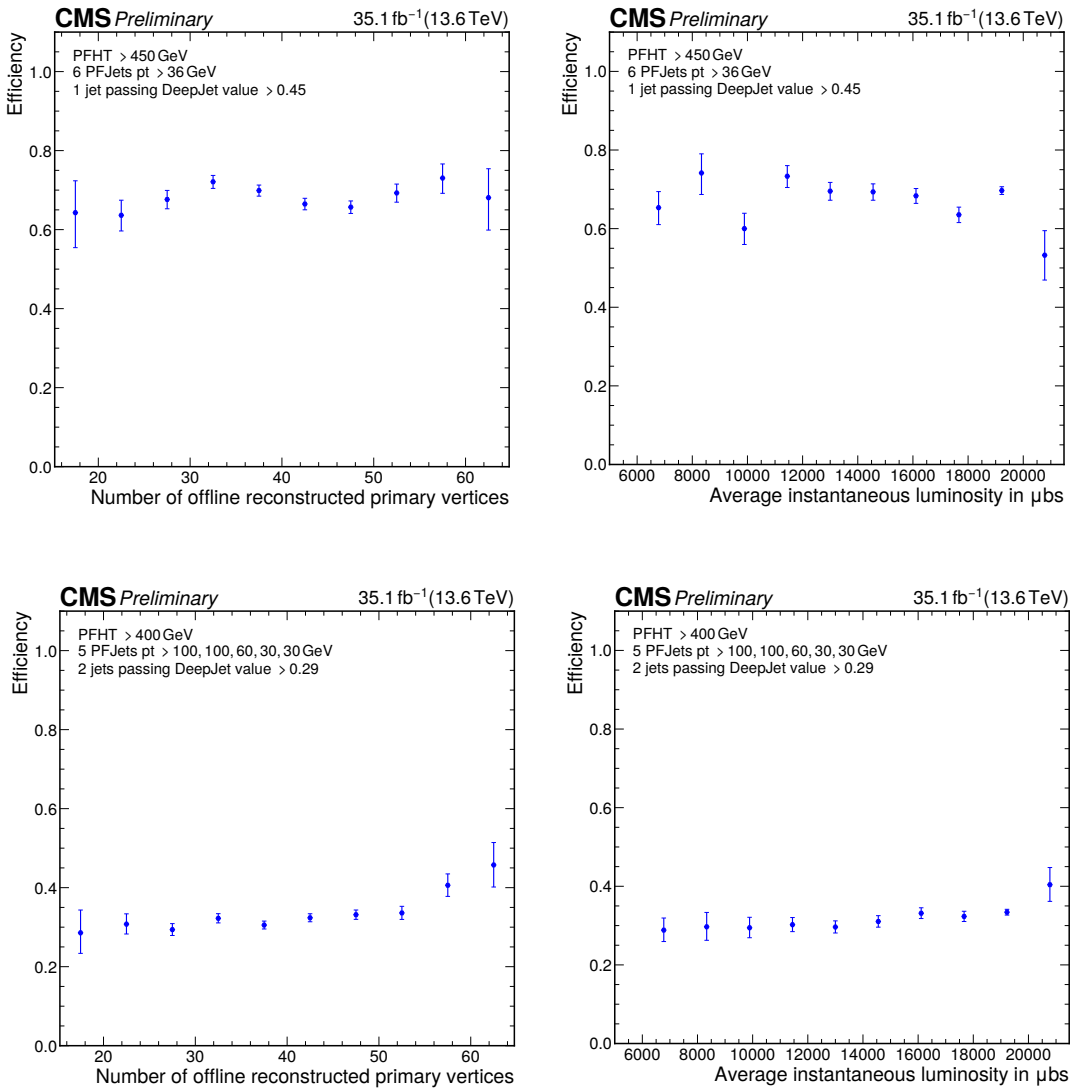


Figure 5.2: Measured pileup dependency for two DeepJet b-tagging sequences. The efficiency is shown in bins of number of offline reconstructed primary vertices (left columns) and in bins of average instantaneous luminosity (right column). No strong pileup dependency can be observed. Published by the CMS collaboration in reference [72].

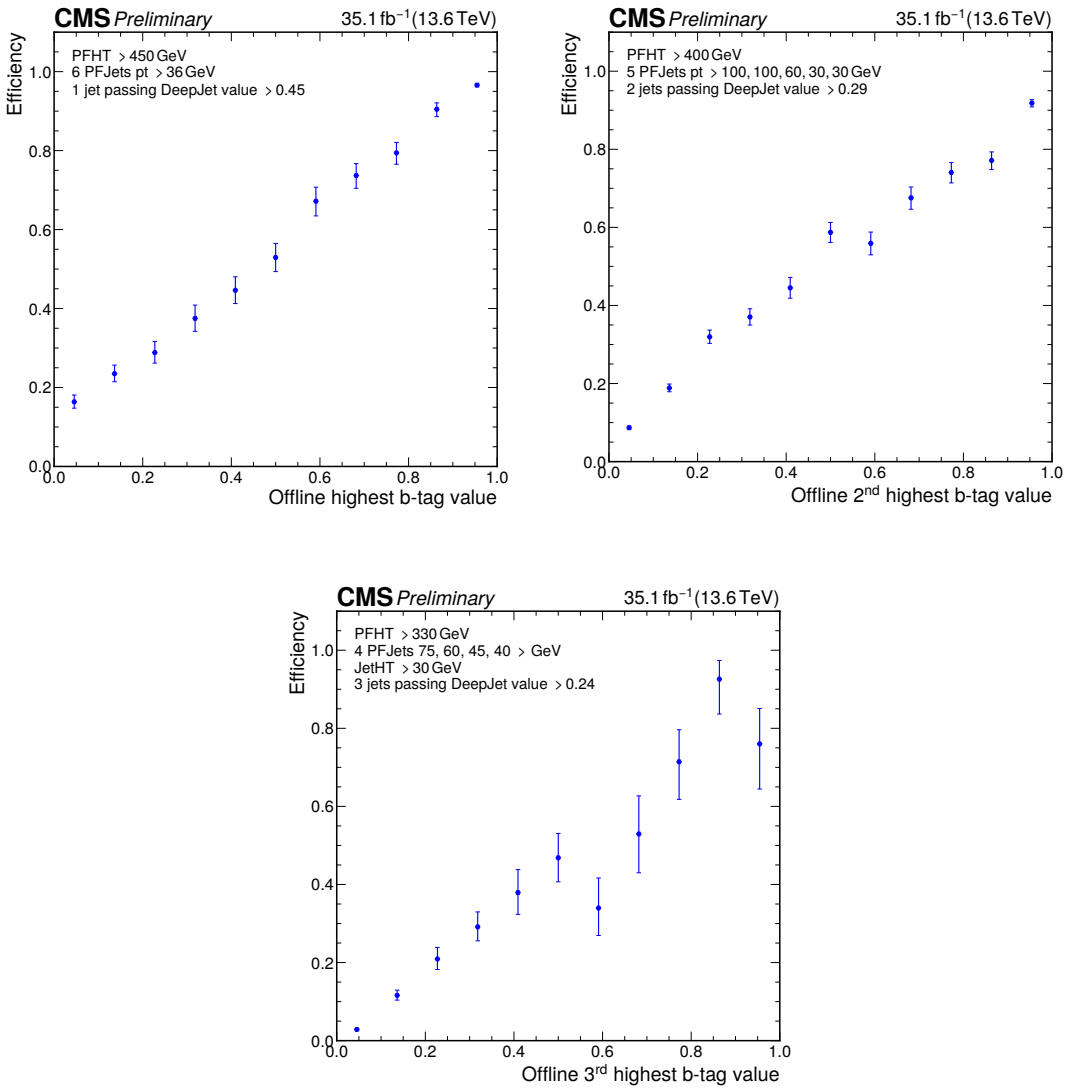


Figure 5.3: Turn-on curves for three different DeepJet b-tagging trigger sequences. The efficiency is shown in bins of n-highest offline b-tag value, where n is the number of b-tags selected in the trigger. Published by the CMS collaboration in reference [72].

6 Summary

The measurement of the top-Higgs coupling in tH and $t\bar{t}H$ production presented in this thesis contains the first measurement of the $\mathcal{CP}$ mixing in the $H \to b\bar{b}$ decay channel at the CMS experiment. Together with the measurement of the top-Higgs coupling in the SM scenario and the search for tH production, it complements the inclusive measurement of $t\bar{t}H$ production. With the performed coupling measurement in the SM scenario, the inverse top coupling scenario ($\kappa_t = -1, \kappa_V = 1$) can be excluded with more than 99 % CL. Due to a down fluctuation of the signal contribution, no strict limits on the $\mathcal{CP}$ mixing of the coupling can be given, but the observation favors a pure SM-like scalar Higgs boson. For the upcoming Higgs combination of the Run 2 measurements by the CMS experiment, the result is going to be combined with the $H \to ZZ/\gamma\gamma$ and multilepton Higgs boson decay channel. An upper limit on the tH production of 14.6 ($19.3^{+9.2}_{-6.0}$) times the SM cross section is observed (expected) with 95 % CL.

The described novel approach to directly measure the top quark width in Wb scattering bypasses systematic limitations and model dependencies present in the current measurements. With the increasing size of data collected during Run 3 and the upcoming HL-LHC, there is the potential to overcome the statistical limitations and precisely measure the top quark decay width totally free of model dependencies.

For the currently ongoing Run 3 of the LHC, the improved performance of the b-tagging algorithm at HLT level is a game changer in terms of identification efficiency and stability against conditions with increased pileup. This is of great importance for future analyses involving top quarks, Higgs bosons or more exotic particles.

With the advancing of technology and the increased amount of data collected by future collider experiments, the adventure to solve natures last mysteries continues. Already now, there is an ongoing effort to measure the Higgs boson self coupling in $t\bar{t}HH$ production, using the (t)tH analysis described here as a starting point. Indirect observation of dark matter and neutrino experiments challenge our knowledge of the universe. The mission to unravel the last secret of nature is still ongoing and a bright future of particle physics lays ahead.

List of Acronyms

ALICE	A Large Ion Collider Experiment
ATLAS	A Toroidal LHC ApparatuS
BDTs	boosted decision trees
BSM	physics beyond the Standard Model
CERN	European Organization for Nuclear Research
CL	confidence level
CMS	Compact Muon Solenoid
CR	control region
CSC	cathode strip chamber
DL	dilepton
DT	drift tube
ECAL	Electromagnetic Calorimeter
ELENA	Extra Low ENergy Antiproton ring
FH	full hadronic
FSR	final-state radiation
ggF	gluon-gluon fusion
HCAL	Hadron Calorimeter
HLT	High Level Trigger
ISR	initial-state radiation
ITC	inverse top coupling
JER	jet energy resolution
JES	jet energy scale
L1	Level 1 trigger
LEP	Large Electron Positron Collider
LHC	Large Hadron Collider

LHCb	Large Hadron Collider beauty
LO	leading order
MC	Monte Carlo
NLL	negative log likelihood
NLO	next-to-leading order
NNLO	next-to-next-to-leading order
PCB	printed circuit board
PDF	parton density function
PF	Particle Flow
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
QCD	quantum chromodynamics
QED	quantum electrodynamics
ROC	receiver operating characteristic
RPC	resistive plate chamber
SL	single lepton
SM	Standard Model of particle physics
SPS	Super Proton Synchrotron
STXS	simplified template cross sections
VBF	vector boson fusion

List of Figures

List of Tables

Bibliography

[1] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235v2 [hep-ex].

[2] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214v2 [hep-ex].

[3] Super-Kamiokande Collaboration. "Evidence for oscillation of atmospheric neutrinos". In: *Physical Review Letters* 81.8 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003v2 [hep-ex].

[4] CMS Collaboration. "Observation of $t\bar{t}H$ production". In: *Physical Review Letters* 120.23 (2018). DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610v2 [hep-ex].

[5] CMS Collaboration. "Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Physical Review D* 99.9 (2019). DOI: 10.1103/PhysRevD.99.092005. arXiv: 1811.09696v2 [hep-ex].

[6] CMS Collaboration. "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state". In: *Physical Review D* 104.5 (2021). DOI: 10.1103/PhysRevD.104.052004. arXiv: 2104.12152v2 [hep-ex].

[7] CMS Collaboration. "Search for CP violation in ttH and tH production in multilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV". 2022. arXiv: 2208.02686v1 [hep-ex].

[8] D0 Collaboration. "Improved determination of the width of the top quark". In: *Physical Review D* 85.9 (2012), p. 091104. DOI: 10.1103/physrevd.85.091104. arXiv: 1201.4156 [hep-ex].

[9] CMS Collaboration. "Measurement of the ratio $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV". In: *Physics Letters B* 736 (2014), pp. 33–57. DOI: 10.1016/j.physletb.2014.06.076. arXiv: 1404.2292 [hep-ex].

[10] ATLAS Collaboration. "Direct top-quark decay width measurement in the $t\bar{t}$ lepton+jets channel at $\sqrt{s}$=8 TeV with the ATLAS experiment". In: *The European Physical Journal C* 78.2 (2018). DOI: 10.1140/epjc/s10052-018-5595-5. arXiv: 1709.04207 [hep-ex].

[11] P. P. Giardino and C. Zhang. "Probing the top-quark width using the charge identification of b jets". In: *Physical Review D* 96.1 (2017). DOI: 10.1103/PhysRevD.96.011901. arXiv: 1702.06996v2 [hep-ph].

[12] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/ptp.49.652.

[13] M. Tanabashi et al. "Review of Particle Physics". In: *Physical Review D* 98.3 (2018). DOI: 10.1103/physrevd.98.030001.

[14] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". In: *Physical Review* 76.6 (1949), pp. 769–789. DOI: 10.1103/physrev.76.769.

[15] Wikimedia Commons. *Standard model of elementary particles: the 12 fundamental fermions and 5 fundamental bosons*. 2019. URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (visited on 04/19/2023).

[16] S. Weinberg. "A Model of Leptons". In: *Physical Review Letters* 19.21 (1967), pp. 1264–1266. DOI: 10.1103/physrevlett.19.1264.

[17] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*. eng. 2013. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347v2 [hep-ph]. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR2014.

[18] CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". In: *Nature* 607.7917 (2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043v3 [hep-ex].

[19] S. Mandelstam. "Determination of the Pion-Nucleon Scattering Amplitude from Dispersion Relations and Unitarity. General Theory". In: *Physical Review* 112.4 (1958), pp. 1344–1360. DOI: 10.1103/physrev.112.1344.

[20] C. Escobar et al. *NLO single-top channel cross sections*. 2017. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec.

[21] Planck Collaboration. "Planck 2015 results. XIII. Cosmological parameters". In: *Astronomy & Astrophysics* 594 (2015), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589v3 [astro-ph.CO].

[22] O. S. Brüning et al. *LHC Design Report*. eng. 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[23] *Public CMS Luminosity Information*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (visited on 04/17/2023).

[24] E. Lopienska. *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*. General Photo. 2022. URL: https://cds.cern.ch/record/2800984.

[25] *Drawing from CMSIM with labels*. 2001. URL: https://cmsdoc.cern.ch/cms/outreach/html/CMSdocuments/DetectorDrawings/DetectorDrawings.html.

[26] V. Karimäki et al. *The CMS tracker system project: Technical Design Report*. Tech. rep. 1997. URL: https://cds.cern.ch/record/368412.

[27] CMS Collaboration. *The CMS tracker: addendum to the Technical Design Report*. Tech. rep. 2000. URL: https://cds.cern.ch/record/490194.

[28] A. Dominguez et al. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. CERN-LHCC-2012-016. CMS-TDR-11. 2012. URL: https://cds.cern.ch/record/1481838.

[29] E. Focardi et al. "The CMS silicon tracker". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 453.1-2 (2000), pp. 121–125. DOI: 10.1016/s0168-9002(00)00616-1.

[30] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Tech. rep. 1997. URL: https://cds.cern.ch/record/349375.

[31] P. Bloch et al. *Changes to CMS ECAL electronics: addendum to the Technical Design Report*. Tech. rep. 2002. URL: https://cds.cern.ch/record/581342.

[32] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech. rep. 1997. URL: https://cds.cern.ch/record/357153.

[33] J. Mans et al. *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*. Tech. rep. CERN-LHCC-2012-015. CMS-TDR-10. 2012. URL: https://cds.cern.ch/record/1481837.

[34] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Tech. rep. 1997. URL: https://cds.cern.ch/record/331056.

[35] D. Campi et al. "Commissioning of the CMS Magnet". In: *IEEE Transactions on Applied Superconductivity* 17.2 (2007), pp. 1185–1190. DOI: 10.1109/tasc.2007.897754. URL: https://cds.cern.ch/record/1102022.

[36] J. G. Layter. *The CMS muon project: Technical Design Report*. Tech. rep. 1997. URL: https://cds.cern.ch/record/343814.

[37] A. Colaleo et al. "The compact muon solenoid RPC barrel detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 602.3 (2009), pp. 674–678. DOI: 10.1016/j.nima.2008.12.234. URL: https://cds.cern.ch/record/1280440.

[38] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Tech. rep. 2000. URL: https://cds.cern.ch/record/706847.

[39] S. Cittolin et al. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Tech. rep. 2002. URL: https://cds.cern.ch/record/578006.

[40] A. Tapper and D. Acosta. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. Tech. rep. CERN-LHCC-2013-011. CMS-TDR-12. 2013. URL: https://cds.cern.ch/record/1556311.

[41] CMS Collaboration. "Commissioning of the CMS High-Level Trigger with cosmic rays". In: *Journal of Instrumentation* 5.03 (2010), T03005–T03005. DOI: 10.1088/1748-0221/5/03/t03005. arXiv: 0911.4889 [physics.ins-det].

[42] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003. arXiv: 1706.04965 [physics.ins-det].

[43] M. Cacciari et al. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (2008). DOI: 10.1088/1126-6708/2008/04/063. URL: http://stacks.iop.org/1126-6708/2008/i=04/a=063.

[44] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (2020), P12012–P12012. DOI: 10.1088/1748-0221/15/12/p12012.

[45] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (2014). DOI: 10.1007/jhep07(2014)079. arXiv: 1405.0301v2 [hep-ph].

[46] J. Bellm et al. "Herwig 7.0 / Herwig++ 3.0 Release Note". In: *The European Physical Journal C* 76.4 (2016). DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178v1 [hep-ph].

[47] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. DOI: 10.1016/s0168-9002(03)01368-8.

[48] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71 (2010). DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727v3 [physics.data-an].

[49] ATLAS Collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Physics Letters B* 784 (2018), pp. 173–191. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425 [hep-ex].

[50] CMS Collaboration. *Combined measurements of the Higgs boson's couplings at* $\sqrt{s} = 13$ *TeV*. Tech. rep. CMS-PAS-HIG-17-031. CERN, 2018. URL: https://cds.cern.ch/record/2308127.

[51] M. Pieri and R. Tanaka. *SM Higgs production cross sections at √s = 13 TeV*. 2017. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV.

[52] A. Denner et al. "Standard Model Higgs-Boson Branching Ratios with Uncertainties". In: *The European Physical Journal C* 71 (2011). DOI: 10.1140/epjc/s10052-011-1753-8. arXiv: 1107.5909v2 [hep-ph].

[53] F. Demartin et al. "Higgs production in association with a single top quark at the LHC". In: *The European Physical Journal C* 75.6 (2015). DOI: 10.1140/epjc/s10052-015-3475-9. arXiv: 1504.00611v2 [hep-ph].

[54] M. Aldaya et al. "Measurement of ttH and tH in the H→bb channel with the full Run 2 data sample". 2022. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/094.

[55] F. Fiedler et al. "The matrix element method and its application to measurements of the top quark mass". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 624.1 (2010), pp. 203–218. DOI: 10.1016/j.nima.2010.09.024. arXiv: 1003.1316 [hep-ex].

[56] M. Link. "Search for single top quark production in association with a Higgs boson on full Run 2 data with the CMS experiment". MA thesis. Institute of Experimental Particle Physics, 2020. URL: https://publish.etp.kit.edu/record/21952.

[57] N. Faltermann et al. "t(t)H jet assignment". 2020. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/193.

[58] R. L. Workman et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2022.8 (2022). DOI: 10.1093/ptep/ptac097.

[59] CMS Collaboration. "Measurement of the $t\bar{t}H$ and tH production rates in the $H \to b\bar{b}$ decay channel with $138\,fb^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,TeV$". 2023. URL: http://cds.cern.ch/record/2868175.

[60] cms-analysis. *HiggsAnalysis-CombinedLimit*. Version v8.2.0. 2021. URL: https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit.

[61] *Automatic statistical uncertainties - Combine*. 2021. URL: https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part2/bin-wise-stats/.

[62] R. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-w.

[63] R. D. Cousins. "Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms". 2013. URL: https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[64] N. Smirnov. "Table for Estimating the Goodness of Fit of Empirical Distributions". In: *The Annals of Mathematical Statistics* 19.2 (1948), pp. 279–281. DOI: 10.1214/aoms/1177730256.

[65] T. W. Anderson and D. A. Darling. "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes". In: *The Annals of Mathematical Statistics* 23.2 (1952), pp. 193–212. DOI: 10.1214/aoms/1177729437.

[66] *LHCHCGModels - KappaVKappaTKappaTTilde*. 2021. URL: https://github.com/cramonal/HiggsAnalysis-CombinedLimit/blob/ktkvkttilde_model/python/LHCHCGModels.py#L1177.

[67] ATLAS Collaboration. "Probing the CP nature of the top-Higgs Yukawa coupling in $t\bar{t}H$ and tH events with $H \to b\bar{b}$ decays using the ATLAS detector at the LHC". 2023. arXiv: 2303.05974v1 [hep-ex].

[68] M. Komm et al. "Progress of top width from WbWb scattering". 2021. URL: https://indico.cern.ch/event/1048004/#3-progress-of-top-width-from-w.

[69] ATLAS Collaboration. *A new tagger for the charge identification of b-jets*. Tech. rep. CERN, 2015. URL: https://cds.cern.ch/record/2048132.

[70] CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (2018), P05011–P05011. DOI: 10.1088/1748-0221/13/05/p05011. arXiv: 1712.07158v2.

[71] CMS Collaboration. "Expected Performance of Run-3 HLT b-quark jet identification". 2022. URL: https://cds.cern.ch/record/2825704.

[72] CMS Collaboration. "Performance of DeepJet b-tagging algorithms at High Level Trigger using Run 3 data from the CMS Experiment". 2023. URL: https://cds.cern.ch/record/2856240.

[73] *Arduino Nano*. URL: https://docs.arduino.cc/hardware/nano.

[74] *telegraf*. URL: https://github.com/influxdata/telegraf.

[75] *influxdb*. URL: https://github.com/influxdata/influxdb.

[76] *grafana*. URL: https://github.com/grafana/grafana.

[77] *kapacitor*. URL: https://github.com/influxdata/kapacitor.

A Additional material from the (t)tH analysis

A.1 Trigger tables

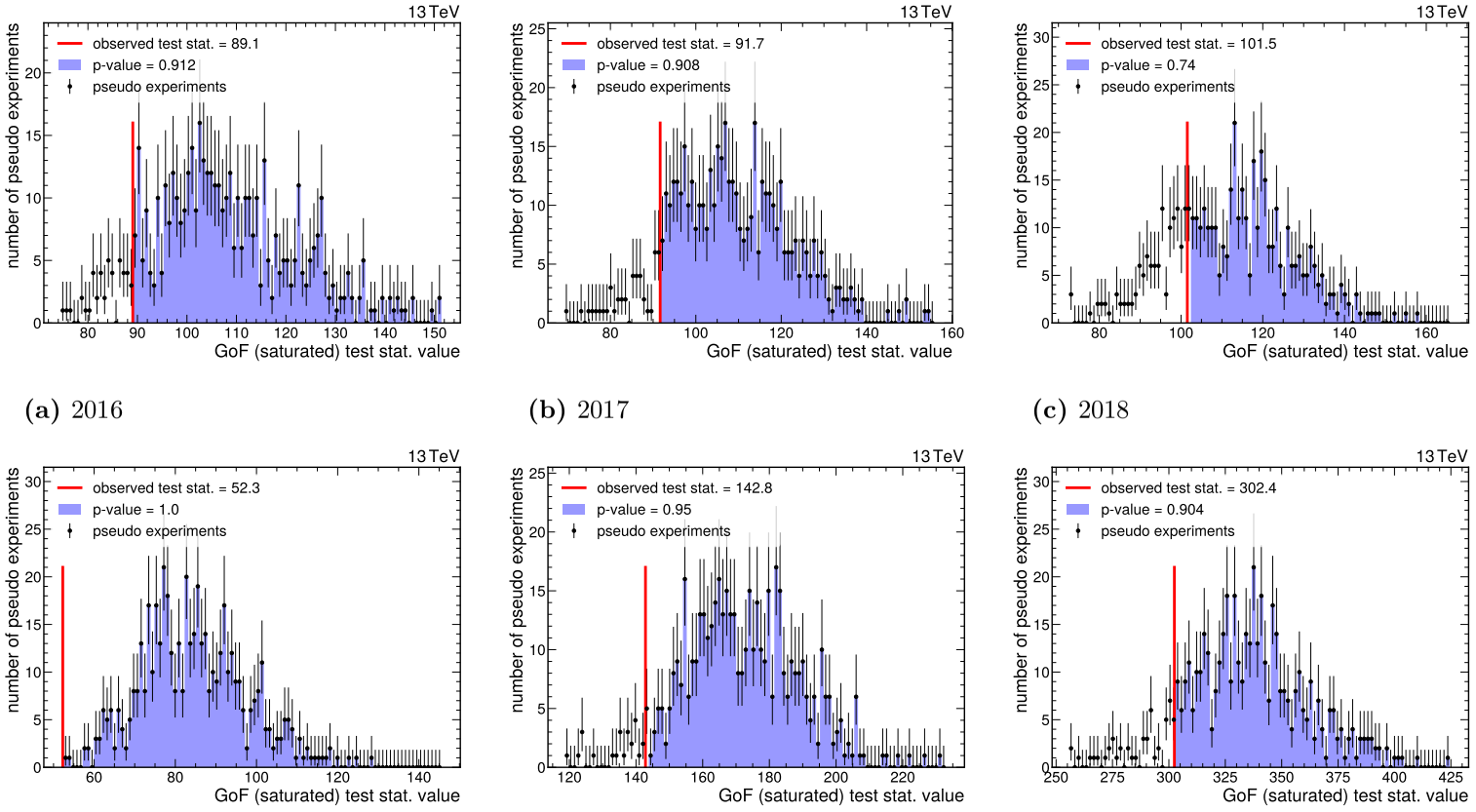
Table A.1: Triggers used in the FH channel of the (t)tH analysis

HLT path	Run era
PFJet450	2016 B–H
PFHT1050	2017 B–F
PFHT1050	2018 A–D
HT300PT30_QuadJet_75_60_45_40_TripleCSV_p07	2017 B
PFHT300PT30_QuadPFJet_75_60_45_40_TriplePFBTagCSV_3p0	2017 C–F
PFHT330PT30_QuadPFJet_75_60_45_40_TriplePFBTagDeepCSV_4p5	2018 A–D
PFHT400_SixJet30_DoubleBTagCSV_p056	2016 B–H
PFHT450_SixJet40_BTagCSV_p056	2016 B–H
PFHT380_SixJet32_DoubleBTagCSV_p075	2017 B
PFHT430_SixJet40_BTagCSV_p080	2017 B
PFHT380_SixPFJet32_DoublePFBTagCSV_2p2	2017 C–F
PFHT430_SixPFJet40_PFBTagCSV_1p5	2017 C–F
PFHT380_SixPFJet32_DoublePFBTagDeepCSV_2p2	2018 A–B
PFHT430_SixPFJet40_PFBTagCSV_1p5	2018 A
PFHT430_SixPFJet40_PFBTagDeepCSV_1p5	2018 B
PFHT400_SixPFJet32_DoublePFBTagDeepCSV_2p94	2018 C–D
PFHT450_SixPFJet36_PFBTagDeepCSV_1p59	2018 C–D

Table A.2: Triggers used in the DL channel of the (t)tH analysis

HLT path	Run era
IsoMu24	2016 B–H
IsoTkMu24	2016 B–H
IsoMu24_eta2p1	2017 B–D
IsoMu27	2017 B–F
IsoMu24	2018 A–D
Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	2016 B–H
Ele27_WPTight_Gsf	2016 B–H
Ele23_Ele12_CaloIdL_TrackIdL_IsoVL	2017 B–F
Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	2017 B–F
Ele32_WPTight_Gsf	2017 B–F
(Emulated via Ele32_WPTight_Gsf_L1DoubleEG)	
Ele23_Ele12_CaloIdL_TrackIdL_IsoVL	2018 A–D
Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	2018 A–D
Ele32_WPTight_Gsf	2018 A–D
Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	2016 B–H
Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	2016 B–H
Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	2016 B–H
Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	2016 B–H
Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	2017 B–F
Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	2017 B–F
Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	2017 B–F
Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	2017 B–F
Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	2018 A–D
Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	2018 A–D
Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	2018 A–D
Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	2018 A–D
Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	2016 B–G
Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ	2016 H
Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_	2016 B–G
Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ	2016 H
Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ	2017 B
Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8	2017 C–F
Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8	2018 A–D
Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8	2018 A–D

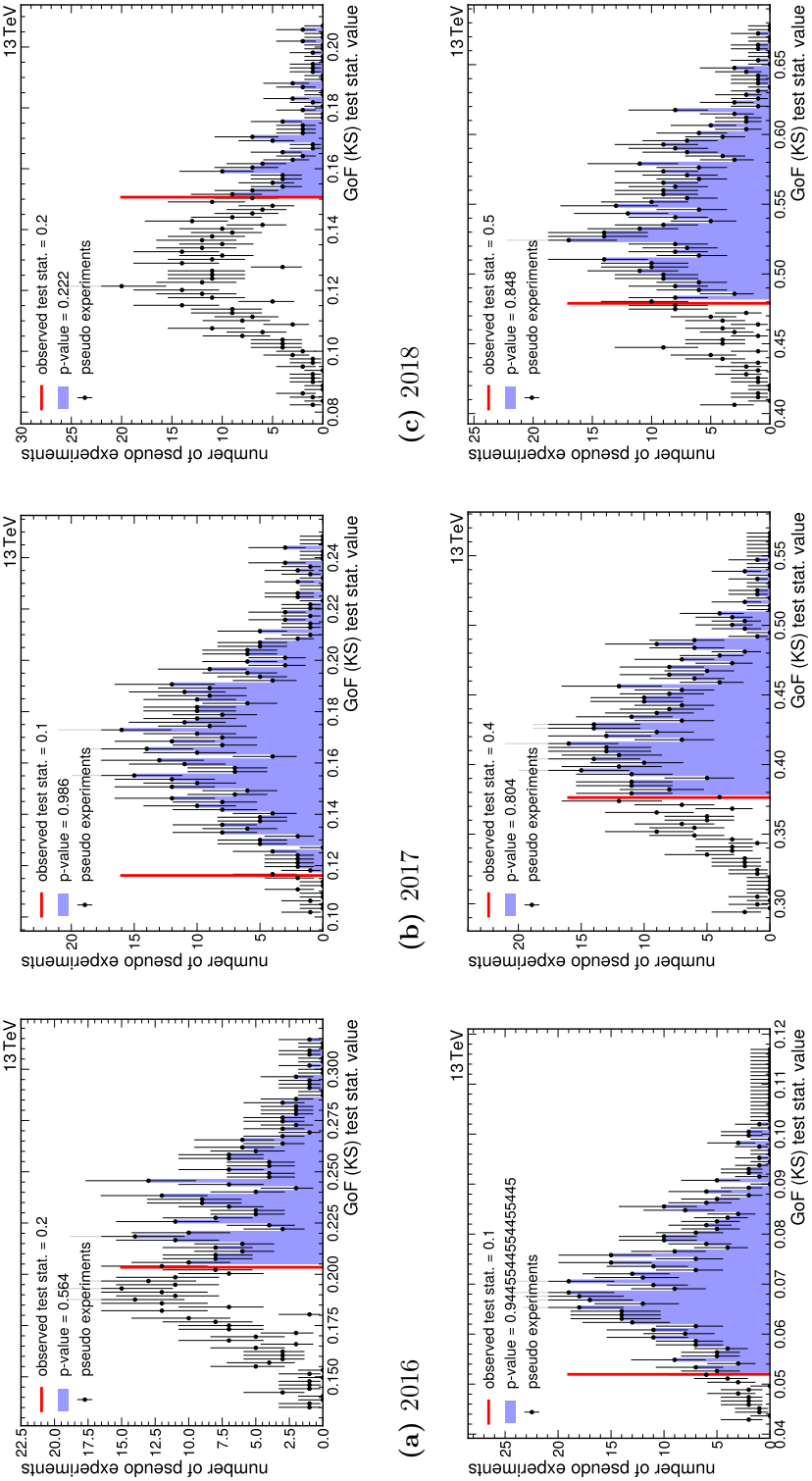
A.2 Additional goodness-of-fit results



(a) 2016 **(b)** 2017 **(c)** 2018

(d) DL channel **(e)** SL channel **(f)** full analysis

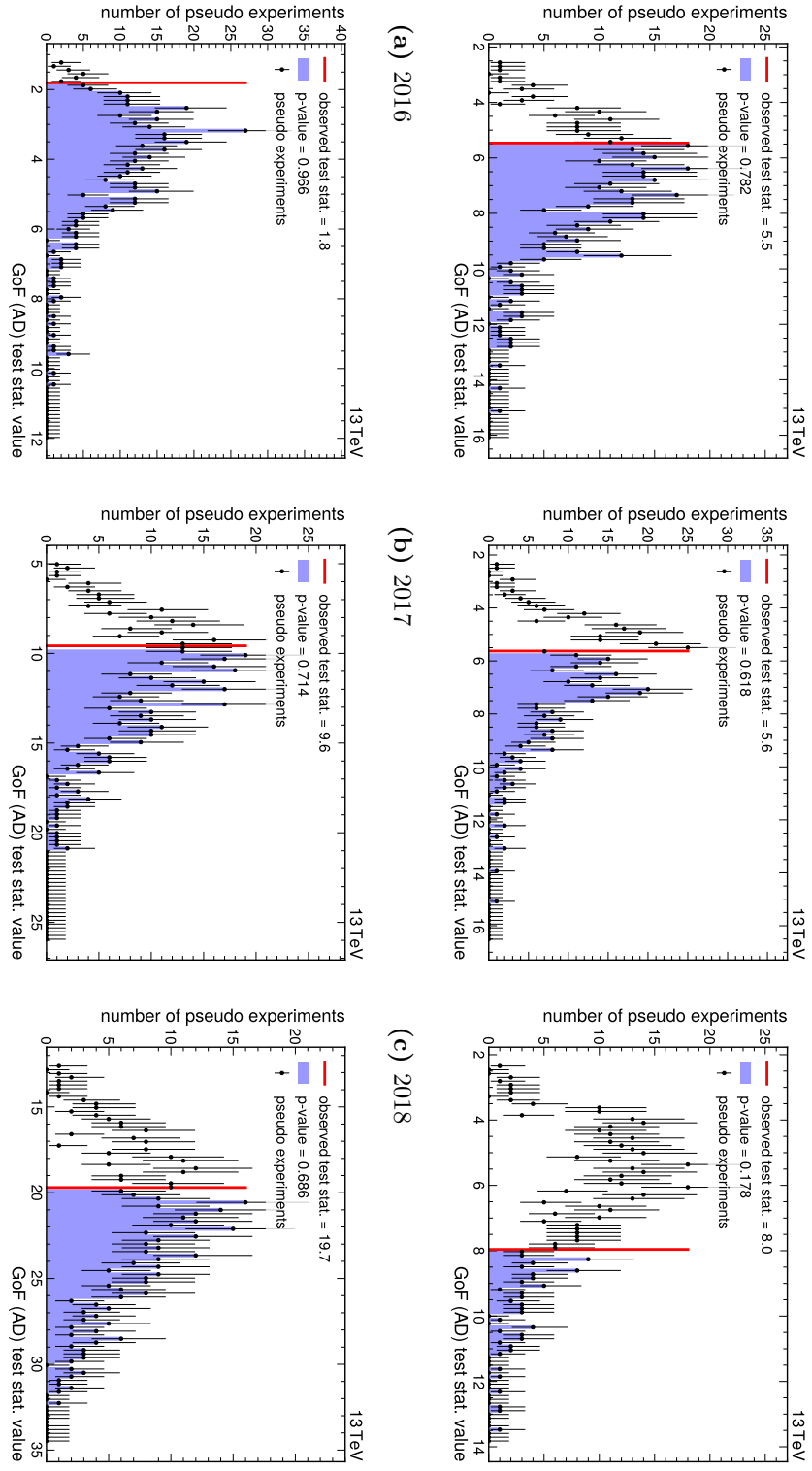
Figure A.1: Distributions of the saturated goodness-of-fit test statistic for 500 pseudo experiments fitting tH signal.



(a) 2016 (b) 2017 (c) 2018

(d) DL channel (e) SL channel (f) full analysis

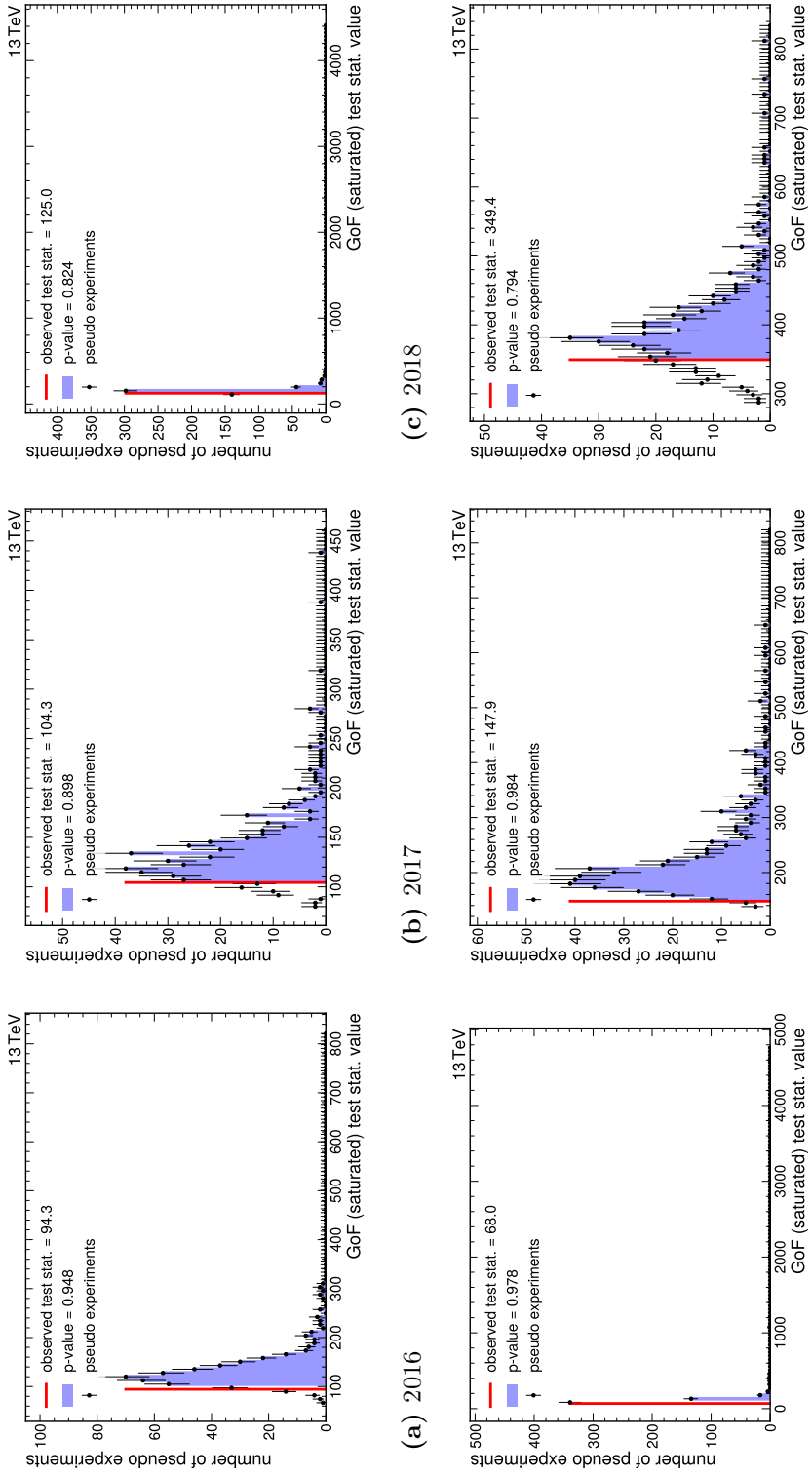
Figure A.2: Distributions of the Kolmogorov-Smirnov goodness-of-fit test statistic for 500 pseudo experiments fitting tH signal.



(a) 2016 (b) 2017 (c) 2018

(d) DL channel (e) SL channel (f) full analysis

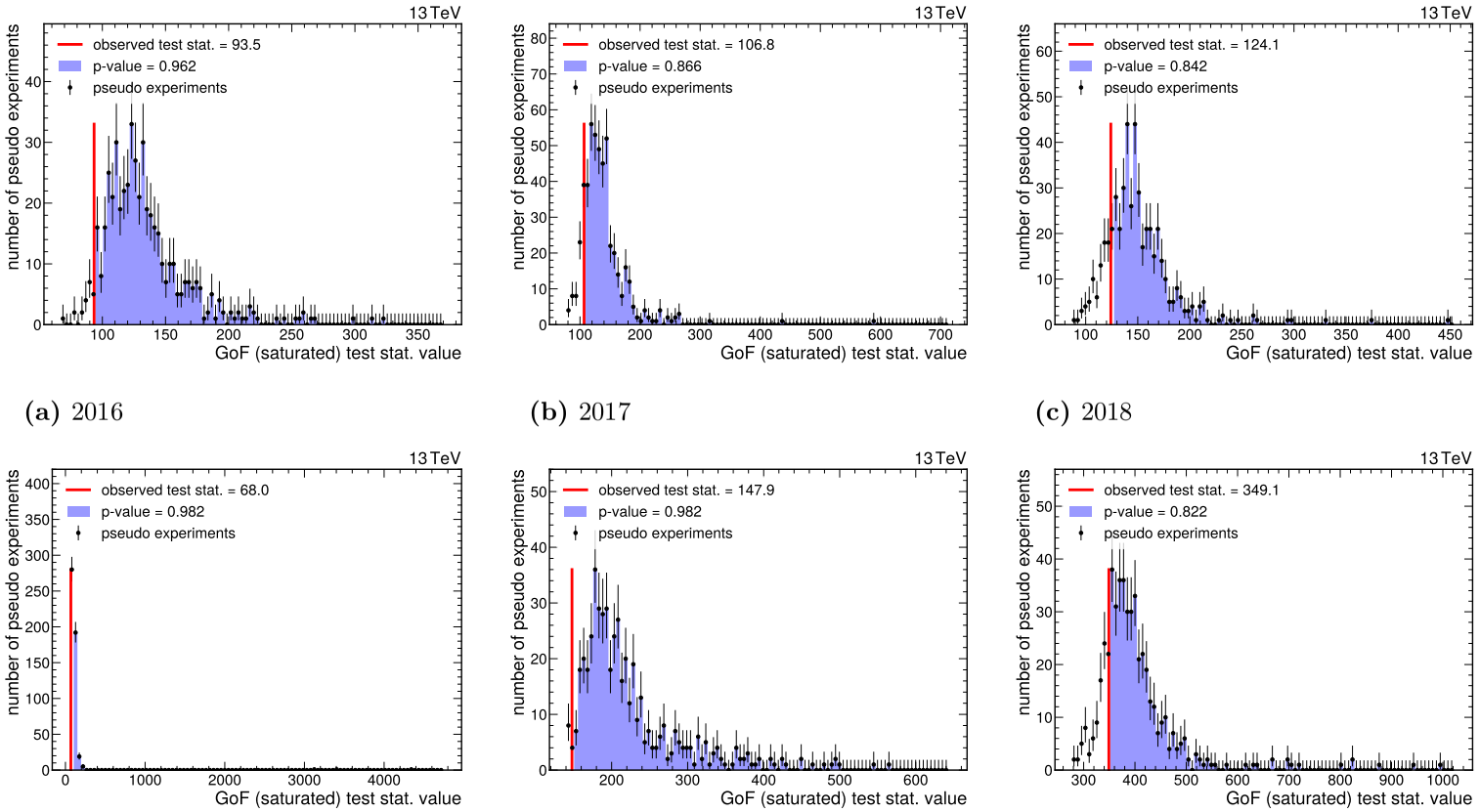
Figure A.3: Distributions of the Anderson-Darling goodness-of-fit test statistic for 500 pseudo experiments fitting tH signal.



(a) 2016 **(b)** 2017 **(c)** 2018

(d) DL channel **(e)** SL channel **(f)** full analysis

Figure A.4: Distribution of the saturated goodness-of-fit test statistic for 500 pseudo experiments fitting in the SM coupling scenario.



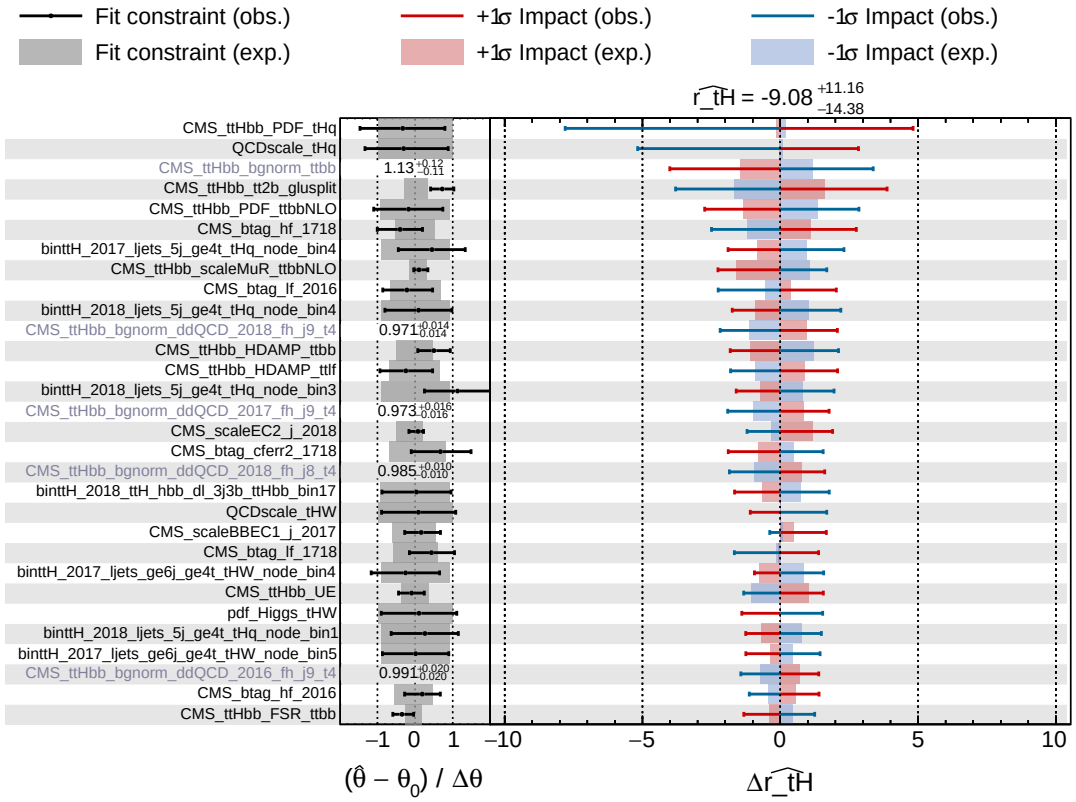
(a) 2016 **(b)** 2017 **(c)** 2018

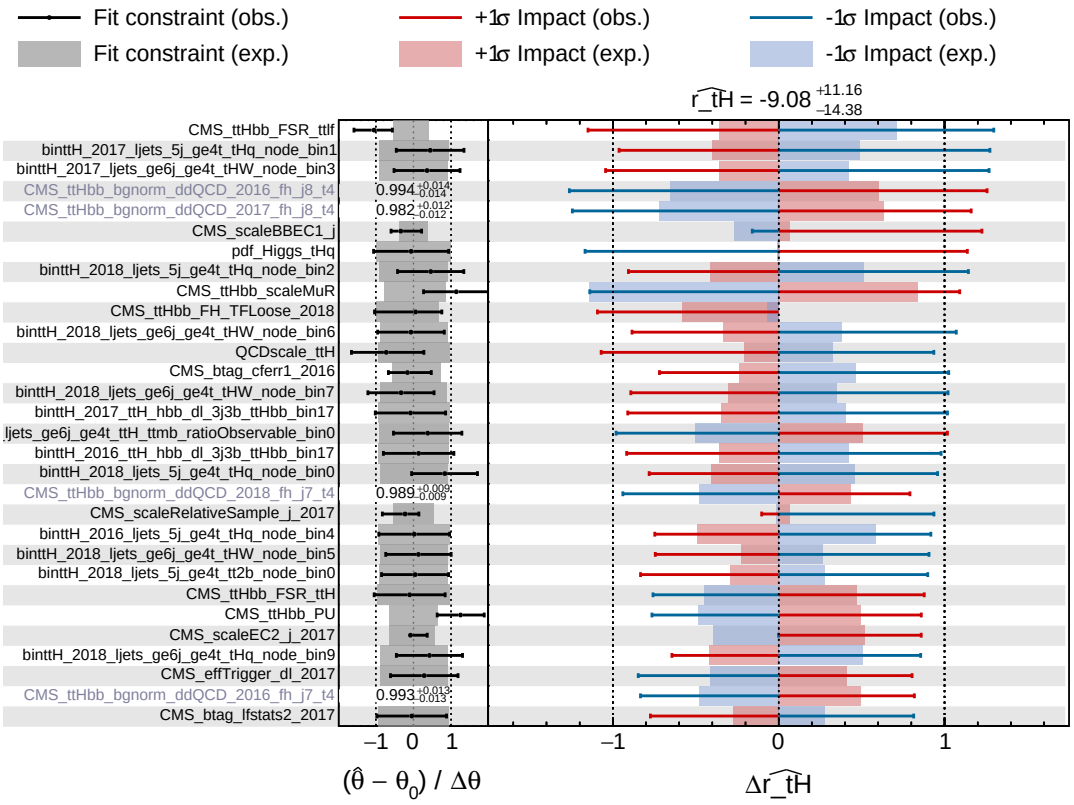
(d) DL channel **(e)** SL channel **(f)** full analysis

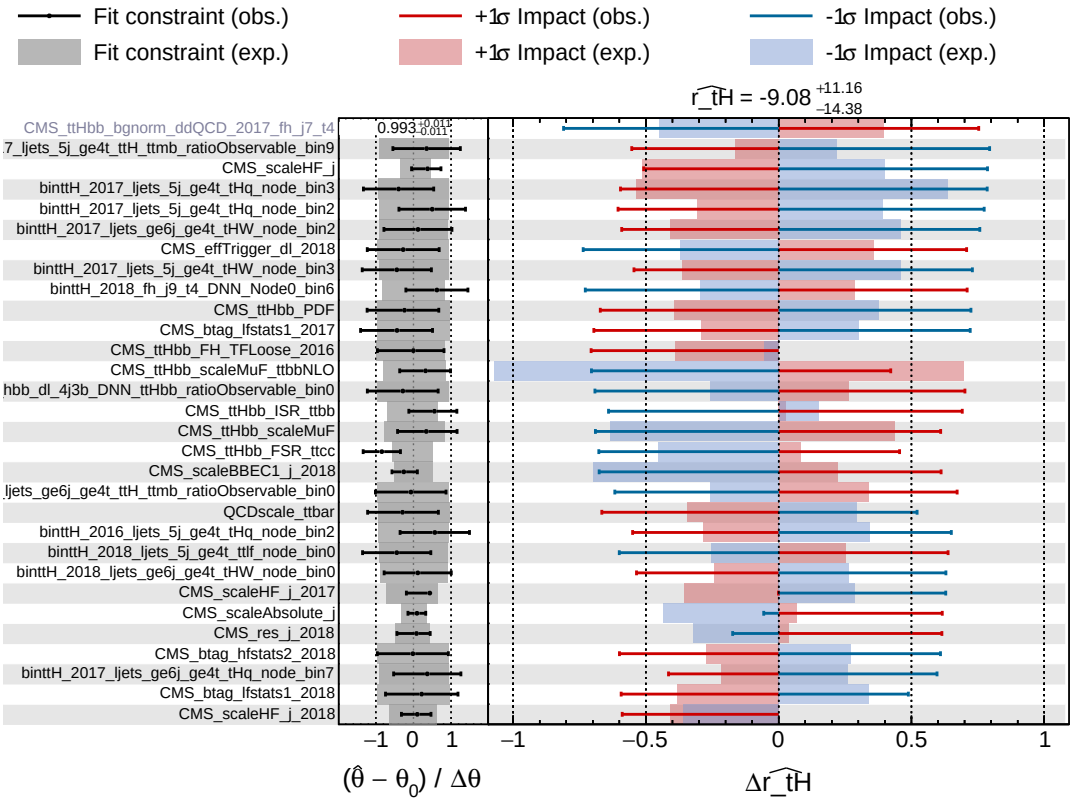
Figure A.5: Distribution of the saturated goodness-of-fit test statistic for 500 pseudo experiments fitting in the $\mathcal{CP}$ coupling scenario.

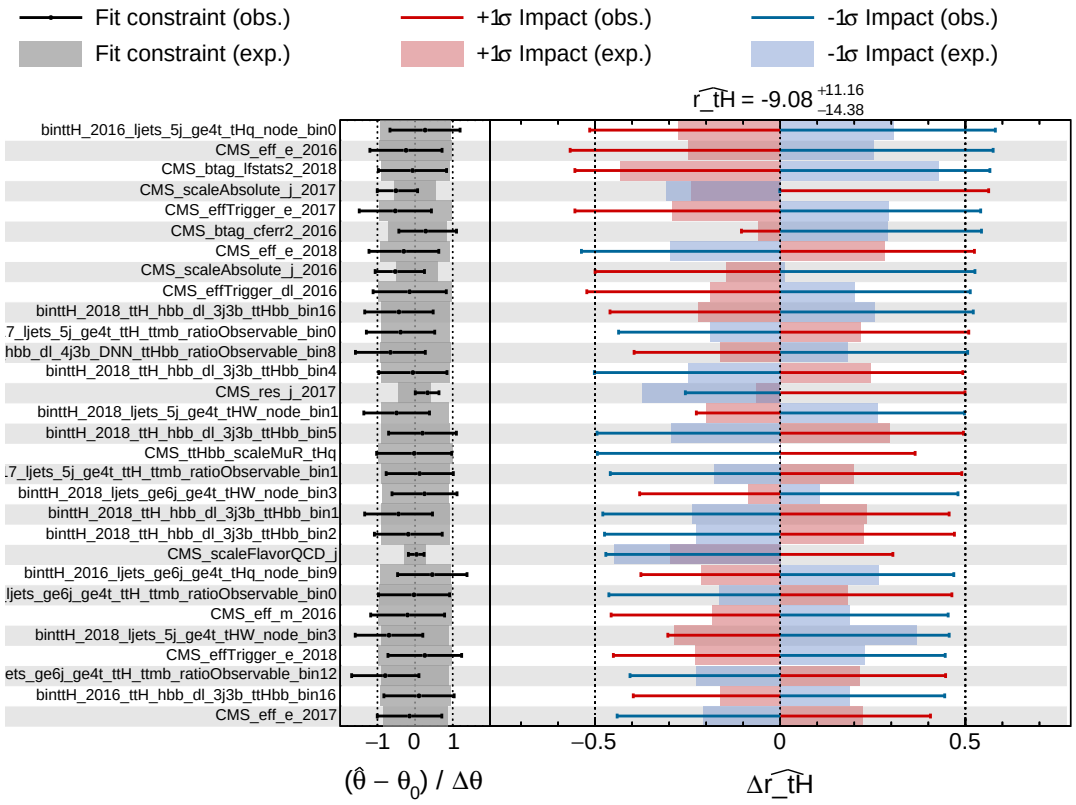
A.3 Complete importance of nuisances for the tH measurement

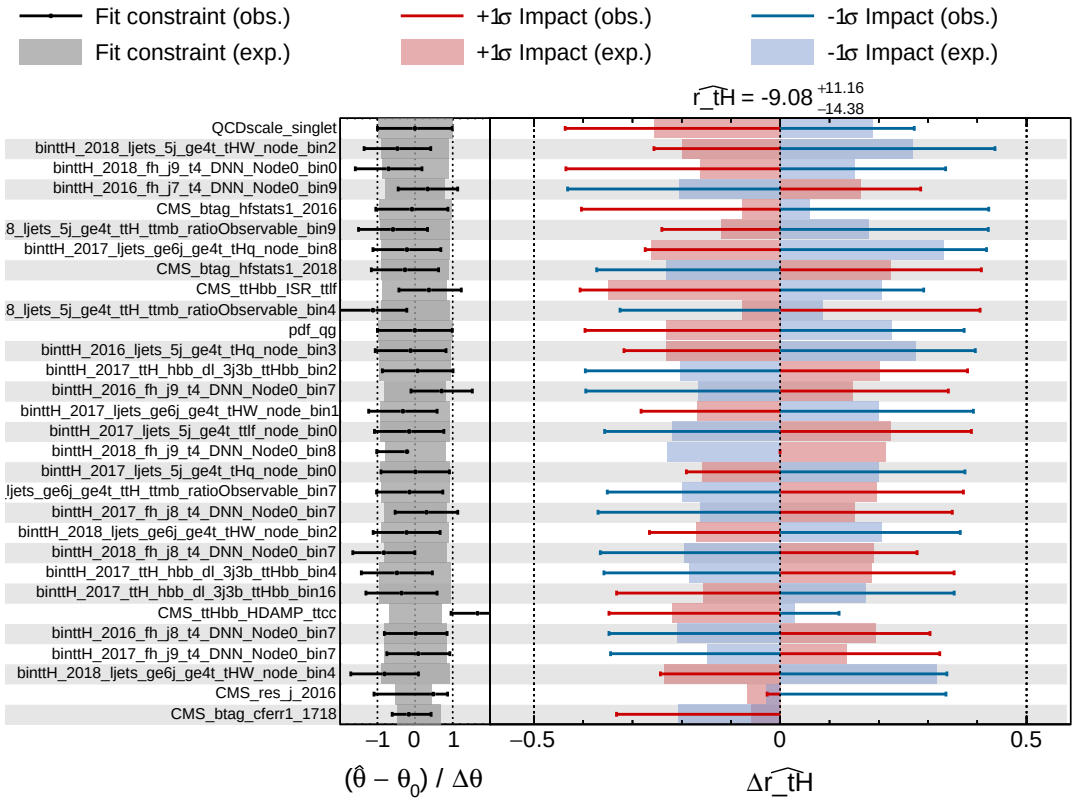
Constraints and importance of all nuisance parameters for the tH measurement, described in section 3.5.

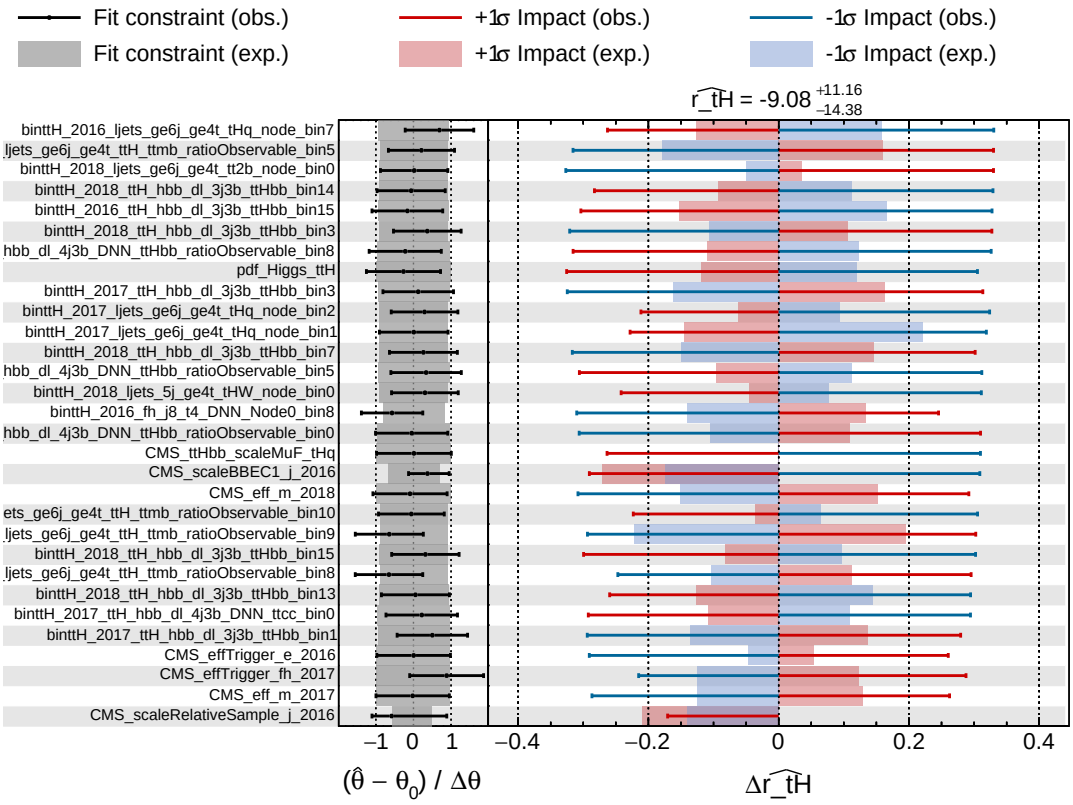


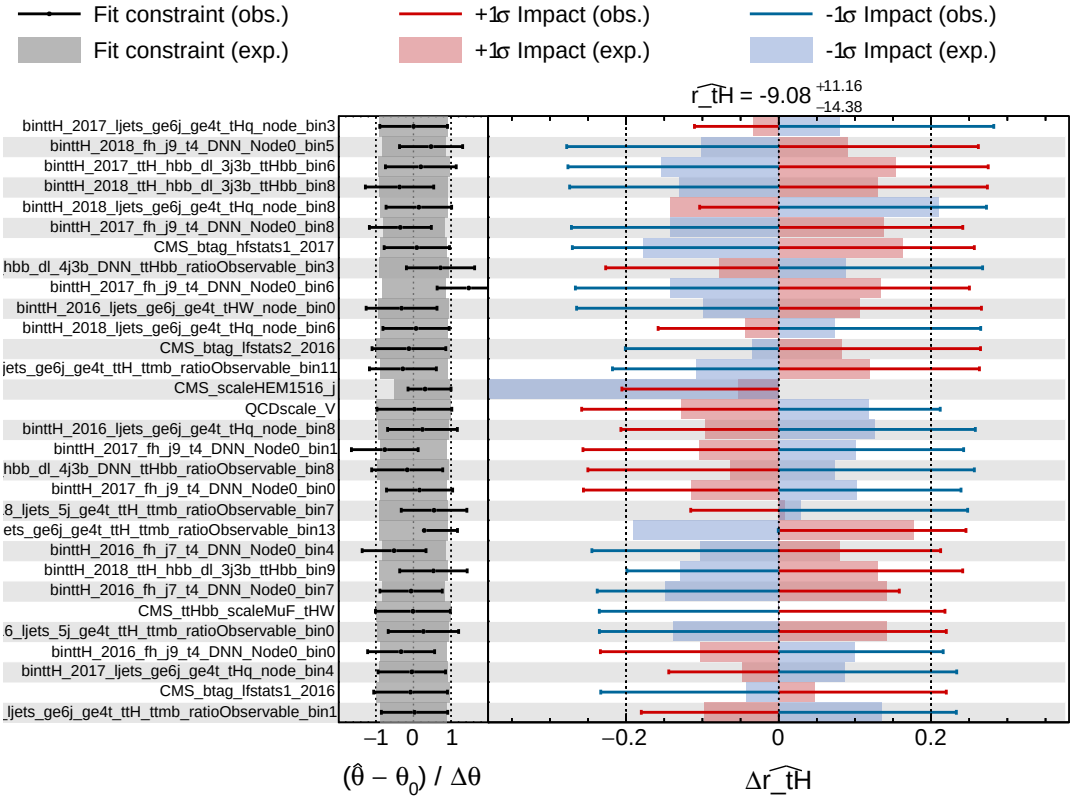

Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
CMS_ttHbb_FSR_ttlf
binttH_2017_ljets_5j_ge4t_tHq_node_bin1
binttH_2017_ljets_ge6j_ge4t_tHW_node_bin3
CMS_ttHbb_bgnorm_ddQCD_2016_fh_j8_t4
0.994 +0.014 −0.014
CMS_ttHbb_bgnorm_ddQCD_2017_fh_j8_t4
0.982 +0.012 −0.012
CMS_scaleBBEC1_j
pdf_Higgs_tHq
binttH_2018_ljets_5j_ge4t_tHq_node_bin2
CMS_ttHbb_scaleMuR
CMS_ttHbb_FH_TFLoose_2018
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin6
QCDscale_ttH
CMS_btag_cferr1_2016
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin7
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin17
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin0
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin17
binttH_2018_ljets_5j_ge4t_tHq_node_bin0
CMS_ttHbb_bgnorm_ddQCD_2018_fh_j7_t4
0.989 +0.009 −0.009
CMS_scaleRelativeSample_j_2017
binttH_2016_ljets_5j_ge4t_tHq_node_bin4
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin5
binttH_2018_ljets_5j_ge4t_tt2b_node_bin0
CMS_ttHbb_FSR_ttH
CMS_ttHbb_PU
CMS_scaleEC2_j_2017
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin9
CMS_effTrigger_dl_2017
CMS_ttHbb_bgnorm_ddQCD_2016_fh_j7_t4
0.993 +0.013 −0.013
CMS_btag_lfstats2_2017
−1 0 1
−1 0 1
(θ̂ − θ0) / Δθ
Δr_tH

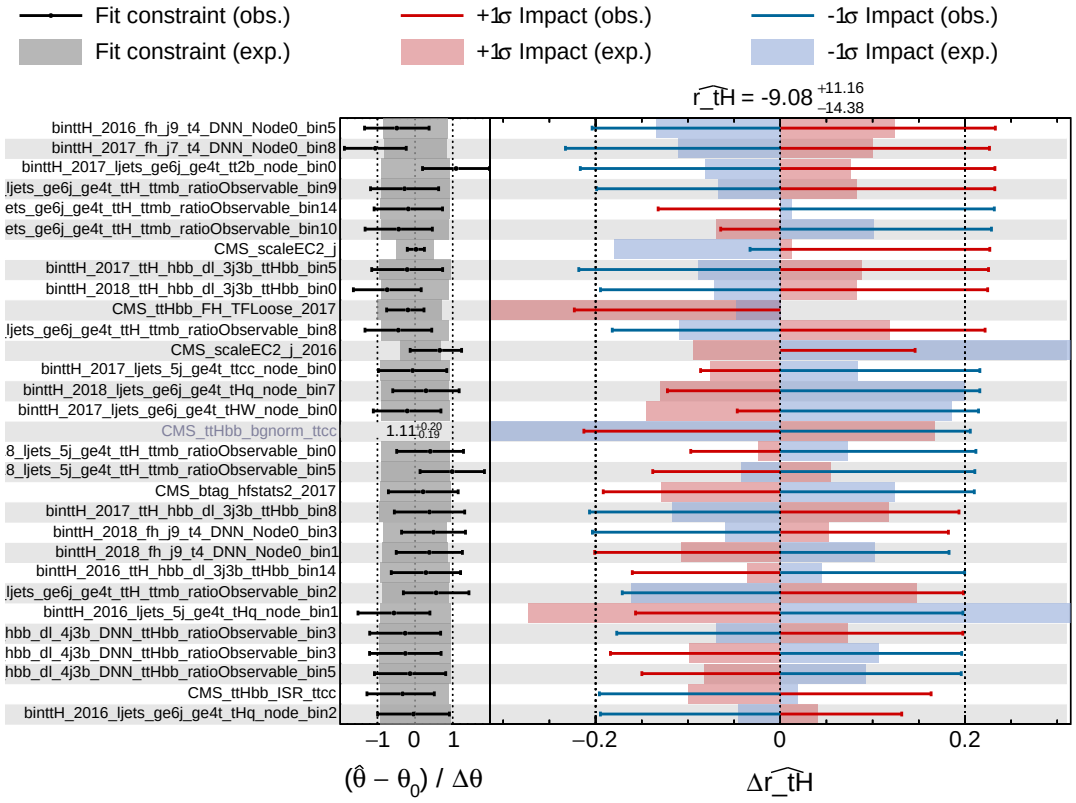

Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
CMS_ttHbb_bgnorm_ddQCD_2017_fh_j7_t4
0.993 +0.011 −0.011
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin9
CMS_scaleHF_j
binttH_2017_ljets_5j_ge4t_tHq_node_bin3
binttH_2017_ljets_5j_ge4t_tHq_node_bin2
binttH_2017_ljets_ge6j_ge4t_tHW_node_bin2
CMS_effTrigger_dl_2018
binttH_2017_ljets_5j_ge4t_tHW_node_bin3
binttH_2018_fh_j9_t4_DNN_Node0_bin6
CMS_ttHbb_PDF
CMS_btag_lfstats1_2017
CMS_ttHbb_FH_TFLoose_2016
CMS_ttHbb_scaleMuF_ttbbNLO
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin0
CMS_ttHbb_ISR_ttbb
CMS_ttHbb_scaleMuF
CMS_ttHbb_FSR_ttcc
CMS_scaleBBEC1_j_2018
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin0
QCDscale_ttbar
binttH_2016_ljets_5j_ge4t_tHq_node_bin2
binttH_2018_ljets_5j_ge4t_ttlf_node_bin0
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin0
CMS_scaleHF_j_2017
CMS_scaleAbsolute_j
CMS_res_j_2018
CMS_btag_hfstats2_2018
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin7
CMS_btag_lfstats1_2018
CMS_scaleHF_j_2018
−1 0 1
−1 −0.5 0 0.5 1
(θ̂ − θ0) / Δθ
Δr_tH

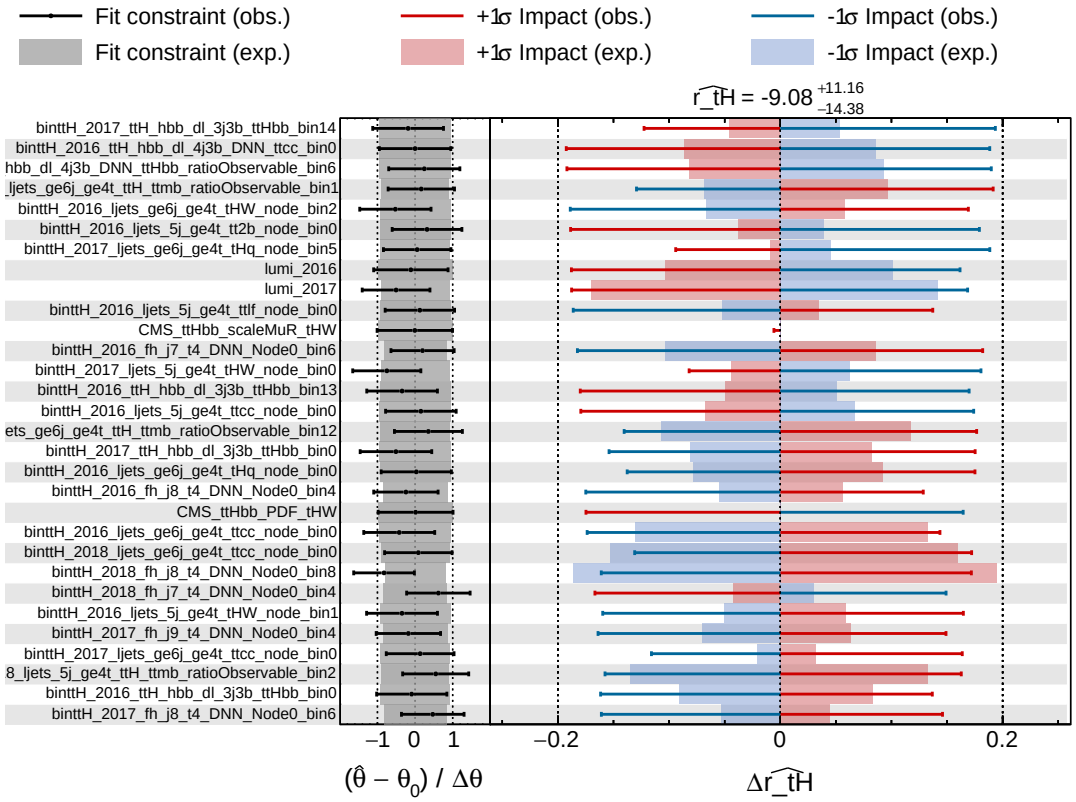

Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2016_ljets_5j_ge4t_tHq_node_bin0
CMS_eff_e_2016
CMS_btag_lfstats2_2018
CMS_scaleAbsolute_j_2017
CMS_effTrigger_e_2017
CMS_btag_cferr2_2016
CMS_eff_e_2018
CMS_scaleAbsolute_j_2016
CMS_effTrigger_dl_2016
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin16
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin0
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin8
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin4
CMS_res_j_2017
binttH_2018_ljets_5j_ge4t_tHW_node_bin1
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin5
CMS_ttHbb_scaleMuR_tHq
.7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin1
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin3
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin1
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin2
CMS_scaleFlavorQCD_j
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin9
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin0
CMS_eff_m_2016
binttH_2018_ljets_5j_ge4t_tHW_node_bin3
CMS_effTrigger_e_2018
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin12
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin16
CMS_eff_e_2017
−1 0 1
−0.5 0 0.5
(θ̂ − θ0) / Δθ
Δr_tH

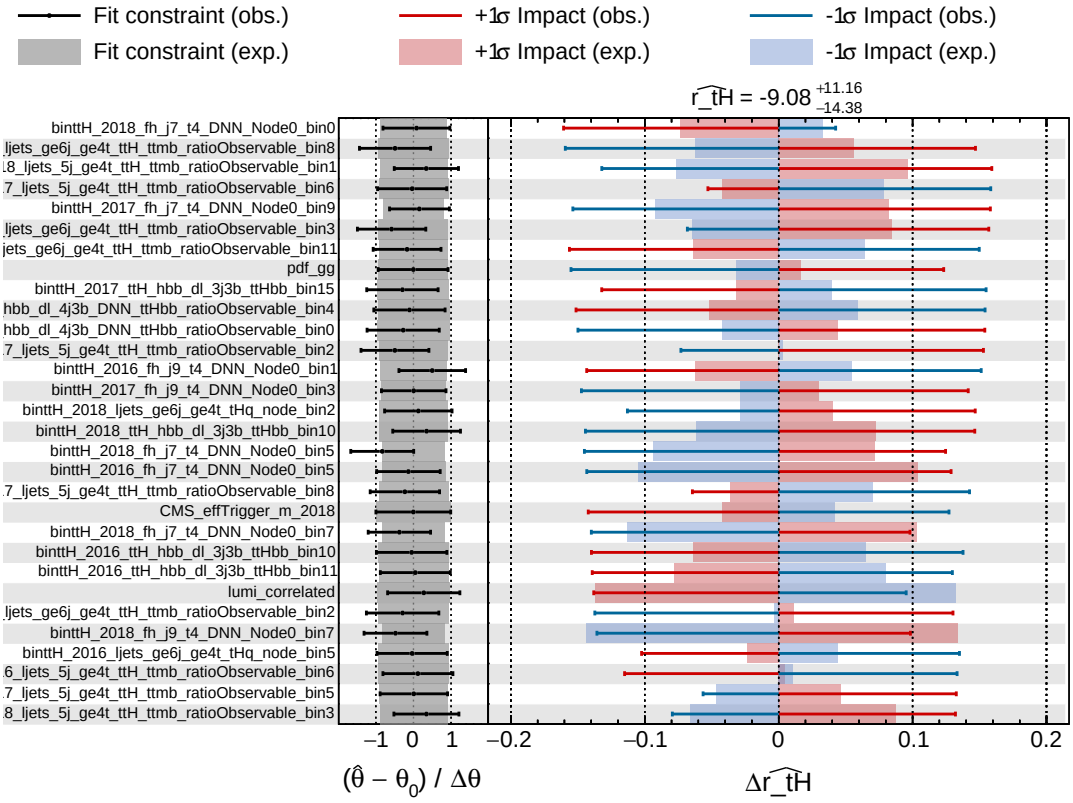

Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
QCDscale_singlet
binttH_2018_ljets_5j_ge4t_tHW_node_bin2
binttH_2018_fh_j9_t4_DNN_Node0_bin0
binttH_2016_fh_j7_t4_DNN_Node0_bin9
CMS_btag_hfstats1_2016
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin9
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin8
CMS_btag_hfstats1_2018
CMS_ttHbb_ISR_ttlf
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin4
pdf_qg
binttH_2016_ljets_5j_ge4t_tHq_node_bin3
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin2
binttH_2016_fh_j9_t4_DNN_Node0_bin7
binttH_2017_ljets_ge6j_ge4t_tHW_node_bin1
binttH_2017_ljets_5j_ge4t_ttlf_node_bin0
binttH_2018_fh_j9_t4_DNN_Node0_bin8
binttH_2017_ljets_5j_ge4t_tHq_node_bin0
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin7
binttH_2017_fh_j8_t4_DNN_Node0_bin7
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin2
binttH_2018_fh_j8_t4_DNN_Node0_bin7
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin4
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin16
CMS_ttHbb_HDAMP_ttcc
binttH_2016_fh_j8_t4_DNN_Node0_bin7
binttH_2017_fh_j9_t4_DNN_Node0_bin7
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin4
CMS_res_j_2016
CMS_btag_cferr1_1718
−1 0 1
−0.5 0 0.5
(θ̂ − θ0) / Δθ
Δr_tH

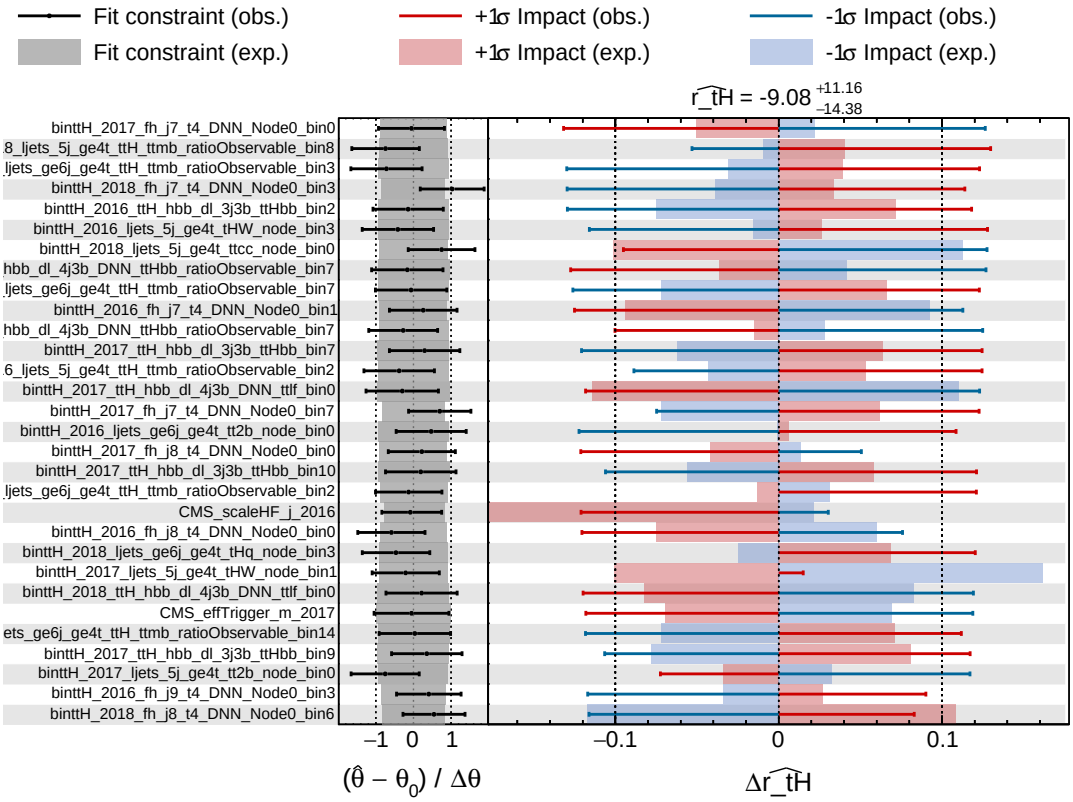

Fit constraint (obs.)
+1σ Impact (obs.)
-1σ Impact (obs.)
Fit constraint (exp.)
+1σ Impact (exp.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin7
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin5
binttH_2018_ljets_ge6j_ge4t_tt2b_node_bin0
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin14
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin15
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin3
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin8
pdf_Higgs_ttH
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin3
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin2
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin1
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin7
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin5
binttH_2018_ljets_5j_ge4t_tHW_node_bin0
binttH_2016_fh_j8_t4_DNN_Node0_bin8
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin0
CMS_ttHbb_scaleMuF_tHq
CMS_scaleBBEC1_j_2016
CMS_eff_m_2018
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin10
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin9
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin15
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin8
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin13
binttH_2017_ttH_hbb_dl_4j3b_DNN_ttcc_bin0
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin1
CMS_effTrigger_e_2016
CMS_effTrigger_fh_2017
CMS_eff_m_2017
CMS_scaleRelativeSample_j_2016
−1 0 1
−0.4 −0.2 0 0.2 0.4
(θ̂ − θ0) / Δθ
Δr_tH

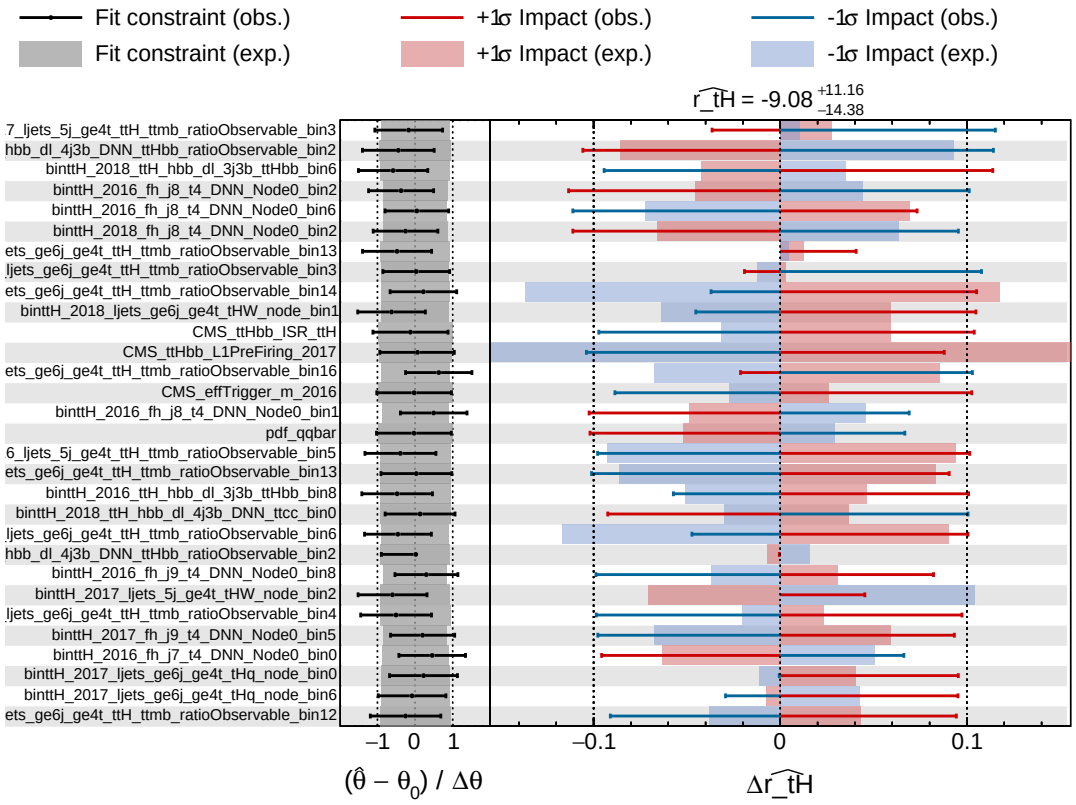

Fit constraint (obs.)
+1σ Impact (obs.)
-1σ Impact (obs.)
Fit constraint (exp.)
+1σ Impact (exp.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin3
binttH_2018_fh_j9_t4_DNN_Node0_bin5
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin6
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin8
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin8
binttH_2017_fh_j9_t4_DNN_Node0_bin8
CMS_btag_hfstats1_2017
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin3
binttH_2017_fh_j9_t4_DNN_Node0_bin6
binttH_2016_ljets_ge6j_ge4t_tHW_node_bin0
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin6
CMS_btag_lfstats2_2016
jets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin11
CMS_scaleHEM1516_j
QCDscale_V
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin8
binttH_2017_fh_j9_t4_DNN_Node0_bin1
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin8
binttH_2017_fh_j9_t4_DNN_Node0_bin0
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin7
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin13
binttH_2016_fh_j7_t4_DNN_Node0_bin4
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin9
binttH_2016_fh_j7_t4_DNN_Node0_bin7
CMS_ttHbb_scaleMuF_tHW
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin0
binttH_2016_fh_j9_t4_DNN_Node0_bin0
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin4
CMS_btag_lfstats1_2016
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin1
−1 0 1
−0.2 0 0.2
(θ̂ − θ0) / Δθ
Δr_tH

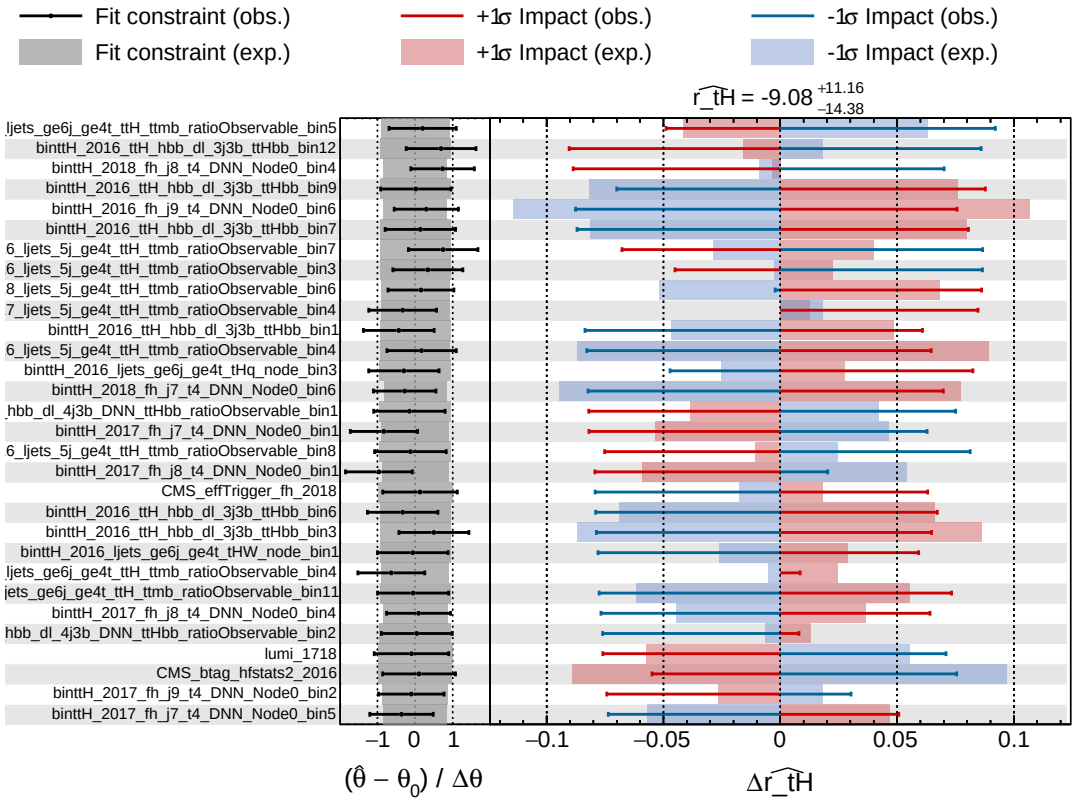

Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2016_fh_j9_t4_DNN_Node0_bin5
binttH_2017_fh_j7_t4_DNN_Node0_bin8
binttH_2017_ljets_ge6j_ge4t_tt2b_node_bin0
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin9
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin14
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin10
CMS_scaleEC2_j
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin5
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin0
CMS_ttHbb_FH_TFLoose_2017
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin8
CMS_scaleEC2_j_2016
binttH_2017_ljets_5j_ge4t_ttcc_node_bin0
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin7
binttH_2017_ljets_ge6j_ge4t_tHW_node_bin0
CMS_ttHbb_bgnorm_ttcc
1.11 +0.20 −0.19
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin0
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin5
CMS_btag_hfstats2_2017
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin8
binttH_2018_fh_j9_t4_DNN_Node0_bin3
binttH_2018_fh_j9_t4_DNN_Node0_bin1
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin14
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin2
binttH_2016_ljets_5j_ge4t_tHq_node_bin1
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin3
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin3
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin5
CMS_ttHbb_ISR_ttcc
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin2
−1 0 1
−0.2 0 0.2
(θ̂ − θ0) / Δθ
Δr_tH

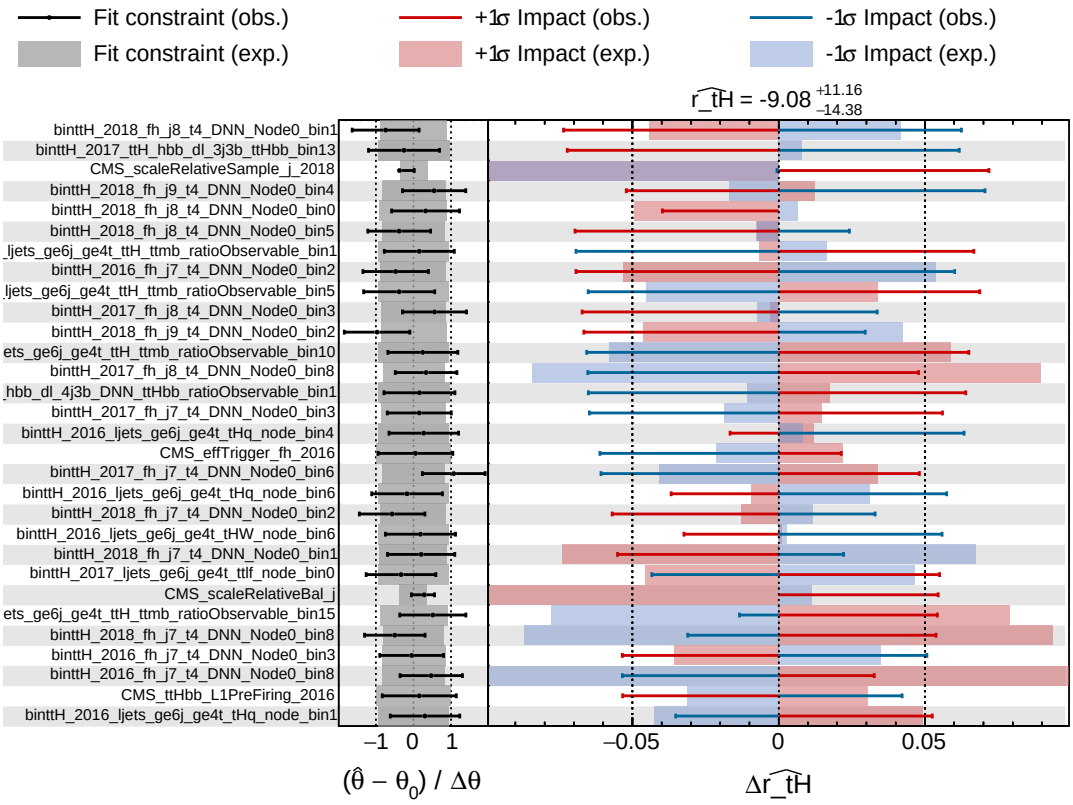

Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin14
binttH_2016_ttH_hbb_dl_4j3b_DNN_ttcc_bin0
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin6
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin1
binttH_2016_ljets_ge6j_ge4t_tHW_node_bin2
binttH_2016_ljets_5j_ge4t_tt2b_node_bin0
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin5
lumi_2016
lumi_2017
binttH_2016_ljets_5j_ge4t_ttlf_node_bin0
CMS_ttHbb_scaleMuR_tHW
binttH_2016_fh_j7_t4_DNN_Node0_bin6
binttH_2017_ljets_5j_ge4t_tHW_node_bin0
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin13
binttH_2016_ljets_5j_ge4t_ttcc_node_bin0
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin12
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin0
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin0
binttH_2016_fh_j8_t4_DNN_Node0_bin4
CMS_ttHbb_PDF_tHW
binttH_2016_ljets_ge6j_ge4t_ttcc_node_bin0
binttH_2018_ljets_ge6j_ge4t_ttcc_node_bin0
binttH_2018_fh_j8_t4_DNN_Node0_bin8
binttH_2018_fh_j7_t4_DNN_Node0_bin4
binttH_2016_ljets_5j_ge4t_tHW_node_bin1
binttH_2017_fh_j9_t4_DNN_Node0_bin4
binttH_2017_ljets_ge6j_ge4t_ttcc_node_bin0
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin2
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin0
binttH_2017_fh_j8_t4_DNN_Node0_bin6
−1 0 1
−0.2 0 0.2
(θ̂ − θ0) / Δθ
Δr_tH

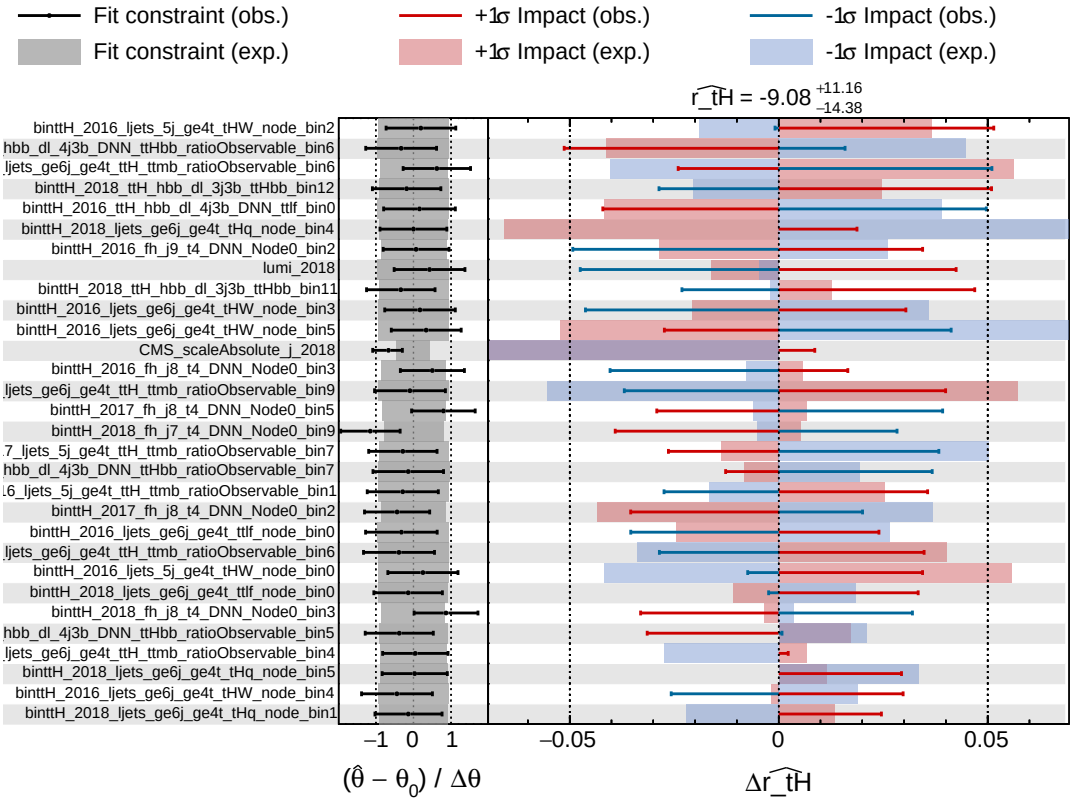

Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2018_fh_j7_t4_DNN_Node0_bin0
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin8
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin1
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin6
binttH_2017_fh_j7_t4_DNN_Node0_bin9
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin3
jets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin11
pdf_gg
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin15
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin4
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin0
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin2
binttH_2016_fh_j9_t4_DNN_Node0_bin1
binttH_2017_fh_j9_t4_DNN_Node0_bin3
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin2
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin10
binttH_2018_fh_j7_t4_DNN_Node0_bin5
binttH_2016_fh_j7_t4_DNN_Node0_bin5
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin8
CMS_effTrigger_m_2018
binttH_2018_fh_j7_t4_DNN_Node0_bin7
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin10
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin11
lumi_correlated
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin2
binttH_2018_fh_j9_t4_DNN_Node0_bin7
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin5
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin6
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin5
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin3
−1 0 1
−0.2 −0.1 0 0.1 0.2
(θ̂ − θ0) / Δθ
Δr_tH


Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2017_fh_j7_t4_DNN_Node0_bin0
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin8
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin3
binttH_2018_fh_j7_t4_DNN_Node0_bin3
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin2
binttH_2016_ljets_5j_ge4t_tHW_node_bin3
binttH_2018_ljets_5j_ge4t_ttcc_node_bin0
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin7
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin7
binttH_2016_fh_j7_t4_DNN_Node0_bin1
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin7
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin7
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin2
binttH_2017_ttH_hbb_dl_4j3b_DNN_ttlf_bin0
binttH_2017_fh_j7_t4_DNN_Node0_bin7
binttH_2016_ljets_ge6j_ge4t_tt2b_node_bin0
binttH_2017_fh_j8_t4_DNN_Node0_bin0
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin10
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin2
CMS_scaleHF_j_2016
binttH_2016_fh_j8_t4_DNN_Node0_bin0
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin3
binttH_2017_ljets_5j_ge4t_tHW_node_bin1
binttH_2018_ttH_hbb_dl_4j3b_DNN_ttlf_bin0
CMS_effTrigger_m_2017
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin14
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin9
binttH_2017_ljets_5j_ge4t_tt2b_node_bin0
binttH_2016_fh_j9_t4_DNN_Node0_bin3
binttH_2018_fh_j8_t4_DNN_Node0_bin6
−1 0 1
−0.1 0 0.1
(θ̂ − θ0) / Δθ
Δr_tH


Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin3
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin2
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin6
binttH_2016_fh_j8_t4_DNN_Node0_bin2
binttH_2016_fh_j8_t4_DNN_Node0_bin6
binttH_2018_fh_j8_t4_DNN_Node0_bin2
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin13
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin3
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin14
binttH_2018_ljets_ge6j_ge4t_tHW_node_bin1
CMS_ttHbb_ISR_ttH
CMS_ttHbb_L1PreFiring_2017
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin16
CMS_effTrigger_m_2016
binttH_2016_fh_j8_t4_DNN_Node0_bin1
pdf_qqbar
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin5
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin13
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin8
binttH_2018_ttH_hbb_dl_4j3b_DNN_ttcc_bin0
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin6
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin2
binttH_2016_fh_j9_t4_DNN_Node0_bin8
binttH_2017_ljets_5j_ge4t_tHW_node_bin2
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin4
binttH_2017_fh_j9_t4_DNN_Node0_bin5
binttH_2016_fh_j7_t4_DNN_Node0_bin0
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin0
binttH_2017_ljets_ge6j_ge4t_tHq_node_bin6
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin12
−1 0 1
−0.1 0 0.1
(θ̂ − θ0) / Δθ
Δr_tH


Fit constraint (obs.)
Fit constraint (exp.)
+1σ Impact (obs.)
+1σ Impact (exp.)
-1σ Impact (obs.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin5
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin12
binttH_2018_fh_j8_t4_DNN_Node0_bin4
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin9
binttH_2016_fh_j9_t4_DNN_Node0_bin6
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin7
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin7
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin3
8_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin6
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin4
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin1
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin4
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin3
binttH_2018_fh_j7_t4_DNN_Node0_bin6
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin1
binttH_2017_fh_j7_t4_DNN_Node0_bin1
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin8
binttH_2017_fh_j8_t4_DNN_Node0_bin1
CMS_effTrigger_fh_2018
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin6
binttH_2016_ttH_hbb_dl_3j3b_ttHbb_bin3
binttH_2016_ljets_ge6j_ge4t_tHW_node_bin1
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin4
jets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin11
binttH_2017_fh_j8_t4_DNN_Node0_bin4
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin2
lumi_1718
CMS_btag_hfstats2_2016
binttH_2017_fh_j9_t4_DNN_Node0_bin2
binttH_2017_fh_j7_t4_DNN_Node0_bin5
−1 0 1
−0.1 −0.05 0 0.05 0.1
(θ̂ − θ0) / Δθ
Δr_tH


Fit constraint (obs.)
+1σ Impact (obs.)
-1σ Impact (obs.)
Fit constraint (exp.)
+1σ Impact (exp.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2018_fh_j8_t4_DNN_Node0_bin1
binttH_2017_ttH_hbb_dl_3j3b_ttHbb_bin13
CMS_scaleRelativeSample_j_2018
binttH_2018_fh_j9_t4_DNN_Node0_bin4
binttH_2018_fh_j8_t4_DNN_Node0_bin0
binttH_2018_fh_j8_t4_DNN_Node0_bin5
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin1
binttH_2016_fh_j7_t4_DNN_Node0_bin2
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin5
binttH_2017_fh_j8_t4_DNN_Node0_bin3
binttH_2018_fh_j9_t4_DNN_Node0_bin2
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin10
binttH_2017_fh_j8_t4_DNN_Node0_bin8
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin1
binttH_2017_fh_j7_t4_DNN_Node0_bin3
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin4
CMS_effTrigger_fh_2016
binttH_2017_fh_j7_t4_DNN_Node0_bin6
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin6
binttH_2018_fh_j7_t4_DNN_Node0_bin2
binttH_2016_ljets_ge6j_ge4t_tHW_node_bin6
binttH_2018_fh_j7_t4_DNN_Node0_bin1
binttH_2017_ljets_ge6j_ge4t_ttlf_node_bin0
CMS_scaleRelativeBal_j
ets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin15
binttH_2018_fh_j7_t4_DNN_Node0_bin8
binttH_2016_fh_j7_t4_DNN_Node0_bin3
binttH_2016_fh_j7_t4_DNN_Node0_bin8
CMS_ttHbb_L1PreFiring_2016
binttH_2016_ljets_ge6j_ge4t_tHq_node_bin1
−1 0 1
−0.05 0 0.05
(θ̂ − θ0) / Δθ
Δr_tH


Fit constraint (obs.)
+1σ Impact (obs.)
-1σ Impact (obs.)
Fit constraint (exp.)
+1σ Impact (exp.)
-1σ Impact (exp.)
r_tH = -9.08 +11.16 −14.38
binttH_2016_ljets_5j_ge4t_tHW_node_bin2
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin6
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin6
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin12
binttH_2016_ttH_hbb_dl_4j3b_DNN_ttlf_bin0
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin4
binttH_2016_fh_j9_t4_DNN_Node0_bin2
lumi_2018
binttH_2018_ttH_hbb_dl_3j3b_ttHbb_bin11
binttH_2016_ljets_ge6j_ge4t_tHW_node_bin3
binttH_2016_ljets_ge6j_ge4t_tHW_node_bin5
CMS_scaleAbsolute_j_2018
binttH_2016_fh_j8_t4_DNN_Node0_bin3
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin9
binttH_2017_fh_j8_t4_DNN_Node0_bin5
binttH_2018_fh_j7_t4_DNN_Node0_bin9
7_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin7
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin7
6_ljets_5j_ge4t_ttH_ttmb_ratioObservable_bin1
binttH_2017_fh_j8_t4_DNN_Node0_bin2
binttH_2016_ljets_ge6j_ge4t_ttlf_node_bin0
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin6
binttH_2016_ljets_5j_ge4t_tHW_node_bin0
binttH_2018_ljets_ge6j_ge4t_ttlf_node_bin0
binttH_2018_fh_j8_t4_DNN_Node0_bin3
hbb_dl_4j3b_DNN_ttHbb_ratioObservable_bin5
ljets_ge6j_ge4t_ttH_ttmb_ratioObservable_bin4
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin5
binttH_2016_ljets_ge6j_ge4t_tHW_node_bin4
binttH_2018_ljets_ge6j_ge4t_tHq_node_bin1
−1 0 1
−0.05 0 0.05
(θ̂ − θ0) / Δθ
Δr_tH

A.4 Additional coupling results

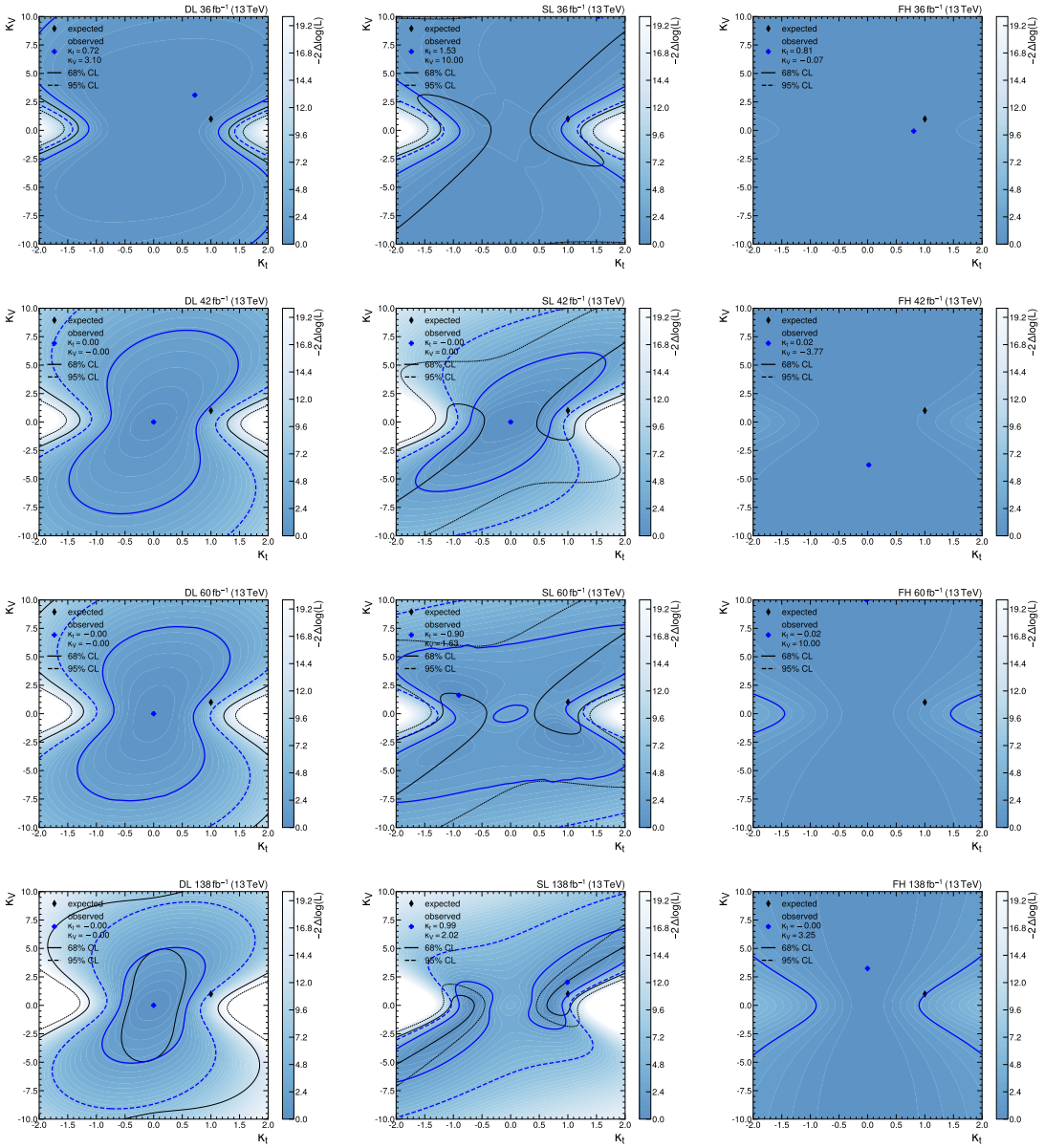


Figure A.6: Expected and observed 2D likelihood in the SM coupling scenario in different parts of the analysis

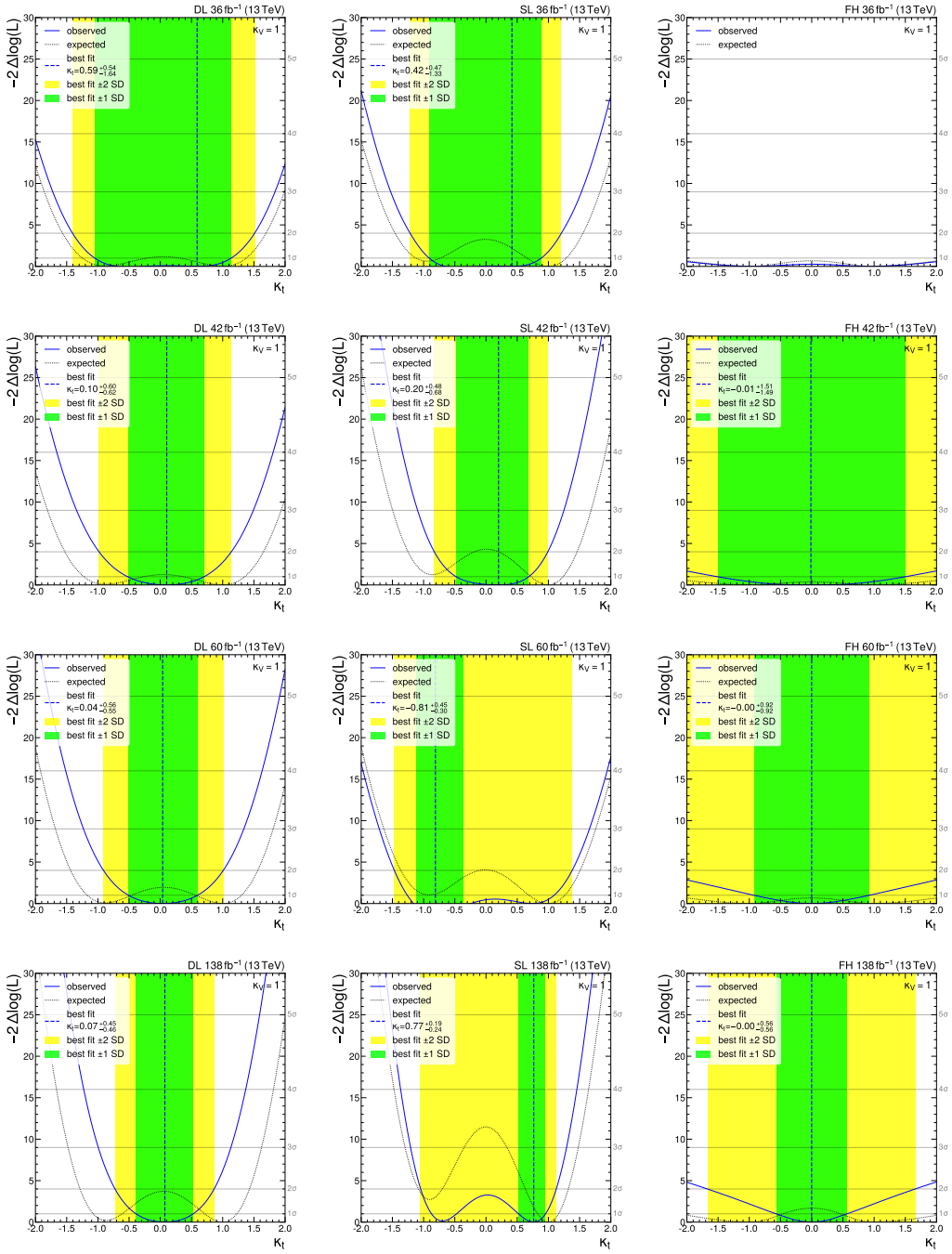


Figure A.7: Expected and observed results for κ_t in the SM coupling scenario in different parts of the analysis

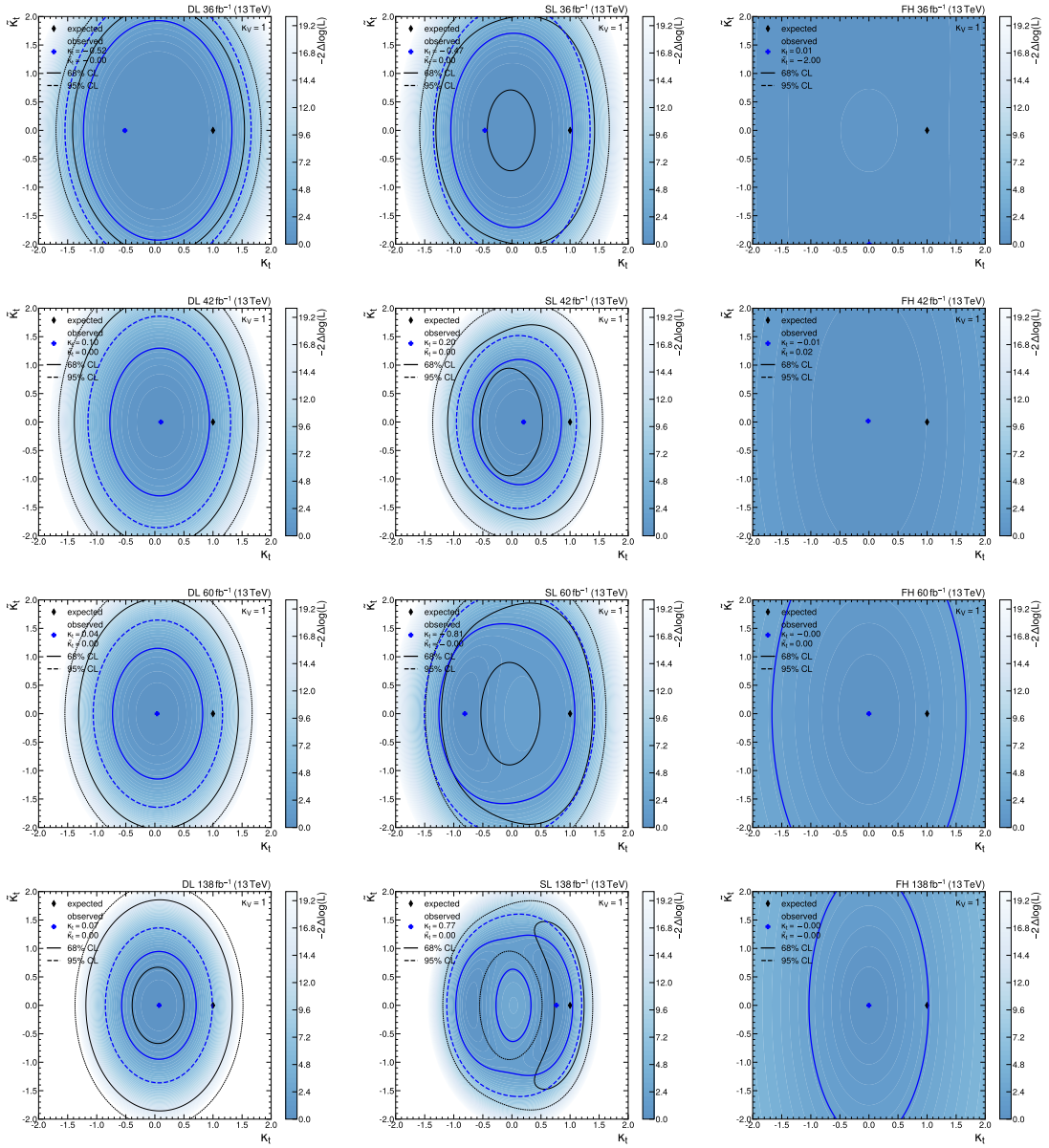


Figure A.8: Expected and observed 2D likelihood in the $\mathcal{CP}$ coupling scenario in different parts of the analysis

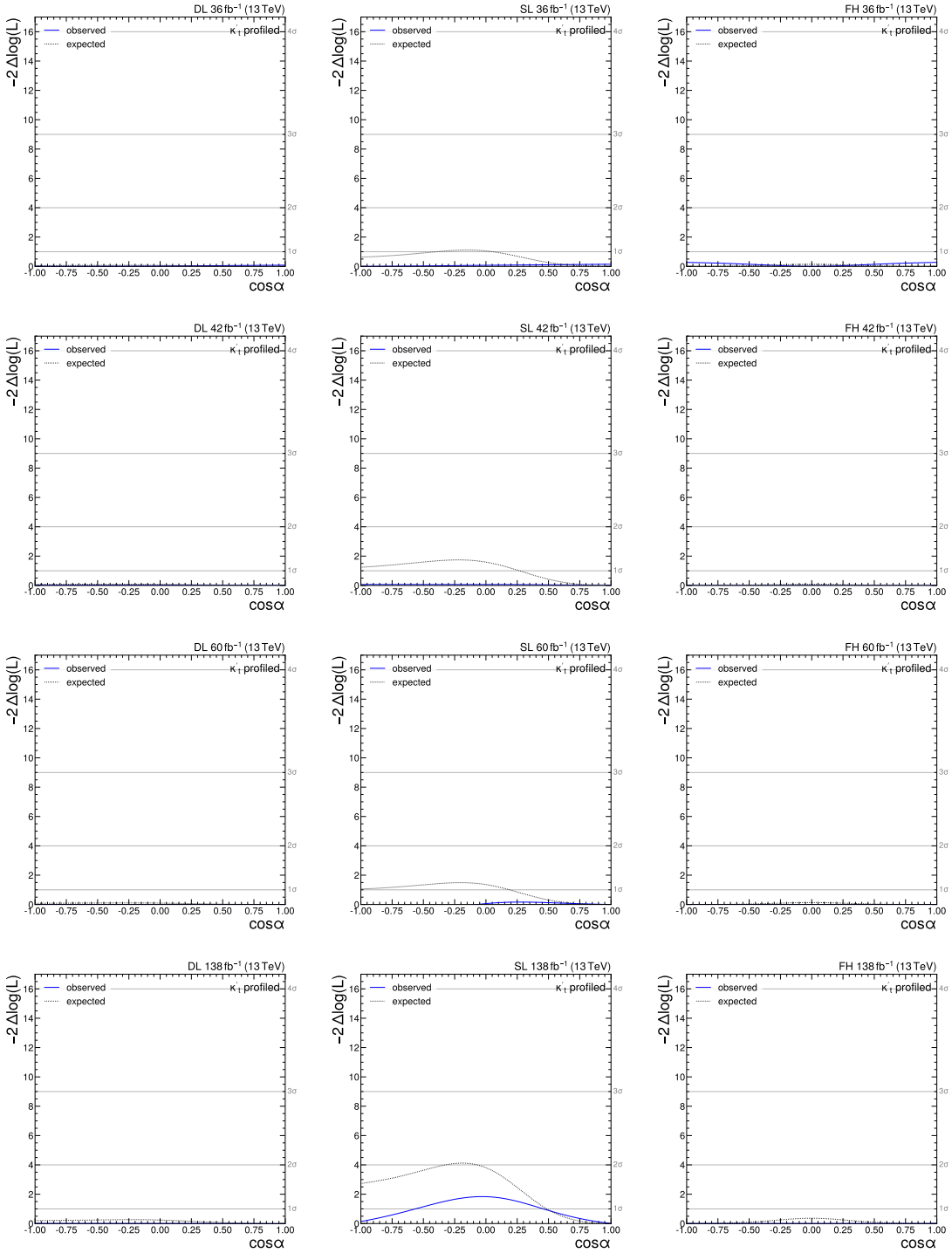
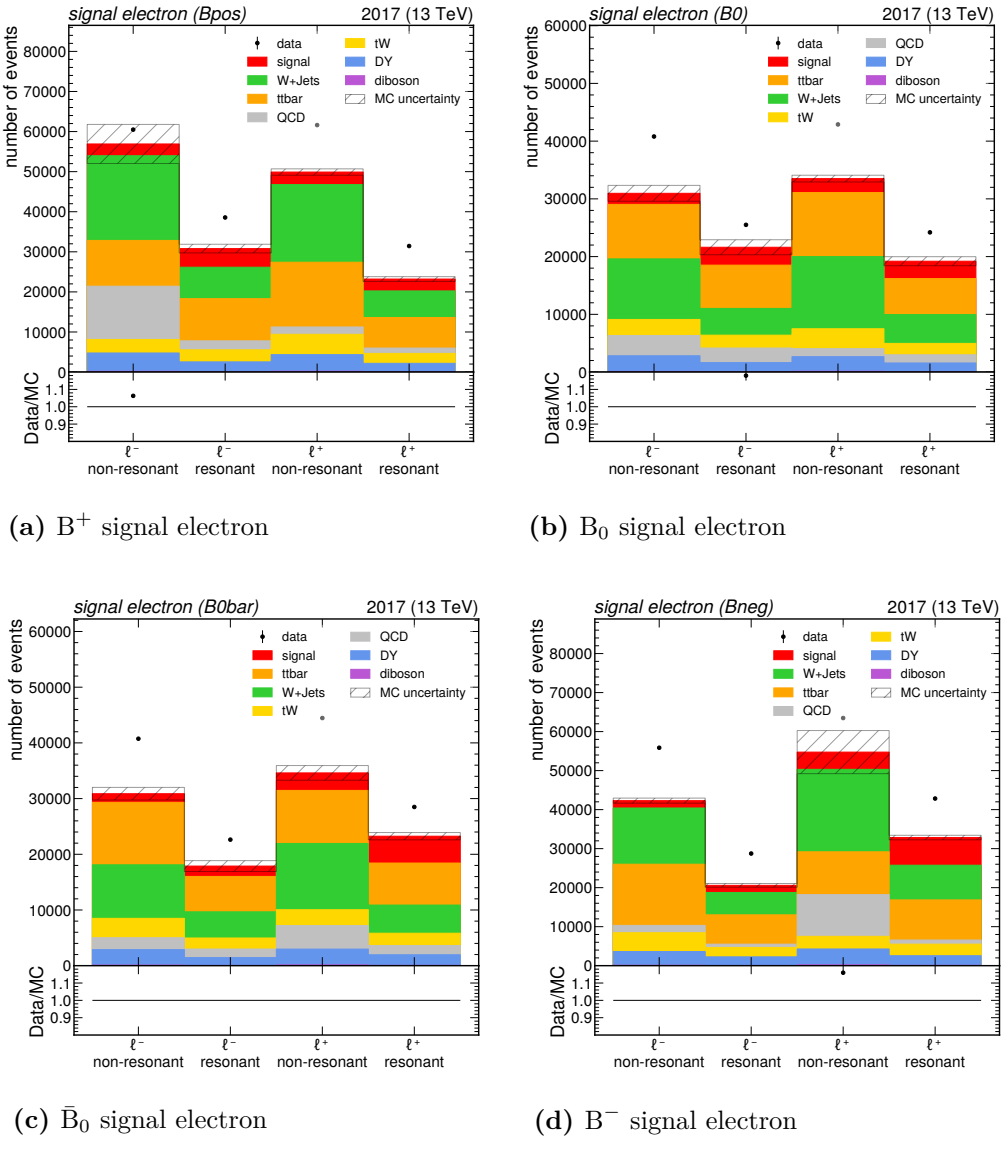


Figure A.9: $\cos\alpha$ best fit results in different parts of the analysis

B Additional material from the Wb analysis



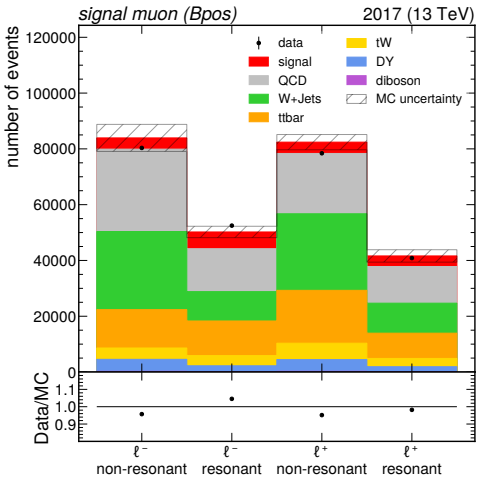
(a) B^+ signal electron

(b) B_0 signal electron

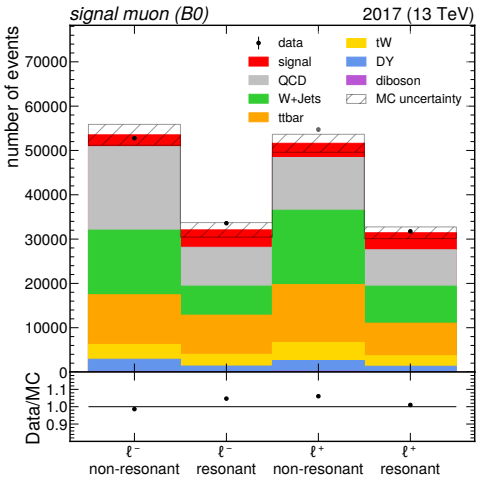
(c) $\bar{B}_0$ signal electron

(d) B^- signal electron

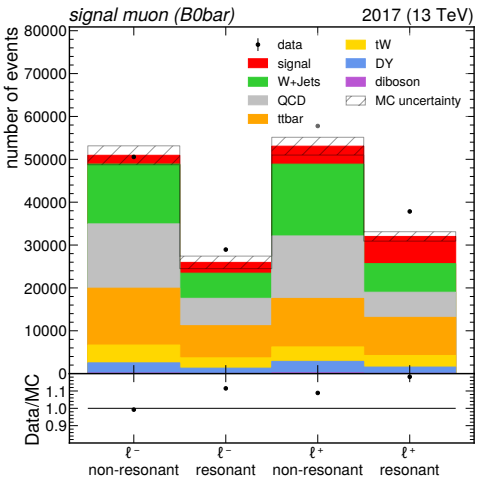
Figure B.1: Pre-fit distributions in the signal electron region of the Wb analysis



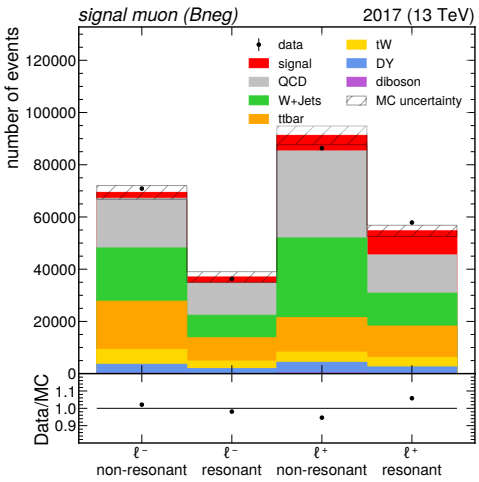
(a) B^+ signal muon



(b) B_0 signal muon

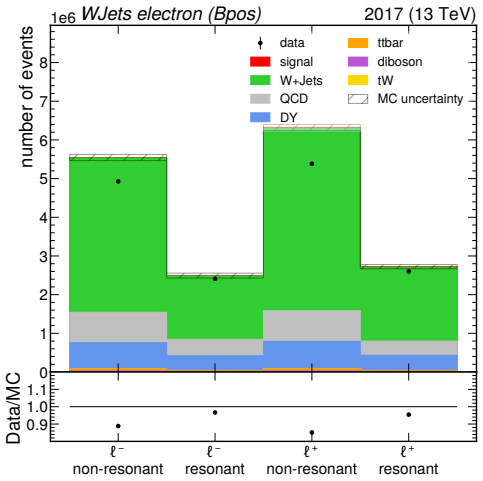


(c) $\bar{B}_0$ signal muon

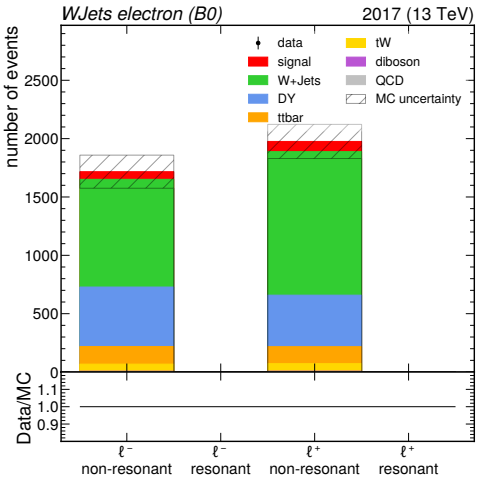


(d) B^- signal muon

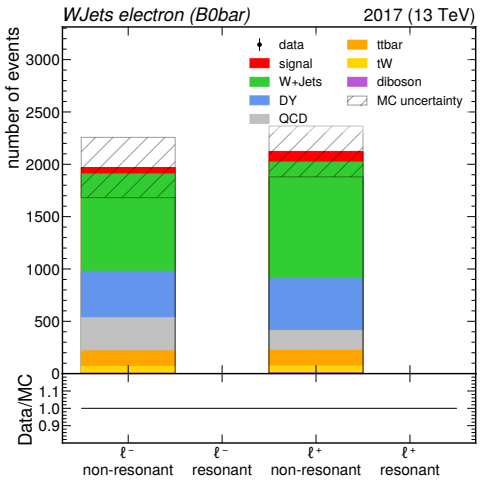
Figure B.2: Pre-fit distributions in the signal muon region of the Wb analysis



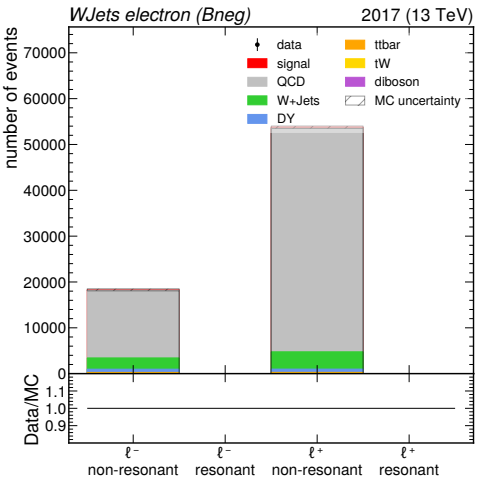
(a) B^+ W + jets electron



(b) B_0 W + jets electron

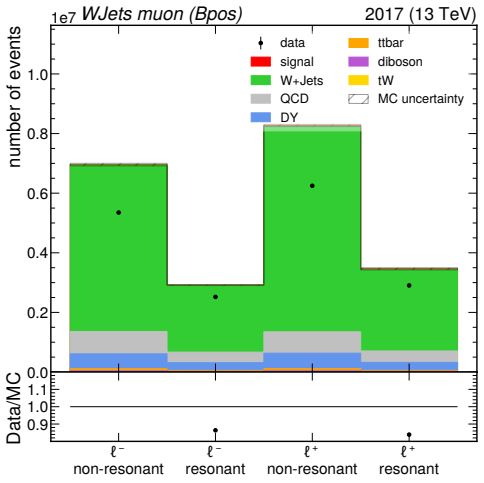


(c) $\bar{B}_0$ W + jets electron

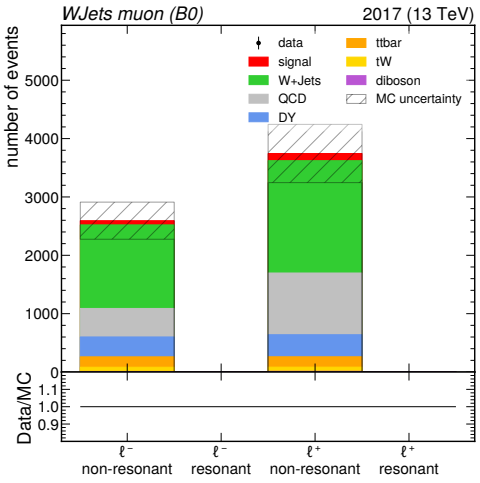


(d) B^- W + jets electron

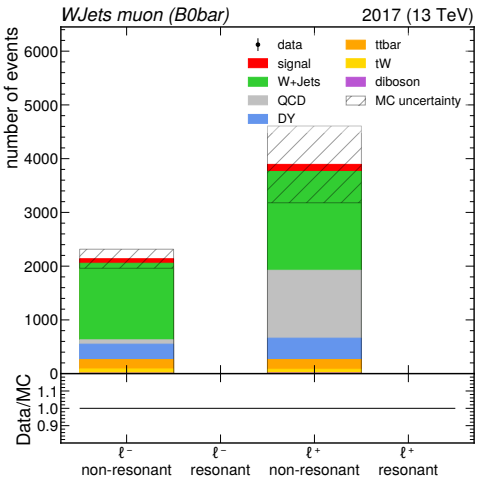
Figure B.3: Pre-fit distributions in the W + jets electron region of the Wb analysis



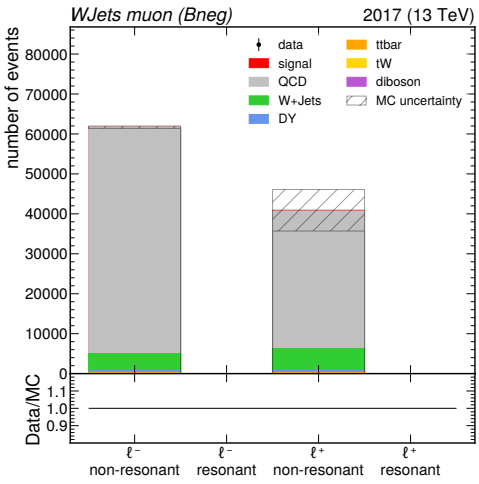
(a) B^+ W + jets muon



(b) B_0 W + jets muon

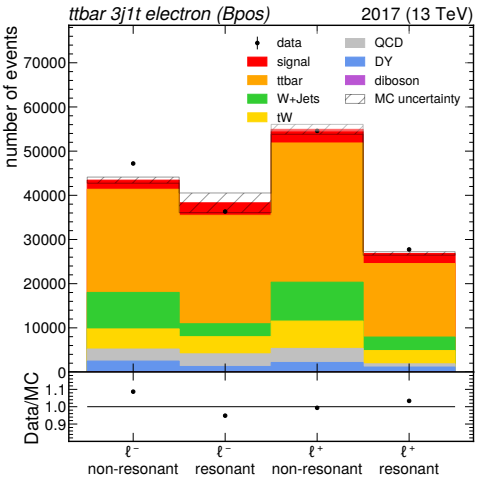


(c) $\bar{B}_0$ W + jets muon

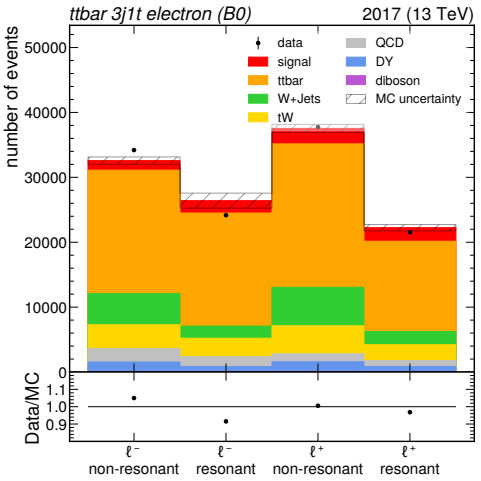


(d) B^- W + jets muon

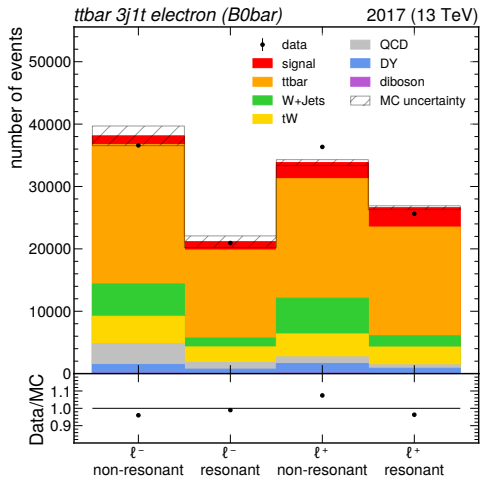
Figure B.4: Pre-fit distributions in the W + jets muon region of the Wb analysis



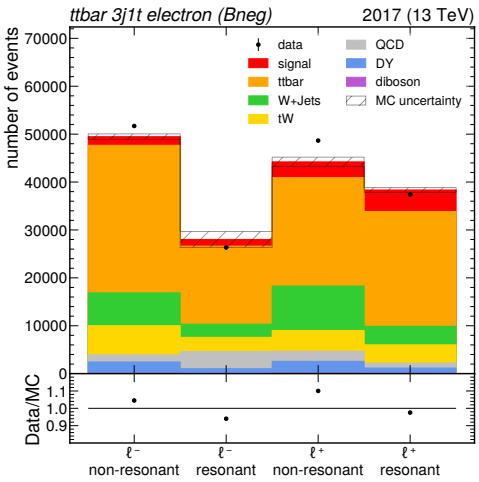
(a) $B^{+}t\bar{t}$ (1 tag) electron



(b) $B_0t\bar{t}$ (1 tag) electron

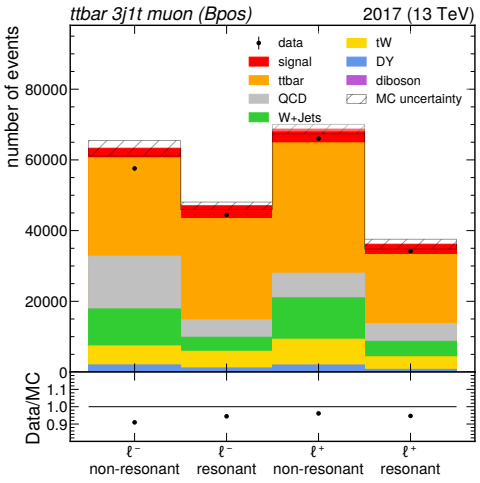


(c) $\bar{B}_0t\bar{t}$ (1 tag) electron

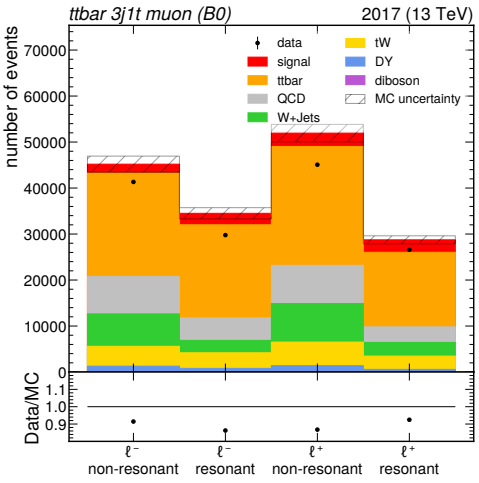


(d) $B^{-}t\bar{t}$ (1 tag) electron

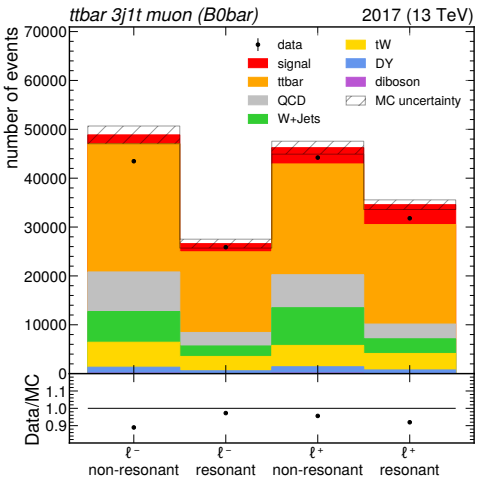
Figure B.5: Pre-fit distributions in the $t\bar{t}$ (1 tag) electron region of the Wb analysis



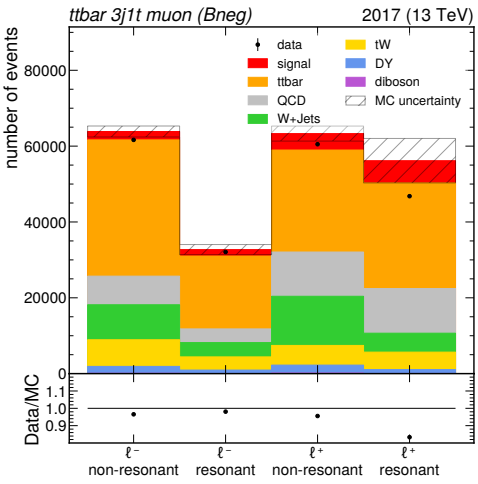
(a) $B^{+}t\bar{t}$ (1 tag) muon



(b) $B_0t\bar{t}$ (1 tag) muon

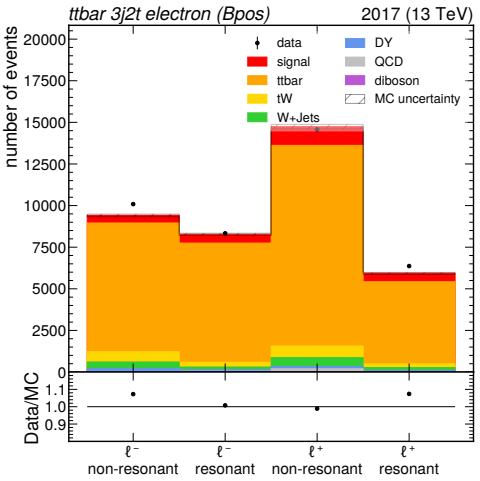


(c) $\bar{B}_0t\bar{t}$ (1 tag) muon

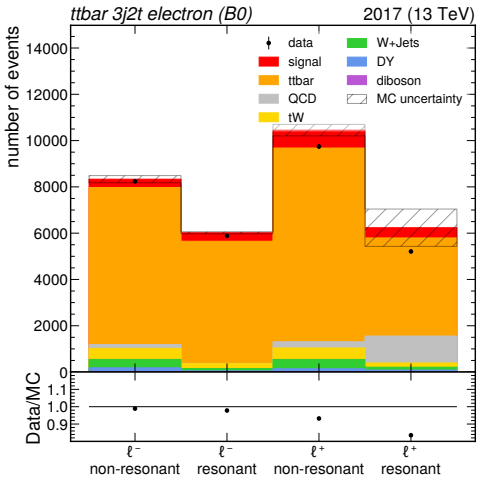


(d) $B^{-}t\bar{t}$ (1 tag) muon

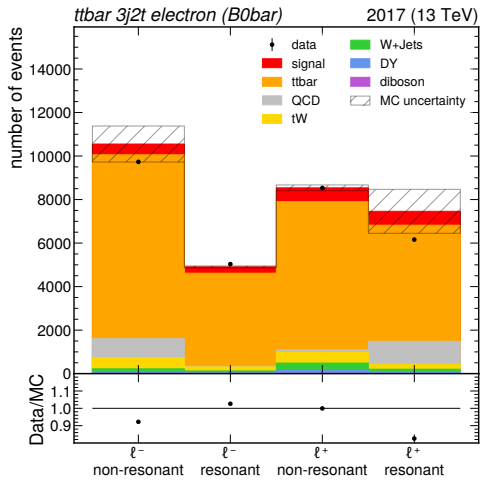
Figure B.6: Pre-fit distributions in the $t\bar{t}$ (1 tag) muon region of the Wb analysis



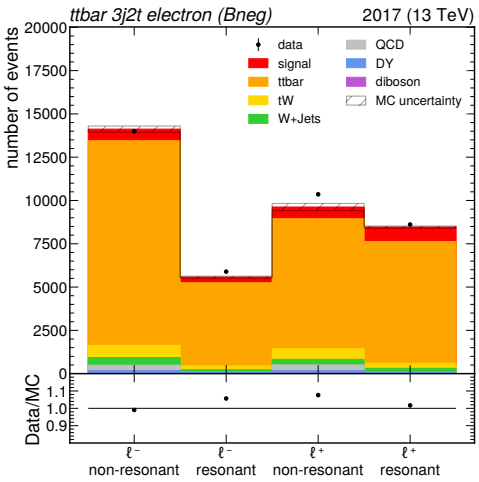
(a) $B^{+}t\bar{t}$ (2 tag) electron



(b) $B_0t\bar{t}$ (2 tag) electron

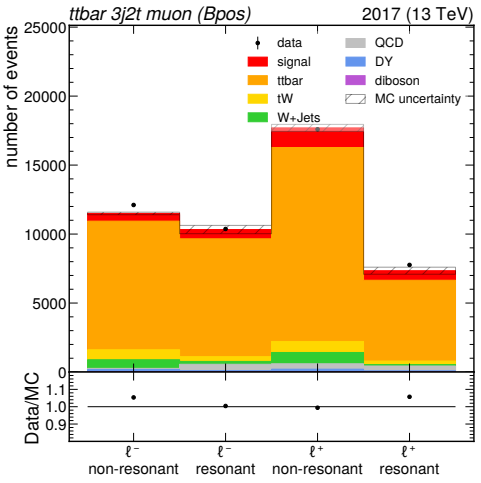


(c) $\bar{B}_0t\bar{t}$ (2 tag) electron

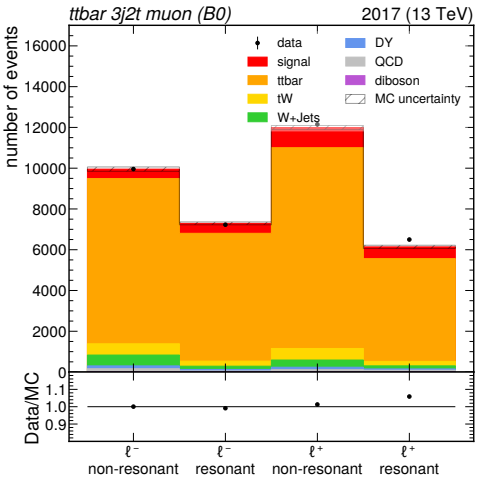


(d) $B^{-}t\bar{t}$ (2 tag) electron

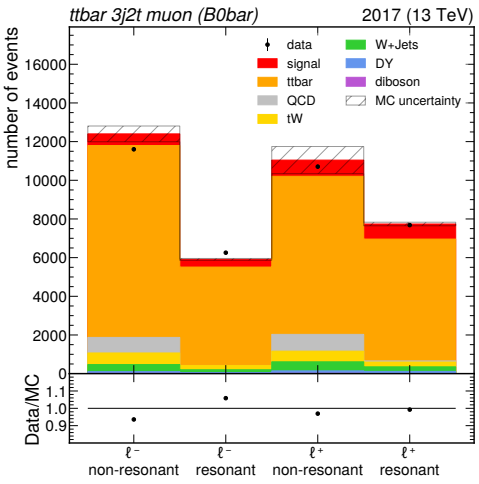
Figure B.7: Pre-fit distributions in the $t\bar{t}$ (2 tag) electron region of the Wb analysis



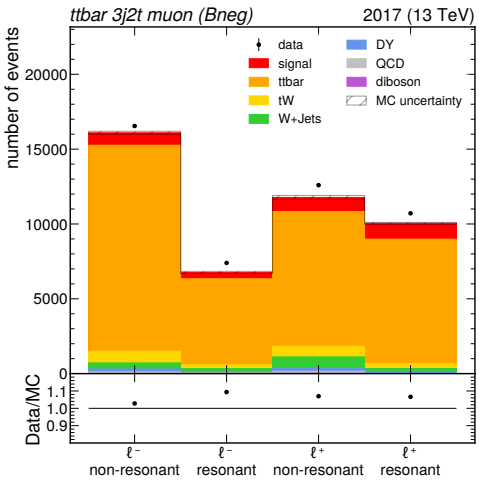
(a) $B^{+}t\bar{t}$ (2 tag) muon



(b) $B_0t\bar{t}$ (2 tag) muon



(c) $\bar{B}_0t\bar{t}$ (2 tag) muon



(d) $B^{-}t\bar{t}$ (2 tag) muon

Figure B.8: Pre-fit distributions in the $t\bar{t}$ (2 tag) muon region of the Wb analysis

C Temperature monitoring

Monitoring the temperature of server rooms is an important task regarding safety and efficiency. If air conditioning fails, it can lead to critical increase of temperature causing overheating of servers or even fire. This can be avoided by quickly responding to temperature rises and fixing the problem, using an alternative cooling method, reducing load on servers or shutting them down.

To replace the previously unreliable temperature sensor, a new temperature monitoring device based on an Arduino Nano microcontroller [73] is developed. For the temperature measurement, multiple digital sensors of the DS18B20-type are used. Their advantage is that they are addressable, allowing the use of the same hardware interface with multiple temperature sensors. The microcontroller, four connectors for the sensors and some other electronic components are mounted on a custom PCB, as depicted in figure C.1. The copper layers of the PCB are shown to scale in figure C.2.

The microcontroller is connected to a server via a USB-C cable. When the server requests a temperature reading via the serial connection, the up to four temperature sensors are read out and the measurement together with the address of the sensor is responded to the server. Requesting the temperature measurement for the monitoring is done via TELEGRAF [74], which is also used to monitor the server himself. The temperature measurement and a timestamp is then written into an INFLUXDB [75] database. The temperature against time is then interactively visualized in the GRAFANA [76] dashboard, shown in figure C.3. In case of temperature readings that exceed defined thresholds, alerts are sent by KAPACITOR [77] via mail and the application interface of the institutes chat service. The response time could be further reduced by setting up a push notification service.

(a) Rendered 3D model **(b)** Photograph of the installed module

Figure C.1: The temperature readout module PCB as 3D render (a) and fully integrated in the basement server room (b).

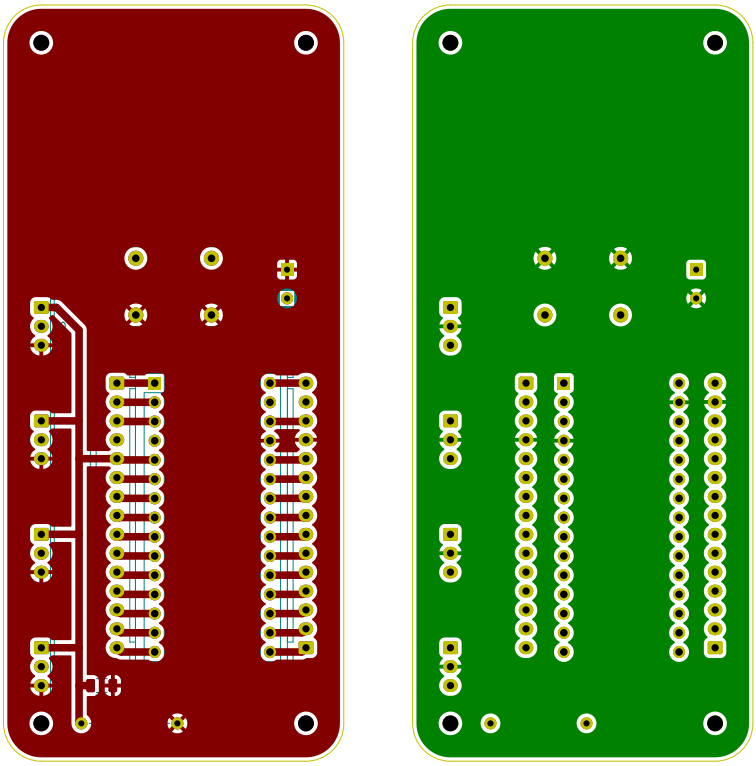


Figure C.2: Front (left) and back (left) copper layers of the PCB, both depicted to scale.

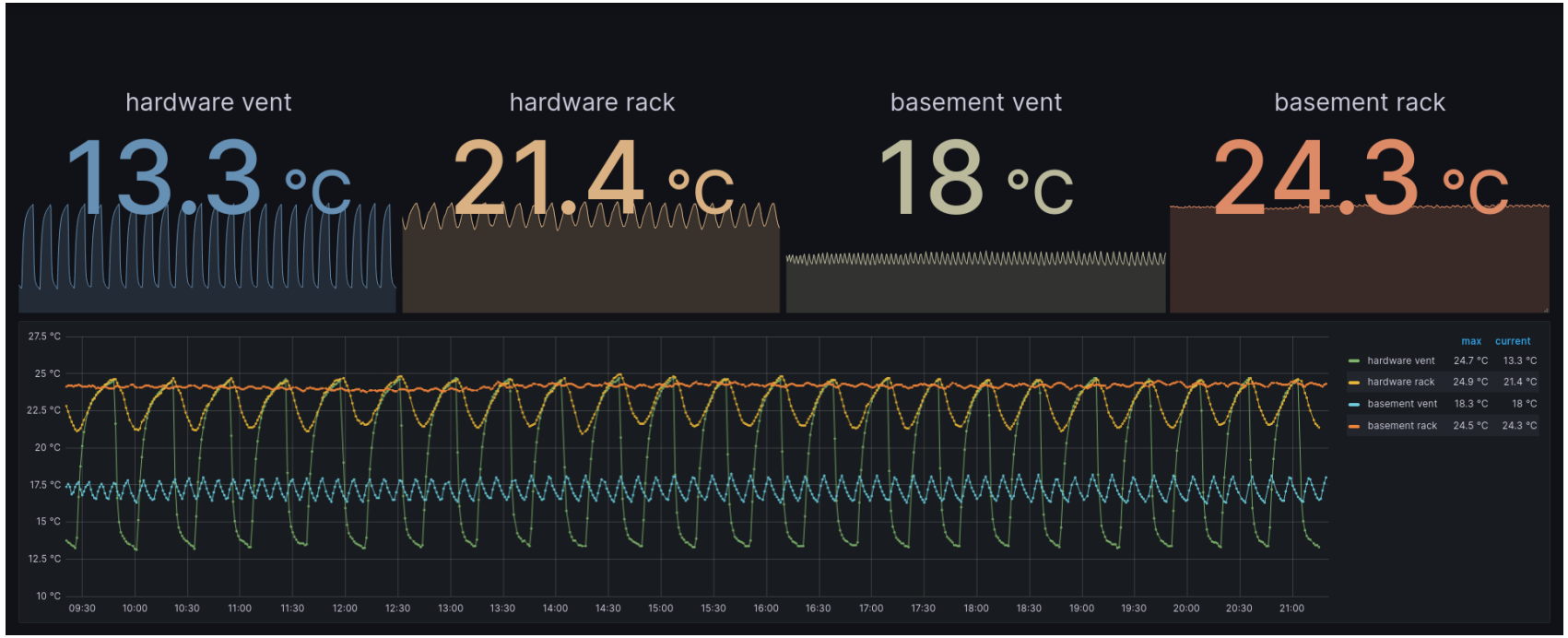
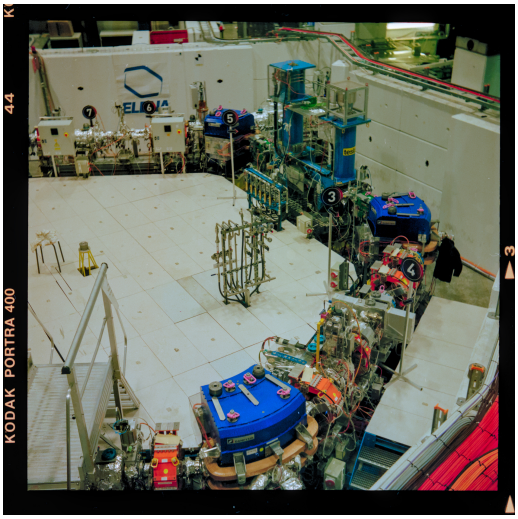
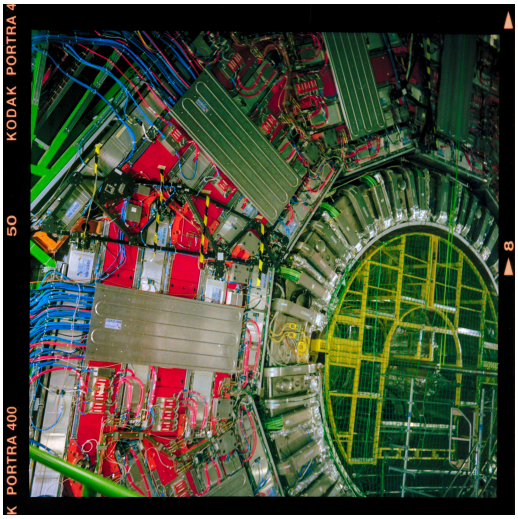


Figure C.3: Temperature monitoring dashboard in GRAFANA showing the measurements of two boards with two temperature sensors each over a time range of 12 hours. The switching of the air conditioning is visible as modulation in the temperature.

D Gallery



(a) The Extra Low ENergy Antiproton ring (ELENA) at the antimatter factory



(b) Barrel slice of the open CMS detector showing the muons system and the solenoid

(c) Barrel and endcap of the CMS detector with the beam pipe in between

(d) Endcap of the CMS detector

Figure D.1: Color pictures from CERN in January 2023

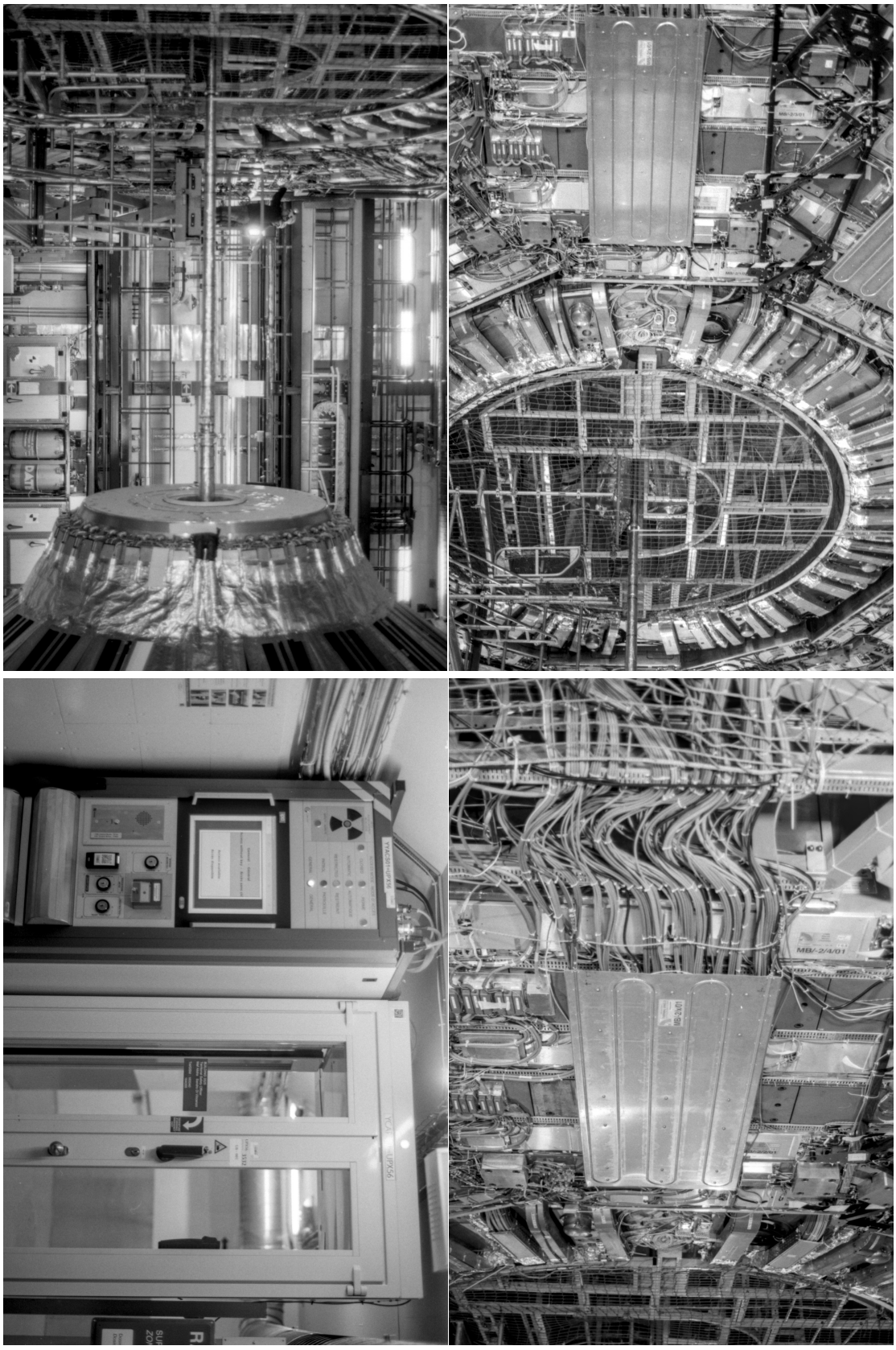


Figure D.2: Pictures from a CMS visit in January 2023 (part 1)

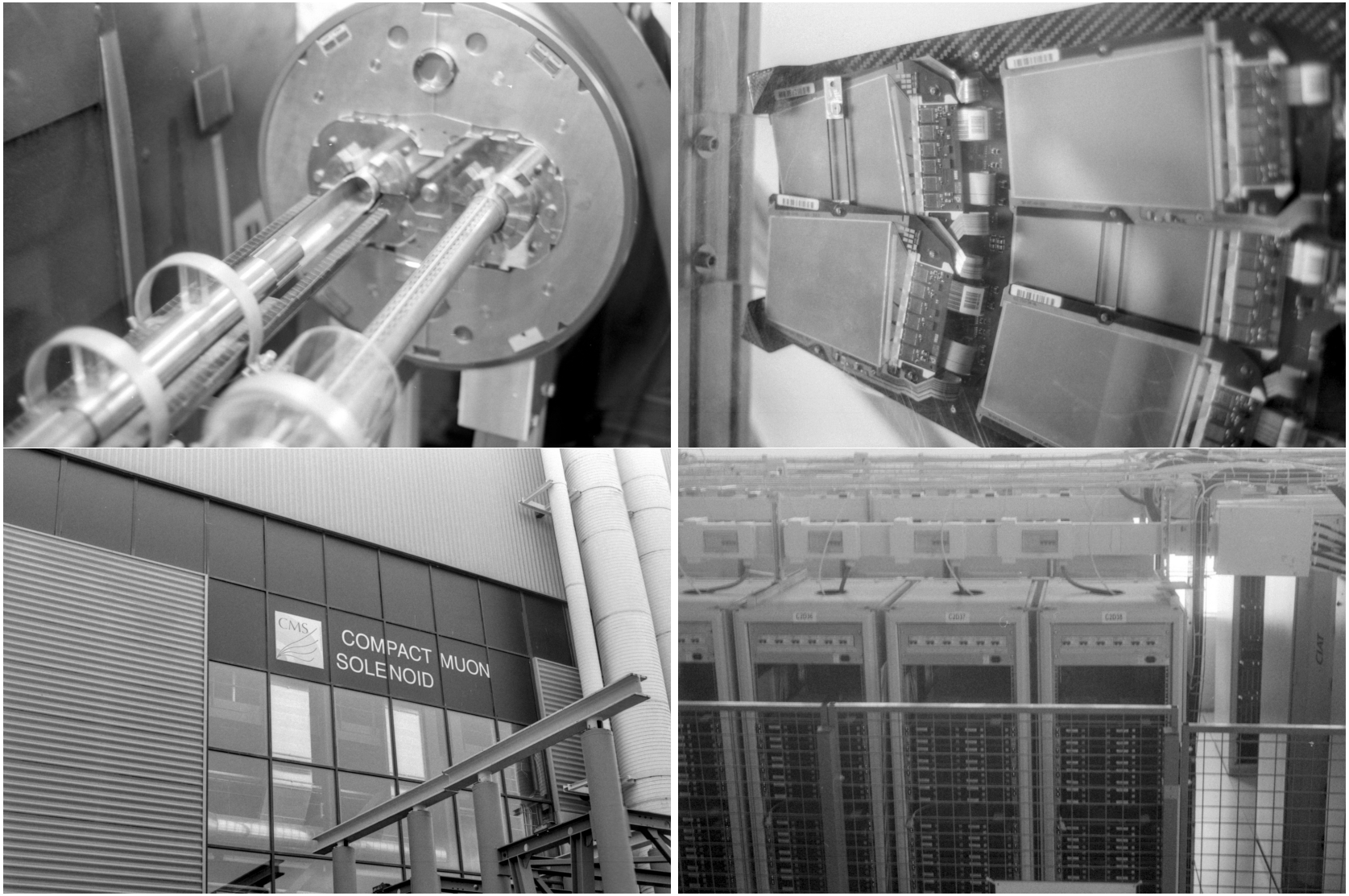


Figure D.3: Pictures from a CMS visit in January 2023 (part 2)

Danksagung

”So Long and Thanks for all the Fish”

Douglas Adams, The Hitchhiker’s Guide to the Galaxy

Teilchenphysik ist Teamwork: Ohne die Hilfe und das Engagement Vieler, wäre diese Arbeit nicht möglich gewesen. Ich danke meinem Referenten Prof. Dr. Thomas Müller, der mir die Möglichkeit gegeben hat frei zu forschen und der trotz Versetzung in den Ruhestand regelmäßig am Institut für Fragen anzutreffen ist. Auch danke ich meinem Korreferenten Prof. Dr. Günter Quast, der mir Einblicke in den Betrieb von GridKa ermöglichte und die Chance gab dort meine Servicearbeit zu leisten. Besonderer Dank gilt meinen beiden Betreuern Nils und Thorsten, die meine vielen Fragen geduldig beantworteten und mich auf dem Weg unterstützten. Bei Soureek bedanke ich mich für die Chance in der BTV-Gruppe viel über die Arbeitsweise der CMS-Kollaboration zu lernen. Dank gilt natürlich auch meinen Mitstreitern am Institut, die den Laden am Laufen halten und mir wertvolle Kommentare zu meiner Arbeit gegeben haben. Vielen Dank auch an die Sekretärinnen insbesondere Frau Bräunling, die mir beim Navigieren durch die Untiefen der Bürokratie geholfen hat.

Ich danke auch der $t\bar{t}H$ und Wb Analysegruppe für die konstruktive Zusammenarbeit und die kritischen Kommentare. Ohne die Arbeit der gesamten CMS-Kollaboration wären die Analysen nicht möglich gewesen. Gruß an meine Bürokollegen und die MIT-Sommerstudenten, die im Sommer 2022 unser Büro mit ihrer aufgeweckten und neugierigen Art bereicherten. Danke an KSETA für die Möglichkeit nach meiner Promotion weiter zu forschen und die Analysen voranzutreiben. Zum Schluss noch ein ganz großes Dankeschön an meine Familie, die mich während meines Studiums und der anschließenden Promotion unterstützt hat.