IEKP-KA/2016-29

MEASUREMENTS OF THE TOP-HIGGS COUPLING WITH THE CMS EXPERIMENT

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Hannes Mildner

aus Bad Soden

Tag der mündlichen Prüfung:
25. November 2016

Referent: Prof. Dr. Ulrich Husemann
Korreferent: Prof. Dr. Günter Quast

Introduction

In the summer of 2012, the ATLAS [1] and CMS [2] experiments at the CERN Large Hadron Collider [3] announced the discovery of a new particle in the search for a Higgs boson [4–6]. Since then, the mass of the particle has been measured with remarkable precision [7], and its couplings [8] and spin and CP-properties [9, 10] have been found to be consistent with the Higgs boson [11–17] predicted by the Standard Model (SM) of particle physics [18–21].

Assuming the particle really is the SM Higgs boson, it completes the SM as a self-consistent theory. Conceiving the SM is one of the greatest accomplishments of science. It is a predictive theory that describes nature on a subatomic level and can be elegantly constructed from symmetry principles. Together with general relativity it is the foundation from which, in principle, all models can be derived that are able to describe the physical phenomena observed on earth and that are relevant in everyday life. It is furthermore compatible with most findings of the particle physics experiments performed over the course of the last 50 years.

However, the theory still has a number of shortcomings, e.g. it cannot explain the existence of dark matter or the baryon anti-baryon asymmetry of the universe. Furthermore, some aspects of the SM are unsatisfying, e.g. that it contains a large number of parameters, some with seemingly peculiar values and that it is not possible to describe the SM and the theory of gravity in a unified framework. The mechanism of electroweak symmetry breaking, its energy scale, and the reason for the apparent mass hierarchy of elementary fermions is also not well understood.

After the Higgs boson discovery, the main tasks of experimental particle physics is to pin down limitations of the SM by testing all of its parts thoroughly and by searching for effects not predicted by it. One way to do this is via precise measurements of the properties of the newly discovered boson. Hereby its interactions, described in terms of couplings, offer a good starting point as they are very peculiar. They originate in the Higgs mechanism that is believed to allow W and Z bosons to acquire mass and that can also explain the mass of of the building blocks of matter, elementary fermions. One of these couplings is the top-Higgs coupling. It is by far the strongest coupling to any fermion and is thus important for many processes and might even play a special role in the mechanism of electroweak symmetry breaking.

So far, the existence and strength of the top-Higgs coupling can be inferred only indirectly, i.e. under certain assumptions, from the observation of Higgs production. The strength of the top-Higgs interaction has been measured this way, but the structure of the coupling has not yet been investigated by the experimental collaborations at the LHC. While there are stringent constraints on anomalous pseudoscalar Higgs interactions with W and Z bosons from the ATLAS and CMS collaborations, no such constraints have been published for the interaction with fermions. They can however be obtained by combining measurements of the rate of Higgs production in gluon fusion and the fraction of Higgs bosons that decay into photons. In this thesis this combination is performed using CMS Higgs signal strength measurements of LHC Run 1.

These deductions of the coupling properties are however rather indirect and modeldepended. The most direct measurement of the coupling strength and structure is possible by studying Higgs boson production in association with a top quark pair ($t\bar{t}H$). In Run 1 of the LHC in 2011 and 2012, CMS [22,23] and ATLAS [24–27] already observed – but did not expect – first hints for a signal in the search for $t\bar{t}H$. This surprising finding and the important role of $t\bar{t}H$ for a direct measurement of the top-Higgs coupling makes a discovery of this process one of the main tasks of LHC Run 2. A contribution to this search is made in this thesis by an analysis of events that are candidates for $t\bar{t}H$ production with the Higgs boson decaying into a $b\bar{b}$-pair ($t\bar{t}H(b\bar{b})$) and which contain one lepton and multiple jets.

As soon as $t\bar{t}H$ production is discovered, the process can be used to investigate not only the coupling strength but also the CP-structure of the coupling. This can be done by studying the kinematics of the Higgs boson and top quarks. In this thesis, a first measurement of the coupling structure is performed with the dataset available in 2015, also with $t\bar{t}H(b\bar{b})$ candidate events. While the current dataset is not expected to yield conclusive results, it is important to lay out a plan for this measurement and to demonstrate its feasibility.

This thesis is organized as follows: Part I gives an overview of the relevant theoretical foundations and techniques, the experimental setup, and the analysis tools and methods used. The foundation is the SM of particle physics described Chapter 1 and the methods to calculate predictions for hadron collider experiments are outlined in Chapter 2. The experimental environment, the CMS detector and the LHC particle accelerator, are explained in Chapter 3 and analysis techniques are detailed in Chapter 4. Part II documents the indirect measurement of the top-Higgs coupling. The current knowledge of the Higgs couplings is summarized in Chapter 5 and indirect constraints on the top-Higgs coupling are derived in Chapter 6. The direct coupling measurement is documented in Part III of the thesis. An introduction to the $t\bar{t}H(b\bar{b})$ signal and its background processes as well as a description of their simulation is given in Chapter 7. In Chapter 8 the definition of analysis objects is detailed. The selection, categorization, and multivariate analysis of events that is used to extract the $t\bar{t}H(b\bar{b})$ signal is documented in Chapter 9. The statistical analysis in the context of the $t\bar{t}H$ search is described in Chapter 10. This analysis is extended to an analysis of the coupling structure in Chapter 11 and combined with $t\bar{t}H$ searches in different final states in Chapter 12.

Contents

I. Theoretical foundation, experimental setup, tools and methods

1 Standard Model

Particle physics aims at unraveling the laws that govern nature on a microscopic scale. The current, generally accepted, status in this endeavor is summarized in the Standard Model (SM). The SM describes subatomic particles and their interactions in terms of quantized fields. Mathematically, it is a relativistic quantum field theory (QFT), more specifically a U(1)×SU(2)×SU(3) gauge theory. QFTs allow to construct models in accordance with the principles of quantum mechanics and special relativity. They can be formulated in terms of Lagrangian densities $\mathcal{L}(\{\phi_i(x)\}, \{\partial_\mu \phi_i(x)\})$ that are functions of field operators $\phi_i(x)$ which describe the probability to find a particle of type i at the space-time point x and their derivatives $\partial_\mu \phi_i(x)$. In the SM, nature is described in terms of elementary fermions, matter particles, that interact via gauge bosons. For the consistency of the SM, the Higgs[1] boson needed to be introduced.

A short overview of the elementary fermions and their interactions will be given in Section 1.1. Because of their relevance for this thesis, discrete symmetries and CP violation are shortly discussed in Section 1.2. Section 1.3 serves as a reminder of the mathematical formulation of the SM, with a focus on the electroweak (EWK) theory, the unified theory of weak and electromagnetic interactions.

1.1 Historical overview and particle content

The historical review these chapters contain loosely follows [28].

1.1.1 Elementary fermions

The fermions of the Standard Model, summarized in Table 1.1, constitute the building blocks of matter. Their intrinsic angular momentum, the spin, has an absolute value of $\frac{1}{2}\hbar$ and they are thus subject to the Pauli exclusion principle. Two fermions cannot be in the same quantum state, which forces them to form more complex and stable structures, e.g. nucleons, atoms, and finally everything we observe as matter.

Properties of elementary fermions

Fermions are classified as either quarks or leptons, of which only quarks take part in the strong interaction. There are three quantum fields for each type of quark, distinguished by their “color”: red, green, and blue. The introduction of three colors was necessary to allow for the existence of baryons that consist of three quarks with otherwise identical quantum numbers without violating the Pauli principle.

Both, leptons and quarks can be further distinguished by their mass and by their electroweak interaction properties. Quarks carry an electric charge of $-\frac{1}{3}$ or $\frac{2}{3}$ times the elementary charge and leptons a charge of either one or zero. Fermions with zero electric

---

[1] Also Anderson, Brout, Englert, Guralnik, Hagen, Kibble, t’Hooft and many more contributed to establishing of the mechanism that predicted the particle that will be called Higgs boson in this thesis.

Table 1.1: The fermions of the Standard Model, together with their mass and interaction-defining properties. The third component of the weak isospin, T_3, is quoted for particles of left-handed chirality, right-handed particles are weak isospin singlets with $T_3 = 0$. Light quark (u, d, s) masses are estimated as current quark masses, for the top-quark mass the mass from direct measurements is quoted, and the remaining quark masses are given in the $\overline{\text{MS}}$ scheme. Uncertainties on the mass values that are below 0.1% are not displayed. The neutrinos masses are given as upper limits as deduced from tritium decay. All numbers are taken from [29].

Particle	Charge/e	Weak isospin T_3	Color	Mass / MeV/c^2
Electron neutrino (ν_e)	0	$\frac{1}{2}$	none	< 0.002
Electron (e^-)	-1	$-\frac{1}{2}$	none	0.511
Muon neutrino (ν_μ)	0	$\frac{1}{2}$	none	< 0.002
Muon (μ^-)	-1	$-\frac{1}{2}$	none	105.7
Tau neutrino (τ_τ)	0	$\frac{1}{2}$	none	< 0.002
Tau (τ^-)	-1	$-\frac{1}{2}$	none	1777
Up quark (u)	$\frac{2}{3}$	$\frac{1}{2}$	r,g,b	$2.3^{+0.7}_{-0.5}$
Down quark (d)	$-\frac{1}{3}$	$-\frac{1}{2}$	r,g,b	$4.8^{+0.5}_{-0.3}$
Charm quark (c)	$\frac{2}{3}$	$\frac{1}{2}$	r,g,b	1275 ± 25
Strange quark (s)	$-\frac{1}{3}$	$-\frac{1}{2}$	r,g,b	95 ± 5
Top quark (t)	$\frac{2}{3}$	$\frac{1}{2}$	r,g,b	$(173.2 \pm 0.87) \times 10^3$
Bottom quark (b)	$-\frac{1}{3}$	$-\frac{1}{2}$	r,g,b	$(4.18 \pm 0.03) \times 10^3$

charge are called neutrinos. The charge determines the particle behavior in the electromagnetic interaction. The weak interaction properties are determined by the weak isospin and, to some extend, the charge, too. Fermion fields can be decomposed into "left-handed" and "right-handed" chirality. The pairs of left-handed fermions that are grouped by the horizontal lines in Table 1.1 form doublets of the weak isospin, which allows them to be transformed into each other by the weak interaction. Right-handed particles do not participate in the weak interaction in this fashion.

For every fermion, an antiparticle with opposite internal quantum numbers, e.g. opposite electric charge, but the same mass exists. The elementary forces allow not only interactions between particles that change their momenta but pairs of particle and antiparticles can also be annihilated or created.

Discovery of the fermions of the Standard Model

The first of the elementary fermions observed was the electron, discovered 1897 by Thomson. It was hypothesized by Pauli in 1930 that the electron has a chargeless partner, the electron neutrino, but confirmed only in 1956 [30]. The mass of neutrinos is extremely small and for a long time they were believed to be massless. This could however be disproven with the discovery of neutrino oscillations in 1998 [31]. A second charged lepton, the muon, was discovered in 1936 in cosmic rays. It is similar to the electron, but its mass is larger by a factor 200. The muon forms a lepton generation together with the muon neutrino, indirectly inferred from muon decays and detected in 1962 [32]. Since then even a third generation of leptons was observed, the heavy tau was discovered in 1975 [33] and the corresponding neutrino only in 2001 [34]. Additional light neutrinos can be excluded from measurements at electron-positron colliders [35]. The observation of charged leptons is relatively simple, electrons and muons leave observable tracks as they ionize matter they traverse and the heavy taus decay characteristically into lighter particles, some of them charged. Neutrinos interact only via the weak interaction. Because of the weakness of this force, neutrinos hardly ever interact and are thus almost undetectable.

In contrast to leptons, quarks cannot be observed individually. They can however form a multitude of composite particles, so called hadrons, and were first theorized as their constituents by Gell-Mann and Zweig in the early sixties. The only hadrons with a lifetime of more than 1 μs, protons and neutrons, are made of three of the two lightest quarks, up quarks and down quarks. The two quarks form an isospin doublet so that they can be transmuted by the weak interaction if kinematically allowed, which is observed as β decay. Together with a third quark, the strange quark, the two quarks can form hadrons whose properties were perceived to be curious at the time of their discovery. Like the down quark is the partner of the up quark in weak interactions, the strange quark can be paired with the charm quark. This was postulated by the GIM mechanism in 1970 [36] and first hadrons containing charm quarks were discovered in 1974 [37, 38]. Even before that, it was realized that a third generation of quarks is necessary to explain the observation of CP violation [39]. After the discovery of the bottom quark in 1977 [40], it took almost 20 years to find its heavy partner, the top quark [41, 42]. In contrast to the lighter quarks, the top quark does not form hadrons but decays almost immediately after its production.

1.1.2 Interactions

In the SM, interactions are described by gauge fields. The quanta of gauge fields have spin one, are called gauge bosons, and act as force carriers propagating interactions between fermions. Three of the four elementary interactions can be described in terms of quantum fields: the electromagnetic force, the weak force, and the strong force. Gravity cannot be included in this framework but is, compared to the other forces, very weak and can usually be safely ignored in particle physics phenomenology.

Electromagnetism

Electric and magnetic phenomena are known since ancient times and have first been described as a unified field theory by Maxwell. The quantum of the electromagnetic field, the photon, was postulated by Einstein in 1905, during the dawn of quantum mechanics. Single photons could first be detected in Compton scattering in 1923. The effort to describe electromagnetic interactions in a relativistic field theory culminated in the formulation quantum electrodynamics by Tomonaga, Schwinger, and Feynman.

Weak interaction

The weak force is of small range and needed to be introduced for the description of flavor-changing phenomena like the β decay of nucleons and muon decay. It was first formulated as a contact interaction by Fermi in 1933. In the SM the interaction propagated by the W^+, W^-, and Z bosons, also called vector bosons. In contrast to the other gauge bosons, which are massless, W and Z bosons have a mass of approximately $80.4\,\mathrm{GeV}/c^2$ [29] and $91.2\,\mathrm{GeV}/c^2$ [29], respectively. For a mathematically consistent formulation of the weak interactions, many challenges had to be overcome. This lead to the introduction of the Higgs mechanism and the Higgs boson [11–17] and finally to a unified description of electromagnetic and weak force [18, 19, 21].

First interactions propagated by the $\mathrm{W}^\pm$ boson, β decay, were observed in the 19th century. However, electroweak theory also describes the interaction of Z bosons with matter. These interactions were observed in 1973 [43]. Together with the proof that the theory is renormalizable [20] and can thus make mathematically sound predictions, this lead to the breakthrough of the SM. The W and Z bosons were produced and detected for the first time in 1983 [44–47] at the CERN Super Proton Synchrotron. It took almost 30 more years until the Higgs boson was observed in 2012 [4, 5] at the LHC.

Strong interaction

The strong interaction is what confines quarks in hadrons and also allows for the formation of protons and neutrons to nuclei. The basis for its theoretical description was already given with the postulation of quarks and color charge in 1964. At the time, two types of hadrons could be observed: baryons like protons or neutrons, consisting of three quarks, and mesons, which consist of a quark and a anti-quark. The strong interactions does not allow quarks to exists outside of these bound states, a phenomenon called confinement. In a sense, the bound states themselves are neutral in terms of the strong interaction. While e.g. protons and neutrons can be found in tightly bound nuclei, their attraction is only a small remnant of the force that binds quarks to nucleons (which is thus aptly called the strong force). This motivates the introduction of “color” as the charge of the strong interaction: the three quarks in baryons have, in this picture, three colors adding up to the neutral white. Similarly, quark and antiquark in a meson carry color and anti-color and thus also combine to white.

The strong force can be described by a gauge theory called quantum chromodynamics (QCD). It became widely accepted when it was shown in 1973 [48, 49] that this theory possesses a feature called asymptotic freedom. It can be interpreted as the force becoming weaker at high energy scales. This then allows to describe QCD interactions in perturbation theory and to make predictions that can be tested in experiments. One of these predictions were three-jet events at electron-positron colliders, in which two jets can be described as originating from quarks and one from a gluon. Such events were first observed at the end of the 1970s at experiments at the PETRA collider of DESY [50].

1.2 CP violation

Symmetries are a powerful concept in particles physics because they are often accompanied by conserved quantities. A system is said to possess a symmetry if it is unaffected by a transformation. In particles physics, several discrete transformations exist that are very useful because they are exact or approximate symmetries of nature.

One example is parity symmetry, i.e. invariance under inversion of coordinates. Strong and electromagnetic interactions are invariant under parity transformations, mathematically described by the parity operator. This means that the mirror images of processes mediated by the strong or electromagnetic force are equally possible. Particles can be eigenstates of the parity operator with parity eigenvalues (also simply referred to as parity) $P = \pm 1$. For example, quarks are assigned a parity of $P = +1$ (even parity) while antiquarks have parity $P = -1$ (odd). For composite systems, the parity is the product of intrinsic parities multiplied by $(-1)^l$, where l is the relative angular momentum. Spin-zero particles with $P = +1$ ($P = -1$) are called scalars (pseudoscalars). Spin-one particles are called vectors for $P = -1$ and pseudovectors for $P = +1$.

Parity is conserved in strong and electromagnetic interactions. However, it was suggested in 1956 [51] and discovered in 1957 [52] that the weak force does not preserve parity. In fact, in weak interactions parity is maximally violated, as the weak force couples, in charged-current interactions, only to particles of left-handed chirality. For massless particles, chirality corresponds to the helicity, the projection of the spin on the momentum direction. Since neutrinos are almost massless, they are almost always of left-handed helicity when emitted in a β decay while the mirror image of the process cannot be observed.

A related discrete transformation is charge conjugation, which reverts all internal quantum numbers of particles. Only particles that are their own antiparticles can thus be eigenstates of this symmetry. Examples of such particles are the π^0 or photons. For photons the eigenvalue is $C = -1$ and for mesons $(-1)^{s+l}$, where s is the total spin. C is also a multiplicative quantum number and conserved in the electromagnetic and strong interaction but not in weak processes. For example, left-handed neutrinos take part in charged current interactions while their charge-conjugated counterparts, the left-handed antineutrinos, do not.

The combination of the two operations, a CP transformation, is an approximate symmetry of weak interactions and violated only in rare processes. It can be used to assign particles a CP quantum number $CP = +1$ (CP even) or $CP = -1$ (CP odd). CP violation in weak interactions was first discovered in 1964 [53] in kaon decays. Neutral kaons and their antiparticles can mix via the weak interaction and propagate as two distinct state, the $\mathrm{K_L}$ and $\mathrm{K_S}$, which would be CP eigenstates if the symmetry was not violated. However, while $\mathrm{K_L}$ decays were observed to decay predominantly into CP-odd final states, rare decays into CP-even two-pion final states indicated that $\mathrm{K_L}$ are not exact CP-odd eigenstates but contain a small CP-even admixture. Since then, additional processes violating the CP symmetry have been discovered. This includes direct CP violation in kaon [54] and B hadron decays [55,56], in which certain decays are more likely than their CP conjugate, as well as direct and indirect CP violation in additional processes involving B hadrons.

In the SM, CP violation is introduced by a complex phase in the Cabibbo-Kobayashi-Maskawa (CKM) matrix [39], which describes the mixing of quark flavors to mass eigenstates. It was in fact the observation of CP violation that lead to the introduction of the three-dimensional CKM matrix and the postulation of a third quark generation because only in three dimensions the CP-violating phase can be incorporated.

The existence of CP violation is one of three criteria necessary to allow baryogenesis [57], i.e. the generation of the apparent baryon-antibaryon imbalance in the universe. However, the CP violation introduced by the CKM matrix is likely not enough to explain the observed asymmetry [58] and additional sources of CP violation might exist beyond the Standard Model. In some BSM models, the Higgs sector can be a source of CP violation [58, 59]. An example are two-Higgs-doublet models, which introduce additional Higgs-bosons, one of which is a neutral pseudoscalar. In the presence of CP violation, the observed state at 125 GeV/c^2 could contain a CP-odd admixture, analogous to the admixture in the kaon system.

1.3 Mathematical formulation

This section can obviously neither give a complete overview of the SM nor be a pedagogical introduction. Instead, is it meant to remind the reader of the concepts and relations most important for this thesis. For brevity, natural units with $\hbar = 1 = c$ will be used in this section.

1.3.1 Quantum electrodynamics

The dynamics of relativistic fermions can be described in terms of the Dirac Lagrangian

$$\mathcal{L}_\mathrm{D} = \bar{\psi}(x)(\mathrm{i}\gamma^\mu\partial_\mu - m)\psi(x), \tag{1.1}$$

where $\psi(x)$ are the Dirac field operators, $\bar{\psi} = \psi^\dagger\gamma^0$ is the adjoint operator, γ_μ the anti-commuting γ-matrices, m the fermion mass, and summation over indices appearing twice is always implied. It can be observed that the Lagrangian is invariant under a global transformation parameterized by θ

$$\begin{aligned} \psi(x) &\to \exp(\mathrm{i}\theta)\psi(x), \\ \bar{\psi}(x) &\to \exp(-\mathrm{i}\theta)\bar{\psi}(x). \end{aligned} \tag{1.2}$$

According to Noether's theorem the existence of this symmetry implies an associated conserved current $j^\mu(x) = \bar{\psi}(x)\gamma^\mu\psi(x)$, which is proportional to electric charge and current density $j^\mu = (\rho, \vec{j})$.

The global symmetry transformation can be promoted to a local one by allowing θ to be a function of space and time, i.e. $\theta(x)$. Without additional modifications, $\mathcal{L}_D$ is no longer invariant under this transformation. This can be overcome by introducing a new field $A_\mu(x)$, a vector under Lorentz transformation and thus associated with a spin 1 particle. It is required to transform

$$A_\mu(x) \to A_\mu(x) - \frac{1}{e}\partial_\mu\theta(x) \tag{1.3}$$

under gauge transformation. In Equation 1.1, ∂_μ is then be replaced by the covariant derivative

$$D_\mu = \partial_\mu + \mathrm{i}eA_\mu(x), \tag{1.4}$$

which ensures that $\psi(x)$ and its derivative $D_\mu\psi(x)$ have the same behavior under gauge transformation. By the replacement, an additional term

$$\mathcal{L}_\mathrm{I} = -eA_\mu\bar{\psi}\gamma^\mu\psi \tag{1.5}$$

enters the Lagrangian, describing interactions between a vector field and a fermion field. The field $A_\mu(x)$ can be identified with the electromagnetic potential, which has exactly the described gauge freedom and interaction properties.

The dynamics of a vector field $A_\mu(x)$ are determined by the Proca equation

$$\mathcal{L}_A = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m_A^2 A^\mu A_\mu. \tag{1.6}$$

The electromagnetic field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is invariant under gauge transformations, but the second term spoils the gauge freedom, so that $m_A = 0$ has to be required. This is in accordance with the observation that the photon is massless. Putting it all together, one obtains the Lagrangian of quantum electrodynamics, describing the electromagnetic interaction of a spin-1/2 particle.

$$\mathcal{L}_\mathrm{QED} = \mathcal{L}_\mathrm{D} + \mathcal{L}_\mathrm{I} + \mathcal{L}_A = \bar{\psi}(\mathrm{i}\gamma^\mu\partial_\mu - m)\psi - eA_\mu\bar{\psi}\gamma^\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{1.7}$$

It is an example of a U(1) gauge theory because the symmetry transformations introduced in Equation (1.2) belong to the group of one-dimensional unitary transformations.

1.3.2 Electroweak interactions

The success of quantum electrodynamics motivates the description of the weak interaction as a gauge theory. Its formulation requires a non-abelian SU(2) gauge theory, introducing more complicated features and subtleties. Furthermore, to describe the observed properties of weak interactions, massive vector bosons need to be introduced via the Higgs mechanism and a unified electroweak theory has to be constructed. The following description is partially taken from [18, 60, 61].

Formulation as gauge theory

For simplicity, the electroweak theory will first be introduced with only the electron field $\mathrm{e}(x)$ and the corresponding neutrino field $\nu(x)$. Fermion fields can be decomposed into left-handed and right handed chirality by applying the corresponding projection operator:

$$\psi_L = \frac{1}{2}(1-\gamma_5)\psi, \ \psi_R = \frac{1}{2}(1+\gamma_5)\psi. \tag{1.8}$$

To obtain a gauge theory that is parity-violating and contains an interaction coupling to charged currents, i.e. allowing for the transformation of left-handed electrons into neutrinos and vice versa, the left-handed chirality states of the two particles are grouped in a doublet that transforms under SU(2) gauge transformations. These weak-isospin doublets represent states in which the third component of the isospin is $T_3 = \frac{1}{2}$ or $T_3 = -\frac{1}{2}$.

Charged-current interactions couple only to left-handed particles, thus right-handed leptons transform as weak-isospin singlets. One can construct a Lagrangian possessing a SU(2) symmetry, in addition to a U(1) symmetry, as long as one omits the fermion mass:[2]

$$\begin{aligned}\mathcal{L}_{\text{free}} &= \bar{\nu}_L(\mathrm{i}\gamma^\mu\partial_\mu)\nu_L + \bar{\mathrm{e}}_L(\mathrm{i}\gamma^\mu\partial_\mu)\mathrm{e}_L + \bar{\mathrm{e}}_R(\mathrm{i}\gamma^\mu\partial_\mu)\mathrm{e}_R \\ &= \bar{\Psi}_L(\mathrm{i}\gamma^\mu\partial_\mu)\Psi_L + \bar{\mathrm{e}}_R(\mathrm{i}\gamma^\mu\partial_\mu)\mathrm{e}_R.\end{aligned} \tag{1.9}$$

Writing $\Psi_L = (\nu_L, \mathrm{e}_L)$ easily allows to identify the invariance of the left-handed doublets under $\mathrm{SU}(2)_L$ transformations, which can be expressed with a matrix $U \in \mathrm{SU}(2)$ that mixes the left-handed fermions:

$$\Psi_L \to U\Psi_L = \exp\left(\frac{\mathrm{i}}{2}\sum_{a=1}^{3}\alpha^a\tau^a\right)\Psi_L(x), \tag{1.10}$$

where τ^a are the Pauli matrices, the generators of the Lie group SU(2) and α^a are three parameters. Mass terms for electrons $m_\mathrm{e}\bar{\mathrm{e}}\mathrm{e} = m_\mathrm{e}\bar{\mathrm{e}}_L\mathrm{e}_R + m_\mathrm{e}\bar{\mathrm{e}}_R\mathrm{e}_L$ would spoil the $\mathrm{SU}(2)_L$ symmetry and cannot be introduced at this point. U(1) transformations can be performed with different hypercharges Y for all fields $\psi_{L/R}$, in contrast to the non-abelian $\mathrm{SU}(2)_L$ that allows only for a universal coupling:

$$\psi_{L/R} \to \exp\left(\frac{\mathrm{i}}{2}Y\theta\right)\psi_{L/R}. \tag{1.11}$$

Again, the symmetries are required to hold locally for $\alpha^a \equiv \alpha^a(x)$ and $\theta \equiv \theta(x)$, which can be achieved by introducing a covariant derivative

$$D_\mu = \partial_\mu - \mathrm{i}\frac{g'}{2}YB_\mu - \mathrm{i}\frac{g}{2}\sum_{a=1}^{3}\tau^a W_\mu^a, \tag{1.12}$$

where g and g' are the SU(2) and U(1) gauge couplings and the third term only acts on left-handed fields. In analogy to A_μ, there are the gauge field B_μ and three fields W_μ^a. Under a local SU(2) gauge transformation described by α^a and a U(1) transformation described by θ, the fields are required to transform

$$B_\mu \to B_\mu + \frac{1}{g'}\partial_\mu\theta, \tag{1.13}$$

$$W_\mu^a \to W_\mu^a + \frac{1}{g}\partial_\mu\alpha^a + \epsilon^{abc}W_\mu^b\alpha^c, \tag{1.14}$$

where ϵ^{abc} is the Levi-Civita symbol. The non-abelian nature of the SU(2) also leads to an additional term in the field strength tensor of the latter fields:

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a + \partial_\nu W_\mu^a + g\epsilon^{abc}W_\mu^b W_\nu^c \tag{1.15}$$

Adding the terms introduced by the covariant derivative and the kinetic terms of the vector fields to $\mathcal{L}_{\text{free}}$, the Lagrangian of the model now reads

$$\begin{aligned}\mathcal{L} =& \bar{\Psi}_L(\mathrm{i}\gamma^\mu(\partial_\mu - \sum_{a=1}^{3}\mathrm{i}\frac{g}{2}\tau^a W_\mu^a - \mathrm{i}\frac{g'}{2}YB_\mu))\Psi_L \\ &+ \bar{\mathrm{e}}_R(\mathrm{i}\gamma^\mu(\partial_\mu - \mathrm{i}\frac{g'}{2}YB_\mu))\mathrm{e}^R \\ &- \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W^{a\mu\nu}W_{\mu\nu}^a,\end{aligned} \tag{1.16}$$

---

[2] For simplicity, neutrinos are assumed to be massless and right-handed neutrinos assumed to be non-existent.

which describes the interaction of massless electrons and muons via two long-range interactions of strength g and g'. However, nature is different, containing fermions with mass and different interactions of which one is short ranged and must thus have heavy force carriers.

Electroweak unification

The off-diagonal elements of $\sum_{a=1}^{3} \mathrm{i}\frac{g}{2}\tau^a$ in Equation (1.16) couple electron and neutrino to a linear combination of W_1 and W_2:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp \mathrm{i}W_\mu^2). \tag{1.17}$$

Comparing the strength of the coupling with the Fermi constant G_F, measured in low-energy interactions yields

$$G_F = \frac{\sqrt{2}}{8}\frac{g^2}{m_W^2}. \tag{1.18}$$

Here m_W, the mass the W field will acquire later, enters via the boson propagator.

Both, the B_μ and W_μ^3 do not change flavor or charge of leptons and can be expressed as linear combinations of the fields representing Z bosons and photons:

$$W_\mu^3 = \sin\theta_W A_\mu + \cos\theta_W Z_\mu, \tag{1.19}$$

$$B_\mu = \cos\theta_W A_\mu - \sin\theta_W Z_\mu, \tag{1.20}$$

where θ_W is the Weinberg angle. Expressing Equation (1.16) in terms of these new fields, one obtains the coupling terms

$$\bar{\psi}\gamma^\mu A_\mu(gT_3 \sin\theta_W + \frac{g'}{2}Y\cos\theta_W)\psi \tag{1.21}$$

and

$$\bar{\psi}\gamma^\mu Z_\mu(gT_3 \cos\theta_W - \frac{g'}{2}Y\sin\theta_W)\psi, \tag{1.22}$$

where ψ can stand for right and left-handed fermion fields and T_3 is the third component of the weak isospin, which is 0 for right-handed singlets and $\pm\frac{1}{2}$ for the doublets. To obtain QED interactions, i.e. a field A_μ that couples equally to left- and right-handed electrons with strength e and does not couple to neutrinos, the following relations have to be fulfilled

$$e = g\sin\theta_W = g'\cos\theta_W, \tag{1.23}$$

$$Q = T_3 + \frac{1}{2}Y, \tag{1.24}$$

relating weak and electromagnetic coupling constants and determining the value of the hypercharge.

1.3.3 Electroweak symmetry breaking

It is possible to give the gauge bosons as well as the fermions mass by introducing a complex scalar isodoublet with hypercharge 1: $\Phi = (\phi^+, \phi^0)^\top$. The most general renormalizable and gauge invariant Lagrangian density for such fields is given by

$$\mathcal{L}_S = (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi), \tag{1.25}$$

where

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda |\Phi^\dagger \Phi|^2 \tag{1.26}$$

is the field's potential. In the SM is $\mu^2 < 0$ and the ground state, i.e. the minimum of the potential, is non-zero. One can choose the vacuum expectation value, which corresponds to one of the minima of the potential, to be $(0, v/\sqrt{2})^\top$, with

$$v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{1.27}$$

and rewrite, without loss of generality,

$$\Phi(x) = \exp\left\{ \mathrm{i} \sum_{a=1}^{3} \theta^a(x) \tau^a \right\} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H(x)) \end{pmatrix}, \tag{1.28}$$

in terms of four fields $\{\theta^a(x)\}$ and $H(x)$. The fields $\theta^a(x)$ do not change the value of the potential of the particle and are thus massless. The $H(x)$ field describes excitations out of the ground state, the behavior of a massive scalar field. The particle corresponding to H is the Higgs boson. If the SU(2) symmetry was global, the $\theta(x)$ would be so called Goldstone bosons, massless bosons that are not observed in nature. However, one can always perform local SU(2) gauge transformations (see Equation (1.10)) such that the $\theta^a(x)$ vanish everywhere, making them unphysical degrees of freedom. With this so called unitary gauge, the first term of (1.25) becomes

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + (v + H)^2 (\frac{g^2}{4} W_\mu^+ W^{-\mu} + \frac{g^2}{8 \cos^2 \theta_W} Z_\mu Z^\mu). \tag{1.29}$$

The terms quadratic in the fields W and Z correspond to mass terms of masses

$$m_W = m_Z \cos\theta_W = \frac{1}{2} v g. \tag{1.30}$$

The photon A_μ remains massless and the $\mathrm{SU}(2)_L \times \mathrm{SU}(1)_Y$ is broken down to the $\mathrm{U}(1)_{\mathrm{em}}$ symmetry.

Fermion masses

Electroweak symmetry breaking can also be used to give mass to fermions, using the same scalar field. It requires the introduction of the conjugate doublet

$$\tilde{\Phi} = \mathrm{i}\tau_2 \Phi^*. \tag{1.31}$$

For the sake of simplicity, this will first be demonstrated for the example of top and bottom quarks t and b, ignoring quark-flavor mixing. A Yukawa coupling of the form

$$\mathcal{L}_Y = -y_\mathrm{b}(\bar{\mathrm{t}}_L, \bar{\mathrm{b}}_L)\Phi \mathrm{b}_R - y_\mathrm{t}(\bar{\mathrm{t}}_L, \bar{\mathrm{b}}_L)\tilde{\Phi}\mathrm{t}_R + \text{ h.c.}, \tag{1.32}$$

where h.c. stands for the Hermitian conjugate, is gauge invariant. After symmetry breaking, in unitary gauge this simply reads

$$\mathcal{L}_Y = -\frac{1}{\sqrt{2}}(v + H)(y_\mathrm{b}\bar{\mathrm{b}}\mathrm{b} + y_\mathrm{t}\bar{\mathrm{t}}\mathrm{t}). \tag{1.33}$$

Comparing the terms to the mass term of the Dirac Lagrangian, one obtains

$$\begin{aligned} m_\mathrm{b} &= y_\mathrm{b} \frac{v}{\sqrt{2}}, \\ m_\mathrm{t} &= y_\mathrm{t} \frac{v}{\sqrt{2}}. \end{aligned} \tag{1.34}$$

CKM matrix

For more than one quark generation, quark doublets and singlets can be combined to

$$\mathrm{Q}_L^{\prime j} = (\mathrm{u}_L^{\prime j}, \mathrm{d}_L^{\prime j})^\top,\ \mathrm{u}_R^{\prime j},\ \mathrm{d}_R^{\prime j} \tag{1.35}$$

where $j \in \{1, 2, 3\}$ enumerates the quark generations and the prime indicates that the quark fields do not need to correspond to the physical fields of definite mass. The gauge-invariant Yukawa Lagrangian for multiple quark generations becomes

$$\mathcal{L}_Y = -\sum_i \sum_j g_{ij} \bar{\mathrm{Q}}_L^{\prime i} \Phi \mathrm{d}_R^{\prime j} - \sum_i \sum_j h_{ij} \bar{\mathrm{Q}}_L^{\prime i} \tilde{\Phi} \mathrm{u}_R^{\prime j} + \text{ h.c.}, \tag{1.36}$$

where g_{ij} and h_{ij} are Hermitian matrices. After symmetry breaking, g_{ij} and h_{ij} generate mass matrices for fermions:

$$M_d^{\prime ij} = \frac{v}{\sqrt{2}} g_{ij}, \tag{1.37}$$

$$M_u^{\prime ij} = \frac{v}{\sqrt{2}} h_{ij}. \tag{1.38}$$

To obtain states of definite mass, the matrices can be diagonalized by two unitary matrices U^{d} and U^{u},

$$M_{\mathrm{d}} = U_L^{\mathrm{d}} M_{\mathrm{d}}' U_R^{\mathrm{d}\dagger}, \tag{1.39}$$

$$M_{\mathrm{u}} = U_L^{\mathrm{d}} M_{\mathrm{u}}' U_R^{\mathrm{d}\dagger}, \tag{1.40}$$

which corresponds to a redefinition of the quark fields to fields with definite mass:

$$\mathrm{u}_{L/R}^l = \sum_k (U_{L/R}^{\mathrm{u}})_{lk} \mathrm{u}_{L/R}^{\prime k}, \tag{1.41}$$

$$\mathrm{d}_{L/R}^l = \sum_k (U_{L/R}^{\mathrm{d}})_{lk} \mathrm{d}_{L/R}^{\prime k}. \tag{1.42}$$

This unitary transformation has no effect on interactions with photons or Z bosons because they do not mix up-type with down-type quarks. However, for interactions with W$^\pm$ bosons the following combination

$$V_{\mathrm{CKM}} = U_L^{\mathrm{u}} U_L^{\dagger\mathrm{d}} \tag{1.43}$$

appears, the CKM matrix. Due to the non-diagonal CKM matrix, a quark of mass eigenstate u^i interacts with the mixture $\sum_j V_{\mathrm{CKM}}^{ij} \mathrm{d}^j$.

1.3.4 Higgs boson

Inserting the vacuum expectation value in the unitary gauge, $v+H$, into the Higgs potential (1.26) yields

$$-\frac{1}{2}(-2\mu^2)H^2 + \frac{1}{4}\mu^2 v^2 (\frac{4}{v^3}H^3 + \frac{1}{v^4}H^4 - 1), \tag{1.44}$$

describing the mass term

$$m_{\mathrm{H}} = \sqrt{-2\mu^2} = \sqrt{2\lambda} v \tag{1.45}$$

of the scalar boson as well as its self interactions coupling with

$$g_{HHH} = \lambda v = \frac{m_{\mathrm{H}}^2}{2v}, \tag{1.46}$$

$$g_{HHHH} = \frac{\lambda}{4} = \frac{m_{\mathrm{H}}^2}{8v^2}, \tag{1.47}$$

and an irrelevant constant term. The coupling g_{HHH} corresponds to a three-Higgs-boson vertex and the g_{HHHH} coupling to one with four. The covariant derivative (1.29) introduces interactions between gauge bosons and Higgs boson that are proportional to the vector boson mass squared, m_V:

$$g_{HVV} = \frac{2m_V^2}{v}, \tag{1.48}$$

$$g_{HHVV} = \frac{2m_V^2}{v^2}. \tag{1.49}$$

Incidentally, this form of the vector boson couplings ensures unitarity in scattering of longitudinally polarized W bosons, an additional motivation for the introduction of the Higgs mechanism.

The coupling to the fermion fields in Equation (1.33) and (1.36) introduces Yukawa interactions

$$g_{Hff} = \frac{y_f}{\sqrt{2}} = \frac{m_f}{v} \tag{1.50}$$

with a strength that is proportional to the fermion's mass m_f.

1.3.5 Quantum chromodynamics

The theory of strong interactions, QCD, can be constructed as a non-abelian gauge theory, in analogy with the weak interaction. The three quark colors are hereby the analogue of the isospin doublets of the electroweak theory and a local SU(3) symmetry is required to hold between them. The Lagrangian of QCD for a single quark flavor reads

$$\mathcal{L}_{\mathrm{QCD}} = \bar{\psi}^i \mathrm{i}\gamma^\mu \partial_\mu \psi^i + g_s \bar{\psi}^i \gamma^\mu t^a_{ij} A^a_\mu \psi^j - m\bar{\psi}^i \psi^i - \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}, \tag{1.51}$$

where t^a are proportional to the eight Gell-Mann matrices, the generators of SU(3) and the analogue to the Pauli-matrices of the SU(2) symmetry. The indices i and j represent the three colors of QCD and $g_S = \sqrt{4\pi\alpha_S}$ the strong coupling constant. The matrices fulfill commutation relations

$$[t^a, t^b] = \mathrm{i} f_{abc} t^c, \tag{1.52}$$

where f_{abc} are the structure constants of SU(3). The dynamics of gluons is described with the field strength tensor

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_S f_{abc} A^b_\mu A^c_\nu. \tag{1.53}$$

The QCD Lagrangian thus contains terms describing interactions of three and four gluons. In contrast to the weak interaction, the SU(3) symmetry is unbroken. Gluons remain massless and quarks of different color are otherwise indistinguishable.

2 Hadron collider physics

The Standard Model can be tested by calculating its predictions and comparing them to experimental observations. In this thesis, this is done at the LHC, a hadron collider. Section 2.1 outlines how observables from such scattering experiments can be related to the physics at high energy and short distance scales. At hadrons colliders, several methods and approximations are necessary to calculate predictions, they are introduced in Section 2.2. Observables constructed specifically for hadron collider experiments are introduced in 2.3. In Section 2.4, the processes that allow studying the Standard Model Higgs boson at the LHC are introduced and the experimental status is discussed.

2.1 Cross sections, branching ratios, and perturbation theory

The physics of short distances can be probed by performing, for example, scattering experiments or studying the decays of short-lived particles. This section will introduce the observables used in that context and their calculation in perturbation theory.

2.1.1 Cross sections and branching ratios

Scattering reactions can be described in terms of cross sections. The cross section $\sigma_{x\to y}$ of a process can be thought of as the effective area incoming particles x have to hit to initiate an interaction creating particles y. The rate of events $\dot{N}$ then depends on the process-specific cross section and the particle flux, described by the luminosity L and determined by the experimental setup:

$$\dot{N} = L\sigma. \tag{2.1}$$

To measure cross sections, events are usually counted after a certain time interval $[t_0, t_1]$, in which the number of events observed is proportional to the integrated luminosity $\mathcal{L}$:

$$N = \int_{t_0}^{t_1} \mathrm{d}t L\sigma = \mathcal{L}\sigma. \tag{2.2}$$

Often not only the rate of events is of interest but also the probability for a certain kinematic configuration Ω of y, e.g. the scattering angles of the outgoing particles. This is described in terms of the differential cross section $\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}$:

$$\frac{\mathrm{d}N}{\mathrm{d}\Omega} = \mathcal{L}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \tag{2.3}$$

Essentially all results in this thesis are extracted by counting events containing a specific final state y, integrated over certain regions of the phase space of kinematic configurations. The initial state x is given by proton-proton collisions and the luminosity L can be independently measured, so that this allows to study differential cross sections.

The probability of a particle x to decay into a final state y per unit time is described in terms of the decay rate $\Gamma_{x\to y}$. Most particles can decay into multiple final states so that the total decay rate is defined as

$$\Gamma_{x\ \text{tot.}} = \sum_i \Gamma_{x\to i}. \tag{2.4}$$

Short-lived particles are observed as resonances, i.e. they are only produced if the energy of incoming particles is close to the rest mass of the particle. The energy dependence is given by a Breit-Wigner distribution of width $\hbar\Gamma$ and thus Γ is also called the width of the particle. Branching ratios

$$\mathcal{B}_{x\to y} = \frac{\Gamma_{x\to y}}{\Gamma_{x\ \text{tot.}}} \tag{2.5}$$

describe the probability of a particle x to decay into a certain final state y.

2.1.2 Calculation of cross sections

For a field theory formulated in terms of a Lagrangian density, the time evolution of arbitrary initial states can, in principle, be calculated. However, while the time evolution can be solved exactly in a free theory, a theory containing interaction terms in the Lagrangian (products of three or more field operators) requires approximate calculations. For scattering reactions these calculations are usually performed in perturbation theory. The basic idea is to use time-independent states of the free theory as a basis of the solution and add the interaction as a small perturbation. The time-evolution of the free-theory states due to the perturbation can be expressed in terms of the interaction-terms in the Lagrangian.

S-matrix, matrix element, and perturbative expansion

In scattering experiments, incoming and outgoing particles can be described as states of definite momenta $|\{\vec{p_i}\}\rangle$ and $|\{\vec{p_f}\}\rangle$ [62]. Their evolution during the interaction can be expressed in terms of a unitary operator S called S-matrix. The probability to observe the final state $|\{\vec{p_f}\}\rangle$ after the scattering is thus

$$p \sim |\langle\{\vec{p_f}\}|S|\{\vec{p_i}\}\rangle|^2. \tag{2.6}$$

The part of the S-matrix that describes non-trivial interactions can, under a few assumptions that are met by typical colliders and detectors, be used to derive an expression for the differential cross section (using natural units with $c = 1 = \hbar$):

$$\begin{aligned} \mathrm{d}\sigma = & \frac{1}{2E_aE_b|v_a - v_b|}(\prod_f \frac{\mathrm{d}^3 p_f}{(2\pi)^3}\frac{1}{2E_f}) \\ & \times|\mathcal{M}(p_a, p_b \to \{p_f\})|^2 \\ & \times(2\pi)^4\delta^{(4)}(p_a + p_b - \sum_f p_f). \end{aligned} \tag{2.7}$$

Here $|v_a - v_b|$ is the relative velocity of incoming particles and E_j the energy of particle j. Moreover, the formula contains the (invariant) matrix element $\mathcal{M}$. The matrix element can be calculated from the Lagrangian and is the only connection to the underlying quantum field theory. The form of Equation (2.7) is also known as Fermi's "golden rule". The cross section is given by the quantum mechanical transition amplitude, encoded in $\mathcal{M}$, and depends on the phase space of the particles exiting the reaction. Furthermore, energy-momentum conservation is required between initial and final state. A similar result can be obtained for the calculation of decay rates.

In perturbation theory, the S-matrix can be calculated as a series containing increasing powers of the interaction terms of the Lagrangian. This then allows to calculate cross sections and other observables as a series in the interaction strength α, which is typically proportional to the square of the coupling between fields:

$$\sigma = \sum_{i=0} \alpha^i \sigma_i. \tag{2.8}$$

The first order for which $\sigma_i \neq 0$ is called leading order. As long as α is sufficiently small, higher order corrections are small and the perturbation series can be stopped after a few terms.

Feynman diagrams

The calculations necessary to calculate matrix elements and with them cross sections and decay rates in this perturbative expansion can be organized in terms of Feynman diagrams. A Feynman diagram corresponds to a mathematical expression that is built of different elements. There are lines, symbolizing particles, and vertices, which correspond to interactions. Incoming and outgoing lines represent initial and final states, lines between vertices symbolize the propagation of virtual particles. For every vertex, a factor proportional to the coupling of the participating fields enters the expression represented by the diagram.

Some of the interaction vertices of the standard model are displayed in Fig. 2.1. The top row corresponds interactions between fermions and gauge bosons. Depending on the time direction, they can represent the annihilation of fermion and antifermion into a boson (left to right), the creation of a fermion-antifermion pair from a boson (right to left), the interaction of a fermion with a boson (top to bottom), or the interaction of an antifermion with a boson (bottom to top). Additional vertices are shown in the bottom row, gluon self interactions, and the interactions between Higgs boson and fermions as well as bosons. They can be interpreted analogous to the fermion-interactions, depending on their orientation.

Calculating the matrix element up to a certain order in the coupling strength α^N in perturbation theory requires to draw and evaluate all diagrams with vertex factors that multiply at most to $\alpha^{N/2}$.

Renormalization, and running coupling

The evaluation of diagrams containing loops involves the calculation of divergent integrals. These divergences can however be regularized and absorbed by a redefinition of fields and constants of the theory, a procedure called renormalization. Renormalization requires the introduction of an apriori arbitrary energy-scale μ_R on which the renormalized parameters depend. For the strong coupling constant α_S these dependence can, at leading order, be written [64]

$$\alpha_S(\mu^2) \approx \frac{1}{b_0 \log(\frac{\mu^2}{\Lambda^2_{\mathrm{QCD}}})}, \tag{2.9}$$

where $b_0 > 0$ and $\Lambda_{\mathrm{QCD}} \approx 200\,\mathrm{GeV}$. The form of this coupling-constant running is responsible for the asymptotic freedom of QCD. At high energy-scales, α_S becomes small

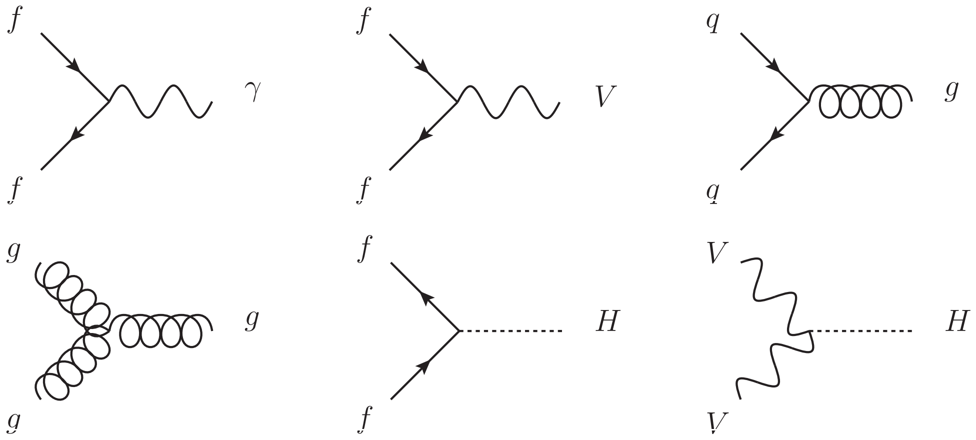


Figure 2.1: Examples of interaction vertices in the Standard Model, drawn with JaxoDraw [63]. Interaction of a charged fermion f and a photon γ, a heavy vector boson V and a fermion, a gluon g and a quark q, three gluons, a fermion and Higgs boson, and a heavy vector boson and a gauge boson. Additional interactions with up to four participating fields are possible, too.

and perturbation theory is possible. Conversely, at low energy scales close to $\Lambda_{\mathrm{QCD}} \approx$ 200 GeV the coupling becomes strong, perturbation theory is no longer valid, and quarks are tightly bound.

Physical quantities like cross sections cannot depend on the arbitrary scale μ_R. However, the truncation of the perturbation series after a few terms introduces a scale dependence of the approximate result. To reduce the influence of higher-order corrections, for a scattering process the scale the α_S has to be evaluated at the energy scale of the process.

2.2 Predictions for hadron collisions

Instead of elementary particles, composite objects, e.g. protons, interact at a hadron collider. Neither the structure of incoming protons nor the formation of outgoing hadrons can be calculated from first principles, which raises the question whether theories like the SM can be tested at a hadron collider. Fortunately, the interactions in hadron-collisions can be factorized into several largely independent parts that can either be calculated or for which parametric models exist, which can be tuned and improved with independent measurements. Their particularities and the approximations necessary to simulate them are described in this section.

The phase-space integration necessary to calculate scattering processes is performed with Monte Carlo (MC) methods because they converge, compared to other integration techniques, relatively fast for high-dimensional integrals [65]. The MC method has the additional advantage that results for arbitrary distributions and phase space cuts can be calculated from phase space vectors that only have to be sampled once. Furthermore, the phase space vectors can be interpreted as simulated events and propagated to additional parts of the event simulation, e.g. the detector simulation.

2.2.1 Hard process

To study, e.g. different models of Higgs interactions or to search for new particles, hard processes are studied, which are characterized by a large invariant mass or large momentum transfer. As they happen at high energy scales, they can be calculated in perturbative QCD. They are connected to the underlying theory as described in the Lagrangian through the matrix element $\mathcal{M}$. To calculate cross sections or generate events, QCD processes are usually calculated in a fixed order of perturbation theory, e.g. leading order (LO) or next-to-leading order (NLO).

However, the description of QCD processes in these perturbative calculations has limitations. Realistic final states consist of hundreds of particles, radiated by the QCD-analogue of bremsstrahlung. Their prediction is not possible in perturbative QCD for a number of reasons. Calculating multi-particle final states increases the complexity of calculations immensely and as the energy scale becomes lower due to the particle radiation, α_S becomes stronger and at some point the perturbative expansion is no longer valid. Moreover, the probability for emission of gluons and for the splitting of gluons into quarks or gluons diverges in the soft and collinear limit [64]. In a collinear splitting, a gluon creates two particles at a low angle and in a soft splitting a gluon is emitted with small transverse momentum to another particle. For the calculation of realistic final states, dedicated heuristics, as they are implemented in parton shower MC programs, are thus necessary. Finally, the scattering protons possess, as composite particles, a substructure that needs to be modeled, too.

2.2.2 Factorization and parton distribution functions

The fact that the hard process happens at very high energy scales and short distance and time scales allows to separate it from the low-energy-scale dynamics inside of the proton in an ansatz motivated by the factorization theorem [66]. In the factorization ansatz, the incoming partons (i.e. the particles that constitute the proton, quarks and gluons) are essentially free, carrying a momentum fraction x of the proton momentum $\vec{p}_i$ [65]. A factorization scale μ_F is defined that separates low-energy dynamics within the proton from the high-energy dynamics of the hard process. The cross section (and differential cross sections) for highly-energetic proton-proton interactions can then be calculated as a convolution of parton distribution functions (PDFs) f_i, which describe the probability to find a parton of flavor i with a momentum fraction x, and the cross section of the hard interaction of free partons $\hat{\sigma}$:

$$\sigma(\vec{p}_1, \vec{p}_2) = \sum_{i,j} \int_0^1 \mathrm{d}x_i \int_0^1 \mathrm{d}x_j \; f_i(x_i, \mu_F^2) f_j(x_j, \mu_F^2) \hat{\sigma}(x_i \vec{p}_1, x_j \vec{p}_2). \tag{2.10}$$

Collinear QCD splittings up the energy scale of μ_F are absorbed in the parton distribution functions. The proton can thus contain, depending on μ_F, large contributions to the proton momentum from quarks and gluons in addition to the three valence quarks.

To reduce the impact of higher-order corrections, μ_F has to be evaluated, similar to μ_R, at the energy scale of the investigated process. Depending on the process, splittings up to different μ_F have thus to be absorbed in the PDF. The evolution of the PDFs to different scales can be calculated using the DGLAP [67–69] equations. Apart from the necessity to evolve them to an appropriate energy scale, PDFs describe universal properties of protons and can be measured in independent experiments. Examples of PDFs at different μ_F are shown in Fig. 2.2. At the higher scale of $\mu_F = 100$ GeV on the right-hand side, the

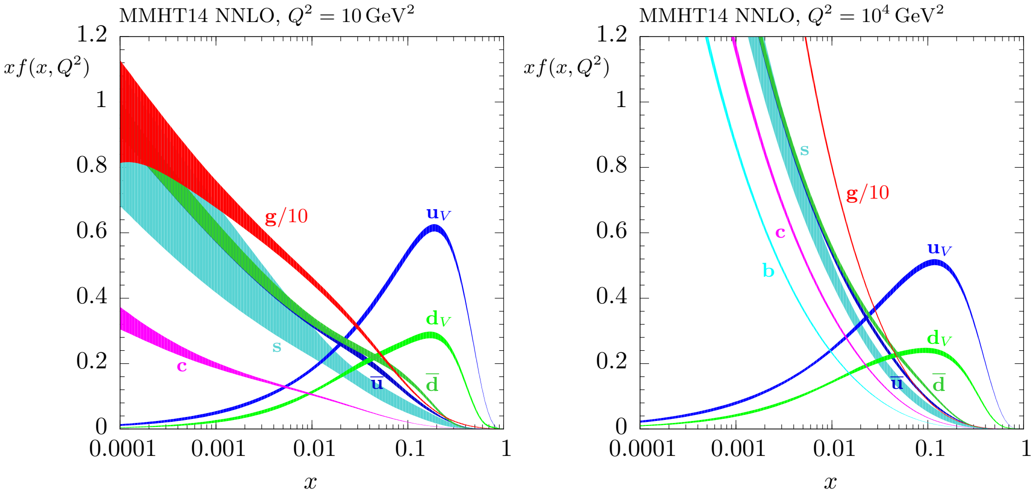


Figure 2.2: Examples of parton distribution function, on the left-hand side at a factorization scale of $\mu_F = 10$ GeV, on the right-hand side at $\mu_F = 100$ GeV. Taken from the MMHT collaboration [70]. Lines show the contributions of different parton flavors, their width the corresponding uncertainty.

fraction of gluons, antiquarks, and heavier quark flavors that carry a significant fraction of the proton momentum increases with respect to the left-hand side because higher-energetic QCD splittings included in the PDF.

2.2.3 Parton shower and hadronization

While the amplitude of QCD splittings diverges in the soft and collinear limit, it has the useful property that it can be factorized from the amplitude of the hard process (see e.g. [64]). This allows simulating such splittings independent of the matrix element calculation as a Markov process. This process is implemented in parton shower MC codes. Parton showers make use of the Sudakov form factor, which describes the probability of not emitting QCD radiation above a certain scale, e.g. above a certain transverse momentum. Making use of the Sudakov form-factor, additional emissions at increasingly lower energies can be added to the final state of a hard process. In a similar way, initial-state radiation is added to the event, with the additional complication that the way initial-state splittings are accounted for in the PDF has to be considered.

The approximations that are made by parton showers work best in the soft and collinear limit. However, they are not able to correctly model hard and large-angle radiation. In that sense, they are complementary to the simulation of particles from matrix element calculation. The benefits of both approaches can be combined with matching algorithms [71]. They define a transverse momentum and angular cutoff above which additional radiation is simulated in the matrix element calculation, thereby avoiding soft and collinear divergences. These events are then showered and clustered with a jet-clustering algorithms with an angular resolution in the order of the angular cutoff. The event is only used if no jets in addition to the ones simulated in the matrix element and above the transverse momentum threshold are generated in the shower. This ensures that only soft and collinear radiation is simulated in the parton shower.

Once the final state, evolved by the parton shower, contains only partons at the energy scale of hadronization, which is of the order Λ_{QCD}, the formation of hadrons is simulated. Hadronization models need to evolve the colored partons into the color singlets that hadrons are made of. For that hadrons are organized into strings (as done in PYTHIA [72]) or clusters (as in HERWIG [73]), which are then broken down to down to hadrons [65]. While these models are motivated by the structure of QCD, they contain a number of free parameters that need to be tuned. After primary hadrons have been formed, they are decayed until the simulated event consists of stable particles.

2.2.4 Underlying event and pileup

The term underlying event is used to describe additional interactions of the two protons initiating the hard process. They are usually simulated as multiple soft parton-parton interactions. However, non-perturbative effects can also play a role and the event is color-connected to the hard interaction, so that it can have a large effect on the hard process, despite involving only low-transverse-momentum phenomena [74]. The simulation of the underlying events requires extensive tuning, using data from electron-positron and hadron colliders [75].

Pileup refers to the effect of proton-proton collisions in addition to the one of the hard process. There can be multiple simultaneous collisions during the same bunch crossing, called in-time pileup. Due to the low response of detectors, also collisions from previous and following bunch-crossings, out-of-time pileup, need to be simulated. Most of these collision are comparatively soft collisions leading to simple QCD processes. At CMS pileup is, in LHC Run 2, simulated by overlaying the hard process with the simulated detector response of events consisting solely of pileup, a method called premixing.

2.2.5 Detector simulation

Hadron collisions are measured with complex detectors, for example the CMS experiment. To be able to compare simulations with the observed data, the interactions of final states with the detector have to be taken into account. This involves interactions with the detector material and the magnetic field, which can lead to scattering or the production of secondary particles, as well as hits and energy deposits in the sensitive material of the detector. In the CMS experiment these interactions are described in a simulation based on the Geant4 toolkit [76]. In the dector simulation, the trajectories of generated particles are traced through the detector, taking into account the particle interactions with the magnetic field, scattering, energy loss, particle decay, and the production of secondary particles, e.g. in bremsstrahlung. Interactions with the sensitive parts of the detector are stored as simulated hits. The final step of the simulation consists of modeling the response of detector readout electronics to the simulated hits. From the simulated response simulated events can then be reconstructed as if they were real data.

2.3 Hadron collider observables

In scattering experiments with hadronic initial states, observables that are invariant under boosts along the beam axis are widely used. For events with hadrons in the final state, jet clustering algorithms play an important role. Both topics are outlined in this section.

2.3.1 Collider kinematics

The kinematics of particles measured in a collider experiment are typically expressed in a coordinate system in which the z-axis is aligned with the beam, a convention that will also be used in this thesis.

At hadron colliders, the two partons initiating the hard interaction can carry very different fractions of the respective proton momenta, creating final states whose center-of-mass is boosted along the z-axis. Thus final states that are similar in their respective center-of-mass frames but have different longitudinal boost can seem very different when described in terms of total momentum or angles between particle trajectories. Thus quantities invariant under longitudinal Lorentz boosts are used preferably.

One example is the transverse momentum

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}, \tag{2.11}$$

which is used to describe how "hard" a jet is. The invariant mass of N particles

$$M^2 = (\sum_{i=1}^{N} E_i)^2/c^4 - (\sum_{i=1}^{N} \vec{p}_i)^2/c^2. \tag{2.12}$$

is a scalar and thus already Lorentz invariant. Because of momentum and energy conservation this observable can be used to calculate the mass of a particle that decayed into multiple particles.

The rapidity of a particle is defined as

$$y = \frac{1}{2} \log\left(\frac{E + cp_z}{E - cp_z}\right). \tag{2.13}$$

Differences in rapidity are invariant under longitudinal Lorentz boost. For massless particles the rapidity is equal to the pseudo rapidity

$$\eta = -\log\left\{\tan\left(\frac{\theta}{2}\right)\right\}, \tag{2.14}$$

where θ measures the angle with respect to the z-axis. Because of the approximate invariance of η-differences under longitudinal boosts, it is often used instead of θ. Instead of calculating angular differences in terms of the solid angle, the measure

$$\Delta R = \sqrt{(\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2} \tag{2.15}$$

is used.

2.3.2 Jets

Quarks or gluons leaving a scattering reaction with high momentum generate collimated streams of particles through the strong interaction, which are called jets. At an experiment, the observed particles (or energy deposits in calorimeter cells) are recombined by jet clustering algorithms to draw conclusions about the kinematic properties of quarks and gluons. Results of calculations in perturbative QCD can also be expressed in terms of jets. Since these calculations are unable to reliably predict the frequency of soft and collinear splittings, it is thus important that the result of a jet-clustering is insensitive to them.

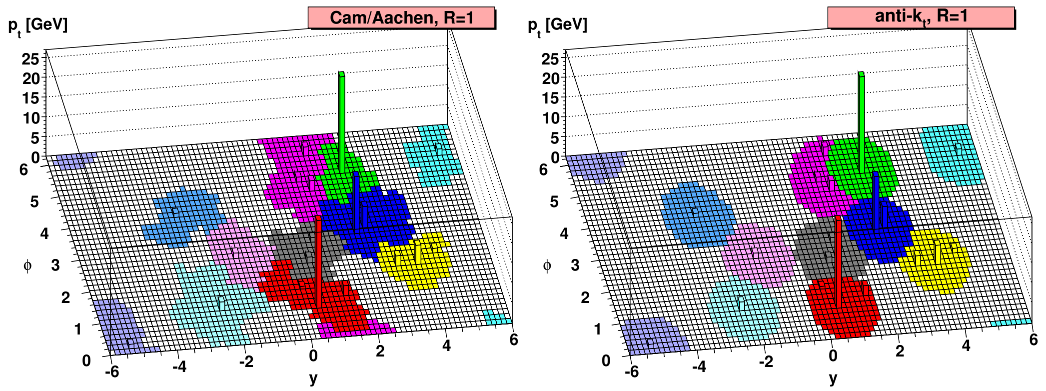


Figure 2.3: Comparison of the C/A (left) and anti-k_T (right) jet clustering algorithms, both with a distance parameter $D = 1$. The energy distribution of particles in the η-ϕ plane is illustrated by the height of bars. In addition to the few hard particles, a homogeneous distribution of soft particles is simulated. Color coding illustrates the area covered by jets. Taken from [79].

Two jet-clustering algorithms will be used in this thesis, the anti-k_T [77] and the Cambridge/Aachen (C/A) algorithm [78]. Both are sequential recombination algorithms and infrared and collinear safe, i.e. not sensitive to soft and collinear splittings [79]. They use a distance measure of

$$d_{ij} = \min(p_{\mathrm{T}i}^{2p}, p_{\mathrm{T}j}^{2p})[(\phi_1 - \phi_2)^2 + (y_1 - y_2)^2] \tag{2.16}$$

between two particles i and j. For the C/A algorithm is $p = 0$ and for the anti-k_T algorithm $p = -1$. The algorithms combine particles to protojets by adding their four-momenta, starting with the two with the smallest d_{ij}, as long as d_{ij} exist that are smaller than

$$d_{iD} = D^2 p_{\mathrm{T}i}^{2p}, \tag{2.17}$$

where D is a parameter determining the jet radius. If the smallest distance is a d_{iD}, the protojet i becomes a jet and does no longer take part in the clustering. The clustering is continued until all particles belong to jets.

Due to the different distance measures in Equation (2.16), the order in which particles are clustered to jets differs between the two algorithms. In anti-k_T jets hard particles or clusters of hard particles are combined with surrounding soft particles first while the C/A algorithm clusters strictly according to the angular distance. The different distance measures also lead to different clustering outcomes (see Fig. 2.3). Jets created with the anti-k_T algorithm are almost circles in the η-ϕ plane, which is beneficial if one wants to estimate experimental corrections, especially in the multi-particle environment of the LHC. The C/A jets have the advantage that the order of the clustering more closely resembles the QCD splitting, in which low-angle branchings are very likely.

2.4 Higgs-boson physics

This section serves as a short introduction to Higgs-boson phenomenology at the LHC.

2.4.1 Higgs-boson production

At the LHC, the 125 GeV/c^2 Higgs boson can be produced in four main production channels, which are described in the following. Exemplary Feynman diagrams for these four

Table 2.1: Main production channels for a 125 GeV/c^2 Higgs boson at 13 TeV center-of-mass energy and their cross section as well as their QCD scale and PDF+α_s uncertainty. [80, 81]

Process	Cross section
Higgs-boson production in gluon-gluon fusion (ggH)	48.6 pb $^{+4.6\%}_{-6.7\%} \pm 3.2\%$
Higgs-boson production in vector boson fusion (VBF)	3.78 pb $^{+0.4\%}_{-0.3\%} \pm 2.1\%$
Associated production of Higgs boson and W boson (WH)	1.37 pb $^{+0.5\%}_{-0.7\%} \pm 1.9\%$
Associated production of Higgs boson and W boson (ZH)	0.88 pb $^{+3.8\%}_{-3.1\%} \pm 1.6\%$
Associated production of Higgs boson and a top quark pair (t$\bar{\text{t}}$H)	0.51 pb $^{+5.8\%}_{-9.2\%} \pm 3.6\%$

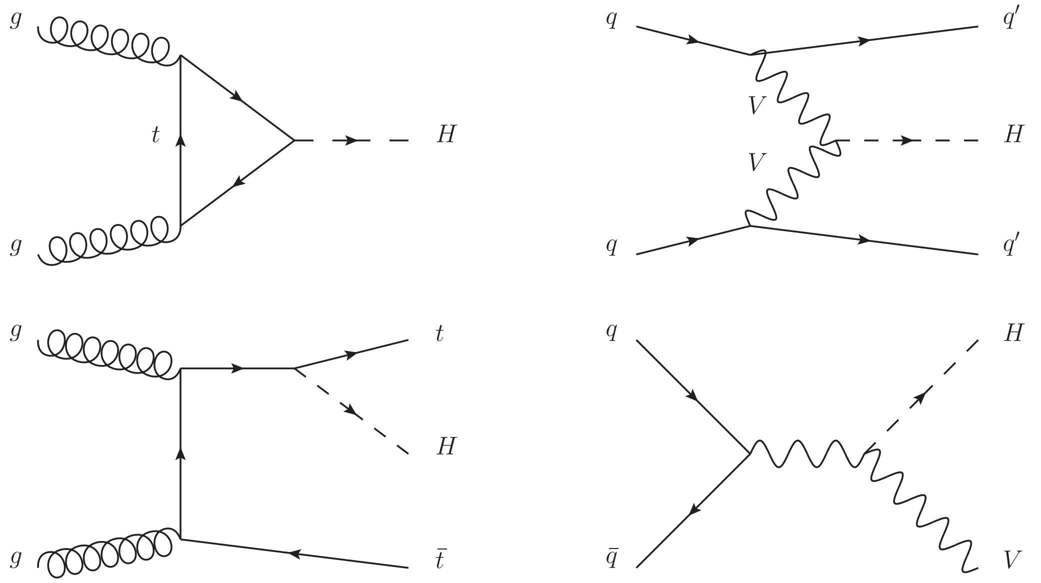


Figure 2.4: Exemplary leading order Feynman diagrams for the four main Higgs-boson production modes at the LHC: Gluon fusion, vector boson fusion, production in association with t$\bar{\text{t}}$, and production in association with a vector boson. Drawn with JaxoDraw [63].

processes are shown in in Fig. 2.4 and their production cross sections at 13 TeV are shown in Table 2.1.

Higgs-boson production in gluon-gluon fusion

The dominant Higgs-boson production mechanism at the LHC is gluon-gluon fusion (also gluon fusion, ggH), in which the Higgs boson is produced with small transverse boost and little associated hard radiation from a gluon initial state. In the SM, this process is enabled by the top-quark and, to lesser extend, the bottom-quark, which strongly couple to both Higgs boson and gluons. The processes can thus be used to indirectly probe the respective Yukawa couplings.

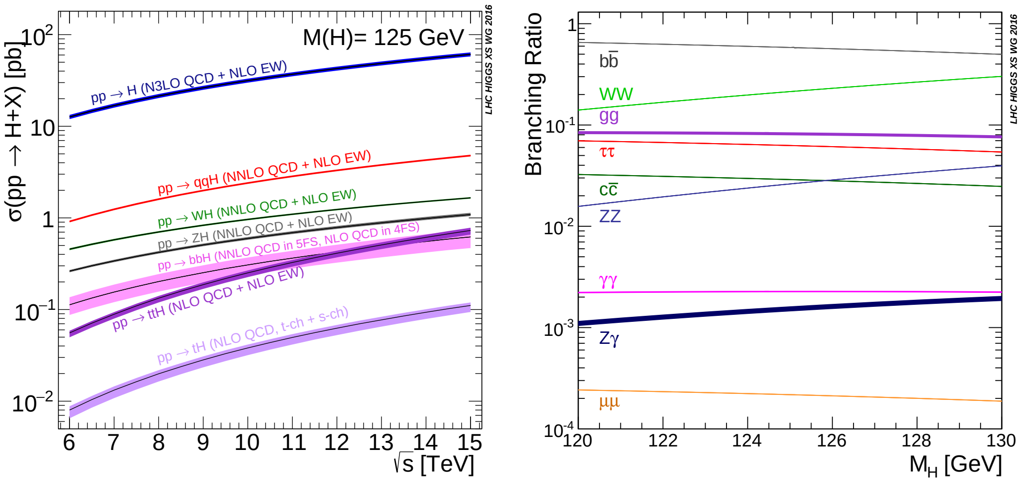


Figure 2.5: Standard model Higgs-boson production cross sections as a function of the proton-proton center-of-mass energy and Higgs-boson branching ratios as a function of the Higgs-boson mass [82].

Higgs-boson production in vector boson fusion

The cross section for vector boson fusion (VBF) production is an order of magnitude smaller than the one for gluon-fusion production. The process is enabled by the Higgs-boson couplings to vector bosons. In VBF, the Higgs boson is accompanied by two jets that have a large gap in η between them and a larger invariant mass, which can be used to tag the events.

Associated production of Higgs boson and W or Z boson

In Higgs-boson production in association with a vector boson (also Higgs strahlung, VH), the Higgs boson is accompanied by either a Z or a W boson, which allows to probe the respective couplings. If the vector bosons decay leptonically, they can help identifying VH candidate events.

Associated production of Higgs boson and a top quark pair

The smallest of the four main Higgs-boson production channels, two orders of magnitude smaller than ggH, is the associated production of Higgs boson and a top quark pair ($t\bar{t}H$). In it, three heavy particles, the Higgs boson and a top quark-antiquark pair, are created and can decay into a multitude of final states. In the SM, the top-quark Yukawa coupling enables this process.

Center-of-mass energy dependence

The center-of-mass-energy dependence of the above production processes is shown in Fig 2.5 on the left-hand side. The $t\bar{t}H$ process profits the most from the increase in center-of-mass energy from LHC Run 1 to Run 2. At 13 TeV, its production cross section is a factor of 3.9 larger than at 8 TeV.

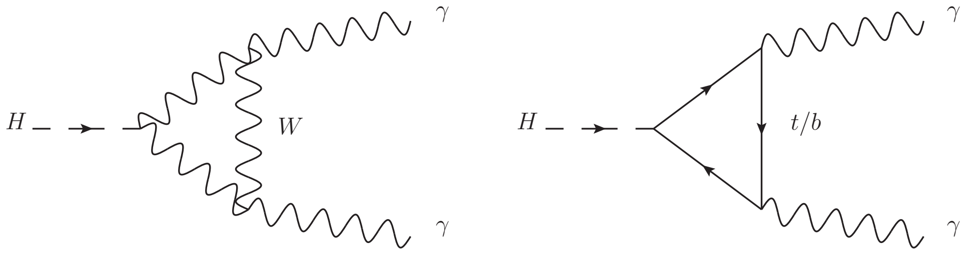


Figure 2.6: Leading order diagrams contributing to the decay of the Higgs boson into photons. Drawn with JaxoDraw [63].

2.4.2 Higgs-boson decays

For a Higgs boson with a mass of 125 GeV/c^2 a large number of decay channels are open. They are depicted on the right-hand side of Fig. 2.5 and outlined in the following.

Vector-boson decays

While the Higgs boson couples strongly to W and Z boson, it is not heavy enough to decay into two on-shell vector bosons (i.e. bosons that fulfill the energy-momentum relation $E^2 = p^2c^2 + m^2c^4$), which suppresses the decay rate for this channel. Nonetheless, the branching ratio is sizable and when the two vector bosons decay into leptons, the final state contains isolated leptons that can be used to distinguish it from background processes.

Fermion decays

Of the Yukawa-coupling induced decays into fermions, the two largest ones into $\mathrm{b\overline{b}}$ pairs (58%) and ττ (6.3%) [80, 81] are the ones that are best-suited for measurements at the LHC. $\mathrm{b\overline{b}}$ decays of the Higgs boson can only be distinguished from the jet backgrounds if they are accompanied by the characteristic final states of VH or $\mathrm{t\bar{t}H}$ production.

The remaining Higgs-boson decays into quarks are impossible to distinguish from the backgrounds. Muon decays have a very small branching ratio and are thus also difficult to detect.

Loop-induced decays

By loop-induced processes, the Higgs boson can decay into the massless gluons and photons. Gluon decays are induced by a top-quark loop but impossible to detect at the LHC. The decay into photons is enabled by the top quark as well as the W boson. Leading order diagrams are shown in Fig. 2.6. In the SM, the amplitudes corresponding to the two diagrams interfere destructively.

2.4.3 Experimental status

In the few years since its discovery, many properties of the Higgs boson have already been investigated. It has been discovered in final states compatible with gluon fusion and VBF production as well as in decays into ZZ, WW, $\mathrm{b\overline{b}}$, and ττ. There is also evidence for VH and (unexpectedly) $\mathrm{t\bar{t}H}$ production [8]. The existence of the Higgs boson is thus clearly established, as well as the fact that it couples to fermions as well as bosons. Measurements

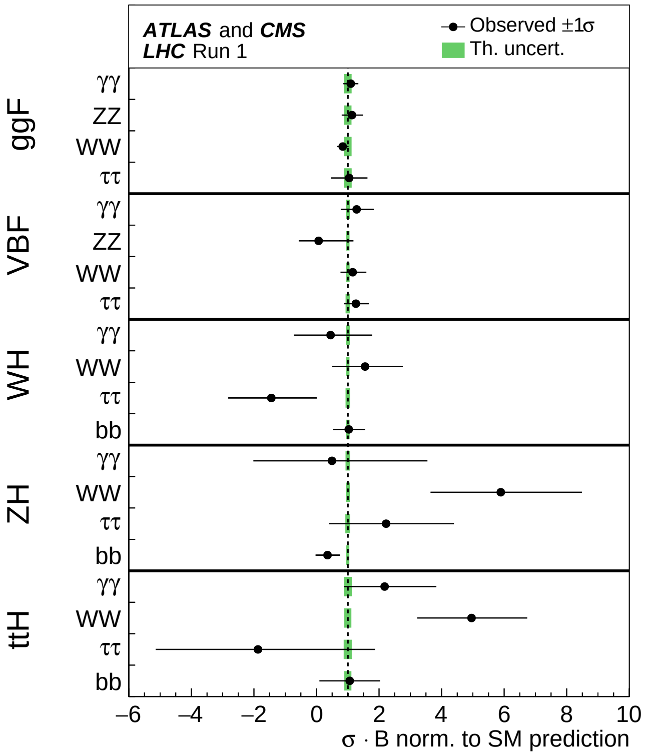


Figure 2.7: Combined measurement of the Higgs-boson signal strength in different combinations of production and decay channels by ATLAS and CMS [8].

of a large number of Higgs-boson production and decay channels by the ATLAS and CMS collaborations are shown in Fig. 2.7.

The mass of the Higgs boson has been measured with high precision [7] as

$$m_{\mathrm{H}} = 125.09 \pm 0.21 \text{ (stat.) } \pm 0.11 \text{ (syst.) } \mathrm{GeV}/c^2,$$

which allows to precisely determine the Higgs self-coupling λ, a parameter that was largely unknown before the Higgs-boson discovery.

The SM predicts that the Higgs boson is a neutral, CP-even, scalar. This hypothesis was tested against many alternatives, e.g. that it is a a pseudoscalar or spin-2 particle, which were all rejected at more than 99% CL [83, 84].

Furthermore, the total width of the particle was measured (in a model-dependent framework) and could be constraint to be below 13 MeV [85, 86], while the standard model prediction for the width is 4.1 MeV. The decays into invisible particles, detected as missing transverse momentum, can also be constrained to be at most 23 % [87, 88]. Searches for Higgs bosons beyond the SM have so far been unsuccessful. These and many more results can be found at [89, 90].

3 Experimental environment

In this thesis, proton-proton collisions, delivered by the LHC and recorded by CMS, are analyzed. This chapter introduces the LHC particle accelerator (Section 3.1) and the CMS experiment (Section 3.2). The CMS subsystems are described in more detail in Section 3.3. In Section 3.4 an overview of the reconstruction algorithms employed by CMS and the typical data analysis workflow is given.

3.1 The LHC

The CERN Large Hadron Collider (LHC) [3] is designed to collide proton beams with a proton-proton center-of-mass energy of 14 TeV and a luminosity of 10^{-34} $\text{cm}^{-2}\text{s}^{-1}$. It is located near Geneva in the circular tunnel built for the LEP collider, which has a circumference of 27 km and lies between 45 and 170 meters underground. At the LHC, several particle physics experiments are situated. Two large multi-purpose experiments study a wide range of physics, ATLAS [1] and CMS [2]. Furthermore, there are two large and specialized detectors, the ALICE experiment [91], studying heavy ion collisions, and the LHCb experiment [92], focusing on the physics of b-flavored hadrons, rare decays, and CP violation. The LHCb experiment requires the LHC to be able to deliver an luminosity that is lower than the high values required by ATLAS and CMS. For the benefit of ALICE, the collider can also be used to accelerate lead ions with an energy of up to 2.8 TeV per nucleon. An overall view of the collider and the experiments is given in Fig. 3.1.

The LHC is designed to achieve a high proton-proton center-of-mass energy and a high luminosity. A high center-of-mass energy increases the cross section for the production of heavy particles, allowing the observation of new phenomena like the Higgs boson. The rate of events is furthermore proportional to the luminosity, which needs to be high to be able to observe rare processes. The LHC is a proton-proton collider because electrons cannot be accelerated to the same energies in a circular collider, as they would lose their energy due to synchrotron radiation. An antiproton-proton collider is also not feasible because antiprotons cannot be produced in the amount necessary to provide high luminosity.

Since the two colliding beams contain particles of the same charge, but are bent in opposite directions, they have to be contained in two separate pipes with opposite magnetic fields. These fields are generated by powerful twin-bore dipole magnets. The beam pipes are required to contain an ultra-high vacuum, below 10^{-10} mbar, to avoid collisions of the accelerated protons with the gas molecules. The momentum of accelerated protons is limited by the requirement that they need to be brought on circular trajectories by the Lorentz force. Since the radius of the LHC is constricted by the LEP tunnel, the maximum beam energy is determined by the strength of magnets. At a beam energy of 7 TeV, dipole magnets are required to provide a magnetic field of more than 8 T. This is only possible by using superconducting technology. The LHC magnets use standard NbTi cables but are cooled by superfluid helium to a temperature of 2 K.

Proton beams are injected into the LHC from a long chain of pre-accelerators, which starts out in the Linac2 linear accelerator. They are further accelerated using the Proton

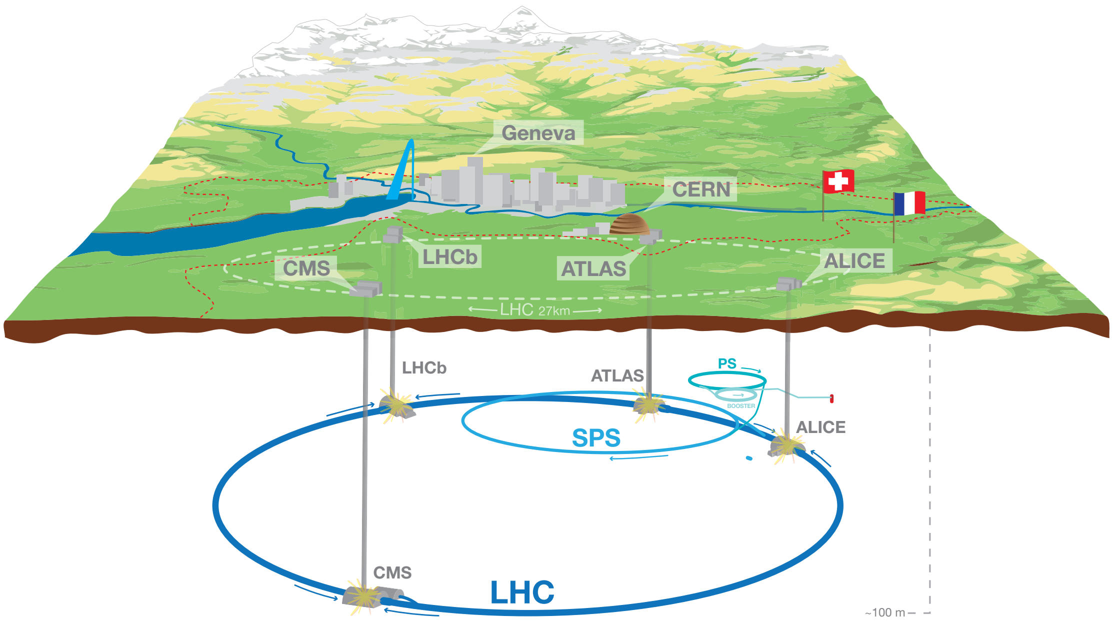


Figure 3.1: Overall view of the LHC and the four main experiments ALICE, ATLAS, CMS, and LHCb [93].

Synchrotron and Proton Synchrotron Booster, and finally reach an energy of 450 GeV in the Super Proton Synchrotron (SPS). In the LHC, the beam is accelerated using superconducting radiofrequency (RF) cavities. In the RF cavities, bunches of protons can be accelerated with resonating electromagnetic waves at a frequency of 400 MHz. Eight cavities per beam deliver 2 MV each with an electric field of 5 MV/m. The acceleration of the beams within the LHC takes about 20 minutes, after which they are collided for up to 12 hours.

The proton beams are organized into bunches with a bunch spacing of 25 ns. Bunches contain approximately 10^{11} protons, limited by the requirement of low beam-beam interactions. Apart from the number of bunches and the number of protons per bunch, the luminosity can be increased by focusing the beams. Focusing and stabilization of the beams is performed by hundreds of quadruple and higher order multipole magnets.

In 2015 and 2016 (Run 2) the LHC operates at a center-of-mass energy of 13 TeV, while in Run 1 (2011 and 2012) the center-of-mass energy was only 7 and 8 TeV. In 2015, the luminosity was below the design value, which was however reached and even exceeded by 50% in 2016.

3.2 Overview of the CMS experiment

The CMS experiment [2] is, together with ATLAS [1], a large multi-purpose experiment at the LHC. It was designed to investigate proton-proton collisions at high center-of-mass energies and luminosities. Collisions of lead ions are also studied but they are not relevant for this thesis and will not be discussed. The CMS detector is located approximately 100 m underground, has a length of 21.6 m, a diameter of 14.6 m, and a mass of 14000 t. It can detect a large range of physics processes. Also owing to the high center-of-mass energy and luminosity of the LHC, most of these processes can be studied with an unprecedented precision. CMS also has the potential to discovery new particles and forces that were out of reach for all previous particle physics experiments. The CMS experiment is carried out in a collaborative effort of 200 institutes from five continents. More than 2000 scientists and students from all over the world are working on improving the detector and analyzing its data.

One of the main motivations to build the CMS detector was to study the mechanism of electroweak symmetry breaking. An intermediate goal was thus already reached with the discovery of the Higgs boson in 2012 [5]. Additional Standard Model physics studied includes electroweak processes, e.g. the production of one or multiple bosons, which is possible with high precision and at high energies. The proton structure and processes enabled by the strong interaction can also be studied at unprecedented energies. Furthermore, a large number of top and bottom quarks are produced at the LHC, allowing precise studies of their properties. All of the above processes can be identified by measuring a limited amount of signatures. In scattering processes involving only the strong interaction, outgoing partons produce jets of hadrons. Electroweak processes, including Higgs production, can be identified from the observation of isolated leptons or photons. In the decays of top quarks and Higgs bosons, bottom quarks are often produced, which can be identified by the presence of heavy hadrons with relatively long lifetimes. Similarly, tau leptons, which are also produced in Higgs boson decays, leave characteristic signatures by their decay into charged particles. Even neutrinos can be observed indirectly, as missing transverse momentum, i.e. a momentum imbalance of the observed particles. As a multi-purpose experiment, the CMS detector has to be able to identify all of these signatures.

The ultimate goal of CMS is to discover physics beyond the Standard Model (BSM). At the LHC, parton interactions with center-of-mass energy of several TeV are possible. Potentially, heavy particles that have never been observed could be produced at these energies. It might be possible to discover supersymmetric particles, new forces, or even microscopic black holes or hints of extra dimensions. BSM particles are expected to either escape the detector without leaving a trace or to decay into particles of the Standard Model. Thus, ultimately, similar particles as in Standard Model measurements have to be identified and measured to discover new physics.

To be able to reach its goals, the following features were required from the detector [2]: It needs to have good muon identification and momentum resolution, enable precise measurements of charged particle momenta, be able to identify tau lepton and b-hadron decays, have good energy resolution for the measurement of electrons and photons, as well as the possibility to reject neutral pions. To enable measurements of jets and missing transverse energy, neutral hadrons have to be measured and the detector needs to be almost hermetic.

All this has to be done in an environment that is experimentally challenging. High bunch crossing rates require a fast detector response and advanced triggers that identify interesting events. At the design luminosity, twenty proton-proton collisions are expected for every bunch crossing. They create a large number of charged particles so that the detector has to be radiation hard. Multiple simultaneous proton-proton collision are furthermore challenging for the identification of particles and for precise energy measurements.

One of the central design features of the CMS detector is the large superconducting solenoid with an inner bore of 6 m. It allows for the placement of the tracker and the calorimeters inside of a 3.8 T strong and almost homogeneous magnetic field. The tracker identifies the paths of charged particles and is based entirely on silicon technology. For the energy measurement of charged particles and photons, a lead-tungstate crystal calorimeter is used that is placed directly outside of the tracking detector. The hadron calorimeter, used for the measurement of strongly interaction particles, is a sampling calorimeter using

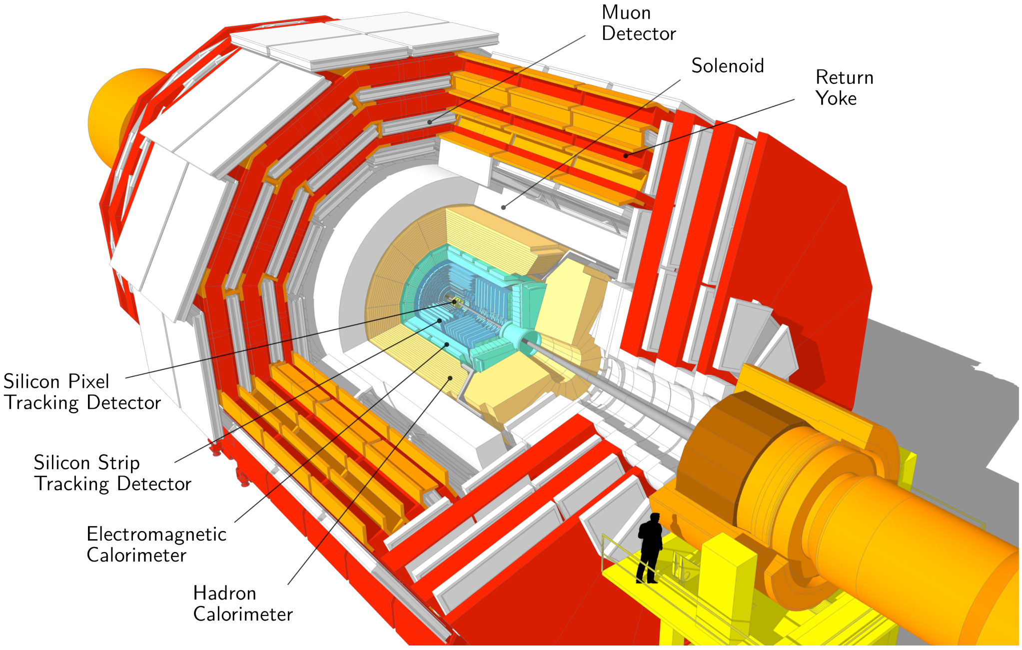


Figure 3.2: Cutaway drawing of the CMS detector, adapted from [94, 95]. The subcomponents that are described in this section are indicated.

brass as the absorber material. The magnetic flux of the solenoid is returned in an iron yoke that also harbors the muon detector. An overview of all subdetectors and their location in the detector is given in Fig. 3.2. They are described in more detail in the following section.

3.3 CMS subdetectors

The description of the subdetectors in this section mostly follows [2]. Information regarding the performance of the subdetectors has been taken from reports that are quoted separately.

3.3.1 Tracker

The CMS tracking detector, or simply tracker, is the part of the detector that is closest to the interaction point. It enables the identification of the path of centrally ($|\eta| < 2.5$) produced charged particles with momenta $\gtrsim 1\ \mathrm{GeV}/c$ with high efficiency and resolution. From the particle trajectories, bent by the magnetic field of the CMS solenoid, the momenta of charged particles can be determined. The reconstruction of tracks also allows identifying primary vertices, i.e. the points of proton-proton collisions, and secondary vertices, which are created in the decay of particles in flight.

At the LHC, more than 1000 charged particles are produced every 25 ns. This requires fast response of the detector as well as fine granularity. Both is possible using silicon detectors. The large flux of charged particles remains a major challenge and introduces radiation damage in the semi-conductors, which is mitigated by using reasonably radiation-hard technology and cooling the detector. Furthermore, keeping a low material budget in the tracker is necessary to minimize particles interactions with the detectors.

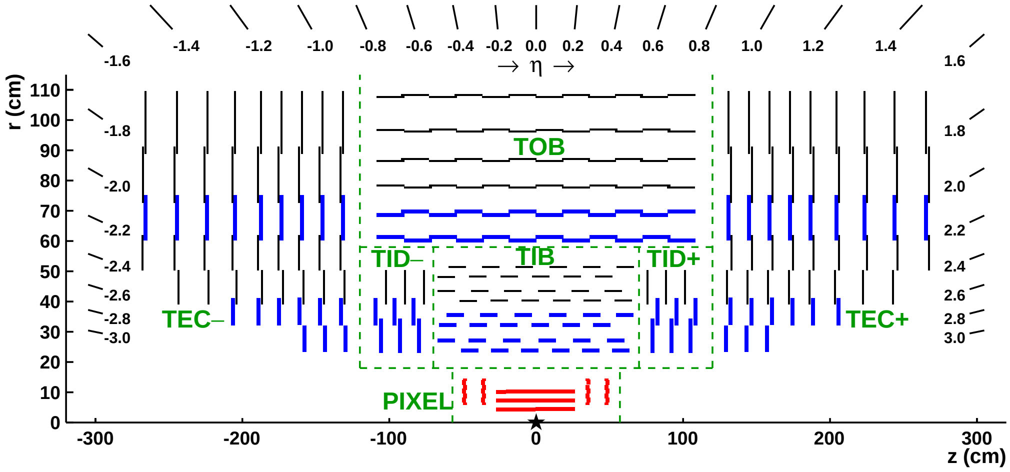


Figure 3.3: Cross section of the top half of the CMS tracker, consisting of pixel detector, tracker inner barrel (TIB), tracker outer barrel (TOB), tracker inner disk (TID), and tracker endcap (TEC). The pixel detector is marked red and strip detector layers allowing for a three-dimensional measurement blue. The nominal collision point is indicated by a star. Taken from [96].

Silicon detectors are, essentially, a collection of pn-junctions in reverse bias. Due to the bias voltage, a depletion zone is created in the semiconductor. Charged particles transversing this zone create electron-hole pairs, creating a measurable current that indicates a hit. These signals are digitized by readout electronics and with sophisticated pattern-recognition algorithms tracks and vertices can be reconstructed.

The total size of the CMS tracker is 5.8 m in length and 2.5 m in diameter. Its layout is sketched in Fig. 3.3. The tracker consists of two parts, a pixel and a strip detector. Both are arranged as cylindrical barrel modules and circular endcaps. The pixel detector is the innermost part of the detector, starting 4.4 cm in radial direction from the nominal interaction point. The high flux of particles and the desired resolution requires high granularity and three-dimensional coordinate measurements, which is achieved by using small pixel sensors to detect transversing particles. The barrel module has three layers at radii of 4.4 cm, 7.3 cm, and 10.2 cm and is complemented by two endcap disks on each side, at ±34.5 cm and ±46.5 cm in longitudinal direction. The sensors are n+ pixels on an n-substrate of a size of 100 μm × 150 μm. The position resolution is improved by charge sharing. Due to the Lorentz drift of electrons within in the strong magnet field, particles transversing the material can be measured in several adjacent pixels. In total, there are 66M pixels on 1440 modules within the detector.

Outside of the pixel detector and extending to 1.2 m in radial direction the strip detector is situated. The lower flux of particles at this radius allows using strip detectors, measuring the trajectories of particles in only two dimensions. The silicon strip tracker consists of different subsystems whose architecture is shown in Fig. 3.3. There is the inner tracker, consisting of barrels (TIB) and disks (TID), the outer barrel part (TOB), and the tracker endcaps (TEC). The barrel parts of the detector have ten and the endcaps a total of twelve layers. In the strip detector, p-on-n type silicon micro-strip sensors are used. The strip pitch is as close as 80 μm in the inner part and increases up to 180 μm for the outer modules. The strips are on modules of a approximate size of 10 cm × 10 cm. Strip detectors allow a two-dimensional measurement of the coordinates of hits but in some modules (marked blue in Fig. 3.3) a second layer of strips is mounted at an angle to the first one, allowing for a three-dimensional measurement. In total, an area of 200 square meters is covered with 9 million strips in 15000 modules.

The CMS tracking detector fulfills the requirements that were identified before its construction. The material budget of the tracker is reasonably low, between 0.4 to 1.8 radiation lengths, depending on the pseudorapidity. Muons can be tracked with an efficiency is of at least 99% for the most relevant η and p_T regions. Due to nuclear interactions the tracking efficiency for charged hadrons is only of the order 85% to 94%, depending on the kinematics [96]. Muons with energies of 100 GeV can be measured with an accuracy of 1% to 3%, where central particles are measured more accurately. At low momentum, for muons below 10 GeV, the resolution is dominated by multiple scattering but usually better than 2% [97]. The transverse and longitudinal impact parameters are measured with an accuracy of 10 μm and 30 μm, respectively, enabling good primary vertex reconstruction and b-tagging.

3.3.2 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is important for the identification and the energy measurement of charged particles, and especially for the detection of photons and electrons. It needs to be fast and radiation hard and be able to operate in a strong magnetic field. Furthermore, it should be compact and have a fine granularity.

These requirements are met by the CMS ECAL. It consists of more than 70000 leadtungstate ($PbWO_4$) crystals, in the shape of truncated pyramids. Owing to the large nuclear charge the material has a radiation length of only 0.89 cm and a Molière radius of 2.2 cm. The calorimeter is located outside of the tracker and is divided into a cylindric barrel that covers $|\eta| < 1.479$ and two endcaps at $1.479 < |\eta| < 3.0$. The ECAL barrel has an inner radius of 1.3 m, the endcaps are 3 m in longitudinal direction from the nominal interaction point. In the barrel, the $PbWO_4$ crystals have a front size of 2.2 cm × 2.2 cm and a length of 23 cm while the size is in the endcaps is 2.86 cm × 2.86 cm × 22 cm. The barrel of the ECAL consists of supermodules, each containing 1700 crystals. The supermodules share parts of the readout electronics and cooling system. In the endcaps, crystals are arranged to supercrystals, mostly consisting of 5 × 5 single modules. These are mounted on semicircular structures called Dees. To improve the rejection of neutral pions, a preshower in installed in front of the crystals in the endcaps. The geometric layout is also illustrated in Fig. 3.4.

Electrons and photons entering the crystals create electromagnetic showers which produce visible scintilation light. The crystals are transparent and the energy of particles can thus be inferred from the brightness of the light. In the barrel, photons are measured using avalanche photodiodes while vacuum phototriodes are used in the the endcaps. The calorimeter is kept at stable operating temperature of nominally 18 °C by a water cooling system.

The preshower detector in front of the ECAL endcaps can be used for the identification of collimated photons from pion decays, which could otherwise be misidentified as single photons. Furthermore, it improves the position measurement and identification of electrons and photons. The detection of particles happens in a sampling calorimeter with a lead absorber and two layers of silicon strip detectors with orthogonal strip orientation.

The energy resolution for 45 GeV electrons is better than 2% is the central region of the calorimeter and better than 5% elsewhere [98]. In general, the relative resolution improves

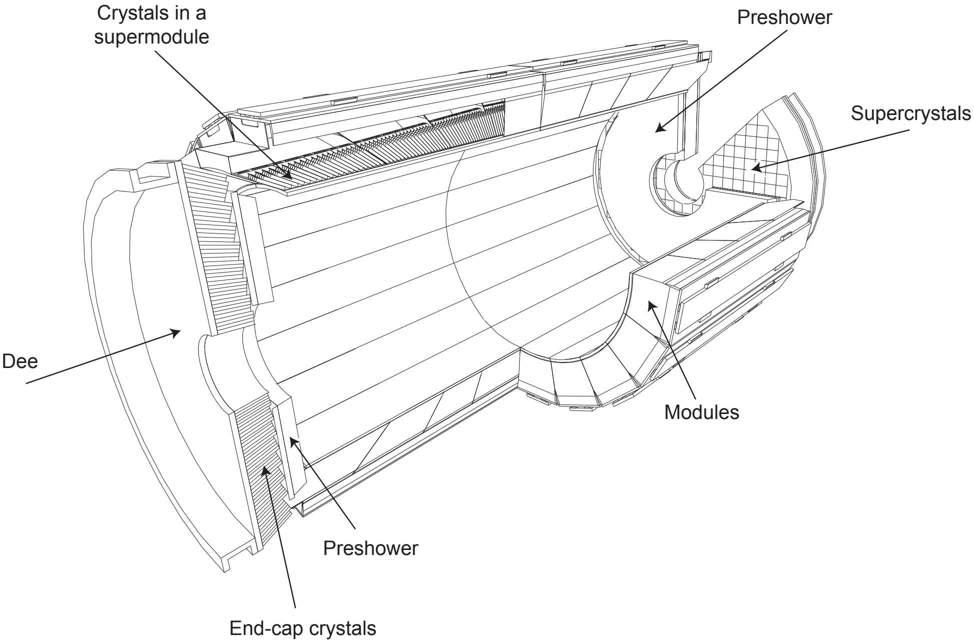


Figure 3.4: Cutaway drawing of the CMS electromagnetic calorimeter [2].

at higher energies. The ECAL has a fast response, 80% of the energy is collected within 25 ns.

3.3.3 Hadron calorimeter

The hadron calorimeter (HCAL) is used for the detection of strongly interacting particles, which are produced at large rate at the LHC. The HCAL allows for an energy measurement of jets and, indirectly, missing transverse momentum. For the latter, it is important that the HCAL is as hermetic as possible. Additionally, the HCAL is used for particle identification as, e.g. electron and photons are stopped before they enter the HCAL. The placement of the calorimeter inside of the solenoid requires it to be very compact and to operate in a strong magnetic field.

The HCAL is a sampling calorimeter. It uses brass plates as absorber material, alternated with tiles of plastic scintillators. Brass is used because it is non-magnetic, has a good stopping power (the nuclear interaction length of brass is 16.4 cm), and is relatively cheap (it is cartridge brass, to a large part of Russian navy shells). The material for the plastic scintillators was chosen to achieve reasonable radiation hardness. Hadrons initiate hadronic showers in the brass that produce light in the scintillators. The light is converted via wavelength shifting fibers and measured with hybrid photodiodes.

The HCAL is located between radii of 1.77 m and 2.95 m. The calorimeter is divided into barrel (HB) part that reaches up to $|\eta| < 1.4$ and a partly overlapping endcap (HE), starting at $|\eta| > 1.3$. The HB (HE) consists of towers of 17 (19) absorber layers. Below $|\eta| < 1.6$ the calorimeter is further segmented into cells of 0.087×0.087 in η and ϕ. For $|\eta| > 1.6$, segments have twice the length in η as well as ϕ. For some towers there is a coarse longitudinal segmentation (see Fig. 3.5), the remaining ones only have one longitudinal readout.

In the barrel, the outer calorimeter (HO) is placed behind the solenoid, to catch the tail

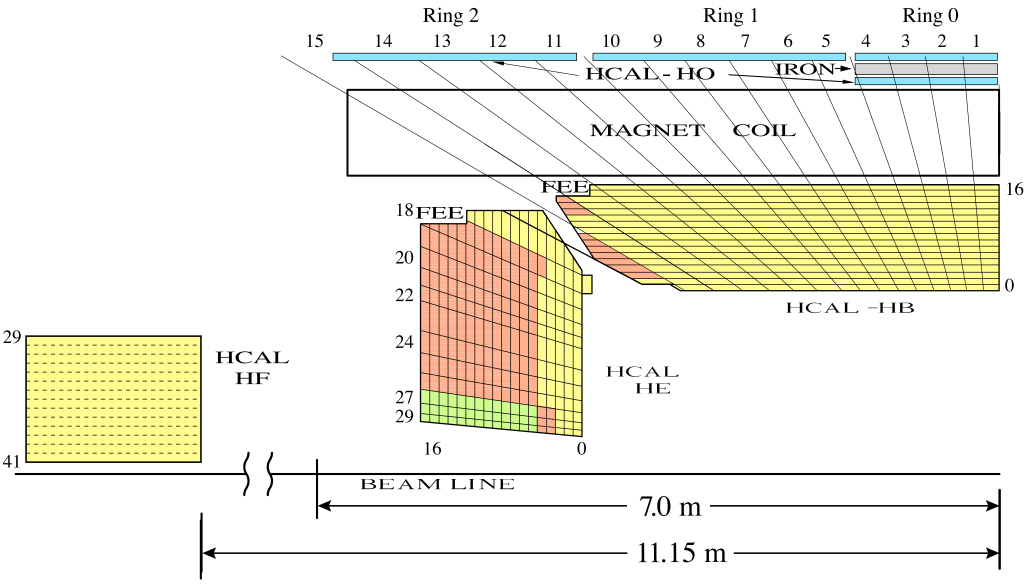


Figure 3.5: Cross section of a quarter of the CMS hadron calorimeter, taken from [99]. Displayed are the barrel, (HB), endcap (HE), outer (HO), and forward (HF) hadron calorimeters. The front end electronics are indicated as FEE. The color indicates independent longitudinal readouts.

of hadronic showers. Additionally, there is a forward calorimeter (HF), in 11.2 m distance from the interaction point, to measure particles at $3 < |\eta| < 5.2$. The forward calorimeter has to be extremely radiation hard, as it is very close to the beam and experiences an extreme particle flux. Here, the detection technique is based on Cherenkov light. Steel absorbers create hadronic showers and the charged particles in that showers can be detected in inserted quartz fibers.

The combined hadronic energy resolution of ECAL and HCAL is approximately [99]

$$\frac{\Delta E}{E} = \frac{85\%}{\sqrt{E(\mathrm{GeV})}} \oplus 7\%.$$

3.3.4 Muon system

The muon detector is primarily important for muon identification. Furthermore, it improves the momentum measurement of high p_T muons and allows triggering events containing them. As the silicon tracker, the muon system is a tracking detector that does not stop particles but identifies their trajectory. Muons hardly lose energy when transversing the detector, a property that is only shared by the undetectable neutrinos. The particle flux through the muon system, outside of the calorimeters and return yokes, is thus much lower than at the center of the detector. At the same time, a large volume has to be instrumented to allow for a robust muon identification and the measurements of the curvature of its trajectory up to $p_\mathrm{T} \gtrsim 1$ TeV/c, which favors the use of large gaseous detectors.

Three different technologies of gaseous detectors are used to track and identify muons in CMS: drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). The measurement principle for all technologies is to create a strong electric field in a cell filled with gas. The gas is ionized at the path of the particle, the ions are accelerated, starting an avalanche of charged particles. The avalanche is then detected as an electric signal.

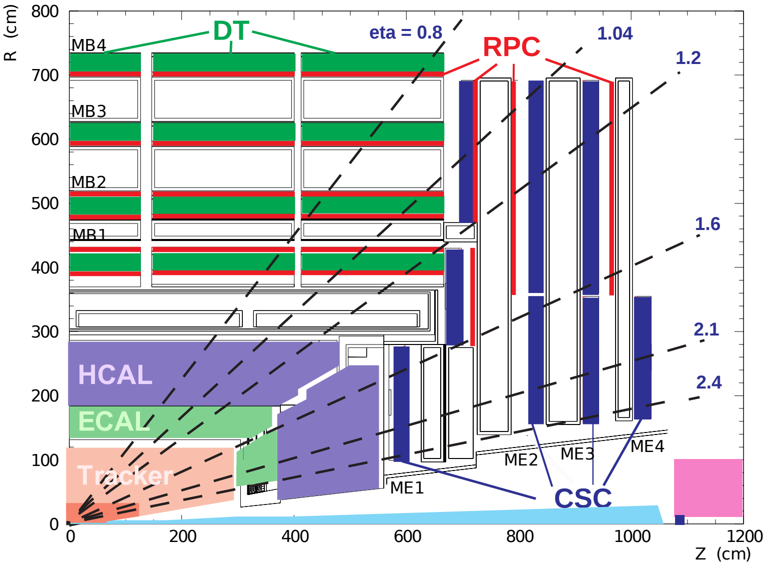


Figure 3.6: Cross section of a quarter of the muon system [100]. Drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC) are indicated.

Like most subdetectors, the detector can be divided into barrel and endcap. In the barrel, drift tube chambers can be employed because the flux density is comparatively low. Four layers of stations are inset between the return yokes (see Fig. 3.6 for an arrangement of the subdetectors). They are filled with a mixture of Ar and CO_2 and contain 172000 sensitive wires. Cells have a length of 2.4 m and a much smaller width, so that the maximum drift path for the ions, 21 mm, can be completed in 380 ns. CSCs are used in the endcaps as they function in the strong and inhomogeneous magnetic field and their shorter response times allow them to cope with higher rates. They are filled with a Ar-CO_2-CF_4 mixture and contain about 2 million wires. CSCs have shorter drift times and cathodes are segmented perpendicular to the wire to allow a three-dimensional position measurement. Both detector types are, up to $|\eta| < 1.6$, complemented by RPCs. Their spatial resolution is only of the order of 1 cm, but they have high electric fields and fast response times and are important for the triggering of events.

The performance goals for the muon detector are met [100]. The spatial resolution is about 100 μm in the CSCs and DTs and 1 cm in RPCs. The efficiency of triggering and identifying muons is, depending on the kinematics, usually above 95%.

3.3.5 Solenoid

The CMS superconducting solenoid provides a strong magnetic field that forces charged particles on bent trajectories and thus enables a momentum measurement. It has a free bore of 6 m diameter and a length of 12.5 m. Its winding is made up of four layers of a NbTi Rutherford cable, reinforced by an aluminum aloy. To allow superconductivity, the solenoid is cooled down to 4 K. It can create a magnetic field of up to 4 T and then stores a magnetic energy of 2.6 GJ. In addition to its exceptional size and strength, the solenoid also comparatively thin. The magnetic flux of the solenoid is returned by a 10 000 t iron yoke.

3.3.6 Trigger and computing infrastructure

In the CMS detector, on average 20 protons collide every 25 ns. This corresponds to a bunch-crossing rate of 40 MHz and a proton collision rate of almost 1 GHz. In 2016 already more than 10^{15} proton-proton collision happened within CMS during more than 10^{14} bunch-crossing events. The data recorded for each event is of the order of MBytes, so obviously not all of these events can be recorded and analyzed. In fact, most collisions happen at a low parton center-of-mass energy. Physics at these energy scales is well-understood, and only a very small fraction of events need to be recorded. The decision which events are recorded is made by the CMS trigger.

The CMS trigger [101] consists of two parts, level one (L1) trigger and high level trigger (HLT). The L1 consists of custom-designed electronics. It uses information from the muon detector and the calorimeters and has to decide within 4 μs whether events are kept. For that it uses primitive objects, e.g. energy deposits of muon candidates or by summing up energies of calorimeter towers. These information can also be combined, to calculate, e.g. the sum of jet transverse momenta. The L1 trigger is designed to reduce the rate of events from 40 MHz to 100 kHz. The HLT is a software trigger, running on a farm of approximately 20k CPU cores. It is based on the offline reconstruction algorithms and further reduces the rate of events to the order of 1 kHz, which are then kept for storage.

The CMS computing infrastructure is part of the Worldwide LHC computing grid [102]. It consists of a large number of computing centers, organized in terms of tiers. There are tier-0 computing centers at CERN and in Budapest, currently 15 tier-1 centers in North America, Europe, and Asia, and 150 tier-2 centers on all five continents. Events are transferred from the HLT to the CMS tier-0 computing center for reconstruction. Reconstructed data is stored there and backups are sent to tier-1 centers. Tier-1 centers are furthermore responsible for storing parts of the reconstructed data, they perform reprocessing tasks and distribute data among tier-1 centers and to and from tier-2 centers. On tier-2 centers, user analyses can be performed and simulated events are generated.

3.4 Event reconstruction and data analysis

Collision events leave signals in the readout electronics of the CMS subdetectors. In this section it is described how they can be combined to the high-level objects like muons, electrons, or jets and how these objects are analyzed by physicists.

3.4.1 Tracks and vertices

The reconstruction of tracks [98] starts as a local reconstruction of hits in the tracker. Hits are clustered from zero-suppressed signals in the silicon sensors. Charge deposits above a certain threshold are combined with adjacent sensors to determine the position of hits as well as the estimated uncertainty. The efficiency to identify transversing particles is well above 99%, limited by the number of defective modules. Hits are combined to tracks in the track reconstruction.

The track reconstruction is performed with the Combinatorial Track Finder (CTF) algorithm, which is based on the Kalman filter [103] algorithm. The CTF algorithm is applied iteratively. First, high-p_T tracks from the interaction vertex with at least three pixel hits are reconstructed. These hits are then removed so that the reconstruction of additional tracks with increasingly worse quality can be performed with simplified combinatorics.

The CTF algorithm starts out by generating seeds from a few hits, for a first estimate of the helical trajectory. For this estimate at least three points are necessary and for most tracks three hits in the layers of the pixel detector can be used. Otherwise beamspot-constraints and hits in the strip tracker have to be considered. Additional hits along the trajectory are then identified with a Kalman-filter based algorithm, in the track-finding step. The track is updated using the information provided by each additional layer, considering the hit and its uncertainty as well as the possibility for multiple scattering and energy loss. For every layer, the track is propagated through multiple hits, increasing the number of track candidates. Ambiguities are resolved by removing track candidates with fewer hits and worse compatibility with the hits and their uncertainty.

The estimates of the track parameter of the track-finding step are combined in a fit using a Kalman filter and smoother. Fake tracks reconstructed by track-finding and fitting procedure are rejected using a number of quality criteria, e.g. requiring a minimum number of layers crossed, a maximum of layers skipped, and a maximum χ^2 of the track fit. The efficiency to reconstruct charged particles with $p_\mathrm{T} \gtrsim 1$ GeV/c as tracks using this algorithm is of the order 90%, with low fake rates and allowing for measurements of 100 GeV/c muons with better than 3% accuracy.

Primary vertices correspond to the points of inelastic proton-proton interactions. They are reconstructed in a three-strep approach. First, tracks with low transverse impact parameters, a minimum number of hits, and a low χ^2 are selected. The tracks are then clustered according to their z-coordinate. The clustering is performed using a deterministic annealing [104] algorithm. A free-energy analogue is constructed containing a χ^2-like expression that considers all possible assignments of tracks to vertices weighted by their probability. As the temperature equivalent of the algorithm is reduced, an increasing number of vertices are identified until a tuned temperature minimum is reached. The set of vertices found is then fit with an adaptive vertex fitter [105]. In the fit, each track is assigned a weight w_i, corresponding to the probability to belong to the vertex. The quality of a vertex can then be measured by its degrees of freedom, defined as $n_\mathrm{dof} = -3 + 2\sum_i^{N_\mathrm{tracks}} w_i$. The position resolution of the vertex fit is of the order 10 μm in all spatial directions.

3.4.2 Muons

The muon reconstruction starts by independently reconstructing tracks in tracker and muon chamber [100]. Two types of muon reconstruction are used at CMS. “Global muons” are first reconstructed in the muon chamber and their track is extrapolated to a matching track in the tracker. Using hits in both subdetectors, a global muon fit is performed, using the Kalman filter approach. The information from the muon chambers improves the muon resolution at large p_T. The second muon-reconstruction approach consists of extrapolating tracker tracks to the muon chambers, considering multiple scattering in the detector, and matching it to at least one muon station. This “tracker muon” reconstruction is more efficient than the global muon reconstruction at low muon momenta.

Muons reconstructed by the two approaches are combined if they share a track. They are then selected applying different criteria. For the particle-flow event reconstruction, which is used in this thesis, quality criteria are applied that also depend on the environment of the muon, e.g. whether it is within a jet or not.

3.4.3 Electrons

Electrons are reconstructed from energy deposits in the ECAL and a matching track [106]. Due to bremsstrahlung, energy deposits of electrons are spread out in ϕ-direction in the ECAL and need to be clustered. The clustering starts from seeds with large energy deposits, which are combined with energy deposits in neighboring calorimeter cells. These energy cluster are then combined to super clusters (SC). The SC energy is the sum of the individual cluster energies, its position is determined by an energy-weighed mean. For the particle-flow reconstruction the clustering is slightly modified to be able to identify individual bremsstrahlung photons, which is achieved by combining significant energy deposits around single seeds to "PF clusters", thereby also allowing for sharing of the crystal energy between multiple clusters.

The electron track is reconstructed using a complex algorithm that takes the changes in curvature due to bremsstrahlung energy losses into account. Because the electron reconstruction is computationally expensive, it can only be performed for selected electron seeds. Seeds are identified by either extrapolating ECAL clusters to track seeds or by extrapolating tracks to the ECAL using the default track finding algorithm. From the seeds the electron track is identified using a modified combinatorial Kalman filter method. Energy loss is considered and the association of tracks is less restrictive than in the standard track fit. In return, the track reconstruction is more restrictive with respect to missing hits, to reject converted photons.

The reconstructed track is then fit using a Gaussian sum filter (GSF). In a GSF, energy loss in tracker layers is modeled by Gaussian distributions. The fit creates a track from the beam spot to the ECAL surface and allows estimating the energy loss due to bremsstrahlung from the curvature difference at the two points. For the PF reconstruction the GSF track is linked to PF clusters in the ECAL, which allows recovering most bremsstrahlung photons.

Finally, electron tracks and ECAL clusters are linked based on their estimates of the ϕ and η position in the ECAL. Electrons can emit bremsstrahlung photons that create electron-positron pairs. Track ambiguities created by such additional electrons are resolved by preferring the tracks of better quality, e.g. the ones with fewer missing hits. The momentum of electrons is measured by combining track and ECAL information with algorithms depending on the bremsstrahlung pattern of the electron. The pattern is determined by considering the bremsstrahlung estimates from PF clusters and the GSF fit.

3.4.4 Particle-flow event reconstruction

The particle-flow (PF) event reconstruction [107] tries to identify the type of all stable particles for an optimal combination of the subdetector information for the reconstruction of their momenta. The PF algorithm identifies five types of particles: charged and neutral hadrons, electrons, muons, and photons. These so called PF candidates can then be used to cluster jets or to calculate the missing transverse momentum in an event. With respect to calorimeter jets, PF jets have the advantage that their constituents, which are usually comparatively soft, benefit from precise charged particle measurements of low-momentum particles in the tracker.

The fundamental elements of the PF reconstruction are tracks, reconstructed as described in Section 3.4.1, and calorimeter cluster. PF ECAL cluster have been discussed in Section 3.4.3, energy deposits in the HCAL are clustered analogous: Seeds are identi-

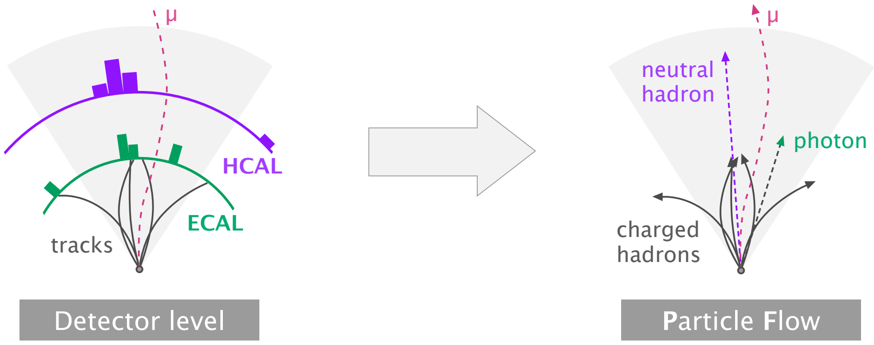


Figure 3.7: Illustration of the particle-flow event reconstruction, taken from [108]. On the left-hand side the detector response for a simplified event is sketched. Tracks are symbolized by black line except for the muon track which is measured in the tracker and the muon chambers and indicated by a dashed red line. Energy deposits in ECAL (HCAL) cells are green (purple). On the right-hand side the interpretation of these objects by the particle-flow algorithm is shown.

fied as local maxima of the energy deposited in the calorimeter. With energy deposits in neighboring calorimeter cells that are two standard deviations over detector noise, they are combined to topological clusters. A topological cluster can contain multiple seeds that share the total cluster energy.

Calorimeter clusters and tracks in tracker and muon detector are linked to blocks of elements. Tracks are extrapolated through pre-shower detector (PS), ECAL, and HCAL. Bremsstrahlung clusters are linked to the tracks. Clusters of the granular PS are linked to the less granular ECAL clusters and ECAL cluster to even less granular HCAL clusters. Tracks in tracker and muon system are combined to global muons, as described in Section 3.4.2. Even for complex elements, the linking results in independent blocks of only a small number of elements.

The particle-flow algorithm then identifies particles from each block. Global muons whose combined momentum measurement is compatible with the tracker measurement become PF muons. Their track is removed from the block as well as their small expected energy deposit in the calorimeters. After muons, electrons are reconstructed with the algorithms described in Section 3.4.3. Their tracks are also removed from the block, together with ECAL deposits compatible with the electrons and bremsstrahlung photons. Of the remaining tracks only those are kept that are expected to provide a more accurate energy measurement than the calorimeters. The calorimeter clusters are therefore recalibrated to allow for a better energy measurement of hadrons. Tracks are linked to ECAL and HCAL clusters and can give rise to charged hadrons. If track momenta are significantly higher than the observed energy deposits, an additional search for muons is performed with looser quality criteria for the tracks. If the energy determined from the track momentum is compatible with the calorimeter energy deposits, the charged hadron energy is determined from a combination of track and calorimeter measurement. In the case that the energy in the calorimeter is higher than what is expected from the tracks or energy clusters exist that are not linked to tracks, this energy deposits give rise to neutral hadrons and photons. The PF algorithm is also illustrated in Fig. 3.7 for a simplified example.

3.4.5 Jets and missing transverse momentum

Most analyses at CMS use jets clustered from particle flow candidates. Usually, candidates from a primary vertex other than the main interaction vertex are excluded from the jet clustering. This is called "charged hadron subtraction" (CHS) and described, e.g. in [109]. The default jet-algorithm at CMS is the anti-k_T [77] algorithm as implemented in FastJet [110].

Even after CHS, particles originating from pileup are clustered to jets, either because they are neutral or because they could not be associated with a pileup vertex. Furthermore, the transverse-momentum measurement of jets is dependent on η and non-linearly dependent on p_T and so the jet-energy has to be corrected. This correction aims to calculate a momentum vector for jets that is as close as possible to the energy of the quark or gluon it originated in. However, due to QCD confinement the concept of a free quark escaping a scattering with a fixed momentum is not well-defined. To obtain quantitative results from simulated events, the measured jet-momentum is thus compared to particle level jets instead. In simulated events, particle level jets are clustered from all stable particles ($c\tau > 1$ cm), except for neutrinos, and then matched to simulated reconstructed (PF) jets. The jet response is then defined as the measured jet-energy divided by the particle level jet energy and the goal of jet energy corrections is to bring it close to unity for jets of different kinematics measured in different environments.

Jet energy corrections follow a factorized approach described in [111]. First, the p_T offset due to pileup is mitigated with the L1 corrections. These corrections are derived from simulated QCD dijet events. They are a function of η, p_T, and the area of the jet as well as the average energy density ρ in the event and are, in simulations, able to make the average jet response independent of the number of pileup interactions.

After the L1 correction, the jet response is however still dependent on p_T and η of the jet. This behavior originates in the non-linear energy measurement of the calorimeters and the geometry of the detector. It is corrected with a p_T- and η-dependent function that ensures that the average jet response is close to one for all jet energies and detector regions. For individual jets, the jet response can of course deviate from unity due to the limited jet energy resolution. The jet response and resolution expected from the CMS detector simulation is displayed in Fig. 3.8.

Missing transverse momentum (usually called missing transverse energy, MET, or $\not{E}_\mathrm{T}$) is defined as

$$\vec{\not{E}}_\mathrm{T} = -\sum_i \vec{p}_{\mathrm{T}i}, \tag{3.1}$$

where the sum includes all particle flow candidates. If all particles were measured perfectly, the sum would correspond to the transverse momentum of the initial state, which is zero. However, neutrinos (and hypothetical BSM particles) can escape the detector unobserved. In this case, $\not{E}_\mathrm{T}$ can be identified with the transverse momentum of the escaped particles. Because the calculation of $\not{E}_\mathrm{T}$ includes the transverse momenta of all particles, its resolution is rather low.

Large contributions to the sum in Equation 3.1 are due to the particles in jets. The resolution can be improved when considering the jet energy corrections in the calculation of $\not{E}_\mathrm{T}$. This method is called type 1 $\not{E}_\mathrm{T}$ correction.

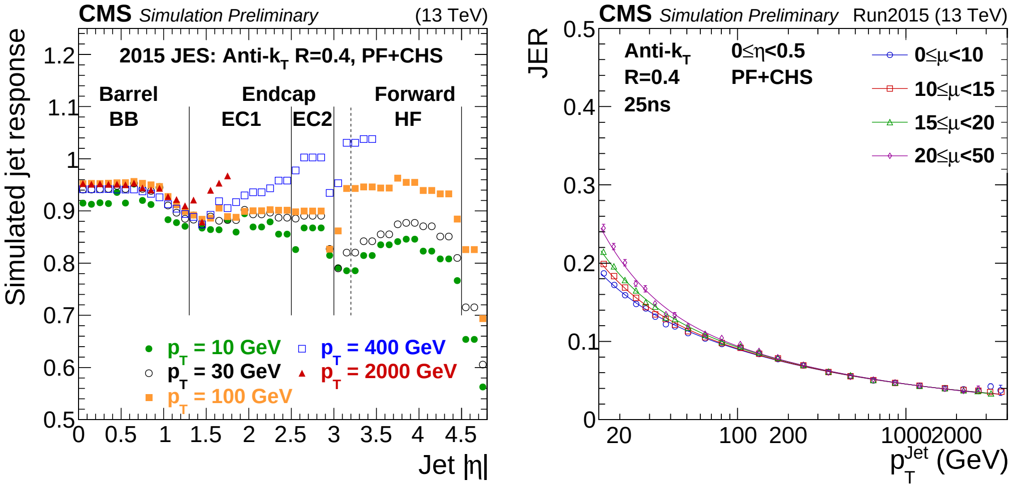


Figure 3.8: Jet response (left) and jet energy resolution (JER, right) in the simulation of the CMS detector used for the analysis in Part III of this thesis. The response is shown as a function of p_T and η, the jet energy resolution is shown as a function of jet p_T for different pileup and central ($|\eta| < 0.5$) jets. Taken from [112].

3.4.6 b-tagging

Jets originating from b quarks can be identified by the presence of b-flavored hadrons (b-hadrons), i.e. hadrons containing a valence b quark. These hadrons have a number of characteristic features. With a rest mass of more than 5 GeV/c^2, b-hadrons are the heaviest hadrons. They decay via the weak interaction into hadrons containing c-quarks, a decay that is suppressed by a small CKM matrix element, which results in a long lifetime of $\tau = 0.5$ mm/c. Their decay products usually include multiple charged leptons and in about 20% of the decays an electron or muon. Furthermore, in the fragmentation of b quarks most of the energy is passed on to the b-hadron, so that they carry a large fraction of the total jet momentum. They can thus have a large Lorentz boost and increased lifetimes of several mm/c. The displaced decay of the hadron gives rise to a secondary vertex that has a significant distance to the primary vertex. Tracks originating from charged particles produced in this decay tend to have large impact parameters d_0, which is the distance of closest approach of the extrapolated track to the primary vertex. A sketch of a b-hadron decay is shown in Fig. 3.9.

Jets originating in b quarks are identified by algorithms called b-taggers. The algorithms identify the characteristic features, the large mass, long lifetime, hard fragmentation, and possibly semi-leptonic decays. Apart from jets, the inputs to b-taggers are tracks, secondary vertices, the primary vertex, and electron or muon candidates. The first step of b-tagging algorithm is to associate tracks with jets. In LHC Run 2, “explicit jet track association” is used at CMS, i.e. only tracks that can be linked to particle-flow candidates in a jet are used to b-tag that jet. Additionally, tracks need to fulfill certain quality criteria, e.g. $p_T > 1$ GeV/c, a track-fit $\chi^2 < 5$, and a large number of tracker hits. Furthermore, several cuts on transverse and longitudinal impact parameters are applied to reduce the influence of tracks from pile-up [113].

Two vertex reconstruction algorithm are used in CMS, the adaptive vertex reconstruction algorithm (AVR) and the inclusive vertex finder (IVF). The latter, introduced in [114], is the default in Run 2. Input to the algorithm are all tracks in the event, with a track

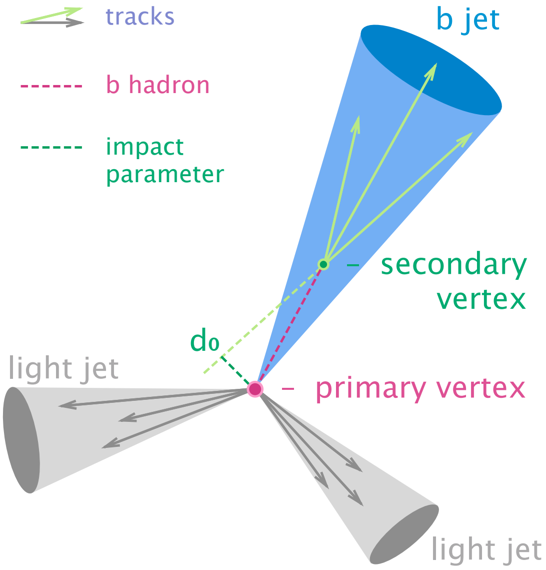


Figure 3.9: Sketch of the decay of a b-hadron in a b-jet, recoiling against two light jets. Taken from [108].

selection that is looser than the default b-tag track selection. For the vertex finding, seed tracks with high impact parameters and impact parameter significances are identified. Nearby tracks are then clustered to the seeds and fitted with the adaptive vertex finder introduced in 3.4.1. After the fit, secondary vertices can share tracks with the primary vertex or other secondary vertices. These ambiguities are resolved and vertices are refitted. Finally, only secondary vertices with significant flight distances, $\Delta R < 0.3$ to jets, masses below 6.5 GeV/c^2 and not compatible with K^0_S decays, and not too many tracks shared with the primary vertices are kept.

The most widely used b-tagging algorithm at CMS in LHC Run 2 is the CSVv2 b-tagger [113], which is a further development of the CSV algorithm which was the Run 1 default [115]. Tracks are selected as described above and furthermore need to be closer than $\Delta R = 0.3$ to the jet and must not be compatible with a K^0_S. The CSVv2 output is a number between 0 and 1 if the algorithm could be run and a negative value if less than two tracks are selected. Higher values correspond to a higher b-jet probability. The output is calculated with the help of artificial neural networks (ANN). In the CSVv2 algorithm, a b-jet candidate can be within one of three vertex categories: vertex, pseudo-vertex, or no vertex. In every category, an ANN distinguishes b-jets from c-jets and light jets. The input variables of the ANNs depend on the category. If a vertex is found, observables related to it are used in the training, e.g. the secondary vertex mass, the flight-distance significance, or its energy compared to the jet-energy. Additionally, track-based variables like the impact parameters of tracks are used. For jets in which no vertex could be reconstructed, a different ANN is trained using only these low-level quantities. Sometimes tracks with large impact parameter significances can be combined to form a pseudo vertex. For jets falling in this category some of the vertex-based quantities can be used in the training. The final output of the CSVv2 algorithm is a likelihood ratio that combines the ANN outputs, considering the expected signal-to-background ratios in the categories. Jets are b-tagged if the b-tagger output is above a certain threshold.

3.4.7 Data analysis workflow

Events are reconstructed at the tier-0 centers and stored in a format called RECO, which already contains most of the high-level objects used in analyses, together with low-level information. Because of the low-level information, it is however too large to be efficiently used for data analysis. For this purpose the Analysis Object Data (AOD) format is used. It contains all high-level information that allow performing almost all physics analyses and is about 1 MB per event. In Run 2, a new format called MINIAOD [116] was introduced that only has a size of the order of 50 kB per event. It contains the most important high-level objects in full precision as well as particle flow candidates and selected generated particles in a compressed format and can be used by most analyses. In contrast to AOD samples, which are only updated with recalibrations once a year, the MINIAOD samples are updated more regularly and contain more recent corrections and updates.

The steps performed after the creation of MINIAOD are highly analysis specific. Many patterns are however shared between most analyses and in this sense the workflow used to obtain the results of Part III of this thesis can be seen as representative for a typical Run 2 CMS workflow. For this analysis, MINIAOD datasets are analyzed using the LHC grid-computing infrastructure. This is done for both real data and several MC samples. Only events fulfilling loose preselection criteria, e.g. a minimum number of jets and a lepton are kept. Some computational expensive high-level information that is not contained in MINIAOD data is also computed in this step. Events passing the preselection are then stored and further analyzed on local computer clusters. With analysis software that was written specifically for this thesis and that is based on the CMSSW software framework (see e.g. [117] for a description), the variables used in this analysis are stored event-wise as simple numbers or arrays of numbers in ROOT [118] N-tuples. ROOT allows a fast evaluation of rows and columns of this data as well as an easy representation in terms of graphs and histograms. Furthermore, it offers a number of data analysis tools widely used in particle physics. For these reasons, it is also used for the data analysis and for the creation of all figures displayed in this thesis, unless stated otherwise.

4 Statistical methods

This chapter introduces the statistical concepts, theorems, and tools that are used in this thesis. In Section 4.1 the concept of likelihoods is introduced and it is explained how it is used to construct discriminants and to estimate confidence intervals. The statistical model and tools typically used in Higgs-boson measurements at CMS are explained in Section 4.2. Section 4.3 introduces the Matrix Element Method and Boosted Decision Trees, two multivariate classifiers used in Part III of this thesis.

4.1 Maximum likelihood estimates, likelihood ratios, and confidence intervals

Let $f(\mathbf{x}|\boldsymbol{\theta})$ be a probability density function parameterized by parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_N)$ that describes the probability of one or multiple observations $\mathbf{x}$. For a given measurement of $\mathbf{x}$, the likelihood function is then defined as

$$L(\boldsymbol{\theta}|\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta}). \tag{4.1}$$

Likelihoods are a useful tool to estimate parameters, perform hypothesis tests, and to calculate confidence intervals.

4.1.1 Maximum likelihood estimates

If the functional form of f is known, but the values of the parameters $\boldsymbol{\theta}$ are unknown, a maximum likelihood fit can be performed to estimate them. In a maximum likelihood fit, the maximum likelihood estimates (or best-fit values) $\hat{\boldsymbol{\theta}}$ of the parameters are calculated by maximizing L with respect to $\boldsymbol{\theta}$. For numerical reasons, usually the negative log-likelihood

$$\mathrm{NLL} = -\log L(\boldsymbol{\theta}|\mathbf{x}) \tag{4.2}$$

is minimized. Maximum likelihood estimates have many useful statistical properties (see, e.g. [119]).

4.1.2 Likelihood ratio and Wilks' theorem

If a hypothesis is tested against one or multiple alternative hypothesis, the most powerful test is, according to the Neyman-Pearson lemma [120], the likelihood ratio test

$$\Lambda(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\theta_1|\mathbf{x})}. \tag{4.3}$$

Here θ_0 represents the null-hypothesis and θ_1 represents alternative hypotheses parameterized (or enumerated, in the case of discrete hypotheses) by θ_1.

For many measurements in this thesis, the difference between null and alternative hypotheses is parameterized by a limited number of parameters of interest (POI) $\boldsymbol{\kappa} =$

Table 4.1: Quantiles of the one and two-dimensional χ^2 distribution for different confidence levels (CL).

CL	1D	2D
0.683	1.00	2.30
0.950	3.84	5.99
0.990	6.63	9.21

$(\kappa_1, ..., \kappa_N)$. The null-hypothesis corresponds to a certain parameter configuration $\boldsymbol{\kappa}$ and alternative hypothesis correspond to arbitrary values. Furthermore, both signal and background hypotheses then share a number of unknown parameters that describe, e.g. experimental uncertainties and are called nuisance parameters, from now on denoted by $\boldsymbol{\theta}$. While the exact form of the likelihood functions $L(\boldsymbol{\kappa})$ is thus unknown, it is still possible to separate hypotheses using a test statistic based on the profile likelihood ratio:

$$q_{\boldsymbol{\kappa}} = -2\log\left\{\frac{L(\boldsymbol{\kappa}, \hat{\hat{\boldsymbol{\theta}}}|\mathbf{x})}{L(\hat{\boldsymbol{\kappa}}, \hat{\boldsymbol{\theta}}|\mathbf{x})}\right\}. \tag{4.4}$$

In the numerator the $\hat{\hat{\boldsymbol{\theta}}}$ maximize the likelihood for fixed $\boldsymbol{\kappa}$ and in the denominator the maximum likelihood estimates $\hat{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\theta}}$ correspond to the global maximum of the likelihood. A very useful property of $q_{\boldsymbol{\kappa}}$ is that, according to Wilks' theorem [121], its distribution can be approximated by a N-dimensional χ^2 distribution in the limit of large sample sizes. Central confidence intervals can then be calculated from the maximum likelihood estimate and its estimated standard deviations, using the properties of χ^2 distributions. For $N = 1$ and $N = 2$ the values of $q_{\boldsymbol{\kappa}}$ that mark the border of confidence intervals are given in Table 4.1.

4.1.3 Upper limits

In this analysis, upper limits are calculated for the $\mathrm{t\bar{t}H}$ signal strength μ, following the conventions used for Higgs-boson searches at the LHC [122]. The incompatibility of the data with a certain signal strength μ is quantified using a profile likelihood ratio test statistic, similar to the one in Equation (4.4). For upper limits it is slightly modified to

$$\tilde{q}_\mu = -2\log\left\{\frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}|\mathbf{x})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}}|\mathbf{x})}\right\}, \text{ with } 0 < \hat{\mu} < \mu, \tag{4.5}$$

meaning that only downward-fluctuations of the data are counted as being incompatible with the signal model. For an observed value $\tilde{q}_\mu^{\mathrm{obs}}$, the probability to find a value as high or higher is

$$p_\mu = \int_{\tilde{q}_\mu^{\mathrm{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\hat{\boldsymbol{\theta}}})\mathrm{d}\tilde{q}_\mu. \tag{4.6}$$

The distributions of $f(\tilde{q}_\mu|\mu, \hat{\hat{\boldsymbol{\theta}}})$ can be calculated with MC methods, by sampling pseudodata from the statistical model with nuisance parameters set to their best-fit values. Alternatively, for a large number of observed events, the distribution can be approximated with an asymptotic formula based on a generalization of Wilks' theorem [123].

To avoid that an experiment is able to exclude a signal just because of a strong downward fluctuation of the data, which is neither compatible with the signal nor the background but leads to small p-values, the modified frequentist (CLs) approach is used. In this approach, also the background p-value p_b is calculated, the probability to observe a value of the test statistic that is as signal-like or more signal-like than $\tilde{q}_\mu^\mathrm{obs}$, assuming the background-only model:

$$1 - p_\mathrm{b} = \int_{\tilde{q}_0^\mathrm{obs}}^{\infty} f(\tilde{q}_\mu | 0, \hat{\hat{\boldsymbol{\theta}}}) \mathrm{d}\tilde{q}_\mu. \tag{4.7}$$

These p-values then used to calculate

$$\mathrm{CL_S} = \frac{p_\mu}{1 - p_b}. \tag{4.8}$$

With this method, a signal strength μ is excluded at the 95% level if $\mathrm{CL_S}$ falls below 5%. The confidence intervals found this way are more conservative than classical confidence intervals.

Expected limits allow to judge the expected sensitivity of an analysis and are usually quoted for the expectation in the absence of a signal. They are derived by sampling pseudodata from the background-only model and calculating the limit for every toy dataset. The distribution of limits can then be used to quote the expected limit, the median of the distribution, and confidence intervals. In the asymptotic limit, it is also possible to derive the expected limit and its uncertainties from asymptotic formulae.

The asymptotic method is based on the properties of an “Asimov dataset” [123]. This dataset removes the need for toy data by replacing them with results from a single dataset, based on the statistical model without including statistical fluctuations. For the calculation of limits, the Asimov dataset consists of the background model with nuisance parameters set to their best-fit value. To compare the values of a test-statistic to the theory expectation, a prefit Asimov dataset can be used, in which all nuisance parameter and parameters of interest are fixed to their prefit estimates.

4.2 Statistical model for Higgs analyses at CMS

In this section, the statistical model as implemented in the so called COMBINE tool and used in most CMS Higgs-boson analyses is introduced. The tool analyzes statistical models built using ROOFIT [124] and performs statistical tests using ROOSTATS [125]. It also allows creating ROOFIT models from text files and ROOT [118] histograms, based on the HISTFACTORY [126] tool. For simplicity, the statistical model is explained only for binned data as it is used this way in Part III of this thesis.

4.2.1 Binned maximum likelihood fit

If the differential cross section of a process as well as luminosity $\mathcal{L}$ and detector acceptance $\mathcal{A}$ and efficiency ϵ is known, the probability to measure a certain final state in a proton-proton collision can be calculated. Since collisions are statistically independent, the observed number of events is Poisson distributed, with a expectation value $\lambda = \mathcal{L}\sigma\mathcal{A}\epsilon$:

$$p(n|\lambda) = \mathrm{Pois}(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}. \tag{4.9}$$

If binned data is analyzed, the likelihood to observe $\mathbf{n} = (n_1, ..., n_{N_\text{bins}})$ events in the N_bins analysis bins is given, expressed in terms of the respective expectation values $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_{N_\text{bins}})$,

$$p(\mathbf{n}|\boldsymbol{\lambda}) = \prod_{i=1}^{N_\text{bins}} \text{Pois}(n_i|\lambda_i). \tag{4.10}$$

In practice neither differential cross sections nor experimental factors are known exactly. However, these systematic uncertainties can often be parameterized by a set of N_p nuisance parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_{N_\text{p}})$:

$$p(\mathbf{n}|\boldsymbol{\theta}) = \prod_{i=1}^{N_\text{bins}} \text{Pois}(n_i|\lambda_i(\boldsymbol{\theta})). \tag{4.11}$$

The uncertainties described by the θ_j can mostly be constrained by auxiliary measurements or theoretical arguments and follow themselves a probability distributions π_j, so that the probability becomes

$$p(\mathbf{n}|\boldsymbol{\theta}) = \prod_{i=1}^{N_\text{bins}} \text{Pois}(n_i|\lambda_i(\boldsymbol{\theta})) \prod_{j=1}^{N_\text{p}} \pi_j(\theta_j). \tag{4.12}$$

Parameters of interest can be treated in the same way as nuisance parameters but are typically unconstrained. For given data the expression (4.12) can be interpreted as a likelihood and be used to derive maximum-likelihood estimates of the parameters $\boldsymbol{\theta}$. This model is close to the one implemented for CMS and differences are mostly of technical origin.

4.2.2 Shape interpolation

Most Higgs-boson searches try to infer the Higgs-boson signal from the shape of the distribution of an observable. This can be an invariant mass distribution, a multi-dimensional distribution that takes multiple kinematic properties into account, or the distribution of the output of a multivariate discriminant. Especially in the last case, the effect of all combinations of systematic uncertainties is not easy to parameterize in terms of $\boldsymbol{\theta}$. For binned analyses a default method based on the HistFactory [126] tool exists in CMS, which is outlined in the following.

With the method the function $\lambda_i(\boldsymbol{\theta})$, introduced above, is implemented in all (histogram) bins of the analysis. The necessary procedure is performed independently for all processes p and nuisance parameters j. It is assumed that the effect of multiple sources of systematic uncertainties on analysis bins corresponds to independent multiplicative corrections:

$$\lambda_i(\boldsymbol{\theta}) = \sum_p \nu_{pi}^0 \prod_j \beta_{jpi}(\theta_j), \tag{4.13}$$

where ν_{pi}^0 is the nominal prediction for process p in bin i and β_{jpi} is a function that additionally depends on the uncertainty j. One can distinguish between rate uncertainties, which change the expected rate in all bins by the same factor β_{jp}, and shape uncertainties, which affect all bins differently.

For a normalization uncertainty, the function $\beta_{jp}(\theta_j)$ can in principle be linear in θ_j for all affected processes p. Depending on the source of the uncertainty, different priors $\pi(\theta_j)$ are used. Often it is assumed that the normalization of a process follows a log-normal distribution whose width is parameterized by Δ_{jp} and whose median is given by the nominal expectation. In this case, the dependence of the normalization on the nuisance parameter is implemented as

$$\beta_{jp}(\theta_j) = (1 + \Delta_{jp})^{\theta_j}. \tag{4.14}$$

and the nuisance parameter θ_j is constrained by a normal distribution, which is equivalent to β_{jp} being log-normal distributed.

Modeling shape uncertainties is slightly more complicated. First, the effect of upwards and downwards variation by one standard deviation, $\nu^{\pm}_{jpi}$, have to be determined for every process and every systematic uncertainty in every bin. Typically, this is done by generating two additional distributions using MC smimulated data with varied assumptions. If the shape-systematic also affects the normalization, this effect is modeled as a log-normal distributed, as described above. The distributions are then normalized and the remaining differences parameterized by interpolating between ν^0_{pi} and $\nu^{\pm}_{jpi}$. The function β_{jpi} interpolates linearly between nominal and up as well as nominal and down value but is smoothed at $\theta_j = 0$ to keep it differentiable.

4.3 Multivariate analysis tools

In this thesis, two multivariate analysis methods are prominently featured: Boosted Decision Trees (BDTs) and the Matrix Element Method (MEM). They are described in this section.

4.3.1 Matrix Element Method

The Matrix Element Method is a multivariate analysis technique used in particle physics. It was introduced in [127, 128] and made famous in top-mass measurements at DØ [129]. With the Matrix Element Method, a likelihood-ratio discriminant based on the hard-scattering cross section formula (2.10), which uses the matrix element $\mathcal{M}$, is constructed. The likelihood-ratio can be used to separate a signal process from a background process or to determine parameters of a model, e.g. a particle mass. The likelihoods depend on the kinematic properties of objects measured in an experiment $\mathbf{p}^{\text{rec}}$, typically the four-momentum vectors of particle candidates, and is a function of one or multiple parameters θ that enumerate signal and background hypothesis or describe parameters of interest.

For a hadron collider experiment, the likelihoods are calculated considering the ingredients necessary to calculate an estimate of the differential cross section of the hard process, i.e. the (leading order) matrix element of the hard process and parton distributions functions. Furthermore, the effects of hadronization, detector resolution, final state combinatorics, and invisible particles can be incorporated. Symbolically, the likelihood can be written as

$$\begin{aligned} L(\theta|\mathbf{p}^{\text{rec}}) \sim & \int \mathrm{d}x_1 \int \mathrm{d}x_2 f_1(x_1) f_2(x_2) \\ & \times \int \mathrm{d}\mathbf{p}^{\text{true}} |\mathcal{M}(x_1, x_2, \mathbf{p}^{\text{true}}, \theta)|^2 \\ & \times W(\mathbf{p}^{\text{true}}, \mathbf{p}^{\text{rec}}). \end{aligned} \tag{4.15}$$

Phase space factors, δ-distributions that ensure momentum conservation, and sums over parton flavors and final state permutations have been omitted for better readability. The $f_i(x_i)$ describe the distribution of the two partons initiating the interaction. The hard process is described by the matrix element $\mathcal{M}$, which is a function of the initial state partons, with proton-momentum fractions x_i, and the final state particles with momenta $\mathbf{p}^{\text{true}}$.

W are transfer functions that describe the probability to observe momenta $\mathbf{p}^{\text{rec}}$ for true momenta $\mathbf{p}^{\text{true}}$. If objects are assumed to be measured perfectly, they are a product of δ-distributions and allow to skip the integration over $\mathbf{p}^{\text{true}}$. For objects with a limited resolution, e.g. jets, which are subject to hadronization and resolution effects, W can, for example, consist of products of Gaussian-like transfer function $\prod_i w_i(\mathbf{p}_i^{\text{rec}}, \mathbf{p}_i^{\text{rec}})$. Objects can also be loosely constrained, e.g. neutrino momenta by $\not{E}_{\text{T}}$ measurements or it can even be decided to assume a flat transfer function. In the case that final state objects (e.g. jets) can be assigned to multiple $\mathbf{p}^{\text{true}}$, all permutations have to be evaluated and summed up.

The phase-space integration of $\mathbf{p}^{\text{true}}$ is the main technical challenge when implementing a Matrix Element Method. Final states can contain many objects and thus high-dimensional integrals have to be calculated. This is usually done using involved Monte Carlo methods that ensure that the relevant part of the phase space, where the transfer functions are non-zero, is covered efficiently.

4.3.2 Boosted Decision Trees

Boosted Decision Trees (BDTs) are supervised learning algorithms. They can be used to classify objects into two classes. From a so called training sample, for which the classes of the objects are known, the algorithm develops a classification heuristic. The response of the BDT is a number that is higher, the higher the probability of an object to belong to a certain class is estimated. In this analysis, BDTs are used to classify events into signal and background.

Decision trees and regression trees

BDTs are based on decision trees. A decision tree can be used to sort a sample event-by-event into two classes, signal and background, based on the features $\mathbf{x} = \{x_i\}$ of events. The tree consists of nodes and leaves. Starting with the root node, at every node a decision is made, to which class y the event more likely belongs. This decision is based on whether one of the features x_i is above or below a cut value c_j. The node splitting divides the sample into two subsamples. Every subsample is split again, based on the same or a new feature. The tree ends at a certain depth and splits the space of $\mathbf{x}$ into regions of different likelihood to contain signal or background. This is illustrated in Fig. 4.1

A decision tree can be built automatically, in a procedure called training. The training is performed on a dedicated sample in which the class y (signal or background, represented by $y = 1$ and $y = -1$, respectively) is known for all events. The goal of the training is to divide the multi-dimensional $\mathbf{x}$-space into regions that are dominated by signal or background. It is assumed that the training sample is representative for data the tree will classify. Beginning with the root node, the variable x_i and the cut value c_j are determined that is best-suited to split the sample into the two categories. A popular choice for the splitting criterion is the Gini index. By the splittings, the space of $\mathbf{x}$ is split into disjoint regions, labeled as signal or background.

In principle, this training procedure could be continued until the phase space is divided in such a way that all training events are correctly classified. However, independent events cannot be expected to behave exactly like the training events and thus this fine splitting of the phase might not work for them. This phenomenon, that a classifier performs only

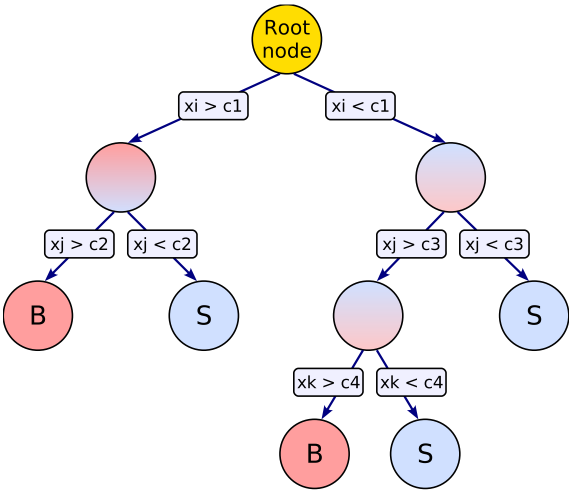


Figure 4.1: Example for a decision tree, that split a sample at values c_i of features x_j into subsamples that are dominated by signal (S) or background (B).

well on the training sample, is called overtraining. One way to mitigate this is to create only “weak learners” by stopping the phase space splitting after only a few steps.

A generalization of the decision tree algorithm is a regression tree. It is used to estimate a target y that can have arbitrary values. The regression tree divides the sample into subsamples, like a decision tree, by a series of binary splits. The splitting criterion is chosen such that the y-values of events in each node is of similar size. In the nodes of a regression tree the target value can be estimated by the average of all training events contained.

Gradient Boosted Decision Trees

The limitations of decision and regression trees can be overcome by combining several weak learners with a boosting algorithm. In this thesis, the gradient boosting algorithm [130] as implemented in TMVA [131] is used to create BDTs. The idea of the algorithm is to estimate a function F of the feature space $\mathbf{x}$ that is an estimate of the signal-background ratio or a related quantity. The function is constructed as a series of regression trees by trying to minimize a loss function $L(F(\mathbf{x}), y)$ for all values of $\mathbf{x}$. The loss function becomes larger the more target y and estimate F deviate. The distribution of the features $\mathbf{x}$ is estimated from a training sample with events $\mathbf{x}_n$.

The initial estimate of F_0 is constant. For every additional tree F_m, the negative gradient of the loss function, the so called the pseudo-residuals, are calculated for a randomly chosen subsample of training events $\mathbf{x}_n$:

$$\tilde{y}_{nm}(\mathbf{x}_n) = -\left[\frac{\partial L(F, y_n)}{\partial F}\right]_{F=F_{m-1}(\mathbf{x}_n)}. \tag{4.16}$$

The pseudo-residuals indicate how the estimate F_{m-1} has to be improved to reduce the loss function. Using only a subsample containing the fraction b of events reduces the algorithms susceptibility to overtraining.

The distribution of $\tilde{y}_{nm}$ is then estimated by a shallow regression tree f_m of a small depth d that can correct for the shortcomings of F_{m-1}. While in principle every function f_m could be used for this purpose, the restriction to shallow regression trees reduces overtraining as they extrapolate the value obtained from the points y_{mn} in the neighboring region of the $\mathbf{x}$-space. To further reduce overtraining effects, only N_{cuts} equidistant cut-values are tried to optimize the node-splitting of the f_m.

For a more robust convergence, a shrinkage parameter $\nu < 1$ can be introduced and the new estimate F_m is defined as.

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \nu f_m(\mathbf{x}). \tag{4.17}$$

This procedure is iterated until a user-defined maximum number of trees $m = N_{\mathrm{trees}}$ is built.

In summary, the training depends, among others, on the parameters N_{trees}, ν, N_{cuts}, b, and d. In addition to these parameters, also the features used in the training have an impact on the BDT performance. As many features that allow separating signal from background as possible have to be identified. However, if too many features are used, the performance can also become worse because the larger the number of features is, the higher the probability is that one of them behaves, by statistical fluctuations, differently in the training sample compared to an independent sample.

II. Indirect top-Higgs coupling constraints from LHC Run 1 data

5 Review of Higgs coupling constraints

In this part of the thesis, an indirect measurement of the top-Higgs coupling is performed. Hereby a potentially anomalous coupling is considered and the coupling strength and structure is inferred from a combined analysis of Higgs signal strength measurements in LHC Run 1. The measurement is indirect in the sense that the sensitivity stems from the distinct role the top-Higgs coupling plays in determining the rate of gluon fusion and $\mathrm{t\bar{t}H}$ production as well as the branching ratios into photons and gluons. Thus this indirect approach is only valid under the assumption that there are no contributions from particles beyond the Standard Model to loop-induced Higgs production and decay processes, which would spoil the assumed top-Higgs coupling dependence. A more direct measurement of anomalous couplings is possible, for example, in the $\mathrm{t\bar{t}H}$ channel and presented in Part III of this thesis.

This chapter is organized as follows: Section 5.1 will introduce the Higgs coupling framework used in LHC Run 1. The coupling measurement published by CMS and ATLAS are summarized in Section 5.2. Measurements of anomalous top-Higgs couplings are introduced in Section 5.3.

5.1 Framework for Higgs coupling analysis in LHC Run 1

As described in Section 1.3.4, the Standard Model makes exact predictions for the couplings between the Higgs boson and all gauge bosons and fermions. Precise measurements of these couplings are one way to search for physics beyond the Standard Model. Every deviation from the predicted couplings is new physics and would need to be explained by a new theory. There are different ways to parameterize deviations from the Standard Model configuration for these couplings. Different parameterizations are motivated by different assumptions on how beyond the Standard Model physics can affect the Higgs sector of the Standard Model.

In principle, the Higgs couplings can affect almost all physical processes. However, because the Higgs is massive and the coupling to most SM particles comparatively weak, most of these effects are hard to detect. To precisely measure the couplings, one has to investigate Higgs boson production and decay at the LHC, processes that are only possible because of the existence of the Higgs couplings. The parameterization of the coupling chosen needs thus be able to model the effect of changes in couplings on Higgs cross sections and branching ratios.

For the main coupling analyses performed by ATLAS [132] and CMS [133] in LHC Run 1 and for the combined measurement [8], the interim framework recommended by the LHC Higgs Cross Section Working Group [81] ("κ-framework") is used. In this framework, it is assumed that there is only one Higgs boson, with a narrow width. This assumption allows to factorize processes in terms of the Higgs boson production cross section $\sigma_{\mathrm{pp}\to\mathrm{H}+x}$ and the Higgs boson partial and total width $\Gamma_{\mathrm{H}\to y}$ and Γ_{total}:

$$\sigma_{\mathrm{pp}\to\mathrm{H}+x,\ \mathrm{H}\to y} = \sigma_{\mathrm{pp}\to\mathrm{H}+x}\frac{\Gamma_{\mathrm{H}\to y}}{\Gamma_{\mathrm{total}}}. \tag{5.1}$$

It is postulated furthermore that BSM physics only affects the size of absolute cross sections and branching ratios of the Higgs production and decay processes in the Standard Model, to allow for an easy re-interpretation of the results of the SM Higgs searches without the need to generate new MC simulations or to recalculate the acceptance of different models. This means that is not possible to change the Lorentz structure of couplings, which would change the kinematics of the process to which these couplings contribute to. As a consequence, the Higgs boson is assumed to be a CP-even scalar.

The framework introduces a number of coupling modifiers κ_i whose effect is closely related to a rescaling of the couplings of the Higgs boson to other particles in the Standard Model (effective) Lagrangian by a constant real factor. Depending on the tests performed, the couplings between Higgs and the massive bosons, the couplings to fermions, or the effective couplings between Higgs and gluons and photons can be modified. However, only effects in leading order in a perturbation series in this coupling are considered. For example, the cross section for $\mathrm{t\bar{t}H}$ production is proportional to the top-Higgs coupling modifier κ_t^2 in this model, even though the dependence on this coupling strength is different when considering electroweak corrections [134]. Because of the small size of the electroweak corrections this is in general a good approximation and simplifies the interpretation of the model.

The scaling of cross sections and branching ratios is defined in such a way that the most precise Standard Model predictions (including electroweak corrections) are recovered by setting all κ_i to unity. For other values the change in cross section is derived from the relative size of the contributions due to the different Higgs couplings contributing to the considered process. These contributions are denoted σ_x^{ii} or Γ_x^{ii}, where i stands for the involved coupling and x for the process under consideration. Interference effects in processes depending on different Higgs couplings are parameterized in the same way, they are denoted σ_x^{ij} or Γ_x^{ij} with $i \neq j$ and can be negative. Such a destructive interference is present, e.g. in the Higgs decay into two photons, whose decay width can be parameterized (omitting minor contributions of fermion couplings other than the top quark): $\Gamma_{\mathrm{H}\to\gamma\gamma} \approx \kappa_\mathrm{t}^2\Gamma_{\mathrm{H}\to\gamma\gamma}^{\mathrm{tt}} + \kappa_\mathrm{W}^2\Gamma_{\mathrm{H}\to\gamma\gamma}^{WW} + \kappa_\mathrm{t}\kappa_\mathrm{W}\Gamma_{\mathrm{H}\to\gamma\gamma}^{\mathrm{tW}}$. Higher order QCD corrections factorize and can be included in the calculation of the subprocess cross sections and decay widths σ_x^{ij} and Γ_x^{ij}.

In this thesis, two sets of κ_i are considered to parameterize deviation from the Standard Model. One is the “resolved” approach, in which the signal strength of loop induced Higgs production and decay processes is expressed in terms of the coupling modifiers of all the Standard Model particles that contribute. In the second approach the loop induced processes are parameterized in terms of effective coupling modifiers κ_g and κ_γ. The introduction of the effective scaling factors is motivated by the fact that one needs physics beyond the Standard Model to explain deviations from the SM coupling. These theories beyond the Standard Model could contain particles that also contribute to the loop induced processes, making the formulas of the resolved approach invalid. For the “effective scaling” approach, it is natural to also allow for the decay into BSM particles which can be either unobserved or invisible. At face value this would make it impossible to constrain the Higgs couplings: increased production cross sections due to larger couplings can always be canceled by increasing the partial width to undetected final states and thereby reducing the size of the observed branching ratios. Thus either no BSM decays are allowed or the assumption that the vector boson couplings fulfill $|\kappa_V| \leq 1$ has to be made. A summary of the cross section modifications defined by the κ_i that are most relevant for this thesis is shown in Table 5.1.

The κ-framework is not a complete theory and does not motivate the origins of the

Table 5.1: Dependence of Higgs production cross sections σ and partial decay widths Γ on the coupling scaling factors κ_i as used in [8]. For reasons of clarity only the most important contributions to the most relevant processes are displayed. In total the five modifiers κ_W, κ_Z, κ_t, κ_b, κ_τ are considered in the resolved approach and in the effective approach κ_g, κ_γ are used in addition.

Process	Resolved	Effective
σ_ggH	$1.06\kappa_\mathrm{t}^2 + 0.01\kappa_\mathrm{b}^2 - 0.07\kappa_\mathrm{t}\kappa_\mathrm{b}$	κ_g^2
σ_VBF	$0.74\kappa_\mathrm{W}^2 + 0.26\kappa_\mathrm{Z}^2$	
σ_WH	κ_W^2	
$\sigma_{\mathrm{q\bar{q}}\to\mathrm{ZH}}$	κ_Z^2	
$\sigma_{\mathrm{gg}\to\mathrm{ZH}}$	$2.27\kappa_\mathrm{Z}^2 + 0.37\kappa_\mathrm{t}^2 - 1.64\kappa_\mathrm{Z}\kappa_\mathrm{t}$	
$\sigma_\mathrm{t\bar{t}H}$	κ_t^2	
σ_tHW	$1.84\kappa_\mathrm{t}^2 + 1.57\kappa_\mathrm{W}^2 - 2.41\kappa_\mathrm{W}\kappa_\mathrm{t}$	
σ_tHq	$3.40\kappa_\mathrm{t}^2 + 3.56\kappa_\mathrm{W}^2 - 5.96\kappa_\mathrm{t}\kappa_\mathrm{W}$	
$\sigma_\mathrm{b\bar{b}H}$	κ_b^2	
Γ_ZZ	κ_Z^2	
Γ_WW	κ_W^2	
$\Gamma_\mathrm{b\bar{b}}$	κ_b^2	
$\Gamma_{\tau\tau}$	κ_τ^2	
$\Gamma_{\gamma\gamma}$	$1.59\kappa_\mathrm{W}^2 + 0.07\kappa_\mathrm{t}^2 - 0.66\kappa_\mathrm{W}\kappa_\mathrm{t}$	κ_γ^2

deviations from the Standard Model. If deviations from unity in one or several of the coupling modifiers are established, the data would need to be confronted with a more well-defined alternative. The framework is also limited in the BSM effects it can detect. Changes in the couplings that change kinematics of Higgs production and decay but do not affect measured signal strengths cannot be detected in this model. The sensitivity to BSM effects that affect several couplings in a correlated way is also not ideal. An example of such an effect is the anomalous top-Higgs coupling discussed later in this chapter, which can be detected as simultaneous modification of κ_t, κ_g, and κ_γ in the effective κ-model.

A conceptually more well defined approach is to parameterize the Higgs interactions in an effective field theory (EFT) framework (see, e.g. [135]). In EFTs, operators are introduced that can be interpreted as the low-energy manifestation of interactions of a theory with particles of large mass. These operators allow to modify particle interactions and their strength can also be parameterized by a number of coefficients. However, in EFTs a large number of parameters is introduced that affect the kinematics of many processes. Simulating these effects and measuring all the operators is much more challenging than using the simpler κ-framework.

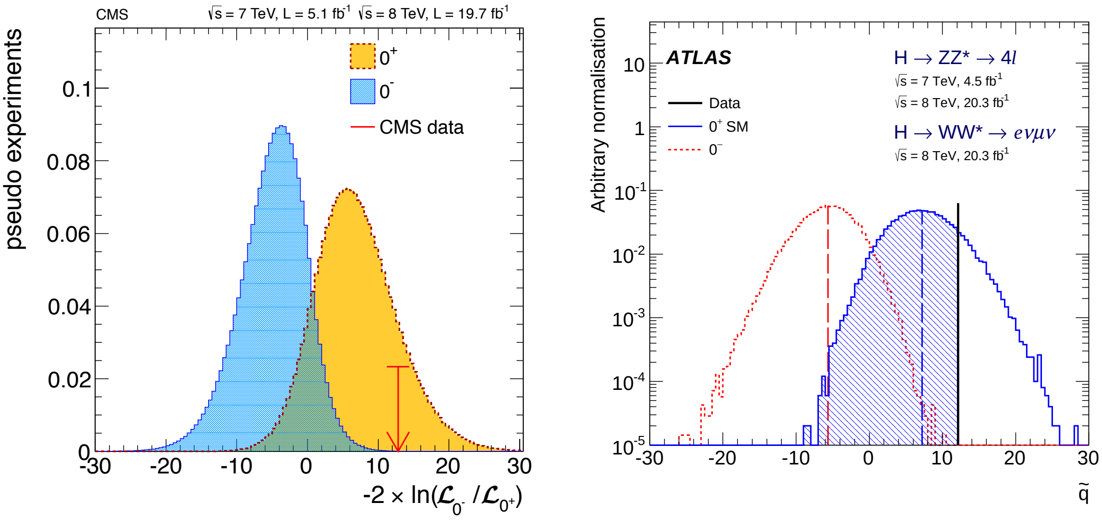


Figure 5.1: The likelihood ratio test statistic used to distinguish a scalar from a pseudoscalar Higgs boson for CMS (left, [136]) and ATLAS (right, [84]) in H → ZZ (CMS) and H → ZZ/WW (ATLAS) events. The red arrow (black line) indicates the value of the test statistics for the data measured by CMS (ATLAS). The expected values from pseudo experiments with a scalar and pseudoscalar Higgs boson are shown in orange and blue for CMS and blue and red for ATLAS.

5.2 Overview of Higgs coupling constraints

5.2.1 Higgs coupling constraints from the LHC experiments

Using the LHC interim framework described above, the Higgs coupling modifiers have been measured by CMS and ATLAS. The most precise results have been extracted from the combination of the two experiments in [8]. The couplings to vector bosons can be stringently constrained under the three different assumptions outlined above, that either no BSM particles couple to the Higgs boson at all, that the Higgs boson does not decay into BSM particles, or that the vector boson coupling modifiers are ≤ 1, but without making other assumptions on the values of the remaining κ_i. The couplings κ_V are measured with a precision of 10% and compatible with SM expectation at the level on one standard deviation.

Using events in which the Higgs boson decays into a pair of vector bosons, its spin and CP structure can be tested with high precision. The hypothesis that the boson behaves like a pseudoscalar is rejected at a confidence level of higher than 99.9%, as shown in Fig. 5.1. Most models with a spin different from have been excluded at a confidence level of more than 99% by CMS [83] and ATLAS [84]. Furthermore, a pseudoscalar admixture to the couplings of the Higgs boson to vector bosons was tested in Higgs decays and large admixtures have been excluded by CMS [9, 136] as well as ATLAS [10] but a sizable admixture is still compatible with the current experimental sensitivity (see Fig. 5.2). By combining ZH production and H → ZZ, CMS could set very tight constraints on this admixture in [137], assuming the absence of other non-standard couplings.

The strength of the coupling to fermions has been measured, too. In general, the precision is worse than in the bosonic sector. So far, only Higgs boson couplings to third generation fermions have been established. The couplings to b quarks and τ-leptons are determined mostly from the respective Higgs decay rates, with a precision between 15%

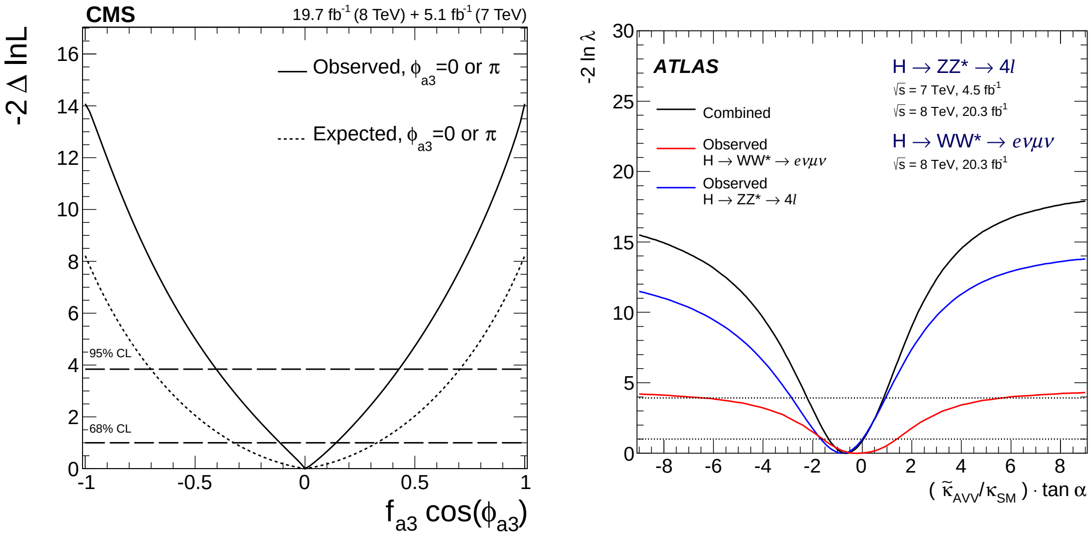


Figure 5.2: Value of the profile likelihood ratio test statistic employed to detect pseudoscalar admixtures to the Higgs coupling to vector bosons. On the left the CMS [83] result determined from H → ZZ decays and on the right the ATLAS [84] result, combining WW and ZZ decays, is shown. For the CMS result the pseudoscalar admixture is measured using f_{a3}, which describes the effective cross section fraction due to a pseudoscalar Higgs boson coupling while for the ATLAS result the ratio of pseudoscalar to Standard Model coupling is used.

and 30%. While the measurement of κ_τ is compatible with the SM, the measurement of the b quark coupling is two standard deviations below the expectation. The results for the top-Higgs coupling modifier κ_{t} strongly depend on the parameterization chosen within the κ-framework. In the “resolved” approach the coupling has a strong effect on gluon fusion production and is measured to be $\kappa_{\mathrm{t}} = 0.87 \pm 0.15$. When the couplings to gluons and photons are allowed to vary indepedently, the top-Higgs coupling can only be measured in $\mathrm{t\bar{t}H}$ production and the measurement yields, depending on the exact model about $\kappa_{\mathrm{t}} = 1.4 \pm 0.2$, driven by an excess of events in the search for $\mathrm{t\bar{t}H}$.

Results of the combined Higgs coupling strength measurement by CMS and ATLAS are shown in Figures 5.3 and 5.4. No measurements of the CP structure of the fermion couplings have been performed so far.

5.2.2 Motivation to search for a non-standard top-Higgs coupling

Investigating the Higgs couplings to fermions, and especially the top-Higgs coupling, is important for a number of reasons. The Standard Model predicts that the size of the coupling between the top quark and Higgs field before electroweak symmetry breaking is $y_t \approx 1$, which is larger by a factor of $\frac{m_t}{m_f}$ than the Yukawa coupling of any other fermion f. The large differences in the size of the fermion couplings are unexplained and studying the largest coupling might give hints for its origin. The large size could be a hint of an involvement in the mechanism of electroweak symmetry breaking, as, e.g. in little Higgs [138] models.

The strength of the coupling is also responsible for the large dependency of quantum corrections to the Higgs mass on the cut-off scale (see e.g. [61]). This dependency means that if physics beyond the Standard Model is present at a high energy scale but no pro-

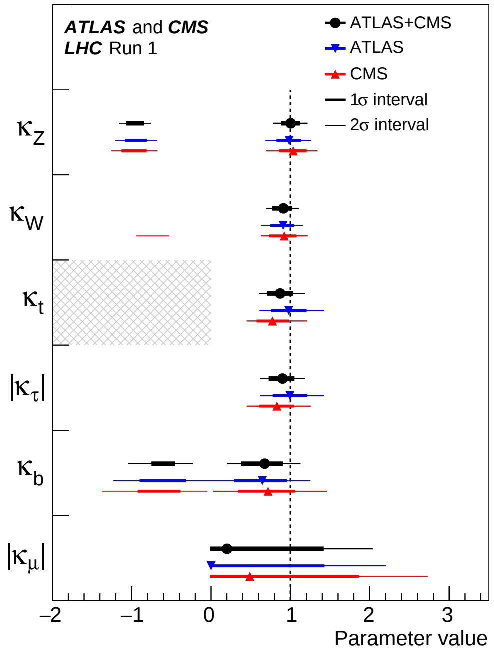


Figure 5.3: Higgs couplings deduced in the "resolved" approach from the CMS and ATLAS Higgs boson measurements. Black dots and fat (slim) lines indicate the best fit values and the 1σ (2σ) confidence level intervals. The results by ATLAS (CMS) are shown in blue (red), their combination black. The shaded area indicates that $\kappa_t \geq 0$ is required. Taken from [8].

tection mechanism as in SUSY exists, fine-tuning is neccessary to explain the small Higgs boson mass. Furthermore, the top-Higgs coupling plays an important role in the calculation of quantum corrections to the Higgs self-interaction λ. A top-Higgs coupling that was a few percent stronger would lead to a Higgs potential in which the vacuum state is unstable (see, e.g. [139]).

The top-Higgs coupling is furthermore important from a phenomenological viewpoint, being (partly) responsible for Higgs decays into photons and Higgs production in gluon fusion and $t\bar{t}H$. As described above, measurement in the $t\bar{t}H$ channel are in some tension with the κ_t values extracted from gluon fusion Higgs production.

The interim framework used to interpret the LHC Run 1 Higgs results requires the Higgs boson to be a CP-even scalar particle. However, it might be also interesting to investigate the possibility of an CP-indefinite Higgs boson, having both scalar and pseudoscalar couplings to the Higgs boson, as is done in the Higgs boson vector-boson coupling measurements referenced above. Such a coupling would lead to a new source of CP violation, in addition to the one obervable due to the CKM matrix. CP violation in the Higgs sector would be new physics, as none is predicted in the SM. While there are many constraints on the presence of CP violation, it is needed, e.g. to explain the baryon asymmetry in the universe, i.e. the fact that there is more matter than anti-matter. In two Higgs-doublet models like supersymmetry, CP odd and even Higgs bosons are predicted. A mixing between the two states could be one possible origin of anomalous coupling [140]. While pseudoscalar admixtures to the boson coupling have been constrained to be rather small, the situation in the fermionic sector could be completely different. In fact, at tree level, a pseudoscalar Higgs boson does not couple to the SM vector bosons, so that one can

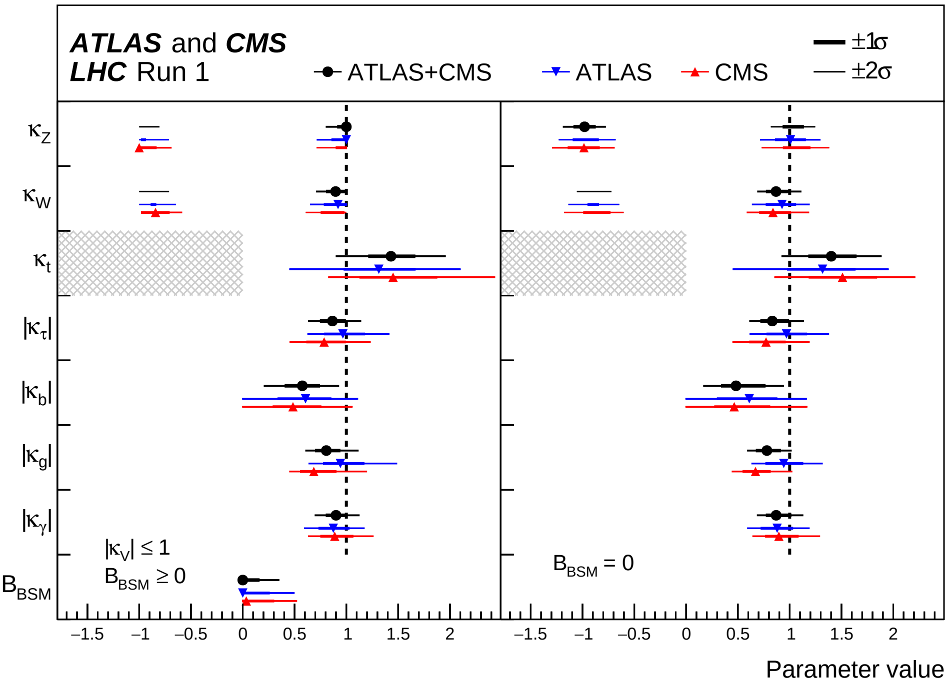


Figure 5.4: Higgs couplings deduced in the effective approach from the CMS and ATLAS Higgs measurements [8]. Black dots and fat (slim) lines indicate the best fit values and the 1σ (2σ) confidence level intervals. The results by ATLAS (CMS) are shown as blue (red) line, their combination black. On the right-hand side decays into BSM particles are assumed to be non-existent, on the left they are allowed but the couplings to vector bosons required to be $\kappa_V \leq 1$.

easily explain pseudoscalar fermionic couplings that are much larger than their bosonic counterparts.

5.2.3 Parameterization of anomalous top-Higgs couplings

The top-Higgs interaction and, in general, any fermion-Higgs interaction can be described by the following two terms in the Lagrangian of a quantum field theory (see, e.g. [141]):

$$\mathcal{L}_{\mathrm{t\bar{t}H}} = -\frac{m_{\mathrm{t}}}{v}\bar{\psi}_{\mathrm{t}}(\kappa_{\mathrm{t}} + \tilde{\kappa}_{\mathrm{t}} i\gamma_5)H\psi_{\mathrm{t}}. \tag{5.2}$$

Here m_t is the top quark mass, v the Higgs vacuum expectation value, κ_{t} and $\tilde{\kappa}_{\mathrm{t}}$ two parameters, and ψ_t the top quark and H the Higgs-boson field operator. In the Standard Model Lagrangian it is $\kappa_{\mathrm{t}} = 1$ and $\tilde{\kappa}_{\mathrm{t}} = 0$. For a purely scalar (pseudoscalar) boson only the first (second) term is present. The parameter κ_{t} modifies the strength of the scalar coupling, $\tilde{\kappa}_{\mathrm{t}}$ the strength of the pseudoscalar coupling. The two parameters are taken to be real, so that the Lagrangian is Hermitian. This assumption is usually, but not always made (e.g. not in [142]). From this extension to the Standard Model Lagrangian a modified κ-model can be built. The framework of Section 5.1 has to be modified only minimally. The paramter κ_{t} is identified with the one from Equation (5.2) and a single additional parameter $\tilde{\kappa}_{\mathrm{t}}$ is introduced.

Other Higgs boson couplings are taken to be purely scalar in the chosen parameteriziation. In the bosonic sector this assumption is motivated by the measurements and theoretical arguments above, which disfavor large pseudoscalar vector boson couplings. The Higgs couplings to fermions other than the top quark contribute significantly only in Higgs decays to τ-leptons and b quarks. Here, the kinematics of τ-leptons or b quarks are not affected by the coupling structure[1]. The only possible effect of the two couplings is thus a change in the signal strength of one process each and one parameter, κ_τ and κ_b respectively, is enough to parameterize this effect, may it originate in a different strength or a different coupling strucure. This situation is different for the top-Higgs coupling, which plays a role in several Higgs production and decay modes.

Both the resolved and the effective scaling scheme can be used with this modified κ-framework. BSM Higgs boson deacys are not considered in this analysis, implying that new particles coupling to the Higgs boson are too heavy to allow for direct decays. The alternative assumption that $|\kappa_V| \leq 1$ yields similar results in the combined Higgs fit by CMS and ATLAS and this is not expected to be different for this study.

Sometimes the following equivalent parameterization of the top-Higgs coupling is useful, too:

$$\mathcal{L}_{\mathrm{t\bar{t}H}} = -\bar{\kappa}_{\mathrm{t}} \frac{m_{\mathrm{t}}}{v} \bar{\psi}_{\mathrm{t}} (\cos(\zeta_{\mathrm{t}}) + \sin(\zeta_{\mathrm{t}}) i\gamma_5) H \psi_{\mathrm{t}}. \tag{5.3}$$

Here the parameter ζ_{t} describes the mixing between scalar and pseudoscalar coupling and $\bar{\kappa}_{\mathrm{t}}$ the strength of the coupling. For direct measurements in $\mathrm{t\bar{t}H}$ production one can also introduce the parameter $\tilde{f}_t$ in analogy to the f_{a3} parameter from boson-coupling measurements described above. It is defined as the relative contribution of the pseudoscalar effective cross section,

$$\tilde{f}_t = \frac{\tilde{\kappa}_{\mathrm{t}} \tilde{\sigma}_{\mathrm{t\bar{t}H}}}{\kappa_{\mathrm{t}} \sigma_{\mathrm{t\bar{t}H}} + \tilde{\kappa}_{\mathrm{t}} \tilde{\sigma}_{\mathrm{t\bar{t}H}}}. \tag{5.4}$$

Hereby $\sigma_{\mathrm{t\bar{t}H}}$ ($\tilde{\sigma}_{\mathrm{t\bar{t}H}}$) is the $\mathrm{t\bar{t}H}$ production cross section for a purely scalar (pseudoscalar) boson, i.e. $\kappa_{\mathrm{t}} = 1$, $\tilde{\kappa}_{\mathrm{t}} = 0$ ($\kappa_{\mathrm{t}} = 0$, $\tilde{\kappa}_{\mathrm{t}} = 1$).

5.3 Existing anomalous top-Higgs coupling constraints

5.3.1 Constraints from low energy measurements

It was pointed out in [144] that if the Higgs boson mediates CP violation, an electric dipole moment (EDM) for the electron can be induced. Using the bounds of the ACME collaboration [145], constraints on the CP violation due to an anomalous top-Higgs coupling are derived in [141] and an upper bound of $|\tilde{\kappa}_t| < 0.01$ is calculated (see Fig. 5.5, left). However, this bound assumes Higgs couplings to electrons as in the Standard Model and the absence of further BSM contributions. It is also possible to derive constraints from neutron EDM measurements, which do not depend on Higgs coupling to first-generation fermions. While these constraints are, at the moment, only of the order $|\tilde{\kappa}_{\mathrm{t}}| \lesssim 1$ (see Fig. 5.5, right), they could become competitive once more precise measurements of the neutron EDM are available. Nonetheless, according to [146], in theories beyond the Standard Model, e.g. in supersymmetric theories, a mechanism could exist that cancels the top-Higgs contribution that makes these constraints invalid.

---

[1] The τ decay products can give insight into the nature of the τ coupling but the published CMS and ATLAS analyses are not able to probe this effect [143].

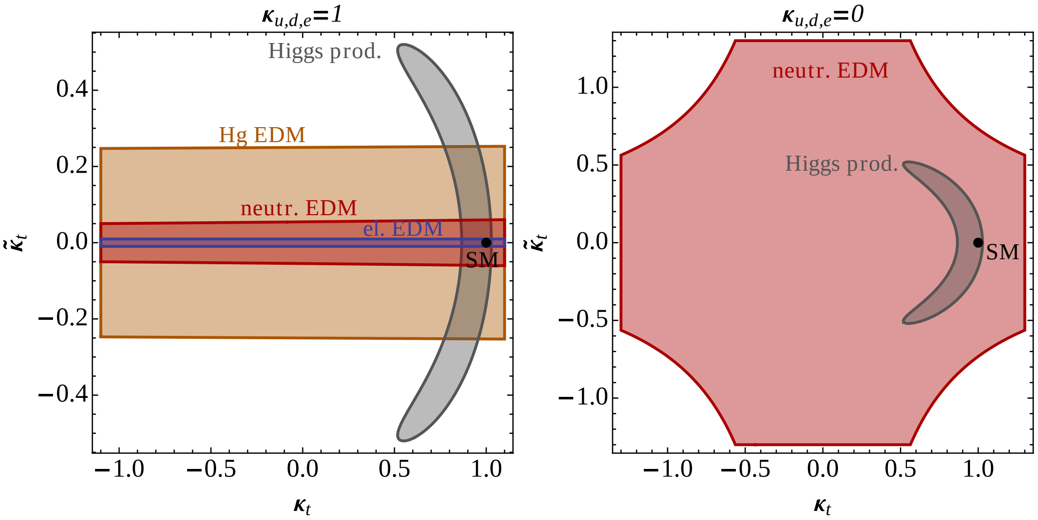


Figure 5.5: Constraints of the top-Higgs coupling modifiers κ_{t} and $\tilde{\kappa}_{\mathrm{t}}$, taken from [141]. In the left plot all couplings except the top-Higgs coupling are assumed to have Standard Model values, in the right plot first-generation couplings are set to zero. The gray area shows constrains from Higgs boson physics at the 68% confidence level and the brown, red, and blue areas constraints from measurements of the mercury, neutron, and electron EDM.

5.3.2 Constraints from Higgs boson physics

A number of cross sections of Higgs boson production and all Higgs boson branching ratios depend non-trivially on the parameters κ_{t} and $\tilde{\kappa}_{\mathrm{t}}$. This allows extracting constraints on these coupling modifiers from the measurements of Higgs interactions by the ATLAS and CMS collaborations. This has been done in several publications, e.g. [140, 141, 146–148].In these five publications it is assumed that the top-Higgs coupling is the only coupling different from the Standard Model prediction. The approach used in [141] is fairly simple, constraints on κ_{t} and $\tilde{\kappa}_{\mathrm{t}}$ are derived from the values of κ_{γ} and κ_g published by the CMS and ATLAS collobarations at the beginning of 2013 [149, 150]. With this simplified approach correlations between these parameters are not considered and the constraints are very likely too tight. The analyses described in [146] and [147] use similar input analyses but derive constraints on the parameters by directly fitting the signal strength measured by the experiments in the different analysis channels, thereby considering the main sources of correlations correctly. The same procedure is used in [140] but the input analyses also include $\mathrm{t\bar{t}H}$ measurements by ATLAS in the $\mathrm{H} \to \gamma\gamma$ [25] and $\mathrm{b\bar{b}}$ channel [24] and by CMS is several channels [22]. A similar analysis is performed in [147], using CMS, ATLAS, CDF, and DØ results that are included the HiggsSignals [151] framework. All results are displayed in Fig. 5.6.

There are no results on anomalous pseudoscalar top-Higgs couplings from ATLAS or CMS available so far. CMS constraints will be derived in the next chapter and there the technique used and its limitations will be explained in more detail. Two remotely related measurements have been published by ATLAS: a measurement of the top-Higgs coupling in the $\mathrm{t\bar{t}H}$, $\mathrm{H} \to \gamma\gamma$ channel, allowing for negative κ_{t} [25] and a measurement constraining anomalous Higgs-gluon interactions in a EFT framework [152]. A negative κ_{t} can also enhance the cross section of t-channel single top plus Higgs production (tHq). Searches for this processes were performed by CMS [153]. In [154], a pseudoscalar top-

Higgs coupling was considered in the search for tHq. The analysis can set constraints on a pseudoscalar Higgs-boson ($\sigma_{\mathrm{tHq}}^{\tilde{\kappa}_{\mathrm{t}}=1,\kappa_{\mathrm{t}}=0} < 26 \times \sigma_{\mathrm{tHq}}^{\mathrm{SM}}$ at 95% CL) but is not able to distinguish scalar from pseudoscalar coupling.

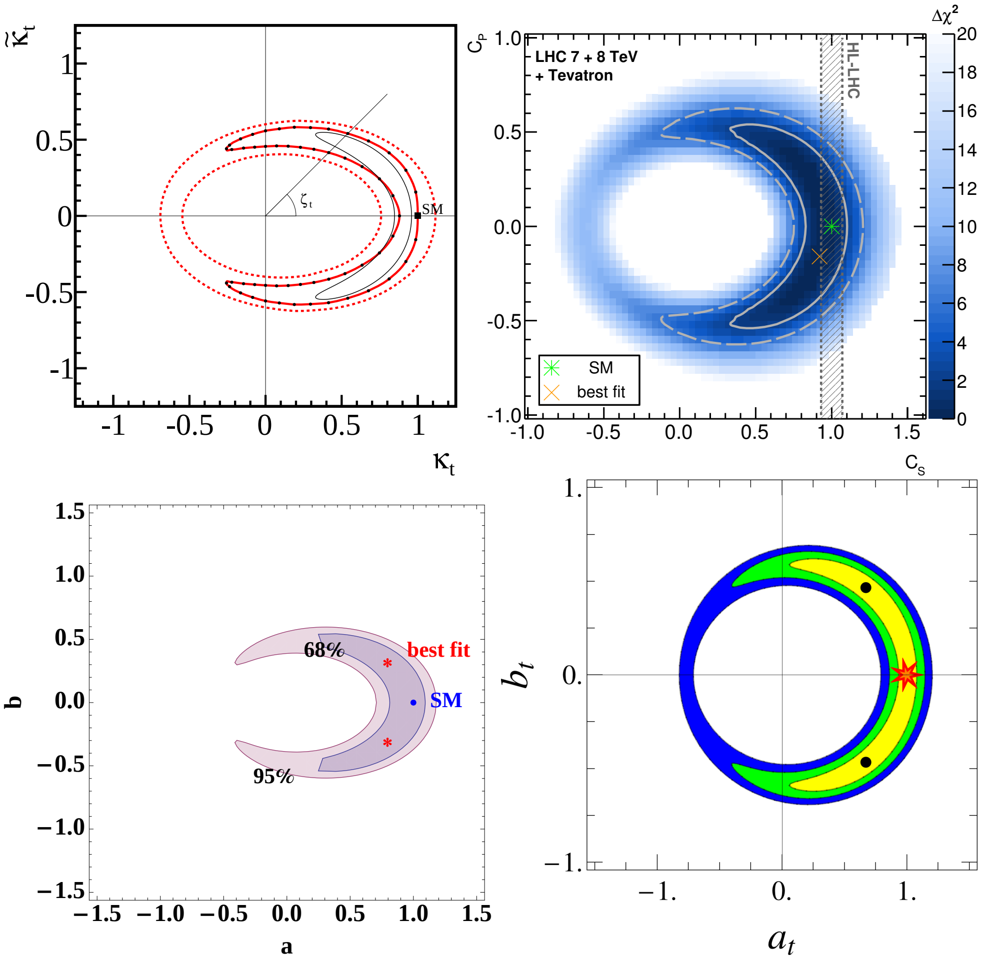


Figure 5.6: Constraints of the top-Higgs coupling modifiers κ_t and $\tilde{\kappa}_t$, taken from [146] (top left), [148] (top right), [147] (bottom left), and [140] (bottom right). The constraints are derived from Higgs signal strength measurements. Only in the bottom right plot $t\bar{t}H$ measurements are included. In the top left plot the solid (dashed) red line corresponds to the regions allowed at 68% (95%) confidence level while the black line is the contour from figure 5.5. In the top right plot the solid (dashed) gray line corresponds to the 68% (95%) confidence level regions and best fit and SM value are indicated by orange and green stars. Here κ_t and $\tilde{\kappa}_t$ are denoted as C_S and C_P. In the bottom left plot 68% (95%) confidence regions are shown as blue (red) areas and best fit (SM) values as red (blue) points. The coupling modifiers are denoted as a and b on the axes. In the bottom right plot the yellow, green, and blue areas correspond to the 68%, 95%, and 99.7% confidence level regions. Best fit values are marked by black dots and the SM value by a red star. The coupling modifiers are denoted as a_t and b_t here.

6 Indirect top-Higgs coupling analysis with CMS data

In this chapter, the extended κ-framework introduced in Section 5.2.3 is implemented within the CMS coupling analysis framework (Section 6.1). It is then used to extract constraints on anomalous top-Higgs couplings under different model assumptions (Section 6.2). These results are summarized in Section 6.3.

6.1 Analysis framework

6.1.1 Production and decay channels considered

A pseudoscalar admixture to the top-Higgs coupling can lead to non-trivial changes in interactions involving the Higgs boson. For physics at the LHC, the most important effects of the modified coupling, and the only effects considered here, are in Higgs boson production and decay channels to which the top-Higgs coupling contributes at leading order. Of these the following processes are considered in this analysis:

- ggH (Higgs boson production in gluon fusion)
- $\mathrm{t\bar{t}H}$ (Associated production of Higgs boson and a top quark pair)
- tHq (t-channel production of Higgs boson and a top quark)
- $\mathrm{H} \to \mathrm{gg}$ (Higgs boson decay into gluons)
- $\mathrm{H} \to \gamma\gamma$ (Higgs boson decay into photons)

The production of a single top quark in association with a Higgs boson is only considered as a background process. In the SM, the cross section for this family of processes is an order of magnitude smaller than $\mathrm{t\bar{t}H}$ production. The cross section for t-channel production of a single top quark and a Higgs boson (tHq) is hereby suppressed by an interference between contributions of the top-Higgs and W-Higgs coupling. An anomalous top-Higgs coupling can increase the tHq cross section significantly for some regions of the coupling parameter space (see, e.g. [155] and [146]) so that dedicated searches exist. Unfortunately, the experimental signature of this process is similar to $\mathrm{t\bar{t}H}$ production and the CMS analysis [153] uses an event selection that has a large overlap with the $\mathrm{t\bar{t}H}$ analysis [22]. Thus a combination of the results is not easily possible. On the other hand, the tHq cross section can become comparable to the $\mathrm{t\bar{t}H}$ production cross section for an anomalous coupling so that it can be relevant as a background or even contribute to the signal in the $\mathrm{t\bar{t}H}$ analyses. This effect is considered here by rescaling the tHq background yield accordingly in analyses in which it can contribute background. For the remaining single top plus Higgs production processes the enhancement with respect to the tiny SM cross section is smaller so that the effect due to a pseudoscalar coupling admixture can be ignored.

The top-Higgs coupling also affects the decay $\mathrm{H} \to \gamma\mathrm{Z}$. Since CMS is far less sensitive to this channel than to $\mathrm{H} \to \gamma\gamma$ (see [156]) and it hardly adds useful information for this analysis, it is not considered. The modification of the ZH gluon fusion production cross section, to which the top-Higgs coupling contributes only sub-dominantly, is neglected because gluon fusion is responsible for less than 10% of ZH production [157].

To calculate the effect of couplings on branching ratios, one has to take into account modifications of the total width of the Higgs boson $\Gamma_{\mathrm{tot}} = \sum_y \Gamma_{H\to y}$:

$$\mathcal{B}_{H\to x} = \frac{\Gamma_{H\to x}}{\Gamma_{\mathrm{tot}}}. \tag{6.1}$$

This indirect effect on branching ratios is dominated by $\mathrm{H} \to \mathrm{gg}$ decays whose top-Higgs coupling induced decay rate is more than three orders of magnitude stronger than the one of $\mathrm{H} \to \gamma\gamma$ and $\mathrm{H} \to \gamma\mathrm{Z}$, which also justifies neglecting the effect of the top-Higgs coupling on the latter process.

6.1.2 Calculation of the top-Higgs coupling dependence of cross sections and partial widths

The inclusion of the anomalous top-Higgs coupling into the κ-framework requires to modify the cross section and branching ratio scaling with respect to the implementation of [8] that is summarized in Table 5.1. Omitting minor contributions due to small couplings, the size of the relevant cross sections and branching ratios is proportional to the following expressions:

$$\Gamma_{\mathrm{H}\gamma\gamma} \propto \kappa_{\mathrm{t}}^2 \Gamma_{\mathrm{H}\gamma\gamma}^{\mathrm{tt}} + \kappa_{\mathrm{W}}^2 \Gamma_{\mathrm{H}\gamma\gamma}^{\mathrm{WW}} + \kappa_{\mathrm{t}}\kappa_{\mathrm{W}} \Gamma_{\mathrm{H}\gamma\gamma}^{\mathrm{tW}} + \tilde{\kappa}_{\mathrm{t}}^2 \tilde{\Gamma}_{\mathrm{H}\gamma\gamma}^{\mathrm{tt}} \tag{6.2}$$

$$\Gamma_{\mathrm{Hgg}} \propto \kappa_{\mathrm{t}}^2 \Gamma_{\mathrm{Hgg}}^{\mathrm{tt}} + \kappa_{\mathrm{b}}^2 \Gamma_{\mathrm{Hgg}}^{\mathrm{bb}} + \kappa_{\mathrm{t}}\kappa_{\mathrm{b}} \Gamma_{\mathrm{Hgg}}^{\mathrm{tb}} + \tilde{\kappa}_{\mathrm{t}}^2 \tilde{\Gamma}_{\mathrm{Hgg}}^{\mathrm{tt}} \tag{6.3}$$

$$\sigma_{\mathrm{ggH}} \propto \kappa_{\mathrm{t}}^2 \sigma_{\mathrm{ggH}}^{\mathrm{tt}} + \kappa_{\mathrm{b}}^2 \sigma_{\mathrm{ggH}}^{\mathrm{bb}} + \kappa_{\mathrm{t}}\kappa_{\mathrm{b}} \sigma_{\mathrm{ggH}}^{\mathrm{tb}} + \tilde{\kappa}_{\mathrm{t}}^2 \tilde{\sigma}_{\mathrm{ggH}}^{\mathrm{tt}} \tag{6.4}$$

$$\sigma_{\mathrm{tHq}} \propto \kappa_{\mathrm{t}}^2 \sigma_{\mathrm{tHq}}^{\mathrm{tt}} + \kappa_{\mathrm{W}}^2 \sigma_{\mathrm{tHq}}^{\mathrm{WW}} + \kappa_{\mathrm{t}}\kappa_{\mathrm{W}} \sigma_{\mathrm{tHq}}^{\mathrm{tW}} + \tilde{\kappa}_{\mathrm{t}}^2 \tilde{\sigma}_{\mathrm{tHq}}^{\mathrm{tt}} \tag{6.5}$$

$$\sigma_{\mathrm{t\bar{t}H}} \propto \kappa_{\mathrm{t}}^2 \sigma_{\mathrm{t\bar{t}H}}^{\mathrm{tt}} + \tilde{\kappa}_{\mathrm{t}}^2 \tilde{\sigma}_{\mathrm{t\bar{t}H}}^{\mathrm{tt}} \tag{6.6}$$

The absolute normalization is given by the requirement that the SM cross section is recovered by setting all $\kappa_i = 1$ and $\tilde{\kappa}_{\mathrm{t}} = 0$. It can be noted that the total cross sections are CP-even observables and do not depend on terms linear in $\tilde{\kappa}_{\mathrm{t}}$. CP-odd observables can be constructed using the decay products and constructing variables sensitive to spin correlations [158].

To implement the behavior described in Equations (6.2) to (6.6), the cross section and decay width contributions $\tilde{\sigma}_{\mathrm{ggH}}^{\mathrm{tt}}$, $\tilde{\sigma}_{\mathrm{t\bar{t}H}}^{\mathrm{tt}}$, $\tilde{\sigma}_{\mathrm{tHq}}^{\mathrm{tt}}$, $\tilde{\Gamma}_{\mathrm{Hgg}}^{\mathrm{tt}}$, and $\tilde{\Gamma}_{\mathrm{H}\gamma\gamma}^{tt}$ of the pseudoscalar top-Higgs coupling have to be calculated. The contributions of the scalar couplings are already implemented in the framework used for the coupling measurements of the LHC Higgs cross section combination [8]. To simplify the calculation of the contribution of the pseudoscalar coupling to cross sections and decay widths, the ratio

$$\lambda = \tilde{\sigma}_{\mathrm{approx}}^{\mathrm{tt}} / \sigma_{\mathrm{approx}}^{\mathrm{tt}} \tag{6.7}$$

Table 6.1: Summary of the contribution of the pseudoscalar top-Higgs coupling to cross sections and branching ratios relevant for this analysis, normalized to the scalar contribution.

Process	$\tilde{\sigma}^{\mathrm{tt}}/\sigma^{\mathrm{tt}}$ or $\tilde{\Gamma}^{\mathrm{tt}}/\Gamma^{\mathrm{tt}}$	Code	Accuracy
$\mathrm{pp} \to \mathrm{H}$	2.35	HIGLU	NNLO QCD
$\mathrm{H} \to \mathrm{gg}$	2.35	HIGLU	NLO QCD
$\mathrm{H} \to \gamma\gamma$	2.35	HDECAY	NLO QCD
$\mathrm{pp} \to \mathrm{t\bar{t}H}$ @ 7 TeV	0.33	MG5_aMC + HC-model	NLO QCD
$\mathrm{pp} \to \mathrm{t\bar{t}H}$ @ 8 TeV	0.35	MG5_aMC + HC-model	NLO QCD
$\mathrm{pp} \to \mathrm{t\bar{t}H}$ @ 13 TeV	0.43	MG5_aMC + HC-model	NLO QCD
$\mathrm{pp} \to \mathrm{tHq}$ @ 8 TeV	0.28	MG5_aMC + HC-model	LO QCD

is calculated with the best accuracy available for both scalar and pseudoscalar coupling at the same time and then multiplied with the contribution of a scalar coupling used in the default framework:

$$\tilde{\sigma}^{\mathrm{tt}} = \lambda \sigma^{\mathrm{tt}}. \tag{6.8}$$

This approach is motivated by the observation that higher order corrections, α_S, and parton distribution functions have similar effects on the scalar and pseudoscalar contributions and thus mostly cancel in the ratio, allowing to calculate precise results for the pseudoscalar process without evaluating multiple PDF sets and even if not all higher order corrections that are available for the scalar process can be calculated.

The cross section for gluon fusion Higgs boson production can be calculated at NNLO QCD with HIGLU [159, 160] for a scalar and a pseudoscalar Higgs boson. The program also allows setting all couplings except for the top-Higgs coupling to zero. The ratio of cross sections $\tilde{\sigma}^{\mathrm{tt}}_{\mathrm{ggH}}/\sigma^{\mathrm{tt}}_{\mathrm{ggH}}$, calculated with the factorization and renormalization scales set to $\mu_F = \mu_R = m_H/2$, the top and Higgs mass $m_H = 125\ \mathrm{GeV}/c^2$ and $m_t = 173\ \mathrm{GeV}/c^2$, and the CTEQ6 [161] parton distribution function is listed in Table 6.1. The leading order result from this program is in agreement with the analytic LO formula given in [140]. When varying the center-of-mass energy between 7 TeV and 13 TeV the ratio changes by less than 1%. The ratio is also very stable with respect to higher oder corrections and hardly changes from LO to NNLO, even though the absolute cross sections changes by 100%, confirming the assumption of canceling higher order corrections. The scale uncertainty, evaluated by varying the renormalization and factorization scale between $m_H/4$ and m_H, also hardly affects the cross section ratio. Finally, it is assumed that the dependency of the ratio on the parton distribution function is also very small, as the scalar as well as the pseudoscalar process are dominantely initiated from a gluon pair with an invariant mass of $125\ \mathrm{GeV}/c^2$.

The HIGLU program also allows to calculate the decay width of the Higgs boson into a pair of gluons, incluing NLO QCD corrections. The mass and scale dependency, evaluated in the same way as above, are found to be neglibile for this process, too. Higgs boson decays into $\gamma\gamma$ can be evaluated with HDECAY [162], also at NLO. The contribution of the top-Higgs coupling to the decay width of a scalar and a pseudoscalar boson are calculated with a small modification to the code.[1]

For the $\mathrm{t\bar{t}H}$ process, cross sections can be calculated with the HIGGS CHARACTERIZA-

---

[1] The recipe was kindly provided by M. Spira.

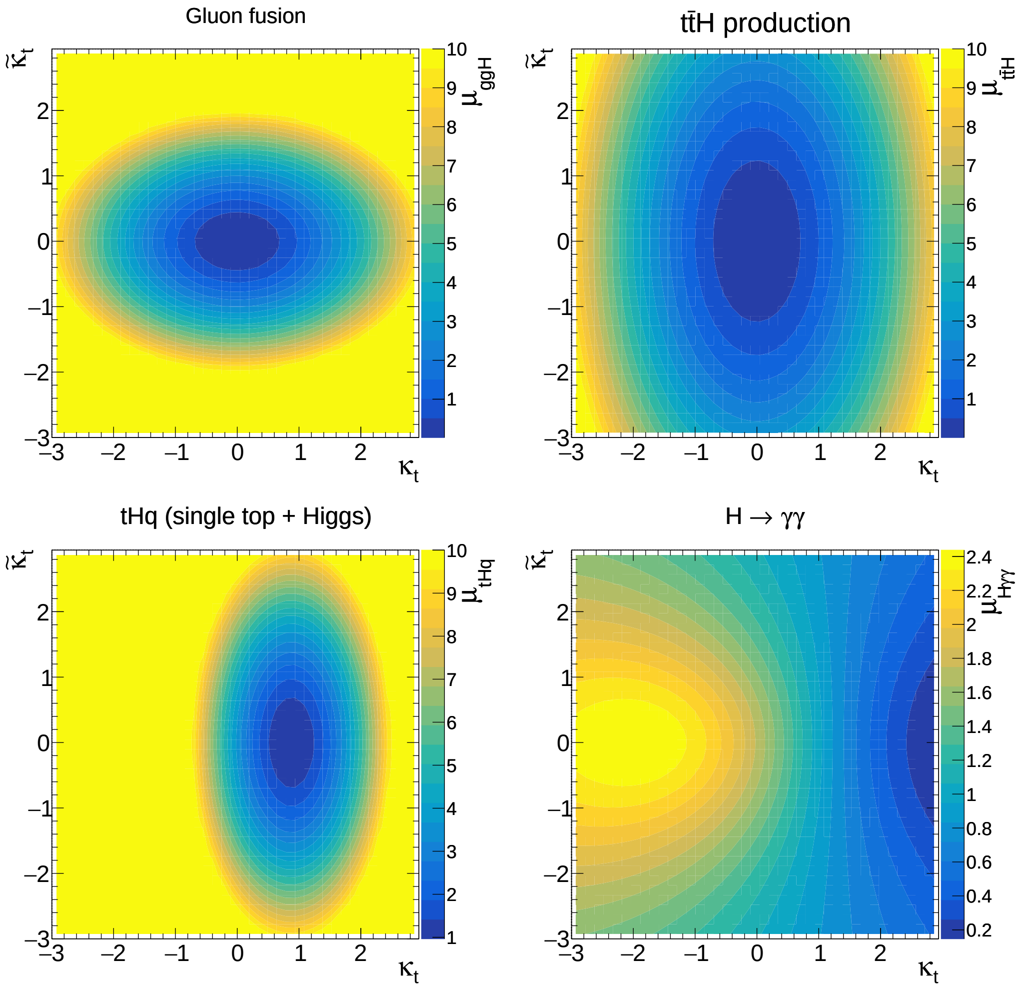


Figure 6.1: Higgs boson cross sections and branching ratios compared to the Standard Model prediction as a function of the top-Higgs coupling modifiers κ_{t} and $\tilde{\kappa}_{\mathrm{t}}$.

tion model [163] in MadGraph5_aMC@NLO at next-to-leading order QCD [164,165]. The cross section ratios are calculated with the factorization and renormalization scales set to $\mu_R = \mu_F = m_t + m_H/2$ and using the PDF4LHC PDF recommendation [166] as implemented in LHAPDF [167] for the proton structure. The result is close to the one calculated in [165] and the result for the scalar cross section is in agreement with the LHC Higgs cross section group result [81]. The cross section ratio $\tilde{\sigma}^{\mathrm{tt}}_{\mathrm{t\bar{t}H}}/\sigma^{\mathrm{tt}}_{\mathrm{t\bar{t}H}}$ is less robust with respect to scale variations than the above results, varying factorization and renormalizations scales up and down by a factor of two changes it by 2% to 3%. However, this uncertainty is small compared to other uncertainties associated with $\mathrm{t\bar{t}H}$ production measurements, so that it seems negligible. Finally, the tHq background scaling is calculated with MadGraph5_aMC@NLO and the Higgs Characterization model, too. Because of its small contribution, the calculation is performed in LO only. The calculated cross sections depend on both the W-Higgs and the top-Higgs coupling but the top contribution can be extracted under the assumption of the structure given in Equation (6.5).

The ratio of the contributions of a pseudoscalar and a scalar top-Higgs coupling to the analyzed cross sections and decay rates is summarized in Table 6.1. In Fig. 6.1, the resulting size of the cross sections and the H($\gamma\gamma$) branching ratio as a function of κ_{t} and

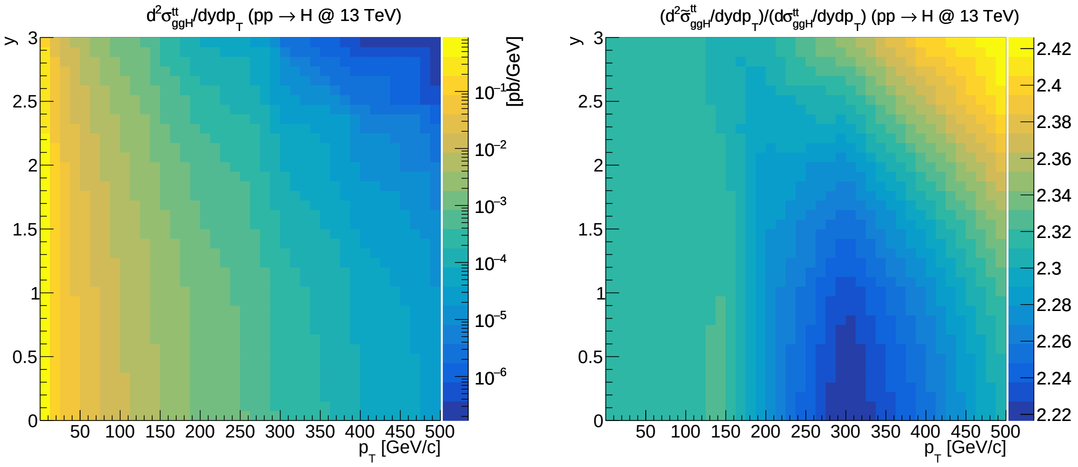


Figure 6.2: In the left plot, the differential cross section for gluon fusion production of a SM Higgs boson at 13 TeV is shown as a function of transverse momentum and rapidity, setting all Higgs-couplings except the coupling to the top quark to zero. In the right plot, the ratio of the cross section of a pseudoscalar boson and the scalar boson is shown, also only considering the top quark contribution. Calculations are performed with HIGLU [159, 160] at NLO QCD with a finite top mass.

$\tilde{\kappa}_\mathrm{t}$, compared to the Standard Model values is visualized. The different shape of contours of equal signal strength will allow to disentangle the effect of κ_t and $\tilde{\kappa}_\mathrm{t}$.

6.1.3 Changes in acceptance due to modified coupling structure

Modifying the Lorentz structure of the top-Higgs coupling not only affects total Higgs boson cross sections and branching ratios but also differential distributions. To be able to re-interpret signal strength measurements that are extracted under the Standard Model assumption, it has to be verified that the kinematics of Higgs boson production are not changed in a way that would affect the acceptance of the analyses and thus the measured signal strength. These changes could be introduced, e.g. by a modified distribution in pseudorapidity which would lead to a different number of events outside of the visible phase space or due to a change in kinematics that changes the output of a multivariate discriminant.

The gluon-fusion Higgs-boson production cross section can be calculated differentially in rapidity and transverse momentum with HIGLU. The calculation includes NLO QCD corrections and top-mass effects. In Fig. 6.2 the differential cross section for a scalar coupling is shown on the left. Higgs boson production in gluon fusion dominantly happens at low p_T. The right figure shows the ratio between the differential cross section for a pseudoscalar boson divided by the distribution for a scalar boson. The ratio is almost flat, especially in the low p_T region where most events are expected. Ignoring the effect on analysis acceptances seems to be justified.

The situation is different in the case of $\mathrm{t\bar{t}H}$ production, in which the top-Higgs coupling is connected to three on-shell particles that can have manifold kinematic configurations. These effects are studied using leading order simulations for different coupling mixings, generated with MadGraph5_aMC@NLO and the Higgs characerization model [163]. Some kinematic distributions, for a center-of-mass energy of 13 TeV, are shown in Fig. 6.3. An important effect is that the transverse boost of the Higgs boson is on average higher

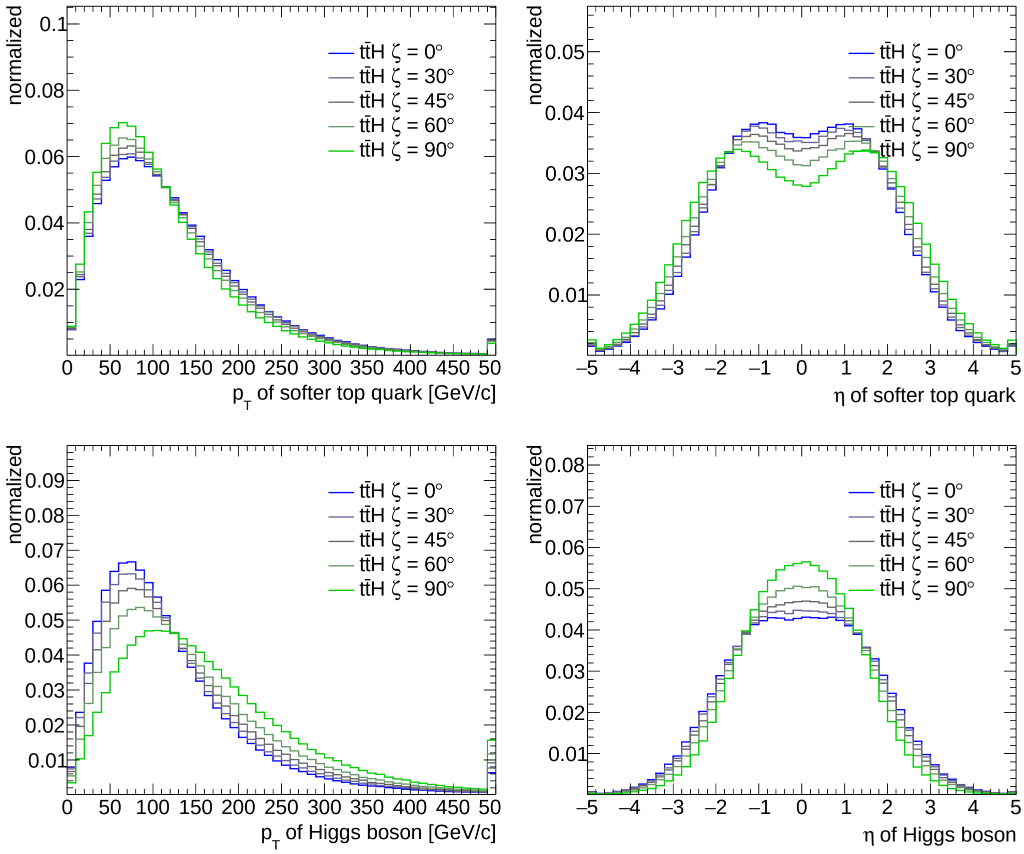


Figure 6.3: Dependence of the shape of differential $t\bar{t}$H production cross sections on the scalar-pseudoscalar mixing angle ζ_t at 13 TeV proton-proton center-of-mass energy. The distributions are generated in LO using MadGraph5_aMC@NLO [168] and the Higgs characterization framework [163], without simulating decay or parton shower. Blue lines correspond to a scalar, green to a pseudoscalar Higgs boson.

and that is it is produced more centrally if the pseudoscalar admixture increases. This can affect the interpreation of measurements, e.g. in the $t\bar{t}$H, H $\rightarrow \gamma\gamma$ channel, in which cuts on the photon momenta are used in the event selection. An opposite effect can be observed for the softer of the two top quarks, which becomes less boosted and central. This can also be important for the event selection of the $t\bar{t}$H analyses which rely on the central reconstruction of high p_{T} jets and leptons from the top quark decay.

That being said, the acceptance differences due to the coupling structure are estimated to be at most 10% to 20%. The apparent kinematic differences are reduced by the decay of Higgs boson and top quarks. For the 13 TeV $t\bar{t}$H($b\bar{b}$) analysis the effect is of the order 10% as shown later in this thesis. Implementing corrections for the coupling-structure dependence of the Run 1 $t\bar{t}$H measurements is hardly feasible but using the $t\bar{t}$H results is interesting as the dependence of the $t\bar{t}$H production cross sections on anomalous top-Higgs couplings is complementary the gluon fusion dependence. Given the large uncertainties of the $t\bar{t}$H measurements using them seems to be justified to get an idea of their impact on a combined fit, keeping the limitations of this approach in mind. The following analysis is performed both including and excluding the $t\bar{t}$H measurements.

6.2 Anomalous top-Higgs coupling constraints

6.2.1 Input analyses

To calculate constraints on anomalous top-Higgs couplings, the following CMS results, which are also part of the LHC combination [8], are re-analyzed:

- H → $\gamma\gamma$ [169]
- H → 4l [136]
- H → WW [170]
- H → ττ [171]
- H → $b\bar{b}$ [172]
- $t\bar{t}H$ [22, 173]

The sensitivity to a 125 GeV/c^2 Higgs boson differs between the analyses: the first three observe excesses with a significace of the order 5σ and more, which is compatible with the SM expectation. The ττ analysis can claim evidence with a 3σ signal while the excess in the $b\bar{b}$ channel is not yet significant. An interesting case is the $t\bar{t}H$ analysis, which sees a 3σ signal even though much less is expected from the SM.

Most of the analyses are grouped by decay channel but can also distinguish between production processes. The $\gamma\gamma$ analysis is expected to be most senstive to the dominating ggH production, less to VBF production, and the analysis categories targeting VH and $t\bar{t}H$ production lack the statistics to make precise measurements. This is similar for the WW analysis, which has however even worse acceptance for $t\bar{t}H$ production. The low-event-count four lepton analysis is sensitive almost only to ggH production. Contrary, the H(ττ) measurements are most precise in categories that are dominated by VBF events while also being able to probe ggH and, to some extent, VH events. The H($b\bar{b}$) analysis is unique in that it is sensitive only to VH produced events because leptons from W or Z boson decays are required to distinguish the events from the overwhelming QCD background. The analyses optimized for $t\bar{t}H$ production have hardly any sensitivity to other production modes and target mainly $b\bar{b}$, $\gamma\gamma$, and WW decays. Since the analyses can observe different combinations of production and decay processes, it is thus possible to constrain both, cross sections and branching ratios, in a combined fit.

These analyses have already been interpreted as part of the LHC Higgs combination [8], yielding among others the coupling strength results discussed in the previous chapter. For the interpretation of the results presented in the next section, the measurements of the signal strengths of the five most important production and decay modes are also instructive. These results are obtained by keeping the branching ratios at the SM values and fitting the production cross sections or, vice versa, fixing production to the SM values and fitting the the most important branching ratios. The results are shown in Table 6.2. To illustrate the sensitivity of the different production and decay modes to the top-Higgs coupling, the expected scaling of the different signal strength in the coupling model constructed in the previous section is shown, too.

6.2.2 Statistical model

The modified κ-model used to measure the $t\bar{t}H$ coupling contains six or eight parameters of interest related to Higgs boson couplings, depending on whether or not effective gluon

Table 6.2: Signal strengths μ normalized to the Standard Model expectation as measured by CMS, from [8]. Cross section (branching ratio) signal strengths are calculated by fixing branching ratios (signal strengths) to the SM Value. In the last column the approximate parameterization used to express these signal strengths in the anomalous top-Higgs coupling measurement is shown to help interpreting the constraints that are calculated for κ_t and $\tilde{\kappa}_t$.

Signal strength	CMS measurement	Parametrization for κ_t-$\tilde{\kappa}_t$ measurement
μ_{ggF}	$0.84^{+0.18}_{-0.16}$	$1.06\boldsymbol{\kappa}_t^2 + 0.01\kappa_b^2 - 0.07\kappa_t\kappa_b + 2.49\tilde{\boldsymbol{\kappa}}_t^2$
μ_{VBF}	$1.14^{+0.37}_{-0.34}$	$0.74\kappa_W^2 + 0.26\kappa_Z^2$
μ_{WH}	$0.46^{+0.57}_{-0.53}$	κ_W^2
μ_{ZH}	$1.35^{+0.58}_{-0.54}$	κ_Z^2
$\mu_{\mathrm{t\bar{t}H}}$	$2.9^{+1.0}_{-0.9}$	$\boldsymbol{\kappa}_t^2 + 0.35\tilde{\boldsymbol{\kappa}}_t^2$
$\mu_{\mathrm{H\gamma\gamma}}$	$1.11^{+0.25}_{-0.23}$	$(1.59\kappa_W^2 + 0.07\boldsymbol{\kappa}_t^2 - 0.66\kappa_W\kappa_t + 0.16\tilde{\boldsymbol{\kappa}}_t^2)/\mu_{\mathrm{tot.width}}$
μ_{HZZ}	$1.04^{+0.32}_{-0.26}$	$\kappa_Z^2/\mu_{\mathrm{tot.width}}$
μ_{HWW}	$0.90^{+0.23}_{-0.21}$	$\kappa_W^2/\mu_{\mathrm{tot.width}}$
$\mu_{\mathrm{H\tau\tau}}$	$0.88^{+0.30}_{-0.28}$	$\kappa_\tau^2/\mu_{\mathrm{tot.width}}$
μ_{Hbb}	$0.81^{+0.45}_{-0.43}$	$\kappa_b^2/\mu_{\mathrm{tot.width}}$
$\mu_{\mathrm{tot.width}}$	not considered	$0.58\kappa_b^2 + 0.22\kappa_W^2 + 0.21\tilde{\boldsymbol{\kappa}}_t^2 + 0.09\boldsymbol{\kappa}_t^2$ $+0.06\kappa_\tau^2 + 0.03\kappa_Z^2 + 0.03\kappa_c^2 - 0.01\kappa_b\kappa_t$

and photon couplings are used: κ_Z, κ_W, κ_t, $\tilde{\kappa}_t$, κ_b, κ_τ (and κ_γ, κ_g). The Higgs boson mass chosen to evaluate the fit is $m_H = 125.1\ \mathrm{GeV}/c^2$, in accordance with the value of $125.09 \pm 0.24\ \mathrm{GeV}/c^2$ obtained in the LHC Higgs boson mass combination [7].

One small modification with respect to the LHC combination is that κ_t is allowed to be negative in this analysis and the remaining couplings are required to be positive. Contrary, in the LHC combination, the κ_t coupling was the only one required to be positve. Of course, only the relative sign of the coupling compared to other couplings is important, as it decides whether interference terms are destructive or constructive. The most important interference for this analysis is the one in the $\mathrm{H} \to \gamma\gamma$ decay, and for this one it is basically irrelevant whether the vector boson or the top quark coupling has a flipped sign. A small difference can arise from the b-quark interference in gluon fusion production. Moreover, in the LHC combination, the rate of $\mathrm{H} \to \mathrm{c\bar{c}}$ is, for simplicity, also controlled by the κ_t coupling. Since this analysis specifically targets a non-standard top-Higgs coupling, a different treatment was chosen and the rate left at the Standard Model value.

In addition to the parameters of interest, the statistical model contains more than two thousand nuisance parameters $\boldsymbol{\theta}$, modeling the effect of systematic uncertainties. Sources of uncertainties modeled by the nuisance parameters are theoretical uncertainties from the predictions of total and differential cross sections of signals and backgrounds, experimental uncertainties, and statistical uncertainties due to the rate predictions determined in control regions as well as due to limited MC statistics for background predictions. Every nuisance parameter models an independent uncertainty. Some uncertainties, like signal cross section uncertainties, affect all analyses but most only certain categories, and some only single analysis bins. The nuisance parameters are constrained by likelihood functions, describing the prior knowledge of these paramters, from measurements in control regions or from

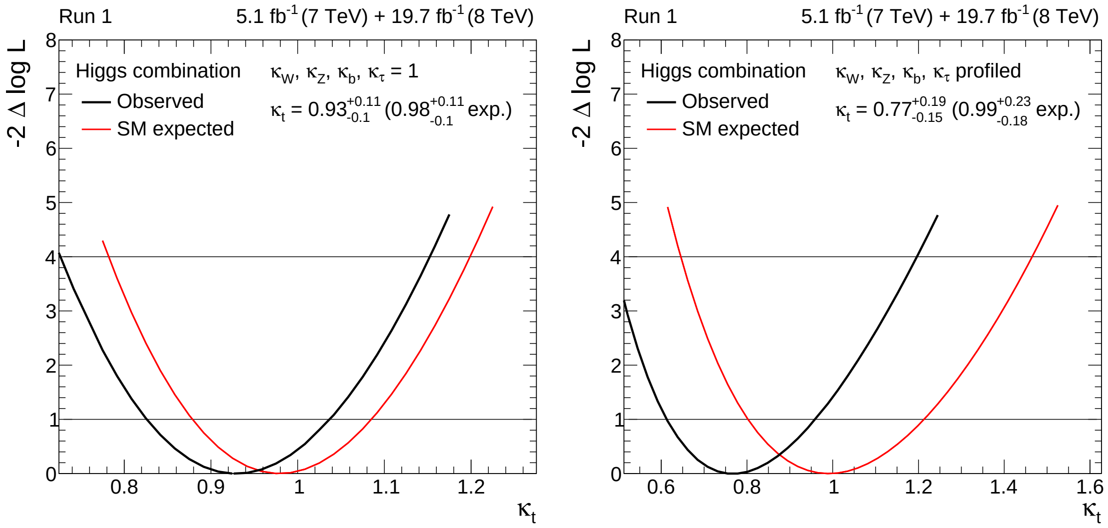


Figure 6.4: Profile likelihood ratio test statistic $-2\log(L(\kappa_\mathrm{t},\hat{\hat{\boldsymbol{\theta}}})/L(\hat{\kappa}_\mathrm{t},\hat{\boldsymbol{\theta}}))$ obtained from an analysis of the CMS Higgs boson measurements in LHC Run 1 with the "resolved" κ-model. For the right figure the likelihood is minimized with respect to the nuisance parameters as well as κ_Z, κ_W, κ_b, κ_τ, for the left plot these coupling modifiers are fixed to 1. The black like is the result for the observed data, red the expected result calculated using the Asimov dataset. The intersection with the horizontal lines at one (four) indicate the borders of the 1σ (2σ) confidence intervals.

theoretical assumptions. An overview of the technical details is given in Section 4.2.

All results in this chapter are obtained by using the profile likelihood ratio test statistic

$$q(\boldsymbol{\kappa}) = -2\log\left\{\frac{L(\boldsymbol{\kappa},\hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{\theta}})}\right\}, \tag{6.9}$$

in which $\boldsymbol{\kappa}$ can be one or multiple parameters associated with the top-Higgs coupling and $\boldsymbol{\theta}$ represents all nuisance parameters and the Higgs boson couplings that are profiled in the fit. In the numerator the likelihood is evaluated with the maximum likelihood estimate of the nuisance parameter for a fixed $\boldsymbol{\kappa}$, in the denominator the likelihood is maximized with respect to both $\boldsymbol{\kappa}$ and $\boldsymbol{\theta}$. Confidence intervals are calculated assuming that $q(\boldsymbol{\kappa})$ follows a N-dimensional χ^2 distribution, N being the dimensionality of $\boldsymbol{\kappa}$. Expected results are obtained from the Asimov dataset, which is essentially the nominal prefit model. The use of these approximations is also motivated in Part I.

6.2.3 Resolved model and scalar coupling

As a first test, scans of all coupling strength parameters are performed in the resolved model. Results are in agreement with the fits performed as part of the LHC coupling combination [8][2]. The measurement of the κ_t parameter under two assumptions is shown in Fig. 6.4, for now without allowing for a pseudoscalar coupling. For the result shown on the left the κ_Z, κ_W, κ_b, κ_τ coupling modifiers are set to their Standard Model values, yielding $\kappa_\mathrm{t} = 0.93^{+0.11}_{-0.10}$. The result can be explained in terms of the signal strengths given

---

[2] The reproduction of this results was possible thanks to the great work done in the CMS Higgs combination group and the instructions kindly provided by Andrew Gilbert.

in Table 6.2. The lack of events in gluon fusion channels and the excess in the $H \to \gamma\gamma$ channel (compared to other decay channels) prefer a coupling smaller than one. Only $t\bar{t}H$ pulls in the other direction, with a larger uncertainty and thus smaller weight. For the right figure no assumptions on the strength of κ_Z, κ_W, κ_b, κ_τ are made, yielding a result of $\kappa_t = 0.77^{+0.19}_{-0.15}$. This reduction of κ_t can be understood from the lower than expected signal strength in fermion decay channels, as it leads to increased boson branching ratios, which leads to an interpretation of the signal strength measured in the $H \to$ bosons and $t\bar{t}H \to$ multilepton analyses as coming from a large branching ratio but relatively smaller κ_t.

The main difference between these results and the ones published in [8] are the confidence intervals quoted. Different to here, in the official result they are corrected in the case of two disjoint allowed regions, which are sometimes possible for positive and negative coupling values. Furthermore, due to a technical problem regarding the construction of the Asimov dataset in the $H \to 4l$ analysis, the "expected" best-fit values obtained using this data set are not exactly at unity.

6.2.4 Resolved model with anomalous top-Higgs coupling

In this subsection the modified resolved coupling framework is employed, allowing for $\tilde{\kappa}_t \neq 0$ to model contributions of a pseudoscalar top-Higgs coupling, but still not allowing for BSM contributions to Higgs production or decay. The remaining Higgs boson coupling modifiers κ_Z, κ_W, κ_b, and κ_τ are included in the maximum likelihood fit. This can be interpreted as a model in which the top-Higgs coupling is anomalous but the remaining Higgs couplings are scalar in nature, with arbitrary strength. As discussed previously, this model can be motivated by the tight constraints on anomalous couplings to vector bosons and the inability of the experiments to distinguish scalar from pseudoscalar couplings in Higgs boson decays to fermions.

Results of the two-dimensional likelihood scan of κ_t and $\tilde{\kappa}_t$ are shown in Fig. 6.5. On the left-hand side $t\bar{t}H$ production measurements are not included in the analysis. The result can be understood by comparison with the signal strength associated with the values of κ_t and $\tilde{\kappa}_t$, displayed in Fig. 6.1. For gluon fusion production, contours with constant cross sections are ellipses with horizontal major axes in the κ_t-$\tilde{\kappa}_t$ plane. Considering only ggH results, allowed regions would thus be expected to be bordered by such ellipses. However, the inclusion of $\gamma\gamma$ measurements gives a preference of $\kappa_t = 1$ over $\kappa_t = -1$, as the latter would lead to a factor of two enhancement of the $H(\gamma\gamma)$ signal strength.

In the right of the figure, $t\bar{t}H$ measurements are included in the fit. They have a large impact on the allowed parameter space. The measured signal strength $\mu_{t\bar{t}H}$ is significantly above the Standard Model (see Table 6.2). This is in tension with the result of the non-$t\bar{t}H$ fit on the left, even if the pseudoscalar admixture is small. A purely pseudoscalar coupling that is allowed by the gluon fusion measurements, e.g. $(\kappa_t,\tilde{\kappa}_t)=(0,0.4)$, would result in a almost vanishing $t\bar{t}H$ production cross section, as can be derived from the formulas of Table 6.2.

The two-dimensional constraints are compared to the SM expectation in Fig. 6.6. The best-fit values are compatible with the SM at the 1σ level. Due to the μ_{ggH} signal strength, which is below the SM expectations, allowed regions are closer to the origin of the coordinate system than expected. As long as $t\bar{t}H$ results are excluded, the comparatively large $\mu_{H\gamma\gamma}$ leads to a small preference of a pseudoscalar coupling over a scalar coupling because only the scalar coupling interferes destructively with the W boson contribution in $H \to \gamma\gamma$, while $\tilde{\kappa}_t$ always increases the signal strength. Because of the large excess in the $t\bar{t}H$

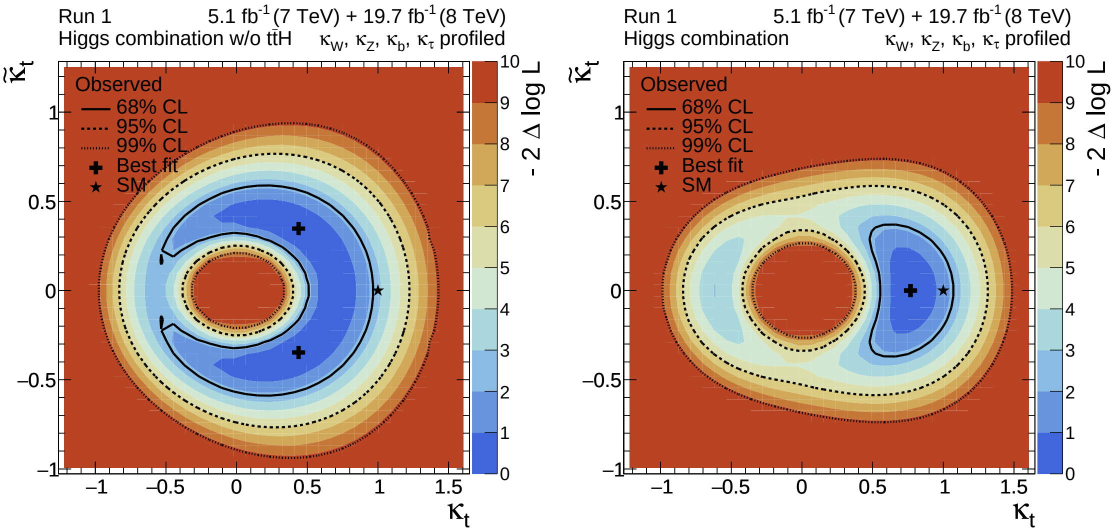


Figure 6.5: Profile likelihood ratio test statistic $q(\kappa_\mathrm{t}, \tilde{\kappa}_\mathrm{t})$ obtained from an analysis of the CMS Higgs boson measurements in LHC Run 1 with the "resolved" κ-model. The likelihood is minimized with respect to κ_Z, κ_W, κ_b, κ_τ as well as the nuisance parameters. For the left plot $\mathrm{t\bar{t}H}$ analyses are excluded from the coupling measurement. Black lines indicate the borders of 68%, 95%, and 99% confidence level regions for κ_t, and $\tilde{\kappa}_\mathrm{t}$. Black crosses indicate the minima of the likelihood and a black star the Standard Model expectation.

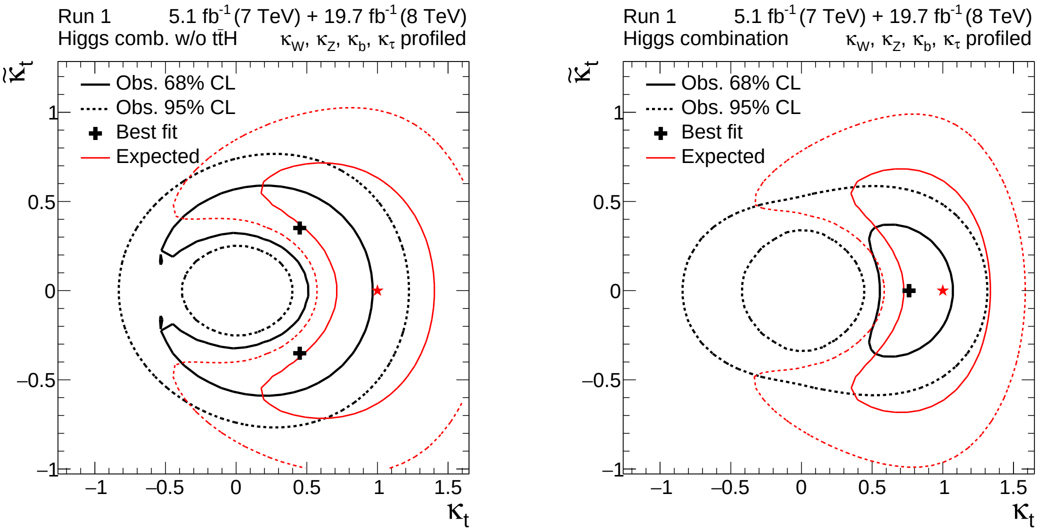


Figure 6.6: Regions allowed at 68% and 95% CL for the top-Higgs coupling parameters κ_t and $\tilde{\kappa}_\mathrm{t}$. The constraints are obtained using the "resolved model" and the likelihood is minimized with respect to κ_Z, κ_W, κ_b, κ_τ as well as the nuisance parameters. Black lines indicate the borders of the regions for the observed data, red lines are obtained from a fit of the Asimov dataset. The best-fit value is marked as a black cross and the SM value as a red star.

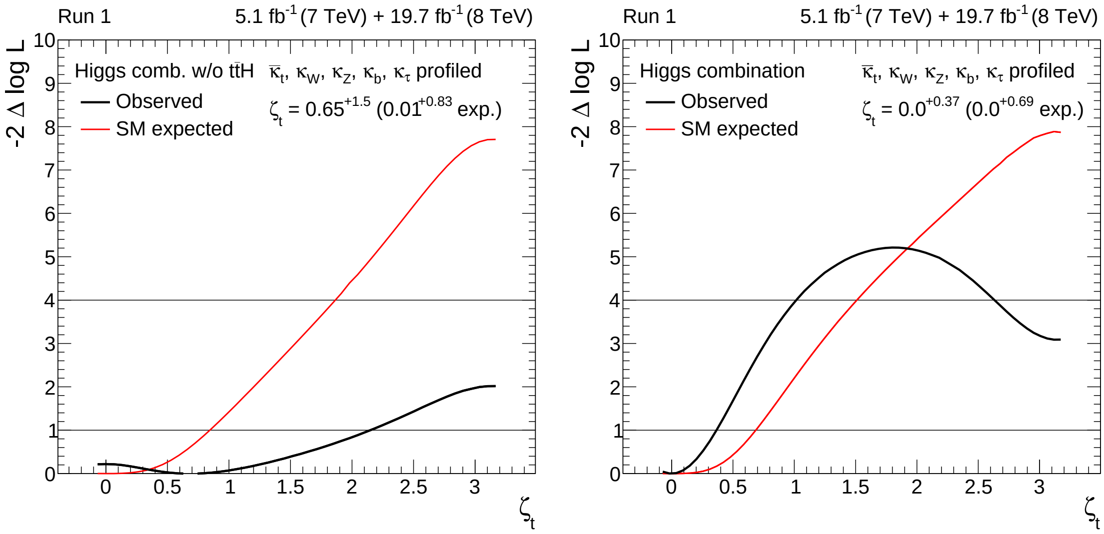


Figure 6.7: Profile likelihood ratio test statistic $q(\zeta_\mathrm{t})$ obtained from an analysis of the CMS Higgs boson measurements in LHC Run 1 with the "resolved" κ-model. The likelihood is minimized with respect to κ_Z, κ_W, κ_b, κ_τ, and $\bar{\kappa}_\mathrm{t}$ as well as the nuisance parameters. For the left plot $\mathrm{t\bar{t}H}$ analyses are excluded from the coupling measurement. The black line is the result for the observed data, red the expected result calculated using the Asimov dataset. The intersection with the horizontal lines at one (four) indicate the borders of the 1σ (2σ) confidence intervals.

channel, its inclusion has a larger impact than expected. As discussed above, the $\mathrm{t\bar{t}H}$ excess favors a purely scalar coupling.

Performing an one-dimensional analysis of $q(\tilde{\kappa}_\mathrm{t})$ and allowing for arbitrary κ_t, one can constrain the pseudoscalar coupling to $|\tilde{\kappa}_\mathrm{t}| < 0.52$ at the 68% CL level without including $\mathrm{t\bar{t}H}$ analyses, while $|\tilde{\kappa}_\mathrm{t}| < 0.57$ is expected. With the inclusion of $\mathrm{t\bar{t}H}$ the constraint is improved to $|\tilde{\kappa}_\mathrm{t}| < 0.25$ ($|\tilde{\kappa}_\mathrm{t}| < 0.54$ expected).

Using the parameterization of Equation (5.3), the scalar-pseudoscalar mixing can be constrained in a fit of the mixing angle ζ_t. An angle $\zeta_\mathrm{t} = 0$ corresponds to a scalar coupling, $\zeta_\mathrm{t} = \pi/2$ to a pseudoscalar coupling, and $\zeta_\mathrm{t} = \pi$ to a scalar top-Higgs coupling with a flipped sign. Not only κ_Z, κ_W, κ_b, and κ_τ are profiled in this analysis but also the strength of top-Higgs coupling $\bar{\kappa}_\mathrm{t}$, parameterizing the radial direction in the κ_t-$\tilde{\kappa}_\mathrm{t}$ plane. The observed and the expected result is displayed in Fig. 6.7. The region of $\zeta_\mathrm{t} < 0$ is not displayed because of the symmetry of the model with respect to positive and negative $\tilde{\kappa}_\mathrm{t}$. Without including $\mathrm{t\bar{t}H}$ measurements the 1σ interval is given by $\zeta_\mathrm{t} < 2.15$ while $\zeta_\mathrm{t} < 0.83$ is expected. A purely pseudoscalar coupling can thus not be excluded, contrary to the expectation. Furthermore, the CMS results exclude $\zeta_\mathrm{t} = \pi$ only at the 1σ level, while more than 2σ is expected. Nonetheless, the result is still compatible with the SM expectation. Again, once the $\mathrm{t\bar{t}H}$ results are added the picture changes drastically, and a pseudoscalar dominated coupling around $\zeta_\mathrm{t} = \pi/2$ is excluded: $|\zeta| < 0.37$ (< 0.59 expexted).

Finally, a fit is performed where κ_Z, κ_W, κ_b, and κ_τ are fixed to unity. This corresponds to a model in which the top-Higgs coupling is the only coupling different from the Standard Model. The result is shown in Fig. 6.8. The parameter space is more constrained in radial direction while the constraint an a pseudoscalar admixture is only moderately tighter in this model: Without considering $\mathrm{t\bar{t}H}$, it is $|\zeta_\mathrm{t}| < 1.25$ (< 0.76 expected) at the 1-σ CL. When including $\mathrm{t\bar{t}H}$ this changes to $|\zeta_\mathrm{t}| < 0.37$ (< 0.59 expected).

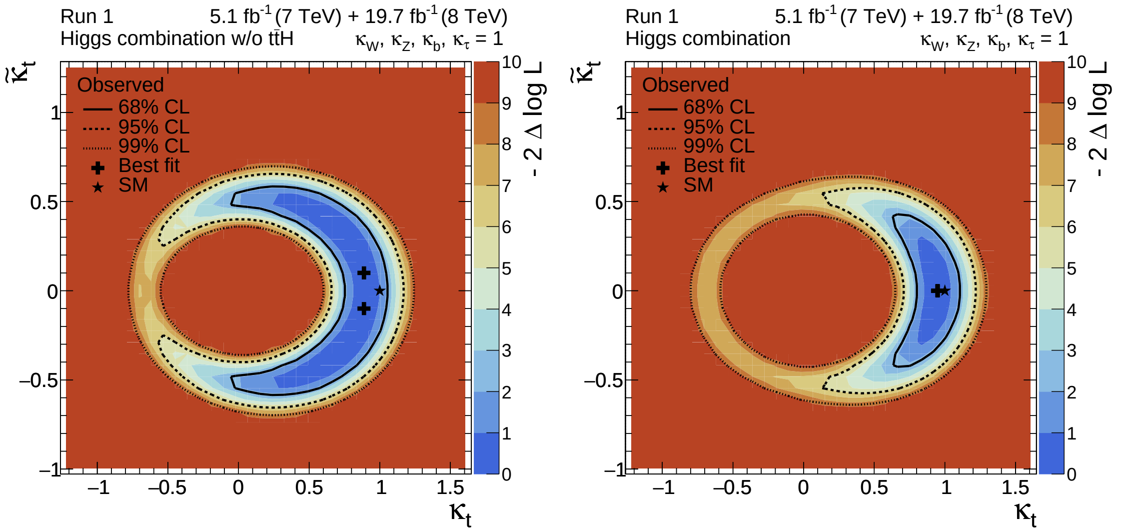


Figure 6.8: Profile likelihood ratio test statistic $q(\kappa_\mathrm{t}, \tilde{\kappa}_\mathrm{t})$ obtained from an analysis of the CMS Higgs boson measurements in LHC Run 1 with the "resolved" κ-model. Higgs boson coupling modifiers not related to the top quarks are fixed to their SM values. The result shown on the left is calculated without including $\mathrm{t\bar{t}H}$ measurements in the analysis. Black lines indicate the borders confidence regions for κ_t, and $\tilde{\kappa}_\mathrm{t}$. Black crosses indicate the minima of the likelihood and a black star the Standard Model expectation.

In general, the result of this fit is closer to the SM expectation. In spirit, the model used to extract these constraints is the same as the ones used to derive the constraints in Fig. 5.6. However, in this analysis only CMS results are used as input and the treatment of systematic uncertainties is more involved. The constraints calculated without the inclusion of $\mathrm{t\bar{t}H}$ measurments are qualitatively in agreement with the ones displayed in Fig. 5.6, keeping in mind that this analysis does not use ATLAS data and the phenomenological publications do not contain all of the CMS results that are analyzed here.

6.2.5 Allowing for BSM contributions to loop-induced processes

Now the model is extended by allowing BSM contributions to loop-induced processes, parameterized by the effective coupling scaling factors κ_γ and κ_g. As discussed before, for the sake of simplicity decays into BSM final states are not considered. The result of a scan of κ_t, with $\tilde{\kappa}_\mathrm{t} = 0$, can be seen on the left in Fig. 6.9. The only real way to constrain κ_t in this parameterization is via direct measurements of $\mathrm{t\bar{t}H}$ production. Since CMS measured $\mu_{\mathrm{t\bar{t}H}} = 2.9^{+1.0}_{-0.9}$, the value measured for κ_t is relatively high. It is slightly smaller than $\sqrt{2.9}$ because the global fit (see Fig. 5.4) also prefers $\kappa_\mathrm{b} < 1$, leading to a larger W branching ratio that in turn is, in this model, responsible for part of the large excess in the $\mathrm{t\bar{t}H}$ multilepton analysis. The small difference with respect to the result in [8] can arise due to the decoupling of $\mathrm{H} \rightarrow \mathrm{c\bar{c}}$ rate from the κ_t parameter.

Once the pseudoscalar coupling parameter $\tilde{\kappa}_\mathrm{t}$ is also allowed to modify the $\mathrm{t\bar{t}H}$ production cross section (right plot in Fig. 6.9), the analysis is no longer able to distinguish between scalar and pseudoscalar coupling, as both can be responsible for $\mathrm{t\bar{t}H}$ production. In this effective coupling model the lack of a gluon fusion or $\mathrm{H}(\gamma\gamma)$ excess over the SM, despite the large $\bar{\kappa}_\mathrm{t}$, is explained as originating in canceling contributions of new physics to

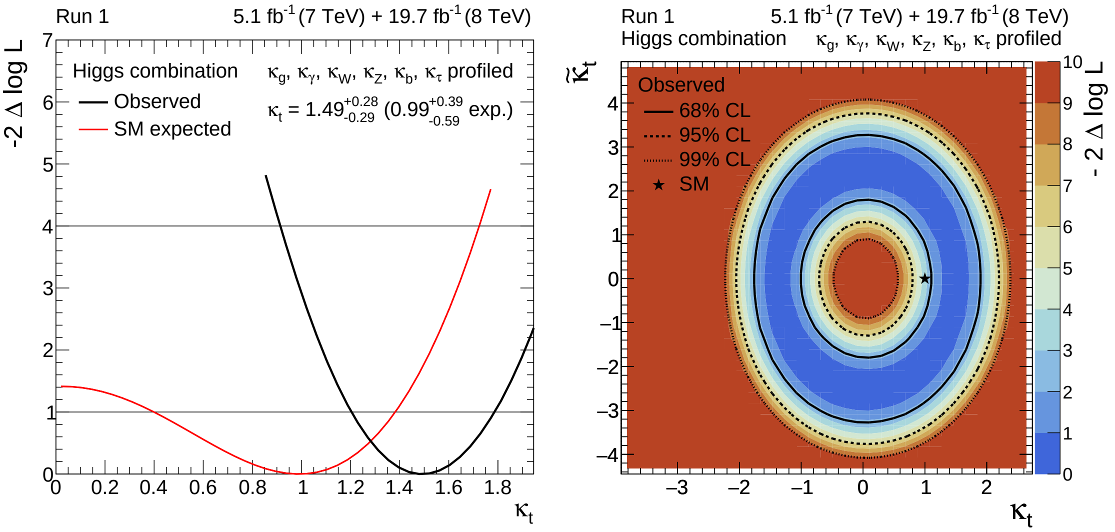


Figure 6.9: The left figure displays the profile likelihood ratio test statistic $q(\kappa_\mathrm{t})$ obtained from an analysis of the CMS Higgs boson measurements in LHC Run 1 with a model including effective couplings κ_g and κ_γ. The likelihood is minimized with respect to all nuisance parameters and scalar non-$\mathrm{t\bar{t}H}$ couplings while $\tilde{\kappa}_\mathrm{t}$ is set to zero. The black like is the result for the observed data, red the expected result calculated using the Asimov dataset. The intersection with the horizontal lines at one (four) indicated the borders of the 1σ (2σ) confidence intervals. The right figure shows the $q(\kappa_\mathrm{t}, \tilde{\kappa}_\mathrm{t})$ test statistic for the same mode. Black lines indicate the borders of confidence regions for κ_t, and $\tilde{\kappa}_\mathrm{t}$.

the loop-induced processes. However, if the top-Higgs coupling was indeed much stronger than predicted by the SM or even pseudoscalar in nature but both gluon fusion and $\mathrm{H} \to \gamma\gamma$ are still as close to the Standard Model as they are observed, a mild conspiracy by the BSM contribution would be required.

Nonetheless, the low precision of the constraints obtained in this model due to large uncertainties from the $\mathrm{t\bar{t}H}$ measurements, the tension between gluon fusion and $\mathrm{t\bar{t}H}$ coupling constraints, and the current inability to constrain the pseudoscalar coupling admixture in $\mathrm{t\bar{t}H}$ should be motivation enough for further analysis of the $\mathrm{t\bar{t}H}$ process with data of the LHC Run 2.

6.3 Summary of constraints

In this chapter anomalous $\mathrm{t\bar{t}H}$ couplings were constrained. In a model in which no assumption on the scalar Higgs boson coupling modifieres κ_Z, κ_W, κ_t, κ_b, and κ_τ are made, the pseudoscalar coupling modifier can be constrained to $|\tilde{\kappa}_\mathrm{t}| < 0.25$ at the 68% CL, while $|\tilde{\kappa}_\mathrm{t}| < 0.54$ is expected. Measurements of κ_t and $\tilde{\kappa}_\mathrm{t}$ are correlated which leads to non-trivial exclusion regions for the two parameters. The pseudoscalar-scalar mixing angle is measured to be $\zeta_\mathrm{t} = 0 \pm 0.37$ ($\zeta_\mathrm{t} = 0 \pm 0.69$ expected). These results and results with the different parameterizations and inputs discussed are summarized in Table 6.3.

With respect to previous publications, e.g. [140, 141, 146–148], several improvements have been made. Cross sections have been calculated including higher-order QCD corrections and considering finite top-mass effects. Since this analysis has been performed in the CMS coupling analysis framework, systematic uncertainties have been treated correlated

Table 6.3: Summary of one-dimensional constraints on an anomalous top-Higgs coupling, expressed either in terms of the pseudoscalar coupling strength $\tilde{\kappa}_t$ or the scalar-pseudoscalar mixing angle ζ_t. The errors correspond to two-sided confidence intervals at 68% confidence level. Three different models discussed in the text are considered, allowing for different degrees of variation in the remaining couplings. The results are quoted with and without considering $t\bar{t}H$ measurements in the analysis.

Non-tH couplings	Input analyses	$\tilde{\kappa}_t$	ζ_t
At SM value	no $t\bar{t}H$	$\lvert\tilde{\kappa}_t\rvert < 0.54$ (< 0.50)	$\lvert\zeta_t\rvert < 1.25$ (< 0.76)
Resolved in loops	no $t\bar{t}H$	$\lvert\tilde{\kappa}_t\rvert < 0.52$ (< 0.57)	$\lvert\zeta_t\rvert < 2.15$ (< 0.83)
Include κ_g, κ_γ	no $t\bar{t}H$	-	-
At SM value	all CMS	$\lvert\tilde{\kappa}_t\rvert < 0.32$ (< 0.46)	$\lvert\zeta_t\rvert < 0.38$ (< 0.63)
Resolved in loops	all CMS	$\lvert\tilde{\kappa}_t\rvert < 0.25$ (< 0.54)	$\lvert\zeta_t\rvert < 0.37$ (< 0.59)
Include κ_g, κ_γ	all CMS	$\lvert\tilde{\kappa}_t\rvert < 3.0$ (< 2.3)	-

wherever appropriate. Furthermore, it is possible to consider the acceptance for all production and decay processes of every analysis channel in this framework. Finally, not only an anomalous top-Higgs coupling was considered in this analysis, but also the couplings to the remaining third-generation fermions as well as to vector bosons were allowed to vary.

For a more precise analysis with a large LHC Run 2 data set it will be neccessary to consider the kinematic effects of a pseudoscalar coupling and to extract the $t\bar{t}H$ signal strength under different coupling mixture assumptions. More precise measurements of all Higgs boson production and decay processes will also increase the sensitivity of this analysis. However, if one really considers new physics in the form of a non-standard top-Higgs couplings, one should also be prepared to expect BSM contributions to, at least, all loop induced processes. In such a scenario anomalous top-Higgs couplings can no longer be constrained by the analysis of signal strengths and even the measurement of a CP-even Higgs boson relies heavily on $t\bar{t}H$ measurements.

III. Direct top-Higgs coupling constraints in the $t\bar{t}H$, $H \rightarrow b\bar{b}$ channel

7 Signal and background processes

In this part of the thesis direct constraints on the top-Higgs coupling are derived by analyzing Higgs boson production in the $\mathrm{t\bar{t}H}$ channel. As discussed in the previous part, measurements of the top-Higgs coupling strength modifier κ_t already exist from the combination of LHC Run 1 results. These measurements were extended in this thesis by re-analyzing the CMS results and deriving constraints on the modifiers of an anomalous coupling, κ_t and $\tilde{\kappa}_\mathrm{t}$. However, these constraints only hold under the assumption that no particles beyond the standard (BSM) model contribute to gluon-fusion Higgs production and decays into photons. Once this assumption is dropped, a measurement in the $\mathrm{t\bar{t}H}$ channel becomes crucial but the $\mathrm{t\bar{t}H}$ measurements performed in LHC Run 1 still have comparatively large uncertainties. For this reason a direct measurement of the $\mathrm{t\bar{t}H}$ coupling is performed with the CMS data recorded at the LHC in 2015 at 13 TeV.

In contrast to the indirect measurement of the previous part, for which existing CMS analyses have been re-analyzed with a new statistical model, the complete analysis described in this part was performed as part of this thesis. The search of $\mathrm{t\bar{t}H}$ production at CMS is very complex and several analysis groups contribute to it. For this thesis, contributions to the search for $\mathrm{t\bar{t}H}$ in the lepton+jets channel, which targets Higgs boson decays into $\mathrm{b\bar{b}}$, have been made. Two results from an analysis of this channel are presented. In Chapter 10, the search for $\mathrm{t\bar{t}H}$ also published in the CMS analysis [174] is described. In addition, this analysis has been modified to also be sensitive to the kinematic differences induced by anomalous top-Higgs couplings. This is presented in Chapter 11. In Chapter 8 and Chapter 9 the preparation that is common for the extraction of both results is described. Finally, in Chapter 12 the combination of the result from Chapter 10 with other $\mathrm{t\bar{t}H}$ analyses is described. The remainder of this chapter serves as an introduction of $\mathrm{t\bar{t}H(b\bar{b})}$ process and its backgrounds and describes how they are simulated with Monte Carlo (MC) methods.

7.1 The $\mathrm{t\bar{t}H}$, $\mathrm{H \to b\bar{b}}$ channel

The predicted $\mathrm{t\bar{t}H}$ production cross section is small compared to the dominant Higgs-boson production channels, only approximately 0.5 pb^{-1} at a 13 TeV proton-proton center-of-mass energy. Because of the very small expected event yield, the Higgs-boson discovery channels, in which the boson decays into four leptons or a pair of photons, are less useful for the search for $\mathrm{t\bar{t}H}$. A more promising alternative is searching for Higgs bosons decaying to $\mathrm{b\bar{b}}$, which is the channel with the largest branching ratio for a 125 GeV/c^2 Higgs-boson, with a branching ratio of 58%. Indeed, in Run 1 CMS [22] and ATLAS [24–27] performed $\mathrm{t\bar{t}H}$ searches with $\mathrm{b\bar{b}}$, WW, $\gamma\gamma$, ZZ, and $\tau\tau$ final states and the $\mathrm{b\bar{b}}$ channel was, together with WW-decays, the most senstive one.

Top quarks decay almost always into a W boson and a b quark. W bosons couple equally strong to all weak isospin doublets of the Standard Model. Since the decay into a quark-pair is three times more likely due to the number of colors, but the decay into third generation quarks is kinematically impossible, the ratio of leptonic to hadronic decays is

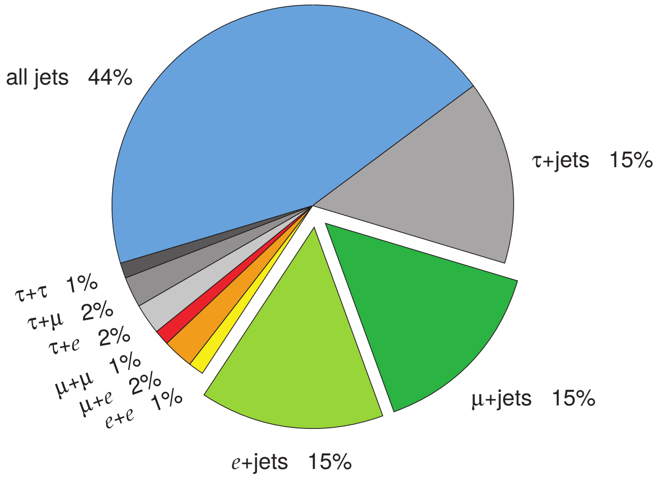


Figure 7.1: Relative size of the $\mathrm{t\bar{t}}$ decay channels, classified based on the number of charged leptons from the two W boson decays. Taken from [129].

three to six. To be able to distinguish $\mathrm{t\bar{t}H(b\bar{b})}$ events from the large multijet backgrounds at the LHC, i.e. from the production of multiple quarks or gluons via processes involving only the strong interaction, it is beneficial to search for final states with at least one isolated electron or muon. In this thesis the lepton+jets channel, in which one W boson decays into a muon or an electron and the other one into hadrons, is analyzed. It has a much larger branching ratio compared to the dilepton channel, in which both W bosons decay leptonically, as illustrated in Fig. 7.1.

Exemplary leading-order Feynman diagrams for the production and decay channel are shown in Fig. 7.2. Diagrams in which the particles are replaced by their antiparticles or in which the Higgs-top vertex connects at the hadronically instead of the semi-leptonically decaying top quark are also possible and have the same amplitude. The events contain one charged lepton, one neutrino, two light quarks, and four bottom quarks. The quarks are visible as jets in the detector while the neutrino escapes undetected, hence the name lepton+jets channel. Typically, $\mathrm{t\bar{t}H}$ events are also accompanied by additional jets, mostly from gluon radiation.

7.1.1 Previous analyses

During the LHC Run 1, searches for $\mathrm{t\bar{t}H}$ with $\mathrm{H \to b\bar{b}}$ have been performed by CMS using lepton+jets and dilepton events. Two analyses were performed, first a multivariate analysis relying on artificial neural networks and boosted desicion trees [22,173], analyzing the 7 TeV and 8 TeV datasets. A limit on $\mathrm{t\bar{t}H}$ production cross section in the $\mathrm{b\bar{b}}$ decay channel of 4.1 times the Standard Model cross section was set at 95% CL, while 3.5 were expected. Later an analysis of the 8 TeV dataset was performed using the Matrix Element Method [23], setting an upper limit on $\mu_{\mathrm{t\bar{t}H}}$ of 4.2, with 3.3 expected. ATLAS performed a search in the dilepton and lepton+jets channel [27] as well as the all-hadronic channel [24]. Both analyses use artificial neural networks in combination with the Matrix Element Method. An upper limit of 3.3 (2.1 expected) was set.

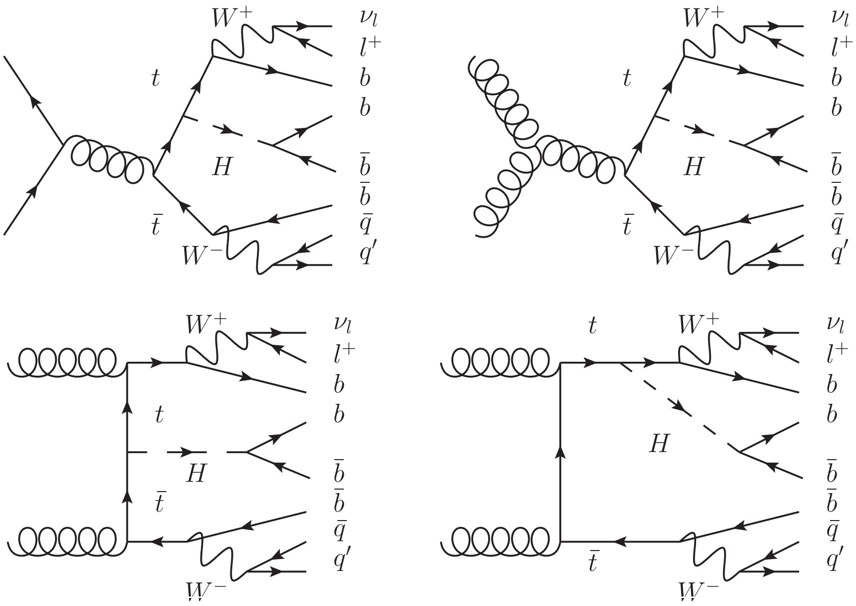


Figure 7.2: Exemplary Feynman diagrams for $\mathrm{t\bar{t}H}$ production with the Higgs-boson decaying into a $\mathrm{b\bar{b}}$ pair and lepton+jets $\mathrm{t\bar{t}}$ decays, drawn with JaxoDraw [63].

7.1.2 Importance of the channel

The main motivation for a search for $\mathrm{t\bar{t}H}$ is the measurement of the top-Higgs coupling. The importance of this coupling has been explained in Chapter 5. As stated before, the coupling measurement in $\mathrm{t\bar{t}H}$ production is more direct than the coupling deduction from other Higgs-boson production and decay channels. In the presence of BSM physics it is the channel with the best chance to obtain insight into the nature of the top-Higgs coupling.

A search for $\mathrm{t\bar{t}H}$ production is also able to detect signs of new physics apart from a modified top-Higgs coupling. One example is the pair production of heavy vector-like top quark partners (see, e.g. [175]). If one of the vector-quark decays into a top quark and a Higgs-boson and the other one into a b quark and a W boson, the final state mimics $\mathrm{t\bar{t}H}$ production. An anomalous trilinear Higgs coupling could also increase or decrease the $\mathrm{t\bar{t}H}$ production cross section [176].

Using the $\mathrm{H \to b\bar{b}}$ channel has additional advantages. It allows to observe Higgs-boson decays into $\mathrm{b\bar{b}}$, which are difficult to distinguish from the overwhelming backgrounds in the gluon and vector boson fusion production channels and only have a chance to be identified if they are produced association with $\mathrm{t\bar{t}}$ or vector bosons. In the SM, the Higgs boson decay into b quarks is the most frequent one. A modified $\mathrm{b\bar{b}}$ decay width thus also has a large effect on the branching into other final states (see Equation 5.1) and complicates the interpretation of branching ratios in terms of couplings. Only Higgs-fermion couplings (indeed only Higgs-quark couplings) contribute to $\mathrm{t\bar{t}H(b\bar{b})}$ at LO and neither production nor decay of the Higgs boson are loop-induced, which furthermore simplifies the interpretation of the result in terms of Higgs-coupling measurements.

The large $\mathrm{t\bar{t}H}$ signal strength of $2.9^{+1.0}_{-0.9}$ extracted in the LHC Run 1 combination is remarkabely more than two standard devations above the Standard Model expectation. It is also remarkable that the $\mathrm{t\bar{t}H}$ measurements of CMS ($\mu_{\mathrm{t\bar{t}H}} = 0.7 \pm 1.9$ [22]) and ATLAS ($\mu_{\mathrm{t\bar{t}H}} = 1.5 \pm 1.1$) [27] in the $\mathrm{H \to b\bar{b}}$ channel do not support the large observed excess. It will be interesting to see whether the excess in the $\mathrm{t\bar{t}H}$ channel can be confirmed in LHC Run 2 and whether it continues to be visible only in some channels. Fortunately, at 13 TeV the $\mathrm{t\bar{t}H}$ production cross section is expected to increase by almost a factor of four with respect to 8 TeV (see Figure 2.5), so that even with comparetively little integrated luminosity similar number of $\mathrm{t\bar{t}H}$ events are expected.

7.1.3 Background processes

The most important background to $\mathrm{t\bar{t}H}$ production with $\mathrm{H \to b\bar{b}}$ is the production of $\mathrm{t\bar{t}}$ in association with a $\mathrm{b\bar{b}}$ pair via the strong interaction, denoted $\mathrm{t\bar{t}b\bar{b}}$. Even when the two b quarks are required to fulfill typical jet cuts, the cross section is an order of magnitude larger than the $\mathrm{t\bar{t}H}$ production cross section. Because the $\mathrm{t\bar{t}b\bar{b}}$ production final state contains the same particles as the one from $\mathrm{t\bar{t}H(b\bar{b})}$, this background is called irreducible. In fact, because the two final states cannot be distinguished, in principle interference between $\mathrm{t\bar{t}b\bar{b}}$ production and the signal has to be considered. However, these effects have been found to be small (see, e.g. [177]), so that $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}b\bar{b}}$ can be effectively treated as two independent processes until measurements have a higher precision.

Similar to the $\mathrm{t\bar{t}b\bar{b}}$ background produced purely by the strong interaction, the additional $\mathrm{b\bar{b}}$ pair can also be produced via electroweak interactions. This process can be kinemetically similar to $\mathrm{t\bar{t}H}$ and often has an invariant $\mathrm{b\bar{b}}$ mass at the Z boson peak, close to the Higgs-mass. However, while the production cross section for this process, $\mathrm{t\bar{t}Z}$ production, is similar to the one of $\mathrm{t\bar{t}H}$, the branching ratio into $\mathrm{b\bar{b}}$ is four times smaller.

The main reducible backgrounds in this analysis are $\mathrm{t\bar{t}c\bar{c}}$ production and the production of $\mathrm{t\bar{t}}$ in association with light jets ($\mathrm{t\bar{t}+lf}$). These backgrounds can be reduced with the help of b-tagging algorithms but not completely erased. The reason for this is that mistag rates for c-hadrons are comparatively large, because c-flavored hadrons have relatively high mass and long life-time and are thus similar to b hadrons. Since the production of $\mathrm{t\bar{t}+lf}$ has a much larger cross section, not all $\mathrm{t\bar{t}+lf}$ events can be eliminated by b-tagging either.

Minor contributions to the reducible backgrounds are expected from other processes containing isolated leptons and b quarks. This includes single top, W+jets, Z+jets, $\mathrm{t\bar{t}W}$, and diboson production. Multijet events do not contain isolated leptons and are mostly eliminated by the event selection. The dominating (accroding to MadGraph5_aMC@NLO, at LO) leading-order Feynman diagrams for the main backgrounds are shown in Fig. 7.3.

7.2 Simulation of standard model processes

The processes that the events selected for this analysis are expected to originate from are simulated using MC event generators. The simulation of collision events is a complex and computationally expensive process, as described in Section 2.2. The effort required to configure and tune these simulations is split within the CMS collaboration and, to save computation time and storage space, MC samples are shared at CMS and produced centrally. For all Standard Model processes samples produced by CMS during the “Fall 15” MC campaign are used. They are listed in Table 7.1. Not shown in the table are multijet simulations because less than 1% of the background events originate from multijet processes and the simulation of this process is thus neglected for the remainder of the analysis.

Table 7.1: MC samples used for the simulation of SM processes in this analysis. The first column names the process and the second the generator used to calculate the hard interaction. The number of events simulated in each sample as well as the cross section σ and branching ratio $\mathcal{B}$ used to normalize the samples are shown in the last two columns.

Process	Generator	Events/10^6	$\mathcal{B} \times \sigma$ [pb]
$\mathrm{t\bar{t}H}$, $\mathrm{H \rightarrow b\bar{b}}$	Powheg v2	3.8	0.295
$\mathrm{t\bar{t}H}$, $\mathrm{H \rightarrow}$ other	Powheg v2	3.9	0.212
$\mathrm{t\bar{t}}$, all decays	Powheg v2	98	831.8
$\mathrm{t\bar{t}}$, lepton+jets decays	Powheg v2	350	378
$\mathrm{t\bar{t}}$, dilepton decays	Powheg v2	110	88.2
t (t-channel)	Powheg v2	3.3	45.3
$\mathrm{\bar{t}}$ (t-channel)	Powheg v2	1.6	28.0
$\mathrm{tW^-}$	Powheg v2	1.0	35.9
$\mathrm{\bar{t}W^+}$	Powheg v2	1.0	35.9
$\mathrm{t/\bar{t}}$ (s-channel)	aMC@NLO	1.0	3.44
W+jets, 100 GeV $< H_\mathrm{T} <$ 200 GeV	MadGraph	10	1345
W+jets, 200 GeV $< H_\mathrm{T} <$ 400 GeV	MadGraph	4.9	360
W+jets, 400 GeV $< H_\mathrm{T} <$ 600 GeV	MadGraph	1.9	48.9
W+jets, 600 GeV $< H_\mathrm{T} <$ 800 GeV	MadGraph	3.8	12.1
W+jets, 800 GeV $< H_\mathrm{T} <$ 1200 GeV	MadGraph	1.6	5.50
W+jets, 1200 GeV $< H_\mathrm{T} <$ 2500 GeV	MadGraph	0.25	1.32
W+jets, 2500 GeV $< H_\mathrm{T}$ GeV	MadGraph	0.25	0.032
Z/γ+jets, 10 GeV $< m_{ll} <$ 50 GeV	aMC@NLO	93	22635
Z/γ+jets, 50 GeV $< m_{ll}$	aMC@NLO	150	6025
$\mathrm{t\bar{t}W}$, leptonic W decays	aMC@NLO	0.25	0.210
$\mathrm{t\bar{t}W}$, hadronic W decays	aMC@NLO	0.83	0.435
$\mathrm{t\bar{t}Z}$, hadronic Z decays	aMC@NLO	0.75	0.611
WW	Pythia 8	0.99	119
WZ	Pythia 8	1.00	44.9
ZZ	Pythia 8	0.98	15.4
$\mathrm{t\bar{t}}$ scale up, all decays	Powheg v2	48	831.8
$\mathrm{t\bar{t}}$ scale up, lepton+jets decays	Powheg v2	96	378
$\mathrm{t\bar{t}}$ scale up, dilepton decays	Powheg v2	30	88.2
$\mathrm{t\bar{t}}$ scale down, all decays	Powheg v2	49	831.8
$\mathrm{t\bar{t}}$ scale down, lepton+jets decays	Powheg v2	97	378
$\mathrm{t\bar{t}}$ scale down, dilepton decays	Powheg v2	29	88.2

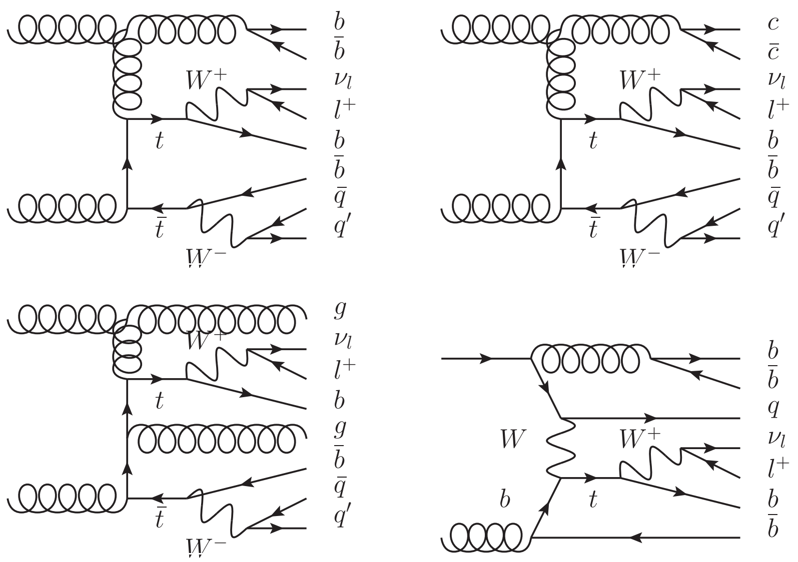


Figure 7.3: Dominant diagrams for the main backgrounds of $\mathrm{t\bar{t}H(b\bar{b})}$ in the lepton+jets channel. From top left to bottom right: $\mathrm{t\bar{t}b\bar{b}}$, $\mathrm{t\bar{t}c\bar{c}}$, $\mathrm{t\bar{t}+lf}$, and t-channel single top + $\mathrm{b\bar{b}}$ production. Drawn using JaxoDraw [63].

7.2.1 Choice and configuration of event generators

Most of the MC samples are simulated using a dedicated matrix element generator for the hard process, interfaced with Pythia 8 [178] to generate parton shower and hadronization. The underlying event is modeled in Pythia 8 with the CUETP8M1 tune, which is similar to the Monash tune [75]. The proton structure is simulated with the NNPDF 3.0 parton distribution functions [179]. All events are interfaced with the Geant4 [76] based CMS detector simulation described in 2.2.5.

The $\mathrm{t\bar{t}H}$ signal and the main backgrounds, $\mathrm{t\bar{t}}$ and single top production, are generated using Powheg v2 [180–182]. The Higgs mass parameter is set to 125 GeV/c^2 and the top mass to 172.5 GeV/c^2. Powheg simulates the hard process at NLO QCD, yielding a better accuracy in the simulation of the top and Higgs kinematics compared to a leading-order generator. An advantage with respect to the aMC@NLO method [183] is that no events with negative weights are generated in the Powheg scheme. Such events reduce the statistical precision of distributions generated with the sample significantly[1] and are also problematic in the training of multivariate methods.

A disadvantage of the Powheg simulation is that at most one additional quark or gluon is generated in the matrix element. All further partons are simulated with the parton shower. As described in Section 2.2, the shower MC is designed to simulate soft and collinear emissions and cannot describe harder parton emissions accurately. Especially for the $\mathrm{t\bar{t}}$ background an accurate predictions of these additional hard partons is important,

---

[1] The relative statistical uncertainty of result obtained using N MC events is

$$\frac{\Delta N}{N} = \frac{\sqrt{\sum_i^{N_\text{events}} w_i^2}}{\sum_i^{N_\text{events}} w_i} = \frac{1}{\sqrt{N}(1-2f_\text{neg})}, \tag{7.1}$$

where the first equation is derived e.g. in [184] and the second part holds true if the event weights are always either +1 or −1 (as they are in MadGraph5_aMC@NLO) and f_neg is the fraction of negative-weight events.

as they can create jets that can be mistaken for Higgs-boson decay products. With a leading-order generator, a large number of additional partons can be generated easily and the resulting simulation can be matched to a parton shower (see e.g. [71]). A similar scheme is also possible at NLO, for example using MadGraph5_aMC@NLO and FxFx-merging [185]. Nonetheless, because of the advantages described above Powheg is chosen as the default MC generator.

Compared to $\mathrm{t\bar{t}}$ production, the production of W bosons has a large cross section in most of these events the W boson is accompanied by comparatively small hadronic activity. In the presented analysis only events with at least four jets with $p_\mathrm{T} > 30$ GeV are considered. Thus dedicated samples that simulate W boson production with H_T of 100 GeV and more are used. Here H_T is the scalar sum of the p_T of all quarks and gluons generated with the matrix element generator. This has the advantage that for events with high jet multiplicity, which are the most important background events, enough MC events are available to reliably model the background. It is checked with an inclusive W+jets sample that hardly any events with $H_\mathrm{T} < 100$ GeV are accepted by the event selection. The W+jets samples are generated using the MadGraph5_aMC@NLO generator in the LO mode, matched to Pythia 8 using MLM matching (see [186]).

The minor backgrounds $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ and s-channel single top production are generated with MadGraph5_aMC@NLO in the NLO mode while diboson production is simulated with Pythia 8.

7.2.2 Total cross section

Cross section calculations for most of the above processes are available at NNLO QCD accuracy and with NLO electroweak corrections. To profit from the accuracy of these predictions, the MC events generated from LO and NLO simulations are reweighted with a constant factor so that the sum of the weights w_i of all events corresponds to the number predicted by the most precise cross section calculation available:

$$\sum w_i = N_{\mathrm{exp.}} = L\sigma. \tag{7.2}$$

The cross section values used for normalization are also given in Table 7.1.

For the $\mathrm{t\bar{t}H}$ signal the cross section recommended by the LHC Higgs cross section working group [187] is used. It includes NLO QCD corrections, calculated first in [188–192] and the more recent NLO EWK corrections [134, 193, 194]. The factorization scale is $\mu = m_t + m_H/2$ and PDFs are evaluated according to the PDF4LHC [166] prescription, with the exception of γ-induced processes for which NNPDF2.3QED [195] is used.

The main background $\mathrm{t\bar{t}}$ was scaled to the NNLO cross section calculated with the Top++ 2.0 program [196], using the MSTW [197], NNPDF 2.3 [198], and CT10 [199] PDF sets according to the 2011 PDF4LHC recommendations [200]. The single top t-channel and s-channel production cross sections are calculated with HATHOR [201, 202], using the same parton distribution function prescription as for $\mathrm{t\bar{t}}$ production. For the associated production of top quark and W boson results from [203] are used.

The W+jets cross sections in H_T-bins are determined using the LO cross section predicted by MadGraph5_aMC@NLO multiplied with a NNLO "k-factor", the ratio of the NNLO prediction calculated from FEWZ [204, 205] divided by the inclusive cross section predicted in LO. The FEWZ program is also used to calculate the Z+jets cross section for the two dilepton [206] invariant mass cuts specified in Table 7.1. The PDF set CTEQ6.6M [161] is used in FEWZ. Finally, the cross sections for associated production of $\mathrm{t\bar{t}}$ and vector bosons are taken from [194], for WW-production NNLO cross sections from [207], and for WZ and ZZ NLO cross sections from [208] are used.

7.2.3 Modeling of the $\mathrm{t\bar{t}}$+ heavy flavor background

The main background processes $\mathrm{t\bar{t}b\bar{b}}$, $\mathrm{t\bar{t}c\bar{c}}$, and $\mathrm{t\bar{t}+lf}$ are all generated as part of the inclusive $\mathrm{t\bar{t}}$ sample. To be able to assign different uncertainties to the different $\mathrm{t\bar{t}}$ subprocesses, this sample is divided based on the heavy flavor content in addition to the $\mathrm{t\bar{t}}$ decay products. With the algorithm described in [209] the sample is split into five subsamples: $\mathrm{t\bar{t}b\bar{b}}$, $\mathrm{t\bar{t}2b}$, $\mathrm{t\bar{t}b}$, $\mathrm{t\bar{t}c\bar{c}}$, $\mathrm{t\bar{t}+lf}$.

The $\mathrm{t\bar{t}b\bar{b}}$, $\mathrm{t\bar{t}2b}$, and $\mathrm{t\bar{t}b}$ events all contain at least four b quarks but only $\mathrm{t\bar{t}b\bar{b}}$ contains two well-separated b quarks that both have a momentum that allows reconstruction as a b-jet. In $\mathrm{t\bar{t}b}$ events, only one of the two additional b quarks is hard and central enough to be reconstructed in a jet. This distinction allows including an additional degree of freedom in the statistical model used to extract the $\mathrm{t\bar{t}H}$ signal and described in 4.2. Events containing b-quarks that are low-p_{T} or outside of the tracker acceptance are not considered as they are experimentally indistinguishable from light jets. In contrast to $\mathrm{t\bar{t}b}$ events, $\mathrm{t\bar{t}2b}$ contains two b quarks with small angular separation, which can be interpreted as coming from a high p_{T} gluon splitting. Such events cannot be accurately described in perturbation theory and have associated uncertainties that are different from events with well-separated b quarks. While all of the above is also relevant for $\mathrm{t\bar{t}c\bar{c}}$ events, this background is less important and no distinction between the subcategories is made.

The first part of the heavy flavor splitting algorithm operates on particle level, i.e. on the simulation including parton shower and hadronization but without the detector simulation. All heavy flavor hadrons before their electroweak decay are identified in simulated $\mathrm{t\bar{t}}$ events. The energy and momentum of these particles is rescaled so that it is practically zero, thereby leaving the direction and velocity unchanged. These hadrons are called “ghost hadrons”. Together with all stable ($c\tau > 1$ cm) final-state particles of the event, including the hadron’s decay products but excluding neutrinos they are then clustered to anti-k_{T} [77] jets with a distance parameter of 0.4. Particle jets containing a ghost b hadron or c hadron are called b-jets and c-jets, respectively. This jet-flavor association has the advantage, that it can be easily generalized for arbitrary jet-algorithms. Furthermore, it does not depend on quarks, which are only intermediate particles in the scattering reactions, and it thus yields results that can be compared for different shower MC and configurations. This part of the algorithm is the default flavor identification tool at CMS in Run 2 and called “ghost hadron injection”.

For this thesis, the second part of the algorithm is also important. For all b- and c-flavored hadrons, the origin of the flavor is searched backwards in the particle history up to the hard process. In this way it is decided whether the b hadron originated in a top quark decay or is from a b quark produced in addition to the top-decay products. The concept of a particle history is not physical, it depends on the model used for hadronization and the scheme used to save intermediate results of the parton shower. Nonetheless, the algorithm has been shown to yield reasonable results for a number of parton showers , e.g. PYTHIA 8, SHERPA [210], HERWIG 6 [73], and PYTHIA 6 [72].

Every event is classified based on the number of particle jets containing hadrons not originating in the $\mathrm{t\bar{t}}$ decay. Only particle jets fulfilling similar kinematic cuts of $p_{\mathrm{T}} > 20$ GeV/c and $|\eta| < 2.4$ are considered. If there are at least two b-jets, containing at least one additional b hadron each but no b hadron from the $\mathrm{t\bar{t}}$ decay, the event is classified as $\mathrm{t\bar{t}b\bar{b}}$. The event is classified as $\mathrm{t\bar{t}b}$ if there is only one such jet. In case that there is

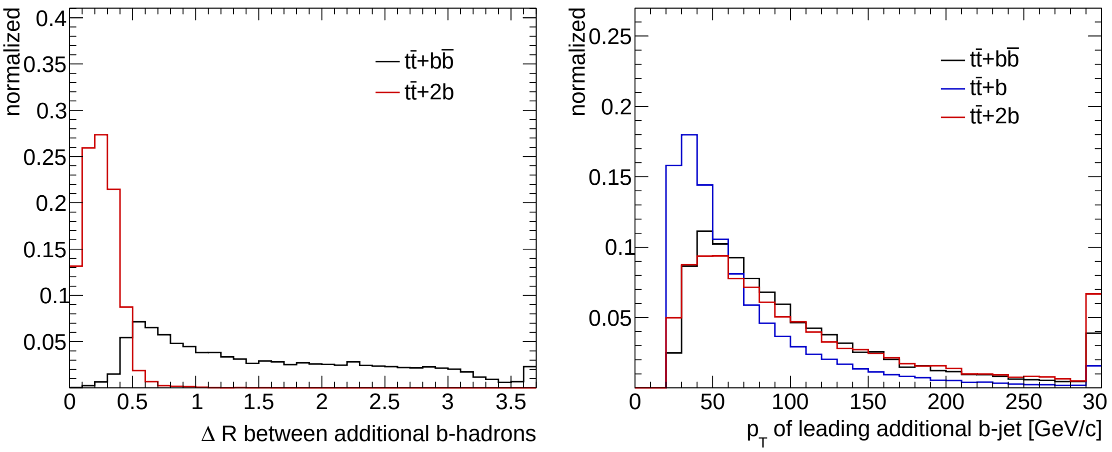


Figure 7.4: Kinematic differences between the $t\bar{t}+b+X$ classes in Powheg $t\bar{t}$ events at 13 TeV center-of-mass energy. In the left-hand plot the expected separation of the two additional b hadrons (i.e. hadrons that cannot be traced back to the $t\bar{t}$ decay) in the η-ϕ plane is shown. In the right-hand plot the transverse momentum of the leading additional b-jet (a particle level jet that contains at least one hadron that cannot be traced back to the $t\bar{t}$ decay) is displayed. Only jets with a transverse momentum of $p_{\mathrm{T}} > 30$ GeV$/c$ are considered and the last bin of the histogram also represents, in both cases, values larger than the x-axis range.

only one additional b-jet, but it contains at least two b hadrons, the event falls in the $t\bar{t}$2b class. If there are no particle level b hadrons but jets with at least one c-hadrons inside, the event is classified as $t\bar{t}c\bar{c}$. Finally, in the case the event falls in none of the classes above, it is called $t\bar{t}+$lf. Some of the kinematic differences between the $t\bar{t}+b+X$ classes can be seen in Fig. 7.4. By construction, $t\bar{t}$2b events contain b hadrons that are very close to each other. The leading b-jet is harder in $t\bar{t}b\bar{b}$ events than in $t\bar{t}b$ events. The jet that contains the b hadron pair in a $t\bar{t}$2b event typically also has a high transverse momentum.

In general, it should be possible to obtain a better description of the $t\bar{t}b\bar{b}$ process by a dedicated simulation with a matrix element generator. Usually, $t\bar{t}b\bar{b}$ is generated at matrix element level with phase space cuts and then interfaced with a parton shower program. For a sample produced in the five-flavor scheme with massless b quarks, such phase space cuts are even mandatory to avoid divergences associated with soft and collinear splittings. Combining such a $t\bar{t}b\bar{b}$ sample with the inclusive $t\bar{t}$ sample as was used in this analysis is not trivial. It would require to identify and remove events from the inclusive sample that contain b quarks fulfilling the same definition as the definition used to generate the inclusive $t\bar{t}b\bar{b}$ sample. Since b quarks only exist as intermediate states in the event simulation, it is not guaranteed that the b quarks simulated, e.g. in Pythia 8 matched to Powheg, are the same as the ones from the MadGraph5_aMC@NLO matrix element calculation.

The best description of $t\bar{t}b\bar{b}$ over a large part of the phase space should be possible in the four flavor scheme, with massive b quarks. With massive b quark the complete $t\bar{t}b\bar{b}$ phase space can in principle be simulated, including regions where the b quarks are collinear. The production of such a sample is even possible at NLO accuracy with MadGraph5_aMC@NLO and Sherpa, as shown in [211]. In principle, these simulations can be combined with any inclusive $t\bar{t}$ simulation. This requires removing all events with b quarks (excluding b quarks from $\mathrm{t\bar{t}}$ decays and underlying event) from the inclusive sample and replacing them by the $\mathrm{t\bar{t}b\bar{b}}$ sample, as described in [212]. Removing all b quarks does not suffer from the problems associated with just removing b quarks in a specific phase space as described above. However, it is unclear if the description of such an approach would really improve the $\mathrm{t\bar{t}b\bar{b}}$ description for the full phase space and an implementation of the approach has not been tested within the context of this analysis.

7.3 Simulation of $\mathrm{t\bar{t}H}$ with anomalous top-Higgs coupling

For the simulation of $\mathrm{t\bar{t}H}$ production with an anomalous top-Higgs coupling no MC samples are centrally produced by CMS. Thus dedicated samples are produced for this analysis. The anomalous top-Higgs coupling is implemented in the Higgs characterization model [163] that can be used with MadGraph5_aMC@NLO [164] at leading-order QCD and at next-to-leading order [165].

7.3.1 Generator configuration

The simulation of a $\mathrm{t\bar{t}H}$ event, including parton shower, CMS detector simulation, and reconstruction, takes about 3 minutes. For the multivariate analysis techniques applied in this thesis, a large number of generated events is necessary to reduce statistical uncertainties. As mentioned above, in the matching scheme employed in MadGraph5_aMC@NLO at NLO, events with negative weights are generated and these events reduce the statistical power of the sample significantly. To generate an MC-simulated data sample with a low statistical uncertainty in reasonable time, the leading order mode of MadGraph5_aMC@NLO is used. NLO corrections have little effect on the features that allow to distinguish scalar form pseudoscalar coupling [142] and it will be shown later in this thesis that the MadGraph5_aMC@NLO LO sample shows no large deviations from an NLO simulation.

Events are decayed with the MadSpin [213] module to preserve the spin correlation of the two top quarks. Parton shower and hadronization is simulated consistently to the other MC samples, using Pythia 8 with the CUETP8M1 tune. The set of parton distribution functions chosen for the generation of events is also the same as for the other MC simulations, NNPDF3.0. The factorization and renormalization scales are set to the MadGraph5_aMC@NLO default, which for this process is $\mu_F = \mu_R = \sqrt[3]{m_\mathrm{T}(t) m_\mathrm{T}(\bar{t}) m_\mathrm{T}(H)}$, where $m_\mathrm{T}(X)$ is the transverse mass of the object X, with $X = \mathrm{t}, \bar{\mathrm{t}}, \mathrm{t}$. Finally, the interaction with the detector is simulated with the full Geant4 based CMS detector simulation.

For the analysis of the top-Higgs coupling a large κ_t and $\tilde{\kappa}_\mathrm{t}$ parameter space needs to be analyzed. Generating dedicated MC simulations for all relevant points is not feasible. Fortunately, from the parameterization in equation 5.3 it is clear that the parameter $\bar{\kappa}_t$ only affects the total rate of $\mathrm{t\bar{t}H}$ production, not the kinematics. Thus, only the dependence on the mixing angle ζ_t needs to be considered, the effect of $\bar{\kappa}_t$ can always be incorporated by scaling distributions by a constant factor.

7.3.2 Interpolation scheme

For differential $\mathrm{t\bar{t}H}$ cross section estimates at leading order in the top-Higgs coupling, three types of terms, proportional to either κ_t^2, $\tilde{\kappa}_\mathrm{t}^2$, or $\kappa_\mathrm{t}\tilde{\kappa}_\mathrm{t}$, are relevant. In [158] it is outlined in some detail that the terms proportional to $\kappa_\mathrm{t}\tilde{\kappa}_\mathrm{t}$ become only relevant when also considering the decay of the top quarks. They can be understood as the effect the

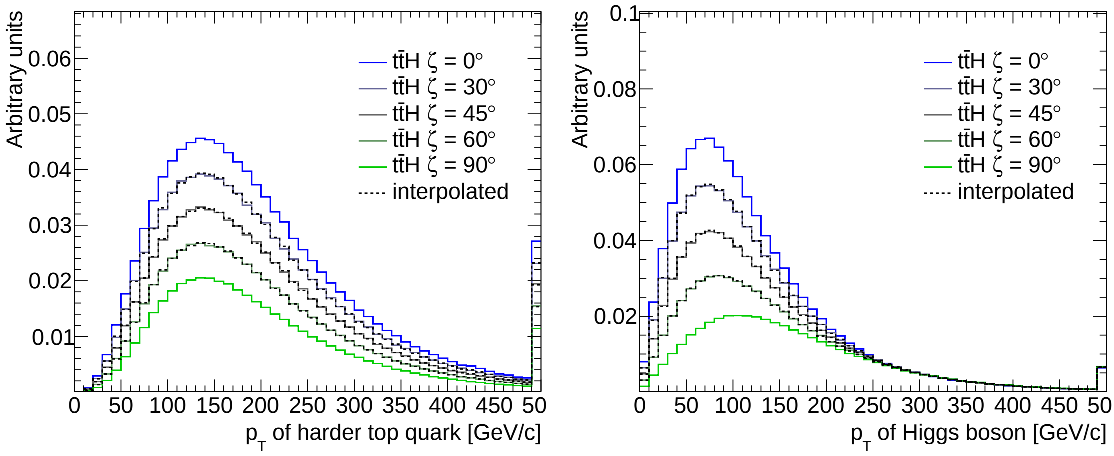


Figure 7.5: Validation of the interpolation used to simulate $\mathrm{t\bar{t}H}$ production with a mixed scalar-pseudoscalar coupling. In blue and green distributions of Higgs-boson and top quark p_{T}, generated with MadGraph5_aMC@NLO and the Higgs characterization model [163] for different coupling mixtures (parameterized by ζ_{t}), are shown. The result of the interpolation between the two extreme cases for the coupling mixture is indicated with dashed lines.

top-Higgs coupling has on the correlation of the top quark spins. Observables sensitive to this correlations have been explored, for example in [146] and [140], to construct variables that are sensitive to the $\kappa_{\mathrm{t}}\tilde{\kappa}_{\mathrm{t}}$-contribution. In [142] a matrix element discriminant D_{CP} is constructed that should achieve the optimal discrimination between $\mathrm{t\bar{t}H}$ production with different ζ_t. Effects due to the interference terms are very subtle for these observables and far beyond the current experimental precision. This feature makes the simulation of the anomalous coupling almost trivial. Terms proportional to $\tilde{\kappa}_{\mathrm{t}}\kappa_{\mathrm{t}}$ can be ignored and differential cross sections can be written as

$$d\sigma_{\mathrm{t\bar{t}H}} = \kappa_{\mathrm{t}}^2 d\sigma^{+}_{\mathrm{t\bar{t}H}} + \tilde{\kappa}_{\mathrm{t}}^2 d\sigma^{-}_{\mathrm{t\bar{t}H}}, \tag{7.3}$$

where $d\sigma^{+}_{\mathrm{t\bar{t}H}}$ ($d\sigma^{-}_{\mathrm{t\bar{t}H}}$) is the differential cross section for a purely scalar (pseudoscalar) coupling. Thus all distributions can be calculated from just two MC samples. Parton shower, hadronization, and the detector response are only simulated for $\zeta_t = 0$ and $\zeta_t = \pi/2$ and the remaining values are simulated by interpolating between distributions generated from the two samples.

To validate this interpolation, distributions generated from MadGraph5_aMC@NLO simulations with different ζ_t are compared to the interpolation between just two values, $\zeta_t = 0$ and $\zeta_t = \pi/2$, in Fig. 7.5. As long as only top quark and Higgs boson kinematics are considered, no statistically significant differences between the dedicated simulation and the interpolation can be observed.

7.4 Uncertainties of the MC simulations

The simulation of events is associated with a number of uncertainties. In this section the uncertainties due to cross section calculations and the modeling of the hard process, parton shower, and hadronization are discussed. Experimental uncertainties introduced by the modeling of the interaction of particles with the CMS detector will be considered in the next chapter.

7.4.1 Rate uncertainties

The calculations of the total cross sections have limited precision, mainly due to the description of the proton structure with parton distribution functions and due to the choice of the factorization scale μ_F and the renormalization scale μ_R. Scale uncertainties are introduced by truncating the perturbative series after a finite number of steps and are, as a convention, evaluated by varying the scales μ_R and μ_F by a factor of two up and down and interpreting the result as 68% confidence intervals.

For the evaluation of the PDF and α_S uncertainties different schemes are used. The signal process cross section uncertainty is evaluated following the Run 2 PDF4LHC prescription [166]. The $\mathrm{t\bar{t}}$ and single top t-channel and s-channel PDF+α_S uncertainties are calculated with the older PDF4LHC interim recommendation [200]. The tW production uncertainty is taken from [203]. For the minor backgrounds only the uncertainties associated to the PDF sets used to calculate the central value are considered.

Some of the uncertainties of the cross section calculations are correlated. For example, processes that are produced at the same center-of-mass energy in the same initial state tend to have correlated PDF uncertainties. The exact correlation between the uncertainty has not been studied, instead a simplified approach is used. The $\mathrm{t\bar{t}H}$ production cross section uncertainties are assumed to be uncorrelated from the backgrounds. All processes originating, at leading-order approximation, in one of the three initial states gg, gq, and qq are treated as fully correlated. Scale uncertainties might also be correlated between similar processes because higher order corrections can have a similar effect of processes with similar topologies. They are also estimated to be either 0% or 100%. A summary of rate uncertainties and their correlations can be found in Table 7.2.

7.4.2 Heavy flavor production uncertainties

Additional uncertainties are considered for the $\mathrm{t\bar{t}} + \mathrm{hf}$ background. To the heavy flavor subsamples $\mathrm{t\bar{t}b\bar{b}}$, $\mathrm{t\bar{t}2b}$, $\mathrm{t\bar{t}b}$, and $\mathrm{t\bar{t}c\bar{c}}$, an additional rate uncertainty of 50% is attributed. This uncertainty is assumed to be uncorrelated between the subprocesses. For $\mathrm{t\bar{t}b\bar{b}}$ the uncertainty can be motivated by the typical size and uncertainty of the NLO QCD corrections to $\mathrm{t\bar{t}b\bar{b}}$ as they are calculated in [214–216]. The size of the NLO QCD correction with respect to LO ranges from up to 100% in [216] for a fixed scale to 20% for a dynamic scale choice as in [214]. Scale uncertainties in all publications are of the order 20% to 30%. From an experimental standpoint the assigment of the rate uncertainty can be justified with the CMS Run 1 measurement of $\mathrm{t\bar{t}b\bar{b}}$ [217]. Here $\mathrm{t\bar{t}b\bar{b}}$ was also simulated as part of an inclusive $\mathrm{t\bar{t}}$ sample and the treatment of $\mathrm{t\bar{t}b\bar{b}}$ as three subprocesses was flexible enough to fit the data. The total cross section was approximately 50% larger than the one predicted by MADGRAPH matched to PYTHIA 6.

As an additional cross-check of the 50% uncertainty, the number of events predicted by three different parton showers matched to different matrix element generators is compared in Fig. 7.6. For this figure the heavy-flavor splitting scheme was extended to include double gluon splitting events into a separate category. The number of events depends mostly on the parton shower and PYTHIA 6 and Herwig++ [218] are within 50% of the PYTHIA 8 prediction.

Table 7.2: Rate uncertainties of theory predictions considered in this analysis and their assumed correlation among each other.

Process	scale uncertainty [%]	PDF uncertainty [%]
$\mathrm{t\bar{t}H}$	$-9.2/+5.8$ [1]	± 3.6 [1]
$\mathrm{t\bar{t}b\bar{b}}$	± 50 [2] $\oplus -4/+2$ [a]	± 3 [e]
$\mathrm{t\bar{t}2b}$	± 50 [2] $\oplus -4/+2$ [a]	± 3 [e]
$\mathrm{t\bar{t}b}$	± 50 [2] $\oplus -4/+2$ [a]	± 3 [e]
$\mathrm{t\bar{t}c\bar{c}}$	± 50 [2] $\oplus -4/+2$ [a]	± 3 [e]
$\mathrm{t\bar{t} + lf}$	$-4/+2$ [a]	± 3 [e]
t (t-channel)	± 3 [b]	± 4 [f]
tW	± 3 [b]	± 4 [f]
t (s-channel)	± 3 [b]	± 4 [f]
W+jets	± 1 [c]	± 4 [g]
Z/γ+jets	± 1 [c]	± 4 [g]
$\mathrm{t\bar{t}W}$	$-12/+13$ [a]	± 2 [e]
$\mathrm{t\bar{t}Z}$	$-12/+13$ [a]	± 3 [e]
WW	± 2 [d]	± 2 [g]
WZ	± 2 [d]	± 2 [g]
ZZ	± 2 [d]	± 2 [g]

[1] Signal uncertainties are uncorrelated from the background uncertainties.

[2] Each $\mathrm{t\bar{t} + hf}$ process is assigned an uncorrelated uncertainty.

[a] Correlated among all $\mathrm{t\bar{t}}$-like processes.

[b] Correlated among all single top processes.

[c] Correlated among all vector boson plus jets processes.

[d] Correlated among all diboson processes.

[e] Correlated among all gg-initiated processes.

[f] Correlated among all qg-initiated processes.

[g] Correlated among all qq-initiated processes.

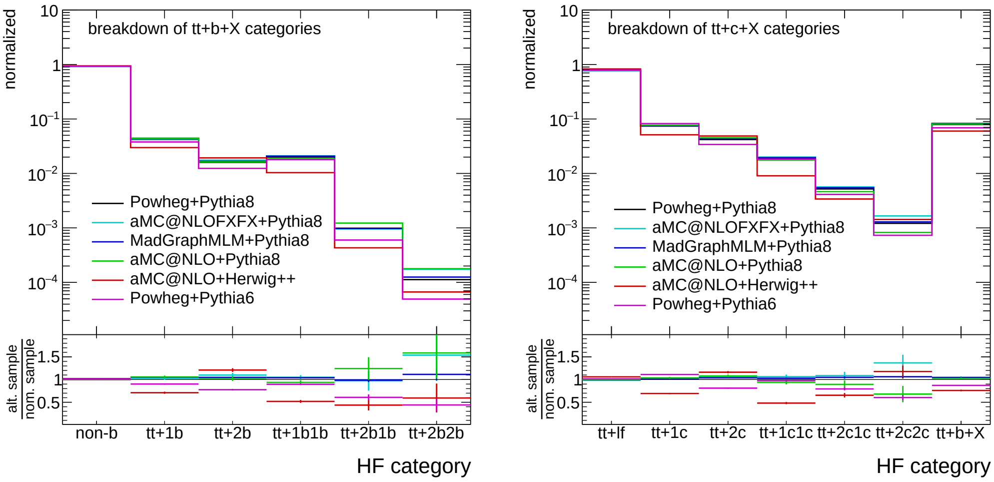


Figure 7.6: Fraction of $\mathrm{t\bar{t}}$ events belonging to the $\mathrm{t\bar{t}+hf}$ classes, predicted by different matrix element and shower MC combinations. The classes are defined based on the number of heavy flavor hadrons within jets not originating from the $\mathrm{t\bar{t}}$ decay. The matrix element generators compared are Powheg v2 and MadGraph5_aMC@NLO in LO and NLO mode. For two of the MadGraph5_aMC@NLO samples, $\mathrm{t\bar{t}}$ + additional parton processes are simulated at matrix element level and merged via FxFx and MLM merging. Three parton showers are compared, Pythia 6, Pythia 8, and Herwig++.

7.4.3 Shape uncertainties

In addition to the uncertainty of the fixed-order calculations that are used to determine the MC normalization, the PDF and scale dependency of the MC event generators introduce uncertainties that affect the shape of distributions. Shape uncertainties are only evaluated for the dominant $\mathrm{t\bar{t}}$ background. When naively evaluated, the shape-changing uncertainties described below also affect the total normalization of the MC simulation, before any cuts. Since this normalization is known with high precision and should only be affected by the normalization uncertainties described in the previous subsection, normalization-changing effects of shape uncertainties are canceled by normalizing the simulation to the event yield predicted by fixed-order calculations.

During the calculation of the hard process, weights are calculated that correspond to the ratio of the probability to obtain a simulated event with a varied scale choice and the probability to obtain it with the nominal scale. Hereby μ_F and μ_R are both independently and jointly varied by factors of two. These weights allow to transform every distribution simulated with the nominal MC to a distribution that corresponds to a changed scale in the matrix element generator. In general, the correlated variation of μ_F and μ_R leads to the largest shape difference and is used to estimate the influence of the matrix element scale on the shape of distributions. The main effect is the change in jet multiplicity and p_T that is shown in Fig. 7.7.

PDF uncertainties can be evaluated in the same fashion. During the calculation of the hard processes, the ratio of the nominal PDF to several systematically changed PDFs is calculated. For this analysis the 100 PDF replicas of the NNPDF 3.0 PDF set are evaluated. While this reweighting procedure indeed has an effect on the normalization of processes, shape effects are small compared to other uncertainties and are neglected.

Reweighting the shower scale is not (yet [219]) possible. Instead, samples with a varied

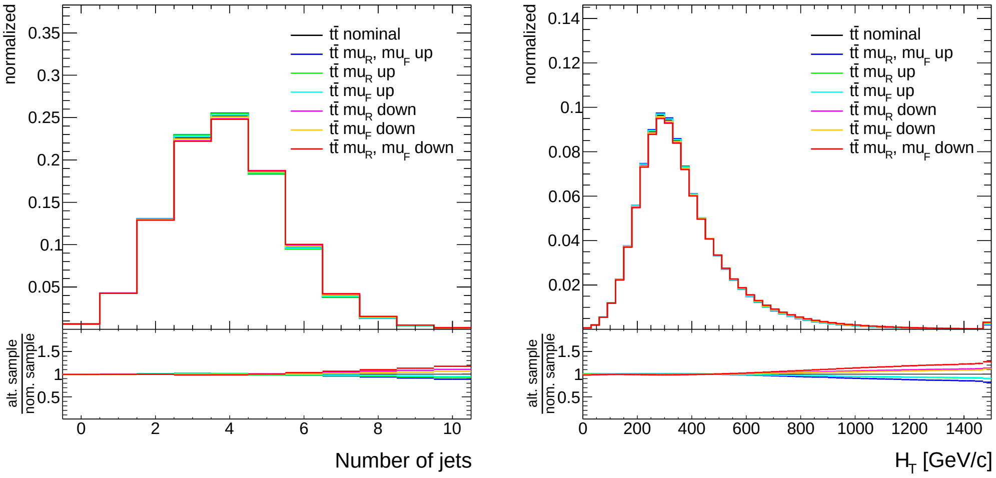


Figure 7.7: Effect of varying the factorization scale μ_F and renormalization scale μ_R up and down by a factor of two in the matrix element generator POWHEG V2 during the simulation of the $\mathrm{t\bar{t}}$ process. Distributions shown are the number of reconstructed jets with $p_\mathrm{T} > 30$ GeV/c and H_T, the scalar sum of jet p_T.

scale in the parton shower are generated. In principle, the scale for initial state and final state radiation can be varied independently but to reduce the number of events that have to be generated both scales are either varied up or down by a factor of two. The effect of the varied shower scale is shown in Fig. 7.8. The effect is larger than the one from varying the scale on matrix element level. Additionally, it affects more parts of the simulation, e.g. jet-shapes and b-tagging.

The variations of normalization and factorization scale are each treated as an independent uncertainty for the five $\mathrm{t\bar{t}}$ subsamples, so that ten independent shape uncertainties are introduced from the variation of matrix element and parton shower scale.

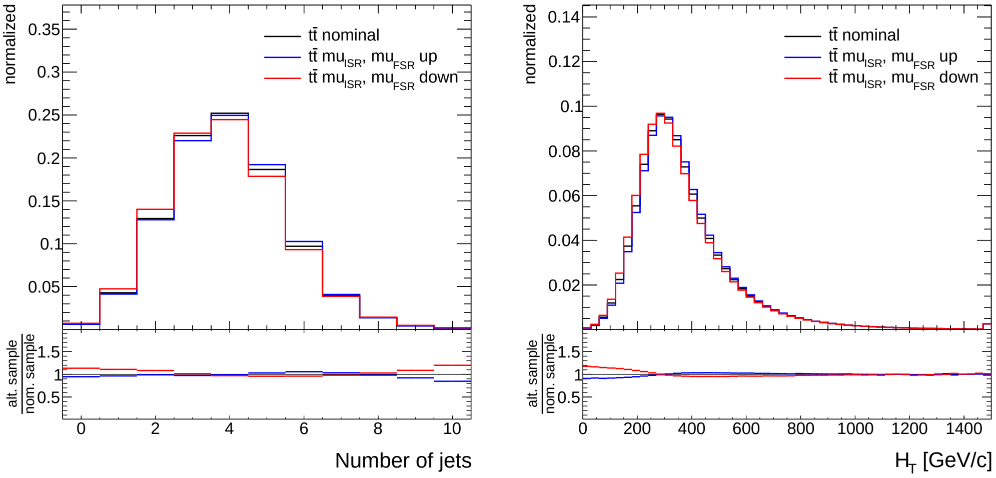


Figure 7.8: Effect of varying the factorization scale μ_F and renormalization scale μ_R up and down by a factor of two in the PYTHIA 8 parton shower during the simulation of the $t\bar{t}$ process. Distributions shown are the number of reconstructed jets with $p_T > 30$ GeV/c and H_T, the scalar sum of jet p_T.

8 Physics objects and calibration measurements

The final state of $t\bar{t}H(b\bar{b})$ lepton+jet events is expected to contain four b-jets (two from the Higgs boson and two from the top quark decays), two light jets, one prompt electron or muon, and missing transverse momentum due to the neutrino. In Section 8.1 the identification of these final state physics objects is described.

The use of high-level objects with a physics motivation has the advantage that objects can be calibrated in events whose properties are well-known. The corrections from calibration measurements used in this analysis are explained in Section 8.2.

8.1 Physics objects

Physics objects are reconstructed using the CMS reconstruction techniques described in Section 3.4. Wherever appropriate, the particle candidates returned by the particle flow (PF) algorithm (see Section 3.4.4) are used. The objects are considered or discarded after applying additional selection criteria. These criteria are chosen according to several principles: the probability to misidentify an object has to be low, the kinematics of the selected objects should be close to the one expected from $t\bar{t}H$ events, and the cuts must be stringent enough so that the events pass the triggers used in the analysis.

8.1.1 Vertices

The inelastic collision of two protons can be reconstructed as a primary vertex with the algorithms described in Section 3.4.1. The correct identification of primary vertices is important as it is relevant for the identification of prompt leptons, it allows for the exclusion of charged particles from pileup vertices to mitigate the influence of pileup, and it is relevant to calculate the distance of secondary vertices for b-tagging.

The following quality criteria are required to be fulfilled by primary vertices used in this and most CMS analysis: The number of degrees of freedom in the vertex fit has to be greater than or equal to four and the distance to the nominal collision point has to be less than 24 cm (4 cm) in longitudinal (radial) direction. The primary vertex associated with the hard interaction is identified as the one for which the sum of the transverse momenta squared of the particles from tracks associated to the vertex is the highest. Additional vertices are attributed to pileup.

8.1.2 Isolated muons

The muon selection is designed to select prompt muons, which are directly produced in the decay of heavy bosons, and to reject muons from hadron decays, cosmic radiation as well as particles mimicking muons. Details on the CMS muon reconstruction and identification can be found in [100].

For all muon definitions in this analysis, even if they are called "loose", the tight muon identification criteria (tight ID) described in [100] are applied. They require that the muon is reconstructed as a muon with the particle flow algorithm and that it is reconstructed as a global muon with at least two muon station hits, at least one pixel hit, and that its passage through more than five tracker layers is measured. The muon track fit must have $\chi^2/n_{\mathrm{d.o.f.}} < 10$ and the radial (longitudinal) distance of the track to the primary vertex must be less than 2 mm (5 mm). The efficiency of this ID is, depending on the muon kinematics, approximately 96%.

The muons are required to be isolated from the remaining particles of the event. This is motivated by the fact that a large hadronic activity close to the muon increases the probability of the muon to originate from a hadron decay or to be a hadron that was not stopped by the calorimeters and left a track in the muon chamber. As a measure of the activity α in the vicinity of the muon, the transverse momentum of all particles in a cone of $\Delta R < 0.4$ around it is summed up. Contributions from particles produced in pileup interactions should not be considered in the sum. Thus charged particles with a track originating from a pileup vertex are ignored. For neutral particles, i.e., photons and neutral hadrons, no track can be reconstructed. Instead, the expected pileup contribution is estimated from the pileup due to charged particles. In CMS that ratio of the energies of neutral to charged particle flow candidates in pileup has been observed to be 1:2. Consequently, half of the transverse momentum sum of charged particles from pileup vertices is subtracted from the activity to account for the contribution of neutral particles from pileup. However, this correction must never be larger than the neutral activity. In summary, the pileup-corrected activity is defined as:

$$\alpha_\mu = \sum_{\text{charged, not PU}} p_{\mathrm{T}} + \max\left(0, \sum_{\text{neutral}} p_{\mathrm{T}} - 0.5 \sum_{\text{charged,PU}} p_{\mathrm{T}}\right), \tag{8.1}$$

where only PF candidates with $\Delta R < 0.4$ to the muon are considered.

Two kinds of muon IDs are used in this analysis. The one with more stringent cuts is called "tight" and is used to identify the muon from the leptonic W decay in the $\mathrm{t\bar{t}H}$ candidate events. Its cuts of $p_{\mathrm{T}} > 25$ GeV/c and $|\eta| < 2.1$ are more restrictive than the cuts required by the muon trigger used in this analysis. The p_{T} cut is chosen to achieve a large $\mathrm{t\bar{t}H}$ efficiency while rejecting events close to the trigger p_{T} threshold. Using such events would require exact knowledge of the trigger efficiency for leptons close to the threshold, which is difficult to model and has thus large associated uncertainties. The activity compared to the lepton p_{T} (counter intuitively called relative isolation) has to be $I_\mu^{\mathrm{rel}} = \alpha_\mu / p_{\mathrm{T}} < 0.15$, which is a compromise between a large prompt muon selection efficiency and a good rejection of backgrounds.

The loose lepton ID is utilized in the analysis of dilepton decays. It is similar to the tight ID, but the kinematic and isolation requirements are relaxed to $p_{\mathrm{T}} > 15$ GeV/c, $|\eta| < 2.4$, and $I_\mu^{\mathrm{rel}} < 0.25$. In the lepton+jets channel this ID is used to reject any lepton in addition to the one with the tight ID to ensure that the event selection does not overlap with the dilepton channel. The cuts of the two IDs are summarized in Table 8.1.

8.1.3 Isolated electrons

The selection of isolated electrons follows the same rationale as the muon selection. It is set up to identify prompt electrons from W decays and two different IDs are used in this analysis. Backgrounds that can be misidentified as prompt electrons are photons

Table 8.1: Summary of the cuts defining the two sets of muon identification criteria used in this analysis.

Observable	Tight cut	Loose cut
ID	tight	tight
p_T	> 25 GeV/c	> 15 GeV/c
$\lvert\eta\rvert$	< 2.1	< 2.4
I_μ^rel	< 0.15	< 0.25

converting to an electron-positron pair, jets misidentified as electrons, and electrons from hadron decays. Electrons are more difficult to distinguish from their backgrounds than muons, therefore the selection is more complex. A multivariate electron ID, a Boosted Decision Tree similar to the one used for the triggering MVA ID from Run 1 [106], identifies electrons. It is trained with simulated Z+jets events. Electrons from the Z-decay were used as signal, additional electron candidates as background.

In this and many other analyses, the prompt electron is also typically the one causing the electron HLT to record the event. Because the electron trigger already applies cuts on electron identification variables, recorded signal and background electrons have rather signal-like characteristics. For this reason the MVA ID used in this analysis is trained after the application of a set of preselection cuts that mimic the trigger requirements and it is thus specialized in removing misidentified electrons candidates that passed the trigger. This preselection, consisting of loose electron identification cuts, is also applied to electrons used in this analysis before applying the MVA ID.

The electron ID BDT uses twenty input variables that can be classified as follows. The first class are calorimetric variables, using the facts that the electromagnetic shower shapes are different for signal and background and that electrons deposit most of their energy in the ECAL. Furthermore, tracking variables are used that take the characteristic shape of the electron track due to bremsstrahlung into account. Finally, observables comparing ECAL and tracker measurements that make sure the electron track is compatible with the energy deposit in the calorimeter are included in the training. For all electron candidates the electron identification BDT is evaluated. If the BDT output is larger than the working point defined as “tight”, the electron is kept. The working point depends on the η-region of the electron and is chosen such that the electron identification efficiency is 80%.

Like muons, electrons are required to be isolated. The activity around the electron α_e is defined as the sum of all particle flow candidates closer than $\Delta R < 0.3$ to the electron. The contribution of pileup from neutral particles is estimated to be proportional to the average energy density ρ in the event. An effective area A is derived in simulation so that ρA corresponds to the energy of the neutral particles in the cone of $\Delta R < 0.3$:

$$\alpha_\mathrm{e} = \sum_{\text{charged,notPU}} p_\mathrm{T} + \max\left(0, \sum_{\text{neutral}} p_\mathrm{T} - \rho A\right). \tag{8.2}$$

Again, two IDs are defined: one to identify the electron from the W-decay and one to guarantee a selection that does not share events with the dilepton channel. Both IDs require the electron to pass the tight MVA ID working point and to have a relative isolation of $I_\mathrm{e}^\mathrm{rel} = \alpha_\mathrm{e}/p_\mathrm{T} < 0.15$. The kinematic cuts are again chosen to stay above the trigger threshold and are summarized in Table 8.2.

Table 8.2: Summary of the cuts defining the two sets of electron identification criteria used in this analysis.

Observable	Tight cut	Loose cut
ID	MVA, tight	MVA, tight
p_T	> 30 GeV/c	> 15 GeV/c
$\|\eta\|$	< 2.1	< 2.4
$I_\mathrm{e}^\mathrm{rel}$	< 0.15	< 0.15

Table 8.3: Summary of the cuts defining the jet identification criterion.

Observable	Cut
ID	loose
p_T	> 30 GeV/c
$\|\eta\|$	< 2.4

8.1.4 Jets and missing transverse energy

Two types of jets clustering algorithms are used, anti-k_T with a distance parameter $D = 0.4$ and C/A jets with $D = 1.5$, both introduced in Section 3.4.5. They are used to cluster jets from particle-flow candidates not from pileup vertices ("charged hadron subtraction" [109]) using FastJet [110]. The C/A jets have a large radius and can contain all decay products of a boosted Higgs boson or hadronically decaying top quark. These jets will be referred to as fat jets and are described in Section 8.1.6. More commonly used at the LHC are jets clustered with the anti-k_T algorithm and a radius parameter of $D = 0.4$. Thus they will be referred to simply as "jets" throughout this thesis.

To ensure that jets do not originate in single particles or detector noise, quality cuts corresponding to a "loose" jet selection are applied. The contribution to the total jet energy of particle flow candidates identified as either neutral hadrons, charged leptons, or photons to the jet energy must each not be larger than 99%. Charged hadrons are required to contribute to the jet energy and every jet should contain at least two particles and at least one charged particle. Finally, to avoid double counting isolated electrons or muons, jets closer than $\Delta R = 0.4$ to loose or tight leptons are removed.

The kinematic cuts described in Table 8.3 are applied to identify jets that originate in $\mathrm{t\bar{t}H}$ decays. Most $\mathrm{t\bar{t}H}$ decays produce comparatively hard and central jets, so that jet p_T and $|\eta|$ thresholds of 30 GeV/c and 2.4 are chosen. Jets of lower p_T or higher $|\eta|$ are measured with low resolution and are difficult to b-tag. Furthermore, lower jet-p_T thresholds increase the number of jets from pileup or initial state radiation almost exponentially so that the 30 GeV/c threshold keeps the combinatorial background low while still allowing to reconstruct most quarks from the $\mathrm{t\bar{t}H}$ decay as jets.

The jet response and resolution is corrected using the factorized approach described in Section 3.4.5 and in [111]. For all jets fulfilling the loose jet ID described above and the very basic requirements of $|\eta| < 5.4$ and $p_\mathrm{T} > 10$ GeV/c the jet energy corrections are used to update the $\not\!\!E_\mathrm{T}$ calculation with the improved response for particles within jets, as also described in Section 3.4.5.

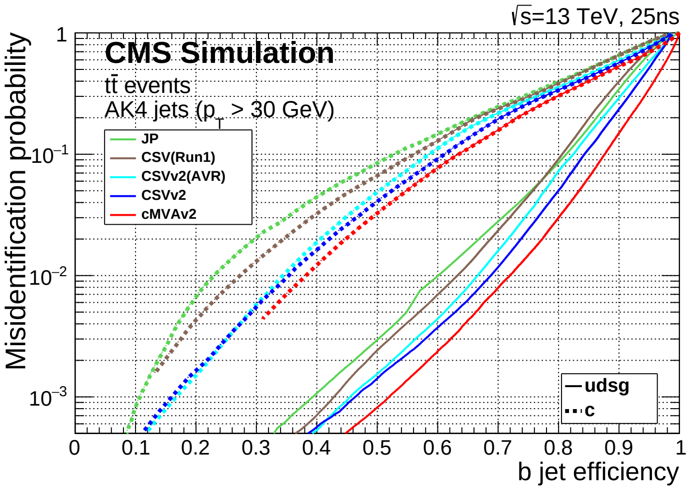


Figure 8.1: Comparison of the performance of b-tagging algorithms used at CMS in simulations of $\mathrm{t\bar{t}}$ events [113]. The graphs with solid (dotted) lines display the mistag probability of light jets (c-jets) for a given b-tagging efficiency. The color indicates the b-tagging algorithm.

8.1.5 b-tags

In this analysis the Combined Secondary Vertex b-tag algorithm version 2 is used (CSVv2) [113], described in more detail in Section 3.4.6. If not enough tracks are associated to a jet to use the CSVv2 algorithm, a default b-tag value of -0.1 is used in this analysis. The b-tagger output is used in two different ways in this analysis. It is employed to classify events according to the number of b-tags. Hereby a jet is classified as a b-jet if it passes the so called medium working point at a value of 0.8. For the $\mathrm{t\bar{t}}$ simulation and the jet cuts used in this analysis, this working point has a b-jet efficiency of 69% and a misidentification probability of 18% for c-jets and 1.3% for light jets. In addition to this use case, the full shape information of the b-tagger is used in the MVA discriminants described in the next chapter. The performance of the CSVv2 compared to other b-taggers at CMS can be deduced from Fig. 8.1. The only b-tagger that performs better is the CMVAv2 tagger, which is a combination of multiple b-tag discriminants and still in a more experimental stage, and thus not considered.

8.1.6 Fat jets

One of the innovations in this analysis with respect to the Run 1 version [22] is the use of fat jets and jet substructure techniques for the search for $\mathrm{t\bar{t}H}$ as first proposed in [220]. A more detailed description of the use of substructure techniques is given in [221], for which the setup used in this analysis has been implemented and optimized. The fat jets used are C/A jets clustered with a distance parameter $D = 1.5$. Goal of the large cone size is to be able to capture all decay products of a hadronically decaying top quark or a Higgs boson decaying into bottom quarks within the jet. This is possible if the top quark or, respectively, the Higgs boson carries a large transverse momentum. The angular distance of the decay products of a heavy resonance of mass m with transverse momentum p_T can be estimated as

$$\Delta R = \frac{1}{\sqrt{z(1-z)}}\frac{m}{p_\mathrm{T}}, \tag{8.3}$$

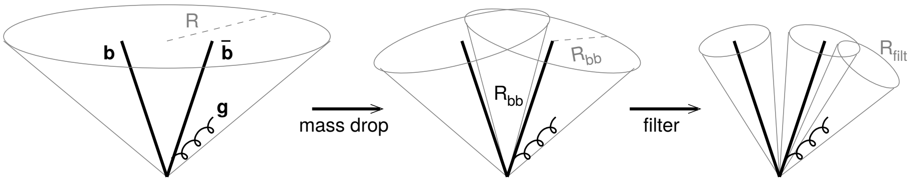


Figure 8.2: Illustration of the BDRS subjet algorithm, taken from [222]. Using a mass-drop criterion two subjets belonging to an $H(b\bar{b})$ decay are identified. In a filtering step radiation not belonging to the Higgs decay is removed.

where $0 < z < 1$ is the momentum fraction of the two decay products [222]. To be able to capture moderately boosted ($p_T \gtrsim 200$ GeV) top quarks and Higgs boson, a cone size parameter of $D = 1.5$ is used.

The C/A algorithm has advantages with respect to the anti-k_T algorithm when analyzing the jet substructure in terms of the clustering history. In the latter algorithm the distance measure is smaller if particles are harder. If two hard cores of particles are close to each other, particles tend to be clustered to the harder of both and only particles further apart than the distance parameter D are clustered into an independent jet. With the C/A algorithm close particles are clustered first, independent of p_T. This way the recombination of particles can be seen as a reversal of the QCD splitting. If two hard cores exist, with this algorithm there is a good chance that each forms one independent protojet and these jets are only recombined in a late step of the clustering.

Like the standard jets, fat jets are clustered from all particles not associated to pileup vertices. Additionally, all leptons identified with the loose electron or muon ID defined above are ignored in the clustering. These leptons are expected to originate in W boson decays and are thus irrelevant for the substructure analysis of hadronically decaying resonances and can even disturb the employed substructure algorithms. The same loose jet ID as for the default jets is applied. Furthermore, a selection is applied to identify jets that can contain top quark or Higgs decay products. Only fat-jets with $p_T > 200$ GeV and $|\eta| < 2.0$ are selected. The first requirement is motivated by Equation (8.3). The second requirement ensures that most of the particles in the jet are measured in the central region of the detector, where p_T measurements are more precise and b-tagging is possible. All of the so-selected fat jets are analyzed with the two substructure algorithms described below.

To subtract the effect of pileup and improve the fat jet response, corrections derived for anti-k_T 0.8 jets are used. Using jet energy corrections derived for jets with half the radius and a different clustering algorithm is a large extrapolation but the jet-area is taken into account in the correction and it can thus be hoped that it improves the jet-energy response.

8.1.7 Higgs identification from jet substructure

The BDRS subjet algorithm [222] serves as the basis of the Higgs identification. An illustration of the algorithm is given in Fig. 8.2. It is applied to all fat jets and can be described as follows.

1. Undo the last clustering step of the jet j with mass m_j and break it into two subjets j_1 and j_2 with $m_{j_1} > m_{j_2}$.

2. If there was a significant mass drop and the splitting is not too asymmetric:
   $m_{j_1} < \mu m_j \ \wedge \ \frac{\min\{p_T^2(j_1), p_T^2(j_2)\}}{m_j^2} \Delta R^2(j_1, j_2) > y_{\text{cut}}$,

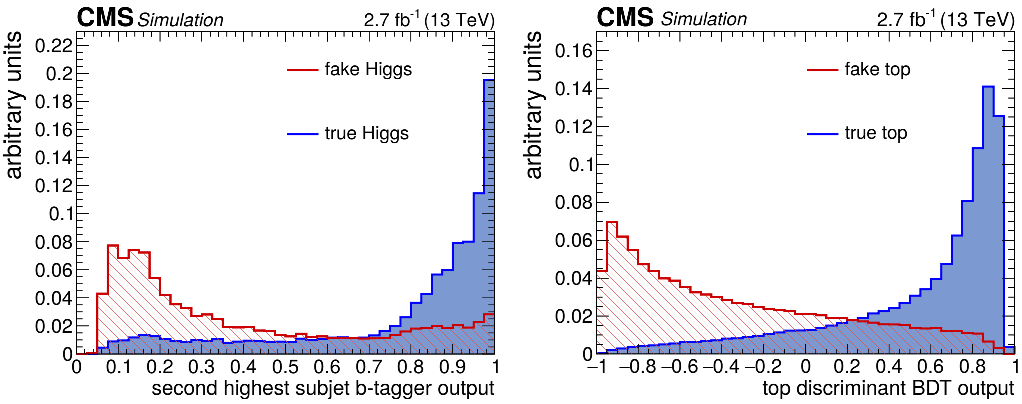


Figure 8.3: Discriminants used for the identification of boosted Higgs bosons and top quarks, as shown in [174]. The distinction between signal-jets and background-jets is made from angular matching between the two b quarks from the Higgs-boson decay and the fat jet. In the left-hand figure the distribution of the second highest b-tag among the b-tags of the filtered jets of the Higgs-boson candidate is shown. Signal and background jets are simulated using a $t\bar{t}H(b\bar{b})$ sample. In the right-hand figure the distribution of the BDT discriminant that is used to distinguish hadronically decaying top quarks from combinatorial background is shown. For this plot, signal and background shapes are simulated with a $t\bar{t}$ sample.

the undone recombination step is interpreted as the splitting of the heavy Higgs boson into two comparatively light b quarks. Go to step 4.

3. Otherwise, interpret j_2 as not belonging to the Higgs-boson decay products. Remove it, define $j := j_1$, and go to step 1.

4. Recluster the particles of j_1 and j_2 with the C/A algorithm and a distance parameter of $\min\{R_{\text{filt}}, \Delta R(j_1, j_2)/2\}$. This is called filtering and will yield at least two filtered jets.

5. Keep only the three hardest filtered jets. Additional soft jets are interpreted as coming from pileup or the underlying event.

The mass drop parameter was set to $\mu = 0.67$, the asymmetry is $y_{\text{cut}} = 0.3$, and the filter jet radius $R_{\text{filt}} = 0.3$, as recommended in [222].

The BDRS algorithm is further refined for the use in this analysis. The filter jets are corrected with the anti-k_{T} 0.4 jet-energy corrections and only filtered jets passing cuts of $p_{\text{T}} > 20$ GeV$/c$ and $|\eta| < 2.4$ are considered. Outside of this range jets are not well-measured and cannot be b-tagged at CMS. Furthermore, the CSVv2 b-tagging algorithm is run with the tracks associated to the particles of the filter jets and the output is used later in the analysis.

Additionally, the N-subjettiness τ_N [223] is calculated for the fat jets. It is a measure of how compatible the jets are with a N-subjet interpretation. The τ_2/τ_1-ratio is used together with the b-tags and invariant masses of the filtered jets in a multivariate analysis. The most powerful observable to distinguish fat jets from a Higgs boson is however the second highest b-tag among the filtered subjets. Its distribution for signal and background jets is displayed in Fig. 8.3 on the left-hand side.

8.1.8 Top-quark identification from jet substructure

The HEP top-tagger v2 [224, 225], including additional features described in [226], is used to analyze whether the substructure of a C/A 1.5 fat jet is compatible with the decay products of hadronically decaying top quark. The algorithm is similar to the Higgs-identification but specialized in identifying the features of top quark decays and can be summarized as follows:

1. Undo the last clustering step of the jet j with mass m_j and break it into two subjets j_1 and j_2 with $m_{j_1} > m_{j_2}$.
2. If there was a significant mass drop $m_{j_1} < \mu' m_j$, keep the jets j_1 and j_2, otherwise keep only j_1.
3. For all jets kept: if $m_{j_i} < m_{\max}$, add it to the list of valid subjets. Otherwise define $j := j_1$, and decluster from step 1 until all subjets fulfill $m_{j_i} < m_{\max}$ or are discarded.
4. For all triplets in the list of valid subjets: recluster the particles of the three jets with a radius of $\min\{R_{\mathrm{filt}}, \Delta R(j_1, j_2)/2\}$ to multiple filter jets. Use the triplet for which the five filter jets combined have the largest p_{T}.
5. Cluster the five hardest filtered jets to three subjets, again using the C/A algorithm. These subjets are interpreted as the three top quark decay products.

The parameters are kept at their default values of $\mu' = 0.8$ and $R_{\mathrm{filt}} = 0.3$.

The three subjets returned are corrected using the anti-k_{T} 0.4 jet energy corrections. While the original version of the HEP top-tagger imposes a number of cuts on the subjets, in this analysis a BDT is trained to identify top quarks. The training is performed on simulated $\mathrm{t\bar{t}}$ events. Jets that can be angular-matched to the three hadronic top decay products are taken as signal, background jets are those for which this is not possible. The CSVv2 b-tagging algorithm is evaluated with the three subjets and the one with the highest output is identified as the b quark candidate, the remaining two are identified with the light quarks from the W-boson decay. The invariant masses of pairs of subjets as well as the invariant mass of all three subjets and the fat jet are among the BDT inputs.

Further inputs are the N-subjettiness ratios τ_2/τ_1 and τ_3/τ_2 and $\Delta R_{\mathrm{opt.}}$ [226]. The last observable is determined by repeating the top-tagger algorithm (steps 1. to 5.) for decreasing fat-jet cone size. At a radius of $R_{\mathrm{opt.}}^{\mathrm{calc.}}$ the cone-size becomes too small to be expected to catch the top-quark decay products. The value is compared with the radius $R_{\mathrm{opt.}}$ at which a significant drop in the three-subjet mass is observed, indicating that the top-quark is indeed no longer contained in the jet and $\Delta R_{\mathrm{opt.}} = R_{\mathrm{opt.}} - R_{\mathrm{opt.}}^{\mathrm{calc.}}$ is calculated.

The shape of the BDT discriminant for signal and background jets is shown in Fig. 8.3 on the right-hand side.

8.2 Correction of physics objects and associated uncertainties

The CMS detector simulation describes the interaction of stable particles with the CMS detector with high accuracy. Nonetheless, a few differences between data and simulation can be observed. Most of the differences are corrected by measuring the relevant properties of physics objects in dedicated control samples. Even though these events can be very different from $\mathrm{t\bar{t}H}$ candidate events, it is usually a good approximation that physics objects with the same kinematic properties show the same data-MC differences.

8.2.1 Jets

The correction of jets is a complicated process and relies on the correct simulation of the jet response, i.e. the difference between reconstructed and generated jet-energy. Because this simulation is imperfect, the corrected jet energy in real data is slightly different from the one for equivalent jets from simulation. Two aspects of this disagreement are corrected using measurements. One is the correction of the jet energy scale, i.e. the average jet response for jets in a certain kinematic region. The second correction is a correction of the resolution, the difference between the jet energy on particle level and the measured energy. Both are described in more detail in [111] and are outlined below.

First, the energy scale of jets reconstructed from real data is corrected using the so called L2L3 residual corrections. They consist of a correction of the relative scale and a correction of the absolute scale. The relative correction takes into account relative differences of the jet response in η. For its derivation, jets in dijet events in which one jet falls in the central and the other in a different η-region are compared and it is required that they are balanced in the transverse direction. The absolute correction is derived by comparing the jet energy of central jets with well-measured objects in events in which they are expected to be balanced, e.g. Z(ee)+jets, Z($\mu\mu$)+jets, and γ+jet events. An example of the size and uncertainty of the jet energy scale (JES) correction is shown in Fig. 8.4.

The jet energy resolution is corrected in a second step. The idea of the derivation technique is that the average transverse imbalance of the momenta of jets in dijet events can be seen as a measure of the jet energy resolution. The larger the imbalance, the larger the mismeasurement of jets (see [227]). From a comparison of the resolution in data and simulated data, correction factors are derived as a function of η (see Fig. 8.4). In this analysis these corrections are accounted for by increasing the difference of the reconstructed jet to the particle level jet by the appropriate resolution correction factor. This correction factor was determined with an older version of the CMS detector simulation so that the associated uncertainties were conservatively doubled.

8.2.2 b-tags

In this analysis the full shape of the CSVv2 b-tagging discriminant is used. The b-tagging discriminant is calculated from a number of observables based on tracking information. These observables are very sensitive to the correct modeling and alignment of the tracking detector. Due to the imperfect modeling, the discriminant output is different between data and MC. For this analysis the full discriminant shape is corrected using a reweighting technique that is described in [113].

The shape correction is derived in two control regions. A region enriched in dilepton $\mathrm{t\bar{t}}$ decays is used to calibrate the discriminant shape of b-jets while the discriminant shape for light jets is calibrated in a Z+jets control region. Different corrections are derived dependent on η and p_T of the jet. The correction is constructed iteratively. In the Z+jets ($\mathrm{t\bar{t}}$) control region the b-jet (l-jet) background is subtracted from the data. From the difference between data and MC a correction for b-jets and light jets is derived as a function of the b-tagger output. The procedure is repeated, performing the background subtraction with corrected discriminant shapes, and new correction factors are derived

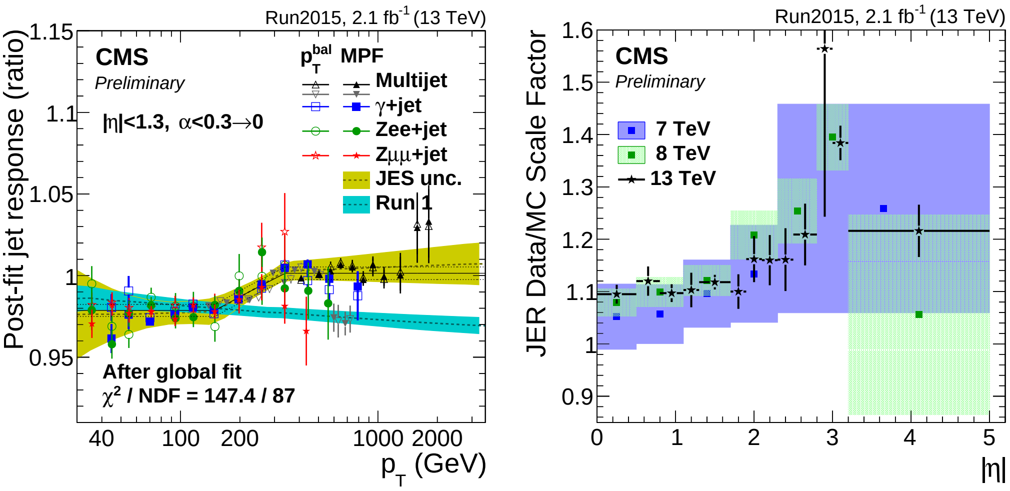


Figure 8.4: In the left plot the residual correction of the jet energy for jets with $|\eta| < 1.3$ is displayed as a function of p_T (solid black line). The yellow band corresponds to the uncertainty of the correction, the cyan band to the value and uncertainty determined in LHC Run 1. In the right plot the jet energy resolution corrections and their uncertainties as a function of η are displayed as black crosses and compared to the values obtained in measurements in LHC Run 1 at center-of-mass energies of 7 TeV and 8 TeV. Both taken from [112].

until the derived correction factors converge. The correction factor for one η-p_T region as well as the discriminant shape when applied in a region that is dominated by lepton+jets $\mathrm{t\bar{t}}$ decays is shown in Fig. 8.5.

The application of the b-tagger shape correction introduces a number of uncertainties. One is the uncertainty of the size of the light flavor (heavy flavor) background in the heavy flavor (light flavor) control region. Furthermore, the jet energy scale influences the result of the procedure. The statistical uncertainty due to the limited number of events in the control region is taken into account by allowing two kinds of shape shifts, for the heavy flavor and light flavor correction each. The “linear” shape uncertainty is the largest tilt of the correction function in upward and downward direction that can be parameterized by a linear function and is compatible with the statistical uncertainties. Similar, “quadratic” shape deviations are defined by a parabola centered in the middle of the distribution and allow to change the rate of extreme discriminant values compared to more central ones.

For c-jets no dedicated scale factors are derived. The default discriminant shape is assumed to be valid and as uncertainty twice the b-jet uncertainties are applied.

The flavor-dependent correction factors c_jet allow to reweight events containing N jets with b-tags b_j, flavors f_j, and kinematics $p_{\mathrm{T}j}$ and η_j so that events with specific b-tag and kinematic configuration in the MC simulation are weighted according to the probability to observe them in data. The event weight is computed as:

$$w_\text{b-tag} = \prod_{j=1}^{N} c_\mathrm{jet}(b_j, f_j, p_{\mathrm{T}j}, \eta_j). \tag{8.4}$$

8.2.3 Muons, electrons, and triggers

The simulation of both electrons and muons has shortcomings in terms of energy scale and resolution. However, the differences between data and simulation are smaller than for jets

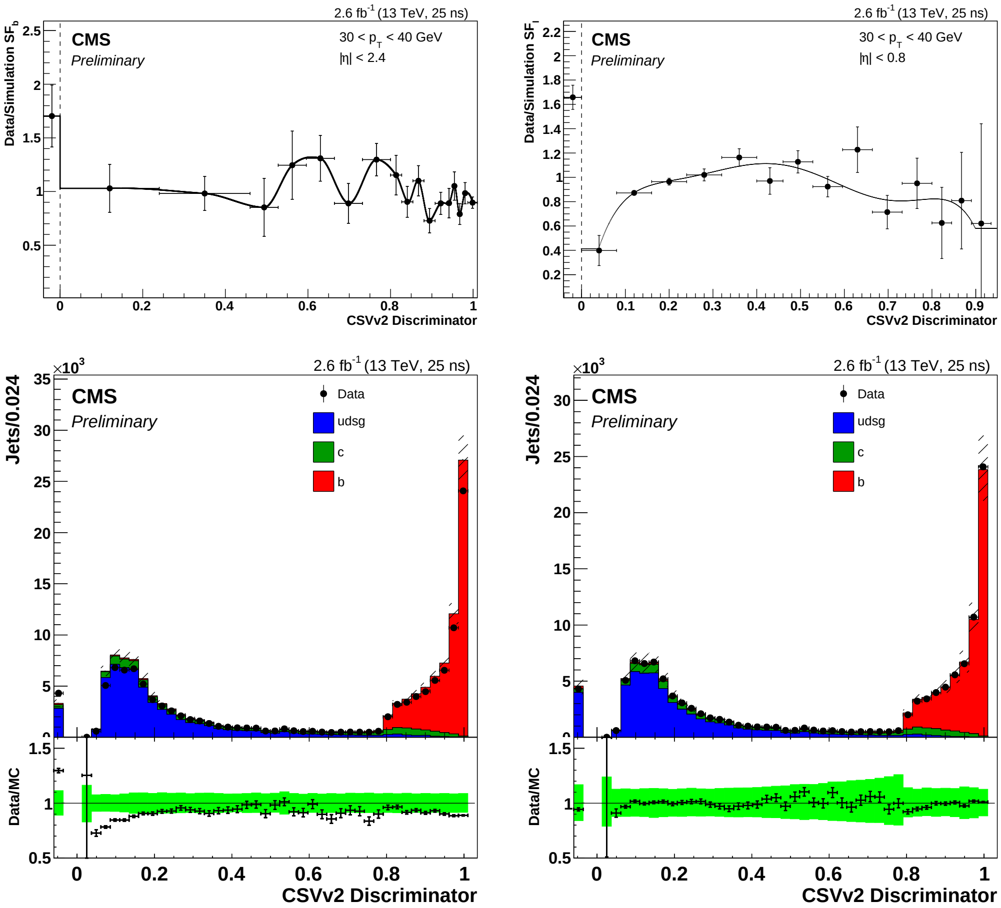


Figure 8.5: Top row: correction factor for b-jets (left) and central light jets (right) with 30 GeV/c < p_{T} < 40 GeV/c that is used in the b-tag reweighting. Bottom row: b-tag discriminant shape in a lepton+jets $t\bar{t}$ control region before the application of b-tag reweighting (left) and after (right). All figures are taken from [113].

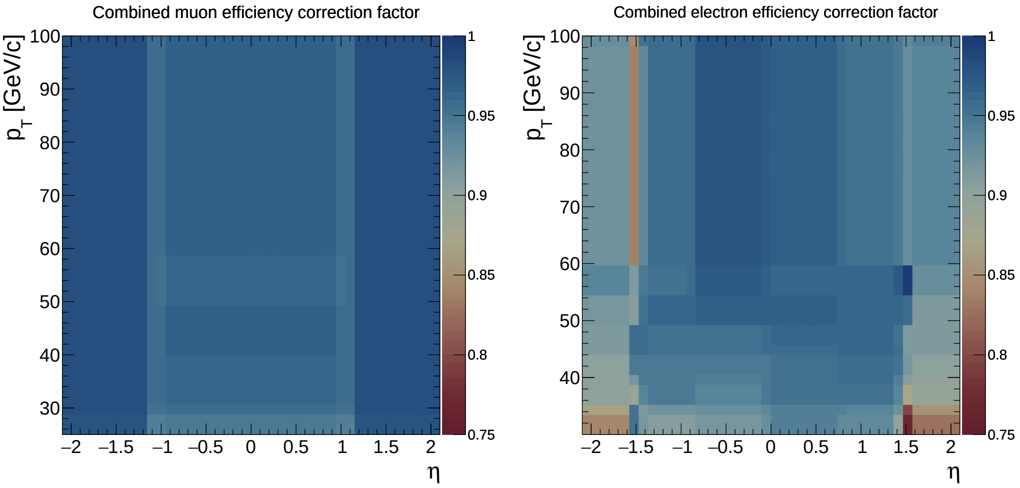


Figure 8.6: Combined efficiency correction for muons (left) and electrons (right) as a function of η and p_{T}. Differences between data and simulation for the efficiency of the trigger, the lepton identification, and the lepton isolation are corrected with the factors indicated by the color code.

and less important for this analysis. Therefore, no lepton scale or resolution correction is applied.

More important is the correction of the simulation of lepton identification and isolation efficiency as well as the trigger efficiency, as the assumed efficiency is inversely proportional to the measured cross section. These efficiencies are measured in Z(ee) and Z(μμ) events with a tag and probe method very similar to the ones described in [106] and [100]. The method is performed in events with one well-identified tag lepton, and a loosely identified probe lepton that forms the Z-boson mass with the first one. Correction factors for the identification, isolation, and trigger probability are then derived by comparing the fraction of events in that the probe lepton passes one of the three requirements in simulation with the fraction in data. The correction factors are derived as a function of p_{T} and η of the lepton and displayed in Fig. 8.6. Close to the p_{T} threshold and, for electrons (muons), in the barrel-endcap transition region of ECAL (muon detector), the correction is the largest.

Uncertainties of these measurements are estimated by varying the tag lepton selection, the statistical model to extract the result, and the bins in which the correction factors are calculated. The uncertainties of the three different measurements are added linearly because the uncertainties are conservatively estimated to be correlated. For muons the total rate uncertainty is 2%. For electrons the identification efficiency measurement was performed with simulations and data that was reconstructed with a different version of the CMS software. For that reason the uncertainty was increased, so that the total electron efficiency uncertainty is assumed to be 4%.

8.2.4 Fat jets and subjets

The b-tags of subjets reconstructed by the BDRS subjet algorithm and the HEP top-tagger behave very similar to the b-tags of anti-k_{T} 0.4 jets with the same kinematic properties and have a similar level of agreement between data and simulation. Thus it is assumed that they can be corrected by the scale factors derived for anti-k_{T} 0.4 jets and the same uncertainties are assumed. This is supported by similar observations made in [113]. In practice this means that the b-tag discriminant correction is calculated using only anti-k_T 0.4 jets. About 90% of the subjets used in this analysis can be matched to anti-k_T jets with very similar properties. The discriminant of these subjets is thus already corrected by the reweighting based on anti-k_T jets.

A similar approach is chosen for the jet energy corrections. It is assumed that the corrections derived with anti-k_T 0.4 jets are also valid for subjets and the associated uncertainties are applied correlated to the standard jet uncertainties.

8.3 Luminosity, data certification, and pileup

8.3.1 Luminosity

The time-dependency of the instantaneous luminosity at CMS is measured using observables that are proportional to the number of proton-proton interaction. The most accurate measurement can be performed using the mean number of pixel cluster in the pixel detector, as described in [228]. Roughly speaking, the more hits in the pixel detector are observed, the higher is the number of proton-proton interactions, the higher is the luminosity. The proportionality factor between pixel clusters and luminosity is determined from a Van der Meer scan. In such a scan the transverse position of the two beams is varied, which allows to determine the beam profile. From this profile and machine parameters the luminosity can be determined. Fig. 8.7 shows the integrated luminosity from proton-proton collisions in 2015. The data analyzed in this analysis was taken in the same time-frame.

8.3.2 Data certification

Not all of the 4.2 fb^{-1} of integrated luminosity delivered by the LHC in 2015 can be used for all physics measurements. The presented analysis uses a large number of physics objects. The CMS solenoid and most of the CMS subdetectors need to be operational during data-taking to allow reconstructing all of them with high quality.

CMS recorded data corresponding to 3.8 fb^{-1} of integrated luminosity. Due to problems with the cooling system of the solenoid, for only 3.0 fb^{-1} the magnetic field was at 3.8 T. Of these events an integrated luminosity of $\mathcal{L} = 2.67\ \mathrm{fb}^{-1} \pm 2.7\%$ was recorded with fully operational subdetectors and is analyzed in this theses. In about 20% of events problems with the hadronic forward calorimeters were observed that could however be mitigated in reconstruction and are not relevant for this analysis.

8.3.3 Pileup

The number of proton-proton collisions per bunch crossing is determined by two factors: the inelastic proton-proton cross section $\sigma_\mathrm{inel.}$ and the instantaneous luminosity during data-taking $L(t)$. The number of pileup interactions is then Poisson distributed with a mean of

$$\langle N^\mathrm{PU}_\mathrm{data}(t)\rangle = \sigma_\mathrm{inel.} L(t). \tag{8.5}$$

In MC samples pileup is simulated by overlaying additional inelastic proton-proton collisions with the process of interest. The distribution of the instantaneous luminosity is only known after data-taking, so that the pileup profile from the simulation has to be corrected. Using the instantaneous luminosity, the fraction of events f_data expected to

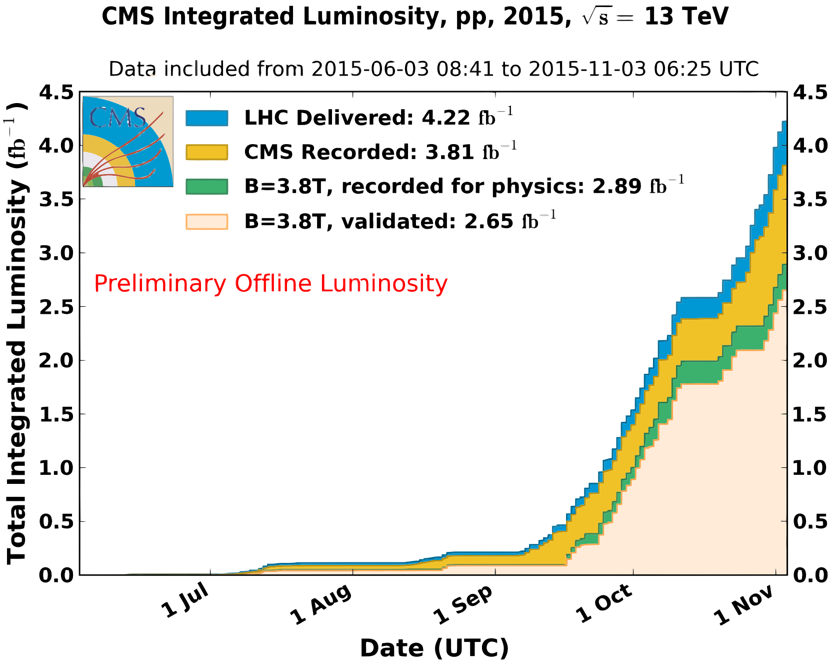


Figure 8.7: Integrated luminosity recorded by the CMS experiment in 2015 with a preliminary calibration, taken from [228]. The blue histogram shows the integrated luminosity that was delivered by the LHC, the orange histogram the fraction of the data that was recorded by CMS, the green histogram the fraction that was recorded with a magnetic field of 3.8 T, and the light yellow histogram the fraction for which all subdetectors were fully operational.

contain $\langle N^{\mathrm{PU}} \rangle$ proton-proton interactions can be calculated. For the MC simulation the number of interactions is sampled from a distribution p_{MC}. Events are then weighted by

$$w_{\mathrm{PU}} = \frac{f_{\mathrm{data}}(\langle N^{\mathrm{PU}} \rangle)}{p_{\mathrm{MC}}(\langle N^{\mathrm{PU}} \rangle)}, \tag{8.6}$$

For this analysis, a cross section of $\sigma_{\mathrm{inel.}} = 69.4$ mb has been assumed, which correctly predicts the number of reconstructed primary vertices per event. The cross section is varied by 5% to estimate the uncertainty introduced by this procedure.

9 Identification of signal events

Signal events are identified by different techniques, described in this chapter. Events are selected according to the criteria described in Section 9.1. In Section 9.2 the background model is validated with these selected events. Further signal-background separation is achieved by categorization (Section 9.3) and the use of Matrix Element Method (MEM, Section 9.4) and Boosted Decision Tree (BDT, Section 9.5) discriminants. The two discriminants are combined in Section 9.6.

9.1 Event selection

Goal of the event selection is to reject a large fraction of the background events while keeping most of the $\mathrm{t\bar{t}H(b\bar{b})}$ signal events. The selection can be summarized as follows:

1. **The primary vertex with the highest sum of squared track p_T is required to fulfill the vertex quality cuts.**
   This ensures that the vertex in which $\mathrm{t\bar{t}H}$ is most likely produced is well-reconstructed. Most events pass this cut.

2. **The event is required to contain exactly one tight lepton (electron or muon).**
   Tight leptons are defined in Chapter 8.1. This step greatly reduces the multijet background.

3. **For events with a tight muon the HLT_IsoMu20 and for events with a tight electron the HLT_Ele27_eta2p1_WPLoose_Gsf trigger must have fired.**
   These are the two single-lepton triggers with the largest efficiency for central and isolated leptons. The first requires the presence of an isolated muon with $p_\mathrm{T} > 20\ \mathrm{GeV}/c$, the latter an electron with $p_\mathrm{T} > 27\ \mathrm{GeV}/c$, $|\eta| < 2.1$, and fulfilling loose identification criteria. While there are triggers combining lepton and jet measurements that have a larger $\mathrm{t\bar{t}H}$ efficiency, they are not used because of their potential to distort distributions of jet kinematic observables which are important for the $\mathrm{t\bar{t}H}$ search.

4. **The event must not contain any loose leptons in addition to the tight lepton.**
   Loose leptons are defined in Chapter 8.1. This ensures that no dilepton $\mathrm{t\bar{t}}$ or Z+jet events are selected.

5. **The event must contain at least four jets of which three are b-tagged or at least six jets of which two are b-tagged.**
   This selection step reduces non-$\mathrm{t\bar{t}}$ background. There is one exception to this selection step: a small fraction of events that are categorized as “boosted” are also selected if the have two b-tags and four or five jets. The boosted category will be described in the next section.

Approximately 8% of all $t\bar{t}H(b\bar{b})$ events pass the event selection. 72% do not pass the lepton selection (mostly because the top-quarks decay hadronically), the remaining events mostly fail to meet the jet and b-tag requirements. Of the $t\bar{t}$ background only 0.7% of events pass the selection. W-boson events in which the W boson decays into leptons pass the event selection with a probability of only 0.001%. Of the trillions of events containing no prompt leptons only about 100 are expected to pass the selection.

To study the modeling of the $t\bar{t}$ background in the simulation, the last step of the event selection is relaxed to ≥ 4 jets and ≥ 2 b-tags. The correct description of the Z+jet background is checked for events with two tight leptons and for a W+jets enriched control region the b-tag cut can be dropped.

Relaxing the event selection for all of the statistical analysis is possible, too and is expected to lead to a small improvement in the sensitivity to the $t\bar{t}H$ signal. However, including a large number of events in the fit requires to be able to model a lot of background uncertainties. The model used in this analysis is not believed to be flexible enough to describe the background for a more inclusive event selection.

9.2 Background model validation

In the following section the ability of the background model to predict the observed data is demonstrated. It is beneficial if the nominal MC simulation already describes the data well. However, the statistical model used to extract the $t\bar{t}H$ signal also incorporates uncertainties. To visualize the flexibility of the model to fit the data, the event-yield change due to a variation of each systematic by one standard deviation is calculated for every histogram bin. The upward and downward changes due to the different systematics are added quadratically and shown as hashed bands.

The largest contribution to the total uncertainty is introduced by b-tagging. At the medium working-point, the uncertainty on the b-tagging efficiency is about 5%. Since two b-tags are required by the event selection, the $t\bar{t}$ event yield has an uncertainty of about 10%. The jet energy scale uncertainty allows to vary the jet p_{T} and can thus change the number of events containing four jets with more than 30 GeV/c transverse momentum, an effect that is of the order 5%. Furthermore, the number of events expected is uncertain due to the limited precision of the total cross section predictions. The $t\bar{t}$ rate uncertainty and the 50% additional uncertainty of the heavy flavor processes are the largest effect. B-tagging and JES also introduce relatively large uncertainties for the shape of b-tag and jet kinematic distributions. Likewise, the theoretical uncertainties due to scale choices in matrix element and parton shower generator can cause shape changing effects.

Comparisons of data and simulation are shown in Fig. 9.1 and Fig. 9.2, additional distributions can be found in Appendix A.1. After applying the corrections discussed in the previous chapter, the simulation predicts the distribution of the observed events well. The data is well within the statistical and systematic uncertainties and hardly any shape deviation between data and MC is visible.

One important exception is the distribution of the number of jets and the p_{T} distribution of soft jets. While the number of events with four and five jets is correctly predicted, there are 10% fewer six-jet events observed than predicted. This difference increases to a 20% overprediction of eight-jet events. One reason for this difference could be a systematically shifted JES. However, the reconstructed masses of the hadronically decaying W bosons and top quarks in $t\bar{t}$ events show no data-MC disagreement. Furthermore, in Z+jets events the distribution of jets is well-described so that this explanation is unlikely. A

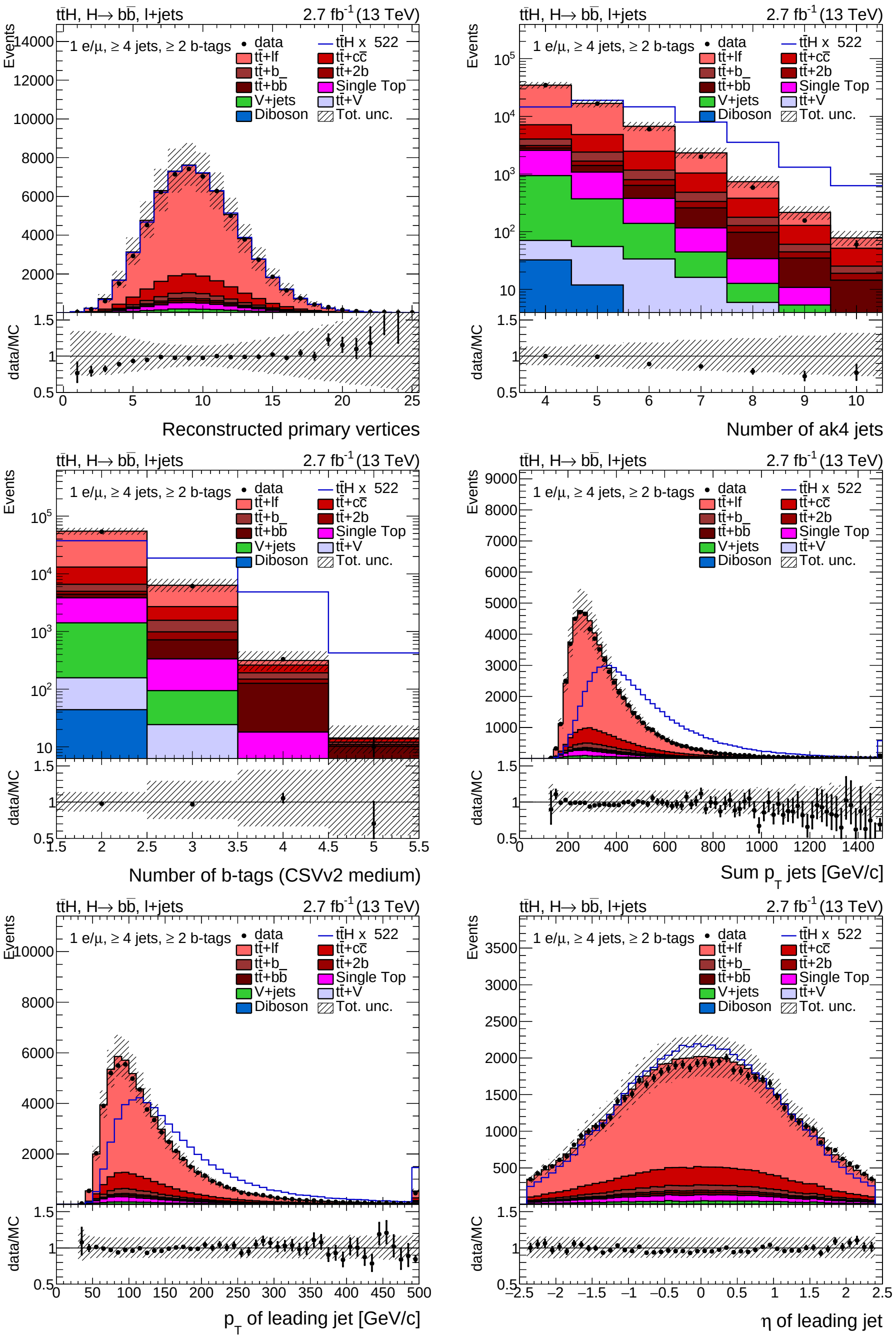


Figure 9.1: Comparison of the observed distribution of data recorded in 2015 (black dots) with the expected background predicted by simulation (stacked histograms) for different observables. Events are selected by requiring at least one lepton and four jets ($p_T >$ 30 GeV/c) of which two are b-tagged. The expected $t\bar{t}$H signal is drawn as a blue line and scaled to the same integral as the sum of backgrounds. The most important systematic uncertainties are added in quadrature and shown as hashed error bands. Statistical uncertainties due to the limited number of observed events are drawn as black horizontal lines on the data points.

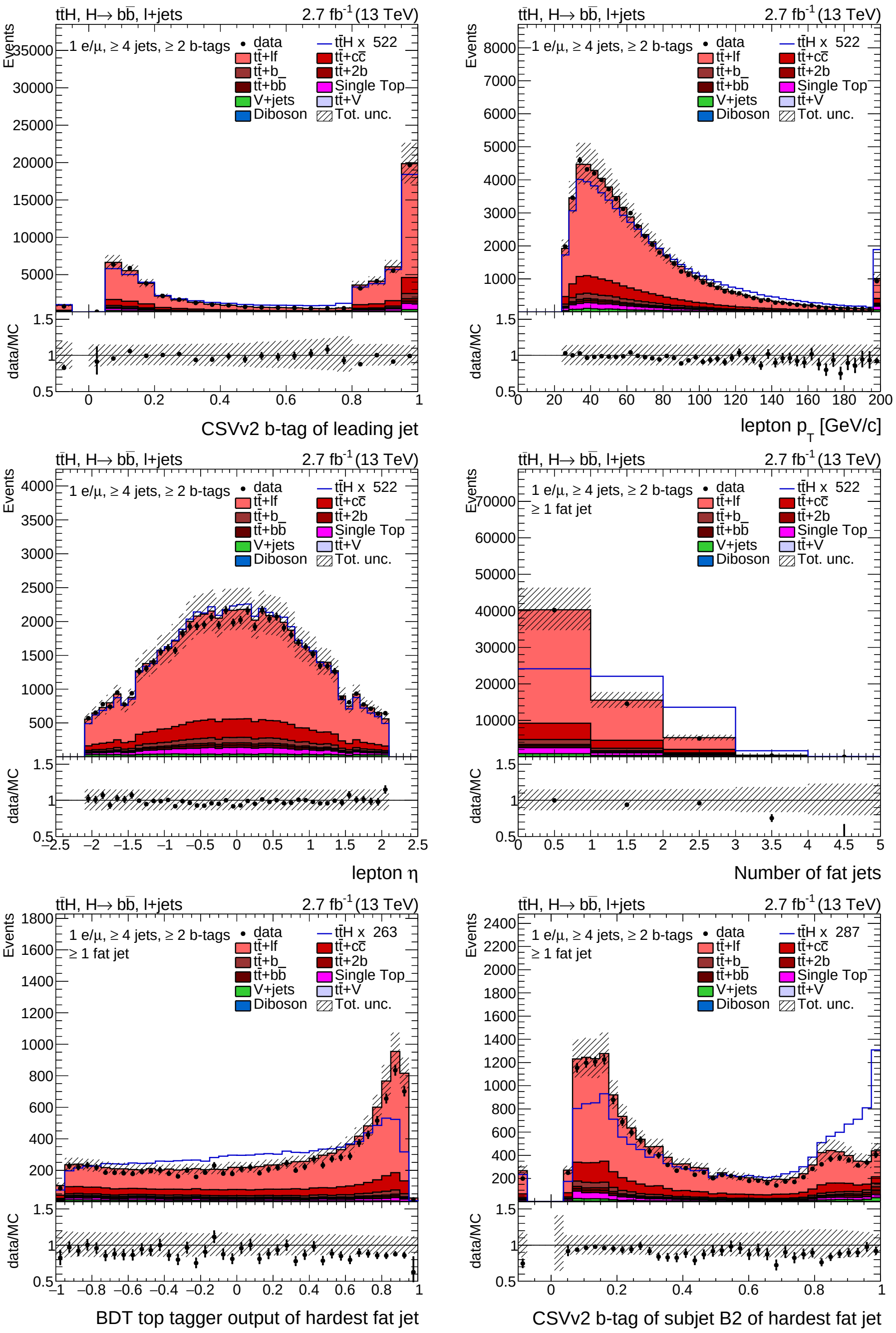


Figure 9.2: Comparison of the observed distribution of data recorded in 2015 (black dots) with the expected background predicted by simulation (stacked histograms) for different observables. Events are selected by requiring at least one lepton and four jets ($p_\mathrm{T} > 30\,\mathrm{GeV}/c$) of which two are b-tagged. The expected $\mathrm{t\bar{t}H}$ signal is drawn as a blue line and scaled to the same integral as the sum of backgrounds. The most important systematic uncertainties are added in quadrature and shown as hashed error bands. Statistical uncertainties due to the limited number of observed events are drawn as black horizontal lines on the data points.

more probable explanation is the subpar tuning of the MC generators used to simulate the $\mathrm{t\bar{t}}$ background. An upward shift of the matrix element scale (see Fig. 7.7) and, more importantly, an upward shift of the parton shower scale (Fig. 7.8) can reduce the number of six jet events by almost 10% while leaving the total cross section unchanged.

Most likely, the value chosen for the parton shower scale is responsible for the effect. This is supported by the observation that different matrix element generators interfaced with Pythia 8 show similar behavior. Furthermore, the data-MC differences increase up to 10 jets, while a maximum of 5 jets can be attributed to Powheg and the $\mathrm{t\bar{t}}$ decay in lepton+jets events. In fact, the value used for α_S for the production of initial state radiation with the Pythia 6 parton shower, which was the default generator at CMS in the LHC Run 1, is set to a lower value than the one used Pythia 8 in Run 2. This is expected to have a similar effect as an increased scale. With Pythia 6 the data-MC agreement is much better. The fact that the bad modeling of soft jets is not present in Z+jets events could mean that it only becomes important at the high energy scale of $\mathrm{t\bar{t}}$ production.

Ultimately, the data-MC differences are judged to be covered by the systematic uncertainties incorporated in the background model. The jet multiplicity distribution in data can be reproduced when changing the parton shower and matrix element scale by a factor of two, which corresponds to one standard deviation.

9.3 Categorization

After the event selection, the signal-to-background ratio is about 1/200. The background rate uncertainties are far too large to extract the $\mathrm{t\bar{t}H}$ production cross section by simply counting the number of selected events. To separate $\mathrm{t\bar{t}H}$ from the backgrounds, as a first step, events are categorized according to the number of anti-k_T jets and b-tags. The separation power of these variables should become clear from comparing the jet and b-tag multiplicity distribution between signal and background in Fig. 9.1. One class of events is excluded from this categorization: events containing two fat jets that can be interpreted as boosted Higgs-boson and top-quark candidates. The selection of these events is described below.

9.3.1 Boosted category

The search for $\mathrm{t\bar{t}H}$ in the boosted regime as proposed in [220] is performed as part of this analysis. Top quark and Higgs boson will be referred to as being boosted if they have transverse momenta $\gtrsim 200\ \mathrm{GeV}/c$. The search in the boosted regime is performed in a dedicated “boosted” category. The selection aims at picking events in which both, the Higgs boson and the hadronically decaying top quark are boosted. This is the case for approximately 7% of all simulated lepton+jets $\mathrm{t\bar{t}H}$ events.

Such events have the advantage that the combinatorial problem that is present when trying to reconstruct the Higgs boson is easier to solve. In a regular $\mathrm{t\bar{t}H(b\bar{b})}$ event, the four b-jets from the Higgs-boson and top-quark decay have very similar kinematic properties and scatter in all directions. Furthermore, the light quarks from the W-boson decay are very similar to initial state radiation jets. Contrary, in a boosted event the three decay products from a hadronically decaying top quark are collimated in the direction of the hadronically decaying top quark. The boosted Higgs boson then typically decays into the opposite detector hemisphere.

Candidates for boosted events are required to contain a tight lepton and two C/A 1.5

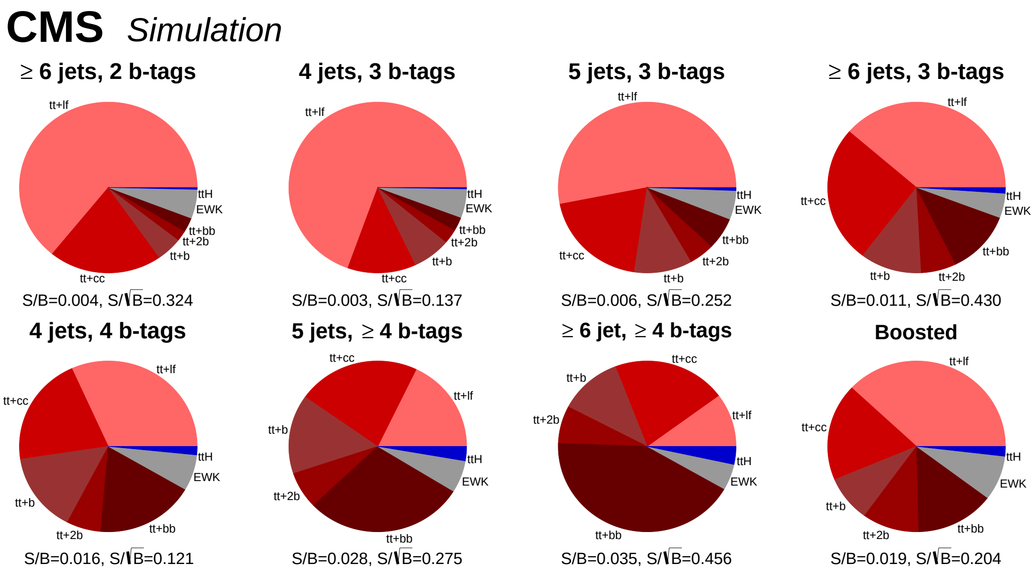


Figure 9.3: Expected signal (blue) and background (red and gray colors) composition in the analysis categories. Below the pie charts the expected $t\bar{t}H$ to background ratio S/B and the expected statistical significance of the signal $S/\sqrt{B}$ is displayed. Taken from [229].

jets with $p_T > 200\ \mathrm{GeV}/c$. The events are further selected employing the Higgs-boson identification and top-quark identification discriminants described in Section 8.1.6 and displayed in Fig. 8.3. First, the C/A jet with the largest top-tagger output is identified as the top-quark candidate. The event is only selected if the top-tagger output is above a certain threshold. Of the the remaining C/A jets the one with the highest Higgs-boson discriminant (defined as the second highest b-tag among the subjets) is identified as the Higgs-boson candidate. The Higgs-boson discriminant has to be above an optimized value for the event to be accepted in the “boosted” category. Picking the top-quark candidate first improves the efficiency of selecting boosted $t\bar{t}H$ events because (parts of) the top-quark decay products are more likely to be misidentified as Higgs-boson candidates than Higgs-boson decay products are misidentified as top-quark candidates. The thresholds of the two taggers are determined such that the signal efficiency is maximized for a given background efficiency. The working point chosen requires the top-tag to be greater than -0.485 and the two Higgs-candidate subjets b-tags to be greater than 0.8925 (see [221] for a derivation of these working points).

Approximately one hundred background events and two signal events are expected in the boosted category. The probability that the true momentum vectors of boosted top quark and boosted Higgs boson are both within $\Delta R < 0.5$ of the respective candidate fat jets is 43%. This is at least twice as high as the values achieved by reconstructing the $t\bar{t}H$ system in the remaining categories, using anti-k_T 0.4 jets, which demonstrates the improved handling of the combinatorial problem.

9.3.2 Analysis categories

Events that pass the basic event selection described in the beginning of this chapter but which are not in the boosted category are categorized according to the number of anti-k_T jets and b-tags into seven additional categories. Giving the boosted category preference over the other categories improves the overall analysis sensitivity. The categories are, ordered by the ratio of expected $\mathrm{t\bar{t}H}$ and background events (S/B), starting from lowest S/B:

- 4 jets, 3 b-tags
- $\geq$ 6 jets, 2 b-tags
- 5 jets, 3 b-tags
- $\geq$ 6 jets, 3 b-tags
- 4 jets, 4 b-tags
- boosted
- 5 jets, $\geq$ 4 b-tags
- $\geq$ 6 jets, $\geq$ 4 b-tags

In addition to the different S/B, the categories differ in the composition of backgrounds which can be seen in Fig. 9.3. While it is possible to extract a signal in the background-dominated categories with low significance, they mainly act as control regions for the $\mathrm{t\bar{t}}+\mathrm{hf}$ backgrounds. None of the categories is completely dominated by one background type but the combination of the categories allows, in principle, to constrain all of them.

9.3.3 Event yields

The number of events predicted by the simulation of the different background processes and the number of events observed in 2015 are shown in Table 9.1. Most of the events (≈ 8000) are expected in the $\geq$ 6 jets, 2 b-tags category. 6000 are expected in the three-b-tag categories and only 400 in the four b-tag and boosted category. In the latter categories the S/B ratio is better, between 1/60 and 1/30 compared to 1/300 to 1/100 in the former. Comparing the prediction to the observation, a small overprediction of six jet events and large underprediction of events in the 4 jets, 4 b-tags category by the simulation is visible. These differences are covered by the systematic uncertainties.

9.4 Matrix Element Method and b-tag likelihood ratio

In addition to the categorization of events according to jets and tags, signal and background events are separated further with the use of multivariate techniques. For this analysis, the Boosted Decision Trees (BDTs) and the Matrix Element Method (MEM) are combined to identify the $\mathrm{t\bar{t}H}$ signal. This is an innovation with respect to Run 1, in which BDTs and MEM were used in two separate analyses [22, 23].

The implementation of the Matrix Element Method (see Section 4.3.1 for an introduction to the concept) used in this analysis is similar to the one described in the Run 1 analysis [23][1]. The hard scattering amplitude is evaluated with OpenLoops [230] and the integration performed with the VEGAS [231] algorithm. Only for jet p_T and for $\not{E}_\mathrm{T}$

---

[1] For the presented analysis the Matrix Element Method and the b-tagging likelihood ratio have been implemented as a self-contained software package by Joosep Pata (ETH Zürich) and are thus only described briefly here.

Table 9.1: Number of events expected in the eight analysis categories due to the signal and background processes and observed events in 2015. The given uncertainties correspond to the influence of the most important systematic and statistical uncertainties of the MC simulation, added in quadrature.

Process	≥ 6 jets, 2 b-tags	4 jets, 3 b-tags	5 jets, 3 b-tags	≥ 6 jets, 3 b-tags
$t\bar{t}$ + lf	5359.3 ± 1226.3	2026.1 ± 651.4	1000.2 ± 352.9	589.5 ± 199.7
$t\bar{t}$ +$c\bar{c}$	1722.2 ± 849.5	363.2 ± 190.9	368.1 ± 191.3	396.6 ± 209.5
$t\bar{t}$ +b	393.7 ± 188.2	203.1 ± 92.5	199.6 ± 90.8	170.8 ± 81.4
$t\bar{t}$ +2b	165.2 ± 81.2	78.9 ± 38.0	87.2 ± 40.7	97.3 ± 46.8
$t\bar{t}$ +$b\bar{b}$	226.4 ± 113.2	75.8 ± 35.3	114.1 ± 52.3	183.7 ± 86.7
Single Top	283.0 ± 49.0	115.3 ± 30.8	76.2 ± 19.5	47.5 ± 12.7
V+jets	130.5 ± 35.2	38.6 ± 17.8	22.8 ± 10.4	13.6 ± 6.4
$t\bar{t}$ +V	43.5 ± 8.2	4.3 ± 1.2	6.4 ± 1.8	10.0 ± 2.7
Diboson	2.8 ± 1.3	2.1 ± 1.3	0.9 ± 0.5	0.2 ± 0.3
Total bkg	8326.7 ± 1788.6	2907.4 ± 836.5	1875.5 ± 534.7	1509.1 ± 423.7
$t\bar{t}H$	29.6 ± 2.1	7.4 ± 1.0	10.9 ± 1.2	16.7 ± 2.1
Data	7185	2793	1914	1386
S/B	0.0036	0.0026	0.0059	0.011
Data/B	0.9 ± 0.2	1.0 ± 0.3	1.0 ± 0.3	0.9 ± 0.3

Process	4 jets, 4 b-tags	5 jets, ≥ 4 b-tags	≥ 6 jets, ≥ 4 b-tags	boosted
$t\bar{t}$ +lf	17.8 ± 10.8	17.7 ± 10.9	17.6 ± 11.3	45.1 ± 9.4
$t\bar{t}$ +$c\bar{c}$	11.6 ± 8.2	22.1 ± 15.4	35.9 ± 24.9	21.8 ± 12.0
$t\bar{t}$ +b	8.4 ± 4.4	14.8 ± 7.7	20.0 ± 10.9	10.3 ± 5.5
$t\bar{t}$ +2b	3.5 ± 1.9	6.9 ± 3.7	12.3 ± 6.9	12.3 ± 6.6
$t\bar{t}$ +$b\bar{b}$	10.1 ± 4.9	28.8 ± 13.9	73.4 ± 36.6	17.0 ± 8.4
Single Top	2.5 ± 1.1	4.3 ± 1.4	5.5 ± 2.0	7.0 ± 1.7
V+jets	1.0 ± 0.8	0.9 ± 0.8	1.4 ± 0.7	2.5 ± 0.8
$t\bar{t}$ +V	0.3 ± 0.1	0.7 ± 0.3	1.6 ± 0.6	0.9 ± 0.3
Diboson	0.0 ± 0.0	0.1 ± 0.1	0.0 ± 0.0	0.1 ± 0.1
Total bkg	55.2 ± 23.0	96.5 ± 37.6	167.6 ± 65.7	117.0 ± 24.9
$t\bar{t}H$	0.9 ± 0.2	2.7 ± 0.6	5.9 ± 1.4	2.2 ± 0.3
Data	75	104	150	104
S/B	0.017	0.028	0.035	0.019
Data/B	1.4 ± 0.5	1.1 ± 0.4	0.9 ± 0.4	0.9 ± 0.2

transfer functions are used, the lepton momentum and jet-direction is thus assumed to be measured perfectly. For a given quark transverse momentum of $p_{\mathrm{T}_{\mathrm{gen}}}$ the probability to measure a jet p_{T} is modeled by the sum of two normal distributions of different width:

$$f(p_{\mathrm{T}}|p_{\mathrm{T}_{\mathrm{gen}}}) = N\left[0.7\exp\left(\frac{p_{\mathrm{T}}-\lambda_1}{\lambda_2}\right)^2 + 0.3\exp\left(\frac{p_{\mathrm{T}}-\lambda_3}{\lambda_2+\lambda_4}\right)^2\right]. \tag{9.1}$$

The parameters λ_i are first determined from the $\mathrm{t\bar{t}H}$ MC simulation in bins of η and p_{T}. Then a continuous p_{T} dependence of the λ_i is extracted as a polynomial in a fit of the values obtained in the p_{T} bins. The transfer function for $\not{E}_{\mathrm{T}}$ is a two-dimensional Gaussian distribution with a variance of $(20\ \mathrm{GeV})^2$ in x and y direction.

The discriminant constructed with the matrix element method is the likelihood ratio

$$d_{\mathrm{MEM}} = \frac{L_{\mathrm{t\bar{t}H}}(\{p_i\})}{L_{\mathrm{t\bar{t}H}}(\{p_i\}) + cL_{\mathrm{t\bar{t}b\bar{b}}}(\{p_i\})}, \tag{9.2}$$

where p_i are energy and momentum of jets, lepton, and $\not{E}_{\mathrm{T}}$ and c is chosen to be 0.15 so that the discriminant is more evenly distributed between 0 and 1.

Only the four jets with the highest compatibility with a b quark interpretation are considered in the calculation of the MEM discriminant. The compatibility with the b quark interpretation is determined with the help of a b-tagging likelihood ratio similar to the one already used in [23]. It is the ratio of the likelihood that the event is an event with four b-jets (e.g. $\mathrm{t\bar{t}b\bar{b}}$ or $\mathrm{t\bar{t}H}$) and the two b-tag likelihood (for example for a $\mathrm{t\bar{t}}+\mathrm{lf}$ event). The likelihood ratio is implemented as a function of the b-tags b_i and jet momenta $\vec{p}_i$:

$$d_{\text{b-tag}} = \frac{L_{\mathrm{t\bar{t}+hf}}(\{d_i\},\{\vec{p}_i\})}{L_{\mathrm{t\bar{t}+hf}}(\{d_i\},\{\vec{p}_i\}) + L_{\mathrm{t\bar{t}+lf}}(\{d_i\},\{\vec{p}_i\})} \tag{9.3}$$

The likelihoods $L_{\mathrm{t\bar{t}+hf}}$ and $L_{\mathrm{t\bar{t}+lf}}$ are calculated by adding up all permutation of jets for which two, respectively four, are interpreted as b-jets, e.g. for the heavy flavor interpretation of a 6-jet event,

$$L_{\mathrm{t\bar{t}+hf}}(\{d_i\},\{\vec{p}_i\}) = \sum_{i_1}\sum_{i_2\neq i_1}\cdots\sum_{i_6\neq i_1,\ldots,i_5}\prod_{j\in\{i_1,i_2,i_3,i_4\}} L_{\text{hf-jet}}(d_j,\vec{p}_j)\prod_{k\in\{i_5,i_6\}} L_{\text{lf-jet}}(d_k,\vec{p}_k). \tag{9.4}$$

The jet-likelihood $L_{\text{hf-jet}}$ ($L_{\text{lf-jet}}$) is read from histograms of the CSVv2 b-tag distribution of b-jets (light and c-jets) obtained from a $\mathrm{t\bar{t}}$ simulation. To account for the dependency of b-tagging on the jet kinematics, different reference histograms are used depending on p_{T} and η of the jets. Finally, only for events in which the likelihood ratio $\lambda_{\text{b-tag}}$ passes a certain threshold the MEM discriminant is calculated. The assignment of jets to b quarks is the one yielding the largest contribution to the sum in Equation (9.4). The MEM discriminant is then evaluated for all possible permutations of the four jets, but ignoring permutations in which the b quarks assigned to the Higgs boson are swapped, requiring thus $4!/2 = 12$ evaluations per event.

An important difference with respect to the Run 1 analysis is that the MEM discriminant only depends on the four b-jets. Light jets are irrelevant for the MEM discriminant used in this analysis as light quarks from W boson decays are marginalized in the phase space integration. This makes the application of the MEM more flexible and possible in all events with at least four jets. Matching light jets to the W boson decay products in events with six or more jets does not improve the discriminant. The main reason for this is that light jets that originate in W boson decays oftentimes have low p_{T} so that in about 50%

of $\mathrm{t\bar{t}H}$ events with six jets and four b-tags at least one of the light jets is from initial state radiation while the jet from the W-decay does not pass the jet cuts. Identifying an initial state radiation jet with the W boson degrades the performance of the MEM discriminant more than is justified by the improvement that correctly matching a W-decay quark to a jet yields.

9.5 Boosted Decision Trees

In addition to the MEM discriminant, Boosted Decision Trees (BDTs) are used to separate $\mathrm{t\bar{t}H}$ from its background processes. Gradient boosted BDTs of the TMVA [131] software package, trained with MC simulations of the $\mathrm{t\bar{t}H}$ signal and the inclusive $\mathrm{t\bar{t}}$ background are used for this purpose.

9.5.1 BDT input candidates

BDTs are trained separately in the eight categories described in Section 9.3. The training of a dedicated BDT in each category is motivated by the different topologies and background compositions because of which different features are important to distinguish the signal from the backgrounds. The training of the BDTs starts by identifying such features in all categories. The features can be classified as follows.

Object kinematics

A first class of features are simple functions of the jet, $\not{E}_{\mathrm{T}}$, and lepton energy and momentum. Examples are the jet-p_{T} of the leading jets, the sum of all jet-p_{T}, or the invariant mass of all physics objects. Using these observables is motivated by the observation that $\mathrm{t\bar{t}H}$ events happen at a higher center-of-mass energy than most of the background events and distribute more energy in the transverse direction, due to the production and decay of the three heavy particles.

Angular correlations

A second class of features are generated by creating more complex observables that identify angular correlations between the momentum vectors of the physics objects. Examples are the distance between jets in η or in the η-ϕ plane. For example, the largest difference in η between two jets is on average larger in $\mathrm{t\bar{t}}$ +jets events than in $\mathrm{t\bar{t}H}$ events.

$\mathrm{t\bar{t}H}$ reconstruction

Another class of features can be constructed by reconstructing the top-quark and Higgs-boson momenta and especially the Higgs-boson mass. There is no unambiguous assignment of two of the many final state jets to the Higgs-boson candidate so that several reconstruction methods are employed. Two simple methods assign the two b-tagged jets closest in the η-ϕ plane to the Higgs boson or the two jets forming the invariant mass closest to 125 GeV/c^2. A more complex method is the χ^2-reconstruction of the $\mathrm{t\bar{t}}$ system. In this method, the leptonically decaying W boson $\mathrm{W}^{\mathrm{lep}}_{\mathrm{rec}}$ is reconstructed from the isolated lepton and the transverse components of the neutrino momentum, measured as missing transverse momentum. Two solutions for the longitudinal neutrino momentum are calculated from a quadratic equation by imposing the constraint $(E_l + E_\nu)^2 - (\vec{p}_l + \vec{p}_\nu)^2 = m_W$. After that, the assignment of the six jets to the remaining $\mathrm{t\bar{t}H}$ decay products, two light quarks and four b quarks, is chosen as the one yielding the smallest χ^2, defined as

$$\chi^2 = \left(\frac{m^{\mathrm{rec}}_{\mathrm{W_{had}}} - m^{\mathrm{MC}}_{\mathrm{W_{had}}}}{\sigma_{\mathrm{W_{had}}}}\right)^2 + \left(\frac{m^{\mathrm{rec}}_{\mathrm{t_{had}}} - m^{\mathrm{MC}}_{\mathrm{t_{had}}}}{\sigma_{\mathrm{t_{had}}}}\right)^2 + \left(\frac{m^{\mathrm{rec}}_{\mathrm{t_{lep}}} - m^{\mathrm{MC}}_{\mathrm{t_{lep}}}}{\sigma_{\mathrm{t_{lep}}}}\right)^2 . \tag{9.5}$$

Two jets form the reconstructed W-mass $m^{\mathrm{rec}}_{\mathrm{W_{had}}}$ and together with a third jet they form the mass of the hadronically decaying top quark $m^{\mathrm{rec}}_{\mathrm{t_{had}}}$. Hypotheses for the leptonic top-mass $m^{\mathrm{rec}}_{\mathrm{t_{lep}}}$ are reconstructed from one of the two leptonic W hypotheses and another jet. Furthermore, it is required that all b-tagged jets are assigned to b quarks. No Higgs-mass constraint is applied to avoid a signal-like sculpting of the background. If several reconstruction hypotheses yield the same χ^2, the one for which the two Higgs-boson jets have the highest p_{T} are chosen. The expectation values and standard deviations m^{MC} and σ in Equation (9.5) are determined from the MC simulation by a Gaussian fit to the core of the invariant mass distributions for correctly reconstructed objects.

Event shape

Another class of features is event shape variables.In this analysis the sphericity S and aplanarity A [232] are used which are calculated from the eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$ of the energy momentum tensor of the physics objects:

$$S = \frac{3}{2}(\lambda_2 + \lambda_3), \tag{9.6}$$

$$A = \frac{3}{2}\lambda_3. \tag{9.7}$$

The energy is distributed more evenly in $\mathrm{t\bar{t}H}$ events, leading to larger values of S and A than in $\mathrm{t\bar{t}}$ events. Furthermore, the Fox-Wolfram moments [233] are calculated for all events as a function of the jet energies and momenta as

$$H_l = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{E^2_{\mathrm{tot}}} P_l(\cos(\phi(\vec{p}_i, \vec{p}_j))), \tag{9.8}$$

where the sum is evaluated for all pairs of jets, P_l are the Legendre polynomials, and $\phi(p_i, p_j)$ the angle between the two jet momenta. The Fox-Wolfram moments can be used to describe the energy distribution as a series of rotationally invariant functions.

b-tagging

The last class of features are the b-tags of jets. These are evaluated for standard anti-k_{T} jets and the subjets created by the HEP top-tagger and the BDRS algorithm. For the training, either specific b-tags like the second highest b-tag of all b-tagged jets or a combination of b-tags are used. One more complicated combination is the b-tagging likelihood ratio which was defined in Equation (9.3). For the multivariate analysis, the likelihood ratio is transformed to be more evenly distributed on a linear scale:

$$d_{\mathrm{blr}} \to \log\left(\frac{d_{\mathrm{blr}}}{1 - d_{\mathrm{blr}}}\right). \tag{9.9}$$

Fat-jet substructure

In the “boosted” category, kinematics of the C/A 1.5 jets, their substructure, and their subjets are analyzed. Substructure variables include for example the invariant mass of BDRS subjets or the ratio of the 2-subjettiness to the 1-subjettiness of the fat jet as introduced in Section 8.1.6.

9.5.2 BDT training

As described in 4.3.2, BDTs have multiple configurable parameters and can only be trained with a limited number of features. For the BDTs used for the signal-background separation the following parameters are tuned: the number of decision trees $\mathrm{N_{trees}}$, the learning rate reduction factor “shrinkage”, the bagging fraction, and the number of cut values tried to find the optimal signal-background separation, $\mathrm{N_{cuts}}$. The depth of the decision trees for this analysis is two, meaning that every tree contains four final nodes. For an optimal BDT performance, both the variable set and the BDT parameter configuration have to be optimized at the same time.

Sample splitting

This optimization is performed with one half of the simulated $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}H}$ events, the optimization sample. The remainder of the analysis is performed with the second, statistically independent half. The reason for this precaution is the automated optimization procedure explained later in this section. The procedure can lead to BDTs that perform better on the optimization sample than on an independent sample (e.g. the events that are actually measured) so that the measurement would be biased if the optimization sample was used. The optimization sample is split again into two parts, a training and a test sample, so that the final splitting of events is 50% in the analysis sample, 25% in the training sample, and 25% in the test sample. As the name suggests, the training samples are used for BDT training. Due to overtraining (see Section 4.3.2), the BDT performance is usually better on this samples than on an independent samples and so the BDT performance has to be evaluated with the test samples.

Figure of merit

The goal of the optimization is to increase the sensitivity of the analysis, e.g. the expected limit (as introduced in Section 4.1.3) on the $\mathrm{t\bar{t}H}$ production cross section. However, calculating the expected limit for every new BDT configuration is time-consuming so that a correlated figure of merit is used, the area under the receiver operating characteristic (ROC).

The ROC curve is calculated by simulating the expected BDT output distributions for signal and background with the respective test samples It represents the background rejection for a given signal efficiency. In general, ROC curves can be used to compare the performance of different classifiers: a classifier is better if it achieves a higher background rejection for every signal efficiency. If the ROC curves of two classifiers intersect, the picture is less clear because both classifiers can have advantages under different circumstances. To still be able to compare the discriminants with a single figure, the integral under the curve is calculated.

Particle Swarm Optimization

In every category, the ROC integral is optimized by varying the BDT's inputs and configuration, then training it on the training sample and calculating the output and ROC integral from the test sample. The optimization is performed using an iterative algorithm based on the Particle Swarm Optimization [234] which has been adapted for the use with BDTs in [235]. The "particles" in this algorithm are BDTs that are reconfigured iteratively and thus move through the multi-dimensional space of BDT configurations. Multiple BDTs are initialized at different points $\mathbf{x}_i$ in the configuration space and they have different initial velocities $\mathbf{v}_i$, i.e. the parameters between BDT iterations are modified with a certain step size. In addition to the different configurations, the BDTs start out with a randomly chosen subset of input variables.

In every step of the algorithm, the effect of removing input variables and adding additional variables of the pool of inputs is tested. The BDTs are trained and the ROC integral is calculated on the test sample. If it improves, a Kolmogorov-Smirnov test (KS test) is performed with the classifier distributions of test and training sample to check for significant difference between the performance on training and test sample. If the result of the test is a probability of less than 5%, it is assumed that the difference is due to overtraining or extreme statistical fluctuations and thus only accidental. Otherwise the variable set is updated including the new variable or without the removed variable, respectively. Dropping the KS-test requirement can lead to a selection of BDTs that perform accidentally well on the given test sample but not on an independent sample. This "over optimization" is also the reason for splitting the samples into three parts – due to over optimization only events that are not in test sample can be compared to data and due to overtraining the training events may not be used.

Once the variable sets of the BDTs are updated, the particle velocities are updated. The new velocity vector is calculated from the old velocity by adding two terms: one tends to move the particle in direction of the point in the configuration space where the best ROC integral was achieved by this particle and one term that pulls the particle in direction of the global maximum of ROC values:

$$\mathbf{v}_{i,\mathrm{new}} = \gamma \mathbf{v}_{i,\mathrm{old}} + U(0,a)(\mathbf{x}_{i,\mathrm{best}} - \mathbf{x}_i) + U(0,b)(\mathbf{x}_{\mathrm{global\ best}} - \mathbf{x}_i). \tag{9.10}$$

Hereby $U(x, y)$ stands for a number generated randomly, following a uniform distribution between x and y. The parameters γ, a, and b and the initial positions and velocities are chosen so that a reasonable part of the configuration space is scanned and but the particles converge to a common point. The BDT configuration is then updated to

$$\mathbf{x}_{i,\mathrm{new}} = \mathbf{x}_{i,\mathrm{old}} + \mathbf{v}_{i,\mathrm{new}}. \tag{9.11}$$

The optimization is stopped after N iterations and the configuration yielding the best ROC integral is chosen.

The parameters found by the optimization are listed in Table 9.2. Roughly speaking, the product of $\mathrm{N}_{\mathrm{trees}}$ and shrinkage increases the learning rate and the risk of overtraining and is thus larger in categories with more training events. The values chosen for the bagging fraction and $\mathrm{N}_{\mathrm{cuts}}$ do not follow clear trends and might be less important for the BDT performance.

9.5.3 Selection of BDT input variables

A large number of BDT input candidates can be constructed for the use in the BDT training. As discussed above they fall in the following classes: object kinematics, an-

Table 9.2: BDT configurations used in the eight analysis categories, optimized by the Particle Swarm Optimization algorithm.

Category	N_{trees}	shrinkage	bagging fraction	N_{cuts}	depth
≥6 jets + 2 tags	642	0.05	0.37	20	2
4 jets + 3 tags	1210	0.03	0.40	63	2
5 jets + 3 tags	845	0.02	0.50	43	2
≥6 jets + 3 tags	518	0.04	0.37	26	2
4 jets + 4 tags	668	0.02	0.42	20	2
5 jets + ≥4 tags	374	0.04	0.23	30	2
≥6 jets + ≥4 tags	1233	0.01	0.48	21	2
boosted	737	0.03	0.63	22	2

gular correlations, $\mathrm{t\bar{t}H}$ reconstruction, event shape, b-tagging, and fat-jet substructure. Additionally, the MEM discriminant is a possible input for the BDT training.

Because the $\mathrm{t\bar{t}}$ simulation has a few known shortcomings, only observables not sensitive to these problematic features are used in the final training. Observables related to b-tagging are well modeled and the fit allows to incorporate their shape uncertainties so that this class is the least problematic. Even though they are less commonly used at CMS, fat jets and substructure variables were also observed to be decently modeled by the default simulation. Some distributions of substructure observables and the corresponding MC predictions are compared in Appendix A.1. Problematic are observables related to light (anti-k_{T} 0.4) jets, especially at large η and low p_{T}. Thus variables like the largest η-difference between two light jets or transverse momenta of soft jets are not used in the BDT training. Furthermore, the total invariant mass of the hardest objects is not well-described in MC, likely due to a mismodeling of the $\mathrm{t\bar{t}}$ system invariant mass.

The distribution of all input variable candidates left is compared between data and MC and the difference between prediction and observation has to be within the uncertainties. Observables showing trends in data that are not described in the simulation are excluded. A few examples of these comparisons for well-modeled variables are shown in Fig. 9.4 and all BDT inputs are displayed in Appendix A.2.

Additionally, the observables should not be modeled significantly different in an alternative $\mathrm{t\bar{t}}$ simulation based on MadGraph5_aMC@NLO at LO. This simulation is complementary in the sense that, on matrix element level, it can describe more additional partons (via MLM merging [71]) but it lacks virtual higher order corrections. Finally correlations between all pairs of input variables are also inspected to ensure no drastic differences are hidden by only looking at one-dimensional distributions.

Only observables passing all tests are included in the final BDT optimization. The definition of all input variables is summarized in Table 9.3. The inputs used in the different categories as well as the BDT configuration is shown in Table 9.4. The MEM discriminant is not used for the training in categories with events with four b-tags. This decision will be discussed in the next section.

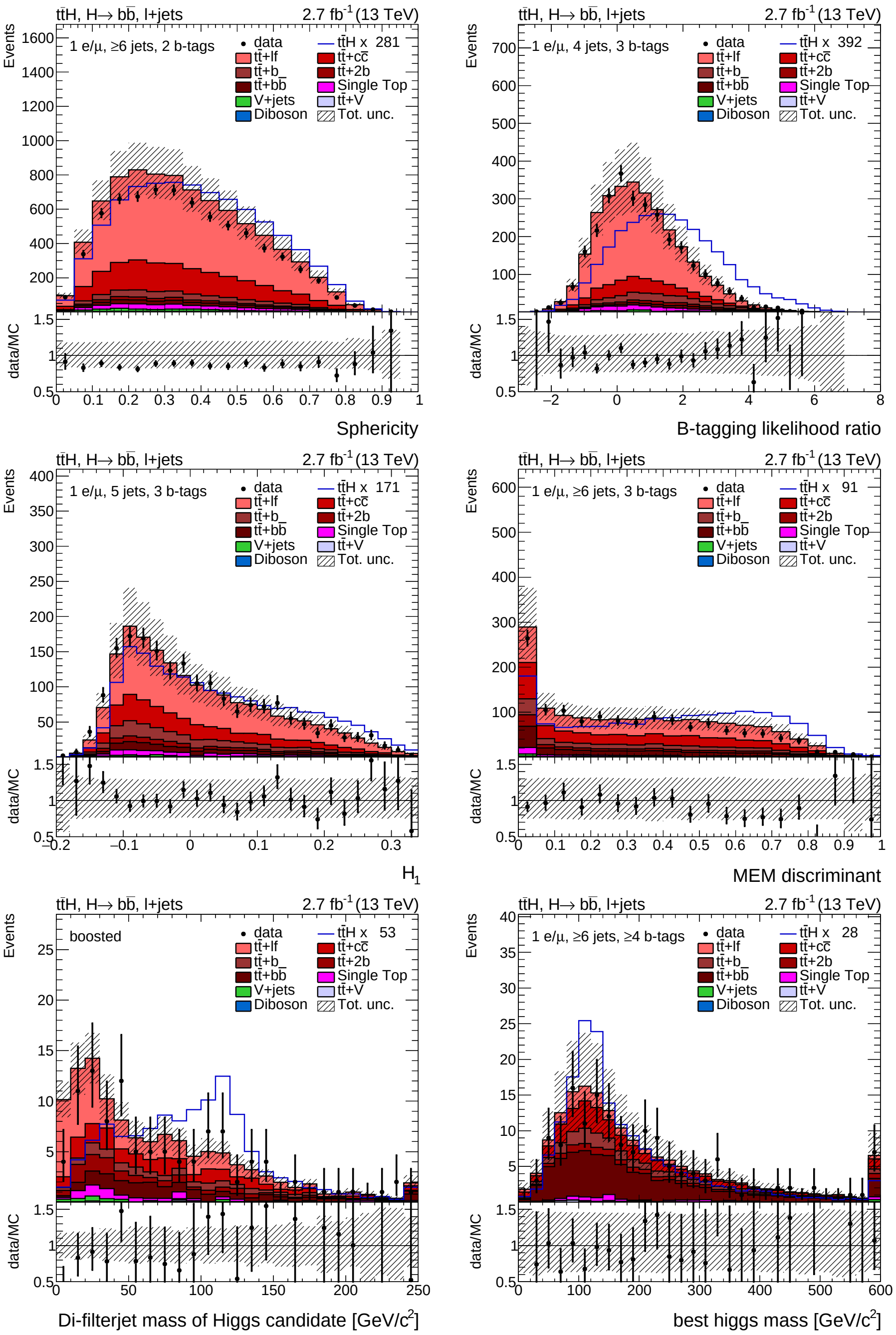


Figure 9.4: Examples of BDT inputs in six of the eight analysis categories (see also [229]). The category is indicated in the top-left corner of each figure. The expected distribution of the background is displayed as stacked histograms, the observed events as black markers. The $t\bar{t}$H signal is superimposed as a blue line and scaled to the background integral to allow for a better shape-comparison. Systematic uncertainties are indicated with hashed bands, statistical uncertainties as horizontal black lines. Invariant masses are displayed in units of GeV/c^2.

Table 9.3: Descriptions of the BDT input variables that are selected by the BDT optimization in at least one category.

Observable	**Description**
Object kinematics	
jet 1, 2, 3, 4 p_T	Transverse momenta of jets, ordered by p_T.
HT	Scalar sum of transverse momenta of all jets.
$\sum p_T$(jets,leptons,MET)	Sum of the p_T of all jets, leptons, and MET.
$(\Sigma$ jet $p_T)/(\Sigma$ jet E)	Ratio of the sum of p_T and the sum of energies of all jets.
Angular correlations	
ΔR(jet 1, jet 2)	Distance in the ϕ-η plane between the two hardest jets.
avg ΔR(tag,tag)	Average distance in the ϕ-η plane between b-tagged jets.
avg $\Delta\eta$(jet,jet)	Average $\Delta\eta$ between jets.
max $\Delta\eta$(tag, avg jet η)	max η-difference between any b-tagged jet and avg η of all jets.
max $\Delta\eta$(tag, avg tag η)	max η-difference between any b-tagged jet and avg η of all b-tagged jets.
min ΔR(tag,tag)	ϕ-η distance between the two closest b-tagged jets.
min ΔR(lepton,jet)	ϕ-η distance between the lepton and the closest jet.
$\mathrm{t\bar{t}H}$ reconstruction	
Closest tagged dijet mass	Invariant mass of the two b-tagged jets that are closest in ΔR.
Tagged dijet mass closest to 125	Invariant mass of the b-tagged jet pair closest to 125 GeV/c^2.
Best Higgs mass	The invariant mass of the two b-jets not used in the χ^2 $\mathrm{t\bar{t}}$ reconstruction.
$\sqrt{\Delta\eta(\mathrm{t_{lep}},\mathrm{b\bar{b}}) \times \Delta\eta(\mathrm{t_{had}},\mathrm{b\bar{b}})}$	Square root of $\Delta\eta$ (leptonic top, $\mathrm{b\bar{b}}$) times $\Delta\eta$ (hadronic top, $\mathrm{b\bar{b}}$), where the $\mathrm{b\bar{b}}$ pair and the leptonic and hadronic top vector are reconstructed with the χ^2 reconstruction.
b-tagging	
second- to fifth-highest CSV	Second- to fifth-highest highest CSVv2IVF discriminant value of all jets.
avg CSV of all jets	Average b-tag discriminant value for all jets.
avg CSV of tagged jets	Average b-tag discriminant value for b-tagged jets.
dev from avg CSV (tags)	Variance of the CSVv2IVF discriminant for b-tagged jets.
b-tag likelihood ratio	Transformed b-tag likelihood radio discriminant.
Event shape	
Sphericity	$\frac{3}{2}(\lambda_2+\lambda_3)$, where λ_i is the ith-largest eigenvalue of the momentum tensor.
Aplanarity	$\frac{3}{2}\lambda_3$, where λ_3 is the 3rd-largest eigenvalue of the momentum tensor.
H_1, H_3	Fox-Wolfram moments.
Fat jet substructure	
τ_2/τ_1 Higgs cand.	2-subjettiness to 1-subjettiness ratio of Higgs-boson candidate fat jet.
m(Higgs, di-filterjet)	Invariant mass of boosted Higgs-boson candidate reconstructed from the two b-tagged filtered subjets.
$\Delta\eta$(top,Higgs)	Pseudo-rapidity difference between boosted top candidate and boosted Higgs-boson candidate.
MEM discriminant (using subjets)	MEM discriminant using the subjets from the reconstructed boosted top candidate instead of anti-k_T jets.
MEM discriminant	
MEM discriminant	MEM discriminant.

Table 9.4: Variables used for BDT training in the analysis categories. Variables are ordered by type and described in Table 9.3. The abbreviations 6j2t to boosted stand for the eight analysis categories, from $\geq 6\,\text{jets}, 2\,\text{b-tags}$ to the boosted category.

Observable	**6j2t**	**4j3t**	**5j3t**	**6j3t**	**4j4t**	**5j4t**	**6j4t**	**boosted**
Object kinematics								
jet 1 p_T					✓			
jet 2 p_T		✓						
jet 3 p_T			✓			✓		
jet 4 p_T		✓					✓	
HT				✓	✓			
$\sum p_T$(jets,leptons,MET)		✓				✓	✓	
(Σ jet p_T)/(Σ jet E)				✓			✓	
Angular correlations								
avg ΔR(jet 1,jet 2)	✓							
avg ΔR(tag,tag)	✓		✓			✓		✓
avg $\Delta\eta$(jet,jet)	✓		✓			✓		
max $\Delta\eta$ (tag, avg jet η)	✓		✓	✓				
max $\Delta\eta$ (tag, avg tag η)	✓						✓	
min ΔR(tag,tag)	✓							✓
min ΔR(lepton,jet)			✓					
$\mathrm{t\bar{t}H}$ reconstruction								
closest tagged dijet mass					✓		✓	
tagged dijet mass closest to 125						✓	✓	
best Higgs mass							✓	
$\sqrt{\Delta\eta(t^{lep}, bb) \times \Delta\eta(t^{had}, bb)}$				✓				
b-tagging								
2nd highest CSV							✓	
3rd highest CSV	✓				✓			✓
fourth-highest CSV	✓		✓	✓				✓
fifth-highest CSV						✓	✓	
avg CSV of all jets		✓	✓					✓
avg CSV of tagged jets		✓	✓	✓				
dev from avg CSV (tags)			✓					
b-tag likelihood ratio	✓	✓	✓	✓	✓	✓	✓	✓
Event shape								
Sphericity	✓						✓	
Aplanarity								✓
H_1		✓	✓	✓		✓		
H_3					✓		✓	
Fat jet substructure								
τ_2/τ_1 Higgs cand.								✓
m(Higgs, di-filterjet)								✓
$\Delta\eta$(top,Higgs)								✓
MEM discriminant (using subjets)								✓
MEM discriminant								
MEM discriminant		✓	✓	✓				

9.6 Combination of BDT and MEM discriminants

9.6.1 Comparisons of discriminants

Both, the BDT discriminant and the discriminant created with the Matrix Element Method have advantages and disadvantages.

The MEM discriminant is constructed by calculating the event likelihood according to a physically motivated model. This allows for a clearer interpretation the output and a more intuitive dependence of the output on the input kinematics than is the case for the BDT. Furthermore, it can make use of arbitrary correlations between final state products, as long as they are modeled in the leading order matrix element. The BDT cannot make optimal use of complicated correlations if it is given primitive input variables. If, e.g. two of the four b-jets form a certain invariant mass or have a distinct angular distance, this feature has to be reconstructed by the analyst and can only then be used in the training. Catching all relevant features of a final state is not always possible.

A disadvantage of the MEM discriminant is that it requires the time-intensive computation of integrals for every event. For the algorithm and CPUs used in this analysis, the calculation of the MEM for one $\mathrm{t\bar{t}H}$ event takes 2 minutes. In contrast to that, the evaluation of the BDT requires less than 1 ms. As a result, BDTs can be evaluated for all measured and simulated events available while the MEM can only be employed on a subset. On the other hand, the BDT performance increases with the number of available training events while the MEM requires no training and can be used for small numbers of events. Furthermore, the Matrix Element Method as implemented for this analysis benefits from a large number of b-tagged jets which increases the likelihood that all b quarks have been reconstructed as b-jets.

The main differences between the discriminants as they are implemented in this analysis is that the matrix element is only sensitive to the differences of the jet, $\not{E}_\mathrm{T}$, and lepton energy and momentum between $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}b\bar{b}}$. Additional variables like the b-tags of the jets are not used to discriminate between signal and background. The MEM is also not designed to separate $\mathrm{t\bar{t}H}$ from non-$\mathrm{t\bar{t}b\bar{b}}$ backgrounds like $\mathrm{t\bar{t}} + \mathrm{lf}$. Conversely, the BDT can use all kinds of observables and is trained to distinguish $\mathrm{t\bar{t}H}$ from all backgrounds. The fact that the MEM discriminant performs well against $\mathrm{t\bar{t}b\bar{b}}$ and the BDT discriminant better against $\mathrm{t\bar{t}} + \mathrm{lf}$ can also be seen from the top row in Fig. 9.5.

From the above discussion it should already be clear that the MEM discriminant should preferably be calculated in categories that have a large number of b-tags, a low number of events, and that are dominated by $\mathrm{t\bar{t}b\bar{b}}$ background. Thus the the MEM discriminant is not calculated in the 6-jet, 2-tag category, which would need a lot of processing time and yields little benefit, as the background is mostly $\mathrm{t\bar{t}} + \mathrm{lf}$.

It is also interesting to study the correlation between the two discriminants. Two-dimensional distributions of the two discriminants for signal and background processes are displayed in Fig. 9.5 for the $\geq 6\,\mathrm{jets}, \geq 4\,\mathrm{b\text{-}tags}$ category. The two discriminants are correlated, there are more events expected on the diagonal than in the off-diagonal edges. The correlation coefficient is 0.36 for the signal, 0.42 for the $\mathrm{t\bar{t}b\bar{b}}$ background, and 0.40 for the remaining $\mathrm{t\bar{t}}$ background. The correlation is thus not complete and a combination of the discriminants might be beneficial. It can also be observed that the two different backgrounds have dissimilar distributions in the discriminant plane. The non-$\mathrm{t\bar{t}}$-$\mathrm{b\bar{b}}$ background can be found at low BDT values, almost independent of the MEM discriminant value. Contrary, the $\mathrm{t\bar{t}b\bar{b}}$ background is dominant in the region in which the MEM discriminant has low values.

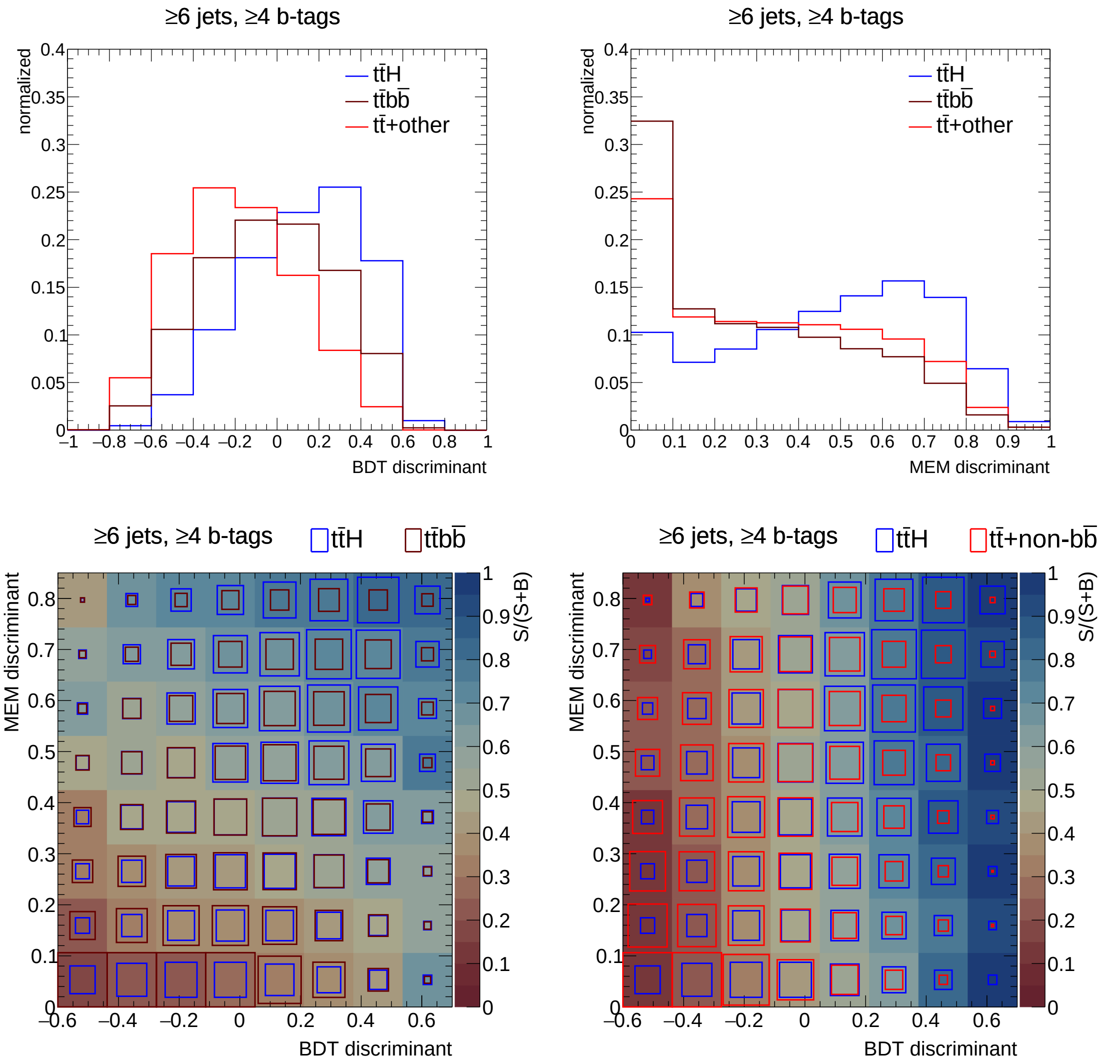


Figure 9.5: Top row: one-dimensional distributions of simulated $t\bar{t}H(b\bar{b})$ (blue), $t\bar{t}b\bar{b}$ (dark red), and $t\bar{t}$-non-$b\bar{b}$ (red) events in the ≥ 6 jets, ≥ 4 b-tags category. On the left-hand side for the BDT discriminant (without the MEM as input) and on the right-hand side for the MEM discriminant. Bottom row: Distribution of the same events in the space spanned by the MEM discriminant and the BDT discriminant. In the left plot, $t\bar{t}H$ is compared with the $t\bar{t}b\bar{b}$ background and in the right plot with the part of the $t\bar{t}$ background not classified as $t\bar{t}b\bar{b}$. The relative contribution of the processes is indicated by the area of the rectangles. All processes are normalized to the same total integral. The background color is more blue in regions in which the fraction of $t\bar{t}H$ events is larger and more red in regions where background dominates.

9.6.2 Combination of discriminants

For the combination of the discriminants two different methods are explored.

The first is to use the MEM discriminant as an additional input to the BDT. As long as there are enough MC events to train the BDT sufficiently well, this should yield the best separation between signal and background. However, in some regions of the phase space the number of MC events available for the training is rather small. Furthermore, even if the discriminant is constructed so that the separation between signal and background is optimal, it does not necessarily lead to the most sensitive analysis. This is due to the fact that systematic uncertainties affect the various background processes differently. The most important example are the $\mathrm{t\bar{t}+hf}$ backgrounds, which have a larger rate uncertainty than the $\mathrm{t\bar{t}+lf}$ background so that it is more important to separate these backgrounds from the signal.

The second method explored is to apply cuts in the two-dimensional space spanned by the MEM and BDT outputs[2]. To reduce the correlation of the two classifiers, for this approach the BDT is trained without the MEM as an input. Here, the disadvantage is that the BDT is unable to use correlations between the MEM discriminant and the remaining input variables to separate signal from background. A benefit is that one is not limited by the training statistics for the combination of the two powerful discriminants. Furthermore, well chosen cuts can improve the discrimination against the $\mathrm{t\bar{t}+hf}$ backgrounds, which are difficult to predict, by sacrificing some of the $\mathrm{t\bar{t}+lf}$ discrimination. As discussed, this can be beneficial for the sensitivity of the analysis. Optimizing the cuts in the two-dimensional space can again introduce an over-training effect because of statistical fluctuations of the MC simulation. For this reason the two-dimensional cuts are kept very simple. A first cut is applied at the median of the $\mathrm{t\bar{t}H(b\bar{b})}$ distribution for the BDT discriminant. The MEM discriminant is then split into equidistant bins in the so constructed subcategories. The splitting is illustrated in Fig. 9.6, again for the ≥ 6 jets, ≥ 4 b-tags category. Additional simple cut combinations, e.g. splitting with the MEM discriminant and using the BDT discriminants within the categories leads to worse expected limits.

It is demonstrated in Fig. 9.7 that this combination indeed incorporates the advantages of both discriminants. On the left-hand side the $\mathrm{t\bar{t}b\bar{b}}$ rejection and on the right-hand side the $\mathrm{t\bar{t}}$ non-$\mathrm{b\bar{b}}$ rejection is shown for a given signal efficiency for MEM, BDT, and their combination with the 2D-approach. The MEM (BDT) discriminant performs better against $\mathrm{t\bar{t}b\bar{b}}$ ($\mathrm{t\bar{t}}$ non-$\mathrm{b\bar{b}}$). The combined discriminant performs against $\mathrm{t\bar{t}b\bar{b}}$ like the MEM and against $\mathrm{t\bar{t}}$ non-$\mathrm{b\bar{b}}$ like the BDT. For the combined discriminant the ROC curve is constructed by combining the bins of the two histograms in the two subcategory to one histogram, with bins ordered by increasing signal-background ratio.

Using the two combination methods outlined above, three combination schemes are defined for the categories containing events with three or more b-tags:

1. Using the MEM as BDT input in the 3 b-tag and ≥ 4 b-tag region ("MEM in BDT").
2. Using the 2D method in the 3 b-tag and ≥ 4 b-tag region ("MEM vs. BDT").
3. Using the first method in the 3 b-tag region and the second in the ≥ 4 b-tag region.

The third scheme is motivated by two arguments. The lack of training events might make using the MEM discriminant as BDT input in the 4-tag region less effective. Additionally, the dominant $\mathrm{t\bar{t}b\bar{b}}$ background can benefit from the increased separation due to applying

---

[2] This method has been first proposed and tested by Gregor Kasieczka (ETH Zürich).

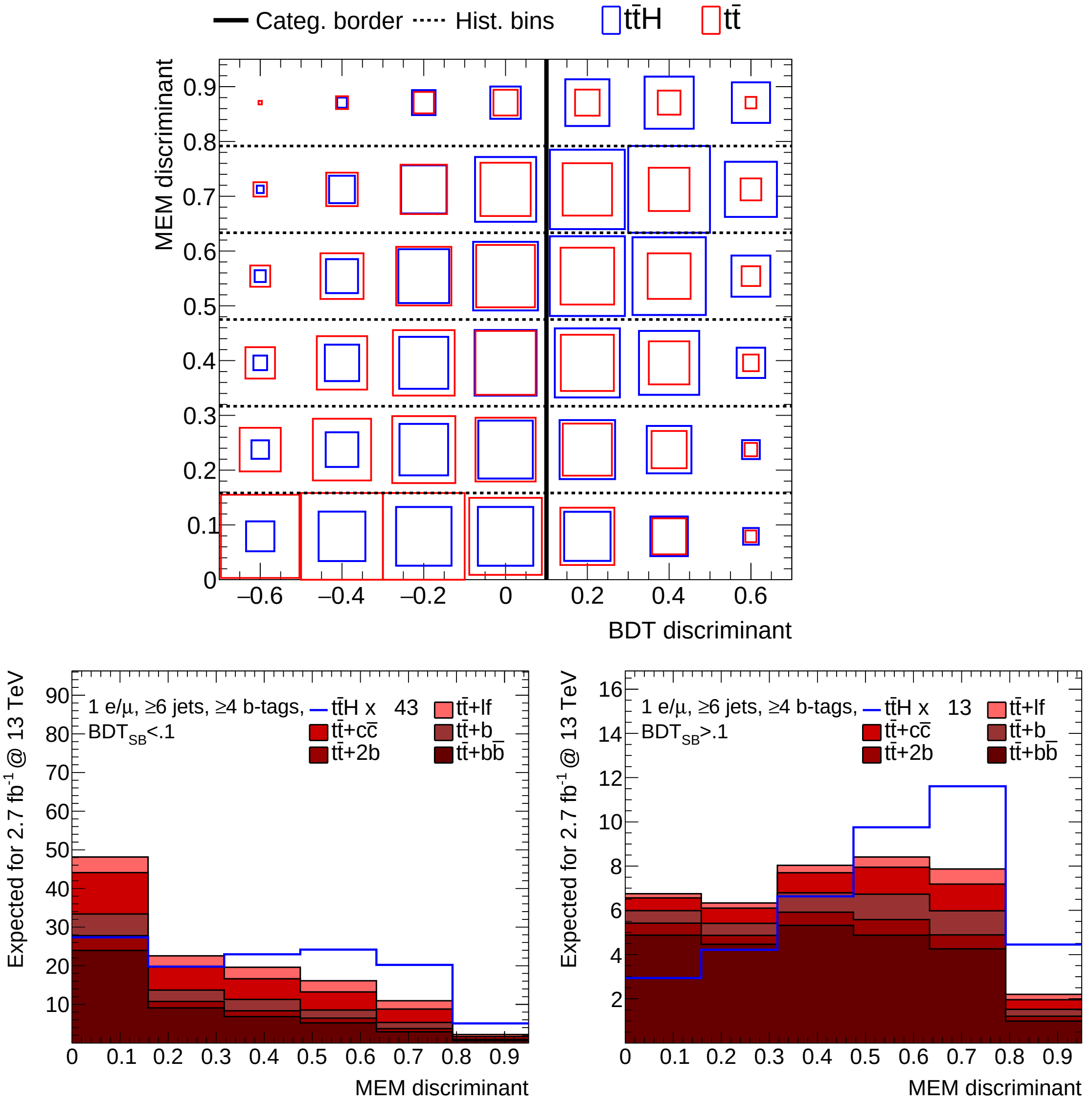


Figure 9.6: Combination of BDT and MEM discriminant in the ≥ 6 jets, ≥ 4 b-tags category with the 2D-approach. Top row: The fat black line indicates the value at which the category is split into two subcategories, dashed lines correspond to histogram-bin borders. The size of blue (red) squares corresponds to the relative contributions of $t\bar{t}H$ signal (inclusive $t\bar{t}$ background). Bottom row: distribution of signal (blue line) and $t\bar{t}$ backgrounds (stacked histograms) in the two subcategories. The background yield expected for this analysis is displayed. To allow for a shape comparison, $t\bar{t}H$ is scaled to the background yield.

.

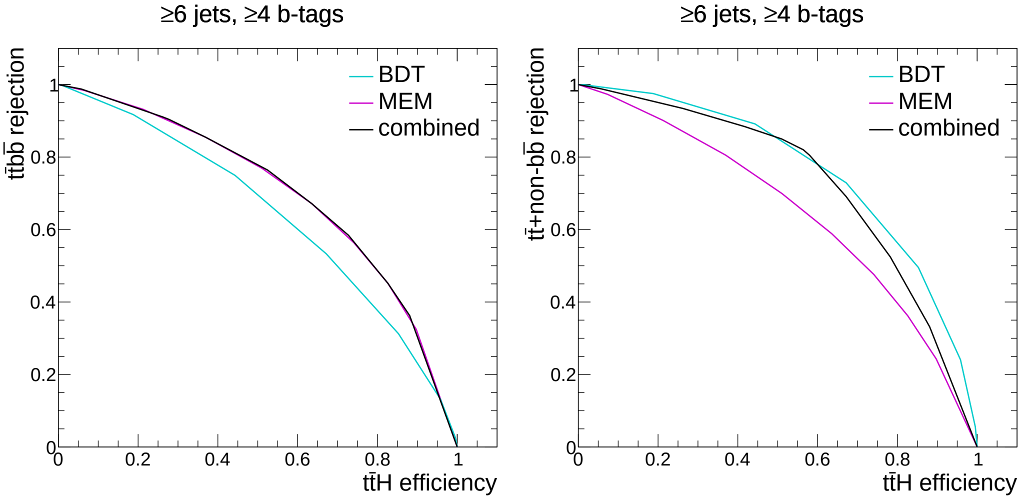


Figure 9.7: The $t\bar{t}b\bar{b}$ (left) and $t\bar{t}$ non-$b\bar{b}$ rejection (right) for different $t\bar{t}H$ efficiencies and discriminants is shown. The cyan (magenta) line represents the efficiency of the BDT (MEM) discriminant. The black line shows the performance of the combined discriminant, constructed in the 2D-approach.

.

the MEM discriminant on top of the BDT. Ultimately, all combination schemes lead to expected limits that differ only by a few percent. These differences are not statistically significant and the third combination scheme is chosen, being a compromise of the two methods.

The new subcategories created by splitting the four b-tag categories increase the number of categories from eight to eleven.

10 Direct constraints on the top-Higgs coupling strength

In this chapter, the shape of the discriminant distributions constructed in the previous chapter is analyzed to measure the strength of the $\mathrm{t\bar{t}H}$ signal. The statistical model used for this analysis is described in Section 10.1 and the results are documented in Section 10.2.

10.1 Statistical model and systematic uncertainties

10.1.1 Final discriminant distributions

The expected distribution for the signal and the $\mathrm{t\bar{t}}$ backgrounds in the eleven analysis categories is shown in Fig. 10.1 and Fig. 10.2.

For the statistical analysis, the binning of the distributions is determined so that at least two background events are predicted in the most signal-like histogram bins. In the background-dominated categories the requirement is increased to approximately ten events. This reflects the disconfidence in the ability of the model to describe the background in extreme regions of the phase space, especially in regions that are very signal-sensitive. If the bins in these regions have a very low background content, a small systematic uncertainty that is not accounted for in the model could shift just a small fraction of events from the next-to-outermost bin in the outermost bin and fake a signal.

10.1.2 Statistical model

For the measurement of the $\mathrm{t\bar{t}H}$ signal strength, the observed distribution of data for the discriminants in the eleven analysis categories is analyzed with the statistical model described in more detail in Section 4.2. A binned maximum likelihood fit is performed simultaneously in all categories. The expected distributions of the signal and background processes, the so called templates, are determined from the MC simulation introduced Chapter 7, including all corrections described in Chapter 8. The parameter of interest of the fit is the $\mathrm{t\bar{t}H}$ signal strength

$$\mu_{\mathrm{t\bar{t}H}} = \sigma_{\mathrm{t\bar{t}H}} / \sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}}, \tag{10.1}$$

which allows to scale the $\mathrm{t\bar{t}H}$ templates, corresponding to a change of the rate of $\mathrm{t\bar{t}H}$ production. Throughout this chapter, SM Higgs boson branching ratios are assumed. However, since the event selection is dominated by $\mathrm{t\bar{t}H(b\bar{b})}$ events, the signal strength can also be interpreted as $\mu_{\mathrm{t\bar{t}H(bb)}} = \mu_{\mathrm{t\bar{t}H}} \times \mu_{\mathrm{Hbb}}$.

10.1.3 Systematic uncertainties

Systematic uncertainties are included as nuisance parameter in the fit. All nuisance parameters are constrained using normal distributions. There are two classes of systematic

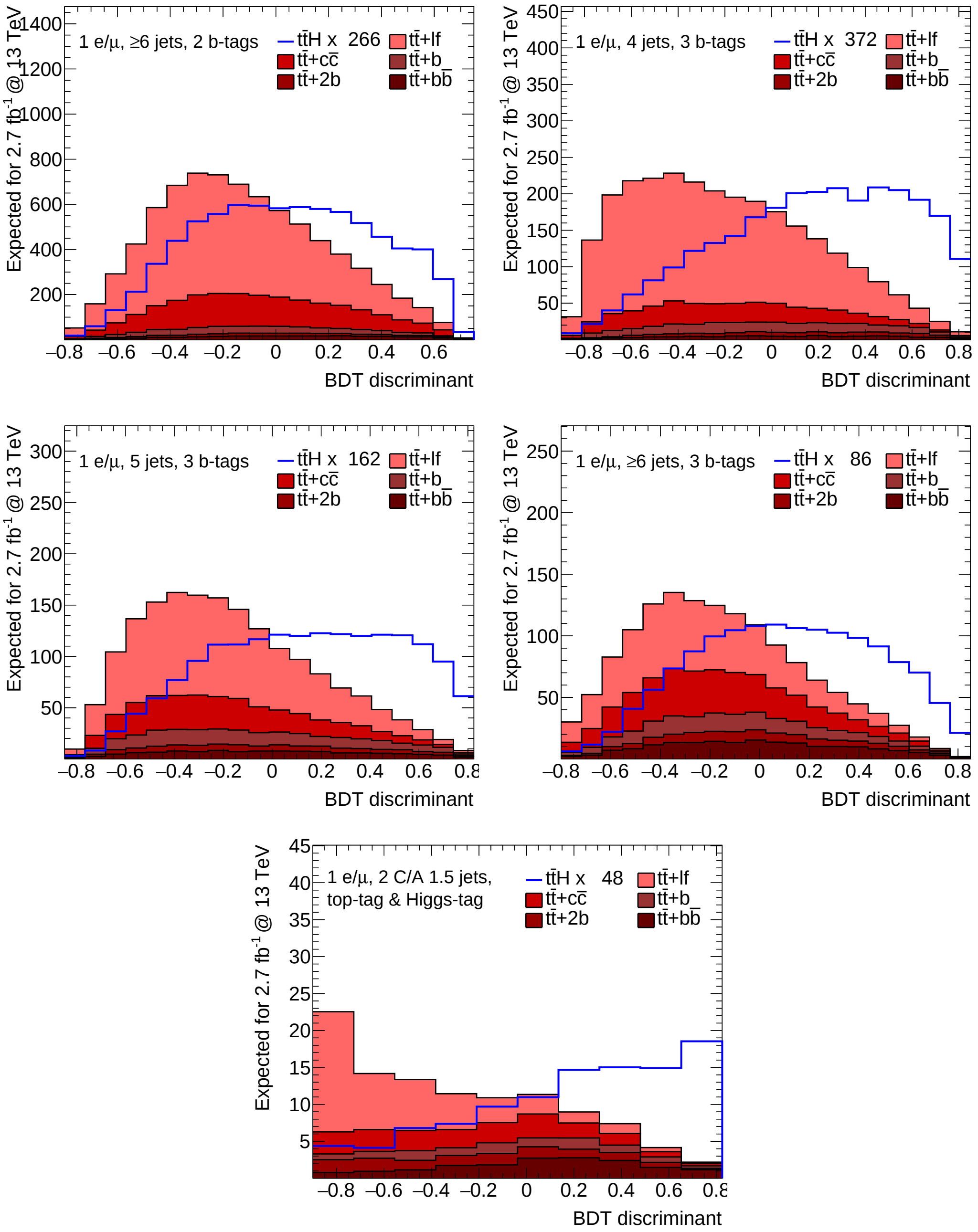


Figure 10.1: Distribution of simulated $t\bar{t}H$ events (blue line) compared to the expected $t\bar{t}$ background (stacked red histograms) for the five categories in that a BDT discriminant containing the MEM discriminant as input is used for signal-background separation. The $t\bar{t}$ backgrounds are scaled to the number of events expected in the 2015 data, the $t\bar{t}H$ signal is scaled to the background to allow for shape comparisons. The scaling factor is indicated in the legend.

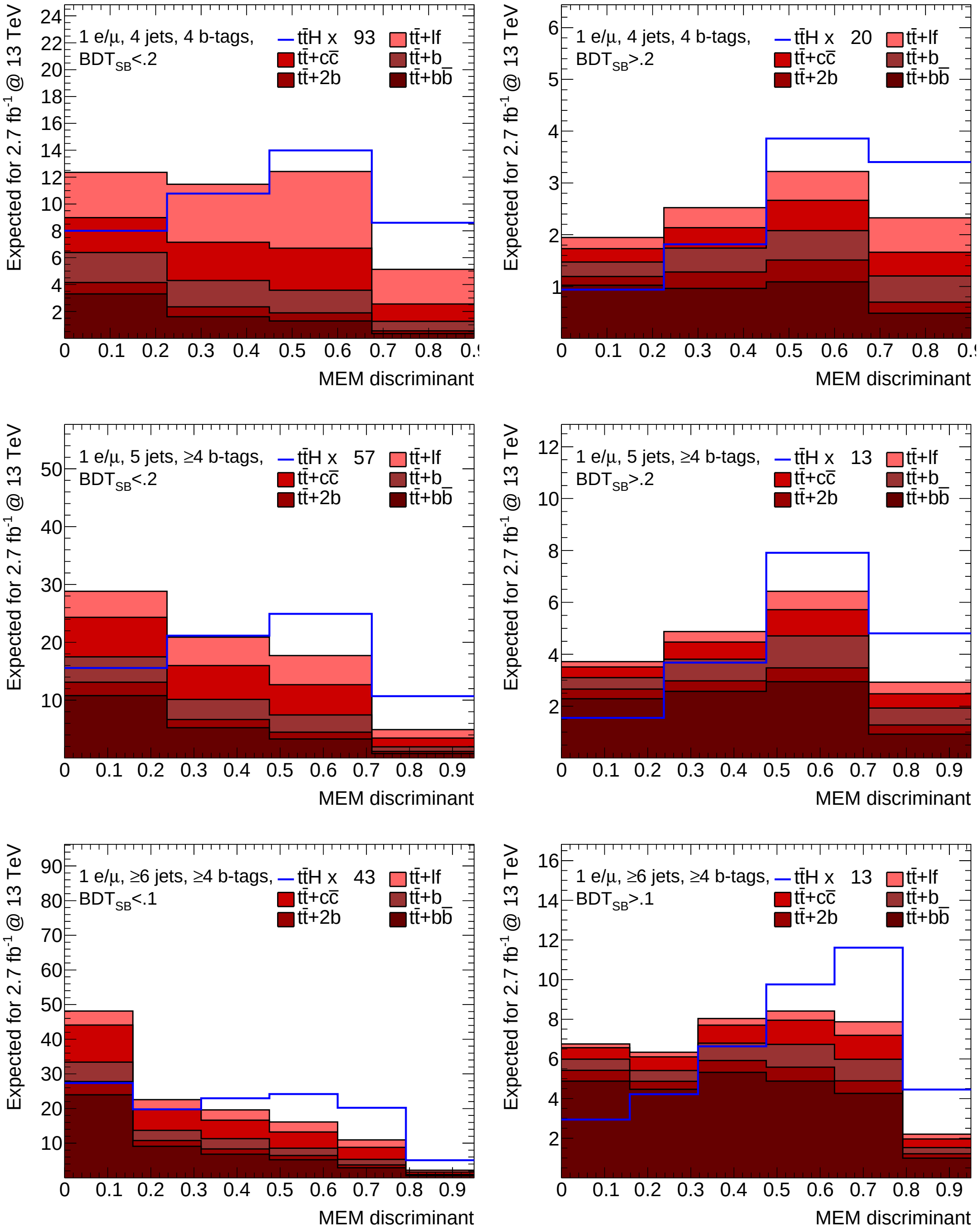


Figure 10.2: Distribution of simulated $t\bar{t}H$ events (blue line) compared to the expected $t\bar{t}$ background (stacked red histograms) for the categories that are split by the BDT and use the MEM as final discriminant. On the left-hand side the background-dominated and on the right-hand side the signal-enriched categories are shown. The $t\bar{t}$ backgrounds are scaled to the number of events expected in the 2015 data, the $t\bar{t}H$ signal is scaled to the background to allow for shape comparisons. The scaling factor is indicated in the legend.

uncertainties, those that change the rate of processes and those that also modify the shape of distributions.

The scaling of templates due to rate systematic uncertainties is implemented using Equation (4.14) so that the rate is effectively log-normal constrained. The most important rate uncertainties are the cross section uncertainties, which have been introduced in Section 7.4 and which are summarized in Table 7.2. As a reminder, they are of the approximate size of 10% for $\mathrm{t\bar{t}H}$, 5% for $\mathrm{t\bar{t}}$, and an additional 50% for the $\mathrm{t\bar{t}+hf}$ processes. When applicable, the rate uncertainties are implemented taking their given uncertainties in both directions into account.

Shape uncertainties are implemented by simulating distributions of the background and signal under modified assumptions. For every shape uncertainty, two additional templates are generated, corresponding to upward and downward variations of the systematic by one standard deviation. The fit interpolates between the two templates to allow for arbitrary shifts. In the case of the JES and parton shower scale systematic the additional templates are generated using modified MC samples. For JES, the jet four-vectors are modified within the uncertainties determined by the JES measurements described in 8.2. This can also lead to migration between categories. For the parton shower scale systematic samples simulated with a different parton shower configuration are used (see Section 7.4). The remaining shape uncertainties are implemented by modifying the weights of MC events by multiplicative corrections. Examples of such uncertainties are b-tagging uncertainties, the matrix element generator scale, and lepton efficiency uncertainties. The effect of JER uncertainties is approximated by reweighting the jet multiplicity distributions of all samples to avoid having to recalculate the MEM discriminant for jets with only slightly shifted momenta. A shifted JER hardly affects the MEM discriminant distribution but requires CPU-intensive computations.

A summary of all systematic uncertainties and their effect, as an example, on the signal and the $\mathrm{t\bar{t}}$ background in the $\geq 6\,\mathrm{jets}, 3\,\mathrm{b\text{-}tags}$ category is given in Table 10.1. While the table only shows the expected effect on the rate, most of the uncertainties can also change the shape of the discriminant distributions.

Finally, another kind of shape uncertainty is caused by the finite number of MC events used to generate the templates. It introduces a statistical uncertainty that is approximated by a Gaussian of width

$$\frac{\Delta N}{N} = \frac{\sqrt{\sum_i^{N_{\text{MC-events}}} w_i^2}}{\sum_i^{N_{\text{MC-events}}} w_i} \tag{10.2}$$

for the prediction of every processes in every histogram bin. In principle, the uncertainty in every bin could be modeled as a systematic uncertainty, as proposed in [236]. However, the presented analysis is performed in 118 histogram bins and with 19 processes (counting different Higgs decay channels as separate processes), which would lead to a huge number of additional nuisance parameters that would make any fit very slow and potentially unstable. Instead of considering the uncertainty in all bins, a pruning procedure is performed. An MC statistics uncertainty of a certain sample in a certain bin is ignored if:

- The statistical uncertainty on the sum of background MC is at least five times smaller than the square root of the observed data in the bin.
- The expected signal-to-background ratio in the bin is smaller than 0.01.
- A sample is expected to contribute less than 0.01 events in the bin.

Table 10.1: Summary of the effect of systematic uncertainties on the event yield of $\mathrm{t\bar{t}H(b\bar{b})}$ signal and the inclusive $\mathrm{t\bar{t}}$ background in the $\geq$ 6 jets, 3 b-tags category. The effect is evaluated by varying them within their prefit confidence intervals.

Systematic	**Effect on yield in $\geq$ 6 jets, 3 b-tags**			
Process	$\mathrm{t\bar{t}}$		$\mathrm{t\bar{t}H(b\bar{b})}$	
Direction of shift	up	down	up	down
Rate (experimental)				
Luminosity	+2.6%	−2.5%	+2.6%	−2.5%
Shape (experimental)				
Jet energy scale	+11.3%	−10.1%	+7.7%	−7.0%
Jet energy resolution	−0.1%	+0.1%	−0.1%	+0.1%
Pileup	−0.1%	+0.0%	+0.1%	−0.2%
Electron efficiency	+1.6%	−1.6%	+1.6%	−1.6%
Muon efficiency	+1.2%	−1.2%	+1.2%	−1.2%
b-tag b-SF lf-contamination	+7.1%	−5.1%	+5.5%	−4.2%
b-tag b-SF stats (linear)	−6.4%	+6.2%	−5.3%	+4.9%
b-tag b-SF stats (quadratic)	+4.2%	−4.4%	+3.3%	−3.6%
b-tag l-SF hf-contamination	−3.5%	+8.4%	+0.2%	+0.6%
b-tag l-SF stats (linear)	−3.2%	+6.5%	−0.6%	+1.1%
b-tag l-SF stats (quadratic)	+0.5%	+1.2%	−0.8%	+1.1%
b-tag charm uncertainty (linear)	−12.6%	+16.9%	−0.6%	−0.7%
b-tag charm uncertainty (quadratic)	+1.4%	−1.4%	+0.0%	−0.0%
Rate (theory)				
Renorm. & fact. scale $\mathrm{t\bar{t}H}$	–	–	+5.8%	−9.2%
Renorm. & fact. scale $\mathrm{t\bar{t}}$	+2.0%	−4.0%	–	–
Renorm. & fact. scale $\mathrm{t\bar{t}c\bar{c}}$	+13.8%	−9.2%	–	–
Renorm. & fact. scale $\mathrm{t\bar{t}b}$	+5.9%	−4.0%	–	–
Renorm. & fact. scale $\mathrm{t\bar{t}2b}$	+3.4%	−2.3%	–	–
Renorm. & fact. scale $\mathrm{t\bar{t}b\bar{b}}$	+6.4%	−4.3%	–	–
Renorm. & fact. scale single top	–	–	–	–
Renorm. & fact. scale V+jets	–	–	–	–
Renorm. & fact. scale diboson	–	–	–	–
PDF $\mathrm{t\bar{t}H}$	–	–	+3.6%	−3.5%
PDF gg	+3.0%	−2.9%	–	–
PDF qq	–	–	–	–
PDF qg	–	–	–	–
Shape (theory)				
ME generator scale $\mathrm{t\bar{t}+lf}$	−1.9%	+2.8%	–	–
ME generator scale $\mathrm{t\bar{t}b}$	−0.6%	+0.9%	–	–
ME generator scale $\mathrm{t\bar{t}2b}$	−0.5%	+0.8%	–	–
ME generator scale $\mathrm{t\bar{t}b\bar{b}}$	−0.9%	+1.3%	–	–
ME generator scale $\mathrm{t\bar{t}c\bar{c}}$	−1.6%	+2.4%	–	–
Parton shower scale $\mathrm{t\bar{t}+lf}$	+4.4%	−8.7%	–	–
Parton shower scale $\mathrm{t\bar{t}b}$	−1.3%	+0.8%	–	–
Parton shower scale $\mathrm{t\bar{t}2b}$	−1.0%	+0.4%	–	–
Parton shower scale $\mathrm{t\bar{t}b\bar{b}}$	−2.0%	+1.3%	–	–
Parton shower scale $\mathrm{t\bar{t}c\bar{c}}$	−4.3%	+2.3%	–	–

- The removal of the uncertainty reduces the bin uncertainty by less than 5%.

This approximation has a negligible effect on the expected limit. Technically, the MC statistics uncertainties are implemented as shape uncertainties with templates in which only one bin is shifted up or down.

In total the statistical model contains 129 nuisance parameters. 92 of the parameters describe MC fluctuations in single bins. 13 parameters describe experimental shape uncertainties, 8 of these alone due to b-tagging. 10 parameters describe the theoretical shape uncertainties of the five $\mathrm{t\bar{t}}$ subprocesses. Finally, there are 14 rate uncertainties, one due to luminosity and 13 due to scale and PDF uncertainties.

10.2 Results

10.2.1 Maximum likelihood fit

The data analyzed in this thesis, compared to the signal and background expectation before the fit, is shown in Fig. 10.3 and 10.4. There are small normalization differences visible between data and expectation that are covered by the uncertainties. Shape differences are small and no signal is visible by eye.

The negative log-likelihood introduced in Section 4.2 is minimized with respect to all nuisance parameters $\boldsymbol{\theta}$ using MINUIT [237]. That the minimum found corresponds to a global minimum is verified by scanning one parameter at a time while minimizing the likelihood with respect to the remaining parameters. The scan of the parameter of interest $\mu_{\mathrm{t\bar{t}H}}$ is shown in figure 10.7. The best fit value for this parameter is

$$\mu_{\mathrm{t\bar{t}H}}^{\mathrm{obs.}} = -0.4 \pm 2.1.$$

The uncertainty is determined from the profile-likelihood ratio q_μ, assuming a χ^2 distribution of the test statistic (see Section 4.1.2 for a justification of this method).

10.2.2 Postfit distributions

Postfit distribution, for which the background templates are scaled to the event yields and morphed to the shapes corresponding to the best-fit values of nuisance parameters, are shown in Fig. 10.5 and 10.6.

The error bands are different from the previous plots. For the prefit plots the effect of all of the uncertainties is added in quadrature, which assumes no correlations and a Gaussian distribution. After the fit nuisance parameter estimates are correlated. For example, the data is almost equally compatible with a larger $\mathrm{t\bar{t}2b}$ rate – but only if the $\mathrm{t\bar{t}b}$ rate would be reduced and the total event yield thus be almost unchanged. Furthermore, parameters are constrained. For example, the $\mathrm{t\bar{t}b\bar{b}}$ rate has a large 50% theory uncertainty before the fit, but the fit finds that such large deviations from the expected value are not compatible with the events observed. The approximate size of correlations and constraints on the fit parameters is encoded in the covariance matrix. It can be used to construct a probability distribution for the nuisance parameters assuming a multivariate normal distribution. Sampling from this distribution and generating "toy" shapes by calculating the corresponding change of the fit model allows determining a probability distribution for the rate in every histogram bin. The size of the error bands on the postfit histograms corresponds to 68% confidence intervals for the rate in the bins.

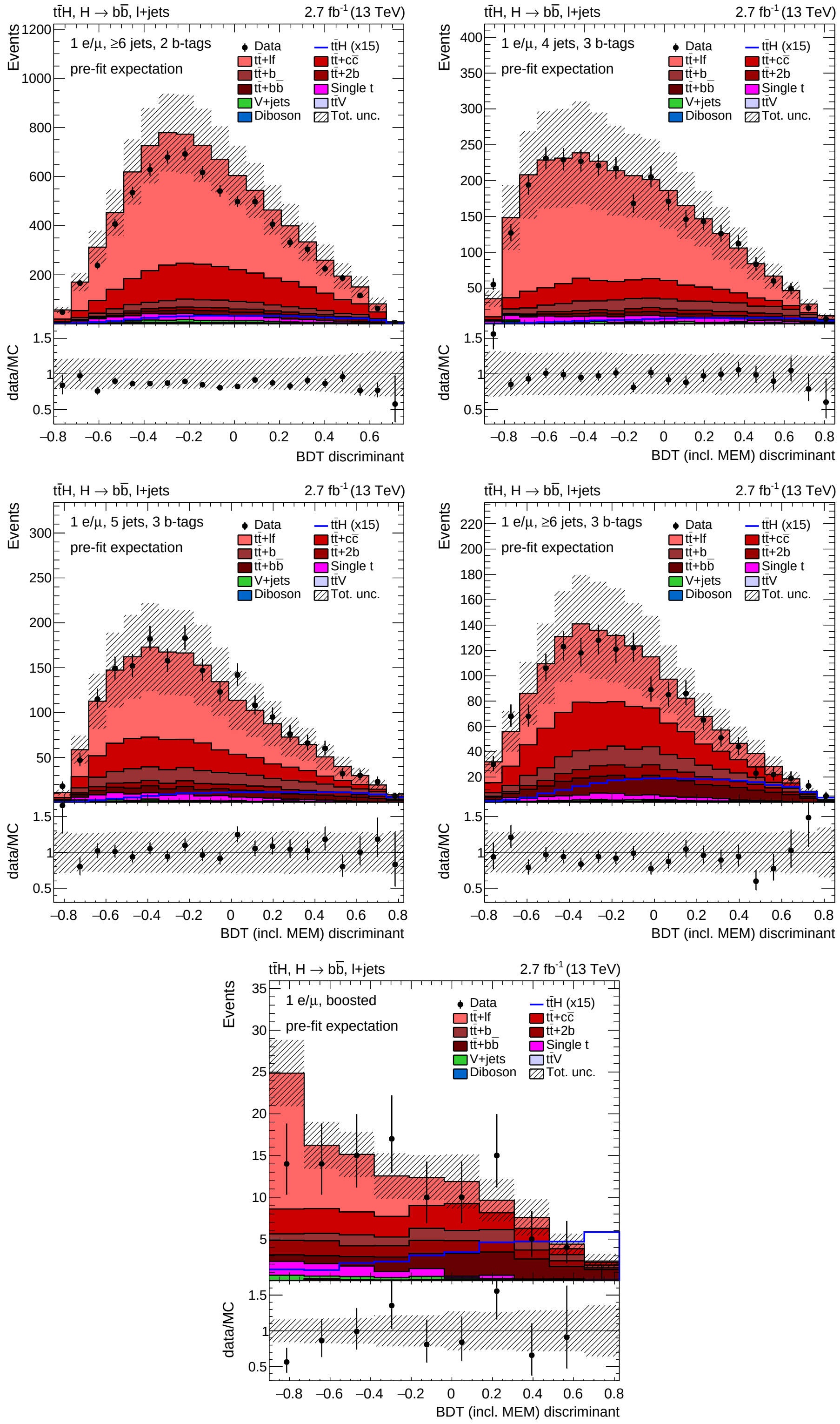


Figure 10.3: Distribution of the data collected in 2015 compared to the expectation from simulation, before the fit, in the five categories in which BDT discriminants are used for signal-background separation. The expected tt̄H signal is drawn as a blue line and scaled up by a factor of 15 w.r.t. the SM expectation for better visibility. The expected effect of systematic uncertainties are added in quadrature and shown as hashed error bands.

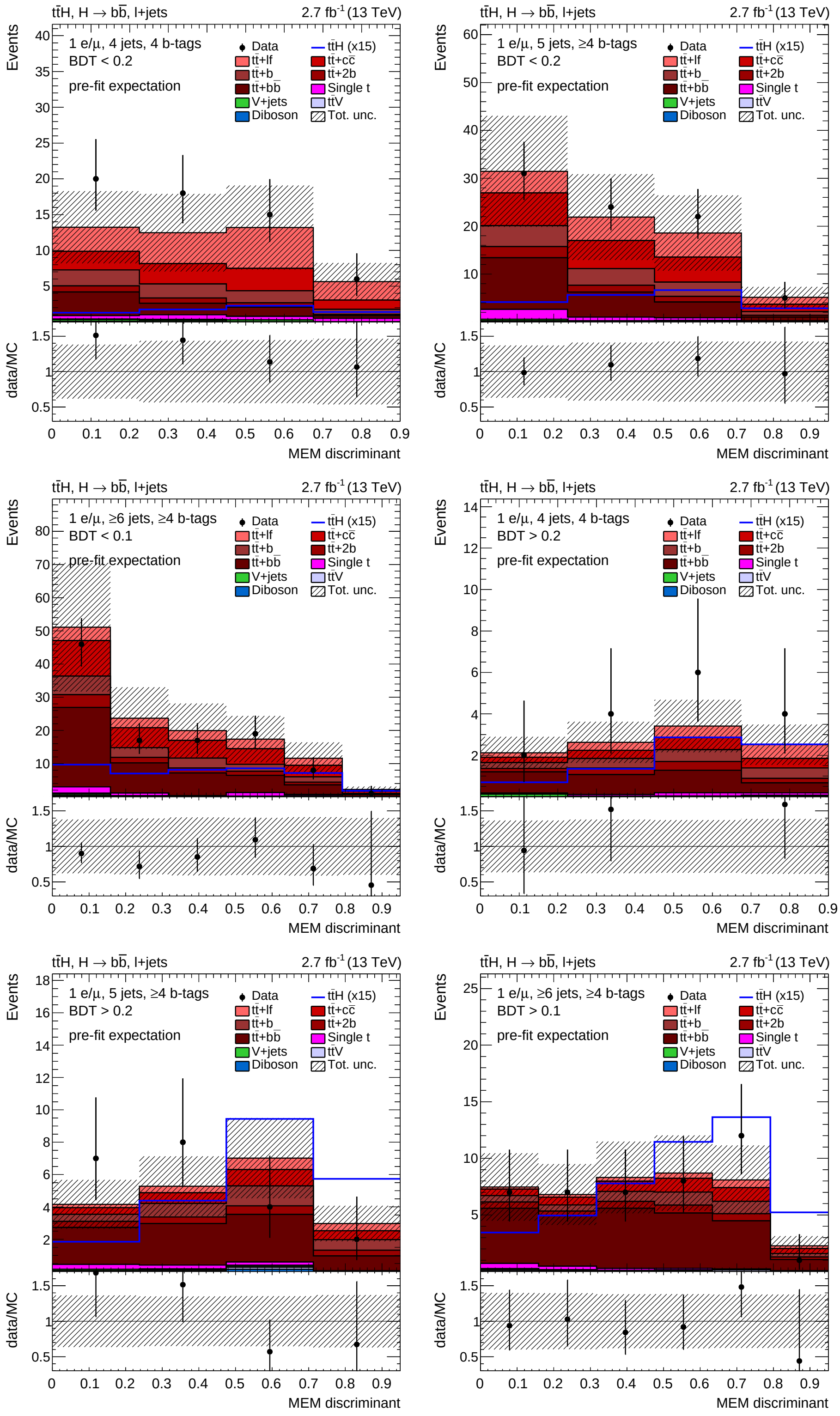


Figure 10.4: Distribution of the data collected in 2015 compared to the expectation from simulation, before the fit, in the categories split by the BDT discriminant, in whose subcategories the MEM discriminant is employed. The expected tt̄H signal is drawn as a blue line and scaled up by a factor of 15 w.r.t. the SM expectation for better visibility. The expected effect of systematic uncertainties are added in quadrature and shown as hashed error bands.

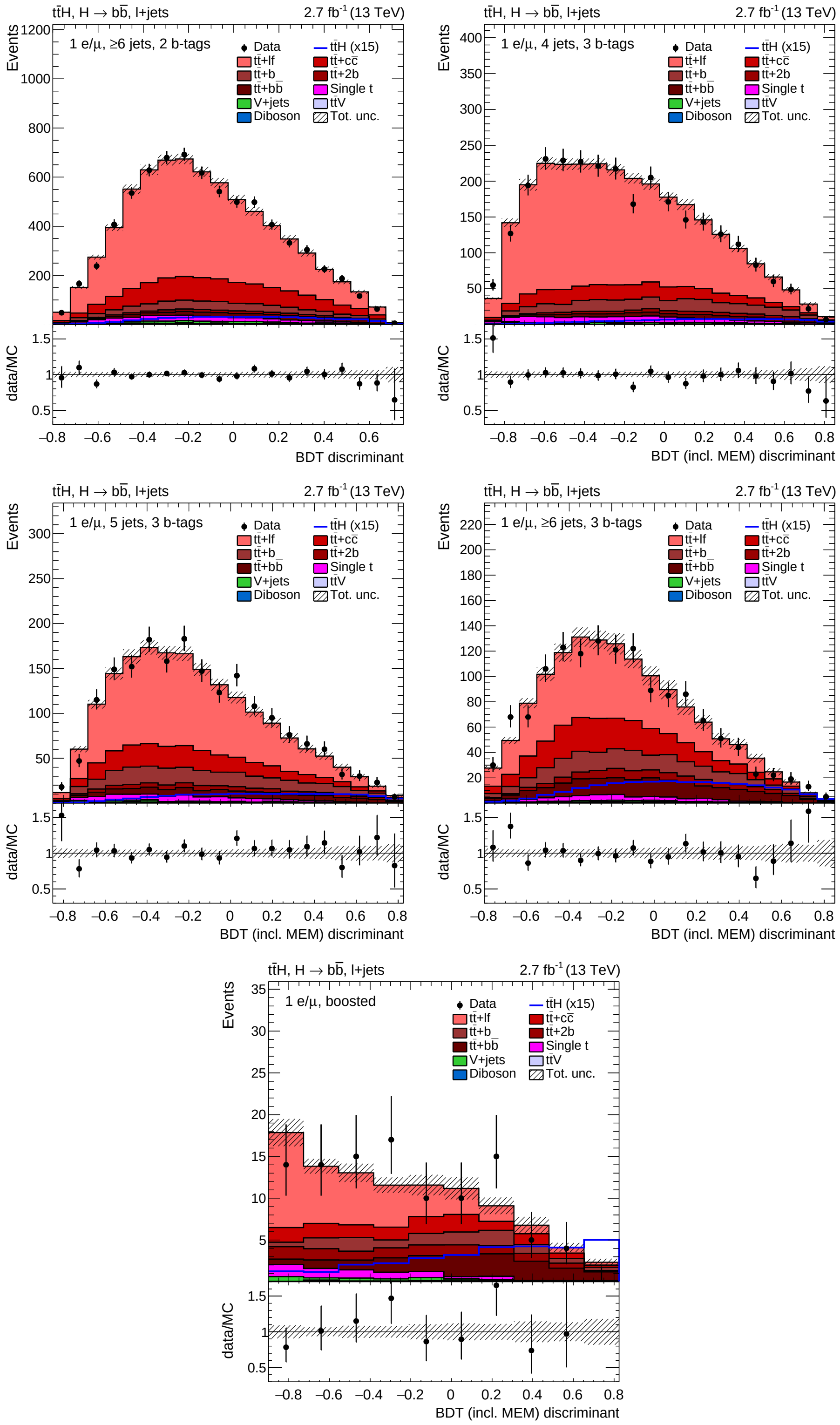


Figure 10.5: Distribution of the data collected in 2015 compared to the model after the fit, in the five categories in which BDT discriminants are used for signal-background separation. The expected tt̄H signal is drawn as a blue line and scaled up by a factor of 15 w.r.t. the SM expectation for better visibility. The postfit uncertainties are evaluated by covariance matrix sampling and displayed as hashed error bands.

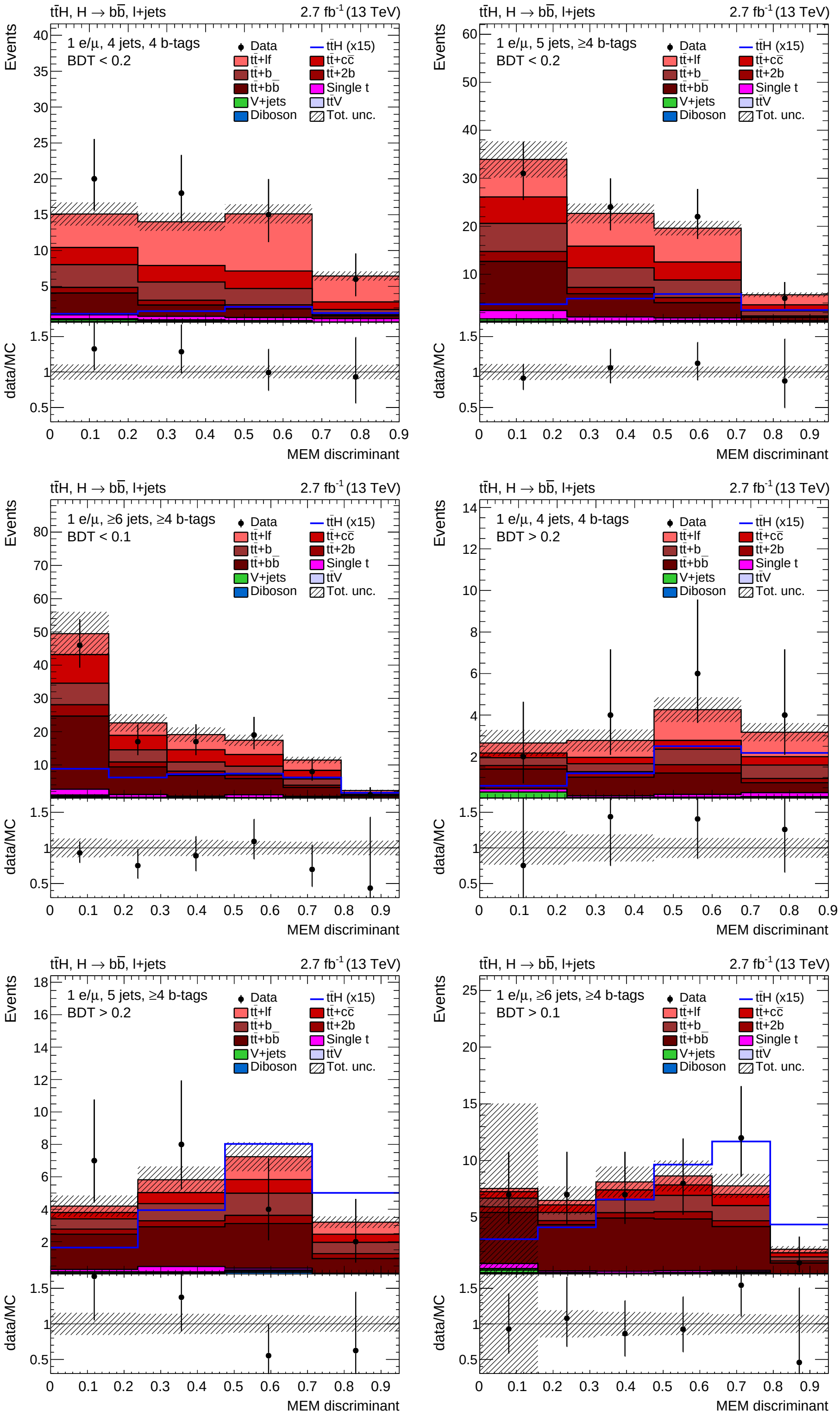


Figure 10.6: Distribution of the data collected in 2015 compared to the model after the fit, in the categories split by the BDT discriminant, in whose subcategories the MEM discriminant is employed. The expected tt̄H signal is drawn as a blue line and scaled up by a factor of 15 w.r.t. the SM expectation for better visibility. The postfit uncertainties are evaluated by covariance matrix sampling and displayed as hashed error bands.

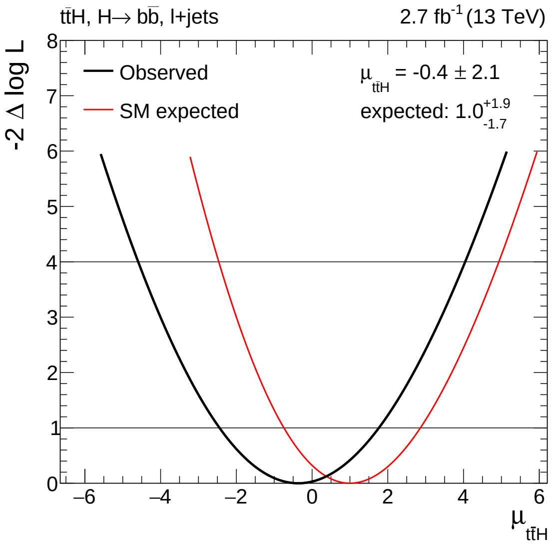


Figure 10.7: Likelihood profile obtained from varying the signal strength parameter $\mu_{\mathrm{t\bar{t}H}}$ while minimizing the likelihood with respect to the nuisance parameters. The result from fitting real data is drawn in black, the expected result in red. Horizontal lines correspond to the values associated with the 1σ and 2σ confidence intervals of $\mu_{\mathrm{t\bar{t}H}}$.

10.2.3 Constraints and correlations of nuisance parameters

A more complete picture of nuisance parameter pulls and constraints can be obtained by studying likelihood profiles. Nuisance parameter pulls are hereby defined as the deviation of the best-fit value from the initial value normalized to the uncertainty assumed before the fit (denoted $(\hat{\theta}-\theta_0)/\Delta\theta_0$) and the constraints correspond to 68% confidence intervals in the asymptotic χ^2 approximation, also normalized to the prefit range.

In Fig. 10.8 the pulls and constraints on the most important (as defined below) nuisance parameters are displayed on the left-hand side. For nuisance parameters correlated to the signal, tighter constraints improve the analysis sensitivity because the parameter has less freedom left to explain signal-like deviations from the background model as the effect of the associated systematic uncertainty. As long as the constraints are not too tight, this reflects the improved knowledge about these uncertainties after the fit. However, tight constraints can also be introduced if the statistical model is too simplistic to describe the effect of uncertainties. For example, without reasonable b-tagging uncertainties the exact rates of the $\mathrm{t\bar{t}}+\mathrm{hf}$ processes could be determined from the b-tag distributions. This would then allow for a too precise prediction of the background in signal dominated regions.

The parameter constrained to values furthest from the initial parameter is the one corresponding to the $\mathrm{t\bar{t}}+\mathrm{lf}$ parton shower scale uncertainty. As discussed before and shown in Fig. 7.8, an upward shift of this parameter corresponds to a reduced jet multiplicity compared to the nominal simulation. This is compatible with the observation of an overprediction of high-jet-multiplicity events by the nominal model. The JES uncertainty is likely constrained to low values because it mimics the effect of the parton shower scale systematic, by reducing the jet energies the number of events passing jet p_{T} cuts is also reduced.

Large pulls for nuisance parameters describing the $\mathrm{t\bar{t}b}$ background are also observed. They might be explained by the underestimation of the event yield in the 4 jets, 4 b-tags category by the nominal model. This is the category with the largest $\mathrm{t\bar{t}b}$ fraction and

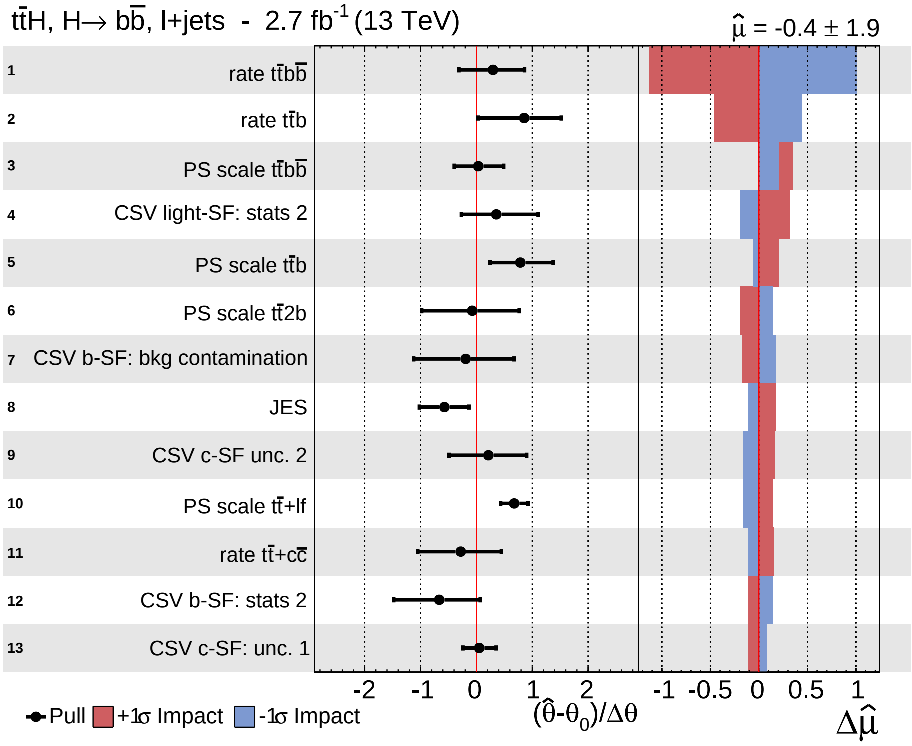


Figure 10.8: Pulls and impacts of all nuisance parameters with an impact of $\Delta\mu_{t\bar{t}H} > 0.1$. Pulls are defined as the difference between the initial and best-fit values of the parameters, divided by their prefit uncertainty. They are shown as black points. 68% CL constraints on the nuisance parameters are drawn as horizontal black lines. The impact of nuisance parameters is evaluated by fixing them to the borders of their 68% CL intervals, repeating the fit with the remaining parameters and calculating the change of the signal strength measured this way. The effect of fixing the parameter to the higher (lower) edge of the interval is shown as red (blue) bars.

with the postfit nuisance parameter configuration the $t\bar{t}b$ event yield in this category is increased. The uncertainties associated with c-flavor b-tagging correction factors are also strongly constrained. The reason for this are the large initial uncertainties on the correction factors which, in lack of a charm-dominated control region, are taken to be twice as large as the b-flavor uncertainties. The remaining b-tagging uncertainties are hardly constrained in the fit. This is important because of the simplified model used to describe the statistical b-tagging uncertainties. The correction factors are derived in multiple bins of p_T and η and in every bin statistical uncertainties are derived. However, in the fit only two nuisance parameters are used per jet flavor to describe the shape uncertainty of the b-tag discriminant, implicitly treating the statistical uncertainties of different bins correlated. In principle this would allow to constrain, e.g. the statistical uncertainty of b-tagging scale factors of high-p_T jets with a sample of low-p_T jets, even though they are derived in different control samples.

In addition to the nuisance parameter constraints, their impact on the measured signal strength is interesting to observe (see the right part of Fig. 10.8). It is determined by fixing the parameters to the edges of their postfit confidence intervals while minimizing the likelihood with respect to the remaining parameters. If a parameter is correlated to the signal strength, this will change signal strength parameter compared to the nominal fit,

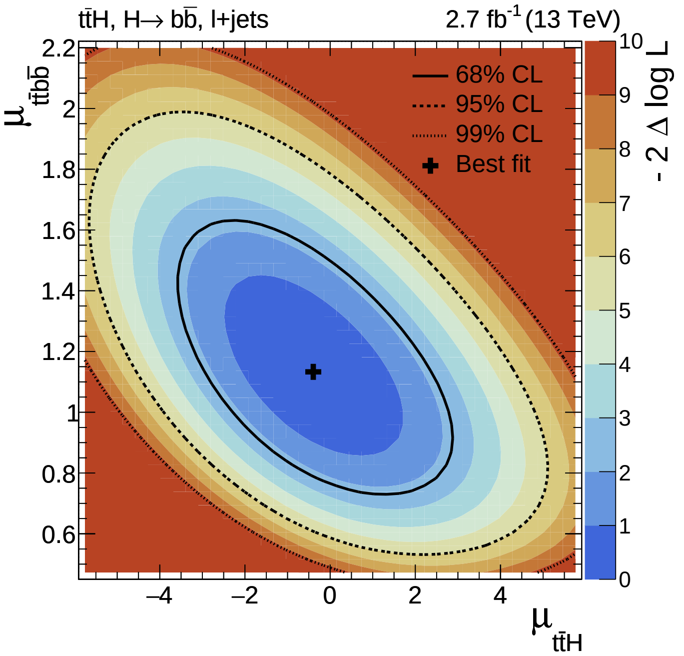


Figure 10.9: Profile of the test statistic $q(\mu_{\mathrm{t\bar{t}H}}, \mu_{\mathrm{t\bar{t}b\bar{b}}})$ used to measure both, the rate of $\mathrm{t\bar{t}H(b\bar{b})}$ and the $\mathrm{t\bar{t}b\bar{b}}$ rate. Solid (dashed/dotted) lines indicate the border of the 68% (95% / 99%) confidence level interval, assuming a two-dimensional χ^2 distribution for the test statistic. The best-fit value is indicated as a black cross.

allowing for the identification of important uncertainties. The most important uncertainty is the $\mathrm{t\bar{t}b\bar{b}}$ rate uncertainty, which affects the measured signal by more than $\Delta\mu_{\mathrm{t\bar{t}H}} = \pm 1$. The impact of the $\mathrm{t\bar{t}b}$ rate, which introduces an uncertainty of 0.4 on $\mu_{\mathrm{t\bar{t}H}}$, is also sizable. Moreover, mostly uncertainties related to the $\mathrm{t\bar{t}} + \mathrm{hf}$ modeling and and b-tagging have a large impact. The thirteen parameters that affect the signal strength by more than 0.1 are shown in Fig. 10.8.

The large impact of the $\mathrm{t\bar{t}b\bar{b}}$ rate uncertainty raises the question to which extend the measurement of the $\mathrm{t\bar{t}H(b\bar{b})}$ is determined by the assumed $\mathrm{t\bar{t}b\bar{b}}$ rate and its prior. This can be tested by treating the rates of both processes as parameters of interest in the fit, dropping the $\mathrm{t\bar{t}b\bar{b}}$ constraint. A scan of the profile likelihood ratio $q(\mu_{\mathrm{t\bar{t}H}}, \mu_{\mathrm{t\bar{t}b\bar{b}}})$ is shown in Fig. 10.9. The confidence intervals corresponds to a two-dimensional χ^2 distribution so that the one-dimensional projections of the intervals are larger than the interval from a one-dimensional fit. The uncertainty of the measurement thus becomes even larger, values of $\mu_{\mathrm{t\bar{t}H}} = 2$ and $\mu_{\mathrm{t\bar{t}b\bar{b}}} = 0.8$ as well as $\mu_{\mathrm{t\bar{t}H}} = -3$ and $\mu_{\mathrm{t\bar{t}b\bar{b}}} = 1.5$ are compatible with the observation at the 68% confidence level. The correlation of the two parameters is clearly visible. Nonetheless, to a certain extend, it is possible to measure both the $\mathrm{t\bar{t}b\bar{b}}$ and the $\mathrm{t\bar{t}H}$ rate at the same time.

10.2.4 Limits

Since no $\mathrm{t\bar{t}H}$ signal is observed, the measurement can be used to set upper limits on the $\mathrm{t\bar{t}H}$ production cross section. Limits are given at the 95% confidence level using the modified frequentist approach (CLs limits, see, e.g. [238]). The technical details of the limit setting procedure are outlined in Section 4.1.3. Due to the large number of nuisance parameters,

Table 10.2: Observed and expected 95% CLs upper limit on the $\mathrm{t\bar{t}H}$ signal strength, calculated with the asymptotic method. The limits are obtained from fits of the individual categories of this analysis, except for the last row, which is calculated from a simultaneous fit of all categories. For expected limits the 68% CL interval is given as superscripts and subscripts.

Category	Observed	Expected
4 jets, 3 b-tags	14.5	$18.6^{+8.2}_{-5.5}$
4 jets, $\geq$ 4 b-tags high BDT output	35.7	$25.6^{+13.4}_{-8.1}$
4 jets, $\geq$ 4 b-tags low BDT output	86.6	$84.2^{+41.3}_{-25.8}$
5 jets, 3 b-tags	16.0	$12.3^{+5.5}_{-3.6}$
5 jets, $\geq$ 4 b-tags high BDT output	7.5	$10.3^{+5.6}_{-3.4}$
5 jets, $\geq$ 4 b-tags low BDT output	35.2	$31.9^{+16.1}_{-9.9}$
$\geq$ 6 jets, 2 b-tags	25.4	$41.1^{+21.1}_{-13.1}$
$\geq$ 6 jets, 3 b-tags	9.6	$7.6^{+3.3}_{-2.2}$
$\geq$ 6 jets, $\geq$ 4 b-tags high BDT output	9.2	$8.3^{+4.4}_{-2.7}$
$\geq$ 6 jets, $\geq$ 4 b-tags low BDT output	15.4	$18.3^{+9.6}_{-5.8}$
$\geq$ 4 jets, $\geq$ 2 b-tags, boosted	7.5	$10.7^{+5.9}_{-3.5}$
lepton+jets combined	4.0	$4.1^{+1.8}_{-1.2}$

the calculation of limits using MC methods is computationally expensive. To avoid these computations, limits are calculated using the asymptotic method [123]. The limit on the $\mathrm{t\bar{t}H}$ production cross section is

$$\mu^{\mathrm{obs.}}_{\mathrm{t\bar{t}H}} < 4.0, \text{ with } \mu^{\mathrm{exp.}}_{\mathrm{t\bar{t}H}} < 4.1^{+1.8}_{-1.2} \text{ expected.}$$

Results derived from single categories are given in Fig. 12.1 and Table 12.1. The highest sensitivity is expected from the $\geq$ 6 jets, 3 b-tags and the $\geq$ 6 jets, $\geq$ 4 b-tags categories. One has to keep in mind that categories with only few observed events profit from the constraints on nuisance parameters in the background-dominated categories so that the combination of categories yields a result that is better than expected from a naive combination of its parts. The per-category results are very compatible with the combined result, they are within the 1σ uncertainties of the background-only expectation. As a test of the accuracy of the asymptotic method, the observed limit (but not the distribution of expected limits) is also calculated using the frequentist approach, but without the MC statistic uncertainties, yielding a limit of $\mu^{\mathrm{obs.,freq.}}_{\mathrm{t\bar{t}H}} < 4.1$, compared to 4.0 with the asymptotic method.

10.2.5 Discussion

The upper limit set by this analysis, $\mu^{\mathrm{obs.}}_{\mathrm{t\bar{t}H}} < 4.0$, is close to the background-only expectation of 4.1 but far from able to exclude SM $\mathrm{t\bar{t}H}$ production. The best-fit value of $\mu^{\mathrm{obs.}}_{\mathrm{t\bar{t}H}} = -0.4 \pm 2.1$ is both compatible with no $\mathrm{t\bar{t}H}$ and an SM-$\mathrm{t\bar{t}H}$ signal. The measurement is subject to large statistical and systematic uncertainties due to the low expected $\mathrm{t\bar{t}H}$ event yield compared to the large and difficult-to-model $\mathrm{t\bar{t}}$ background.

The main systematic uncertainty of this analysis is the $\mathrm{t\bar{t}b\bar{b}}$ rate uncertainty (see Fig. 10.8). A better understanding of the $\mathrm{t\bar{t}b\bar{b}}$ background is of paramount importance for future $\mathrm{t\bar{t}H(b\bar{b})}$ analysis as it would improve the confidence in the result and might

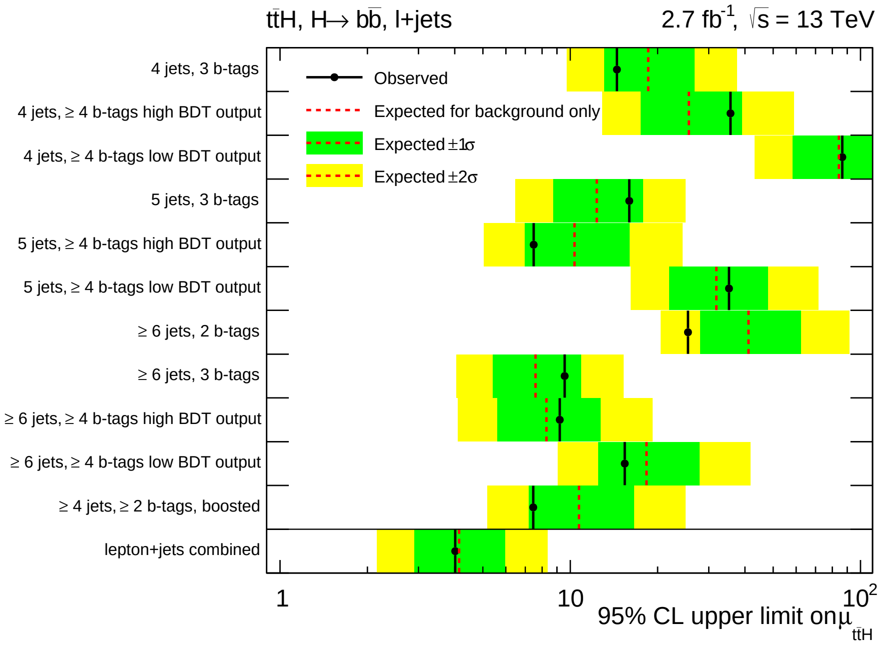


Figure 10.10: Observed and expected 95% CLs upper limit on the $\mathrm{t\bar{t}H}$ signal strength, calculated with the asymptotic method. The limits are obtained from the individual categories of this analysis, except for the last one with is calculated from a simultaneous fit of all categories. Observed limits are drawn as black lines and points, median expected limits as dashed red lines. The 1σ (2σ) confidence interval is shown as green (yellow) bars.

also allow to decrease the background uncertainty. In general, the large susceptibility to a mismodeling of the $\mathrm{t\bar{t}} + \mathrm{hf}$ background is one of the main problems of this analysis. As discussed in Section 7.2.3, improved predictions for $\mathrm{t\bar{t}b\bar{b}}$ are available and should be utilized in future analyses. Furthermore, more systematic and quantitative tests of the background model in control regions are required for more precise $\mathrm{t\bar{t}H(b\bar{b})}$ measurements. For the current iteration of the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis, the subpar modeling of the jet-multiplicity of the $\mathrm{t\bar{t}}$ background is problematic (see Fig. 9.1), which has to be mitigated by a better tuning of the MC simulation.

It might be possible to further improve the sensitivity to $\mathrm{t\bar{t}H(b\bar{b})}$ by introducing additional advanced analysis techniques. Examples are using differences in the $\mathrm{t\bar{t}}$ spin correlations between signal and background [239], or utilizing observables that are sensitive to the different QCD-color flow between a $\mathrm{b\bar{b}}$ pair of a Higgs-boson decay compared to a QCD-produced pair [240]. An improved (multivariate) $\mathrm{t\bar{t}H}$-system reconstruction, further use of boosted topologies, and a more complex categorization scheme might also be beneficial. However, while some gain can be obtained from these techniques, studies suggest that none of them are expected to revolutionize the $\mathrm{t\bar{t}H}$ search so that the focus of future analyses should lie on a better understanding of the background and systematic uncertainties.

The presented analysis is already of higher sensitivity than the Run 1 lepton+jets analyses. For comparison, the 8 TeV analyses of the lepton+jets channel, utilizing a BDT and a MEM discriminant were able to set an upper limit of $\mu^{\mathrm{obs.}}_{\mathrm{t\bar{t}H}} < 4.9$, with 4.7 expected [22] and $\mu^{\mathrm{obs.}}_{\mathrm{t\bar{t}H}} < 5.5$, with 4.2 expected [23], respectively. While the $\mathrm{t\bar{t}H}$ cross section is a factor

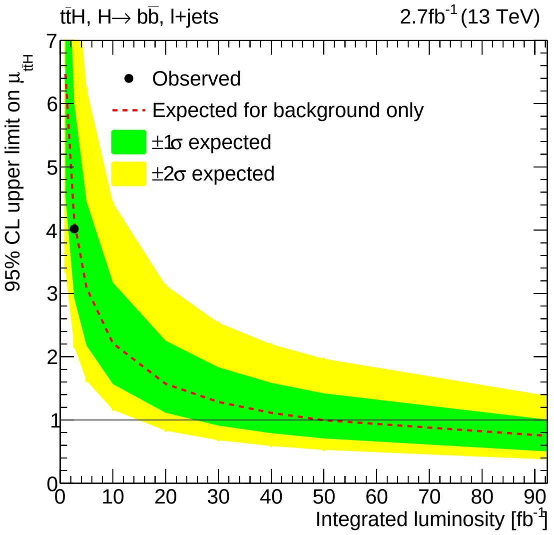


Figure 10.11: Limits obtained by extrapolating the statistical model of this analysis to what is expected at different integrated luminosities. For that, event yields are increased linearly with the luminosity while statistically dominated systematic uncertainties are decreased with the square root of it. The observed limit of this analysis is drawn as a black point, median expected limits as dashed red lines. The 1σ (2σ) confidence intervals of the expected limit are shown as green (yellow) bars.

of 3.9 smaller at 8 TeV compared to 13 TeV, a dataset of 19.5 fb^{-1} was analyzed for these analyses so that almost twice as many $t\bar{t}$H events were expected. The fact that the Run 2 analysis still has a higher sensitivity can be attributed to multiple factors. One of the main reasons is the improved b-tagging algorithms at CMS in LHC Run 2, which achieve better b-tag efficiency for the same mistag rate. The smaller increase of the background compared to the signal production cross sections ($t\bar{t}$ increases by a factor of 3.3) can play a role, too. Finally, analysis improvements like the combination of MEM and BDT discriminants, the use of substructure techniques, and the automated BDT optimization also contribute to the better sensitivity.

10.2.6 Extrapolation

From the postfit model, expected limits for integrated luminosities $\mathcal{L}$ different from the one used in this analysis, $\mathcal{L}_0$, can be calculated. For this, the expected number of events is increased by $\mathcal{L}/\mathcal{L}_0$. Furthermore, systematic uncertainties that are dominated by statistical uncertainties of calibration measurements are expected to decrease by $\sqrt{\mathcal{L}/\mathcal{L}_0}$. In this analysis this behavior is assumed for the b-tagging uncertainties, except the influence of background contamination, and the JES and JER uncertainty. The Asimov dataset for the limit calculation is generated from the background model with nuisance parameters set to the values found in the fit to real data as this describes the best knowledge of their true value and it is also this method that was used to calculate the expected limit for the nominal luminosity. The result of the extrapolation is shown in Fig. 10.11. For between 25 fb^{-1} and 100 fb^{-1} the analysis is expected to reach Standard Model sensitivity, i.e. be able to exclude $t\bar{t}$H at the 95% confidence level if it does not exist.

One important caveat is that the extrapolation assumes that the statistical model used to fit data corresponding to 2.7 fb^{-1} is flexible enough to describe the distributions expected for 50 fb^{-1}. This is most likely not the case. For example for the fit of the 50 fb^{-1} Asimov dataset, the JES uncertainty is constrained to 20% of the prefit uncertainty, even though the prefit uncertainty is already factor four smaller than the one assumed in the 2.7 fb^{-1} analysis. Working at such a precision would require additional flexibility in the fit to allow to model uncertainties of jets in different η regions or of different flavor partly uncorrelated. Also problematic is the constraint on the CSV statistical uncertainties to 30%. While the analysis is in principle able to constrain the b-tagging uncertainty, this constraint seems questionable because of the simplified treatment of the statistical uncertainties. Lastly, the parton shower scale for $\mathrm{t\bar{t}+lf}$ is constrained at the 5% level for the 50 fb^{-1} extrapolation. While it is plausible that the fit can constrain this parameter from the jet multiplicity distribution alone, there is certainly parton shower mismodeling imaginable that leads to a shape change that is different from the one parameterized in this analysis.

From this, a task for future analyses can be derived. With an increased dataset, statistical uncertainties will become smaller and the analysis will become sensitive to more subtle systematic effects. It will thus be important to improve the statistical model. For the experimental uncertainties this is rather straightforward (but not trivial), uncertainties like the ones associated with the measurement of jet energy scale and b-tag discriminant have to be divided into components and their correlation has to be analyzed. For theoretical uncertainties a starting point could be to improve the scheme used to calculate the $\mathrm{t\bar{t}b\bar{b}}$ background. Furthermore, it will be necessary to introduce additional degrees of freedom to the model of all $\mathrm{t\bar{t}}$ backgrounds to allow for shape changes that are beyond the ones introduced by varying matrix element and parton shower scale by a single constant factor.

11 Direct constraints on anomalous top-Higgs couplings

The analysis of the $\mathrm{t\bar{t}H}$ signal strength documented in the previous chapter is optimized to detect $\mathrm{t\bar{t}H}$ production as predicted by the Standard Model. In Section 11.1 it will be shown that the analysis is also able to detect anomalous $\mathrm{t\bar{t}H}$ production.

With the signal-strength analysis setup of the previous section it is however not possible to distinguish between a SM and an anomalous top-Higgs coupling. The ability to do so can be obtained by modifying the analysis such that it allows for both signal-background and scalar-pseudoscalar discrimination, as is shown in Section 11.2. The constraints on an anomalous top-Higgs coupling that can be derived with this modified analysis are presented in Section 11.3.

11.1 Signal-strength analysis with non-standard signal

11.1.1 Signal simulation

The anomalous coupling $\mathrm{t\bar{t}H}$ simulation introduced in section 7.3 is used to test the sensitivity of the coupling strength analysis of the previous chapter to a non-standard signal. Events generated with this simulation are used instead of events from the Standard Model $\mathrm{t\bar{t}H}$ simulation throughout the analysis.

For the anomalous coupling MC simulation of $\mathrm{t\bar{t}H}$, a leading order matrix element generator is used. To ensure that the leading order simulation is an adequate description of the signal, its discriminant distributions are compared with the distributions of the default POWHEG simulation. The main difference is the jet multiplicity distribution and with it a 4% difference in the acceptance of the event selection. To avoid inconsistencies with the measurement presented in the previous chapter, reweighting factors are determined so that the jet multiplicity distribution of the leading order SM $\mathrm{t\bar{t}H}$ simulation matches the one predicted by the default POWHEG simulation. Distributions before and after reweighting are shown in Fig. 11.1. The same jet-multiplicity dependent weights are then applied to the pseudoscalar simulation, too.

As discussed in Section 7.3, the effect of an anomalous coupling that has a scalar and a pseudoscalar component can be simulated by interpolating between the distributions of a purely scalar (CP even, $\mathrm{t\bar{t}H_+}$) and a purely pseudoscalar signal (CP odd, $\mathrm{t\bar{t}H_-}$). Shape differences between the two extreme cases are shown in Fig. 11.2. Distributions relevant for the $\mathrm{t\bar{t}H}$ signal-strength analysis, i.e. jet and b-tag multiplicity and the BDT and MEM discriminants are rather similar for both extremes. The analysis is thus able to detect an anomalous $\mathrm{t\bar{t}H}$ signal but not able to distinguish it well from a SM signal.

11.1.2 Modification to the statistical model

The statistical model already utilized in the previous chapter is used with only small modifications. Instead of only one parameter of interest, the signal is described by two

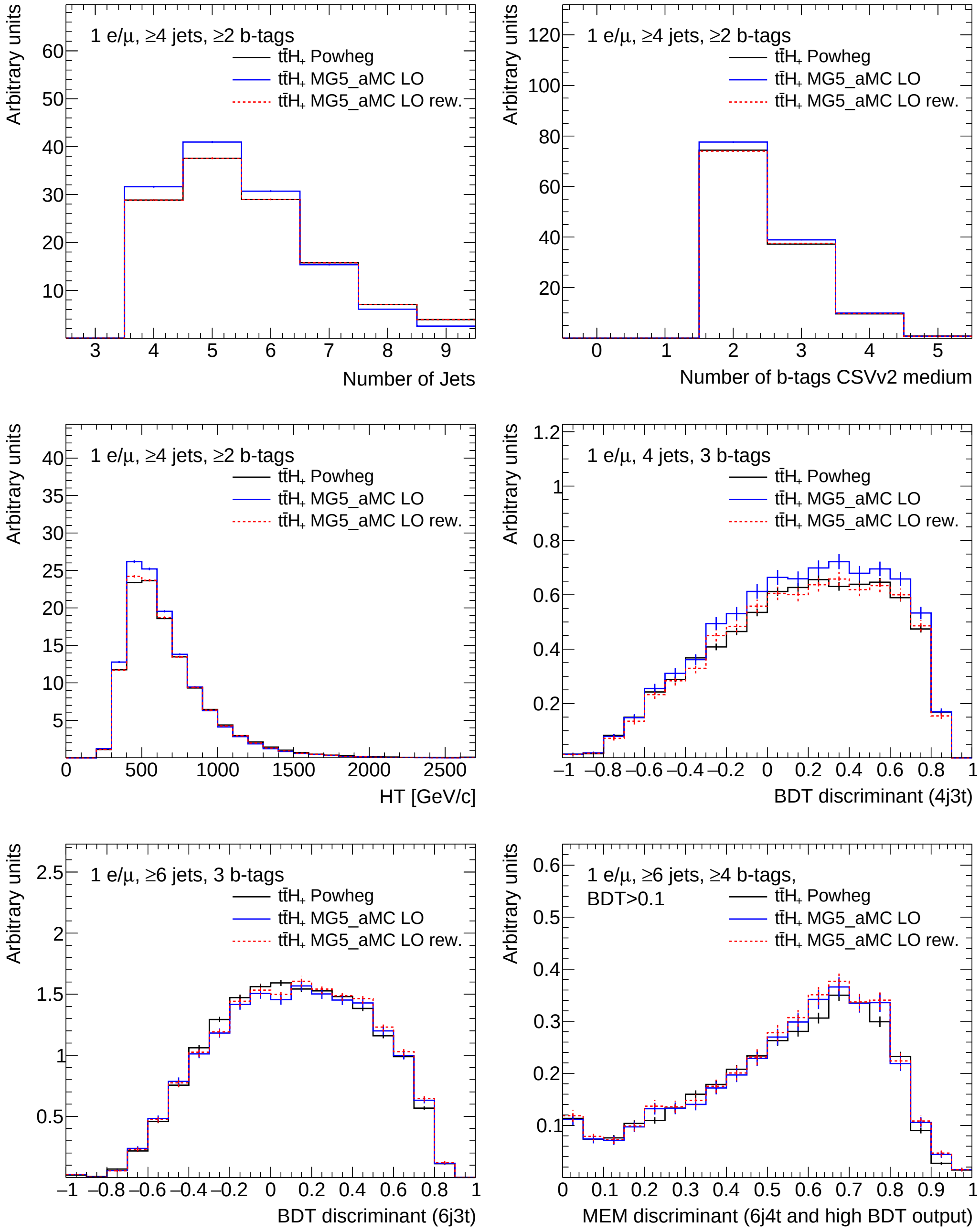


Figure 11.1: Comparison of the Powheg $t\bar{t}H$ simulation (black line) with the privately produced MadGraph5_aMC@NLO leading order sample of $t\bar{t}H_{-}$ (solid blue line) as well as to the latter sample after jet-multiplicity reweighting (dashed red line). Statistical uncertainties are indicated as vertical lines on the histograms. In the top row, jet and b-tag multiplicities are shown and the scalar sum of jet p_{T} is displayed in the middle row on the left-hand side. Moreover, three discriminants of the signal-strength analysis described in the previous chapter are shown in the middle row on the right-hand side and in the bottom row.

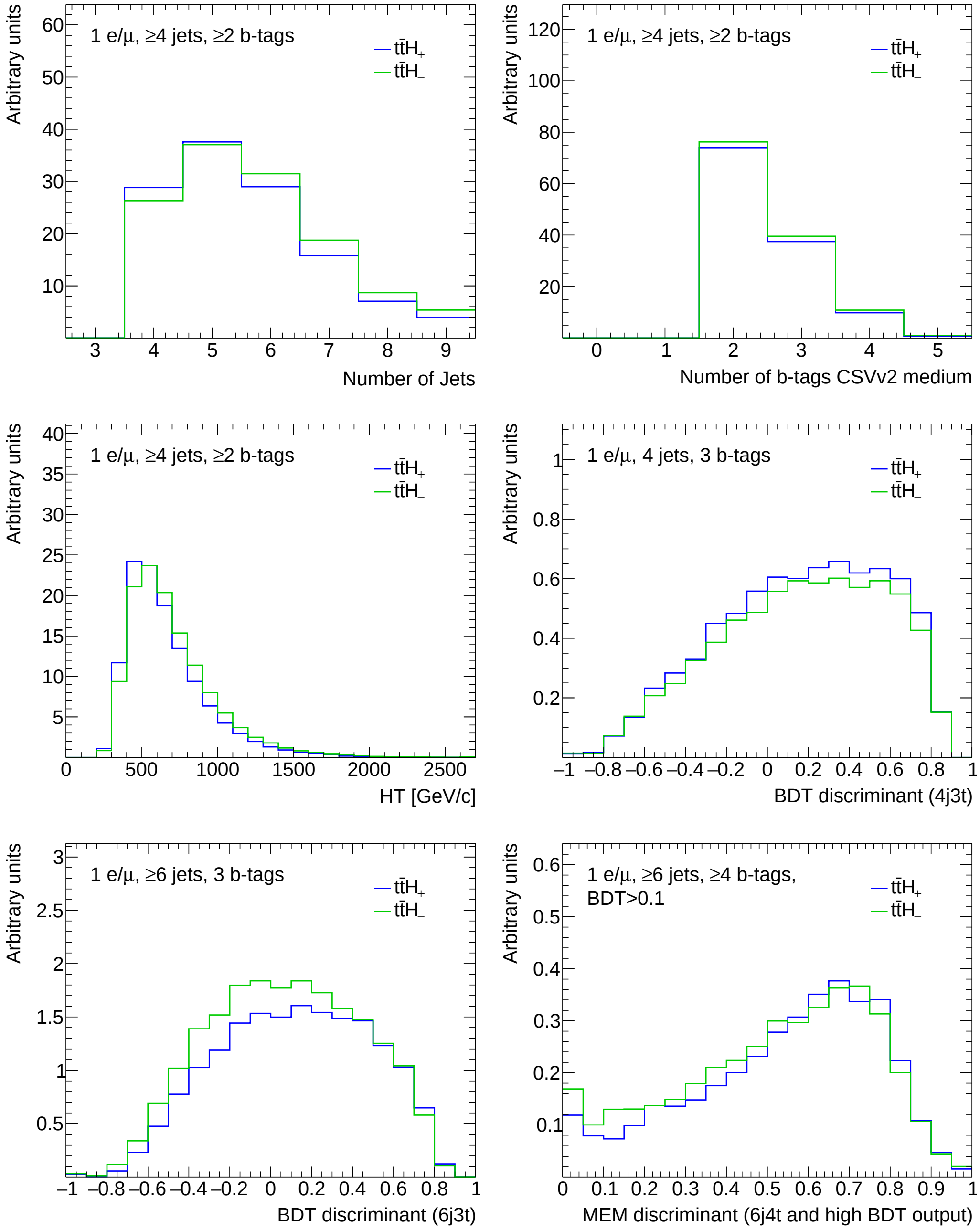


Figure 11.2: Comparison between the MadGraph5_aMC@NLO + Higgs Characterization Model $t\bar{t}H$ simulations with a scalar (blue) and pseudoscalar (green) Higgs boson. Statistical uncertainties are indicated as vertical lines on the histograms. In the top row, jet and b-tag multiplicities are shown and the scalar sum of jet p_T is displayed in the middle row on the left-hand side. Moreover, three discriminants of the signal-strength analysis described in the previous chapter are shown in the middle row on the right-hand side and in the bottom row.

parameters, e.g. the coupling modifiers κ_t and $\tilde{\kappa}_\mathrm{t}$.

One additional nuisance parameter is introduced, corresponding to a 10% log-normal rate uncertainty of the yield associated with a pseudoscalar coupling. This value is meant to cover two sources of uncertainties: the one introduced by the the jet-multiplicity reweighting scheme described above and the increase of the scale and PDF uncertainty with a pseudoscalar admixture discussed below. The SM uncertainty,

$$\Delta\sigma_{\mathrm{t\bar{t}H}} =^{+5.8\%}_{-9.2\%} \text{ (scale) } \pm 3.3\% \text{ (PDF)},$$

is applied independent of the admixture, as it is strongly correlated between $\mathrm{t\bar{t}H_-}$ and $\mathrm{t\bar{t}H_+}$. For the pseudoscalar top-Higgs coupling the total uncertainty, calculated using the MadGraph5_aMC@NLO configuration described in Section 7.3 and the PDF4LHC [166] prescription, is

$$\Delta\tilde{\sigma}_{\mathrm{t\bar{t}H}} =^{+6.4\%}_{-10.5\%} \text{ (scale) } \pm 4.1\% \text{ (PDF)},$$

so that the small increase is generously covered by the 10% uncertainty.

The Higgs-boson branching ratios are assumed to be SM-like for this analysis. Strictly, this is an inconsistent treatment, at least in the “resolved” model introduced in Chapter 6. The effect of an anomalous coupling is considered in the production but not in the decay of the Higgs boson, e.g. into gluons, which would then reduce the $\mathrm{b\bar{b}}$ branching ratio of the Higgs boson. However, for the sake of easy interpretability of the result such effects are not considered.

11.1.3 Upper limits on a non-standard signal, using the Standard Model analysis

Upper limits are evaluated in terms of the anomalous coupling parameterization introduced in Equation (5.3). The mixing between scalar and pseudoscalar coupling and thus the signal-shape changing effect is parameterized by ζ_t. Limits on the $\mathrm{t\bar{t}H}$ production cross section are calculated for different signal shapes and the observed and expected results are displayed on the left-hand side in Fig. 11.3. The observed limit is between $\mu_{\mathrm{t\bar{t}H}} < 4.0$ for a scalar and $\mu_{\mathrm{t\bar{t}H}} < 3.5$ for a pseudoscalar signal. The data is thus slightly less compatible with a large pseudoscalar signal than it is with a scalar signal. The sensitivity for a scalar and a pseudoscalar signal is almost independent of ζ_t: the median expected limit for the background-only hypothesis varies between 4.1 and 4.0.

For the coupling modifier $\bar{\kappa}_\mathrm{t}$ from Equation (5.3) an upper limit can be calculated as a function of ζ_t, too. It is shown in Fig. 11.3 on the right. The observed limit varies between 2.0 for the SM signal and 2.9 for a purely pseudoscalar Higgs boson. The fact that larger values of $\bar{\kappa}_\mathrm{t}$ are allowed in the latter case can be explained by the associated cross section, which is a factor of 2.3 smaller for the same $\bar{\kappa}_\mathrm{t}$ and $\zeta_\mathrm{t} = \pi/2$ compared to the SM at $\zeta_\mathrm{t} = 0$.

The best fit value for the signal strength of a purely pseudoscalar signal is

$$\mu^{\mathrm{obs.}}_{\mathrm{t\bar{t}H_-}} = -1.5 \pm 2.2,$$

compared to $\mu^{\mathrm{obs.}}_{\mathrm{t\bar{t}H_+}} = -0.4 \pm 2.1$. for the scalar signal model. The fit can also be performed as a two-dimensional fit for the scalar and pseudoscalar top-Higgs coupling modifiers κ_t and $\tilde{\kappa}_\mathrm{t}$. With this parameterization no negative signals are possible and the best fit is at $\kappa_\mathrm{t} = 0$ and $\tilde{\kappa}_\mathrm{t} = 0$. Confidence regions are displayed in Fig. 11.4, calculated under the assumption that the profile likelihood ratio test statistic $q(\kappa_\mathrm{t}, \tilde{\kappa}_\mathrm{t})$ (see equation 6.9)

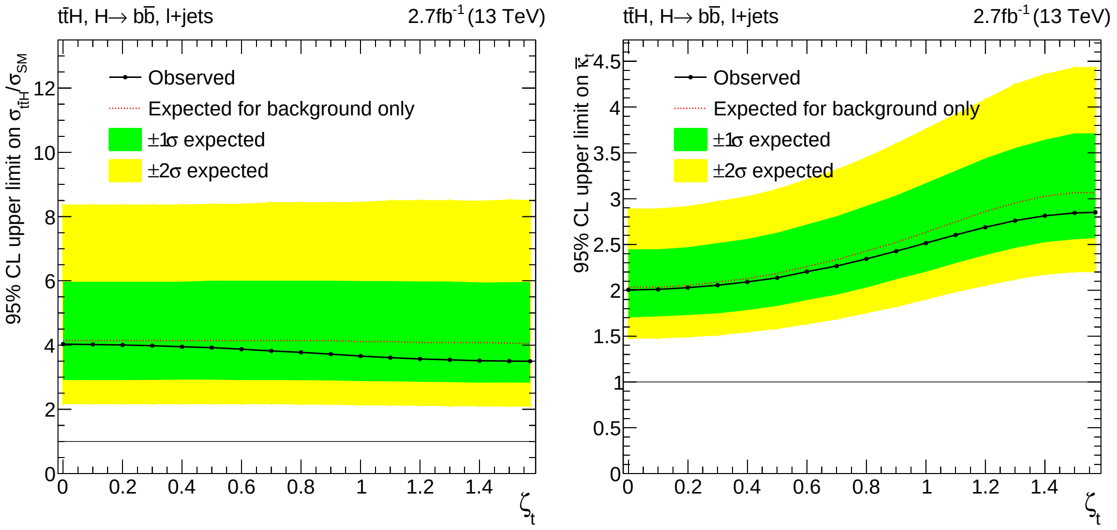


Figure 11.3: Observed and expected 95% CLs upper limit on the $\mathrm{t\bar{t}H}$ signal strength (left) and coupling strength $\tilde{\kappa}_\mathrm{t}$ (right) as a function of the scalar-pseudoscalar mixing angle ζ_t, calculated with the asymptotic method. Observed limits are drawn as black lines and points, median expected limits as dashed red lines. The 1σ (2σ) confidence intervals are shown as green (yellow) bands. $\zeta_\mathrm{t} = 0$ corresponds to the Standard Model, $\zeta_\mathrm{t} = \pi/2$ to a pseudoscalar boson.

follows a two-dimensional χ^2-distribution. As stated before, larger values for $\tilde{\kappa}_\mathrm{t}$ than for κ_t are allowed by the data because of the smaller cross section associated with the pseudoscalar coupling modifier. While the SM expectation for the best-fit value is at $(\kappa_\mathrm{t}, \tilde{\kappa}_\mathrm{t}) = (1, 0)$, the $(0, 1.5)$-point is almost equally likely. The analysis is thus blind to the coupling structure, which is not expected to change with higher integrated luminosity. The expected confidence regions and best-fit values are plotted in the right plot of Fig. 11.4.

11.2 Analysis of the top-Higgs coupling structure

11.2.1 Kinematic differences

To be able to measure a pseudoscalar admixture to the top-Higgs coupling from the analysis of $\mathrm{t\bar{t}H}$ events, the signal-strength analysis has to be modified. The first step is to construct a discriminant that can detect the kinematic differences between the two signals. In Fig. 11.5 it is shown that the top-quark and Higgs-boson kinematics clearly depend on the size of the pseudoscalar admixture to the $\mathrm{t\bar{t}H}$ coupling. Some of the main effects are that, on average, for a CP odd Higgs boson $\mathrm{t\bar{t}H}$ is produced at higher center-of-mass energies, the Higgs boson has a stronger transverse boost and is less aligned with the softer top quark. Maybe the most striking feature is that there is a larger gap in pseudorapidity between the two top quarks for a pseudoscalar coupling. Some of those effects have also been pointed out in, e.g. [140, 142, 146–148, 165, 241, 242].

It is also remarkable that while the amplitudes associated with the first, third, and fourth leading order diagram displayed in 7.2 contribute similarly to the SM $\mathrm{t\bar{t}H}$ production cross section, for the pseudoscalar boson the third diagram, in which the Higgs boson is connected to a top-quark exchanged in the t-channel, is dominant by far.

Hence it should clearly be possible to construct a discriminant that is sensitive to these

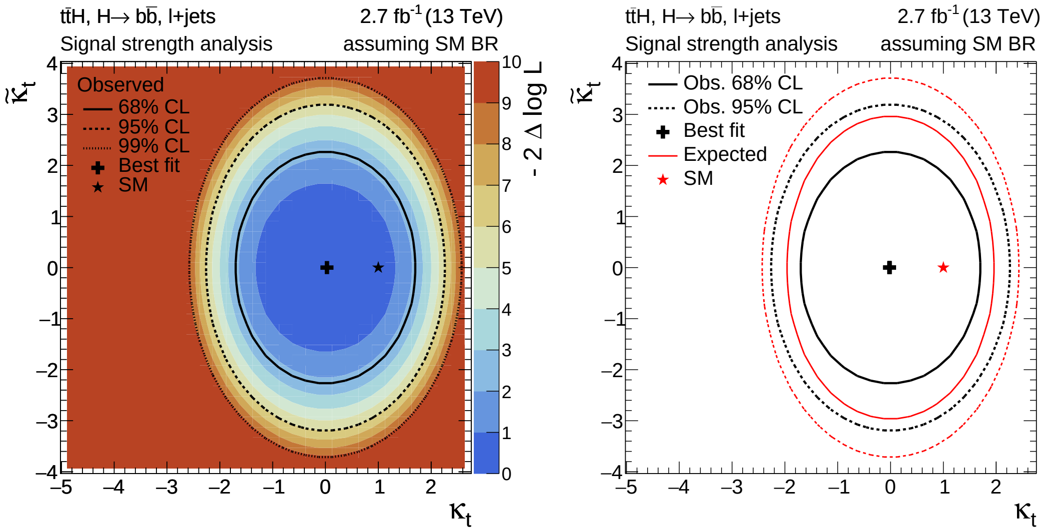


Figure 11.4: Profile likelihood ratio test statistic $q(\kappa_\mathrm{t}, \tilde{\kappa}_\mathrm{t})$ obtained from the coupling strength measurement of the previous chapter, fitted with an anomalous coupling signal. The left plot shows the observed result and the right plot the SM expectation. In the left plot, black lines indicate the borders of 68% to 99% confidence level regions for κ_t, and $\tilde{\kappa}_\mathrm{t}$, assuming $q(\kappa_\mathrm{t}, \tilde{\kappa}_\mathrm{t})$ follows a two-dimensional χ^2 distribution while black crosses indicate the minima of the likelihood and a black star the Standard Model expectation. In the right plot black (red) lines indicate the confidence intervals calculated from the observation (SM expectation).

differences. Ultimately Boosted Decision Trees are trained for this purpose. The most important input for the BDTs is a matrix element discriminant.

11.2.2 Matrix element CP-discriminant

At leading order perturbation theory the complete probability density function that describes the observation of a tt̄H event with a certain kinematic configuration can be calculated from the leading order matrix element, the parton distribution functions, and phase space factors (see Section 2.1). For a given observation of top-quark and Higgs-boson momenta p_t, $p_{\bar{\mathrm{t}}}$, and p_H, the likelihoods $L_{\mathrm{t\bar{t}H}_+}$ and $L_{\mathrm{t\bar{t}H}_-}$ to observe tt̄H with a scalar ($\mathrm{t\bar{t}H}_+$) or pseudoscalar Higgs boson ($\mathrm{t\bar{t}H}_-$) can thus be calculated. Under the leading order approximation the most powerful discriminant (see Section 4.1.2) is then the likelihood ratio

$$\lambda_\mathrm{CP} = \frac{L_{\mathrm{t\bar{t}H}_+}(p_\mathrm{t}, p_{\bar{\mathrm{t}}}, p_\mathrm{H})}{L_{\mathrm{t\bar{t}H}_+}(p_\mathrm{t}, p_{\bar{\mathrm{t}}}, p_\mathrm{H}) + cL_{\mathrm{t\bar{t}H}_-}(p_\mathrm{t}, p_{\bar{\mathrm{t}}}, p_\mathrm{H})}. \tag{11.1}$$

The normalization constant $c = 2.5$ does not affect the discrimination power but is chosen to cancel the effect of the smaller cross section of $\mathrm{t\bar{t}H}_-$ compared to $\mathrm{t\bar{t}H}_+$.

The calculation of λ_CP is simplified by only considering gluon-initiated tt̄H production. This initial state dominates with 80% for the scalar and almost 100% for the pseudo scalar coupling. For a given observation of final state momenta, PDFs and phase space factors are thus the same for nominator and denominator in Equation (11.1) and cancel, so that it simplifies to

$$\lambda_\mathrm{CP} \approx \frac{|\mathcal{M}_{\mathrm{t\bar{t}H}_+}(p_\mathrm{t}, p_{\bar{\mathrm{t}}}, p_\mathrm{H})|^2}{|\mathcal{M}_{\mathrm{t\bar{t}H}_+}(p_\mathrm{t}, p_{\bar{\mathrm{t}}}, p_\mathrm{H})|^2 + c|\mathcal{M}_{\mathrm{t\bar{t}H}_-}(p_\mathrm{t}, p_{\bar{\mathrm{t}}}, p_\mathrm{H})|^2}. \tag{11.2}$$

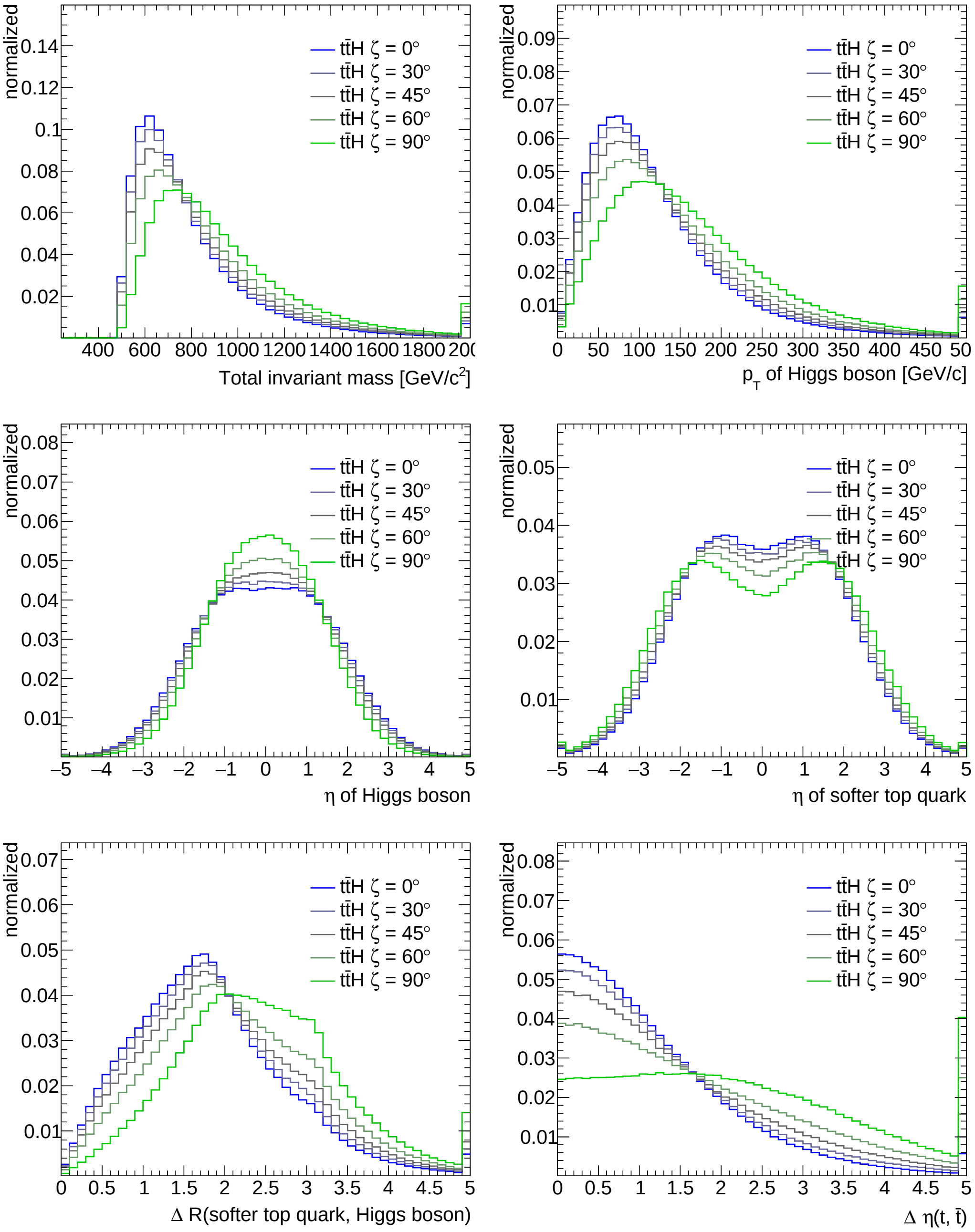


Figure 11.5: Kinematic differences between different coupling mixtures on parton level, calculated with MadGraph5_aMC@NLO and the Higgs Characterization Model. The predicted distribution of the variables for a purely scalar (pseudoscalar) boson is displayed in blue (green).

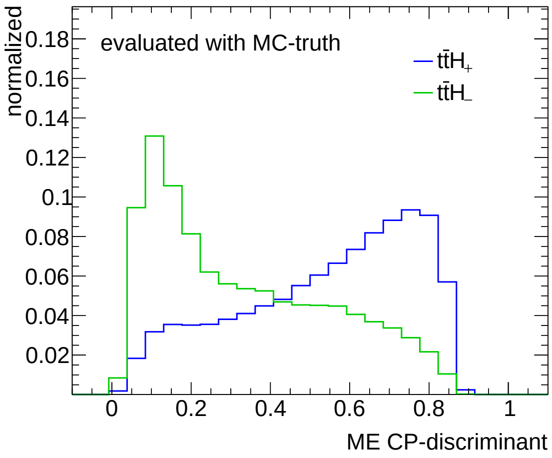


Figure 11.6: Distribution expected for $t\bar{t}H$ production via a purely scalar (blue line) and purely pseudoscalar coupling (green line) for the matrix element CP-discriminant evaluated with the generator level top-quark and Higgs-boson momenta.

This simplification only holds if resolution effects are neglected and the four-momentum vectors are assumed to be measured perfectly.

Code for the calculation of the matrix elements $|\mathcal{M}_{t\bar{t}H_+}|^2$ and $|\mathcal{M}_{t\bar{t}H_-}|^2$ is generated using, again, MadGraph5_aMC@NLO and the Higgs Characterization Model. For simplicity, the matrix element is evaluated in the $t\bar{t}H$ center-of-mass frame. In this frame, the two initial gluon momenta are given by

$$p_g = \left(\frac{1}{2}\sqrt{(p_t + p_{\bar{t}} + p_H)^2}, 0, 0, \pm\frac{1}{2}\sqrt{(p_t + p_{\bar{t}} + p_H)^2}\right). \tag{11.3}$$

The performance of the discriminant under optimal circumstance, calculated with the parton level four vectors from the $t\bar{t}H_+$ and $t\bar{t}H_-$ samples, is shown in Fig. 11.6. A similar, independently developed discriminant, constructed within the MELA framework [243–245] is used in [142] and seems to have a comparable performance.

CP matrix element discriminant on reconstruction level

It is not completely straightforward to evaluate the discriminant with reconstructed physics objects in $t\bar{t}H(b\bar{b})$ events. For this, some modifications have to be implemented. One problem is that the reconstructed energy and momentum vectors of the top quarks and the Higgs boson suffer from the imprecise energy measurements of jets and $\not{E}_T$ and can have invariant masses m_{rec} far from the pole masses m_{true}. Before evaluating the matrix elements, energy and momentum of the three objects are therefore scaled by a constant factor:

$$p^{\text{rec}} \rightarrow p^{\text{corr}} = \frac{m_{\text{true}}}{m_{\text{rec}}} p^{\text{rec}} \tag{11.4}$$

Secondly, the transverse component of the $t\bar{t}H$ momentum is non-zero because of jet energy mismeasurement and initial state radiation not considered in the matrix element. This is solved by evaluating the momenta of the Higgs boson and $t\bar{t}$ in the center-of-mass frame, ignoring the transverse boost.

Most importantly, there is a combinatorial problem when trying to assign jets to quarks, and the missing transverse momentum can only identified with two of the neutrino momentum components. Because of this ambiguity, reconstruction hypotheses are created similar to the method described in Section 9.5.1. Two solutions for the longitudinal momentum of the neutrino are calculated from $\not{E}_{\mathrm{T}}$, the lepton, and a W-boson mass constraint. Additionally, all possible assignments of jets to the six quarks from the $\mathrm{t\bar{t}H(b\bar{b})}$ decay are considered. The only constraint to limit the number of permutations is that at most one jet not interpreted as a b quarks can be b-tagged. The likelihood that the reconstructed $\mathrm{t\bar{t}H}$ event corresponds to a scalar or pseudoscalar coupling is evaluated using not only one but all possible associations of jets to quarks as well as both neutrino momentum solutions.

Weighting of hypotheses

The reconstruction hypotheses are not all equally likely, which is reflected in the reconstructed masses and b-tags of jets. Correct hypotheses tend to have masses closer to the true masses of the resonances in $\mathrm{t\bar{t}H}$. Furthermore, b-tag discriminant values of jets associated with b quarks tend to be higher than those of jets associated with light jets. To consider this, two likelihood functions are introduced. The first one,

$$L_{\text{b-tag}}(\{b_i\}) = \prod_{i=1}^{4} L_b(b_i) \prod_{i=5}^{6} L_l(b_i), \tag{11.5}$$

describes the likelihood that the four jets associated with the b quarks from the $\mathrm{t\bar{t}H}$ decay, with b-tags b_1 to b_4, really originate from b quarks and the remaining two jets, with b-tags b_5 and b_6, originate from light quarks. The expected b-tag discriminant distribution for light jets L_l and b-jets L_b is determined from the $\mathrm{t\bar{t}H}$ MC simulation and displayed in Fig. 11.7.

The mass likelihood,

$$L_m(m^{\mathrm{rec}}_{\mathrm{t_{lep}}}, m^{\mathrm{rec}}_{\mathrm{t_{had}}}, m^{\mathrm{rec}}_{\mathrm{W_{had}}}, m^{\mathrm{rec}}_{\mathrm{H}}) = L_{m_{\mathrm{t_{lep}}}}(m^{\mathrm{rec}}_{\mathrm{t_{lep}}}) L_{m_{\mathrm{t_{had}}}}(m^{\mathrm{rec}}_{\mathrm{t_{had}}}) L_{m_{\mathrm{W_{had}}}}(m^{\mathrm{rec}}_{\mathrm{W_{had}}}) L_{m_{\mathrm{H}}}(m^{\mathrm{rec}}_{\mathrm{H}}), \tag{11.6}$$

is determined from the probability to observe reconstructed resonance masses, $m^{\mathrm{rec}}_{\mathrm{t,lep}}$, $m^{\mathrm{rec}}_{\mathrm{t,had}}$, $m^{\mathrm{rec}}_{\mathrm{W,had}}$, and $m^{\mathrm{rec}}_{\mathrm{H}}$ under the assumptions that the jets truly originate from the corresponding quarks. The probability densities are determined from the $\mathrm{t\bar{t}H(b\bar{b})}$ simulation by evaluating the reconstructed invariant masses of jets that can be matched to the quarks from the resonance decays. They are also shown in Fig. 11.7. For the construction of L_m, the individual likelihoods of the four masses are simply multiplied, assuming no correlation between the masses, which is of course only a bad approximation in the case of the hadronic W-boson and top-quark masses. The mass of the leptonically decaying W boson is not part of this likelihood as it is already assumed to be 80.4 GeV/c^2 in the reconstruction of the neutrino momentum.

With the two likelihoods and the matrix element at hand, the likelihood for a single reconstruction hypothesis h of a $\mathrm{t\bar{t}H}$ candidate event to belong to $\mathrm{t\bar{t}H_+}$ or $\mathrm{t\bar{t}H_-}$ can be calculated as

$$L^h_{\mathrm{t\bar{t}H_\pm}} = L_{\text{b-tag}}(\{b_i\}) \times L_m(m^{\mathrm{rec}}_{\mathrm{t_{lep}}}, m^{\mathrm{rec}}_{\mathrm{t_{had}}}, m^{\mathrm{rec}}_{\mathrm{W_{had}}}, m^{\mathrm{rec}}_{\mathrm{H}}) \times |\mathcal{M}_{\mathrm{t\bar{t}H_\pm}}(p^{\mathrm{corr}}_{\mathrm{t_{lep}}}, p^{\mathrm{corr}}_{\mathrm{\bar{t}_{had}}}, p^{\mathrm{corr}}_{\mathrm{H}})|^2, \tag{11.7}$$

where the b-tags, $\{b_i\}$, reconstructed masses $\{m^{\mathrm{rec}}_i\}$, and $\{p^{\mathrm{corr}}_i\}$ depend on the reconstruction hypothesis h. Parton distribution functions and phase space factors are dropped

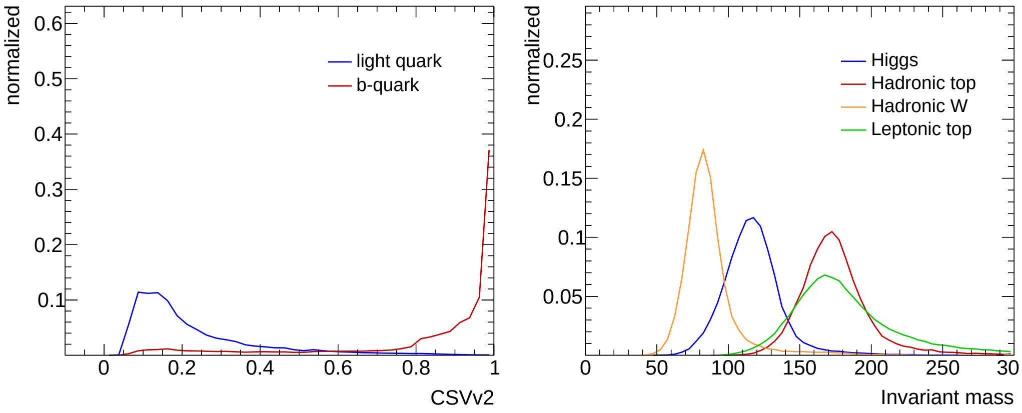


Figure 11.7: Distributions used to evaluate the likelihood of $\mathrm{t\bar{t}H}$-system reconstruction hypotheses. On the left the probabilities to observe b-tagging discriminant outputs of jets originating from a b quark (red) and light quark (blue) is displayed. On the right multiple probability density functions for the reconstructed masses of hadronic resonances in $\mathrm{t\bar{t}H}$ events are shown. The functions are determined from the SM $\mathrm{t\bar{t}H}$ simulation by angular matching of jets to the quarks from the resonance decays.

as they are approximately the same for all hypotheses and thus cancel in a likelihood ratio[1]. To calculate the $\mathrm{t\bar{t}H_+}$ and $\mathrm{t\bar{t}H_-}$ likelihood of an event, the likelihoods for all interpretations are summed up:

$$\lambda_{\mathrm{CP}}^{\mathrm{reco}} = \frac{\sum_h L^h_{\mathrm{t\bar{t}H_+}}}{\sum_h L^h_{\mathrm{t\bar{t}H_+}} + c\sum_h L^h_{\mathrm{t\bar{t}H_-}}}. \tag{11.8}$$

By this procedure much of the discrimination power of the discriminant can still be utilized with reconstructed events, as can be seen by comparing Fig. 11.8, for which this procedure is performed with six-jet events, with the discriminant for perfectly reconstructed objects in Fig. 11.6.

The performance of the matrix element CP-discriminant $\lambda_{CP}^{\mathrm{reco}}$ is superior to all alternative constructions tested, for example evaluating 11.1 only for the interpretation with the largest mass and b-tag likelihood $L_{\text{b-tag}}L_m$ or compared to using the interpretation that maximizes $L^h_{\mathrm{t\bar{t}H_+}}$ ($L^h_{\mathrm{t\bar{t}H_-}}$) in the nominator (denominator) of the likelihood ratio in Equation (11.8).

11.2.3 BDT CP-discriminant

The matrix element (ME) CP-discriminant is powerful in events with at least six jets. In the remaining events machine learning can be used to identify a pseudoscalar admixture to the top-Higgs coupling. Gradient boosted BDTs are trained in the seven most sensitive categories of the signal-strength analysis. For the sake of consistency, BDTs are also used in the six-jet categories, with the ME CP-discriminant as input, even though they hardly improve the performance compared to using the ME discriminant alone.

The training of the BDTs is performed with a heuristic that is less sophisticated than the Particle Swarm Optimization algorithm used for the signal-background separating BDT.

---

[1] Since the phase space depends on the momenta of top quarks and Higgs boson, which depend on the reconstruction hypothesis, this is not exactly true. Phase space factors would cancel if the full process $\mathrm{gg} \rightarrow \mathrm{t\bar{t}H} \rightarrow \mathrm{b}q\bar{q}\mathrm{b}l\nu_l\mathrm{b\bar{b}}$ was considered.

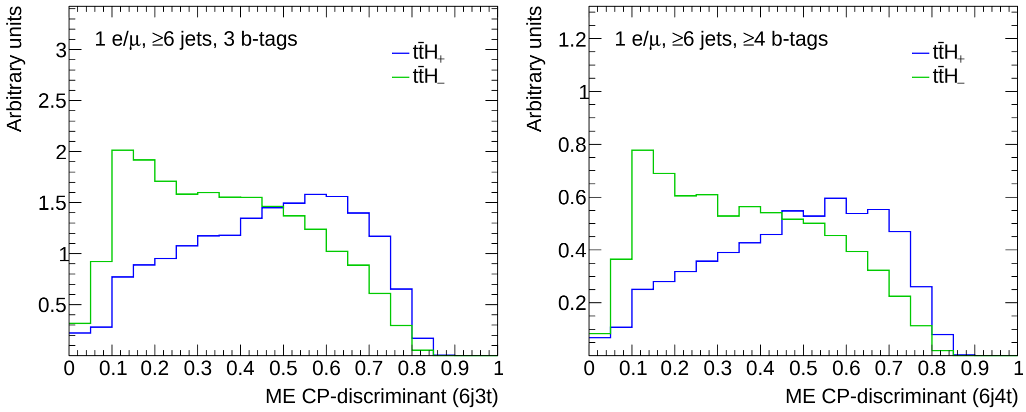


Figure 11.8: Matrix element discriminant designed to detect an anomalous top-Higgs coupling in $t\bar{t}H$. The expected distribution for a CP even (odd) and Higgs boson is shown in blue (green). The distributions are calculated from the simulation of reconstructed objects. For the left plot the discriminant is evaluated for events with at least six jets and three b-tags, for the right plot for events with at least six jets and at least four b-tags.

Table 11.1: Configuration of the BDTs used to distinguish pseudoscalar form scalar coupling.

Category	N_{trees}	shrinkage	bagging fraction	N_{cuts}	N_{vars}
5 jets, 3 b-tags	100	0.02	0.6	60	6
$\geq$ 6 jets, 3 b-tags	200	0.02	0.6	60	6
4 jets, 4 b-tags, $\text{BDT}_{\text{SB}} > 0.2$	30	0.02	0.6	60	4
5 jets, $\geq$ 4 b-tags, $\text{BDT}_{\text{SB}} > 0.2$	100	0.02	0.6	60	4
$\geq$ 6 jets, $\geq$ 4 b-tags, $\text{BDT}_{\text{SB}} < 0.1$	50	0.02	0.6	60	4
$\geq$ 6 jets, $\geq$ 4 b-tags, $\text{BDT}_{\text{SB}} > 0.1$	30	0.02	0.6	60	4
Boosted	100	0.02	0.6	60	6

Most configuration parameters are set to values that have proven to yield a reasonable performance. The number of trees mostly depends on the number of available training events and is set to a comparatively low value, so that overtraining is almost non-existent. The BDT configuration is summarized in Table 11.1.

The observables used in the BDT training are summarized in Table 11.2. Only features well-modeled in the MC simulation are considered. They can be classified by the characteristics of a pseudoscalar admixture they try to identify: the high-p_{T} Higgs boson, the large pseudorapidity gap between the two top quarks, and the Higgs boson being close to one of the top quarks. The ME CP-discriminant is used in the six-jet categories and in the boosted category for events that have at least six jets.

The selection of BDT inputs is performed similar to the signal-strength analysis. First, a BDT is trained with an initial set of features. Then the effect of adding any of the additional input candidates is tested. Only the variable yielding the largest improvement of the ROC integral is kept. This is repeated until the maximum number of inputs for the BDT is reached. After that the effect of removing any input variable is tested, the subset of inputs yielding the best ROC is kept, until a minimum number of inputs is reached.

Table 11.2: Description of input variables of the Boosted Decision Trees separating $\mathrm{t\bar{t}H}$ with scalar top-Higgs coupling from $\mathrm{t\bar{t}H}$ with pseudoscalar coupling.

Observable	**Description**
High p_T Higgs boson	
Highest tagged dijet p_T	b-tagged jets with the highest combined transverse momentum
Highest p_T tagged dijet mass	mass of b-tagged jets with the highest combined p_T
Closest tagged dijet mass	mass of the two b-tagged jets closest in ΔR
Total invariant mass	invariant mass of jets, lepton, and $\not{E}_\mathrm{T}$
Large $\Delta\eta(t,\bar{t})$	
$\max\Delta\eta(\mathrm{jet},\mathrm{lepton})$	maximal η difference of lepton and any jet
$\max\Delta\eta(\mathrm{tag},\mathrm{tag})$	maximal η difference of two tagged jets
avg $\Delta\eta$(tag,tag)	average difference in η between b-tagged jets
avg ΔR(jet,jet)	average η-ϕ distance between jets
ΔR(lepton, next-to-closest jet)	next-to-smallest ϕ-η distance between the lepton any jet
Small $\Delta R(t,H)$	
min ΔR(lepton,tag)	smallest ϕ-η distance between the lepton and b-jet
m_{lb}	invariant mass of lepton and closest b-tagged jet
min ΔR(boosted Higgs,lepton)	ϕ-η distance between lepton and boosted Higgs-boson candidate
ME CP-discriminant	matrix element CP-discriminant

Phases of adding and removing inputs are alternated until a stable set with $\mathrm{N_{vars}}$ variables is found. The variables used in the categories are summarized in Table 11.3. Examples of BDT inputs and comparisons between distributions in data and simulation are shown in Fig. 11.9. The outputs of the CP-discriminating BDT, with comparison between $\mathrm{t\bar{t}H_+}$ and $\mathrm{t\bar{t}H_-}$ as well as between data and simulation can be seen in Fig. 11.10. The BDT output is centered at zero and not distributed between -1 and $+1$ because of the comparatively little discrimination that can be achieved between $\mathrm{t\bar{t}H_-}$ and $\mathrm{t\bar{t}H_+}$, as the BDT output is a function of the expected purity.

11.2.4 Modification of the signal-strength analysis

The CP-discriminating BDT is utilized by splitting the seven most signal-sensitive categories into two subcategories each, one more sensitive for $\mathrm{t\bar{t}H}$ with a scalar Higgs boson and one more sensitive for $\mathrm{t\bar{t}H_-}$. Events with a positive output of the CP-discriminating BDT are analyzed in $\mathrm{t\bar{t}H_-}$-enriched categories, the remaining events in $\mathrm{t\bar{t}H_+}$-enriched categories. Within the subcategories, the S/B-discriminants of the signal-strength analysis (MEM and BDT discriminants) are used to separate signal and background. This is possible without further modifications because the output of the CP and the S/B discriminants is only weakly correlated. With this setup the Standard Model $\mathrm{t\bar{t}H}$ signal is expected to appear in the $\mathrm{t\bar{t}H_+}$-enriched categories while a pseudoscalar admixture can be detected by a signal in the $\mathrm{t\bar{t}H_-}$-enriched categories.

The expected and observed event yields are displayed in Table 11.4. In the $\mathrm{t\bar{t}H_-}$-enriched subcategories the $\mathrm{t\bar{t}H_-/t\bar{t}H_+}$-ratio is approximately twice as high as in the $\mathrm{t\bar{t}H_+}$-enriched categories. The observed number of events is mostly in accordance with the prediction of the nominal model, only in the $\mathrm{t\bar{t}H_+}$-enriched subcategories of the signal-depleted ≥ 6 jets, ≥ 4 b-tags category and the boosted category a significant deficit of data is observed. It is likely introduced by the subpar description of the $\mathrm{t\bar{t}}$ background by the nominal model and covered by the systematic uncertainties.

Table 11.3: Observables used for the training of the CP-discriminating BDT in the most sensitive analysis categories, sorted by the kinematic features of $\mathrm{t\bar{t}H}$ events they target.

Observable	**5j3t**	**6j3t**	**4j4th**	**5j4th**	**6j4tl**	**6j4th**	**boosted**
High p_T Higgs boson							
Highest tagged dijet p_T	✓	✓		✓			✓
Highest p_T tagged dijet mass					✓	✓	
Closest tagged dijet mass			✓				
Total invariant mass	✓			✓			
Large $\Delta\eta(t,\bar{t})$							
$\max\Delta\eta(\mathrm{jet},\mathrm{lepton})$	✓	✓	✓		✓	✓	✓
$\max\Delta\eta(\mathrm{tag},\mathrm{tag})$	✓	✓	✓				
avg $\Delta\eta$(tag,tag)					✓	✓	✓
avg ΔR(jet,jet)	✓			✓			
ΔR(lepton, next-to-closest jet)		✓					
Small $\Delta R(t,H)$							
min ΔR(lepton, tag)	✓	✓		✓			
m_{lb}			✓				✓
min ΔR(boosted Higgs,lepton)							✓
ME CP-discriminant		✓			✓	✓	✓

Table 11.4: Expected and observed event yields in the categories designed to be either enriched in a scalar signal $\mathrm{t\bar{t}H}_+$ or a signal with a pseudoscalar boson $\mathrm{t\bar{t}H}_-$. In the remaining categories yields are unchanged w.r.t. to Table 9.1. The data column corresponds to events observed in 2015, for signal and background systematic uncertainties are added in quadrature.

Category	$\mathrm{t\bar{t}H}_+$	$\mathrm{t\bar{t}H}_-$	Background	Data	Data/Background
5 jets, 3 b-tags					
$\mathrm{t\bar{t}H}_-$-enriched	4.3 ± 0.7	2.8 ± 0.4	1035.0 ± 282.9	1095	1.1 ± 0.3
$\mathrm{t\bar{t}H}_+$-enriched	6.8 ± 1.1	1.8 ± 0.3	840.3 ± 248.3	819	1.0 ± 0.3
≥ 6 jets, 3 b-tags					
$\mathrm{t\bar{t}H}_-$-enriched	6.4 ± 1.1	5.1 ± 0.9	815.2 ± 221.9	792	1.0 ± 0.3
$\mathrm{t\bar{t}H}_+$-enriched	10.4 ± 1.8	3.3 ± 0.6	694.2 ± 196.4	594	0.9 ± 0.3
≥ 6 jets, ≥ 4 b-tags, low					
$\mathrm{t\bar{t}H}_-$-enriched	1.1 ± 0.3	1.0 ± 0.3	68.3 ± 26.0	65	1.0 ± 0.4
$\mathrm{t\bar{t}H}_+$-enriched	1.7 ± 0.5	0.6 ± 0.2	57.7 ± 23.1	43	0.7 ± 0.4
4 jets, 4 b-tags, high					
$\mathrm{t\bar{t}H}_-$-enriched	0.2 ± 0.0	0.1 ± 0.0	4.8 ± 1.7	7	1.4 ± 0.4
$\mathrm{t\bar{t}H}_+$-enriched	0.3 ± 0.1	0.1 ± 0.0	5.9 ± 2.2	9	1.5 ± 0.4
5 jets, ≥ 4 b-tags, high					
$\mathrm{t\bar{t}H}_-$-enriched	0.5 ± 0.1	0.3 ± 0.1	8.1 ± 2.9	9	1.1 ± 0.4
$\mathrm{t\bar{t}H}_+$-enriched	0.9 ± 0.3	0.2 ± 0.1	11.4 ± 4.0	12	1.1 ± 0.4
≥ 6 jets, ≥ 4 b-tags, high					
$\mathrm{t\bar{t}H}_-$-enriched	1.0 ± 0.3	0.8 ± 0.2	16.7 ± 6.3	15	0.9 ± 0.4
$\mathrm{t\bar{t}H}_+$-enriched	2.2 ± 0.7	0.7 ± 0.2	24.9 ± 9.6	27	1.1 ± 0.4
Boosted category					
$\mathrm{t\bar{t}H}_-$-enriched	1.0 ± 0.2	0.8 ± 0.1	55.4 ± 11.4	60	1.1 ± 0.2
$\mathrm{t\bar{t}H}_+$-enriched	1.4 ± 0.2	0.4 ± 0.1	61.9 ± 11.1	44	0.7 ± 0.2

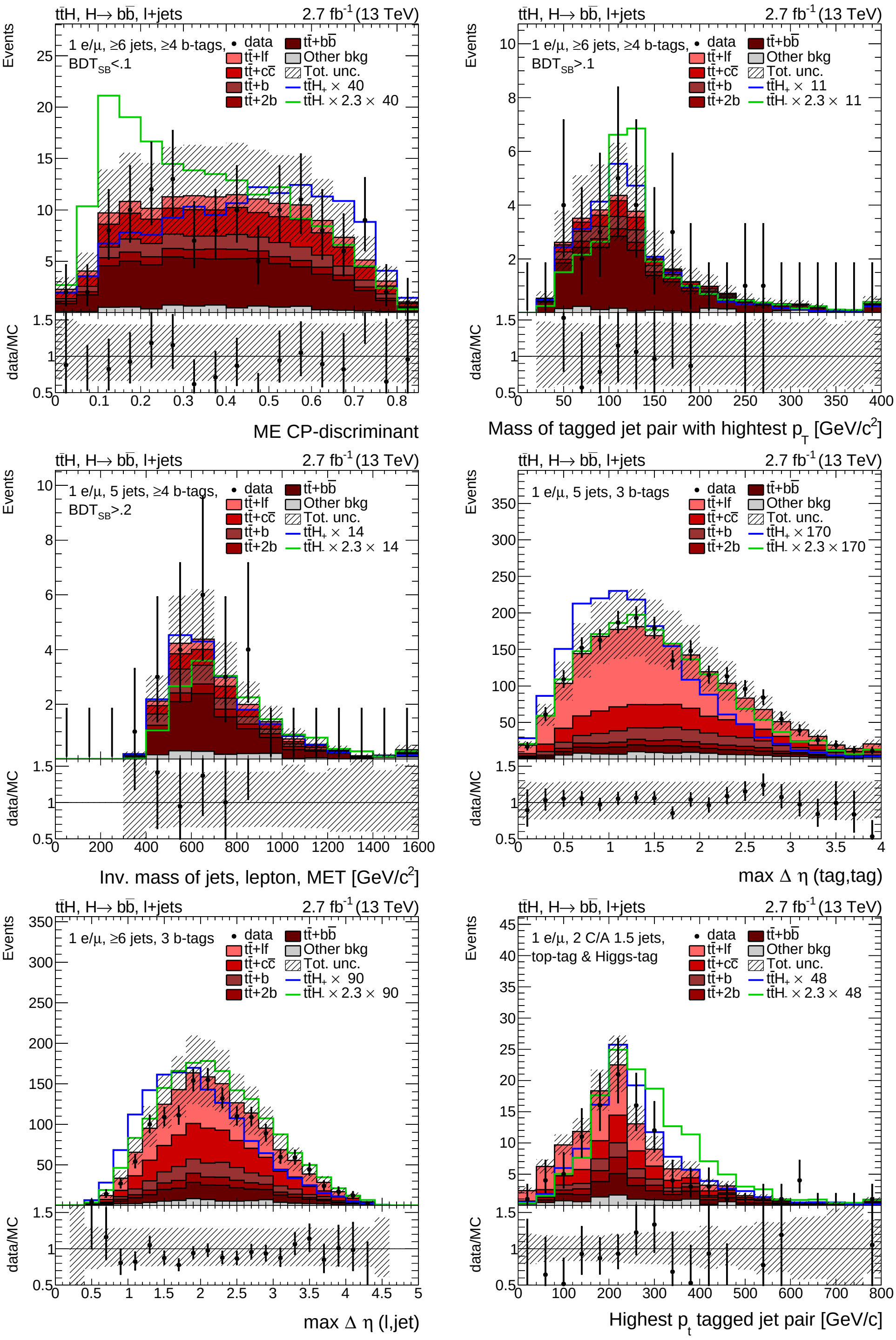


Figure 11.9: Input variables in six of the seven categories in which BDTs sensitive to the coupling structure are trained. The observed distribution of data (black dots) is compared with the expected background predicted by simulation (stacked histograms). The signal with pseudoscalar coupling (green line) is scaled to the same integral as the sum of backgrounds. The expected SM signal is drawn as a blue line and scaled by a factor of 1/2.3 less than the CP odd signal, to account for the difference in total cross section. The most important systematic uncertainties are added in quadrature and shown as hashed error bands. Statistical uncertainties are drawn as black horizontal lines on the data points.

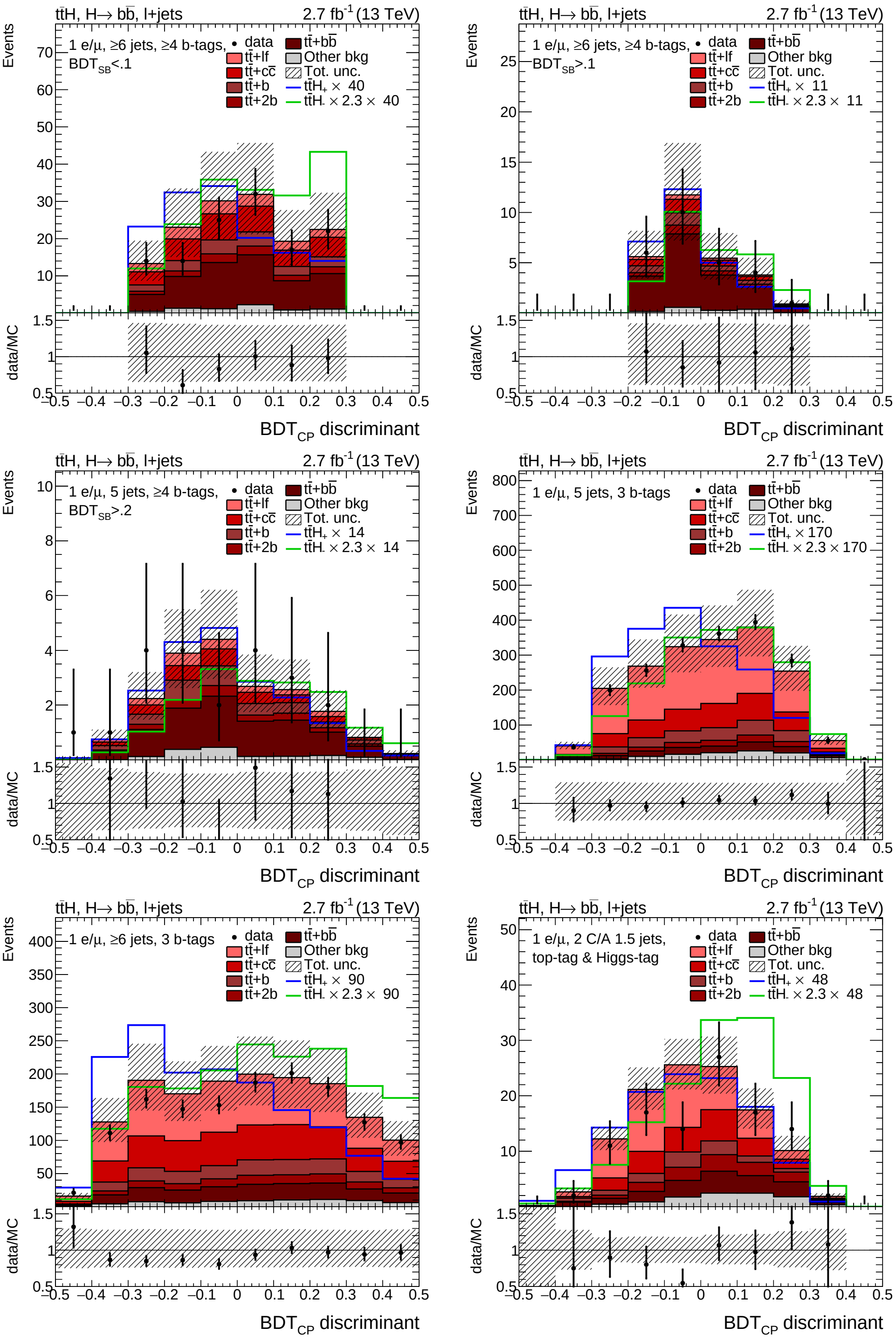


Figure 11.10: Six of the seven BDTs that are sensitive to the tt̄H coupling structure. The observed distribution of data (black dots) is compared with the expected background predicted by simulation (stacked histograms). The signal with pseudoscalar coupling (green line) is scaled to the same integral as the sum of backgrounds. The expected SM signal is drawn as a blue line and scaled by a factor of 1/2.3 less than the CP odd signal, to account for the difference in total cross section. The most important systematic uncertainties are added in quadrature and shown as hashed error bands. Statistical uncertainties are drawn as black horizontal lines on the data points.

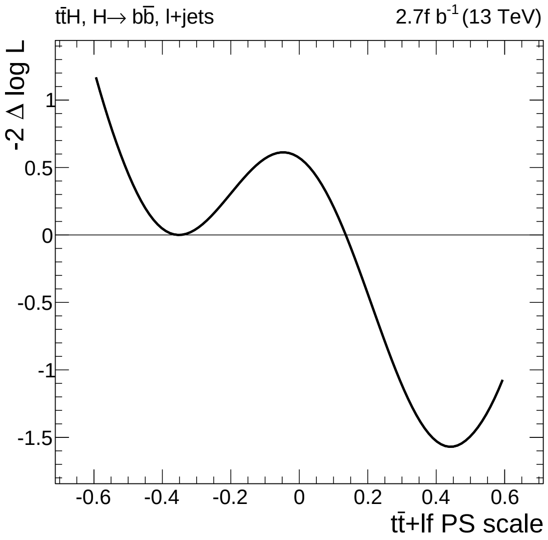


Figure 11.11: Twice the negative log-likelihood relative to the local minimum obtained at a $\mathrm{t\bar{t}}+\mathrm{lf}$ parton shower scale value of -0.35. The likelihood is minimized with respect to all remaining nuisance parameters.

11.3 Statistical interpretation

In this section the statistical interpretation of the modified analysis described in the previous section, which is sensitive to the coupling structure, is discussed.

11.3.1 Maximum likelihood fit

The modified statistical model described in Section 11.1.2 is used and the negative log-likelihood (NLL) is minimized with respect to the two parameters of interest κ_t and $\tilde{\kappa}_\mathrm{t}$ as well as all nuisance parameters. At the minimum, nuisance parameter have values that are close to the fit result of the signal-strength analysis with constraints of similar strength. Interestingly, the NLL function now has at least two minima. This is demonstrated in Fig. 11.11, where a scan of the parameter corresponding to a variation of the parton shower scale for the $\mathrm{t\bar{t}}+\mathrm{lf}$ process is shown, for which the NLL is minimized with respect to the remaining parameters. The changes in the fit model introduced by the variation of the $\mathrm{t\bar{t}}+\mathrm{lf}$ parton scale parameter can be absorbed by moderate shifts of a number of nuisance parameters, making the two points are almost equally likely. As a result, special care is required in a fit of this model to ensure that the best-fit value really corresponds to the global minimum of the likelihood.

Postfit distributions are shown in Fig. 11.12 to Fig. 11.14. The statistical model is flexible enough to describe the data in the large number of categories. Distributions comparing the data to the nominal prefit model are shown in Appendix A.4.

11.3.2 Results

The best-fit corresponds, as for the signal-strength analysis, to $\kappa_\mathrm{t} = 0$ and $\tilde{\kappa}_\mathrm{t} = 0$, i.e. no $\mathrm{t\bar{t}H}$ signal is observed, scalar or pseudoscalar. Confidence regions for the two parameters are shown in Fig. 11.15. On the right-hand side of the figure the observed region is compared with the expectation for different integrated luminosities. The observed confidence

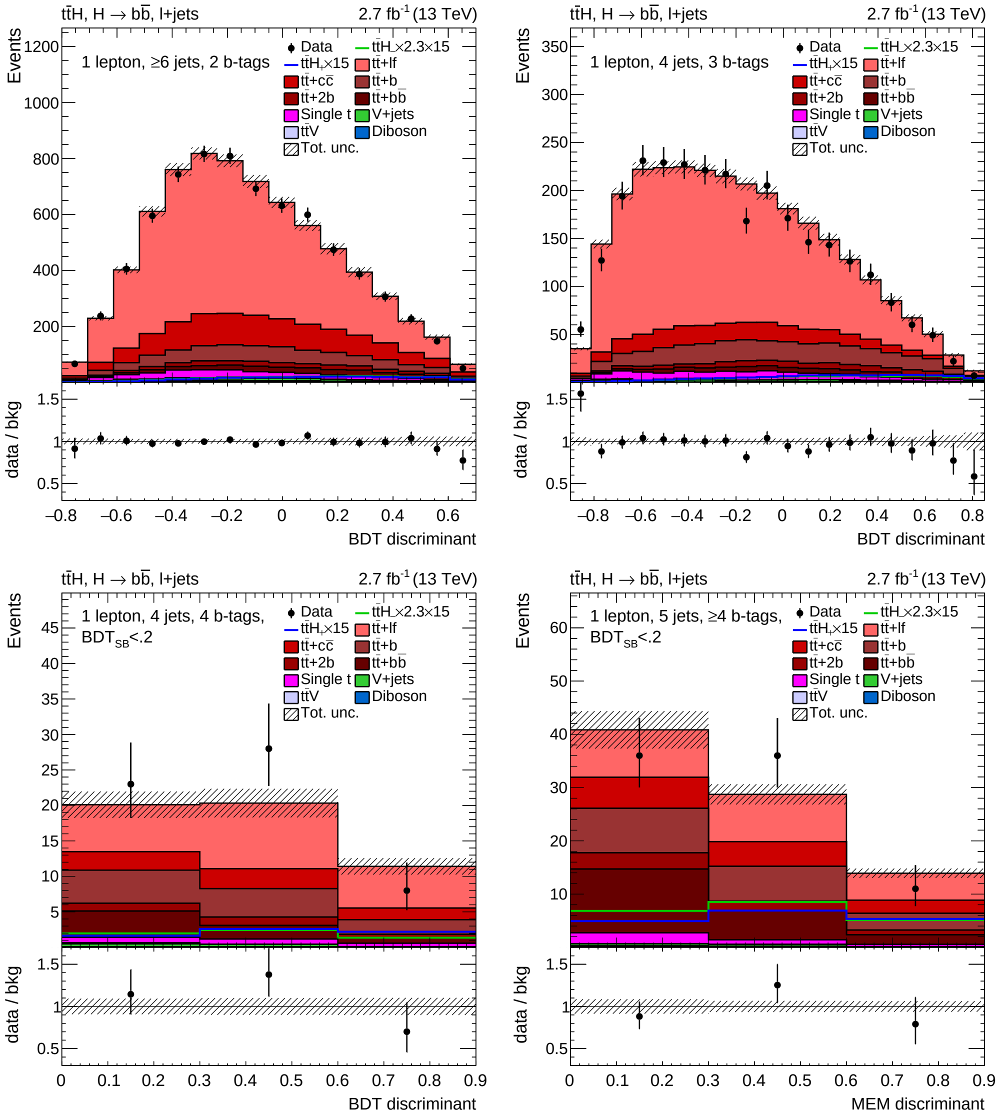


Figure 11.12: Distribution of the data collected in 2015 compared to the model after the fit. On this page the first four background-dominated categories with no scalar-pseudoscalar splitting are shown. On the next two pages the remaining fourteen categories are shown with categories targeting the SM signal on the right and those targeting a CP-odd signal on the left. The expected SM tt̄H signal is drawn as a blue line and scaled by a factor of 15 w.r.t. the expected yield for better visibility. The signal for a pseudoscalar coupling is drawn as a green line and scaled by an additional factor of 2.3 to account for the lower cross section. The postfit systematic uncertainty on rate and shapes of the distributions is evaluated by covariance matrix sampling and displayed as hashed error bands.

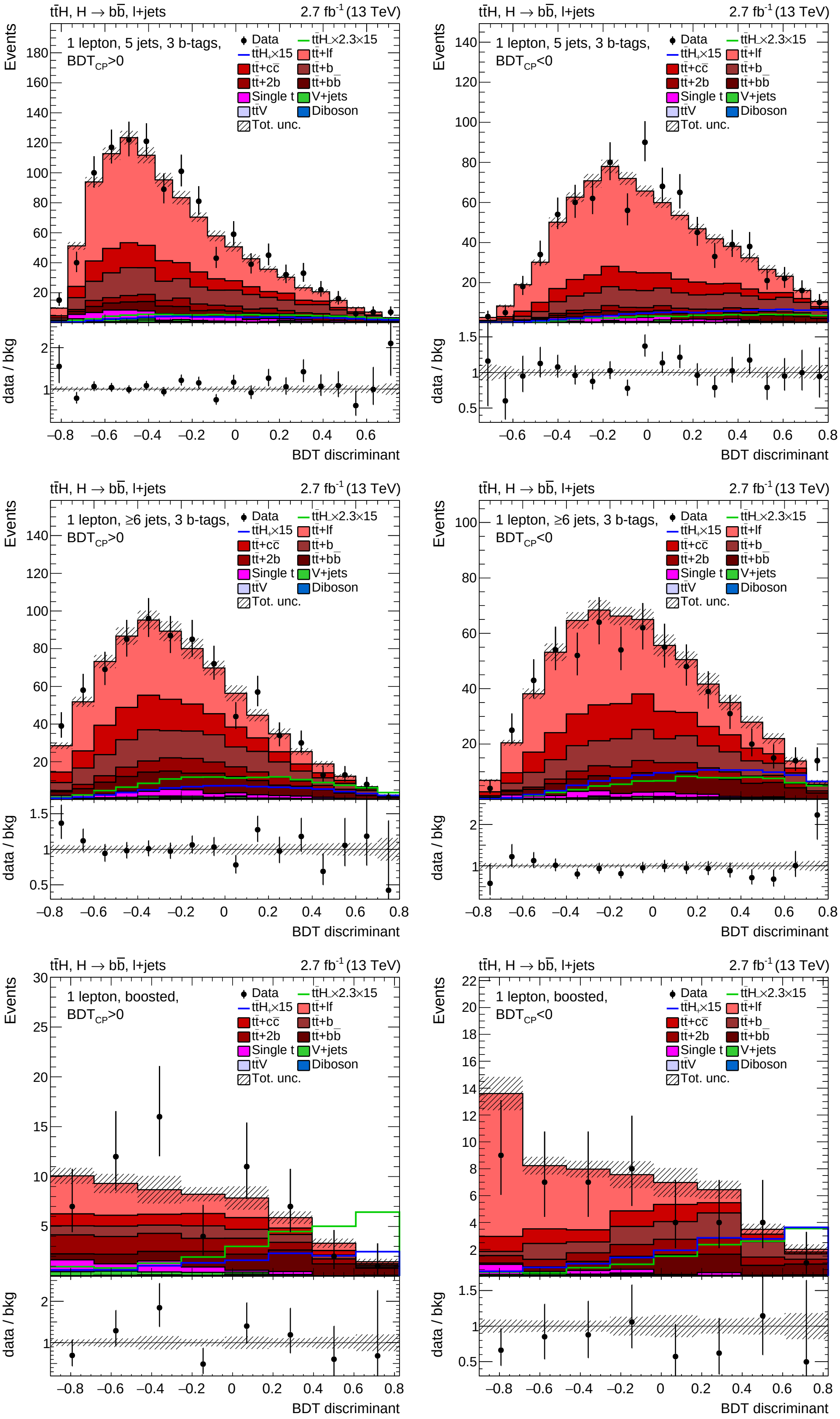


Figure 11.13: See Fig. 11.12 for a description.

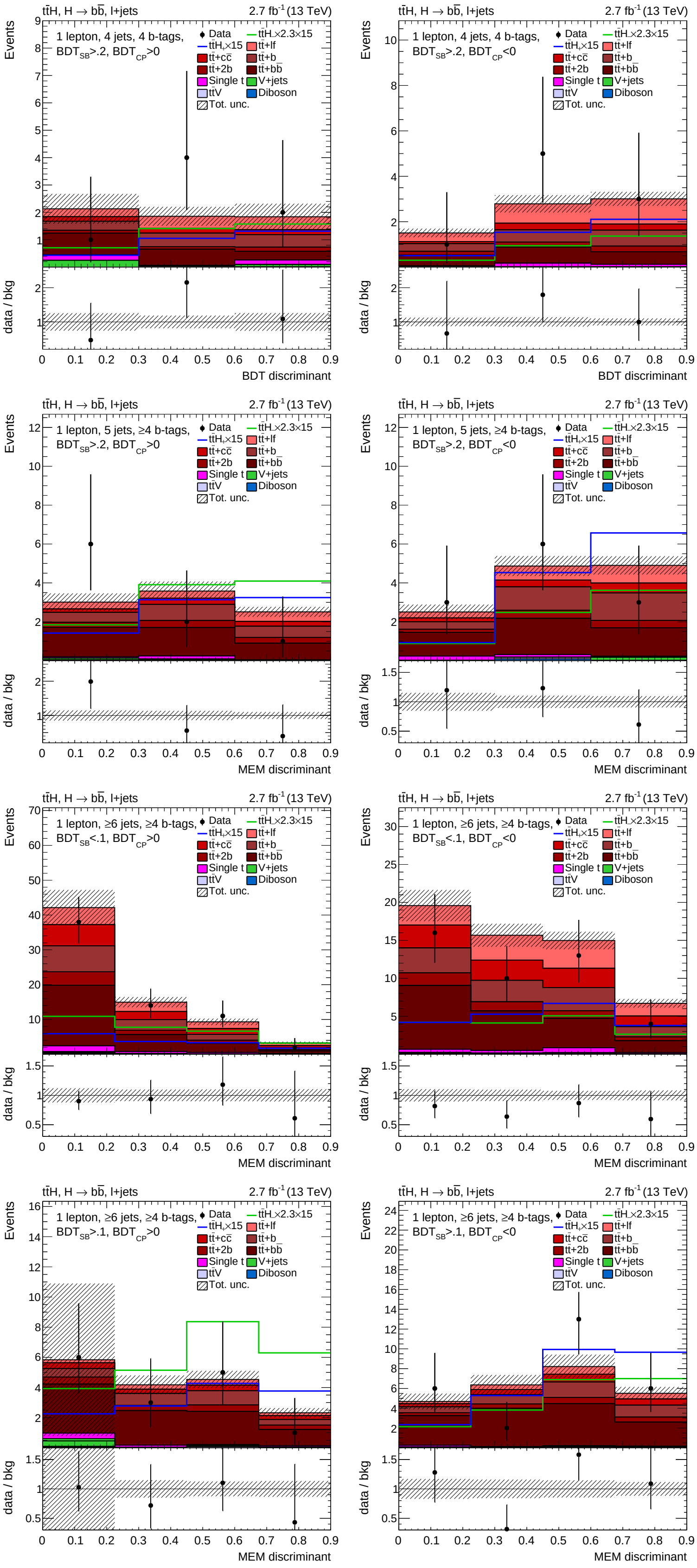


Figure 11.14: See Fig. 11.12 for a description.

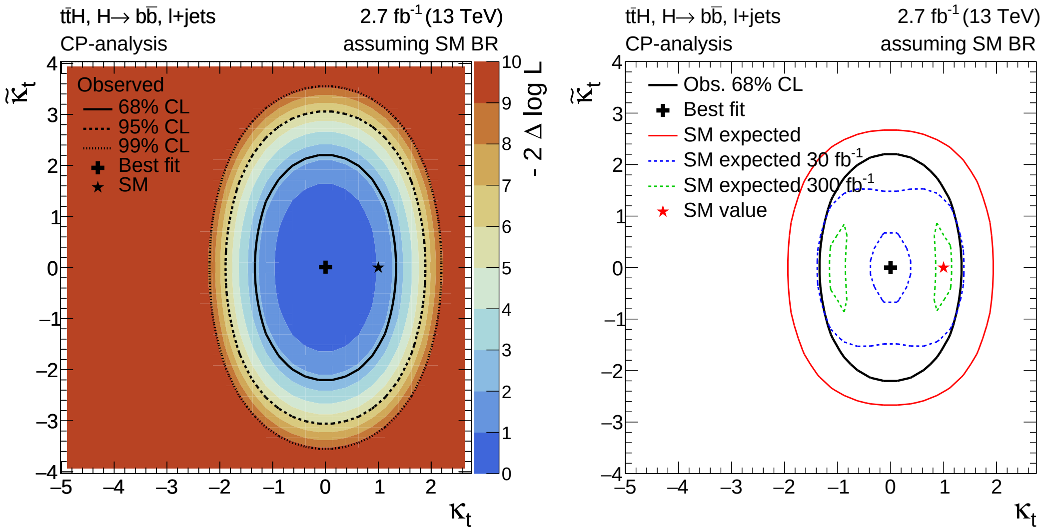


Figure 11.15: Profile likelihood ratio test statistic $q(\kappa_t, \tilde{\kappa}_t)$ obtained from the coupling-structure measurement. The left plot shows the observed result and the right plot the SM expectation. In the left plot, black lines indicate the borders of 68% to 99% confidence level regions for κ_t, and $\tilde{\kappa}_t$, assuming $q(\kappa_t, \tilde{\kappa}_t)$ follows a two-dimensional χ^2 distribution while black crosses indicate the minima of the likelihood and a black star the Standard Model expectation. In the right plot black (colorful) lines indicate the 68% confidence interval calculated from the observation (expectations for different integrated luminosities).

region is smaller than the one expected for an integrated luminosity of 2.7 fb^{-1}, due to a downward fluctuation of the data. With an increased integrated luminosity of 300 fb^{-1} it will be possible to constrain the allowed parameter space closer to $\kappa_t = \pm 1$, $\tilde{\kappa}_t = 0$. This is enabled by the splitting of categories into $\mathrm{t\bar{t}H}_+$- and $\mathrm{t\bar{t}H}_-$-enriched parts and would not be possible in the signal-strength analysis of the previous chapter.

In addition to the two-dimensional fit, a one-dimensional test can be performed by scanning the coupling mixture ζ_t while allowing for arbitrary coupling strengths $\bar{\kappa}_t$. The result of the $q(\zeta_t)$ likelihood scan is shown in Fig. 11.16 on the left-hand side. It is also compared to the expectation for 2.7 fb^{-1} and higher integrated luminosities. Remarkably, no assumptions on the strength of any Higgs-boson couplings are made with this setup and the measurement is indeed directly sensitive to the kinematics of the final state. Furthermore, this measurement is of course also independent of assumption regarding the structure of loop induced process, as it is performed only with $\mathrm{t\bar{t}H(b\bar{b})}$ events. Implicit assumptions are that the top-Higgs and Higgs-bottom coupling strength is non-zero and that the couplings can be described in the framework introduced in Chapter 5. The test statistic is also expressed in terms of the effective pseudoscalar cross section fraction, defined in Equation (5.4) and shown on the left-hand side of Fig. 11.16. With the current data, no statement about a possible pseudoscalar admixture can be made. However, with an integrated luminosity of 30 fb^{-1} larger pseudoscalar admixtures to the coupling can be excluded:

$$\tilde{f}_t \lesssim 0.7 \text{ (SM expected for 30 fb}^{-1}\text{, at 68\% CL).}$$

With 300 fb^{-1}, admixtures of the order 40% (65%) can be excluded at the 2σ (3σ) level.

The most important systematic uncertainties for the ζ_t measurement are determined with the technique described in Section 10.2.3. Since no sensitivity is expected with the

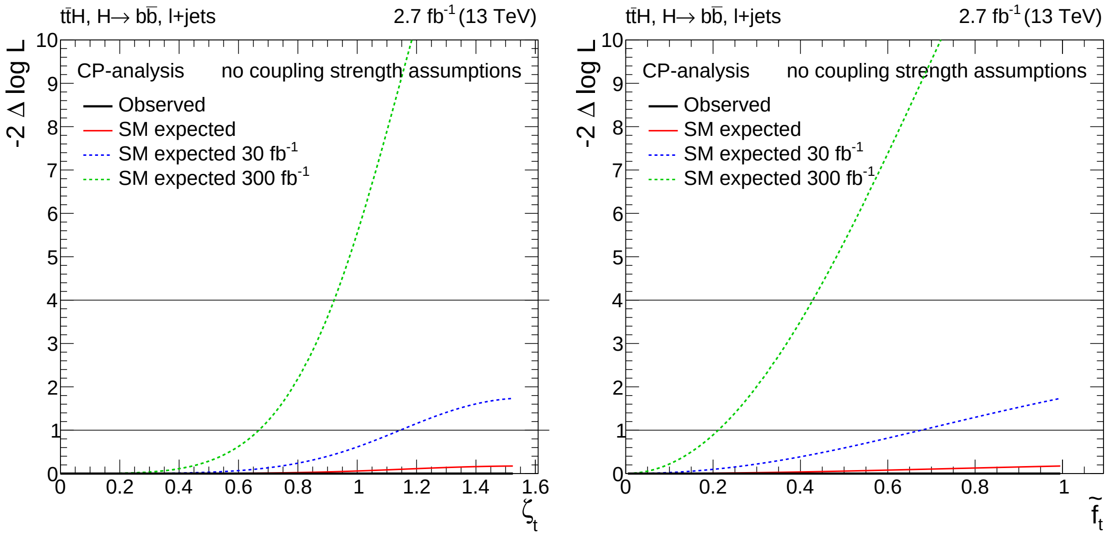


Figure 11.16: Profile likelihood scan of the scalar-pseudoscalar mixing angle ζ_t (left) and, equivalently, of the effective pseudoscalar cross section admixture $\tilde{f}_\mathrm{t}$, obtained from the coupling-structure analysis. No assumptions regarding the $\mathrm{t\bar{t}H}$ signal strength are made, the corresponding parameter is profiled. The intersection with the horizontal lines at one (four) indicated the borders of the 1σ (2σ) confidence intervals.

current dataset, the study is performed with simulated data expected for an integrated luminosity of 30 fb^{-1}. In contrast to the signal-strength analysis, the $\mathrm{t\bar{t}b\bar{b}}$ rate is not the dominant systematic. Instead, a large number of nuisance parameter have a very similar impact on the measurement. This can be understood by the fact that the category-splitting that introduces the sensitivity to the coupling structure is performed around the median of the discriminant distribution. The measurement is thus less susceptible to small effects in extreme regions of the phase space and more to effects that affect the median of the CP-discriminant distribution. Main uncertainties include uncertainties associated with b-tagging, the $\mathrm{t\bar{t}b}$ rate uncertainty, and the JES uncertainty.

The expected sensitivity to a pseudoscalar admixture $\tilde{f}_\mathrm{t}$ of this analysis is higher than the one expected from feasibility studies in the $\mathrm{t\bar{t}H}$, $\mathrm{H} \to \gamma\gamma$ and $\mathrm{t\bar{t}H}$, $\mathrm{H} \to \mathrm{ZZ}$ channels in [142]. Nonetheless, performing similar analyses in the $\gamma\gamma$ or lepton decay channels, that have been proven to be of similar sensitivity in the search for $\mathrm{t\bar{t}H}$ production, can be expected to improve the coupling-structure constraints that can be obtained in the $\mathrm{t\bar{t}H(b\bar{b})}$ channel. A combination of these analyses could then allow to exclude a pseudoscalar coupling, maybe already with the dataset available in 2016.

12 Other $\mathrm{t\bar{t}H}$ analyses at 13 TeV

In this chapter, other $\mathrm{t\bar{t}H}$ analysis performed by ATLAS and CMS with data recorded in 2015 and 2016 are summarized. The measurements that have already been combined with the results presented in this thesis are outlined in Section 12.1 and in Section 12.2 the remaining $\mathrm{t\bar{t}H}$ measurements at 13 TeV are discussed.

12.1 Combination with other $\mathrm{t\bar{t}H}$ analyses

12.1.1 Combination with the dilepton channel

The measurements of the $\mathrm{t\bar{t}H}$ signal strength performed in Chapter 10 can be combined with other measurements at CMS. The most natural combination is the one with the dilepton $\mathrm{t\bar{t}H(b\bar{b})}$ analysis at 13 TeV, as it targets the same Higgs production and decay modes. It is included in the same publication [174] as the results presented in this thesis. An analysis of the dilepton channel alone yields an upper limit at 95% CL of $\mu_{\mathrm{t\bar{t}H}} < 5.2$ (7.7 expected), and a best fit value of $-4.7^{+3.7}_{-3.8}$.

For the combination, a simultaneous fit of both analysis channels is performed. Correlated systematic uncertainties are parameterized by the same nuisance parameters for both channels. Most uncertainties are treated as fully correlated, the exception being uncertainties related to lepton-trigger and lepton-identification efficiency, which are modeled by separate nuisance parameters in the two channels. Both analyses are susceptible to the same main systematic effects: b-tagging uncertainties and potential mismodeling of the $\mathrm{t\bar{t}}$ + hf background. The limit calculated from a combined fit of both channels is $\mu_{\mathrm{t\bar{t}H}} < 2.6$ (with 3.6 expected) at 95% CL, using the asymptotic approximation also employed in the individual analyses. The best fit value of the combined analysis is $\mu_{\mathrm{t\bar{t}H}} = -2.0 \pm 1.8$, which is 1.7 standard deviations below the SM expectation. The individual and combined results are summarized in Table 12.1 and Fig. 12.1.

Because of the correlated systematic uncertainties, the combined result is not completely intuitive. A naive combination of the dilepton and lepton+jets results, i.e. assuming uncorrelated Gaussian uncertainties, would yield $\mu^{\mathrm{naive}}_{\mathrm{t\bar{t}H}} = -1.4 \pm 1.8$ instead of $\mu_{\mathrm{t\bar{t}H}} = -2.0 \pm 1.8$. Even more striking is the fact that the combined limit is, like the dilepton limit, one standard deviation below the background-only expectation while the lepton+jets

Table 12.1: Best-fit values and upper limits of the $\mathrm{t\bar{t}H}$ signal strength modifier $\mu_{\mathrm{t\bar{t}H}}$. Upper limits are calculated with the asymptotic method and given at 95% CL.

Channel	Best-fit of $\mu_{\mathrm{t\bar{t}H}}$	Upper limit at 95% CL (observed)	(68% expected)
Lepton+jets	$-0.4^{+2.1}_{-2.1}$	4.0	$4.1^{+1.8}_{-1.2}$
Dilepton	$-4.7^{+3.7}_{-3.8}$	5.2	$7.7^{+3.6}_{-2.3}$
Combined	$-2.0^{+1.8}_{-1.8}$	2.6	$3.6^{+1.6}_{-1.1}$

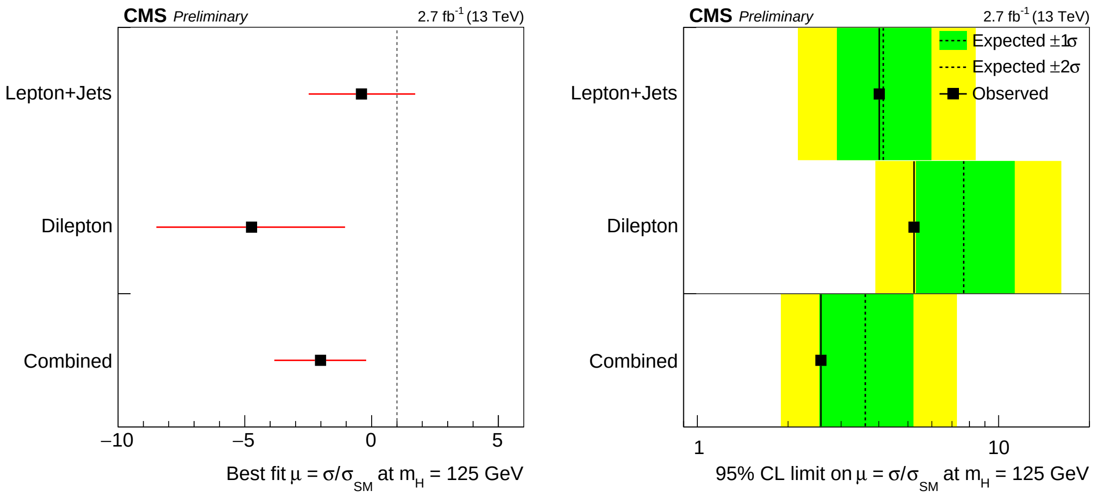


Figure 12.1: Visualization of the best-fit values (left) and upper limits (right) of the $\mathrm{t\bar{t}H}$ signal strength modifier $\mu_{\mathrm{t\bar{t}H}}$. Upper limits are calculated with the asymptotic method and given at 95% CL. Both also displayed in [174].

channel, the channel with the higher precision, is close to the expectation. The origin of this behavior is that multiple uncertainties can have a similar effect on event yields and shapes of distributions. Compared to the lepton+jets fit, the combined fit prefers a different nuisance parameter configuration that describes the data almost equally well, but leads to a small deficit of events in the most signal-enriched regions. The most extreme example of such a nuisance parameter is the one associated with the $\mathrm{t\bar{t}b}$ rate uncertainty, which is constrained to 1.6 ± 0.5 in the combined fit but only to 0.9 ± 0.9 in the lepton+jets fit. The effect this introduces in the background description in the lepton+jets channel is canceled by smaller shifts in several other nuisance parameters. The most important one is a reduction of the c-jet mistag rate in the combined fit (modeled by the linear c-mistag uncertainty listed in Table 10.1). Effectively this means that $\mathrm{t\bar{t}c\bar{c}}$ or $\mathrm{t\bar{t}+lf}$ (with $\mathrm{W} \to \mathrm{cs}$) events with a c-jet mistag are exchanged with $\mathrm{t\bar{t}b}$ events. This is demonstrated in Fig. 12.2. The left plot is the postfit model of the lepton+jets fit, the right plot the one of the combined fit. The $\mathrm{t\bar{t}b}$ background, marked in yellow, increases from left to right. At the same time the light flavor background is reduced so that both models describe the data well. The combination of channels thus leads to a reduction of the $\mathrm{t\bar{t}+lf}$ background and sees an increase of the more signal-like $\mathrm{t\bar{t}b}$ and $\mathrm{t\bar{t}b\bar{b}}$ background. In turn, this leads to a smaller signal strength.

The fact that the change in signal strength is related to the different systematic uncertainties can also be demonstrated by fitting lepton+jets and dilepton channel simultaneously but introducing separate signal strength parameters in each channel. The best fit values of these are then $\mu_{\mathrm{DL}} = -2.0^{+3.2}_{-3.0}$ and $\mu_{\mathrm{LJ}} = -2.0 \pm 1.8$. Both signal strength parameters are in perfect agreement with the combined result, indicating that the changed background prediction is indeed responsible for the counter-intuitive combination result.

12.1.2 Combination with the diphoton and multilepton analyses

The combined $\mathrm{t\bar{t}H(b\bar{b})}$ result is further combined with two additional channels. One is the search for $\mathrm{t\bar{t}H}$ in the multilepton channel [246], i.e. $\mathrm{t\bar{t}H}$ events in which both $\mathrm{t\bar{t}}$ and the

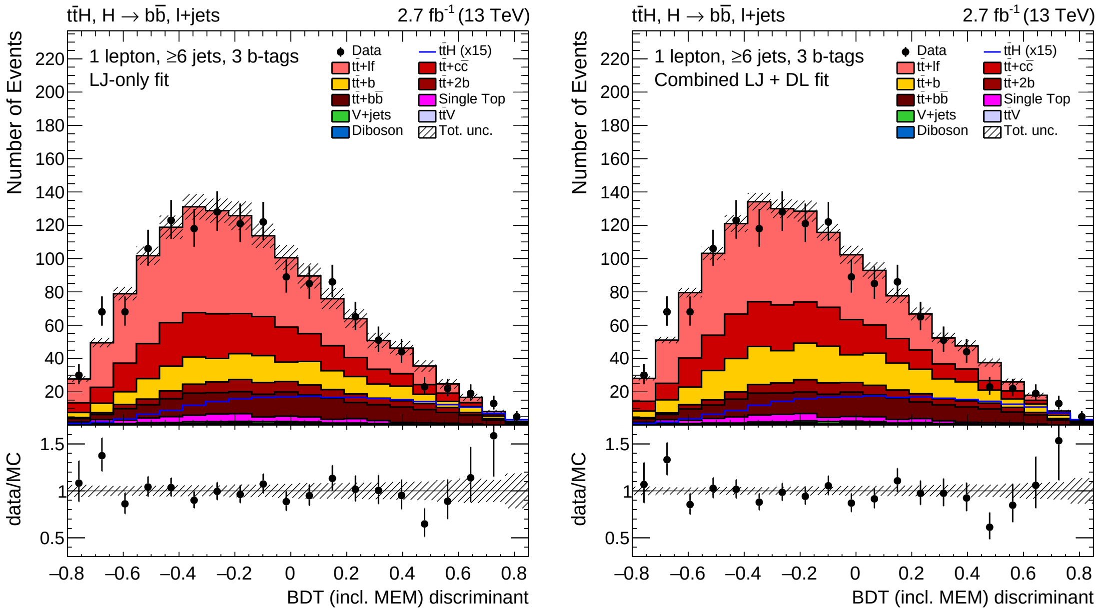


Figure 12.2: Interpretation of the data observed in the ≥ 6 jets, 3 b-tags category by the fit of the lepton+jets analysis (left) and by the combined fit with the dilepton categories (right). The $t\bar{t}b$ background is marked in yellow.

Higgs boson decay (semi-)leptonically, leading to final states containing at least two leptons (electrons or muons) of the same charge. The second, the search for $t\bar{t}H$ in the $H(\gamma\gamma)$ channel, is documented in [247]. In both channels, events recorded in 2015 are analyzed. Of the three analysis, the multilepton analysis is most sensitive to a $t\bar{t}H$ signal, with an expected upper limit of $\mu_{t\bar{t}H} \lesssim 3$ while the diphoton analysis only expects to set a limit of $\mu_{t\bar{t}H} \lesssim 7$. The $t\bar{t}H$ search in the $\gamma\gamma$ channel is still dominated by statistical uncertainties, with only one $t\bar{t}H$ event expected. In the multilepton channel, systematic uncertainties related to the estimation of non-prompt lepton background are also important. Only the $H(\gamma\gamma)$ analysis observes a small excess, but is also compatible with no signal. The result of a combined fit of the three analysis, fixing the branching ratios to their SM-values is (also shown in Fig. 12.3)

$$\mu_{t\bar{t}H}^{\text{CMS 2015}} = 0.2^{+0.9}_{-0.8}$$

and the 95% CL upper limit on the $t\bar{t}H$ production cross section, normalized to the SM is $\mu_{t\bar{t}H} < 2.1$ (1.9 expected). The large excess of Run 1 can thus not be confirmed.

12.2 Most recent $t\bar{t}H$ results

While this thesis was written, first results results on $t\bar{t}H$ using data collected at the LHC in 2016 have been published. At CMS, the multilepton and the $\gamma\gamma$ analysis have been updated with events recorded in the first half of 2016, corresponding to 12.9 fb^{-1} [249,250]. In the multilepton channel a signal strength of $\mu_{t\bar{t}H} = 2.0^{+0.8}_{-0.7}$ was measured by combining the 2015 and 2016 datasets, while the $\gamma\gamma$ channel reports $\mu_{t\bar{t}H} = 1.9^{+1.5}_{-1.2}$ from the data recorded in 2016.

ATLAS also presented first results on $t\bar{t}H$, using events collected at 13 TeV and corresponding to up to 13.3 fb^{-1}. The measurement in the $b\bar{b}$ decay channel [251] targets

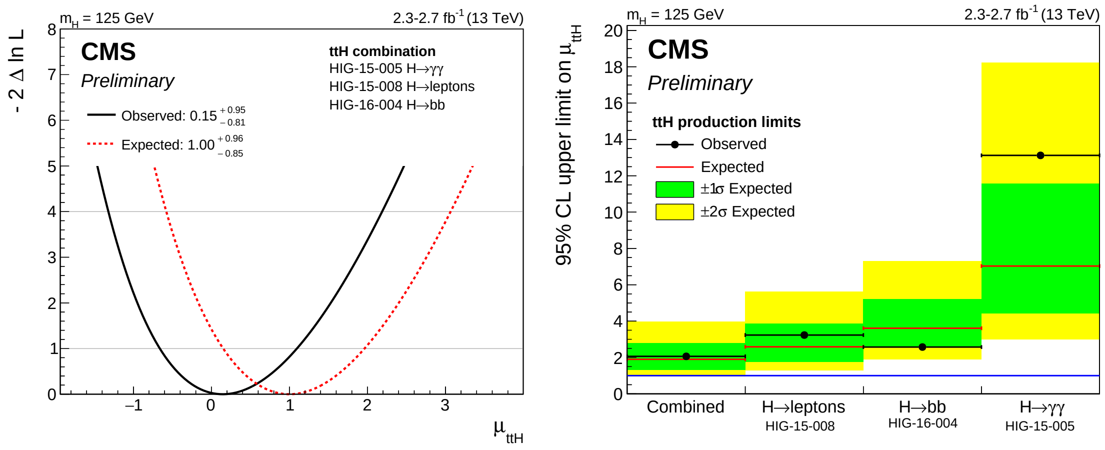


Figure 12.3: Combination of the CMS $\mathrm{t\bar{t}H}$ analyses using data recorded at the LHC at a center-of-mass energy of 13 TeV in 2015 [248]. On the left-hand side a scan of the profile likelihood for the $\mathrm{t\bar{t}H}$ signal strength is shown. The observed (expected) result is drawn as a solid black (dashed red) line, the intersections with the horizontal lines indicate the borders of the 1σ and 2σ confidence intervals. On the right-hand side the combined upper limit at 95% CL, using the asymptotic method, is shown for the three individual analyses and their combination.

dilepton and lepton+jets $\mathrm{t\bar{t}}$ decays. The lepton+jets analysis is comparable to the corresponding CMS analysis, which is presented in Chapter 10 of this thesis. For the ATLAS analysis, events are also categorized according to jet and b-tag multiplicity. The event selection includes, in addition to the seven categories of the CMS analysis, a 4 jets, 2 b-tags and a 5 jets, 2 b-tags category. However, the signal is only searched for in three of the nine categories. In the remaining categories, observables that provide little signal-background separation are analyzed to constrain systematic uncertainties. In contrast to CMS, ATLAS does not utilize the Matrix Element Method or a search in the boosted regime. Instead, a two-stage BDT approach is employed to identify the $\mathrm{t\bar{t}H}$ signal. A first BDT is used to reconstruct the $\mathrm{t\bar{t}H}$ system by identifying jets, lepton, and $\not{E}_{\mathrm{T}}$ with the $\mathrm{t\bar{t}H}$ decay products. Observables constructed this way, e.g. the Higgs-boson candidate mass, are combined with additional kinematic observables in a final BDT that separates the signal from the backgrounds. The background simulation of the ATLAS analysis is more involved than the one in the CMS analysis. For the $\mathrm{t\bar{t}}$ background, a two-dimensional reweighting is performed to obtain better agreement with the NNLO QCD predictions for the p_{T} of the top quarks and the $\mathrm{t\bar{t}}$ system p_{T} [252, 253]. The $\mathrm{t\bar{t}b\bar{b}}$ background is corrected to match a prediction of Sherpa and OpenLoops at NLO QCD [210,211,230]. The $\mathrm{t\bar{t}}$ sample is split into fewer $\mathrm{t\bar{t}}$ + hf classes but $\mathrm{t\bar{t}}$-modeling uncertainties are assigned based on comparisons with a number of alternative event generators and generator configurations. An upper limit of $\mu_{\mathrm{t\bar{t}H}} < 3.6$ at 95% CL, with $\mu_{\mathrm{t\bar{t}H}} < 2.2$ expected, can be set in the lepton+jets channel. The higher sensitivity compared to CMS can be attributed to the larger dataset, corresponding to 13.2 fb^{-1} instead of 2.7 fb^{-1} of integrated luminosity. With a dataset of the same size, the analysis presented in this theses is of similar sensitivity (see 10.2.6).

The ATLAS $\mathrm{t\bar{t}H}$ measurement in the $\mathrm{b\bar{b}}$ channel is combined with a measurement in multilepton [254] and $\gamma\gamma$ [255] decay channels in [256]. The three ATLAS analyses are of similar precision. The $\gamma\gamma$ channel measures 1σ below the Standard Model expectation while both $\mathrm{b\bar{b}}$ and multilepton are more than one standard deviation above it. The

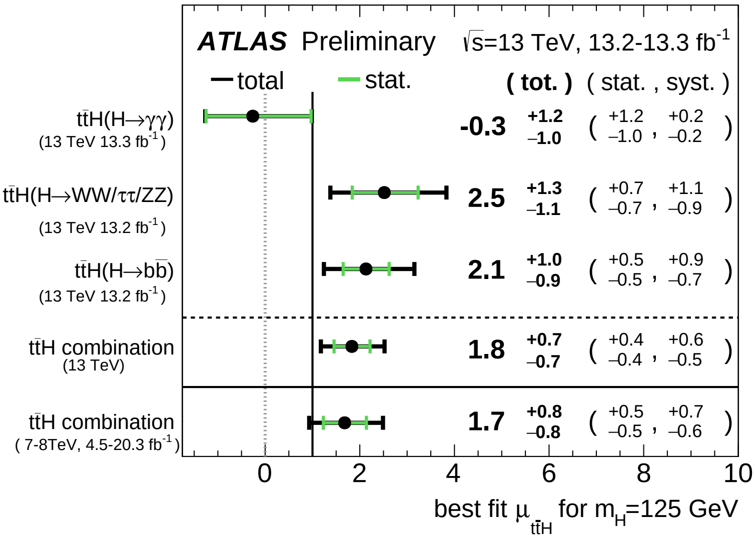


Figure 12.4: Combination of the ATLAS $t\bar{t}H$ analyses using data recorded at the LHC at a center-of-mass energy of 13 TeV in 2015 and 2016 [256]. The signal strength is shown for the combined and the three individual analyses. Statistical (total) uncertainties are indicated as green (black) horizontal line.

combination of the three analyses yields

$$\mu_{t\bar{t}H}^{\text{ATLAS 13 TeV}} = 1.8 \pm 0.7$$

and is, together with the separate results, displayed in Fig. 12.4.

The uncertainties of the CMS and ATLAS analyses are not expected to be strongly correlated among each other and a naive combination using a Gaussian approximation can be expected to yield reasonable results. Combining the ATLAS $t\bar{t}H$ combination [256] and the most recent CMS multilepton [249], $\gamma\gamma$ [250], and $b\bar{b}$ [174] $t\bar{t}H$ analyses this way, symmetrizing the given uncertainties, would lead to the unofficial result of

$$\mu_{t\bar{t}H}^{\text{13 TeV}} \approx 1.4 \pm 0.5.$$

This is compatible with the Standard Model expectation and less so with the background-only hypothesis. The combination would already be of higher precision than the official Run 1 LHC combination of $\mu_{t\bar{t}H}^{\text{run 1}} = 2.3^{+0.7}_{-0.6}$ [8].

Summary and outlook

In Run 1 of the LHC in 2011 and 2012, a particle compatible with the Standard Model Higgs boson has been discovered. Precise measurements of its properties allow testing the electroweak theory and searching for physics beyond the Standard Model. In this thesis, the top-Higgs coupling has been constrained using different model assumptions and data measured with the CMS experiment in LHC Run 1 and 2.

First, constraints on anomalous top-Higgs couplings have been derived by re-analyzing CMS Run 1 measurements. Hereby it is assumed that no particles beyond the Standard Model contribute to the production and decay processes of the Higgs boson. Under this assumption a measurement becomes feasible because the top-Higgs coupling affects gluon-fusion Higgs-boson production, $\mathrm{t\bar{t}H}$ production, and Higgs-boson decays into two photons in a different manner. The constraints derived for the two parameters describing an anomalous top-Higgs coupling, κ_t and $\tilde{\kappa}_\mathrm{t}$, are shown in Fig. 13.1. A purely pseudoscalar top-Higgs coupling can be excluded, the strength of the pseudoscalar coupling is observed to be

$$\tilde{\kappa}_\mathrm{t} = 0.0 \pm 0.25,$$

and the scalar-pseudoscalar mixing angle is measured as

$$\zeta_\mathrm{t} = 0.0 \pm 0.4.$$

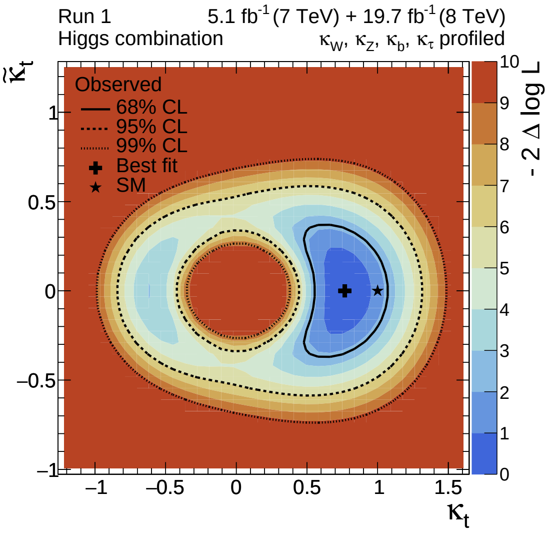


Figure 13.1: Constraints on anomalous top-Higgs couplings, parameterized by a scalar and a pseudoscalar coupling strength modifier κ_t and $\tilde{\kappa}_\mathrm{t}$, and calculated by re-analyzing CMS Higgs-boson measurements using data recorded at the LHC in 2011 and 2012. The best fit value is at the black cross, borders of 68%/95%/99% confidence intervals are drawn as solid/dashed/dotted lines. The value of the profile likelihood ratio test statistic used for this analysis is indicated by the color coding and the Standard Model coupling value represented by a black star.

This analysis is the first of its kind at CMS and similar to the ones presented in [140, 141, 146–148]. With respect to these analyses, which have performed outside of the experimental collaborations at the LHC, several improvements have been made. Cross sections have been calculated with higher precision and the re-analysis of results is much more thorough, especially the treatment of systematic uncertainties. Most importantly, for the presented analysis it was not necessary to assume that the strength of any Higgs-boson coupling corresponds to the Standard Model expectation.

A direct measurement of the top-Higgs coupling that does not rely on assumptions about the structure of loop-induced processes is possible in $\mathrm{t\bar{t}H}$ production. Discovering this process is one of the main goals of LHC Run 2. For this thesis, major contributions to the CMS analysis in the $\mathrm{H} \to \mathrm{b\bar{b}}$ channel [174] have been made. The analysis is the sole successor of two CMS analysis performed at CMS in Run 1 [22, 23], uniting Matrix Element Method and machine learning approaches. The part of the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis presented here investigates final states containing one lepton. The main results, displayed in Fig. 13.2, are an upper limit on the $\mathrm{t\bar{t}H}$ production cross section normalized to the Standard Model value of

$$\mu_{\mathrm{t\bar{t}H}} < 4.0 \text{ (4.1 expected)}$$

at 95% confidence level and signal strength measurement of

$$\mu_{\mathrm{t\bar{t}H}} = -0.4 \pm 2.1.$$

The measurement has also been combined with $\mathrm{t\bar{t}H}$ searches in other decay channels,

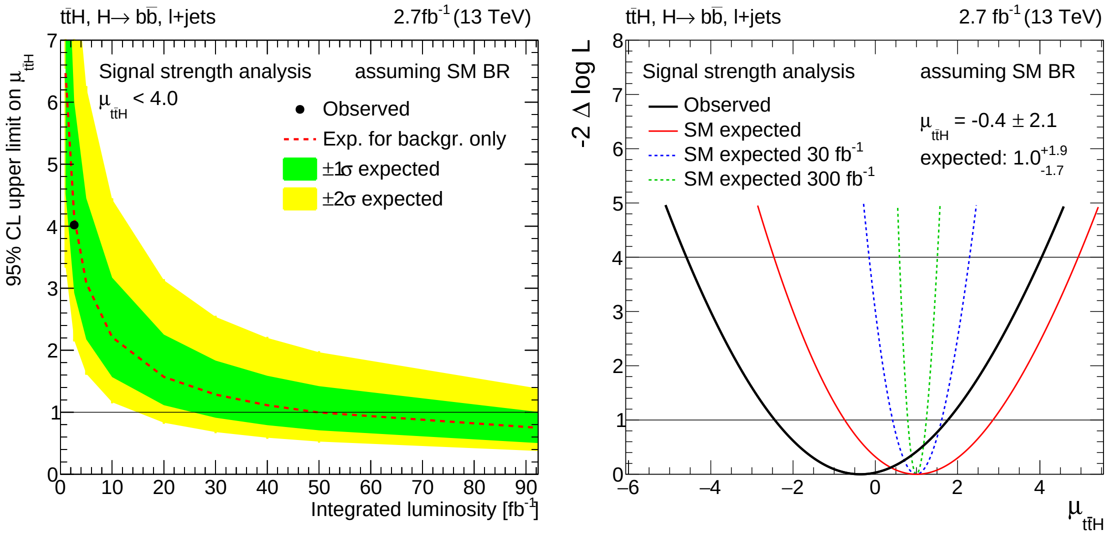


Figure 13.2: Measurement and expected results as a function of luminosity for the Standard Model Higgs-boson search in the lepton+jets channel. On the left-hand side the 95% CL upper limit on the $\mathrm{t\bar{t}H}$ production cross section normalized to the Standard Model expectation is shown. The red line corresponds to the expected value in the absence of the process and is extrapolated to higher luminosities. Green and yellow bands are 68% and 95% confidence level intervals for the expected value. A black dot indicates the observed limit at 2.7 fb^{-1}. On the right-hand side the likelihood profile of the signal strength parameter is shown as a black line. Intersections with the two horizontal lines indicate the borders of the 1σ and 2σ confidence intervals. In red/blue/green the expectation in the presence of $\mathrm{t\bar{t}H}$ for 2.7/30/300 fb^{-1} is shown.

targeting Higgs-boson decays into two photons and $\mathrm{t\bar{t}H}$ decays into two and more leptons [248], yielding $\mu_{\mathrm{t\bar{t}H}} = 0.2^{+0.9}_{-0.8}$. Because of the limited number of collisions recorded in 2015, neither this analysis nor the CMS combination could discover or exclude $\mathrm{t\bar{t}H}$ production. The presented analysis is however a solid basis for further $\mathrm{t\bar{t}H(b\bar{b})}$ measurements in Run 2, containing, in addition to the combination of MEM and BDT discriminants, many new features, e.g. the automated optimization of the multivariate analysis, improved modeling of systematic uncertainties, and a search in the boosted regime as proposed in [220].

$\mathrm{t\bar{t}H}$ events can then also be used to investigate the properties of the top-Higgs coupling. So far, experimental analyses are optimized to discover a Standard Model $\mathrm{t\bar{t}H}$ signal but are unable to exclude an anomalous top-Higgs coupling. Nonetheless, the coupling structure is imprinted in the distribution of kinematic properties of the three final state particles. In this thesis, it was demonstrated for the example of the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis that analyses searching for $\mathrm{t\bar{t}H}$ can be modified to be also sensitive to these effects. In the absence of a signal it is not yet possible to make a statement about an anomalous pseudoscalar admixture to the coupling. This will however change as soon as larger datasets are available. It is for example projected that a first measurement of $\tilde{f}_{\mathrm{t}}$, the effective pseudoscalar cross section fraction, will already be possible with the data collected in 2016:

$$\tilde{f}_{\mathrm{t}} \lesssim 0.7 \ (\text{at } 68\% \text{ CL, expected for } 30\ \ \mathrm{fb}^{-1})$$

This is also illustrated in Fig. 13.3.

All three measurements will benefit from the additional data recorded at the LHC in 2016. It is planned to repeat the indirect coupling measurement as part of the first CMS coupling combination at 13 TeV. For a more exact inclusion of measurements of $\mathrm{t\bar{t}H}$ in this combination, it would be beneficial to evaluate those analyses also using an alternative signal model with an anomalous top-Higgs coupling. The search for $\mathrm{t\bar{t}H}$ is already being updated with new data, and CMS published results targeting Higgs to lepton [249] and photon [250] decays, which will soon be followed by an update of the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis.

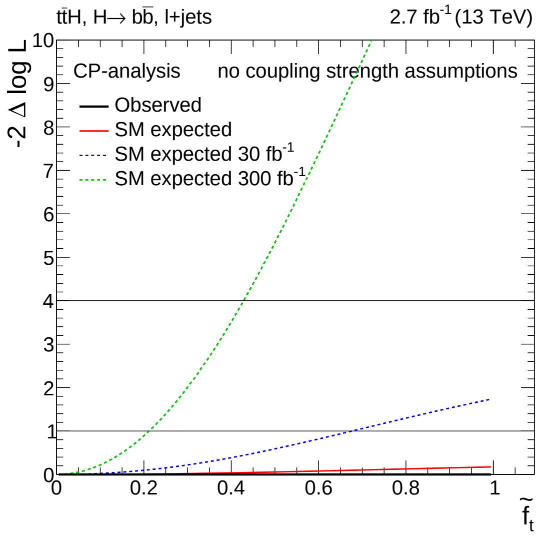


Figure 13.3: Likelihood profile of the effective pseudoscalar cross section fraction $\tilde{f}_{\mathrm{t}}$ in $\mathrm{t\bar{t}H}$, from an analysis of the lepton+jets channel. The observed profile is shown as a black line and red/blue/green the expectation for 2.7/30/300 fb^{-1} is shown. Intersections with the two horizontal lines indicate the borders of the 1σ and 2σ confidence intervals.

The main challenges for this analysis are to adapt to the changing experimental conditions and to improve the background model, so that it describes the data with higher accuracy and in order to incorporate more subtle effects of systematic uncertainties. Once first evidence for $\mathrm{t\bar{t}H}$ is observed, the process should be studied further to gain more insight into the coupling structure. To improve the sensitivity of the coupling structure analysis, it has to be performed simultaneously in multiple Higgs boson decay channels.

Thus exciting times lie ahead in which it will be possible to increase our knowledge of the top-Higgs coupling, which then might provide more insights into the nature of electroweak symmetry breaking.

Bibliography

[1] Aad, G., et al. "The ATLAS Experiment at the CERN Large Hadron Collider". *JINST*, 3:S08003 (2008).

[2] Chatrchyan, S., et al. "The CMS experiment at the CERN LHC". *JINST*, 3:S08004 (2008).

[3] Evans, L. and Bryant, P. "LHC Machine". *JINST*, 3:S08001 (2008).

[4] Aad, G., et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys.Lett.*, B716:1–29 (2012).

[5] Chatrchyan, S., et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Phys.Lett.*, B716:30–61 (2012).

[6] Chatrchyan, S., et al. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". *JHEP*, 1306:081 (2013).

[7] Aad, G., et al. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". *Phys.Rev.Lett.*, 114:191803 (2015).

[8] Aad, G., et al. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV". *JHEP*, 1608:45 (2016).

[9] Khachatryan, V., et al. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". *Phys.Rev.*, D92:012004 (2015).

[10] Aad, G., et al. "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector". *Eur.Phys.J.*, C75:476 (2015).

[11] Higgs, P. W. "Broken Symmetries and the Masses of Gauge Bosons". *Phys.Rev.Lett.*, 13:508–509 (1964).

[12] Higgs, P. W. "Broken symmetries, massless particles and gauge fields". *Phys.Lett.*, 12:132–133 (1964).

[13] Englert, F. and Brout, R. "Broken Symmetry and the Mass of Gauge Vector Mesons". *Phys.Rev.Lett.*, 13:321–323 (1964).

[14] Nambu, Y. and Jona-Lasinio, G. "Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1." *Phys.Rev.*, 122:345–358 (1961).

[15] Guralnik, G. S., Hagen, C. R., and Kibble, T. W. B. "Global Conservation Laws and Massless Particles". *Phys.Rev.Lett.*, 13:585–587 (1964).

[16] Higgs, P. W. "Spontaneous Symmetry Breakdown without Massless Bosons". *Phys.Rev.*, 145:1156–1163 (1966).

[17] Kibble, T. W. B. "Symmetry breaking in non-Abelian gauge theories". *Phys.Rev.*, 155:1554–1561 (1967).

[18] Weinberg, S. "A Model of Leptons". *Phys.Rev.Lett.*, 19:1264–1266 (1967).

[19] Glashow, S. L. "Partial Symmetries of Weak Interactions". *Nucl.Phys.*, 22:579–588 (1961).

[20] 't Hooft, G. and Veltman, M. J. G. "Regularization and Renormalization of Gauge Fields". *Nucl.Phys.*, B44:189–213 (1972).

[21] Salam, A. "Weak and Electromagnetic Interactions". *Conf.Proc.*, C680519:367–377 (1968).

[22] Khachatryan, V., et al. "Search for the associated production of the Higgs boson with a top-quark pair". *JHEP*, 1410, 1409:106, 087 (2014).

[23] Khachatryan, V., et al. "Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method". *Eur.Phys.J.*, C75:251 (2015).

[24] Aad, G., et al. "Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". *JHEP*, 1605:160 (2016).

[25] Aad, G., et al. "Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector". *Phys.Lett.*, B740:222–242 (2015).

[26] Aad, G., et al. "Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector". *Phys.Lett.*, B749:519–541 (2015).

[27] Aad, G., et al. "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". *Eur.Phys.J.*, C75:349 (2015).

[28] Griffiths, D. Introduction to Elementary Particles. Wiley-VCH, Weinheim (2008).

[29] Olive, K. A., et al. "Review of Particle Physics". *Chin.Phys.*, C38:090001 (2014).

[30] Cowan, C. L., et al. "Detection of the free neutrino: A Confirmation". *Science*, 124:103–104 (1956).

[31] Fukuda, Y., et al. "Evidence for oscillation of atmospheric neutrinos". *Phys.Rev.Lett.*, 81:1562–1567 (1998).

[32] Danby, G., et al. "Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos". *Phys.Rev.Lett.*, 9:36–44 (1962).

[33] Perl, M. L., et al. "Evidence for Anomalous Lepton Production in e+ - e- Annihilation". *Phys.Rev.Lett.*, 35:1489–1492 (1975).

[34] Kodama, K., et al. "Observation of tau neutrino interactions". *Phys.Lett.*, B504:218–224 (2001).

[35] Schael, S., et al. "Precision electroweak measurements on the Z resonance". *Phys.Rept.*, 427:257–454 (2006).

[36] Glashow, S. L., Iliopoulos, J., and Maiani, L. "Weak Interactions with Lepton-Hadron Symmetry". *Phys.Rev.*, D2:1285–1292 (1970).

[37] Aubert, J. J., et al. "Experimental Observation of a Heavy Particle J". *Phys.Rev.Lett.*, 33:1404–1406 (1974).

[38] Augustin, J. E., et al. "Discovery of a Narrow Resonance in e^+e^- Annihilation". *Physical Review Letters*, 33(23):1406–1408 (1974).

[39] Kobayashi, M. and Maskawa, T. "CP Violation in the Renormalizable Theory of Weak Interaction". *Prog.Theor.Phys.*, 49:652–657 (1973).

[40] Herb, S. W., et al. "Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions". *Phys.Rev.Lett.*, 39:252–255 (1977).

[41] Abe, F., et al. "Observation of top quark production in $\bar{p}p$ collisions". *Phys.Rev.Lett.*, 74:2626–2631 (1995).

[42] Abachi, S., et al. "Observation of the top quark". *Phys.Rev.Lett.*, 74:2632–2637 (1995).

[43] Hasert, F. J., et al. "Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment". *Phys.Lett.*, B46:138–140 (1973).

[44] Banner, M., et al. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider". *Phys.Lett.*, B122:476–485 (1983).

[45] Bagnaia, P., et al. "Evidence for Z0 —> e+ e- at the CERN anti-p p Collider". *Phys.Lett.*, B129:130–140 (1983).

[46] Arnison, G., et al. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV". *Phys.Lett.*, B122:103–116 (1983).

[47] Arnison, G., et al. "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider". *Phys.Lett.*, B126:398–410 (1983).

[48] Gross, D. J. and Wilczek, F. "Ultraviolet Behavior of Nonabelian Gauge Theories". *Phys.Rev.Lett.*, 30:1343–1346 (1973).

[49] Politzer, H. D. "Reliable Perturbative Results for Strong Interactions?" *Phys.Rev.Lett.*, 30:1346–1349 (1973).

[50] Brandelik, R., et al. "Evidence for Planar Events in e+ e- Annihilation at High-Energies". *Phys.Lett.*, B86:243–249 (1979).

[51] Lee, T. D. and Yang, C.-N. "Question of Parity Conservation in Weak Interactions". *Phys.Rev.*, 104:254–258 (1956).

[52] Wu, C. S., et al. "Experimental Test of Parity Conservation in Beta Decay". *Phys.Rev.*, 105:1413–1414 (1957).

[53] Christenson, J. H., et al. "Evidence for the 2 pi Decay of the k(2)0 Meson". *Phys.Rev.Lett.*, 13:138–140 (1964).

[54] Fanti, V., et al. "A New measurement of direct CP violation in two pion decays of the neutral kaon". *Phys.Lett.*, B465:335–348 (1999).

[55] Chao, Y., et al. "Evidence for direct CP violation in B0 —> K+ pi- decays". *Phys.Rev.Lett.*, 93:191802 (2004).

[56] Aubert, B., et al. "Observation of direct CP violation in $B^0 \to K^+\pi^-$ decays". *Phys.Rev.Lett.*, 93:131801 (2004).

[57] Sakharov, A. D. "Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe". *Pisma Zh.Eksp.Teor.Fiz.*, 5:32–35 (1967).

[58] Shu, J. and Zhang, Y. "Impact of a CP Violating Higgs Sector: From LHC to Baryogenesis". *Phys.Rev.Lett.*, 111:091801 (2013).

[59] Accomando, E., et al. "Workshop on CP Studies and Non-Standard Higgs Physics". CERN-2006-009 (2006).

[60] Novaes, S. F. "Standard model: An Introduction". arXiv:hep-ph/0001283 (1999).

[61] Djouadi, A. "The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model". *Phys.Rept.*, 457:1–216 (2008).

[62] Peskin, M. E. and Schroeder, D. V. An Introduction To Quantum Field Theory. Westview Press, Reading, Mass. (1995).

[63] Binosi, D. and Theussl, L. "JaxoDraw: A Graphical user interface for drawing Feynman diagrams". *Comput.Phys.Commun.*, 161:76–86 (2004).

[64] Salam, G. P. "Elements of QCD for hadron colliders". arXiv:1011.5131 [hep-ph] (2010).

[65] Skands, P. "Introduction to QCD". arXiv:1207.2389 [hep-ph] (2012).

[66] Collins, J. C. and Soper, D. E. "Parton Distribution and Decay Functions". *Nucl.Phys.*, B194:445–492 (1982).

[67] Dokshitzer, Y. L. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." *Sov.Phys.JETP*, 46:641–653 (1977).

[68] Gribov, V. N. and Lipatov, L. N. "Deep inelastic e p scattering in perturbation theory". *Sov.J.Nucl.Phys.*, 15:438–450 (1972).

[69] Altarelli, G. and Parisi, G. "Asymptotic Freedom in Parton Language". *Nucl.Phys.*, B126:298–318 (1977).

[70] Harland-Lang, L. A., et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". *Eur.Phys.J.*, C75:204 (2015).

[71] Alwall, J., et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". *Eur.Phys.J.*, C53:473–500 (2007).

[72] Sjostrand, T., Mrenna, S., and Skands, P. Z. "PYTHIA 6.4 Physics and Manual". *JHEP*, 0605:026 (2006).

[73] Corcella, G., et al. "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)". *JHEP*, 0101:010 (2001).

[74] Cacciari, M., Salam, G. P., and Sapeta, S. "On the characterisation of the underlying event". *JHEP*, 1004:065 (2010).

[75] Skands, P., Rojo, J., and Carrazza, S. "Tuning PYTHIA 8.1: The Monash 2013 Tune". *Eur.Phys.J.*, C74:3024 (2014).

[76] Agostinelli, S., et al. "GEANT4: A Simulation toolkit". *Nucl.Instrum.Meth.*, A506:250–303 (2002).

[77] Cacciari, M., Salam, G. P., and Soyez, G. "The Anti-k(t) jet clustering algorithm". *JHEP*, 0804:063 (2008).

[78] Dokshitzer, Y. L., et al. "Better jet clustering algorithms". *JHEP*, 9708:001 (1997).

[79] Salam, G. P. "Towards Jetography". *Eur.Phys.J.*, C67:637–686 (2010).

[80] LHC Higgs cross-section working group. "SM Higgs production cross sections at $\sqrt{s}$ = 13 TeV (update in CERN Report4 2016)". `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV`. Last visited 2016-10-29.

[81] Heinemeyer, S., et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". CERN-2013-004 (2013).

[82] LHC Higgs cross-section working group. "LHC Higgs Cross Section Working Group". `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG`. Last visited 2016-09-16.

[83] Khachatryan, V., et al. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". *Phys.Rev.*, D92:012004 (2014).

[84] Aad, G., et al. "Determination of spin and parity of the Higgs boson in the $WW^* \to e\nu\mu\nu$ decay channel with the ATLAS detector". *Eur.Phys.J.*, C75:231 (2015).

[85] Khachatryan, V., et al. "Search for Higgs boson off-shell production in proton-proton collisions at 7 and 8 TeV and derivation of constraints on its total decay width". *JHEP*, 1609:051 (2016).

[86] Aad, G., et al. "Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector". *Eur.Phys.J.*, C75:335 (2015).

[87] Khachatryan, V., et al. “Searches for invisible decays of the Higgs boson in pp collisions at $\sqrt{s} = 7$, 8, and 13 TeV”. arXiv:1610.09218 [hep-ex] (2016).

[88] Aad, G., et al. “Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector”. *JHEP*, 1511:206 (2015).

[89] ATLAS Collaboration. “ATLAS Higgs Physics Public Results”. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HiggsPublicResults`. Last visited 2016-11-02.

[90] CMS Collaboration. “CMS Higgs Physics Results”. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsHIG`. Last visited 2016-11-02.

[91] Aamodt, K., et al. “The ALICE experiment at the CERN LHC”. *JINST*, 3:S08002 (2008).

[92] Alves, A. A., et al. “The LHCb Detector at the LHC”. *JINST*, 3:S08005 (2008).

[93] CERN. “Overall view of the LHC”. OPEN-PHO-CHART-2014-006 (2014).

[94] CMS Collaboration. “SketchUpCMS”. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS`. Last visited 2016-10-12.

[95] Sakuma, T. and McCauley, T. “Detector and Event Visualization with SketchUp at the CMS Experiment”. *J.Phys.Conf.Ser.*, 513:022032 (2014).

[96] Chatrchyan, S., et al. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. *JINST*, 9:P10009 (2014).

[97] Khachatryan, V., et al. “CMS Tracking Performance Results from early LHC Operation”. *Eur.Phys.J.*, C70:1165–1192 (2010).

[98] Chatrchyan, S., et al. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. *JINST*, 9:P10009 (2014).

[99] Chatrchyan, S., et al. “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data”. *JINST*, 5:T03012 (2010).

[100] Chatrchyan, S., et al. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”. *JINST*, 7:P10002 (2012).

[101] Khachatryan, V., et al. “The CMS trigger system”. arXiv:1609.02366 [physics.ins-det] (2016).

[102] Bird, I., et al. “LHC computing Grid. Technical design report”. CERN-LHCC-2005-024 (2005).

[103] Fruhwirth, R. “Application of Kalman filtering to track and vertex fitting”. *Nucl.Instrum.Meth.*, A262:444–450 (1987).

[104] Rose, K. “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”. *Proceedings of the IEEE*, 86(11):2210–2239 (1998).

[105] Fruhwirth, R., Waltenberger, W., and Vanlaer, P. “Adaptive vertex fitting”. *J.Phys.*, G34:N343 (2007).

[106] Khachatryan, V., et al. “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{\mathrm{s}} = 8$ TeV”. *JINST*, 10:P06005 (2015).

[107] CMS Collaboration. “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”. CMS-PAS-PFT-09-001 (2009).

[108] Bartosik, N. “HEP sketches”. `http://bartosik.pp.ua/hep_sketches/`. Last visited 2016-10-18.

[109] CMS Collaboration. “Pileup Removal Algorithms”. CMS-PAS-JME-14-001 (2014).

[110] Cacciari, M., Salam, G. P., and Soyez, G. “FastJet User Manual”. *Eur.Phys.J.*, C72:1896 (2012).

[111] Khachatryan, V., et al. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. arXiv:1607.03663 [hep-ex] (2016).

[112] CMS Collaboration. “Jet energy scale and resolution performances with 13 TeV data”. CMS-DP-2016-020 (2016).

[113] CMS Collaboration. “Identification of b quark jets at the CMS Experiment in the LHC Run 2”. CMS-PAS-BTV-15-001 (2016).

[114] Khachatryan, V., et al. “Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV”. *JHEP*, 1103:136 (2011).

[115] Chatrchyan, S., et al. “Identification of b-quark jets with the CMS experiment”. *JINST*, 8:P04013 (2013).

[116] Petrucciani, G., Rizzi, A., and Vuosalo, C. “Mini-AOD: A New Analysis Data Format for CMS”. *J.Phys.Conf.Ser.*, 664:072052 (2015).

[117] CMS Collaboration. “CMS Physics : Technical Design Report Volume 1: Detector Performance and Software”. CERN-LHCC-2006-001 (2006).

[118] Brun, R. and Rademakers, F. “ROOT: An object oriented data analysis framework”. *Nucl.Instrum.Meth.*, A389:81–86 (1997).

[119] Cowan, G. Statistical Data Analysis. Oxford University Press, Oxford : New York (1997).

[120] Neyman, J. and Pearson, E. S. “On the Problem of the Most Efficient Tests of Statistical Hypotheses”. *Phil.Trans.R Soc.Lond.A*, 231(694-706):289–337 (1933).

[121] Wilks, S. S. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. *Ann. Math. Statist.*, 9(1):60–62 (1938).

[122] ATLAS Collaboration, CMS Collaboration, and LHC Higgs Combination Group. “Procedure for the LHC Higgs boson search combination in summer 2011”. ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005 (2011).

[123] Cowan, G., et al. “Asymptotic formulae for likelihood-based tests of new physics”. *Eur.Phys.J.*, C71:1554 (2011).

[124] Verkerke, W. and Kirkby, D. P. “The RooFit toolkit for data modeling”. *eConf*, C0303241:MOLT007 (2003).

[125] Moneta, L., et al. “The RooStats Project”. *PoS*, ACAT2010:057 (2010).

[126] Cranmer, K., et al. “HistFactory: A tool for creating statistical models for use with RooFit and RooStats”. CERN-OPEN-2012-016 (2012).

[127] Kondo, K. “Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models”. *J.Phys.Soc.Jap.*, 57:4126–4140 (1988).

[128] Dalitz, R. H. and Goldstein, G. R. “The Decay and polarization properties of the top quark”. *Phys.Rev.*, D45:1531–1543 (1991).

[129] Abazov, V. M., et al. “A precision measurement of the mass of the top quark”. *Nature*, 429:638–642 (2004).

[130] Friedman, J. H. “Greedy function approximation: A gradient boosting machine.” *Ann. Statist.*, 29(5):1189–1232 (2001).

[131] Hocker, A., et al. “TMVA - Toolkit for Multivariate Data Analysis”. *PoS*, ACAT:040 (2007).

[132] Aad, G., et al. “Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment”. *Eur.Phys.J.*, C76:6 (2015).

[133] Khachatryan, V., et al. “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”. *Eur.Phys.J.*, C75:212 (2014).

[134] Frixione, S., et al. “Weak corrections to Higgs hadroproduction in association with a top-quark pair”. *JHEP*, 1409:065 (2014).

[135] Contino, R., et al. “Effective Lagrangian for a light Higgs-like scalar”. *JHEP*, 1307:035 (2013).

[136] Chatrchyan, S., et al. “Measurement of the properties of a Higgs boson in the four-lepton final state”. *Phys.Rev.*, D89:092007 (2014).

[137] Khachatryan, V., et al. “Combined search for anomalous pseudoscalar HVV couplings in VH(H $\rightarrow b\bar{b}$) production and H → VV decay”. *Phys.Lett.*, B759:672–696 (2016).

[138] Perelstein, M., Peskin, M. E., and Pierce, A. “Top quarks and electroweak symmetry breaking in little Higgs models”. *Phys.Rev.*, D69:075002 (2004).

[139] Alekhin, S., Djouadi, A., and Moch, S. “The top quark and Higgs boson masses and the stability of the electroweak vacuum”. *Phys.Lett.*, B716:214–219 (2012).

[140] Boudjema, F., et al. “Lab-frame observables for probing the top-Higgs interaction”. *Phys.Rev.*, D92:015019 (2015).

[141] Brod, J., Zupan, J., and Haisch, U. “Constraints on CP-violating Higgs couplings to the third generation”. *JHEP*, 1311:180 (2013).

[142] Gritsan, A. V., et al. “Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques”. *Phys.Rev.*, D94:055023 (2016).

[143] Harnik, R., et al. “Measuring CP violation in $h \to \tau^+\tau^-$ at colliders”. *Phys.Rev.*, D88:076009 (2013).

[144] Barr, S. M. and Zee, A. “Electric Dipole Moment of the Electron and of the Neutron”. *Phys.Rev.Lett.*, 65:21–24 (1990).

[145] Baron, J., et al. “Order of Magnitude Smaller Limit on the Electric Dipole Moment of the Electron”. *Science*, 343:269–272 (2013).

[146] Ellis, J., et al. “Disentangling Higgs-Top Couplings in Associated Production”. *JHEP*, 1404:004 (2014).

[147] Nishiwaki, K., Niyogi, S., and Shivaji, A. “ttH Anomalous Coupling in Double Higgs Production”. *JHEP*, 1404:011 (2014).

[148] Kobakhidze, A., Wu, L., and Yue, J. “Anomalous Top-Higgs Couplings and Top Polarisation in Single Top and Higgs Associated Production at the LHC”. *JHEP*, 1410:100 (2014).

[149] CMS Collaboration. “Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV”. CMS-PAS-HIG-13-005 (2013).

[150] Aad, G., et al. “Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC”. *Phys.Lett.*, B726:88–119 (2013).

[151] Bechtle, P., et al. “HiggsSignals: Confronting arbitrary Higgs sectors with measurements at the Tevatron and the LHC”. *Eur.Phys.J.*, C74:2711 (2014).

[152] Aad, G., et al. “Constraints on non-Standard Model Higgs boson interactions in an effective Lagrangian using differential cross sections measured in the $H \to \gamma\gamma$ decay channel at $\sqrt{s} = 8$TeV with the ATLAS detector”. *Phys.Lett.*, B753:69–85 (2016).

[153] Khachatryan, V., et al. “Search for the associated production of a Higgs boson with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ TeV”. *JHEP*, 1606:177 (2016).

[154] Fink, S. “Probing the Top-Yukawa Coupling by Searching for Associated Higgs Boson Production with a Single Top Quark at the CMS Experiment”. Ph.D. thesis, KIT (2016).

[155] Biswas, S., Mele, B., and Gabrielli, E. “Single top and Higgs associated production as a probe of the Htt coupling sign at the LHC”. *JHEP*, 1301:088 (2012).

[156] Chatrchyan, S., et al. “Search for a Higgs boson decaying into a Z and a photon in pp collisions at sqrt(s) = 7 and 8 TeV”. *Phys.Lett.*, B726:587–609 (2013).

[157] Hespel, B., Maltoni, F., and Vryonidou, E. “Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM”. *JHEP*, 1506:065 (2015).

[158] Mileo, N., et al. “Pseudoscalar top-Higgs coupling: Exploration of CP-odd observables to resolve the sign ambiguity”. *JHEP*, 1607:056 (2016).

[159] Spira, M. “HIGLU: A program for the calculation of the total Higgs production cross-section at hadron colliders via gluon fusion including QCD corrections”. arXiv:hep-ph/9510347 (1995).

[160] Spira, M. “HIGLU and HDECAY: Programs for Higgs boson production at the LHC and Higgs boson decay widths”. *Nucl.Instrum.Meth.*, A389:357–360 (1997).

[161] Nadolsky, P. M., et al. “Implications of CTEQ global analysis for collider observables”. *Phys.Rev.*, D78:013004 (2008).

[162] Djouadi, A., Kalinowski, J., and Spira, M. “HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension”. *Comput.Phys.Commun.*, 108:56–74 (1998).

[163] Artoisenet, P., et al. “A framework for Higgs characterisation”. *JHEP*, 1311:043 (2013).

[164] Alwall, J., et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. *JHEP*, 1407:079 (2014).

[165] Demartin, F., et al. “Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction”. *Eur.Phys.J.*, C74:3065 (2014).

[166] Butterworth, J., et al. “PDF4LHC recommendations for LHC Run II”. *J.Phys.*, G43:023001 (2016).

[167] Buckley, A., et al. “LHAPDF6: Parton density access in the LHC precision era”. *Eur.Phys.J.*, C75:132 (2015).

[168] Alwall, J., et al. “MadGraph 5 : Going Beyond”. *JHEP*, 1106:128 (2011).

[169] Khachatryan, V., et al. “Observation of the diphoton decay of the Higgs boson and measurement of its properties”. *Eur.Phys.J.*, C74:3076 (2014).

[170] Chatrchyan, S., et al. “Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states”. *JHEP*, 1401:096 (2014).

[171] Chatrchyan, S., et al. “Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons”. *JHEP*, 1405:104 (2014).

[172] Chatrchyan, S., et al. “Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks”. *Phys.Rev.*, D89:012003 (2014).

[173] Chatrchyan, S., et al. “Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC”. *JHEP*, 1305:145 (2013).

[174] CMS Collaboration. “Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s} = 13$ TeV”. CMS-PAS-HIG-16-022 (2016).

[175] Harigaya, K., et al. “Search for the Top Partner at the LHC using Multi-b-Jet Channels”. *Phys.Rev.*, D86:015005 (2012).

[176] Degrassi, G., et al. “Probing the Higgs self coupling via single Higgs production at the LHC”. arXiv:1607.04251 [hep-ph] (2016).

[177] Denner, A., Feger, R., and Scharf, A. “Irreducible background and interference effects for Higgs-boson production in association with a top-quark pair”. *JHEP*, 1504:008 (2015).

[178] Sjöstrand, T., et al. “An Introduction to PYTHIA 8.2”. *Comput.Phys.Commun.*, 191:159–177 (2014).

[179] Ball, R. D., et al. “Parton distributions for the LHC Run II”. *JHEP*, 1504:040 (2014).

[180] Alioli, S., et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX”. *JHEP*, 1006:043 (2010).

[181] Frixione, S., Nason, P., and Oleari, C. “Matching NLO QCD computations with Parton Shower simulations: The POWHEG method”. *JHEP*, 0711:070 (2007).

[182] Re, E. “Single-top Wt-channel production matched with parton showers using the POWHEG method”. *Eur.Phys.J.*, C71:1547 (2010).

[183] Frixione, S. and Webber, B. R. “Matching NLO QCD computations and parton shower simulations”. *JHEP*, 0206:029 (2002).

[184] Bohm, G. and Zech, G. Einführung in Statistik und Messwertanalyse für Physiker. Springer, Berlin (2006).

[185] Frederix, R. and Frixione, S. “Merging meets matching in MC@NLO”. *JHEP*, 1212:061 (2012).

[186] Hoeche, S., et al. “Matching parton showers and matrix elements”. arXiv:hep-ph/0602031 (2006).

[187] Dittmaier, S., et al. “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”. CERN-2011-002 (2011).

[188] Reina, L. and Dawson, S. “Next-to-leading order results for t anti-t h production at the Tevatron”. *Phys.Rev.Lett.*, 87:201804 (2001).

[189] Dawson, S., et al. “Associated top quark Higgs boson production at the LHC”. *Phys.Rev.*, D67:071503 (2002).

[190] Dawson, S., et al. “Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections”. *Phys.Rev.*, D68:034022 (2003).

[191] Beenakker, W., et al. “Higgs radiation off top quarks at the Tevatron and the LHC”. *Phys.Rev.Lett.*, 87:201805 (2001).

[192] Beenakker, W., et al. “NLO QCD corrections to t anti-t H production in hadron collisions”. *Nucl.Phys.*, B653:151–203 (2002).

[193] Zhang, Y., et al. “QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider”. *Phys.Lett.*, B738:1–5 (2014).

[194] Frixione, S., et al. “Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons”. *JHEP*, 1506:184 (2015).

[195] Ball, R. D., et al. “Parton distributions with QED corrections”. *Nucl.Phys.*, B877:290–320 (2013).

[196] Czakon, M. and Mitov, A. “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”. *Comput.Phys.Commun.*, 185:2930 (2014).

[197] Martin, A. D., et al. “Parton distributions for the LHC”. *Eur.Phys.J.*, C63:189–285 (2009).

[198] Ball, R. D., et al. “Parton distributions with LHC data”. *Nucl.Phys.*, B867:244–289 (2013).

[199] Lai, H.-L., et al. “New parton distributions for collider physics”. *Phys.Rev.*, D82:074024 (2010).

[200] Botje, M., et al. “The PDF4LHC Working Group Interim Recommendations”. arXiv:1101.0536 [hep-ph] (2011).

[201] Aliev, M., et al. “HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR”. *Comput.Phys.Commun.*, 182:1034–1046 (2011).

[202] Kant, P., et al. “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”. *Comput.Phys.Commun.*, 191:74–89 (2015).

[203] Kidonakis, N. “Top Quark Production”. arXiv:1311.0283 [hep-ph] (2014).

[204] Gavin, R., et al. “FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order”. *Comput.Phys.Commun.*, 182:2388–2403 (2011).

[205] Li, Y., et al. “W Physics at the LHC with FEWZ 2.1”. *Comput.Phys.Commun.*, 184:208–214 (2013).

[206] Li, Y. and Petriello, F. “Combining QCD and electroweak corrections to dilepton production in FEWZ”. *Phys.Rev.*, D86:094034 (2012).

[207] Gehrmann, T., et al. “W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD”. *Phys.Rev.Lett.*, 113:212001 (2014).

[208] Campbell, J. M., Ellis, R. K., and Williams, C. “Vector boson pair production at the LHC”. *JHEP*, 1107:018 (2011).

[209] Bartosik, N. “Associated Top-Quark-Pair and b-Jet Production in the Dilepton Channel at $\sqrt{s} = 8$ TeV as Test of QCD and Background to tt+Higgs Production”. Ph.D. thesis, Universität Hamburg (2015).

[210] Gleisberg, T., et al. “Event generation with SHERPA 1.1”. *JHEP*, 0902:007 (2009).

[211] Cascioli, F., et al. "NLO matching for ttbb production with massive b-quarks". *Physics Letters B*, 734:210–214 (2014).

[212] Moretti, N., et al. "Measuring the signal strength in $\mathrm{t\bar{t}H}$ with $\mathrm{H \to b\bar{b}}$". *Phys.Rev.*, D93:014019 (2016).

[213] Artoisenet, P., et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". *JHEP*, 1303:015 (2012).

[214] Bredenstein, A., et al. "NLO QCD Corrections to Top Anti-Top Bottom Anti-Bottom Production at the LHC: 2. full hadronic results". *JHEP*, 1003:021 (2010).

[215] Bredenstein, A., et al. "NLO QCD corrections to pp —> t anti-t b anti-b + X at the LHC". *Phys.Rev.Lett.*, 103:012002 (2009).

[216] Bevilacqua, G., et al. "Assault on the NLO Wishlist: Pp —> t anti-t b anti-b". *JHEP*, 0909:109 (2009).

[217] Khachatryan, V., et al. "Measurement of $\mathrm{t\bar{t}}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV". *Eur.Phys.J.*, C76:379 (2016).

[218] Bahr, M., et al. "Herwig++ Physics and Manual". *Eur.Phys.J.*, C58:639–707 (2008).

[219] Bellm, J., et al. "Reweighting Parton Showers". *Phys.Rev.*, D94:034028 (2016).

[220] Plehn, T., Salam, G. P., and Spannowsky, M. "Fat Jets for a Light Higgs". *Phys.Rev.Lett.*, 104:111801 (2010).

[221] Williamson, S. "Search for Higgs-Boson Production in Association with a Top-Quark Pair in the Boosted Regime with the CMS Experiment". Ph.D. thesis, KIT (2016).

[222] Butterworth, J. M., et al. "Jet substructure as a new Higgs search channel at the LHC". *Phys.Rev.Lett.*, 100:242001 (2008).

[223] Thaler, J. and Van Tilburg, K. "Identifying Boosted Objects with N-subjettiness". *JHEP*, 1103:015 (2011).

[224] Plehn, T., et al. "Stop Reconstruction with Tagged Tops". *JHEP*, 1010:078 (2010).

[225] Anders, C., et al. "Benchmarking an even better top tagger algorithm". *Phys.Rev.*, D89:074047 (2014).

[226] Kasieczka, G., et al. "Resonance Searches with an Updated Top Tagger". *JHEP*, 1506:203 (2015).

[227] Schröder, M. "Quality of Jet Measurements and Impact on a Search for New Physics at CMS". Ph.D. thesis, Universität Hamburg (2012).

[228] CMS Collaboration. "CMS Luminosity Measurement for the 2015 Data Taking Period". CMS-PAS-LUM-15-001 (2016).

[229] CMS Collaboration. "Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H \to b\bar{b}}$ decay channel with $\sqrt{s} = 13$ TeV pp collisions at the CMS experiment". `http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/HIG-16-004/index.html`. Last visited 2016-10-02.

[230] Cascioli, F., Maierhofer, P., and Pozzorini, S. “Scattering Amplitudes with Open Loops”. *Phys.Rev.Lett.*, 108:111601 (2012).

[231] Lepage, G. P. “A New Algorithm for Adaptive Multidimensional Integration”. *J.Comput.Phys.*, 27:192 (1978).

[232] Bjorken, J. D. and Brodsky, S. J. “Statistical Model for Electron-Positron Annihilation Into Hadrons”. *Phys.Rev.*, D1:1416–1420 (1970).

[233] Fox, G. C. and Wolfram, S. “Event Shapes in e+ e- Annihilation”. *Nucl.Phys.*, B149:413 (1979).

[234] Kennedy, J. and Eberhart, R. “Particle swarm optimization”. In “IEEE International Conference on Neural Networks, 1995. Proceedings”, pages 1942–1948 (1995).

[235] El Morabit, K. “A study of the multivariate analysis of Higgs boson production in association with a top quark-antiquark pair in the boosted regime at the CMS experiment”. Master thesis, KIT (2015).

[236] Barlow, R. J. and Beeston, C. “Fitting using finite Monte Carlo samples”. *Comput.Phys.Commun.*, 77:219–228 (1993).

[237] James, F. and Roos, M. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”. *Comput.Phys.Commun.*, 10:343–367 (1975).

[238] Chatrchyan, S., et al. “Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV”. *Phys.Lett.*, B710:26–48 (2013).

[239] Biswas, S., et al. “Enhancing the $t\bar{t}H$ signal through top-quark spin polarization effects at the LHC”. *JHEP*, 1407:020 (2014).

[240] Gallicchio, J., et al. “Multivariate discrimination and the Higgs + W/Z search”. *JHEP*, 1104:069 (2011).

[241] Dolan, M. J., et al. “Determining the quantum numbers of simplified models in $t\bar{t}X$ production at the LHC”. *Phys.Rev.*, D94:015025 (2016).

[242] Buckley, M. R. and Goncalves, D. “Boosting the Direct CP Measurement of the Higgs-Top Coupling”. *Phys.Rev.Lett.*, 116:091801 (2016).

[243] Gao, Y., et al. “Spin determination of single-produced resonances at hadron colliders”. *Phys.Rev.*, D81:075022 (2010).

[244] Bolognesi, S., et al. “On the spin and parity of a single-produced resonance at the LHC”. *Phys.Rev.*, D86:095031 (2012).

[245] Anderson, I., et al. “Constraining anomalous HVV interactions at proton and lepton colliders”. *Phys.Rev.*, D89:035007 (2014).

[246] CMS Collaboration. “Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with $\sqrt{s} = 13$ TeV pp collisions at the CMS experiment”. CMS-PAS-HIG-16-004 (2016).

[247] CMS Collaboration. “First results on Higgs to gammagamma at 13 TeV”. CMS-PAS-HIG-15-005 (2016).

[248] CMS Collaboration. “ttH Combination Moriond 2016”. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/TTHCombMoriond2016`. Last visited 2016-09-05.

[249] CMS Collaboration. “Search for ttH production in multilepton final states at sqrt(s) = 13 TeV”. CMS-PAS-HIG-15-008 (2016).

[250] CMS Collaboration. “Updated measurements of Higgs boson production in the diphoton decay channel at $\sqrt{s} = 13$ TeV in pp collisions at CMS.” CMS-PAS-HIG-16-020 (2016).

[251] ATLAS Collaboration. “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. ATLAS-CONF-2016-080 (2016).

[252] Czakon, M., Heymes, D., and Mitov, A. “High-precision differential predictions for top-quark pairs at the LHC”. *Phys.Rev.Lett.*, 116:082003 (2016).

[253] Czakon, M., Heymes, D., and Mitov, A. “Dynamical scales for multi-TeV top-pair production at the LHC”. arXiv:1606.03350 [hep-ph] (2016).

[254] ATLAS Collaboration. “Search for the Associated Production of a Higgs Boson and a Top Quark Pair in Multilepton Final States with the ATLAS Detector”. ATLAS-CONF-2016-058 (2016).

[255] ATLAS Collaboration. “Measurement of fiducial, differential and production cross sections in the $H \to \gamma\gamma$ decay channel with 13.3 fb^{-1} of 13 TeV proton-proton collision data with the ATLAS detector”. ATLAS-CONF-2016-067 (2016).

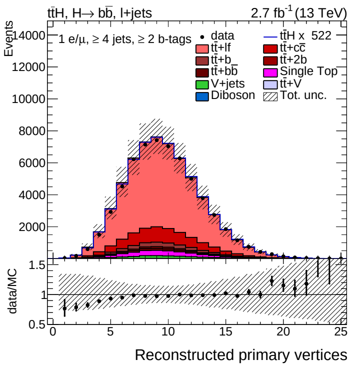
[256] ATLAS Collaboration. “Combination of the searches for Higgs boson production in association with top quarks in the $\gamma\gamma$, multilepton, and $b\bar{b}$ decay channels at $\sqrt{s}$=13 TeV with the ATLAS Detector”. ATLAS-CONF-2016-068 (2016).

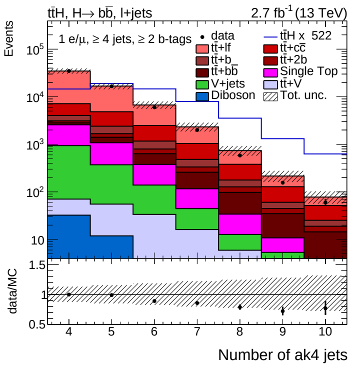
Appendix

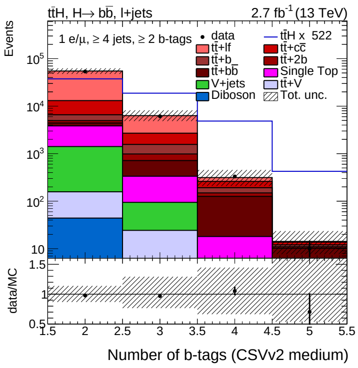
A.1 Background-model validation in control region

In this section, the data recorded in 2015 (black dots) is compared with the expected background predicted by simulation (stacked histograms) for different observables. Events are selected by requiring at least one lepton and four jets ($p_{\mathrm{T}} > 30\ \mathrm{GeV}/c$) of which two are b-tagged. The expected $\mathrm{t\bar{t}H}$ signal is drawn as a blue line and scaled to the same integral as the sum of backgrounds. The most important systematic uncertainties are added in quadrature and shown as hashed error bands. Statistical uncertainties due to the limited number of observed events are drawn as black horizontal lines on the data points. See Section 9.2 for more details.

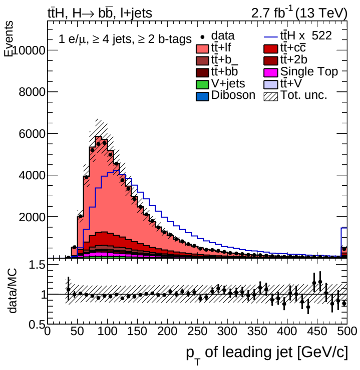
Vertex, jet, and b-tag multiplicity

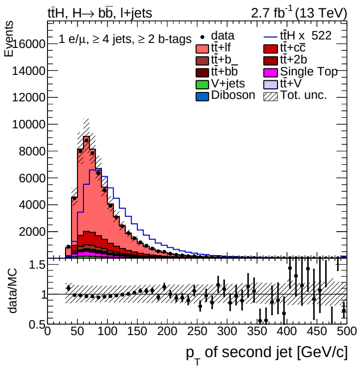


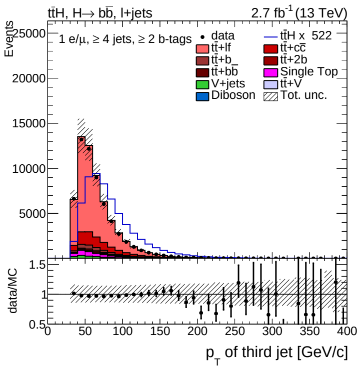


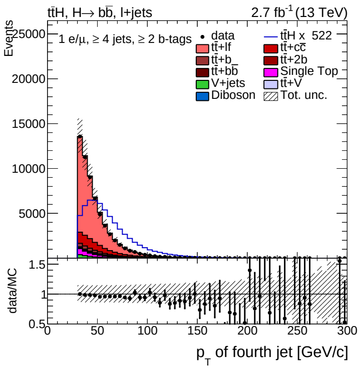


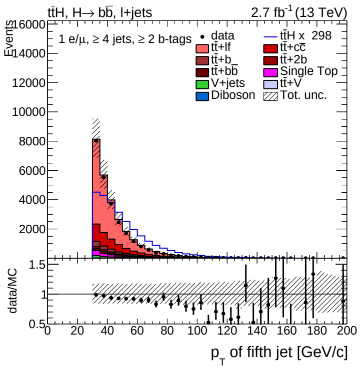
Jets

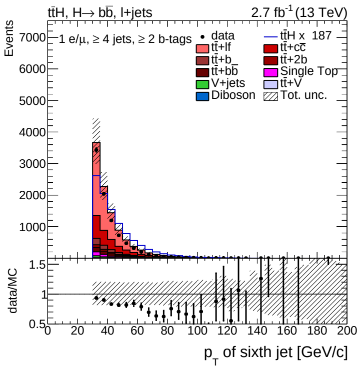


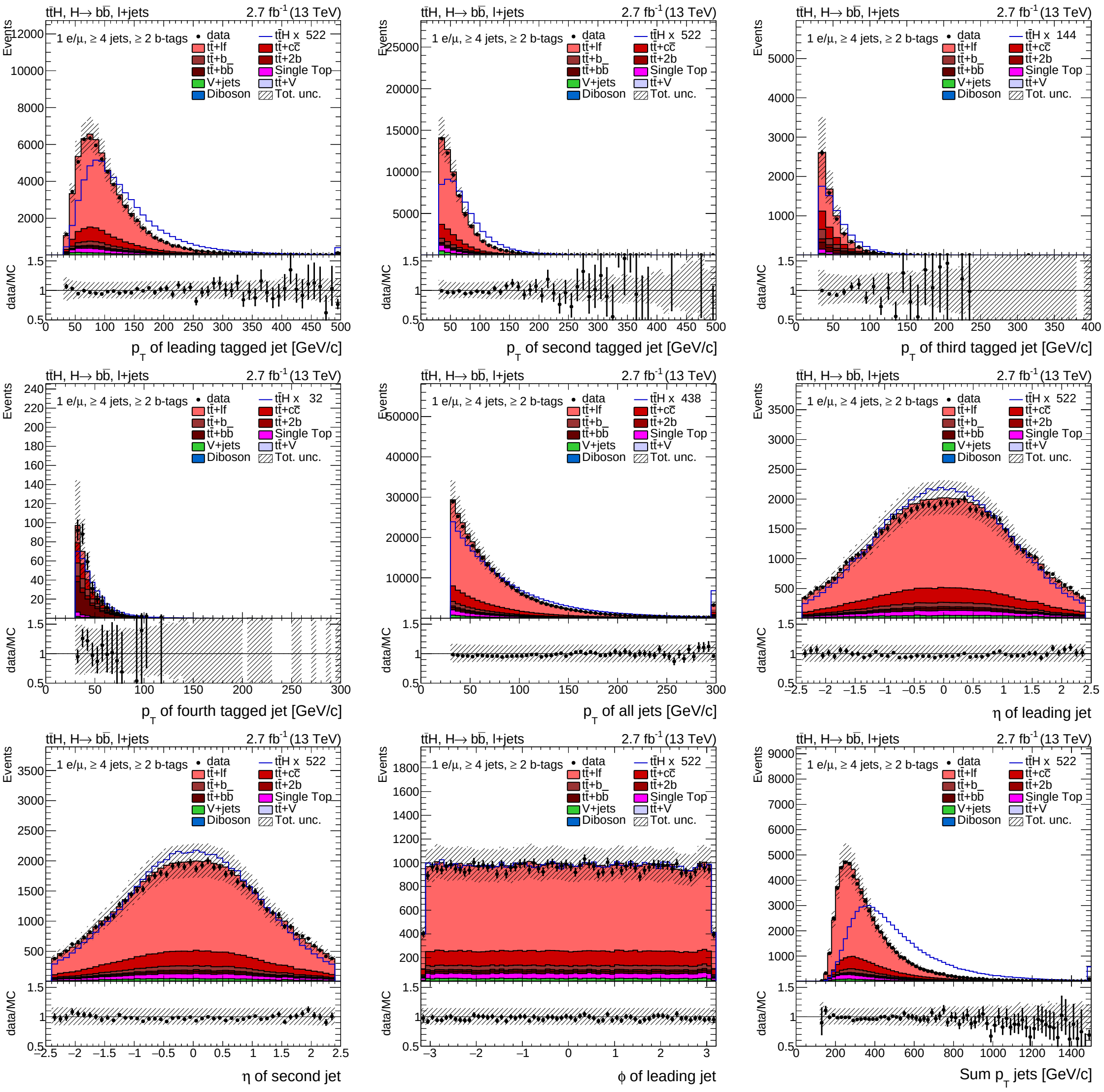




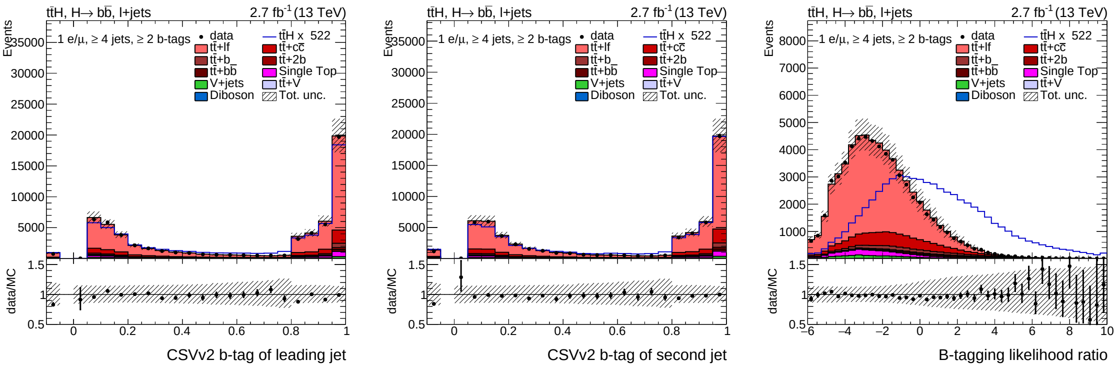




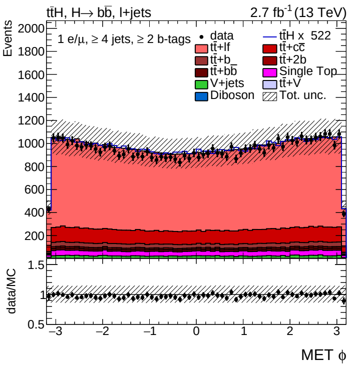


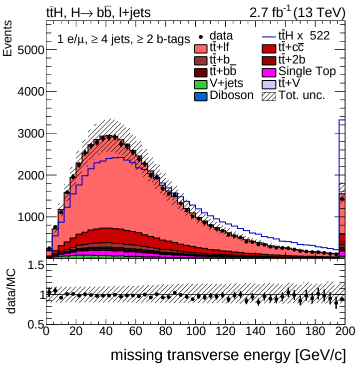

tt̄H, H→ bb̄, l+jets
2.7 fb^{-1} (13 TeV)
1 e/μ, ≥ 4 jets, ≥ 2 b-tags
data
tt̄+lf
tt̄+b
tt̄+bb̄
V+jets
Diboson
tt̄H x 522
tt̄+cc̄
tt̄+2b
Single Top
tt̄+V
Tot. unc.
Events
data/MC
p$_T$ of leading tagged jet [GeV/c]
p$_T$ of second tagged jet [GeV/c]
tt̄H x 144
p$_T$ of third tagged jet [GeV/c]
tt̄H x 32
p$_T$ of fourth tagged jet [GeV/c]
tt̄H x 438
p$_T$ of all jets [GeV/c]
η of leading jet
η of second jet
ϕ of leading jet
Sum p$_T$ jets [GeV/c]

b-tagging

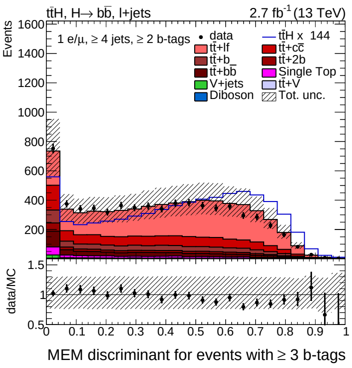

tt̄H, H→ bb̄, l+jets
2.7 fb^{-1} (13 TeV)
1 e/μ, ≥ 4 jets, ≥ 2 b-tags
data
tt̄+lf
tt̄+b
tt̄+bb̄
V+jets
Diboson
tt̄H x 522
tt̄+cc̄
tt̄+2b
Single Top
tt̄+V
Tot. unc.
Events
data/MC
CSVv2 b-tag of leading jet
CSVv2 b-tag of second jet
B-tagging likelihood ratio

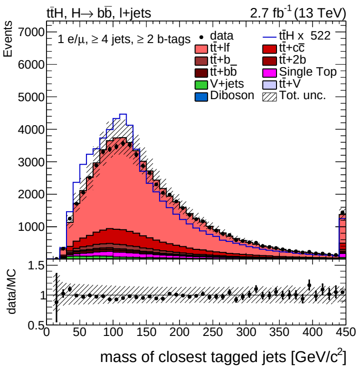
MET

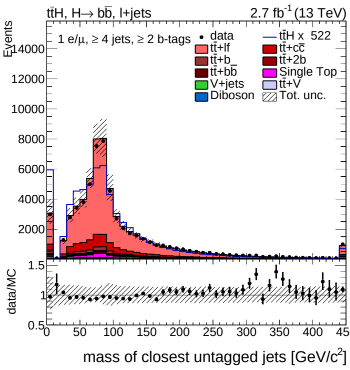




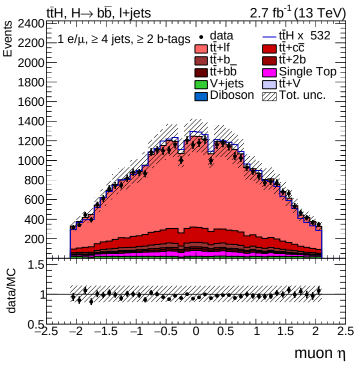
Mass reconstruction and MEM

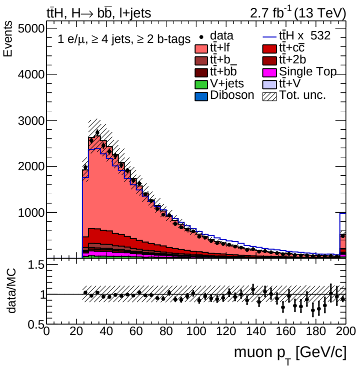


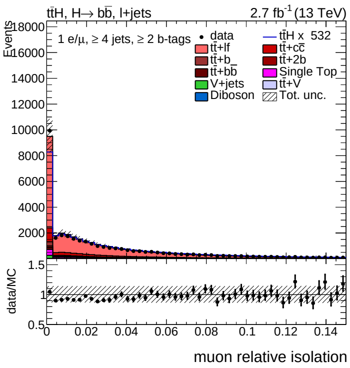


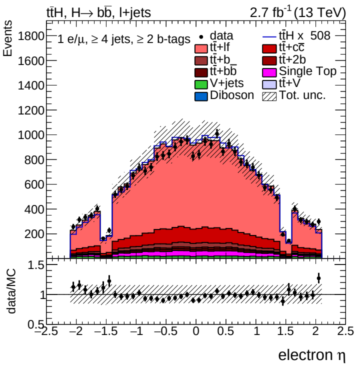


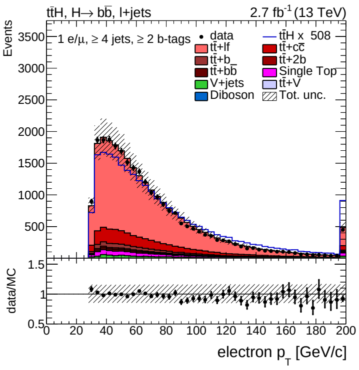
Leptons

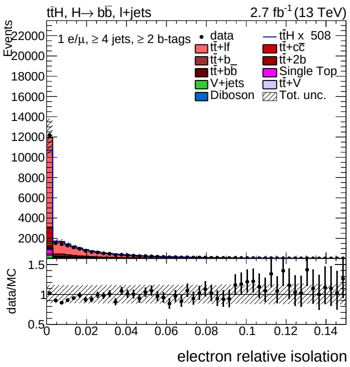




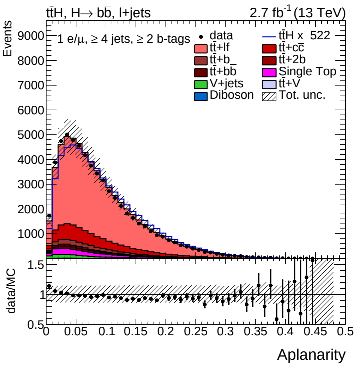


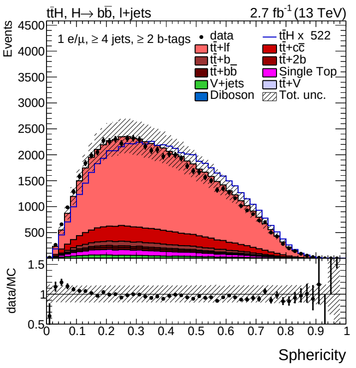


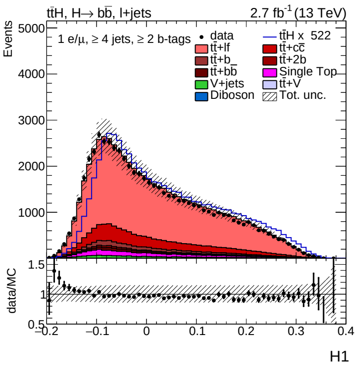


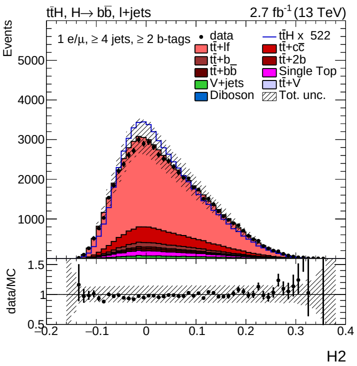


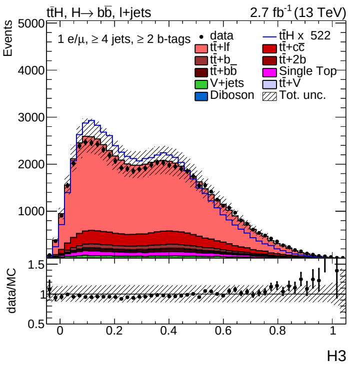
Event shape

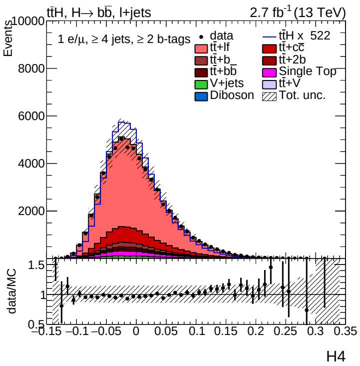




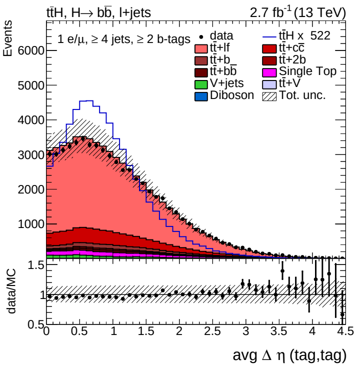


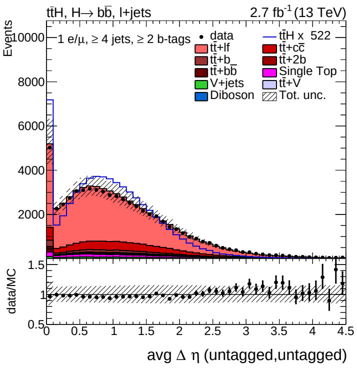


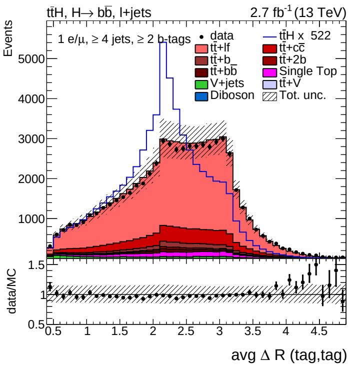


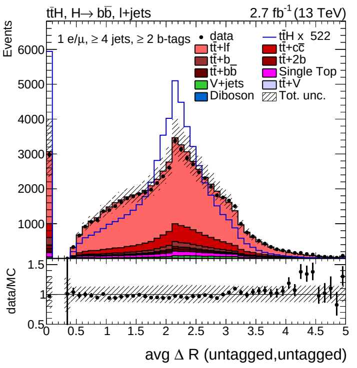


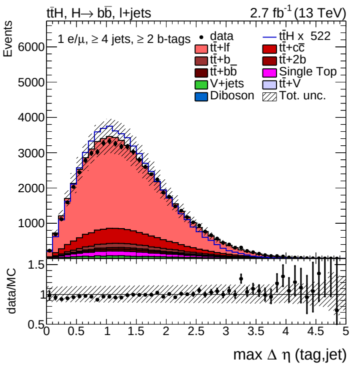
Angular correlations

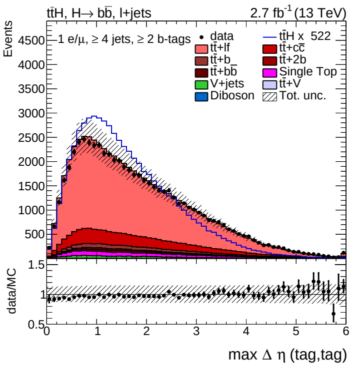




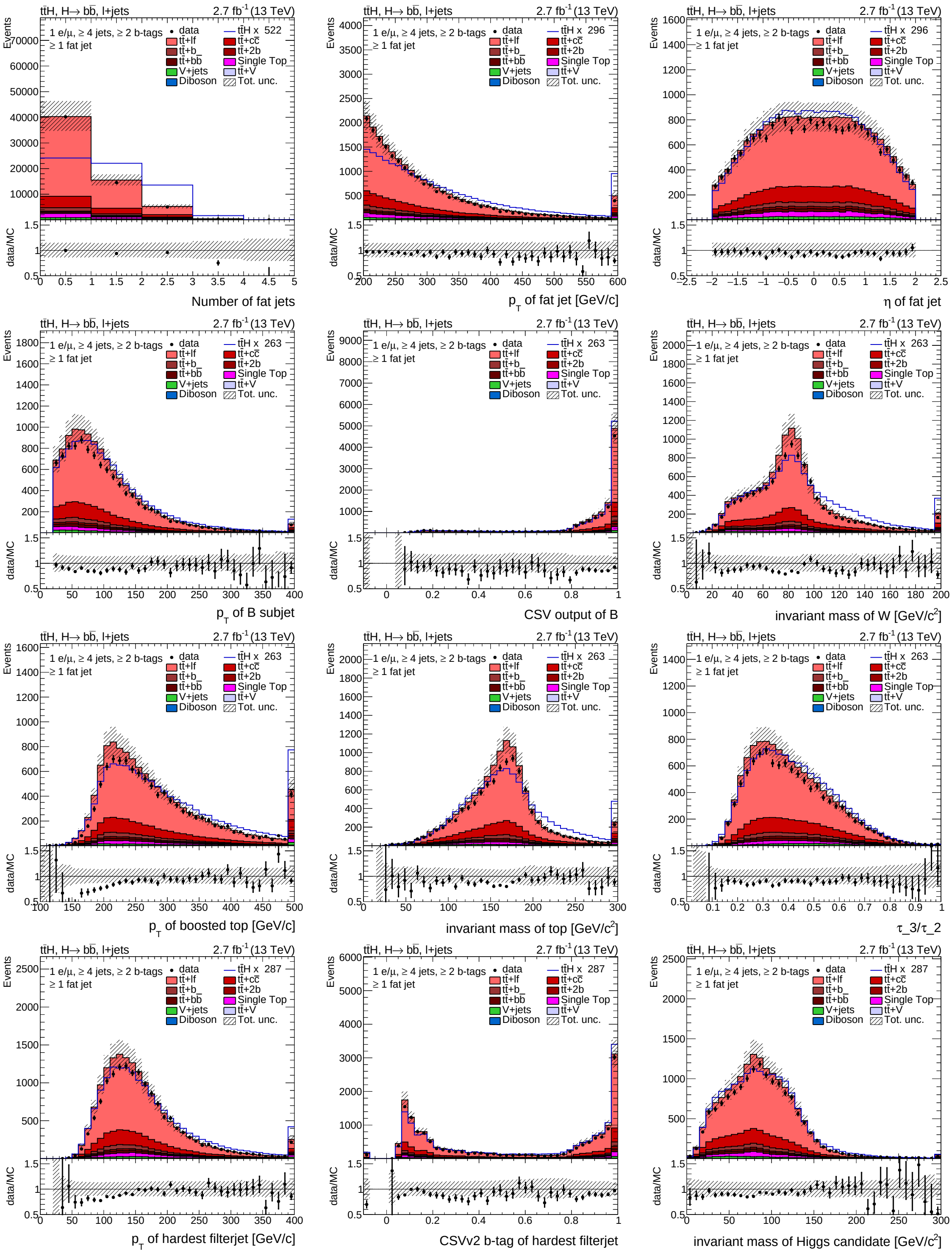








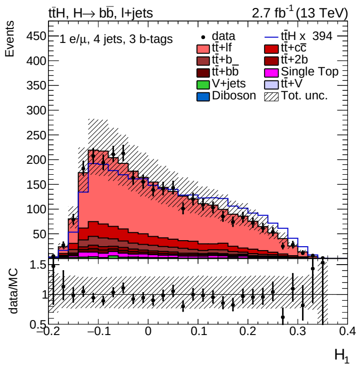
C/A 1.5 jets and substructure

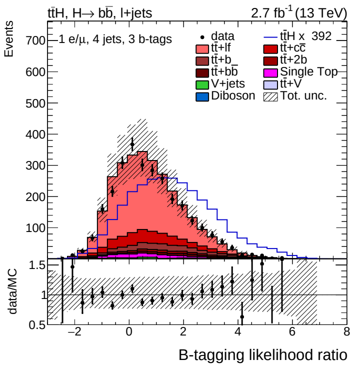


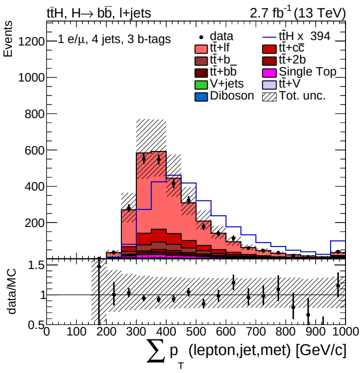
A.2 BDT inputs for the coupling-strength measurement

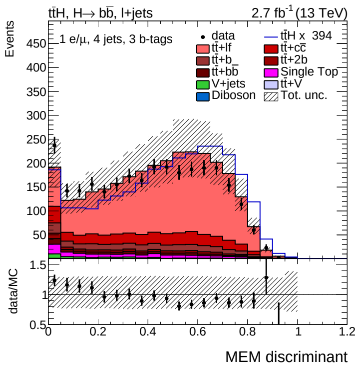
In this section, the BDT input variables for the coupling strength measurement are compared between observed and simulated data. For more details, see Section 9.5. The data recorded in 2015 (black dots) is compared with the expected background predicted by simulation (stacked histograms). The expected $t\bar{t}H$ signal is drawn as a blue line and scaled to the same integral as the sum of backgrounds. The most important systematic uncertainties are added in quadrature and shown as hashed error bands. Statistical uncertainties due to the limited number of observed events are drawn as black horizontal lines on the data points.

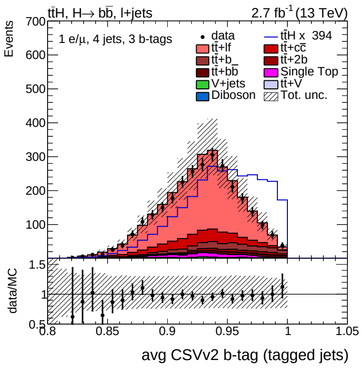
BDT inputs for signal-background separation in the 4 jets, 3 b-tags category

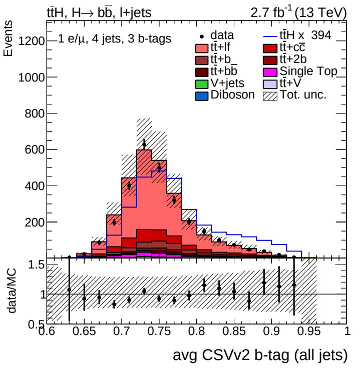


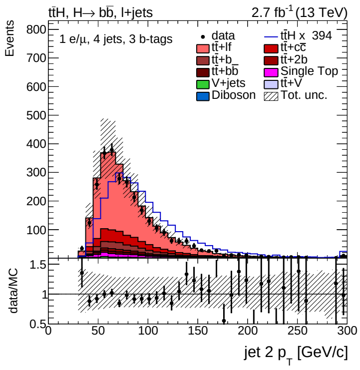


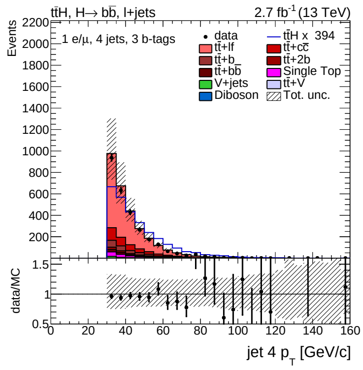




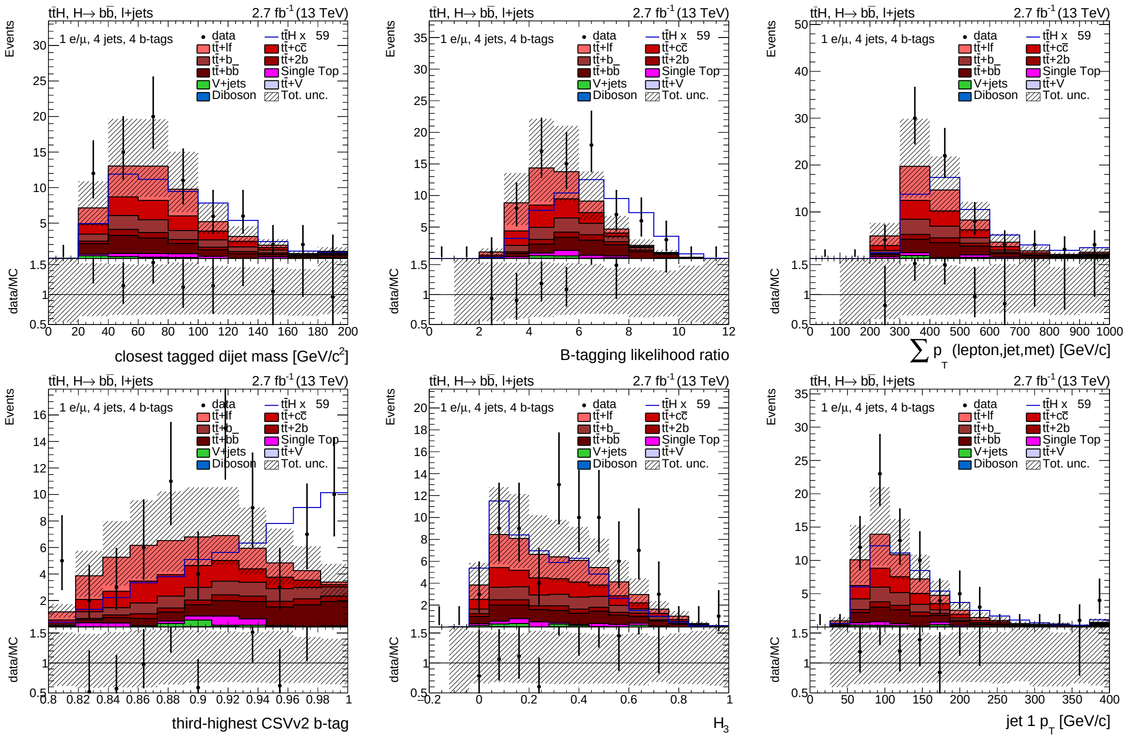




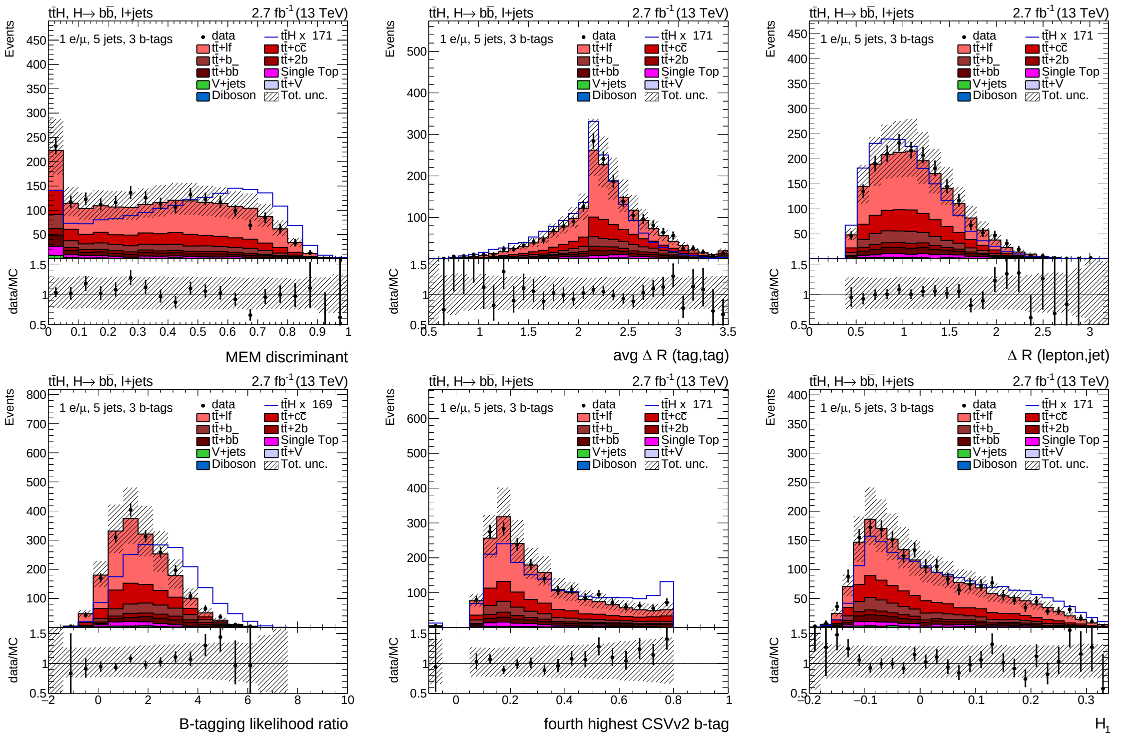


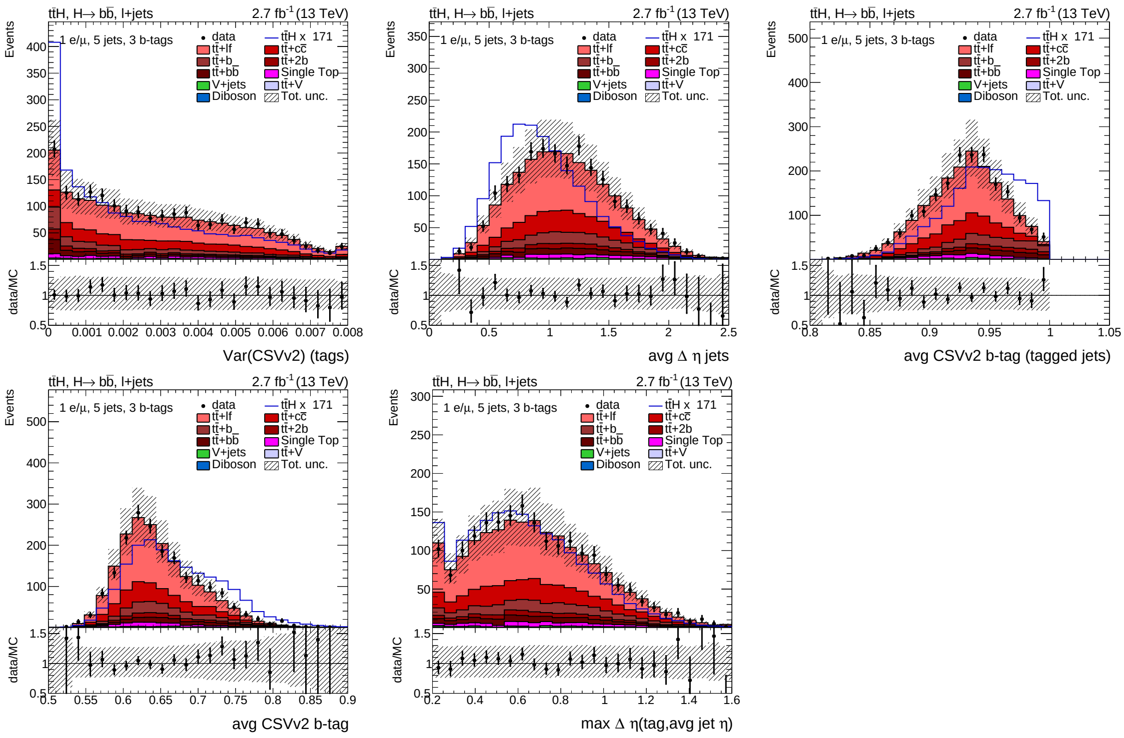


BDT inputs for signal-background separation in the $4\,$jets, $4\,$b-tags category

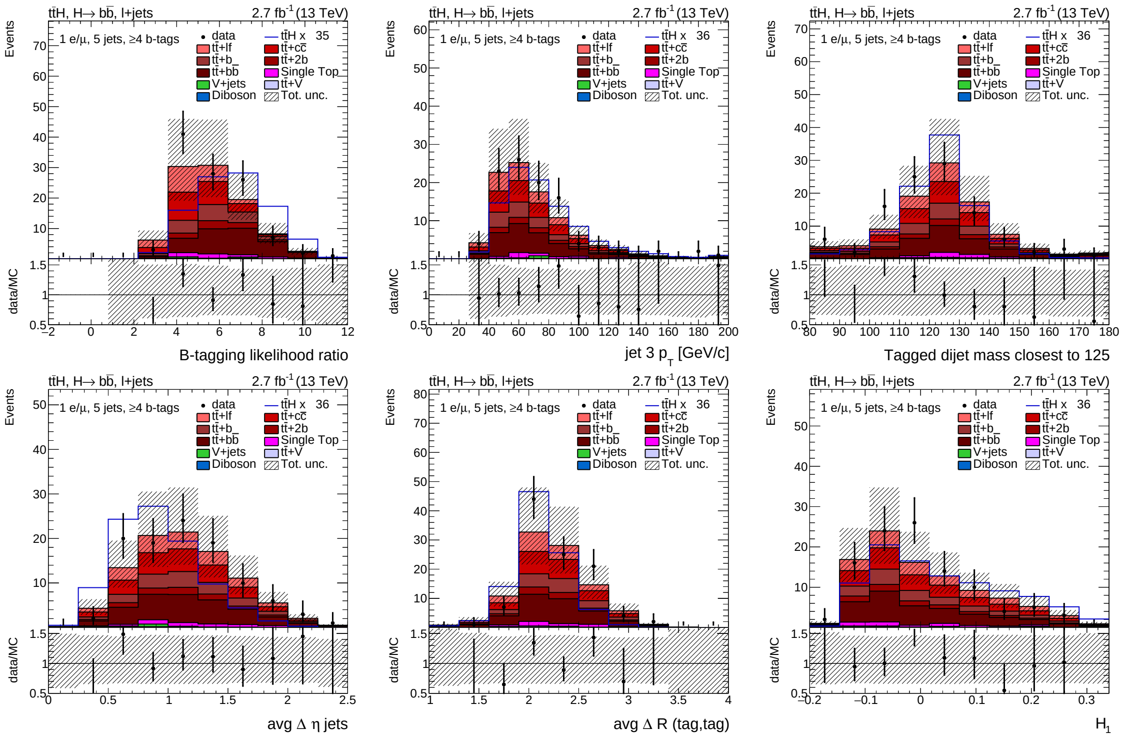


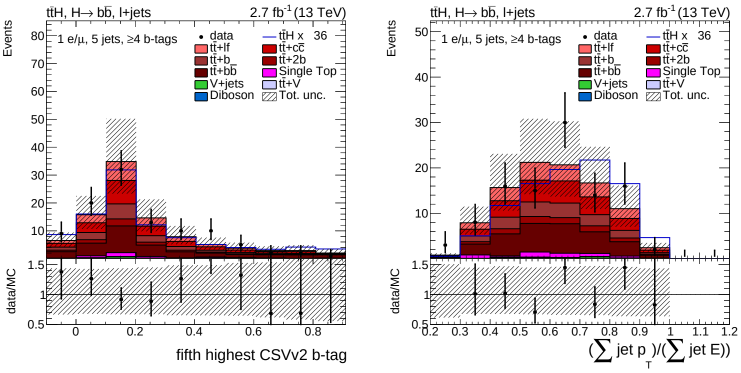
BDT inputs for signal-background separation in the $5\,$jets, $3\,$b-tags category



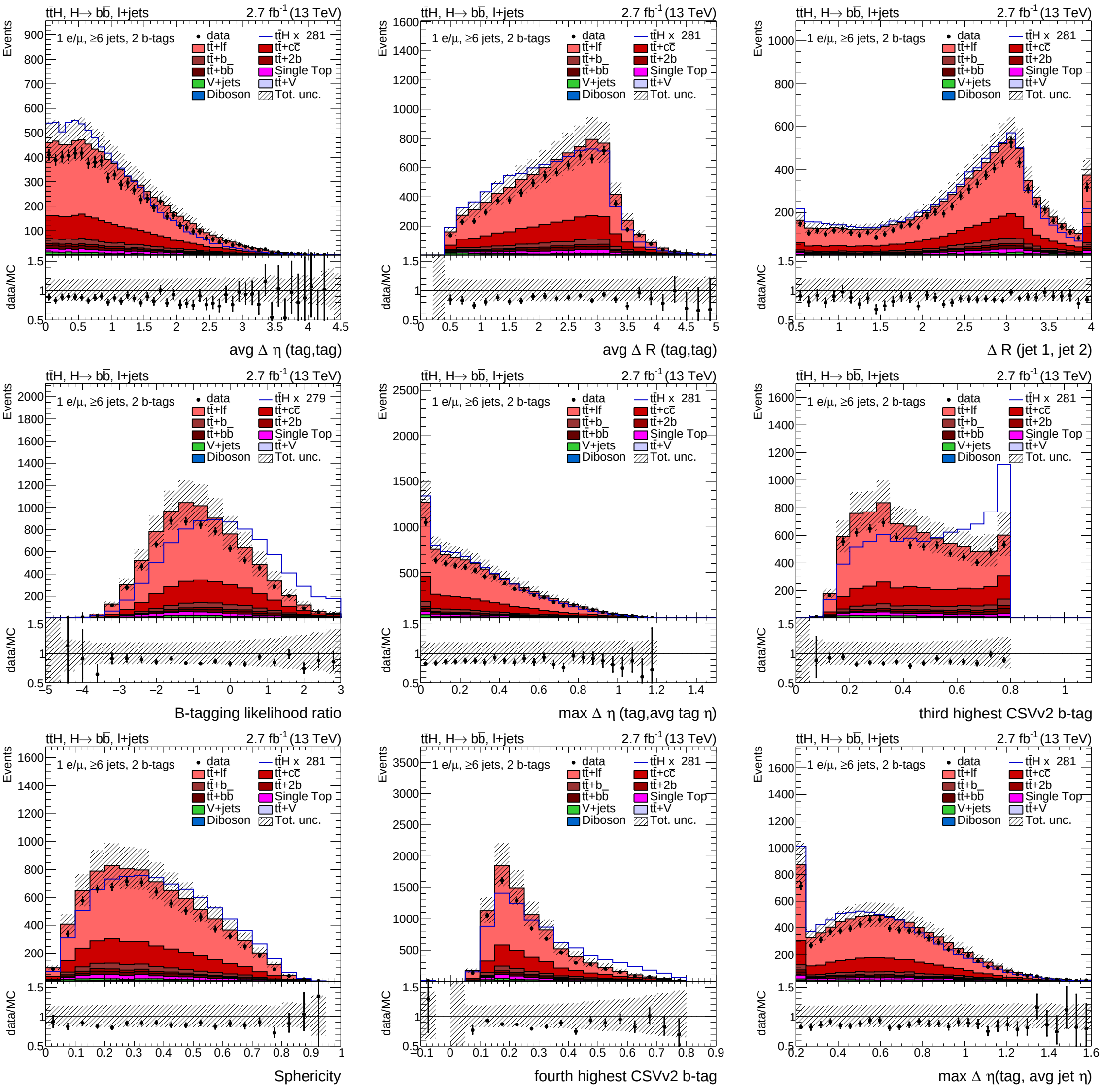


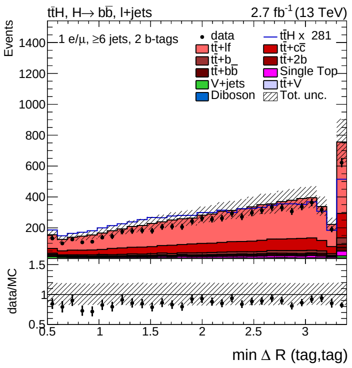
BDT inputs for signal-background separation in the 5 jets, ≥ 4 b-tags category



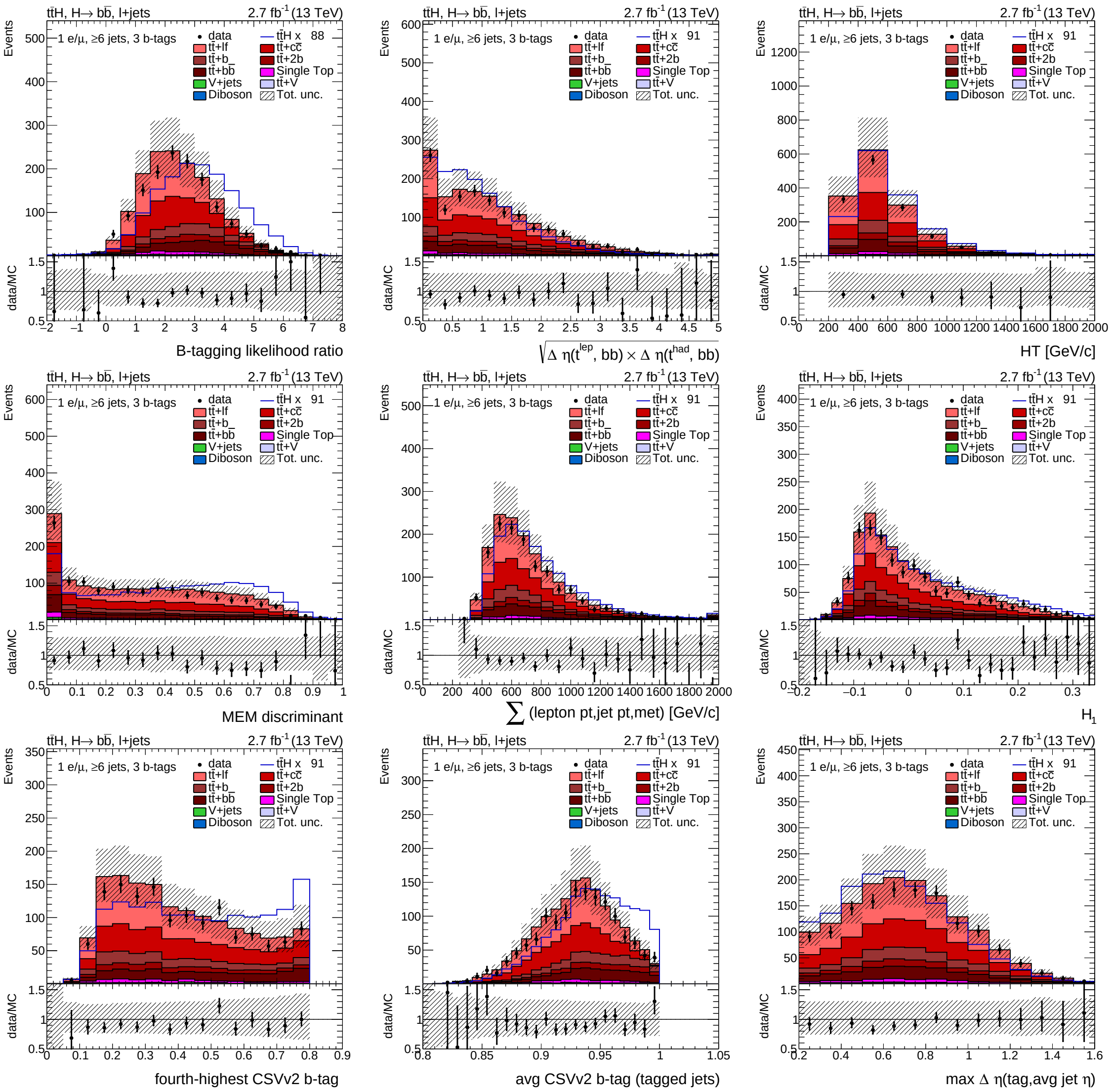


BDT inputs for signal-background separation in the ≥ 6 jets, 2 b-tags category

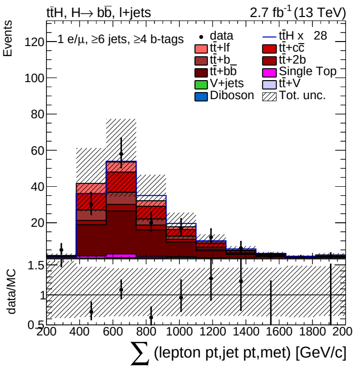


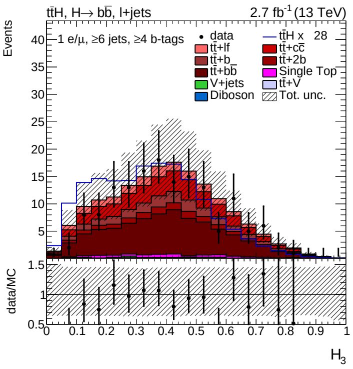


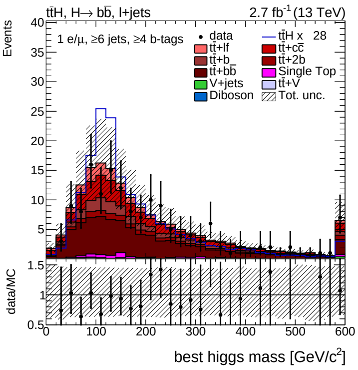
BDT inputs for signal-background separation in the ≥ 6 jets, 3 b-tags category

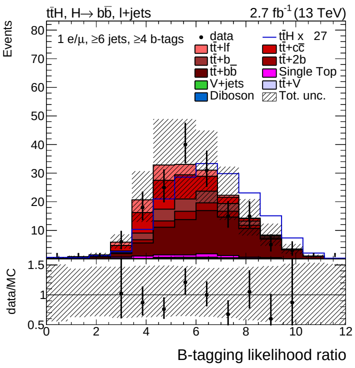


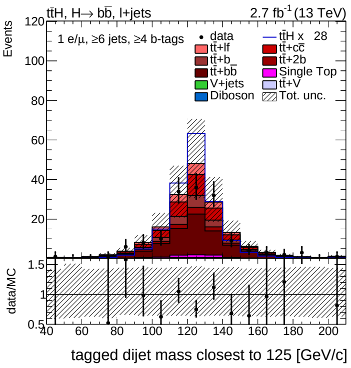
BDT inputs for signal-background separation in the ≥ 6 jets, ≥ 4 b-tags category

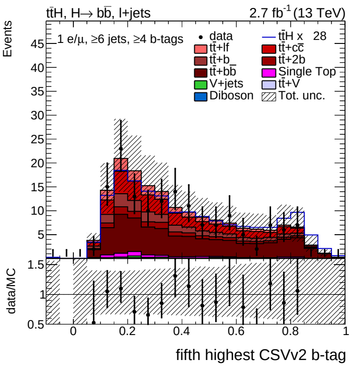


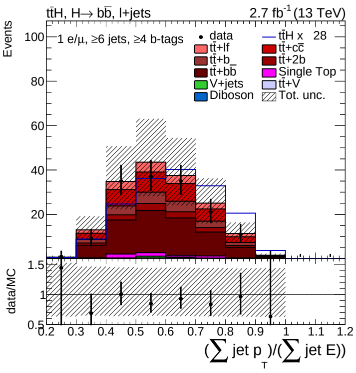


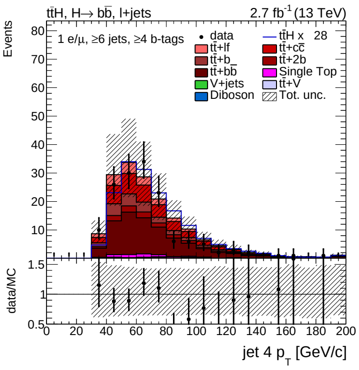


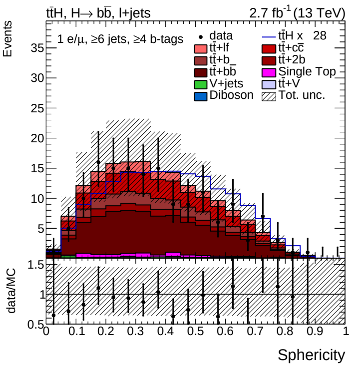


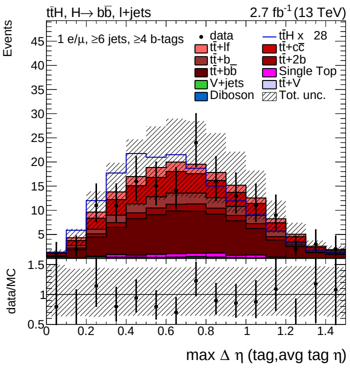


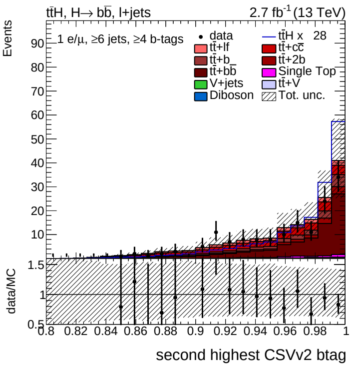




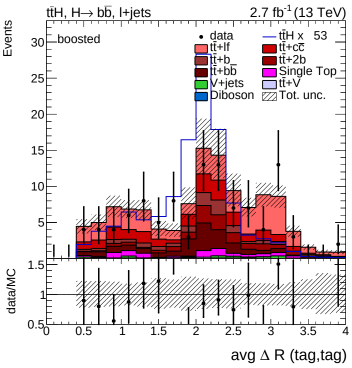


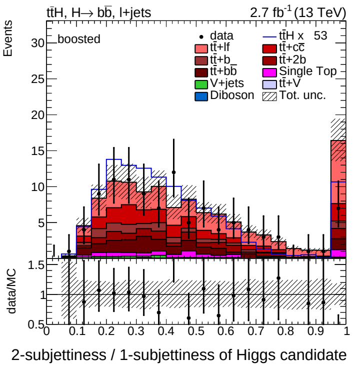


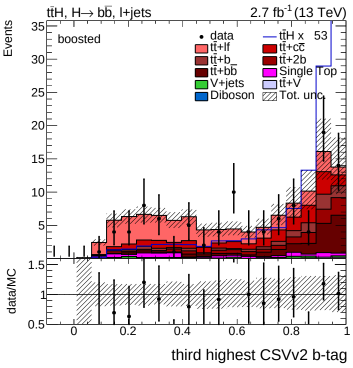


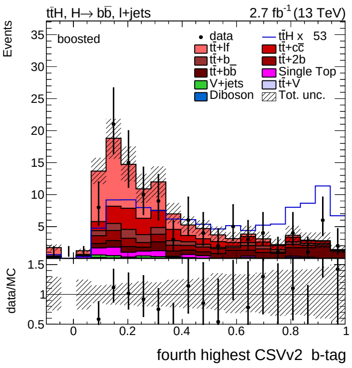


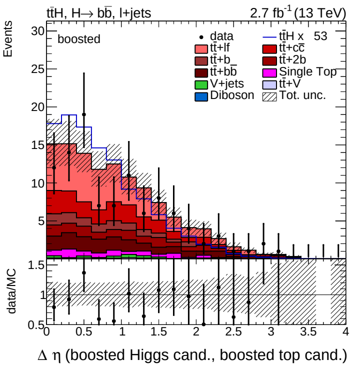
BDT inputs for signal-background separation in the “boosted” category

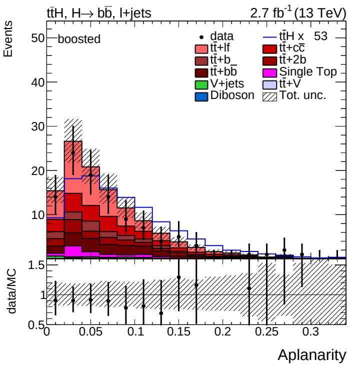


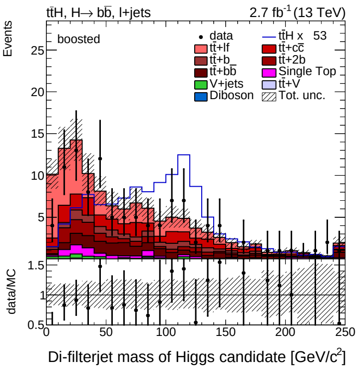


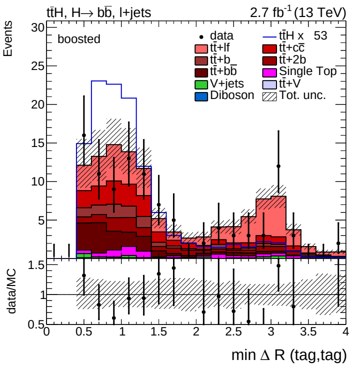


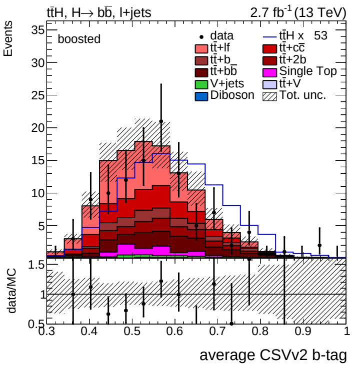


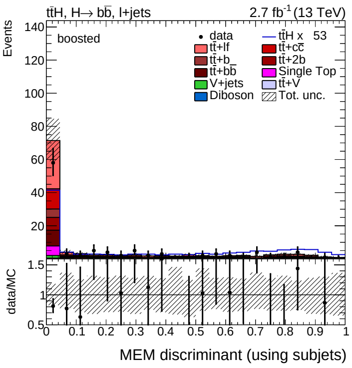


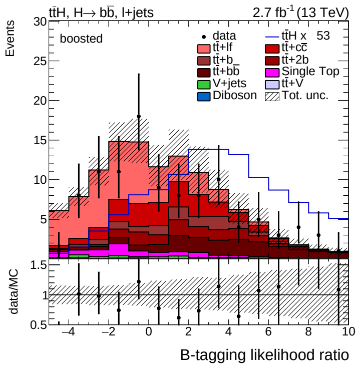








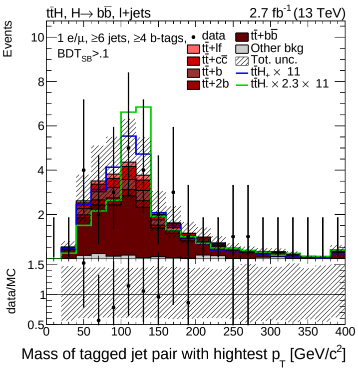


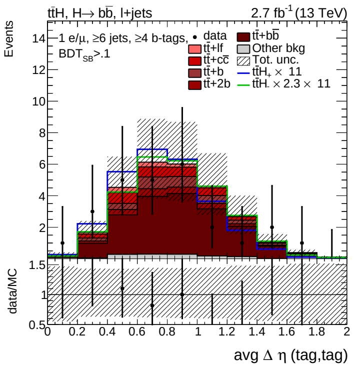


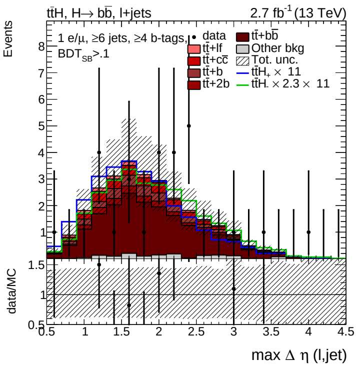
A.3 BDT inputs for the anomalous-coupling measurement

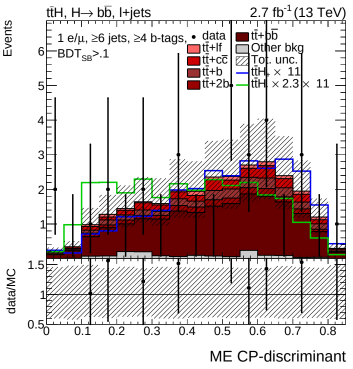
In this section, the input variables for the BDT that is sensitive to the $t\bar{t}H$ coupling structure are compared between observed and simulated data. The observed distribution of data (black dots) is compared with the expected background predicted by simulation (stacked histograms). The signal with pseudoscalar coupling (green line) is scaled to the same integral as the sum of backgrounds. The expected SM signal is drawn as a blue line and scaled by a factor of 1/2.3 less than the CP odd signal, to account for the difference in total cross section. The most important systematic uncertainties are added in quadrature and shown as hashed error bands. Statistical uncertainties are drawn as black horizontal lines on the data points. More details can be found in Section 11.2.

BDT inputs for the CP-discriminating BDT in the signal-enriched ≥ 6 jets, ≥ 4 b-tags category

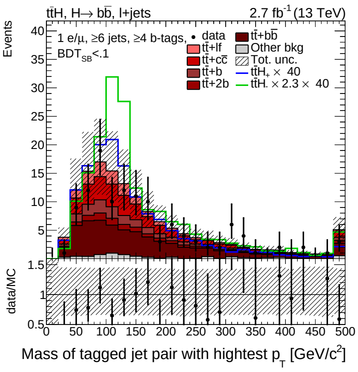


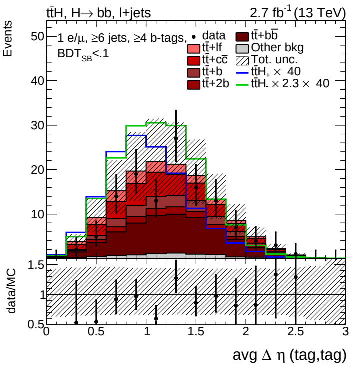


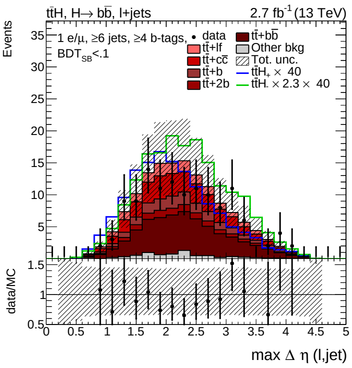


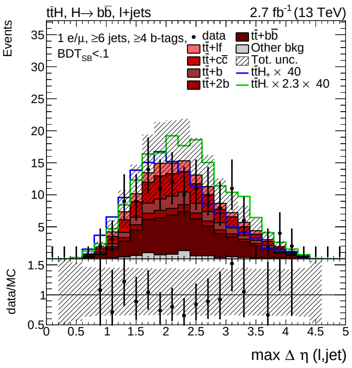


BDT inputs for the CP-discriminating BDT in the signal-depleted ≥ 6 jets, ≥ 4 b-tags category

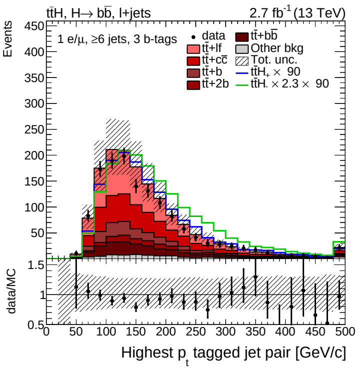


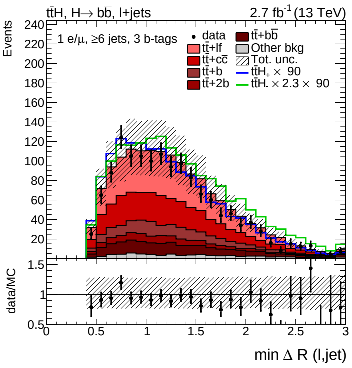


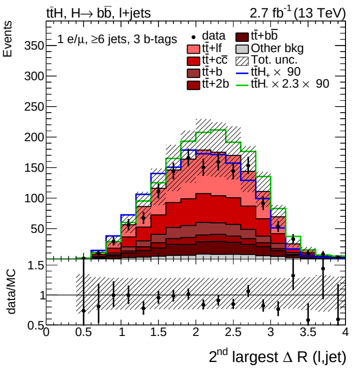


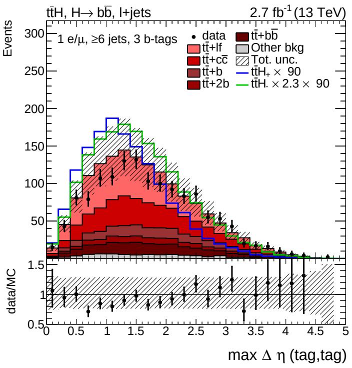


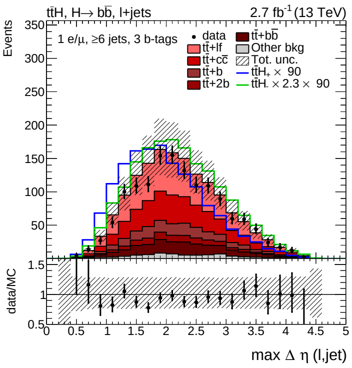
BDT inputs for the CP-discriminating BDT in the ≥ 6 jets, 3 b-tags category

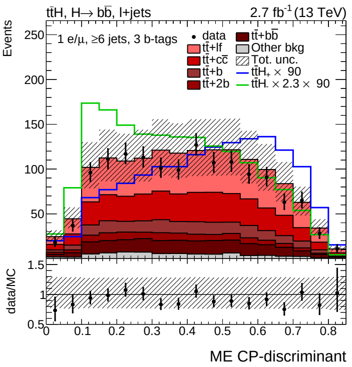




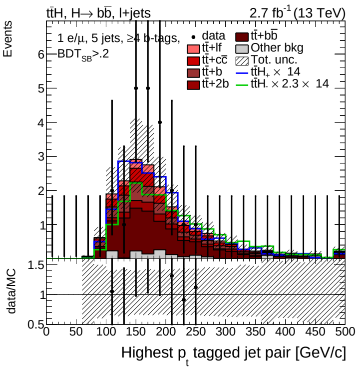


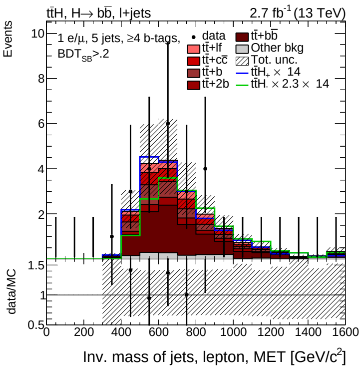


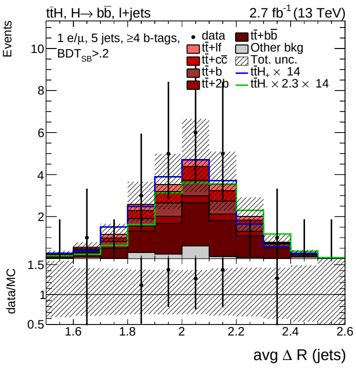


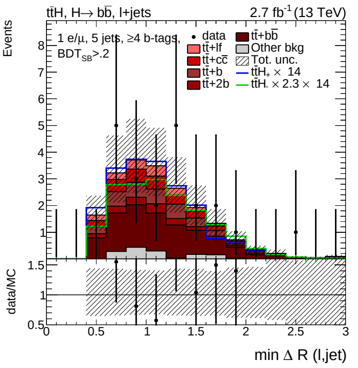


BDT inputs for the CP-discriminating BDT in the signal-enriched $5\,$jets, $\geq 4\,$b-tags category

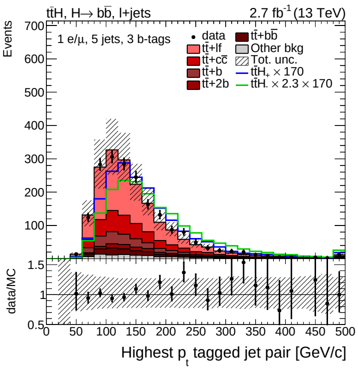


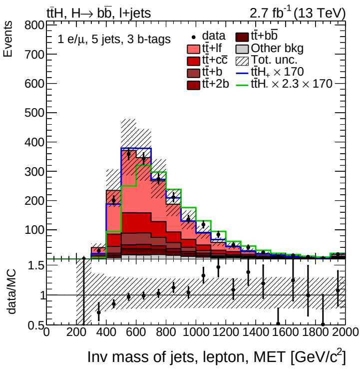


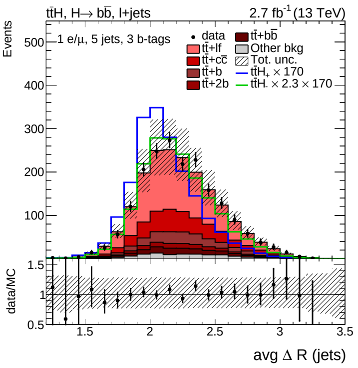


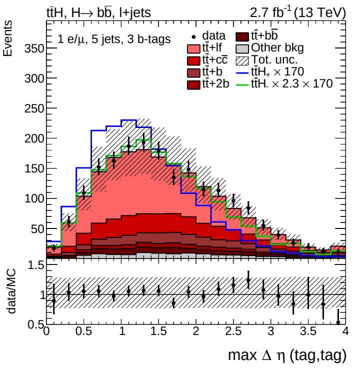


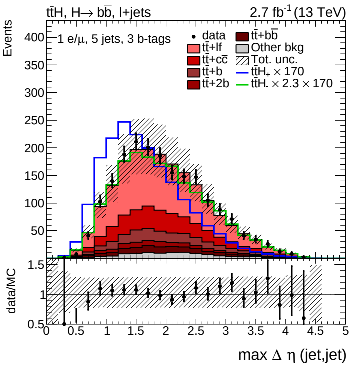
BDT inputs for the CP-discriminating BDT in the $5\,$jets, $3\,$b-tags category

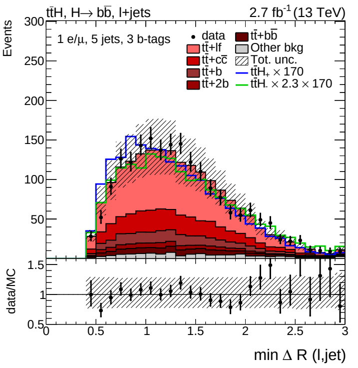




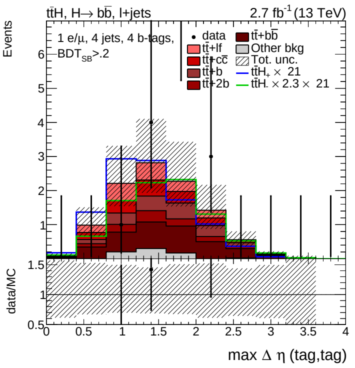


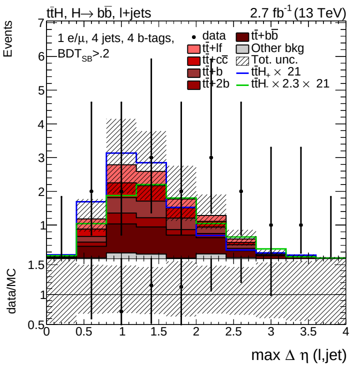


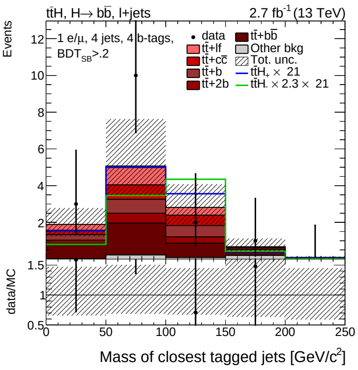


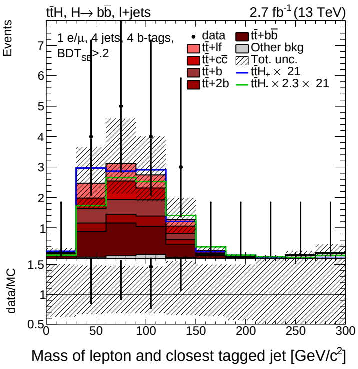


BDT inputs for the CP-discriminating BDT in the signal-enriched $4\,$jets, $4\,$b-tags category

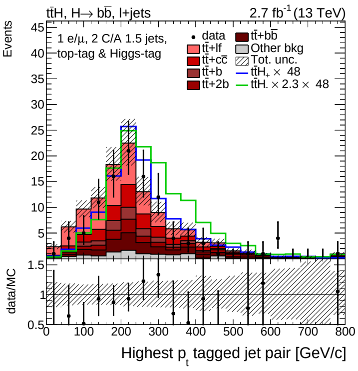


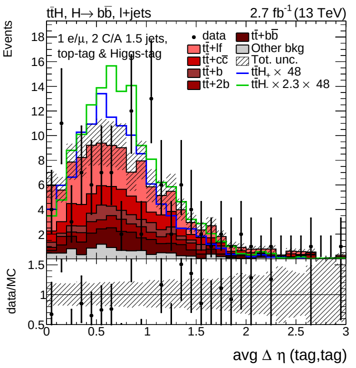


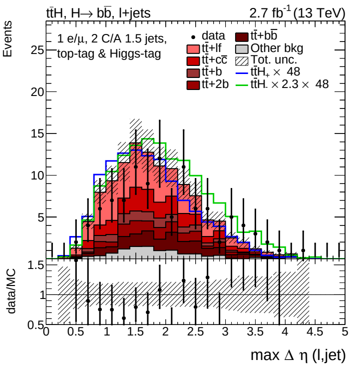


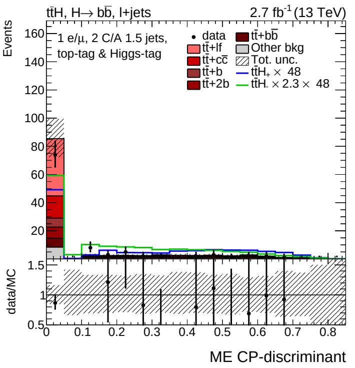


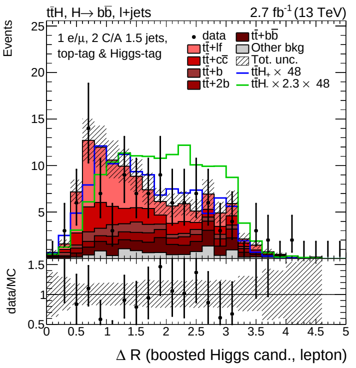
BDT inputs for the CP-discriminating BDT in the “boosted” category

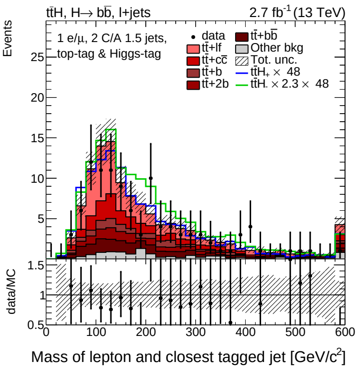






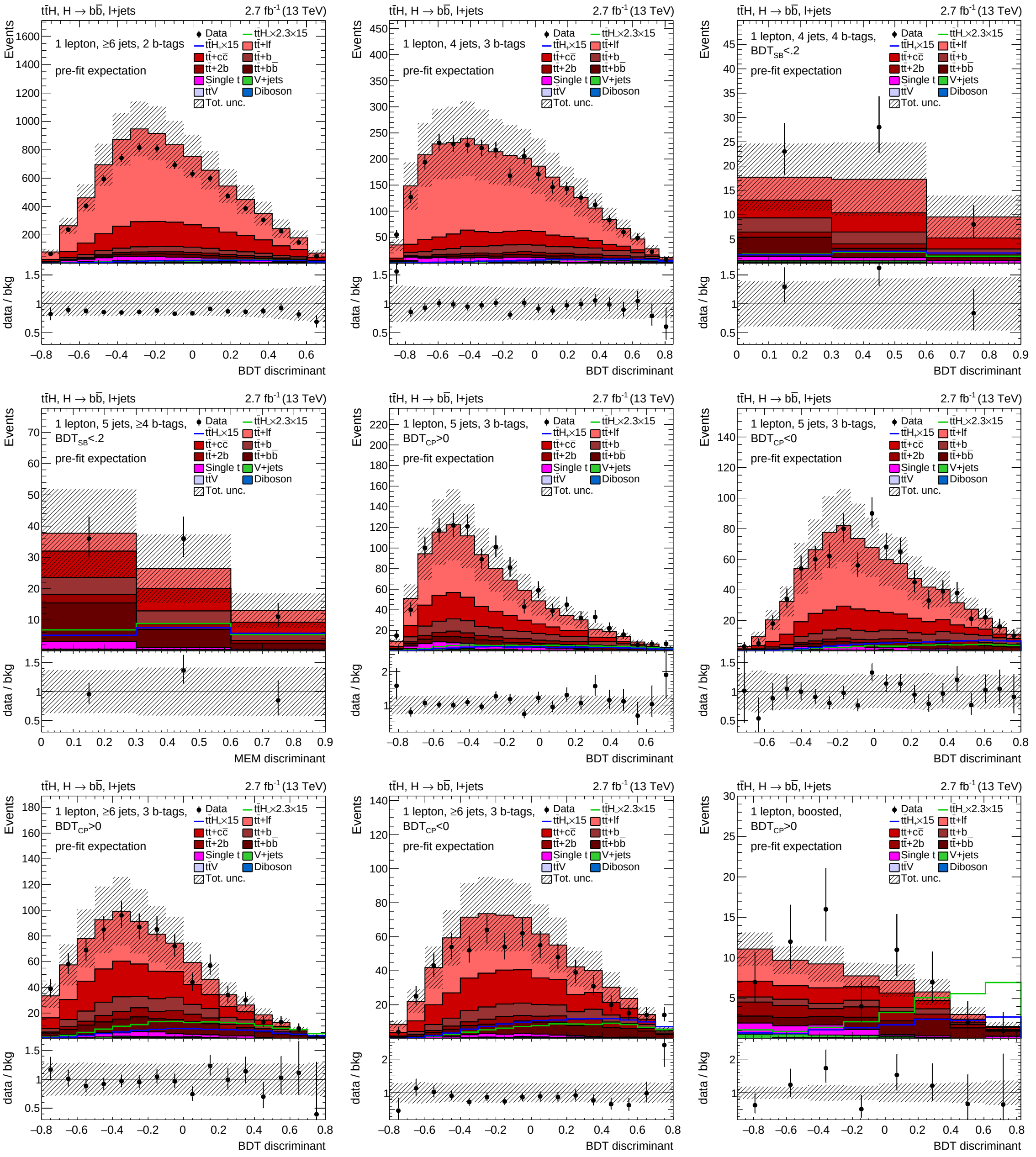


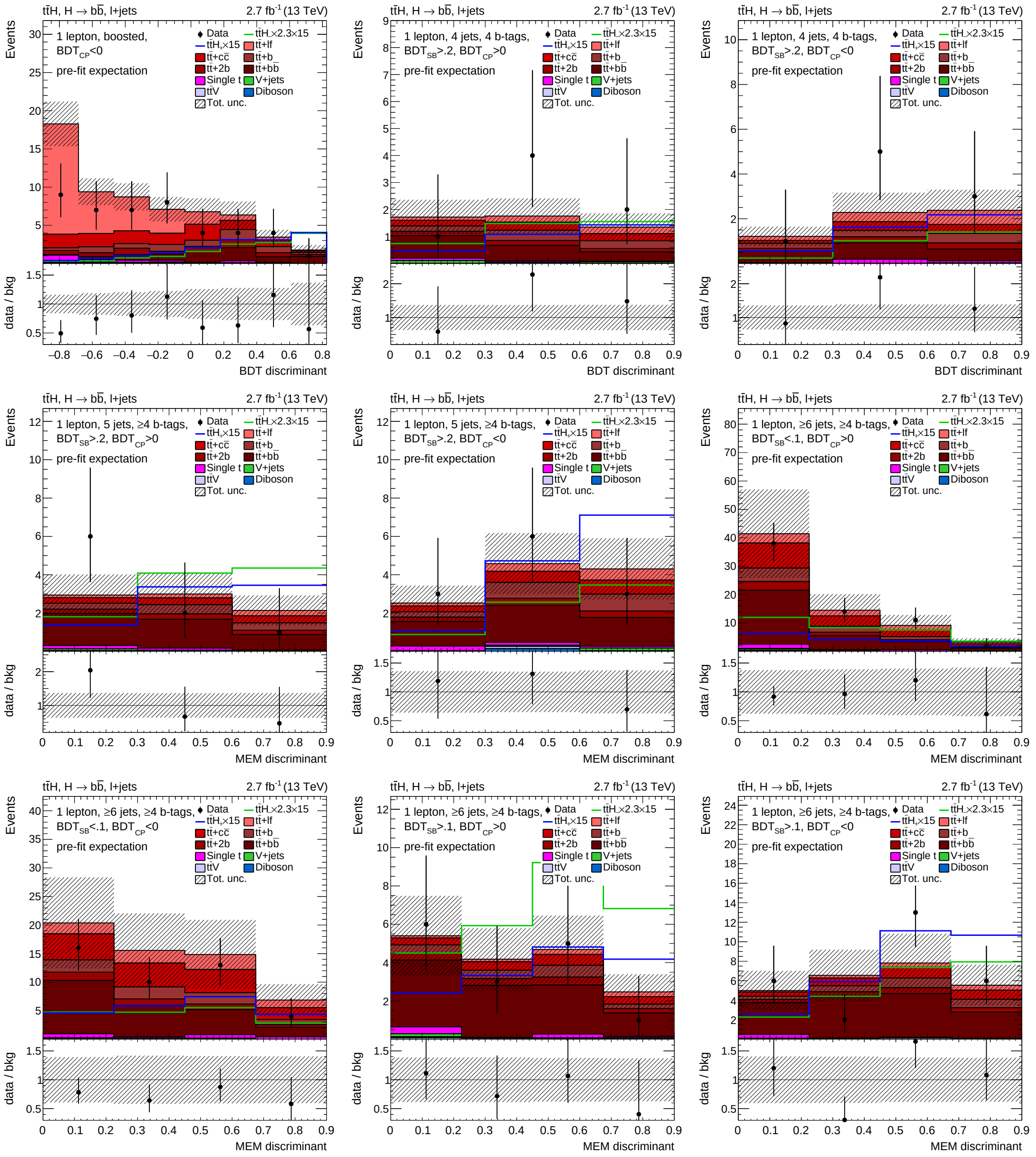




A.4 Prefit distribution for the anomalous-coupling measurement

Distribution of the data collected in 2015 in the categories of the coupling structure analysis, compared to the prefit model. The expected SM $t\bar{t}$H signal is drawn as a blue line and scaled by a factor of 15 w.r.t. the SM expectation for better visibility. The signal for a pseudscalar coupling is drawn as a green line and scaled by an additional factor of 2.3 to account for the lower cross section. The hashed error bands correspond to the effect of all systematic uncertainties added in quadrature. More details can be found in Section 11.2.




tt̄H, H → bb̄, l+jets
2.7 fb⁻¹ (13 TeV)
1 lepton, boosted, BDTCP<0
pre-fit expectation
1 lepton, 4 jets, 4 b-tags, BDTSB>.2, BDTCP>0
1 lepton, 4 jets, 4 b-tags, BDTSB>.2, BDTCP<0
1 lepton, 5 jets, ≥4 b-tags, BDTSB>.2, BDTCP>0
1 lepton, 5 jets, ≥4 b-tags, BDTSB>.2, BDTCP<0
1 lepton, ≥6 jets, ≥4 b-tags, BDTSB<.1, BDTCP>0
1 lepton, ≥6 jets, ≥4 b-tags, BDTSB<.1, BDTCP<0
1 lepton, ≥6 jets, ≥4 b-tags, BDTSB>.1, BDTCP>0
1 lepton, ≥6 jets, ≥4 b-tags, BDTSB>.1, BDTCP<0
Data
ttH×2.3×15
ttH×15
tt+lf
tt+cc̄
tt+b
tt+2b
tt+bb̄
Single t
V+jets
ttV
Diboson
Tot. unc.
Events
data / bkg
BDT discriminant
MEM discriminant

Acknowledgements

First I would like to thank my supervisor Prof. Dr. Ulrich Husemann for giving me the opportunity to write my thesis in his group, for thoroughly reading my work, for many helpful comments, and for his support. I am also grateful for the funding I received with his help and for the great group he created at KIT Campus North. I would like to thank Prof. Dr. Günter Quast for being my co-supervisor and for his encouragement.

I am particularly thankful for the great collaboration with Shawn Williamson, Karim El Morabit, and Dr. Matthias Schröder, which made the hard work and late-night shifts for the $\mathrm{t\bar{t}H}$ analysis not only bearable but even fun. Alexis Descroix I would like to thank for the practical introduction to particle physics analysis. I am thankful to all members of the CMS Campus North group, who contributed to the great atmosphere that allowed us to accomplish a lot over the last years while also having a lot of fun.

For coordinating the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis and giving a lot of helpful input I wish to thank Dr. Darren Puigh. I am grateful for the help received by Dr. Andrew Gilbert in running the CMS Higgs fit.

Finally, I would like to thank my parents, Karla and Robert, and my family who always supported me and without whom this dissertation would not have been possible. My girlfriend Julia I want to thank for her support and our time together in Stuttgart.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und unter ausschließlicher Verwendung der angegebenen Hilfsmittel angefertigt habe.

Karlsruhe, den 8. November 2016,

Hannes Mildner