ETP-KA/2020-31

Measurement of $t\bar{t}H(H \to b\bar{b})$ production in the semi-leptonic $t\bar{t}$ decay channel at the CMS Experiment

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

M.Sc. Karim El Morabit
aus Freiburg im Breisgau

Tag der mündlichen Prüfung: 30. Oktober 2020

Referent: Prof. Dr. U. Husemann
Institut für Experimentelle Teilchenphysik

Korreferent: Prof. Dr. Th. Müller
Institut für Experimentelle Teilchenphysik

Contents

1 Introduction

The field of particles physics aims to understand the fundamental particles in our universe and the manner in which they interact. The most successful theory for their description is called the standard model of particle physics. Developed, tested and refined over many decades, it is not only able to describe the known particles but was also able to predict the existence of, as of then, undiscovered particles, which were all later observed. For a long time, the last of these undiscovered particles was the Higgs boson, whose prediction is a consequence of a mechanism to explain the observed masses of the particles. It was finally discovered in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN [1–3]. In the Higgs mechanism, the fundamental particles acquire their mass through interaction with the Higgs field. However, not all of these interactions have been directly observed yet, leaving open the questions whether the mechanism is the whole source of the particle masses and whether the interaction has the predicted structure. And although, with the observation of the Higgs boson, the standard model is now complete, there are many phenomena in nature that are not explained by the standard model. This suggests that the standard model in its current form has to be extended or superseded by a more fundamental theory of nature.

To study the properties of particles or to discover new particles, experiments have to be conducted. The Large Hadron Collider at CERN is currently the most powerful particle collider, allowing access to the highest energies ever achieved in a laboratory setting. With it, two proton beams are accelerated up to energies of 6.5 TeV and subsequently brought to collision to conduct scattering experiments. The particles created in these collisions are measured and analyzed by the ALICE, ATLAS, CMS and LHCb experiments.

In the standard model, the interaction between the Higgs boson and fermions is a Yukawa interaction with a coupling strength proportional to the mass of the interacting fermions. Since the top quark is the heaviest known particle, processes involving the interaction between a Higgs boson and top quarks are of special interest for a study of the Yukawa interaction. While the decay of a Higgs boson into a top quark-antiquark pair is not possible, the Yukawa coupling can be accessed by studying the production of a Higgs boson in association with one or more top quarks. The cross section for the production of a Higgs boson in association with a top quark-antiquark pair ($t\bar{t}H$) is directly related to the absolute value of the Yukawa coupling constant. A Higgs boson can also be produced in association with a single top quark (tH). Although the cross section for tH production is much smaller than that for $t\bar{t}H$ production, it is sensitive to the sign of the Yukawa

coupling constant relative to the sign of the coupling constant for the interaction between a Higgs boson and vector bosons.

In this thesis, data recorded with the Compact Muon Solenoid (CMS) experiment at the LHC are analyzed to study the production of $t\bar{t}H$ with the subsequent decay of the Higgs boson into a bottom quark-antiquark pair ($t\bar{t}H(b\bar{b})$) and a semi-leptonic decay of the top quark-antiquark pair in proton-proton collisions at a center-of-mass energy of 13 TeV. To measure the production cross section of this process, compatible events are selected from data corresponding to an integrated luminosity of $77.4\,\mathrm{fb}^{-1}$. Top quarks almost always decay to a bottom quark and a W boson, which in turn decays to either a quark-antiquark pair or a lepton and a neutrino. Events are therefore required to contain one lepton and multiple jets, several of which are required to be compatible with originating from bottom quarks. The sample of selected events contains a large background of top quark-antiquark pair production with an irreducible component consisting of top quark-antiquark pairs produced in association with bottom quark-antiquark pairs. To increase the sensitivity to $t\bar{t}H(b\bar{b})$ production, phase space regions enriched in signal events are constructed using multivariate analysis methods. The numbers of events observed in these regions are compared to predictions made by the standard model to extract the cross section for $t\bar{t}H(b\bar{b})$ production relative to its standard model prediction. The analysis in the semi-leptonic decay channel of the top quark-antiquark pair is subsequently combined with analyses performed in other decay channels of the top quark-antiquark pair to further increase the sensitivity to $t\bar{t}H(b\bar{b})$ production. The results of the analysis had already been published prior to this thesis, albeit in less detail, in Refs. [4, 5].

In Chapter 2, the necessary theoretical foundation is established. The standard model and the Higgs mechanism are introduced, the prediction of observables at the LHC is discussed and the relevant production and decay modes of the Higgs boson and the top quark are discussed. In Chapter 3, a brief introduction to the LHC at CERN and the CMS experiment is given. The statistical methodology for the comparison of recorded data to simulated data and the calculation of the results of the analysis is discussed in Chapter 4. This is followed by a description of the multivariate analysis methods relevant to this thesis in Chapter 5. In Chapter 6, the reconstruction, calibration and identification of physics objects, such as jets and leptons, are discussed. Furthermore, the identification of b jets, the triggers used for the selection of events and corrections applied to simulated data are discussed. The details of the $t\bar{t}H(b\bar{b})$ analysis are discussed in Chapter 7. This includes a discussion of the event selection, the samples of recorded and simulated data, as well as the separation of signal and background events with the help of multivariate methods. The systematic uncertainties affecting the analysis are discussed in Chapter 8. The results of the analysis are presented in Chapter 9. First, results of the $t\bar{t}H(b\bar{b})$ analysis in the semi-leptonic decay channel of the top quark-antiquark pair are discussed, followed by a discussion of the combination with the analyses in the other decay channels. Subsequently, combinations with other Higgs boson analyses performed by the CMS Collaboration and results of other analyses of the top-Higgs Yukawa coupling are discussed. In Chapter 10, projections for future $t\bar{t}H(b\bar{b})$ analyses at the planned High Luminosity LHC are performed and studies, performed by students at the Karlsruhe Institute of Technology (KIT), on the improvement of the $t\bar{t}H(b\bar{b})$ analysis are discussed. In the final Chapter 11, a summary of the analyses is given and conclusions are drawn.

2 Theoretical foundations

The best known description of the fundamental particles and their interactions is given by the standard model of particle physics (henceforth only called the standard model or SM). It is an experimentally well tested theory and, as of today, no significant deviations from the predictions made by the SM have been observed. Several particles have been first predicted by the standard model for reasons of mathematical consistency and have been experimentally observed only later. One of these particles is the Higgs boson, which is studied in this thesis.

Section 2.1 aims to give a brief overview over the standard model, motivate the introduction of the Higgs boson and introduce important concepts. In this thesis, the production of a Higgs boson in association with a top quark-antiquark pair ($t\bar{t}H$) is studied using scattering experiments performed at the LHC. A brief description about the calculation of cross sections for processes occurring at hadron colliders such as the LHC will be given in Sec. 2.2. The data produced in scattering experiments is usually compared to theoretical predictions with the help of simulated events. The generation of simulated data will be discussed in Sec. 2.3. For the analysis of the $t\bar{t}H$ process it is necessary to understand how top quarks and Higgs bosons can be produced at the LHC and how they decay. A discussion about top quark processes at the LHC will be given in Sec. 2.4. This is followed by a discussion about Higgs boson production and decay modes in Sec. 2.5.

Throughout this thesis the common convention $c = \hbar = 1$ is followed.

2.1 The Standard Model

The standard model of particle physics (SM) is a theory describing all known fundamental particles and their interactions. The particles are classified into two groups depending on their spin: fermions and bosons. Fermions have half-integer spin and will be discussed in the first part of this section. The particles of the standard model interact through four fundamental forces. These are the gravitational force as well as the electromagnetic, weak and strong interactions. In the SM the electromagnetic, weak and strong interactions are mediated by gauge bosons and mathematically described in a unified theoretical framework. The gauge bosons have a spin of one and will be introduced in the second part of this section. The fourth force, gravity, can, as of today, not be described in the same theoretical framework as the other interactions. The standard model therefore only describes the electromagnetic, weak and strong interactions. The effect of gravitation on the fundamental

particles can be neglected at the energies accessible with currently existing experiments. The discussion of the history and properties of the fermions and bosons is based on Ref. [6]. The mathematical formulation of the SM is in the form of quantum field theories (QFTs). In this framework, particles are represented by quantum fields, quantized excitations of which are interpreted as individual particles. The interactions of the fields are expressed in the Lagrangian density of the theory with which the dynamics of the fields can be derived. In the third part of this section, a brief overview of the QFT describing the electromagnetic, weak and strong interactions is given. As will be seen, the unified description of the electromagnetic and weak interactions will give rise to an additional spin-0 boson: the Higgs boson.

2.1.1 Fermions

The fermions of the SM consist of six quarks, three charged leptons and three neutrinos. They are summarized in Tab. 2.1 and differ in their masses as well as in the types of charges they carry. For each fermion there exists a corresponding antiparticle with the same mass but opposite charges. All particles carrying electric charge interact electromagnetically. Particles carrying color charge interact via the strong interaction. All left-handed fermions and right-handed antifermions carry the charge of the weak interaction called weak isospin. For the right-handed fermions and left-handed antifermions the weak isospin is zero and they do not interact via the weak interaction. The quarks as well as the charged leptons and neutrinos are grouped into three generations based on their behavior with regard to the weak interaction. Charged leptons and neutrinos only interact weakly with charged leptons and neutrinos of the same generation and the weak interaction between quarks of the same generation is stronger than that between quarks of different generations.
For unknown reasons the masses of the fermions differ drastically. The top quark, for example, has a mass of about 173 GeV [7] and is about 170 times heavier than the proton, whereas the masses of the neutrinos are very small. For these, only upper limits can be calculated based on the observation of neutrino oscillations, described in Refs. [8,9] or direct measurements such as for example with the KATRIN experiment, described in Ref. [10], in which an upper limit of 1.1 eV at the 90% confidence level was found for the lightest neutrino.

Only the fermions of the first generation are stable. Those of higher generations ultimately decay to stable fermions. Furthermore, quarks cannot be observed as free particles and always form bound states called hadrons. The most stable ones of these are the proton made out of two up quarks and one down quark, and the neutron made out of one up quark and two down quarks. For this reason, ordinary macroscopic matter consists of protons, neutrons and electrons. Not surprisingly protons, neutrons and electrons were therefore, together with the photon, the first particles that had been known. Later, the electron neutrino was added to explain the energy spectrum observed in β-decays. The list of known particles was further extended by the muon, muon neutrino and the pion, which were all discovered in astroparticle experiments. Following this, a whole "zoo" of new particles was discovered. It was observed that the properties of these particles could be explained if they were bound states of either two or three undiscovered particles with fractional electric charge, which were called up quark, down quark and strange quark as discussed in Refs. [11, 12]. The existence of quarks was later confirmed by deep-inelastic scattering experiments, described in Ref. [13]. The charm quark was postulated to explain observed anomalies of flavor-changing neutral currents and was later discovered by two independent experiments, which is described in Refs. [14–17]. This was followed by the discovery of the τ-lepton, described in Ref. [18] and the bottom quark, described in Ref. [19]. Then in the year 1995 the top quark was discovered by the CDF and DØ Collaborations at

Table 2.1: The three generations of fermions of the standard model. The electric charge, weak isospin T_3 and color determine by which interactions they interact. Only left-handed fermions and right-handed antifermions carry $T_3 \neq 0$. The right-handed fermions and left-handed antifermions are isospin singlets with $T_3 = 0$. For masses without indicated uncertainties, the corresponding uncertainties were too small to reasonably be displayed in this table. Information taken from Ref. [7].

particle	gen.	electric charge	T_3	color	mass
up (u)	1	2/3	1/2	r,g,b	$2.2^{+0.5}_{-0.3}$ MeV
down (d)	1	−1/3	−1/2	r,g,b	$4.7^{+0.5}_{-0.2}$ MeV
electron neutrino (ν_e)	1	0	1/2	none	< 1.1 eV
electron (e)	1	−1	−1/2	none	0.511 MeV
charm (c)	2	2/3	1/2	r,g,b	1.27 ± 0.02 GeV
strange (s)	2	−1/3	−1/2	r,g,b	93^{+11}_{-5} MeV
muon neutrino (ν_μ)	2	0	1/2	none	< 0.19 MeV
muon (μ)	2	−1	−1/2	none	105.7 MeV
top (t)	3	2/3	1/2	r,g,b	172.76 ± 0.30 GeV
bottom (b)	3	−1/3	−1/2	r,g,b	$4.18^{+0.03}_{-0.02}$ GeV
tau neutrino (ν_τ)	3	0	1/2	none	< 18.2 MeV
tau (τ)	3	−1	−1/2	none	1776.86 ± 0.12 MeV

the Tevatron collider as described in Refs. [20,21]. The last known fermion to be discovered was the τ-neutrino, as discussed in Ref. [22].

2.1.2 Gauge bosons and interactions

The particles of the standard model interact via the electromagnetic, weak and strong interactions which are mediated by so-called gauge bosons with spin 1. The properties of the gauge bosons are summarized in Tab. 2.2.

The electromagnetic interaction is described by a relativistic quantum field theory called quantum electrodynamics (QED), which is described in Refs. [23–25]. It is mediated by the massless photon. Since the photon is massless and does not carry electromagnetic charge itself, the range of the electromagnetic interaction is infinite.

The weak interaction is mediated by the Z boson and the $W^\pm$ bosons. The charge of the weak interaction is called weak isospin T_3 and depends on the chirality, i.e. the "the handedness", of the fermions, a phenomenon called parity violation, which was observed in Ref. [26]. The chirality of a fermion is defined by the $1 \pm \gamma_5$ operator and for particles moving at the speed of light it is identical to the helicity which is the projection of the spin on the direction of the momentum. Only left-handed fermions and right-handed antifermions carry weak isospin $T_3 \neq 0$ and can interact with the $W^\pm$ and Z bosons. Therefore the weak interaction also separately violates the charge symmetry. For example, under the parity transformation, a left-handed fermion, which carries weak isospin and therefore interacts via the weak interaction, turns into a right-handed fermion which does not carry weak isospin. Similarly, under the charge transformation the left-handed fermion turns into a left-handed antifermion, which also does not carry weak isospin. Under the

simultaneous transformation of charge and parity, however, a left-handed fermion turns into a right-handed antifermion, which does carry weak isospin and interacts in the same way as the left-handed fermion. Therefore the weak interaction conserves CP-symmetry, at least in the lepton sector as will be discussed in a moment. The $W^{\pm}$ bosons interact either with a lepton and a neutrino of the same generation or with a pair of quarks with different isospins. This leads to the interesting effect that leptons and neutrinos are able to transition into each other given that they are of the same generation. Similarly, up-type quarks and down-type quarks are able to transition into each other. The relative probabilities for the transitions between quark flavors are described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which is discussed in Refs. [7, 27, 28]. It is a unitary 3×3 matrix with a complex phase. The absolute values of the elements of this matrix can be measured in experiments, such as those summarized in Ref. [7] and are found to be

$$\begin{pmatrix} |V_{\mathrm{ud}}| & |V_{\mathrm{us}}| & |V_{\mathrm{ub}}| \\ |V_{\mathrm{cd}}| & |V_{\mathrm{cs}}| & |V_{\mathrm{cb}}| \\ |V_{\mathrm{td}}| & |V_{\mathrm{ts}}| & |V_{\mathrm{tb}}| \end{pmatrix} = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04503^{+0.000083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix} \tag{2.1}$$

The probabilities for transitions in the same generation are shown in the diagonal elements and are significantly larger than those in the off-diagonal elements describing transitions into other generations. This complex phase allows the weak interaction to violate the CP-symmetry in the quark sector since it changes its sign under simultaneous charge and parity transformation. The rates for processes involving interference terms with the CKM matrix elements are then different for the same process after CP-transformation. CP violation can be observed in experiments, for example in the oscillation between neutral kaons and their antiparticles or in the decays of kaons, as described in Refs. [29, 30]. The Z boson always interacts with pairs of particles and corresponding antiparticles. The Z boson carries $T_3 = 0$ while the $W^{\pm}$ bosons carry a weak isospin of $T_3 = \pm 1$ as well as electromagnetic charge. Therefore W bosons can interact with each other, the Z boson and the photon as long as all quantum numbers are conserved, giving rise to interaction vertices involving three or four of these gauge bosons. The W and Z bosons have large masses, which together with their ability to interact with each other, leads to a very short range for the weak interaction. The quantum field theory describing the weak interaction is formulated together with the electromagnetic interaction in a unified theory as described in Refs. [31–33]. This unification also requires the existence of a new particle called the Higgs boson, which was postulated in Refs. [34–36] and later observed by the CMS and ATLAS experiments at the LHC as described in Refs. [1–3]. The Higgs mechanism and the Higgs boson will be discussed further below.

The strong interaction is mediated by gluons and affects all particles carrying color charge. There are three types of color charge, often called red, green and blue. Quarks carry one type of color and antiquarks carry anticolor, while gluons carry a combination of color and different anticolor and can therefore interact with each other. A property of the strong interaction called color confinement implies that only particles in color singlet states can be observed. These can be reached in bound states of a quark and an antiquark with equal color and anticolor (mesons) or in bound states of three quarks with all three colors (baryons) [1]. As a consequence, quarks cannot be observed on their own. The energy required to separate a single quark from such a bound state increases with the separation distance until it exceeds the energy required for the production of a new quark-antiquark pair. These then recombine with the separated quark and the bound state to again form color singlet states. The interactions of gluons with each other also leads to an effect called asymptotic freedom where the strength of the strong coupling constant decreases

[1] This requirement of color neutral singlet states is reminiscent to the behavior of the real primary colors when mixed together.

with increasing energy. For this reason quarks with high energies can be approximately described as free quarks. The strong interaction is described by the theory of quantum chromodynamics (QCD) which is discussed in Refs. [37–39].

2.1.3 Qauntum field theories and the Higgs mechanism

In the following a brief overview of the mathematical formulation of the SM is given, with the aim to motivate the Higgs mechanism. If not otherwise stated, the presented information is based on Refs. [6,40]. A more thorough discussion of the SM and an overview over its current status is given in Ref. [7].

The SM is formulated in terms of relativistic quantum field theories, or more specifically in terms of gauge theories. Particles are interpreted as excitations of quantum fields which are represented by field operators $\phi(x)$, where x denotes a coordinate in four-dimensional space-time. The dynamics and interactions of the fields are described by the scalar Lagrangian density [2]:

$$\mathcal{L} = \mathcal{L}(\phi_i(x), \partial_\mu \phi_i(x)) \text{ with } \partial_\mu \equiv \frac{\partial}{\partial x_\mu}. \tag{2.2}$$

The equations of motion for the fields are obtained by application of Hamilton's principle, analogous to classical mechanics, and can therefore be calculated using the Euler-Lagrange-equation

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{2.3}$$

As an example, in a theory with one scalar field with a mass m and an no interactions the Lagrangian is given by

$$\mathcal{L}(\phi(x), \partial_\mu \phi(x)) = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 \tag{2.4}$$

and application of the Euler-Lagrange equation results in the equation of motion, called Klein-Gordon equation, for such a field,

$$(\partial_\mu \partial^\mu - m^2)\phi = 0. \tag{2.5}$$

In gauge theories it is required that the Lagrangian is invariant under local transformations of some theory specific symmetry group. To make usable predictions, the theories should be renormalizable, which means that all observable quantities can be calculated to finite values. Renormalization and renormalizability will not be discussed here, but it was shown that gauge theories are renormalizable [41, 42]. In the following the gauge theories of the standard model will be introduced.

[2] The Lagrangian density is colloquially also simply called Lagrangian although the Lagrangian is already defined as $L = \int \mathcal{L} \, \mathrm{d}^3 x$.

Table 2.2: The gauge bosons of the standard model. Indicated are which particles mediate which interaction and which charges they carry. The gluons always carry two color charges, one color and a different anticolor. All gauge bosons have spin 1. Information taken from Ref. [7].

particle	interaction	charge	isospin T_3	color	mass
photon (γ)	electromagnetic	0	0	none	–
$W^\pm$	electroweak	$\pm e$	± 1	none	80.379 ± 0.012 GeV
Z	electroweak	0	0	none	91.1876 ± 0.0021 GeV
gluon	strong	0	0	color	–

Quantum electrodynamics

The electrodynamic interaction between charged fermions is described by quantum electrodynamics (QED), which is a gauge theory with the symmetry group U(1). The fermions are represented by four-component Dirac spinors $\psi(x)$, which can be decomposed into operators that create or annihilate particles or antiparticles of different momenta and spin. For one fermion with a mass m, the Lagrangian density describing the dynamics without interactions is given by

$$\mathcal{L}_{\mathrm{D}} = \bar{\psi}(x)(\mathrm{i}\gamma^{\mu}\partial_{\mu} - m)\psi(x), \tag{2.6}$$

where the γ^{μ} are the Dirac matrices. The equation of motion for this field is the well known Dirac equation

$$(\mathrm{i}\gamma^{\mu}\partial_{\mu} - m)\psi = 0. \tag{2.7}$$

The Lagrangian density should be invariant under transformations of the local U(1) symmetry group, i.e. the gauge transformation

$$\psi(x) \to e^{-\mathrm{i}q\alpha(x)}\psi(x) \tag{2.8}$$

and

$$\bar{\psi}(x) \to e^{+\mathrm{i}q\alpha(x)}\bar{\psi}(x). \tag{2.9}$$

To make the Lagrangian density invariant under the local transformation, the derivative ∂_{μ} is replaced by the covariant derivative

$$\partial_{\mu} \to D_{\mu} = [\partial_{\mu} + \mathrm{i}qA_{\mu}(x)], \tag{2.10}$$

which transforms in the same way as the field $\psi(x)$ and adds a new term to the Lagrangian:

$$\mathcal{L}_{\mathrm{I}} = -q\bar{\psi}(x)\gamma^{\mu}A_{\mu}\psi(x). \tag{2.11}$$

Terms with multiple fields correspond to interactions between the fields with strengths corresponding to their numerical coefficients. In this case, it is the interaction between a fermion field, an antifermion field and the vector field A_{μ} with coupling constant q. The introduced field A_{μ} is a vector field (it has spin 1) and has to transform as

$$A_{\mu}(x) \to A_{\mu}(x) + \partial_{\mu}\alpha(x) \tag{2.12}$$

for the Lagrangian density to be invariant under the local transformation. The dynamics for vector fields are described by the Lagrangian density

$$\mathcal{L}_{\mathrm{A}} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{8\pi}m^2A^{\nu}A_{\nu} \tag{2.13}$$

with $F^{\mu\nu} \equiv \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$. Application of the Euler-Lagrange equation results in the Proca equation of motion for the field A^{μ}:

$$\left[\partial_{\mu}\partial^{\mu} + m^2\right]A^{\nu} = 0. \tag{2.14}$$

The Lagrangian density $\mathcal{L}_{\mathrm{A}}$ is invariant under the local gauge transformation if the mass of the particle A_{μ} is zero. The full Lagrangian density of the gauge theory is then

$$\mathcal{L}_{\mathrm{QED}} = \mathcal{L}_{\mathrm{D}} + \mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{I}} = \bar{\psi}(x)(\mathrm{i}\gamma^{\mu}\partial_{\mu} - m)\psi(x) - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} - q\bar{\psi}\gamma^{\mu}(x)A_{\mu}\psi(x). \tag{2.15}$$

Therefore the requirement of invariance under U(1) symmetry leads to the introduction of a massless spin 1 boson A_{μ}, the photon, that couples to a fermion and an antifermion with a strength q, the electric charge.

Quantum chromodynamics

The strong interaction is described by quantum chromodynamics (QCD), which is a non-abelian gauge theory with a SU(3) symmetry group and is discussed in Refs. [37–39]. Quarks are spin $\frac{1}{2}$ fermions and represented by Dirac spinors. In contrast to QED there are now three fermion fields per quark flavor, one for each of the three colors. With the notation

$$\Psi(x) = \begin{pmatrix} \psi_{\mathrm{r(x)}} \\ \psi_{\mathrm{g(x)}} \\ \psi_{\mathrm{b(x)}} \end{pmatrix}, \bar{\Psi}(x) = (\bar{\psi}_{\mathrm{r(x)}}, \bar{\psi}_{\mathrm{g(x)}}, \bar{\psi}_{\mathrm{b(x)}}), \tag{2.16}$$

the Lagrangian density can be written as

$$\mathcal{L}_{\mathrm{D}} = \sum_{\mathrm{f}=0}^{5} \bar{\Psi}_f(x)(\mathrm{i}\gamma^\mu \partial_\mu - m)\Psi_f(x) \tag{2.17}$$

where the sum is performed over all flavors. The SU(3) is generated by the eight Gell-Mann matrices λ_i and local transformations of the fields can be written as

$$\Psi_f(x) \to e^{-\mathrm{i}g_{\mathrm{s}} \sum_{i=0}^{7} \lambda_i \alpha_i(x)} \Psi_f(x) \tag{2.18}$$

and

$$\bar{\Psi}_f(x) \to e^{+\mathrm{i}g_{\mathrm{s}} \sum_{i=0}^{7} \lambda_i \alpha_i(x)} \bar{\Psi}_f(x). \tag{2.19}$$

To make the Lagrangian density gauge invariant, the normal derivative ∂_μ is replaced by a covariant derivative

$$D_\mu = [\partial_\mu + \mathrm{i}g_{\mathrm{s}} \sum_{i=0}^{7} \lambda_i G_{\mu,i}(x)] \tag{2.20}$$

where the $G_{\mu,i}(x)$ are vector fields called gluons. They have to transform as

$$G_{\mu,i}(x) \to G_{\mu,i}(x) + \partial_\mu \alpha_i(x) + g_{\mathrm{s}} f_{ijk} \alpha_j(x) G_{\mu,k}(x) \tag{2.21}$$

where the f_{ijk} are the structure constants of SU(3) and their dynamics are described by the Lagrangian density

$$\mathcal{L}_{\mathrm{G}} = \sum_{i=0}^{7} -\frac{1}{16\pi} F_i^{\mu\nu} F_{i\mu\nu} \tag{2.22}$$

with

$$F_i^{\mu\nu} = \partial^\mu G_i^\nu(x) - \partial^\nu G_i^\mu(x) + g_{\mathrm{s}} f_{ijk} G_{\mu,j}(x) G_{\nu,k}(x). \tag{2.23}$$

The complete Lagrangian density of QCD is then

$$\mathcal{L}_{\mathrm{QCD}} = \mathcal{L}_{\mathrm{D}} + \mathcal{L}_{\mathrm{G}} + \mathcal{L}_{\mathrm{I}}. \tag{2.24}$$

Fully expanding this Lagrangian density leads to terms describing interactions between a gluon, a quark and an antiquark of the same flavor and opposite color, meaning that gluons carry a color and a different anticolor, as well as interactions between three and four different gluons.

Electroweak theory

During the formulation of a gauge theory describing the weak interaction several complications arise. From experimental observation it is known that only left-handed particles (and right-handed antiparticles) have isospin $|T_3| = \frac{1}{2}$ and interact weakly while right-handed particles (and left-handed antiparticles) are isospin singlets with $T_3 = 0$, thereby violating parity. Interactions with the W boson occur between charged leptons and neutrinos of the same family and between up-type quarks and down-type quarks, while interactions with the Z boson only involve pairs of particles of opposite charge. This suggests that the symmetry group should be the SU(2) for the weak isospin T_3. Furthermore, the W and Z bosons have a finite mass.

The chiral components of a Dirac spinor can be projected by

$$\psi^{\mathrm{R/L}}(x) = \frac{1}{2}(1 \pm \gamma^5)\psi(x). \tag{2.25}$$

The fermions are then grouped into left-handed isospin doublets and right-handed isospin singlets. For the first generation for example

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_{\mathrm{L}}, \begin{pmatrix} u \\ d \end{pmatrix}_{\mathrm{L}}, (e)_{\mathrm{R}}, (u)_{\mathrm{R}}, (d)_{\mathrm{R}}. \tag{2.26}$$

Right-handed neutrinos do not interact with anything in the standard model and might therefore be treated as non-existing.

Assuming zero neutrino masses, the Lagrangian density for one generation of leptons can be written as

$$\mathcal{L}_{\mathrm{D,l}} = \bar{\psi}_{\mathrm{l}}^{\mathrm{L}} \mathrm{i}\partial\!\!\!/ \psi_{\mathrm{l}}^{\mathrm{L}} + \bar{\psi}_{\mathrm{l}}^{\mathrm{R}} \mathrm{i}\partial\!\!\!/ \psi_{\mathrm{l}}^{\mathrm{R}} - m\bar{\psi}_{\mathrm{l}}^{\mathrm{R}} \psi_{\mathrm{l}}^{\mathrm{L}} - m\bar{\psi}_{\mathrm{l}}^{\mathrm{L}} \psi_{\mathrm{l}}^{\mathrm{R}} + \bar{\psi}_{\nu_{\mathrm{l}}}^{\mathrm{L}} \mathrm{i}\partial\!\!\!/ \psi_{\nu_{\mathrm{l}}}^{\mathrm{L}} + \bar{\psi}_{\nu_{\mathrm{l}}}^{\mathrm{R}} \mathrm{i}\partial\!\!\!/ \psi_{\nu_{\mathrm{l}}}^{\mathrm{R}} \tag{2.27}$$

where the coordinate dependence of the $\psi(x)$ is omitted and the notation $\partial\!\!\!/ = \gamma^\mu \partial_\mu$ is introduced. The Lagrangian density for quarks is similar with the exception that there are also right-handed quark singlets. It is found that this Lagrangian density cannot be made invariant under local SU(2) symmetry while the masses of the fermions are finite. If the fermions were massless one could introduce an appropriate covariant derivative leading to three vector fields. As before, the vector fields have to be massless to make their own Proca Lagrangian density invariant under local SU(2) transformations. It is however known that the fermions as well as the bosons have finite masses. This problem is solved by first formulating the theory for massless particles and then generating the particle masses through the Higgs mechanism.

By writing the left-handed spinors $\psi_{\mathrm{l}}^{\mathrm{L}}$ and $\psi_{\nu_{\mathrm{l}}}^{\mathrm{L}}$ as a two component field

$$\Psi_{\mathrm{l}}^{\mathrm{L}} = \begin{pmatrix} \psi_{\nu_{\mathrm{l}}}^{\mathrm{L}} \\ \psi_{\mathrm{l}}^{\mathrm{L}} \end{pmatrix} \tag{2.28}$$

the Lagrangian density for the now massless fermions can be rewritten as

$$\mathcal{L} = \mathrm{i}[\bar{\Psi}_l^L \partial\!\!\!/ \Psi_l^L + \bar{\psi}_l^R \partial\!\!\!/ \psi_l^R + \bar{\psi}_{\nu_l}^R \partial\!\!\!/ \psi_{\nu_l}^R]. \tag{2.29}$$

The Lagrangian density is invariant under global SU(2) transformations where right-handed singlets transform as isoscalars

$$\psi^{\mathrm{R}} \to \psi^{\mathrm{R}} \tag{2.30}$$

and the left-handed doublets transform as isospinors

$$\Psi_{\mathrm{l}}^{\mathrm{L}} \to e^{\mathrm{i}\frac{\tau_{\mathrm{j}}}{2}\alpha_{\mathrm{j}}} \Psi_{\mathrm{l}}^{\mathrm{L}} \tag{2.31}$$

$$\bar{\Psi}_{\mathrm{l}}^{\mathrm{L}} \to e^{-\mathrm{i}\frac{\tau_{\mathrm{j}}}{2}\alpha_{\mathrm{j}}}\bar{\Psi}_{\mathrm{l}}^{\mathrm{L}} \tag{2.32}$$

with the τ_{j} denoting the Pauli matrices, which are the generators of SU(2), and the Einstein summation convention being used. The Lagrangian density is also invariant under the global U(1) transformation

$$\psi \to e^{\mathrm{i}\beta Y}\psi, \bar{\psi} \to e^{-\mathrm{i}\beta Y}\bar{\psi}, \tag{2.33}$$

where ψ stands for any lepton field. The quantity Y is called hypercharge and is connected to the electric charge and the third component of isospin via

$$Y = 2(Q - T_3) \tag{2.34}$$

Together with the invariance under global SU(2)×U(1) transformations this suggests that this is a theory describing the weak interaction together with the electrodynamic interaction.

To make the theory invariant under local the SU(2)×U(1) transformations, the transformations are first rewritten as

$$\Psi_{\mathrm{l}}^{\mathrm{L}} \to e^{\mathrm{i}\frac{\tau_{\mathrm{j}}}{2}g\omega_{\mathrm{j}}(x)}\Psi_{\mathrm{l}}^{\mathrm{L}}, \tag{2.35}$$

$$\bar{\Psi}_{\mathrm{l}}^{\mathrm{L}} \to e^{-\mathrm{i}\frac{\tau_{\mathrm{j}}}{2}g\omega_{\mathrm{j}}(x)}\bar{\Psi}_{\mathrm{l}}^{\mathrm{L}} \tag{2.36}$$

$$\psi^R \to \psi^R. \tag{2.37}$$

and

$$\psi \to e^{\mathrm{i}g\prime\beta(x)Y}\psi, \tag{2.38}$$

$$\bar{\psi} \to e^{-\mathrm{i}g\prime\beta(x)Y}\bar{\psi}. \tag{2.39}$$

Then covariant derivatives are introduced as

$$D^\mu\Psi_{\mathrm{l}}^{\mathrm{L}} = [\partial^\mu + \mathrm{i}g\frac{\tau_{\mathrm{j}}}{2}W_{\mathrm{j}}^\mu - \mathrm{i}\frac{g\prime}{2}B^\mu]\Psi_{\mathrm{l}}^{\mathrm{L}} \tag{2.40}$$

for the local SU(2) transformation and

$$D^\mu\psi_{\mathrm{l}}^{\mathrm{R}} = [\partial^\mu - \mathrm{i}g\prime Y B^\mu]\psi_{\mathrm{l}}^{\mathrm{R}}, \tag{2.41}$$

$$D^\mu\psi_{\nu_{\mathrm{l}}}^R = \partial^\mu\psi_{\nu_{\mathrm{l}}}^R \tag{2.42}$$

for the local U(1) transformation of the right-handed electrons and neutrinos. Together with the Proca Lagrangian densities for the four massless vector fields W_j^μ and B^μ the total Lagrangian density is

$$\begin{aligned}\mathcal{L} =&\mathcal{L}_{\mathrm{D}} + \mathcal{L}_{\mathrm{I}} + \mathcal{L}_{\mathrm{G}}\\ =&\mathrm{i}[\bar{\Psi}_{\mathrm{l}}^{\mathrm{L}}\not{\partial}\Psi_{\mathrm{l}}^{\mathrm{L}} + \bar{\psi}_{\mathrm{l}}^{\mathrm{R}}\not{\partial}\psi_{\mathrm{l}}^{\mathrm{R}} + \bar{\psi}_{\nu_{\mathrm{l}}}^{\mathrm{R}}\not{\partial}\psi_{\nu_{\mathrm{l}}}^{\mathrm{R}}]\\ &- \bar{\Psi}_{\mathrm{l}}^{\mathrm{L}}[g\frac{\tau_{\mathrm{j}}}{2}W_{\mathrm{j}}^\mu + \frac{g\prime}{2}B^\mu]\Psi_{\mathrm{l}}^{\mathrm{L}} + \bar{\psi}_{\mathrm{l}}^{\mathrm{R}}g\prime Y B^\mu\psi_{\mathrm{l}}^{\mathrm{R}}\\ &- \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{j\mu\nu}W_j^{\mu\nu},\end{aligned} \tag{2.43}$$

where the field tensors are defined as

$$W_{\mathrm{j}}^{\mu\nu} = \partial^\nu W_{\mathrm{j}}^\mu - \partial^\mu W_{\mathrm{j}}^\nu - g\epsilon_{\mathrm{jkl}}W_{\mu,\mathrm{k}}W_{\nu,\mathrm{l}} \tag{2.44}$$

and

$$B^{\mu\nu} = \partial^\nu B^\mu - \partial^\mu B^\nu. \tag{2.45}$$

This contains terms describing interactions between the vector fields and the left-handed isospin doublets or right-handed leptons, but not with right handed neutrinos. It also contains interactions between three and four vector bosons.

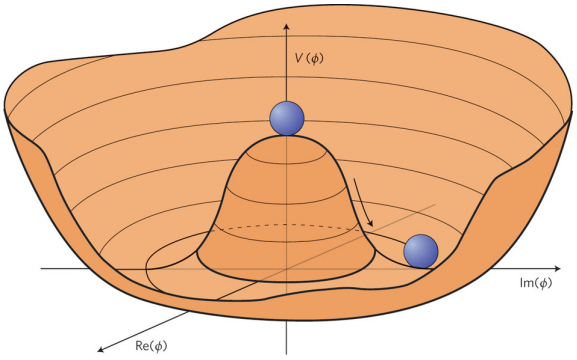


Figure 2.1: Illustration of the "Mexican hat potential", a two dimensional analogue of the potential used in electroweak symmetry breaking. Taken from Ref. [43].

Electroweak symmetry breaking and the Higgs mechanism

The particles described by the above Lagrangian density are still massless. To change this, an isospin doublet of complex scalar fields is introduced as

$$\Phi = (\phi_\mathrm{a}, \phi_\mathrm{b})^\mathrm{T}. \tag{2.46}$$

Complex scalar fields describe spinless charged particles. Under local SU(2)×U(1) transformations this field transforms the same way as $\Psi_\mathrm{l}^\mathrm{L}$ and with a covariant derivative defined as

$$D^\mu \Phi = [\partial^\mu + \mathrm{i}g\frac{\tau_\mathrm{j}}{2}W_\mathrm{j}^\mu - \mathrm{i}\frac{g'}{2}B^\mu]\Phi, \tag{2.47}$$

which is analogous to that for the field $\Psi_\mathrm{l}^\mathrm{L}$ a gauge invariant Lagrangian density for the field Φ can be written as

$$\mathcal{L}_\mathrm{H} = [D^\mu \Phi]^\dagger [D_\mu \Phi] - \mu^2 \Phi^\dagger \Phi - \lambda [\Phi^\dagger \Phi]^2. \tag{2.48}$$

The last two terms in this Lagrangian density correspond to the potential of the field. For $\lambda > 0$ and $\mu^2 < 0$ the potential has infinitely many ground states and for each such ground state the field is non-zero. This is illustrated by a two-dimensional analogue of this potential, the so-called "Mexican hat potential", as shown in Fig. 2.1. Since the vacuum should be at one specific ground state (and particles are the excitations of the fields about that ground state) one specific ground state is chosen. This spontaneously breaks the SU(2)×U(1) symmetry of the Lagrangian density. For the reason of notational clarity the ground state is often chosen to be

$$\Phi_0 = \frac{1}{\sqrt{2}}(0, v)^\mathrm{T}, \text{ with } v = \sqrt{\frac{\mu^2}{\lambda}}. \tag{2.49}$$

This choice is arbitrary and does not change the physics since all ground states are connected through a global phase transformation. The parameter v is called vacuum expectation value. Next, the field Φ is expanded in a perturbation series around the ground state as

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1 + \mathrm{i}\eta_2 \\ v + h + \mathrm{i}\eta_3 \end{pmatrix}. \tag{2.50}$$

The field h is called the Higgs field, has a mass of $m_h = \sqrt{2\lambda v^2}$ and its excitation is called the Higgs boson. The fields η_i are massless scalar fields called Goldstone bosons and are unphysical as discussed in Ref. [44]. The Goldstone bosons can be eliminated by performing a U(1) gauge transformation, which is allowed since the electroweak theory is

invariant under SU(2)×U(1) transformations. However, as a result of this transformation, the degrees of freedom of the Goldstone bosons are passed on to the gauge fields B_μ and $W_{\mu,\mathrm{j}}$ which leads to their masses. The fields B_μ and $W_{\mu,\mathrm{j}}$ are not actually the fields of the gauge bosons observed in experiments. These are given by linear combinations of the fields B_μ and $W_{\mu,\mathrm{j}}$ as

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_{\mu,1} \mp \mathrm{i}W_{\mu,2}), \tag{2.51}$$

$$Z_\mu = \frac{1}{\sqrt{g^2 + g\prime^2}}(-g\prime B_\mu + gW_{\mu,3}), \tag{2.52}$$

and

$$A_\mu = \frac{1}{\sqrt{g^2 + g\prime^2}}(gB_\mu + g\prime W_{\mu,3}). \tag{2.53}$$

After the spontaneous symmetry breaking the Lagrangian density of the electroweak theory

$$\mathcal{L} = \mathcal{L}_\mathrm{D} + \mathcal{L}_\mathrm{I} + \mathcal{L}_\mathrm{G} + \mathcal{L}_\mathrm{H} \tag{2.54}$$

contains terms describing interactions between the $\mathrm{W}^\pm$, Z bosons and the Higgs boson h. It also includes interaction terms $\propto V^\mu V_\mu$ which have the form of mass terms, where V denotes one of the vector bosons. Expansion of the Lagrangian shows that the $\mathrm{W}^\pm$ bosons obtain a mass of $m_W = \frac{1}{2}vg$ and the Z boson obtains a mass of $m_Z = \frac{1}{2}v\sqrt{g^2 + g\prime^2}$. The masses of the W and Z bosons are related through the Weinberg angle θ_W:

$$\cos\theta_\mathrm{W} = \frac{m_\mathrm{W}}{m_\mathrm{Z}} \tag{2.55}$$

with

$$\cos\theta_\mathrm{W} = \frac{g}{\sqrt{g^2 + g\prime^2}} \text{ and } \sin\theta_\mathrm{W} = \frac{g\prime}{\sqrt{g^2 + g\prime^2}} \tag{2.56}$$

and $\sin^2\theta_\mathrm{W} \approx 0.231$ as listed in Ref. [7]. For the photon A there is no such interaction term with the Higgs boson and the photon therefore remains massless (the symmetry of the electromagnetic U(1) has not been broken). Interactions between the fermions and the $\mathrm{W}^\pm$, Z bosons or the A are also present as before. There are also terms for trilinear Higgs boson self-interaction with strength $\propto \lambda v$ and quartic Higgs boson self-interaction with strength $\propto \lambda$. This leads to the observation that all parameters of this model are interconnected via the vacuum expectation value v and can be independently measured. The masses of the W and Z bosons can be measured and their coupling constants g and $g\prime$ can be measured from weak interactions. The mass of the Higgs boson can be measured and the parameter λ can, while difficult, be measured in analyses of the Higgs self-interaction.

This mechanism, in which the masses of the gauge bosons are generated through spontaneous symmetry breaking, is called the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism (or more concise simply Higgs mechanism) and was first developed in 1964, as described in Refs. [34–36], based on the concept of spontaneous symmetry breaking discussed in Refs. [45, 46].

Fermion masses

The masses of the fermions can also be obtained by interaction with the Higgs boson. This is achieved by introducing Yukawa interaction terms (see Ref. [47]) between the fields of the fermions and the isospin doublet Φ. For leptons the Lagrangian density for this is

$$\mathcal{L}_\mathrm{Y} = -y_\mathrm{l}[\bar{\Psi}_\mathrm{l}^\mathrm{L}\psi_\mathrm{l}^\mathrm{R}\Phi + \Phi^\dagger\bar{\psi}_\mathrm{l}^\mathrm{R}\Psi_\mathrm{l}^\mathrm{L}], \tag{2.57}$$

which after the spontaneous symmetry breaking becomes

$$\mathcal{L}_{\mathrm{Y}} = -\frac{1}{\sqrt{2}} y_{\mathrm{l}} (v+h)[\bar{\psi}_{\mathrm{l}}^{\mathrm{L}} \psi_{\mathrm{l}}^{\mathrm{R}} + \bar{\psi}_{\mathrm{l}}^{\mathrm{R}} \psi_{\mathrm{l}}^{\mathrm{L}}] = -\frac{1}{\sqrt{2}} y_{\mathrm{l}} (v+h) \bar{\psi}\psi. \tag{2.58}$$

Therefore, the leptons have a mass of $m_l = \frac{1}{\sqrt{2}} y_l v$, where y_l is called the Yukawa coupling constant, and interact with the Higgs boson with a coupling strength of $g_{\mathrm{Hll}} = \frac{1}{\sqrt{2}} y_{\mathrm{l}} = \frac{m_{\mathrm{l}}}{v}$ while the neutrinos remain massless. Similar mass terms and interactions can be derived for quarks. Since up-type quarks are also massive this is more complicated and requires the introduction of the Cabibbo–Kobayashi–Maskawa (CKM) matrix, described above. With this the flavor eigenstates of the quarks (d', s', b') are not the mass eigenstates (d, s, b) of the quarks but rather a superposition of them:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{\mathrm{ud}} V_{\mathrm{us}} V_{\mathrm{ub}} \\ V_{\mathrm{cd}} V_{\mathrm{cs}} V_{\mathrm{cb}} \\ V_{\mathrm{td}} V_{\mathrm{ts}} V_{\mathrm{tb}} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.59}$$

This mass generation mechanism can be tested by measuring processes in which the Higgs boson interacts with fermions. Since the interaction strength is proportional to the mass of the fermion, the top quark is a good candidate for this.

As of now, the masses of the neutrinos are not generated with this mechanism and their origin is still studied.

2.1.4 Matrix elements and Feynman diagrams

As was pointed out, products of three or more fields in the Lagrangian density describe interactions of particles. The probability for incoming particles to interact in some way resulting in the production of outgoing particles is given by the scattering cross section (this is also the case for particle decays). The scattering cross section depends on the initial state, the phase space available for the final state and on the squared matrix element $|\mathcal{M}|^2$ of the process and can be calculated with Fermi's golden rule, which is explained in for example Ref. [6], as

$$\sigma = 2\pi |\mathcal{M}_{i \to f}|^2 \cdot \rho. \tag{2.60}$$

Here i and f denote the initial and final states and ρ is the final-state phase space. The matrix element for a process can be calculated in perturbation theory from the Lagrangian density. The different terms in the perturbative expansion are often visualized by Feynman diagrams.

During the calculation of the matrix elements a complication arises in that the coupling constants and masses appearing in the Lagrangian densities are not actually the quantities measured in experiments. These depend on the energy scale at which the interaction takes place. As it turns out, the coupling constant in QED increases with increasing energy while for QCD it increases with decreasing energy. This behavior is caused by the group structure of the theories. Since for low energies the QCD coupling constant α_{s} turns large, for sufficiently low energies results depending on QCD cannot be calculated in perturbation theory where the matrix element is expanded in α_{s}. The "running" of the QCD coupling constant also results in the observation, that at low energies, strongly interacting particles are confined to bound states, while they can be treated as asymptotically free at high energies. In the calculation of the matrix elements this running of the coupling constants is accounted for by higher order terms. However, since the series is only calculated to a finite order, it has to be taken into account in another way. This is done by treating the coupling constants as functions of the energy scale at which the calculation is performed

and approximating their value by evolving the coupling to this scale from a set scale called the "renormalization scale" μ_R^2, for which it is set to have a specific value (for example through measurement), as for example explained in Refs. [40, 48]. This, however, results in the observation that the calculated results depend on the chosen renormalization scale, which is usually taken into account as an uncertainty on the results.

2.1.5 Open questions

Although the presented theories describe most of the physics of the standard model, several open questions remain unanswered, and might require explanations beyond the standard model. Observation of galactic rotation curves implies that a large part of galactic matter is made of dark matter which could be explained by yet unknown, weakly interacting particles. Another open question concerns the origin of the matter/antimatter imbalance in our universe. The current cosmological model requires a process of CP-violation to generate this imbalance and the known sources of CP-violation are insufficient to explain the large imbalance. The question whether the hierarchy of the largely different fermion masses is caused by chance or by an underlying theory also remains unanswered. While the SM can explain the electromagnetic, weak and strong interactions, there is no known theory with which gravity can be described in a consistent manner together with the other interactions. Of course there are many more issues and listing them all would exceed the scope of this discussion. These questions suggest that the SM needs to be extended or replaced by a more fundamental underlying theory. A prominent candidate for such a theory beyond the standard model is supersymmetry (SUSY), which predicts a number of new particles that could be produced at the LHC experiments.

2.2 Cross sections for hadron colliders

An important tool for the study of the standard model are particle colliders where particles are accelerated to high energies to conduct scattering experiments. The $\mathrm{t\bar{t}H(b\bar{b})}$ analysis is performed with data of scattering events produced at the LHC, one such particle accelerator, which are recorded by the CMS experiment, a detector system to measure the results of these scattering experiments. Both are discussed in Ch. 3. At the LHC protons are accelerated and brought to collision to induce proton-proton interactions. The rate with which a process occurs in this scattering experiments is given by

$$\frac{\mathrm{dN}}{\mathrm{dt}} = \sigma \cdot L \tag{2.61}$$

where σ denotes the cross section of the process and L is a quantity called instantaneous luminosity describing the flux of protons available for the interactions (see Ch. 3). The cross sections for such processes can in principle be calculated perturbatively up to a finite order, currently usually up to next-to-leading order or sometimes even next-to-next-to-leading order. But since protons are composite particles made up of partons, additional complications arise in that the initial-state particles in the hard scattering process are actually the partons inside the protons. The probability to interact with a specific parton with a momentum fraction x when probing a proton at an energy scale μ^2 is described by parton density functions (PDFs) $f(x, \mu^2)$. The description is usually given either in the so-called four flavor scheme (4FS) or the five flavor scheme (5FS). In the 4FS the PDFs do not include bottom or top quarks in their description of the sea quarks of the proton. In the 5FS the PDFs include the bottom quark but not the top quark. PDFs are constructed from measurements at various scales using data from various scattering experiments. To evaluate them at different scales, they can be evolved using the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations described in Ref. [49–51].

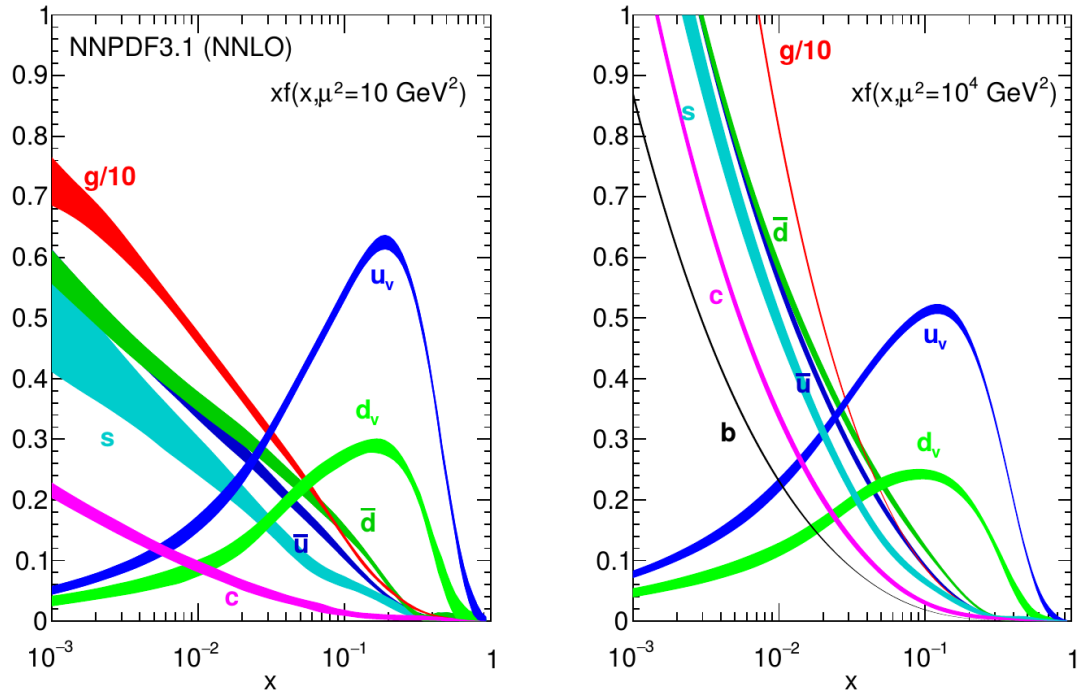


Figure 2.2: Parton distribution functions (PDFs) describing the probability to find a parton with a fraction x of the proton momentum at a scale μ^2 for $\mu^2 = 10\,\text{GeV}$ (left) and $\mu^2 = 10^4\,\text{GeV}$ (right). These PDFs are provided by the NNPDF Collaboration as described in Ref. [53].

PDFs are provided by multiple research groups and in multiple variations differing in the data they use and how they are derived. An overview over the derivation of PDFs and the commonly used sets is given in Ref. [52]. A PDF at different scales, derived by the NNPDF Collaboration as described in Ref. [53], is shown in Fig. 2.2. At both scales, the up and down valence quarks have the highest probabilities to be found for large momentum fractions x. The probabilities to find sea quarks or gluons increases with increasing scale μ.

With the proton structure accounted for, cross sections for processes at the LHC can be calculated as

$$\sigma(p_1 p_2 \to f) = \sum_{i,j} \int_0^1 \mathrm{d}x_i \int_0^1 \mathrm{d}x_j f_i(x_i, \mu_\mathrm{F}^2) f_j(x_j, \mu_\mathrm{F}^2) \hat{\sigma}(\hat{p}_i \hat{p}_j \to f). \tag{2.62}$$

Here p_1 and p_2 are the momenta of the two protons, the x_i are the fractions of the proton momenta carried by parton i, the $\hat{p}_i$ are the momenta of the partons, the $f_i(x_i, \mu_\mathrm{F}^2)$ are the PDFs for this type of parton and $\hat{\sigma}(\hat{p}_i \hat{p}_j \to f)$ denotes the partonic cross section for the process. The sum is performed over all partons described by the PDFs. The scale μ_F is called factorization scale and defines the energy at which the interaction is factorized into a low energy soft part described by the PDFs and a high energy hard part described by the partonic cross section, as described in Ref. [54].

A similar factorization approach is used for the description of the decays of the particles produced in the interactions. The quarks and gluons in the final state can decay into or emit additional quarks and gluons. This leads to a spray of color-charged particles, called parton shower. While the partons in the parton shower have high energies this can be calculated by perturbation theory, but once the energies are too low the description is based on the Altarelli-Parisi splitting functions and Sudakov form-factors, described

in Refs. [51, 54, 55]. Finally, since quarks and gluons cannot be observed on their own they combine to bound states in a process called hadronization. This is usually described using phenomenological models. Due to momentum conservation, the directions of the quarks and gluons created in the partonic interaction impart their momenta on their decay products and their radiated gluons leading to cone-shaped *jets* of hadrons and other decay products around the directions of the original hard partons.

2.3 Simulation of collision events

To make comparisons between the predictions of a theory and the data recorded in scattering experiments, it is often necessary to generate samples of simulated data. For one, the predictions have to be made in the form of observables that can be measured in the scattering experiments, such as the distribution of jet energies. Therefore, the predictions need to include the previously discussed parton shower and hadronization. The calculations of cross sections already include many integrations over internal momenta and the PDFs. Furthermore, the splittings in the parton shower are described by the probabilistic Altarelli-Parisi splitting functions and Sudakov form-factors, described in Refs. [51, 54, 55]. This suggests the use of Monte Carlo (MC) methods, described in Refs. [56, 57] for these calculations. In this scheme individual events are simulated by sampling from probability distribution functions. The number of events in the whole final-state phase space (or some part of it) then corresponds to the total (or fiducial) cross section of the simulated process. This method has the additional advantage that it allows the simulation of the detector response to individual events. Such simulations are performed in multiple steps, as shown in Fig. 2.3. A very simplistic overview over these steps will be given in the following while a more in-depth description of the simulation of events with MC methods is given for example in Refs. [58, 59].

First a partonic final state is generated based on the PDFs and the matrix element of the process. The matrix element is only calculated to finite order, today usually to leading (LO) or next-to-leading (NLO) order in α_{QCD}. The software responsible for this step is called matrix element generator.

Next the parton shower is simulated. First, the radiation of additional gluons from initial- or final-state partons is simulated. Then radiation of gluons or the splitting of gluons into quark-antiquark pairs is simulated for each particle. This process is repeated until the energies of the particles in the shower are low enough that the strong coupling constant is large enough to allow for the formation of hadrons. When the energies are high enough the splittings and radiation can be calculated using matrix elements. For lower energies the parton shower is described by the Altarelli-Parisi splitting functions and Sudakov form-factors. A complication arises in that additional gluon radiation can also arise from the matrix element generator. In such cases it is important to note that hard emissions are usually better described by the matrix element generator while soft or collinear emissions are better described by the parton shower simulation. The matrix element generator and the parton shower simulation are therefore combined using matching and merging schemes such as the MLM and FxFx schemes described in Refs. [60, 61]. It also has to be considered that the remnants of the two protons, i.e. the partons not interacting in the hard scattering, are still color charged particles and contribute additional particles to the parton shower. In addition, the accelerated protons at the LHC travel in packets called *bunches* of about 10^9 protons, as will be discussed in Ch. 3 and instead of only two protons two whole bunches are brought to collision. This leads to multiple proton-proton interactions per bunch collision but usually at most one proton-proton interaction has a large enough momentum transfer to result in an interesting hard scattering event. The other proton-proton interactions are

called pileup and are taken into account by simulating additional proton-proton interactions and adding the resulting particles to the event.

After the parton shower is concluded the hadronization is simulated. All additional quark-antiquark pairs are created from the remaining gluons and the quarks and antiquarks are combined to form hadrons. Since the energies of the particles are very low at this point, phenomenological models are used such as the Lund string model or the cluster hadronization model which are discussed in Refs. [62, 63]. Finally, the decays of all particles are simulated.

After all simulation steps are performed, the simulated event contains the same typical composition of particles as those in real proton-proton collisions. Events simulated in this manner are called *particle level* events.

The steps of the event simulation are implemented in various software packages. The event simulation is usually performed either using one software package for all steps or using one package for the matrix element level simulation and another for the subsequent parton shower and hadronization steps. Commonly used matrix element generators include the MadGraph5_aMC@NLO [64], Powheg [65–67] and OpenLoops [68, 69] generators. General-purpose event generators that can be used for all simulation steps include Sherpa [70], Pythia [71] and HERWIG [72, 73] where the last two are often used only for the parton shower and hadronization steps. Each of these generators has different capabilities and advantages and disadvantages for the simulation of different processes. Even a rudimentary discussion of these would severely exceed the scope of this thesis.

After the samples of the events are generated, there is usually another step that simulates the response of the detector. For the CMS detector this is performed using the GEANT software package described in Ref. [74]. This involves the simulation of the interactions of the particles produced in the collision with a detailed model of the CMS detector and the simulation of the readout electronics. Afterwards the simulated events can be reconstructed and analyzed in the same way as real events. Comparisons are then made between data simulated with the detector response taken into account and data recorded in the real detector. The detector simulation is computationally intensive and has the disadvantage that observables measured and reconstructed by the CMS experiment cannot be easily compared to data simulated by theoreticians outside of the experiments. It is possible to *unfold* the measured data back to the particle level. This is however accompanied by different issues as is explained in for example Ref. [75].

2.4 The top quark

The top quark is the heaviest known particle with a mass of $172.76 \pm 0.30\,\text{GeV}$, as listed in Ref. [7], and was discovered in 1995 by the CDF and DØ Collaborations at the Tevatron collider as described in Refs. [20, 21]. Since its mass is so much higher than the mass of the bottom quark and the relevant element of the CKM matrix $V_{\text{tb}} \approx 1$ top quarks almost immediately decay to a bottom quark and a W boson. It is the only quark not found in bound states since it decays too fast. This also allows the top quark to impart its spin to its decay products. The spin correlation between top quark-antiquark pairs is sensitive to effects caused by physics beyond the standard model and can be studied as was done in Refs. [77, 78]. Furthermore, since the coupling strength between fermions and the Higgs boson is proportional to the fermion mass, the top quark is of special interest for the study of the Higgs sector. The origin of hierarchy of fermion masses with the dominant top mass at the top is itself a mystery. Intriguingly the top-Higgs Yukawa coupling is

$$y_{\text{t}} = \sqrt{2}\frac{m_{\text{t}}}{v} \approx 1 \tag{2.63}$$

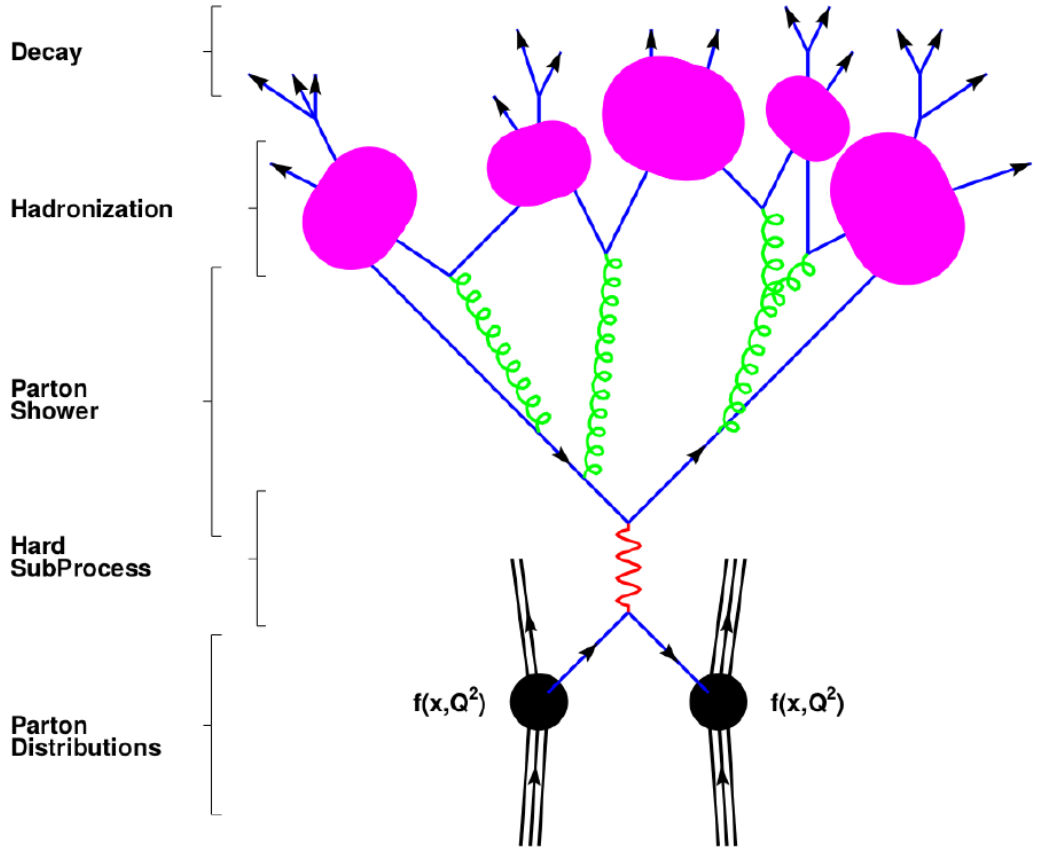


Figure 2.3: Illustration of the event simulation process. Starting from the bottom first partons are sampled from the PDFs then the hard subprocess is generated. This is followed by a simulation of the parton shower, the hadronization and the decay of the resulting particles. Taken from Ref. [76].

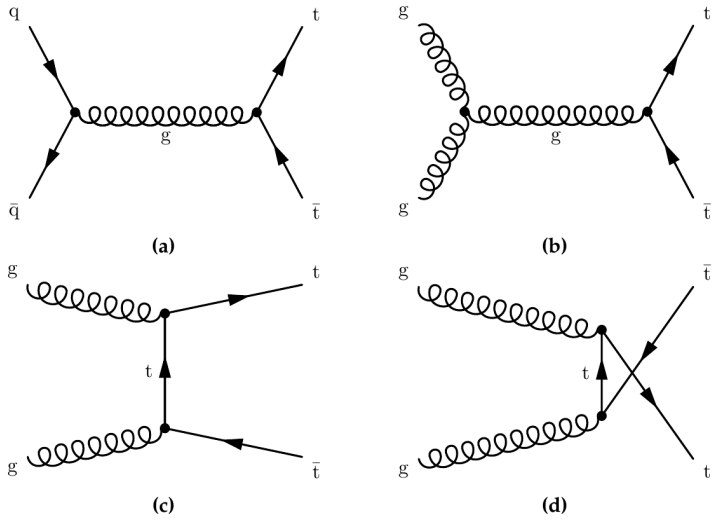


Figure 2.4: The leading order Feynman diagrams for the production of top quark-antiquark pairs. Taken from Ref. [88].

for the observed value of $v \approx 246\,\text{GeV}$ [7]. It is not known whether this value is required by some underlying theory or whether it is a coincidence. Through its interaction with the Higgs boson, the top quark also occurs in loop corrections of many processes. Additionally, many BSM models allow the decay of new postulated particles to top quarks. Finally, the production of top quarks is an important background in many analyses at the LHC. For $\text{t}\bar{\text{t}}\text{H}$ analyses it plays the role of the dominant background.

2.4.1 Top quark production

At the LHC, top quarks can be produced in pairs ($\text{t}\bar{\text{t}}$) via the strong interaction or as single top quarks via the weak interaction. Of special importance for $\text{t}\bar{\text{t}}\text{H}(\text{b}\bar{\text{b}})$ analyses is the case where top quark-antiquark pairs are produced with additional b quarks in the final state. In the following different production channels are discussed.

$\text{t}\bar{\text{t}}$ production

The production of top quark-antiquark pairs ($\text{t}\bar{\text{t}}$) is the dominant top quark production process at the LHC. The leading order Feynman diagrams for this process are shown in Fig. 2.4. At the LHC top quark-antiquark pairs are usually produced in association with additional jets. For a center-of-mass energy of 13 TeV the predicted inclusive $\text{t}\bar{\text{t}}$ production is

$$\sigma_{\text{t}\bar{\text{t}}} = 831.76^{+19.77}_{-29.20}(\text{scale})^{+35.06}_{-35.06}(\text{pdf} + \alpha_\text{s})\,\text{pb}, \tag{2.64}$$

calculated at NNLO+NNLL accuracy as is described in Refs. [79–85] for a top quark mass of 172.5 GeV. Recent measurements of this value by the CMS and ATLAS Collaborations are summarized in Fig. 2.5 where it can be seen that the predicted cross section describes nature well. Differential cross sections for $\text{t}\bar{\text{t}}$ production have also been studied, for example in Refs. [86, 87] and found to be predicted reasonably well.

$\text{t}\bar{\text{t}}$+$\text{b}\bar{\text{b}}$ production

The production of a top quark-antiquark pair in association with a highly energetic bottom quark-antiquark pair is called $\text{t}\bar{\text{t}}$+$\text{b}\bar{\text{b}}$ production. This can occur for example when the hard-scattering is mediated by gluons radiated from sea bottom quarks, the bottom quarks

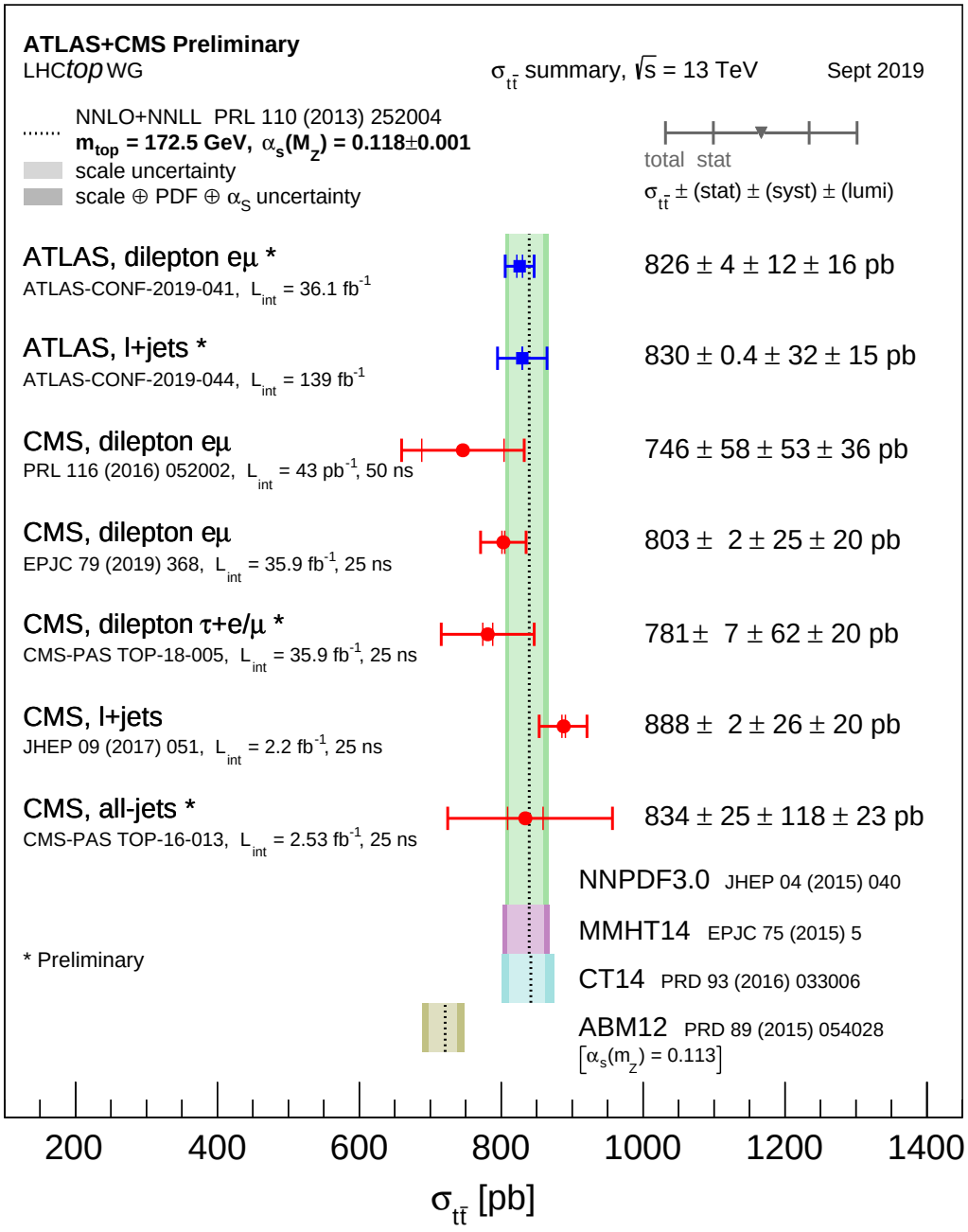


Figure 2.5: Summary of measurements of the top quark-antiquark pair production cross-section at 13 TeV performed by the ATLAS and CMS Collaborations. Results are compared to the prediction calculated in Refs. [79–85]. The uncertainty band on the prediction represents uncertainties due to renormalization and factorization scale, parton density functions and the strong coupling. Taken from Ref. [89].

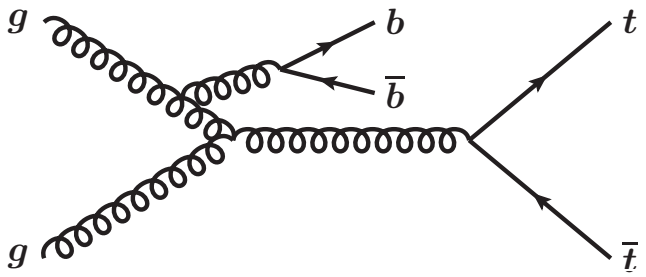
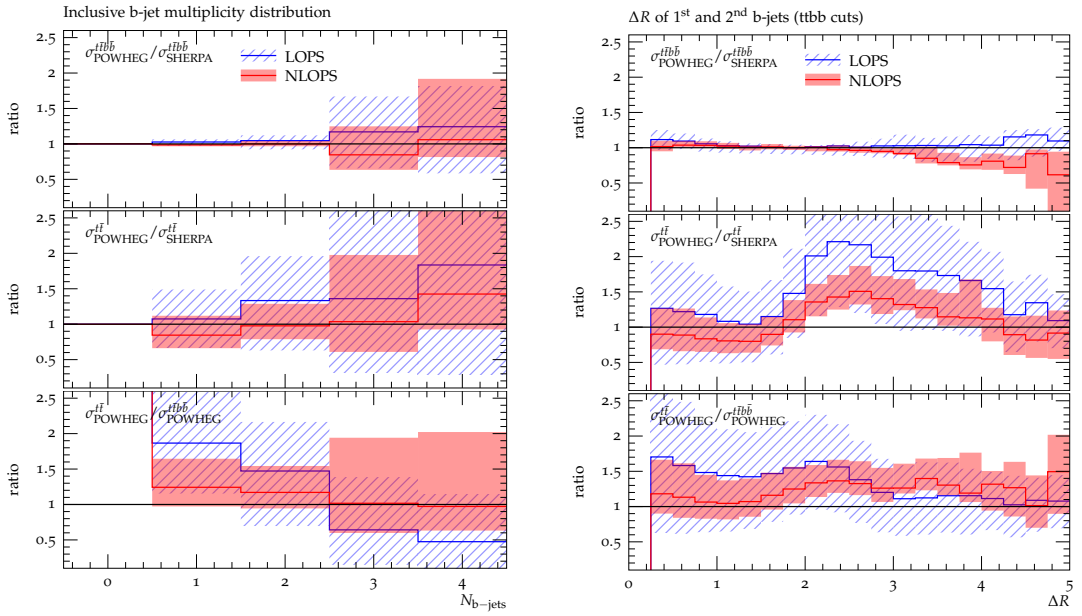


Figure 2.6: A leading order Feynman diagram for the production of $t\bar{t}+b\bar{b}$. Taken from Ref. [90].

then leaving the protons, or when initial- or final-state radiation splits into a $b\bar{b}$ pair. This process is of special importance to $t\bar{t}H(b\bar{b})$ analyses since it leads to the same final-state particle content as the $t\bar{t}H(b\bar{b})$ process. The production of $t\bar{t}$ in association with charm quarks ($t\bar{t}+c\bar{c}$) is also challenging for $t\bar{t}H(b\bar{b})$ analyses since, as will be discussed in Ch. 6, the experimental distinction between bottom quarks and charm quarks is difficult. A representative leading order Feynman diagram for $t\bar{t}+b\bar{b}$ production with an initial-state gluon splitting into the bottom quark-antiquark pair is shown in Fig. 2.6.

The prediction of $t\bar{t}+b\bar{b}$ production is challenging due to multiple reasons that are discussed in Refs [91, 92]. First, the leading order matrix element already depends on $\alpha_s(\mu_R)^4$ making the process very sensitive to variations of the renormalization scale. The process also combines very different energy scales in that the top quarks are heavy and $t\bar{t}$ is produced at energy scale on the order of 500 GeV while the bottom quarks are much lighter and have lower momenta on the order of 50 GeV. Additional complications arise since additional bottom quarks (and b jets) can also arise from final-state gluon splittings during the parton shower. A hard scattering cross section for $t\bar{t}+b\bar{b}$ production can be calculated at NLO accuracy either using the 4FS or the 5FS.
In Ref. [91] it is argued that the calculations involve large terms of the form $\alpha_s \ln(m_t/m_b)$ related to collinear gluon splittings in the initial state. In 5FS calculations these are automatically resummed to all orders inside the PDF. However, since the bottom quarks are massless, the 5FS matrix element diverges due to collinear singularities in cases where the bottom quarks have insufficient momenta or angular separation. For this phase space region the bottom quarks have to be generated during the parton shower, which usually leads to a worse description of their kinematic properties.
In the 4FS on the other hand, the bottom quarks are massive and can be generated in the matrix element in the whole phase space. In Ref. [91] it is also argued that the $\alpha_s \ln(m_t/m_b)$ terms do not have a large effect on the prediction of $t\bar{t}+b\bar{b}$. Therefore it is suggested to calculate the $t\bar{t}+b\bar{b}$ predictions in the 4FS. A similar conclusion is reached in Ref. [92].
The difference between 5FS predictions, where additional b jets (those not coming from the top quark decays) only originate from bottom quarks produced by the parton shower, and 4FS predictions where the bottom quarks and jets originate also[3] from the matrix element, is shown in Fig. 2.7. Shown are ratios of differential pp $\rightarrow t\bar{t}$+b-jets cross sections for the number of additional b jets and the ΔR between the two b jets with the highest momenta, calculated in the 4F ($\sigma_{t\bar{t}b\bar{b}}$) and 5F ($\sigma_{t\bar{t}}$) schemes using the (N)LO+PS generators Sherpa and Powheg+Pythia. As can be seen, the differences between the generators and the 4FS and 5FS calculations are smaller for NLO calculations. In the bottom ratio plots it can also be seen that even at NLO precision differences between 4FS and 5FS calculations remain.

[3] In addition to the bottom quarks from the matrix element, additional bottom quarks can also still be produced from $g \rightarrow b\bar{b}$ splittings in the parton shower



(a) (b)

Figure 2.7: Ratios of differential pp → $t\bar{t}$+b-jets cross section predictions at a center-of-mass energy of 13 TeV for: (a) the number of additional b jets and (b) the ΔR between the two b jets with the highest momenta. The cross sections were calculated in the 4F ($\sigma_{t\bar{t}b\bar{b}}$) and 5F ($\sigma_{t\bar{t}}$) schemes using the (N)LO+PS generators Sherpa and Powheg+Pythia. Uncertainty bands include uncertainties on α_s and parton shower parameters. Taken from Ref. [91].

Multiple calculations of the inclusive $t\bar{t}$+$b\bar{b}$ cross section have been performed for various center-of-mass energies and phase spaces at (N)LO accuracy in the four or five flavor scheme and are discussed in Refs. [91, 93–98]. A comparison between these analyses shows the uncertainties on $t\bar{t}$+$b\bar{b}$ production ranging from 20% to 40%. A recent NLO+PS calculation for pp → $t\bar{t}$+ ≥ 2 b-jets production at a center-of-mass energy of 13 TeV found a value of

$$\sigma_{pp\to t\bar{t}+\geq 2b-jets} = 2.537^{+40\%}_{-29\%}(\text{scale})\ \text{pb} \tag{2.65}$$

where the uncertainty corresponds to the uncertainties on the renormalization and factorization scales, as described in Ref [91].

The production of $t\bar{t}$+$b\bar{b}$ in the context of $t\bar{t}H(b\bar{b})$ analyses was studied in Ref. [99–101]. Measurements of $t\bar{t}$+$b\bar{b}$ production have also been performed by the ATLAS and CMS Collaborations as described in Refs. [102–105] with uncertainties ranging from 13 to 36%. Fig. 2.8 shows the measured cross sections of three CMS analyses in comparison to various predictions. It can be seen that the measured cross sections are consistently higher than the predicted ones. The numerical values of the measured cross sections for the shown analyses are:

$$\sigma_{t\bar{t}+b\bar{b},\text{alljet}} = 5.5 \pm 3.3(\text{ stat.})^{+1.6}_{-1.3}(\text{syst.})\,\text{pb}, \tag{2.66}$$

$$\sigma_{t\bar{t}+b\bar{b},\text{dilep}} = 2.9 \pm 0.1(\text{stat.}) \pm 0.5(\text{syst.})\,\text{pb}, \tag{2.67}$$

$$\sigma_{t\bar{t}+b\bar{b},\text{semilep}} = 4.7 \pm 0.2(\text{stat.}) \pm 0.6(\text{syst.})\,\text{pb}. \tag{2.68}$$

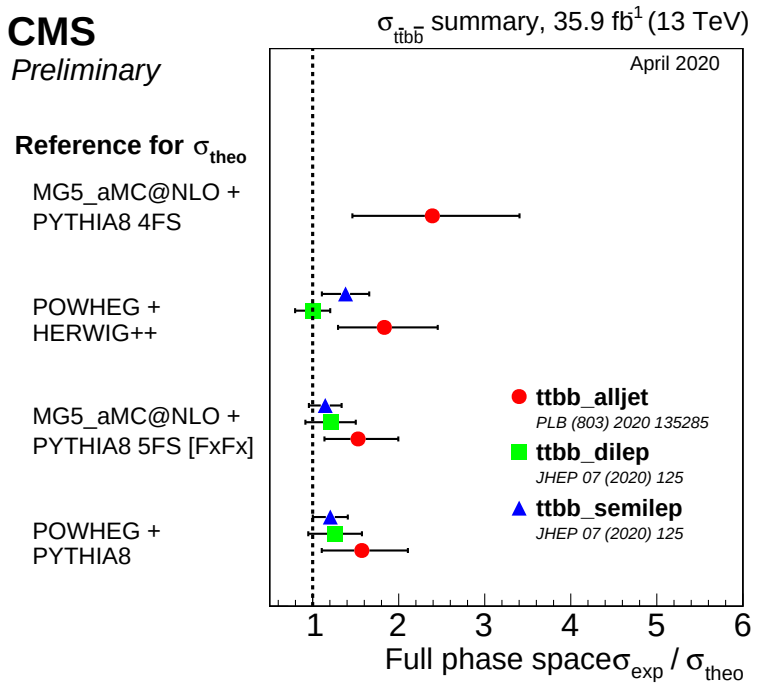


Figure 2.8: Measured $t\bar{t}+b\bar{b}$ cross sections with respect to various predictions. Taken from Ref. [106].

Single t production

The production of a single top quark mediated by a W boson is also possible at the LHC. Three production channels are distinguished by the virtuality of the W boson: t-channel, s-channel and tW-channel production. Leading-order Feynman diagrams for the three channels are shown in Fig. 2.9. For a center-of-mass energy of 13 TeV, the cross sections for the production of single t events in these channels are

$$\sigma_{\mathrm{t,t-channel}} = 216.99^{+6.62}_{-4.64}(\text{ scale})^{+6.16}_{-6.16}(\mathrm{pdf}+\alpha_{\mathrm{s}})\,\mathrm{pb}, \tag{2.69}$$

$$\sigma_{\mathrm{t,s-channel}} = 10.32^{+0.29}_{-0.24}(\text{ scale})^{+0.27}_{-0.27}(\mathrm{pdf}+\alpha_{\mathrm{s}})\,\mathrm{pb}, \tag{2.70}$$

$$\sigma_{\mathrm{t,tW-channel}} = 71.7 \pm 1.80(\text{ scale}) \pm 3.40(\mathrm{pdf}+\alpha_{\mathrm{s}})\,\mathrm{pb}, \tag{2.71}$$

as calculated in Ref. [107–111].

2.4.2 Top quark decays

The top quark has a very short lifetime of about 10^{-25} s and decays almost immediately after being created without forming any bound states. Since the relevant element of the CKM matrix V_{tb} is approximately one, a top quark almost always decays into a bottom quark and a W boson. The W boson subsequently decays into either a quark-antiquark pair with a branching ratio of 67.4% or a lepton and a neutrino with a branching ratio of 32.6%, as stated in Ref. [7]. For this reason the decays of single top quarks and top quark-antiquark pairs are characterized by the decay products of the W bosons. In hadronic $t\bar{t}$ decays, with a branching ratio of 45.7%, both W bosons decay into quarks. In the dileptonic $t\bar{t}$ decay, with a branching ratio of 10.5%, both W boson decay into leptons and neutrinos. The case in which one W boson decays into quarks and the other into a lepton and neutrino is called semileptonic and has a branching ratio of 43.8%.

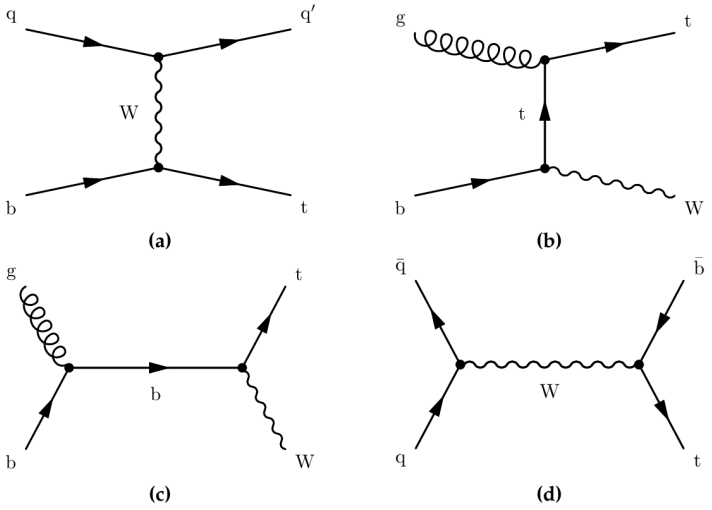


Figure 2.9: Exemplary leading order Feynman diagrams for single t production in the t-channel (a), the tW-channel (b) and the s-channel (d). Taken from Ref. [88].

2.5 The Higgs boson

The Higgs boson, predicted by the electroweak theory, was the last missing piece of the standard model to be discovered. In the standard model the mass of the Higgs boson is not uniquely determined, but can be constrained through the interactions of the Higgs boson with the other particles. Searches for the Higgs boson had been performed at the Large Electron Positron (LEP) and Tevatron colliders but were only able to further constrain the allowed mass range, as discussed in Refs. [112–114]. In 2012 a particle with a mass of around 125 GeV and the characteristics of the Higgs boson was finally discovered at the LHC as described in Refs. [1–3]. Since then, a vast amount of analyses of this particle have been performed leading to the conclusion that it is indeed the Higgs boson predicted by the standard model. The currently best estimate for its mass is $125.10 \pm 0.14\,\text{GeV}$ calculated from combinations of different measurements as stated in Ref. [7].

Several theories of physics beyond the standard model predict changes to the properties of the Higgs boson and its couplings to the standard model particles, such as for example those described in Ref. [115–119].

To study deviations of the coupling constants from the standard model, the so-called κ-framework is introduced in Refs. [92,120]. For a generic process involving the Higgs boson the product of production cross section σ and branching ratio $\mathcal{B}$ is

$$(\sigma \cdot \mathcal{B})_{\mathrm{i}\to\mathrm{H}\to\mathrm{f}} = \frac{\sigma_{\mathrm{i}\to\mathrm{H}} \cdot \Gamma_{\mathrm{H}\to\mathrm{f}}}{\Gamma_{\mathrm{H}}} \tag{2.72}$$

with the initial state i, the final state f and the widths Γ. Scaling parameters are introduced, defined as

$$\kappa_{\mathrm{i}}^2 = \frac{\sigma_{\mathrm{i}}}{\sigma_{\mathrm{i,SM}}} \text{ and } \kappa_{\mathrm{f}}^2 = \frac{\Gamma_{\mathrm{f}}}{\Gamma_{\mathrm{f,SM}}} \tag{2.73}$$

thereby effectively describing the deviation from the standard model. For example the $\sigma \cdot \mathcal{B}$ of the $\mathrm{t\bar{t}H(b\bar{b})}$ process can then be written as

$$\sigma_{\mathrm{ttH}} \cdot \mathcal{B}_{\mathrm{H}\to\mathrm{bb}} = \kappa_{\mathrm{t}}^2 \sigma_{\mathrm{ttH,SM}} \cdot \kappa_{\mathrm{b}}^2 \Gamma_{\mathrm{H}\to\mathrm{bb,SM}}. \tag{2.74}$$

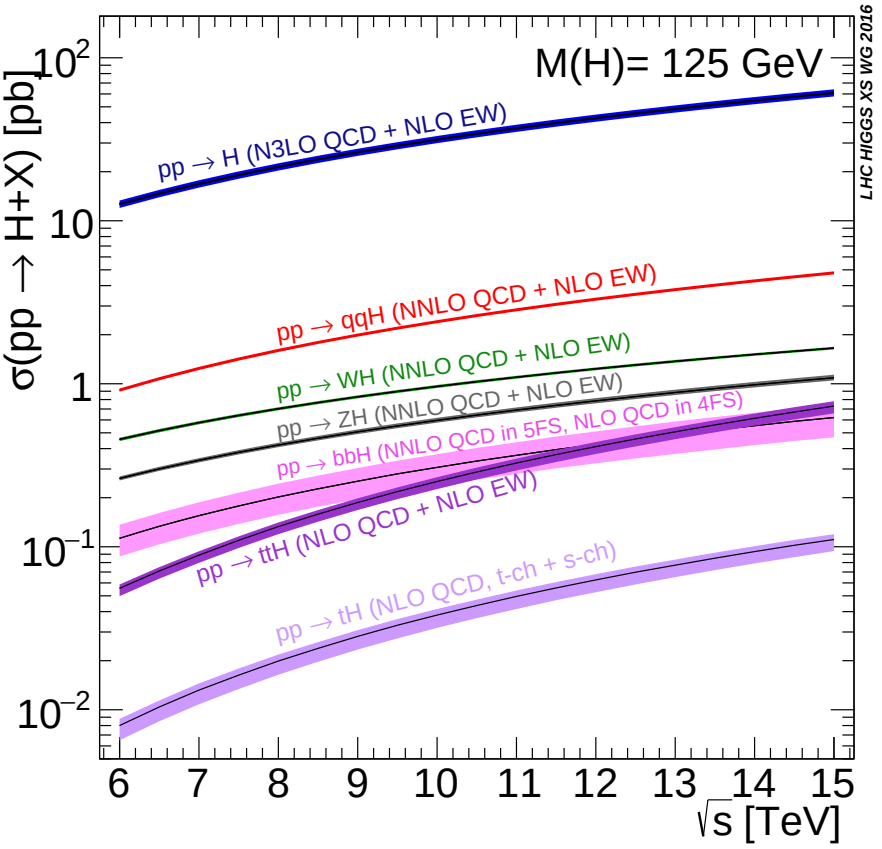


Figure 2.10: Higgs production cross sections as a function of the center-of-mass energy. Taken from Ref. [92].

2.5.1 Higgs production at the LHC

At the LHC the Higgs boson can be produced through multiple modes. A summary of the production cross sections in various modes as a function of the center-of-mass energy is shown in Fig. 2.10. The modes with the highest cross sections are gluon-gluon fusion, vector boson fusion, associated production with a vector boson and associated production with a top quark-antiquark pair. A recent measurement of these cross sections by the CMS Collaboration is shown in Fig. 2.12. The results are expressed in terms of the signal-strength modifier $\mu = \sigma/\sigma_{\mathrm{SM}}$ and are calculated from a combination of various measurements, including those discussed in this thesis. As can be seen, no significant deviations from the SM prediction can be observed. The individual production modes are further discussed in the following.

Gluon-gluon fusion

At a center-of-mass energy of 13 TeV the dominant production mode is gluon-gluon fusion (ggF), shown in the figure as pp → H. A Feynman diagram for this process is shown in Fig. 2.11 (a). The Higgs boson is produced from initial-state gluons via a fermion loop, dominated by the top quark due to its large mass. Since no additional particles are produced in association with the Higgs boson, analysis of this mode can be challenging.

Vector boson fusion

The production via vector boson fusion (VBF) has the second highest cross section and is shown in Fig. 2.11 (b) and denoted as pp → qqH in Fig. 2.10. Here the Higgs boson couples

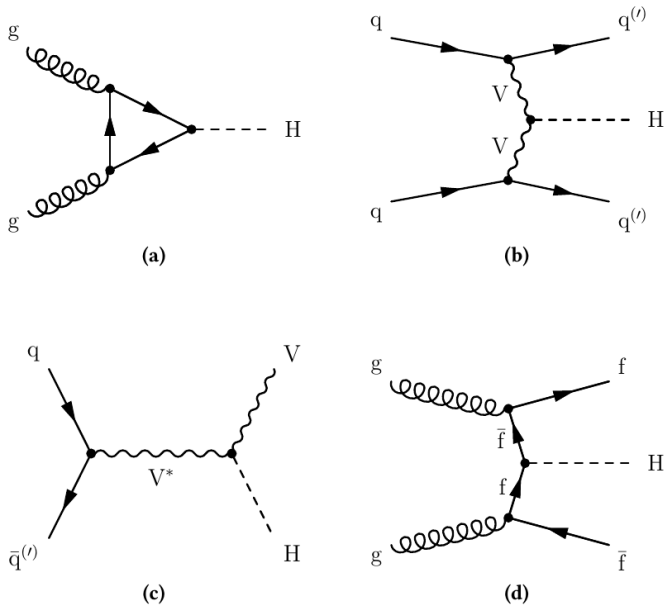


Figure 2.11: Feynman diagrams for the four most important Higgs production channels at the LHC: (a) gluon-gluon fusion (ggF) ($pp \to H$), (b) vector boson fusion (VBF) ($pp \to qqH$), (c) associated production with a vector boson (VH) and associated production with a top quark-antiquark pair ($t\bar{t}H$). Taken from Ref. [121].

to vector bosons radiated from two initial-state quarks. The VBF process results in two additional quarks in the final state.

Associated production with vector bosons

The cross section for production in association with vector bosons (VH) is shown in Fig. 2.10 separately for WH and ZH production. An exemplary Feynman diagram of this process is shown in Fig. 2.11 (c). Two initial-state quarks form a virtual vector boson which then radiates a Higgs boson. This process has the third highest cross section. It can also occur at a hypothetical lepton-antilepton collider with sufficient energies. In that case, if the final-state vector boson can be well reconstructed, for example in case it is a Z boson decaying to muons, the mass of the Higgs boson can be precisely measured.

$t\bar{t}H$ production

The production mode of the Higgs boson with the fourth highest cross section and that most relevant to this thesis is the associated production with a top quark-antiquark pair ($t\bar{t}H$). As shown in Fig. 2.11 (d), in this process the Higgs boson couples to a fermion and the final state includes a fermion-antifermion pair. The leading-order matrix elements for this process, such as the one shown in the Feynman diagram, are proportional to the Yukawa coupling $\mathcal{M} \propto y_f$. A measurement of the cross section of $t\bar{t}H$ is therefore a probe for the top Yukawa coupling constant y_t. Since the Yukawa coupling between the Higgs boson and fermions is proportional to the mass this interaction is most likely for top quarks. The cross section for $t\bar{t}H$ production calculated to NLO accuracy and including electroweak corrections, as discussed in Ref. [92], is found to be:

$$\sigma_{t\bar{t}H} = 0.5071^{+0.0291}_{-0.0462}(\text{scale})^{+0.0181}_{-0.0181}(\text{pdf} + \alpha_s)\,\text{pb}. \tag{2.75}$$

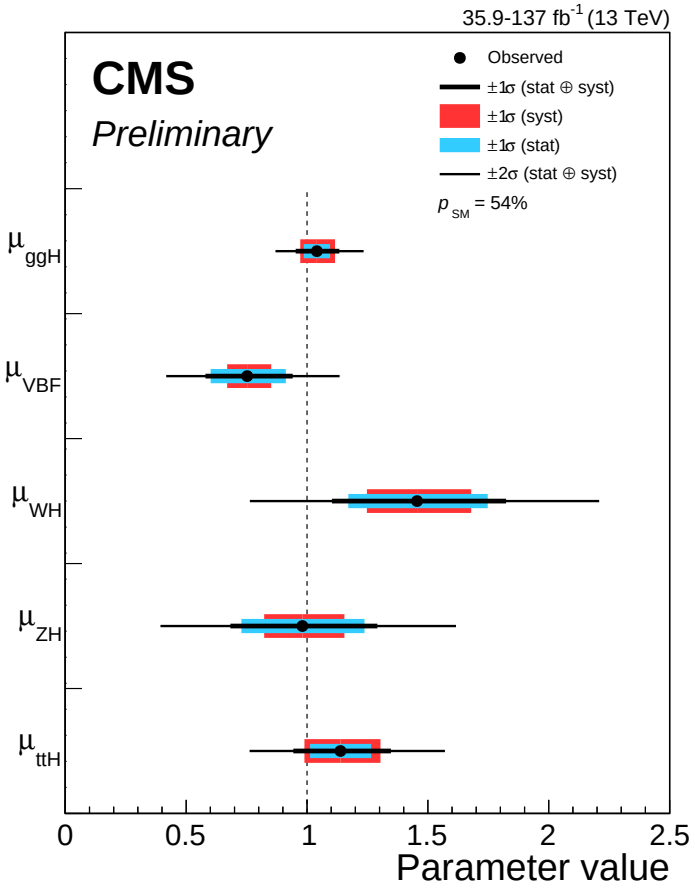


Figure 2.12: Measured signal-strength modifiers $\mu = \sigma/\sigma_{\mathrm{SM}}$ for different production modes of the Higgs boson. Taken from Ref. [122].

Although the mass of the bottom quark is much smaller, for kinematic reasons the production in association with bottom quarks ($b\bar{b}H$) has a similar cross section. It is however much more challenging to analyze. Compared to $b\bar{b}H$, the $t\bar{t}H$ process has the advantage that events including two top quarks can be relatively well identified via the decay products of the top quarks. The signal-strength modifiers for $t\bar{t}H$ production and the top-Higgs Yukawa coupling have been measured by the ATLAS and CMS Collaborations by combining the results of analyses targeting different decay channels, including the analyses described in this thesis. The results of this measurement will be discussed in Ch. 9.

Production in association with a single top quark

It is also possible to produce the Higgs boson in association with a single top quark. This mode has a much smaller cross section than the other production modes. It can occur in three channels: tHq where an additional light quark is in the final state, tHW where the top quark and Higgs boson are produced in association with an additional W boson and tHb in which the final state includes an additional b quark. The cross sections for these processes are calculated at NLO accuracy as described in Ref. [92]. They are:

$$\sigma_{\mathrm{tHq}} = 74.26^{+4.83}_{-10.92}(\mathrm{scale})^{+2.60}_{-2.75}(\mathrm{pdf}+\alpha_{\mathrm{s}})\,\mathrm{fb}, \tag{2.76}$$

$$\sigma_{\mathrm{tHW}} = 15.17^{+0.74}_{-1.02}(\mathrm{scale}) \pm 0.96(\mathrm{pdf}+\alpha_{\mathrm{s}})\,\mathrm{fb}, \tag{2.77}$$

$$\sigma_{\mathrm{tHb}} = 2.875^{+0.069}_{-0.052}(\mathrm{scale}) \pm 0.63(\mathrm{pdf}+\alpha_{\mathrm{s}})\,\mathrm{fb}. \tag{2.78}$$

Measurements of tH production are interesting since the interference effects between contributing Feynman diagrams depend on the relative sign between the couplings of the Higgs boson to the top quark and the W boson. This is further studied in for example Refs. [88, 121].

2.5.2 Higgs boson decays

The Higgs boson can decay in a multitude of ways since it interacts directly with all massive particles and indirectly through loops with massless particles. The branching fractions of these decays depend on the interaction strength and the phase space available to the decay products. The predicted branching fractions for different modes as a function of the Higgs boson mass (the observed Higgs boson mass is $125.10 \pm 0.14\,\mathrm{GeV}$) are shown in Fig. 2.13. At the observed mass, the dominant decay mode is the decay to two bottom quarks $H \to b\bar{b}$ with a branching fraction of about 0.583 ± 0.004 as stated in Ref. [92]. Since over half of all Higgs bosons decay to bottom quarks and since bottom quarks can be relatively well identified, this decay channel is popular for Higgs boson analyses. It is also the decay channel targeted by the analysis in this theses. This is followed by decays to two W bosons and to two gluons, the latter via fermion loops with large contributions from top quarks. While the decays into two Z bosons or photons have relatively small branching fractions, their experimental signatures are rather cleanly identifiable. For this reason these decay channels had an important part in the discovery of the Higgs boson.

Several decay channels have been measured by the CMS Collaboration by combination of several analyses, including the one described in this thesis. The results are shown in Fig. 2.14 where it can be seen that the predictions agree with the observed values within their uncertainties.

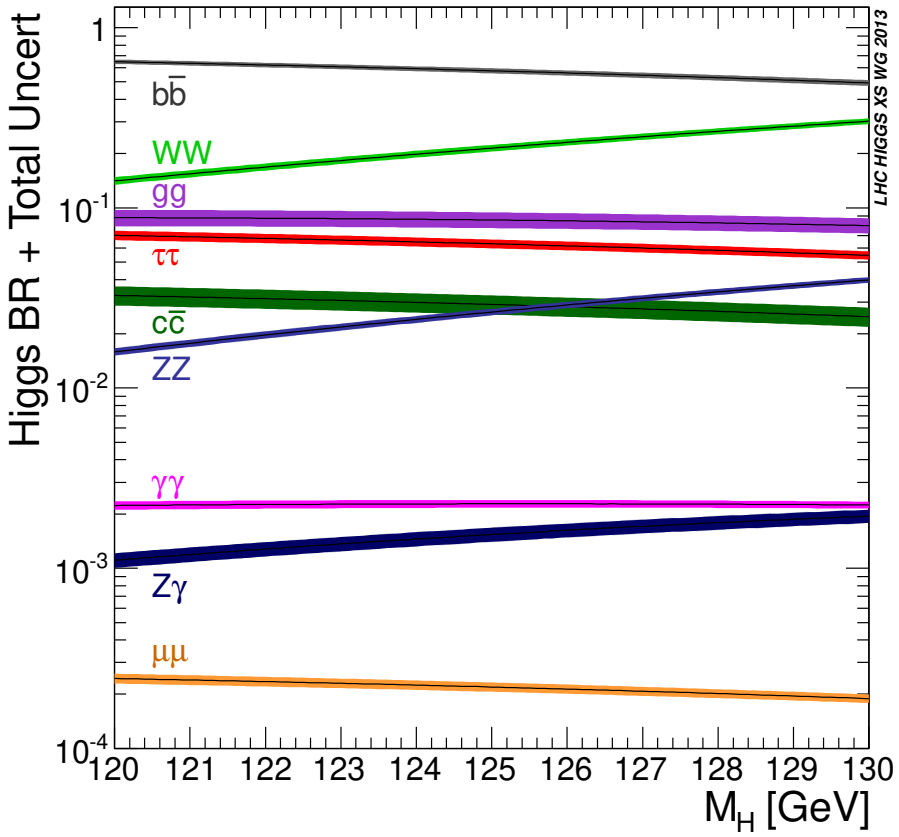


Figure 2.13: Predicted branching fractions of the Higgs boson as a function of the Higgs boson mass. Taken from Ref. [92].

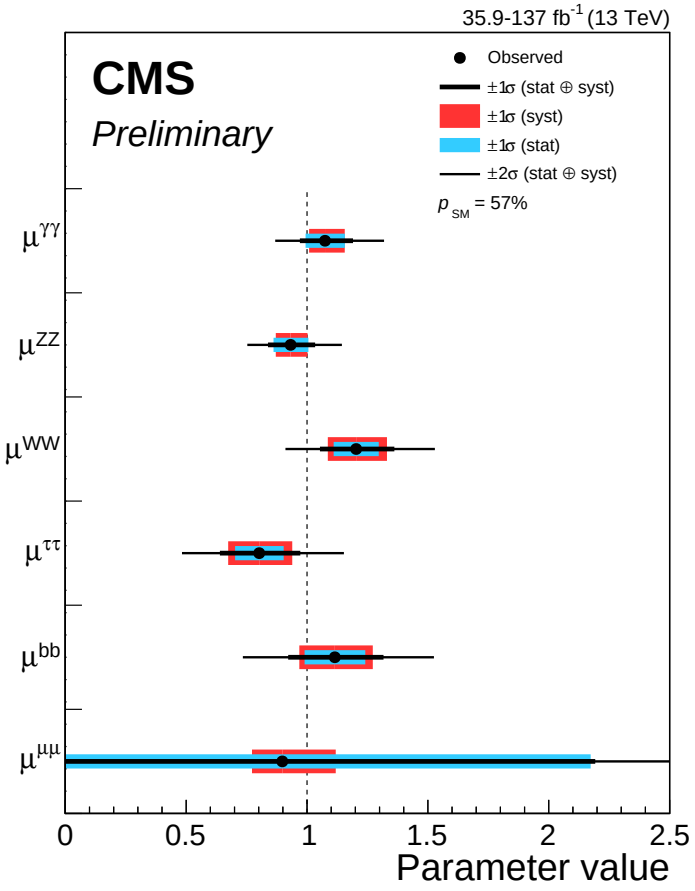


Figure 2.14: Measured signal-strength modifiers $\mu = \mathcal{B}/\mathcal{B}_{\mathrm{SM}}$ for different decay modes of the Higgs boson. Taken from Ref. [122].

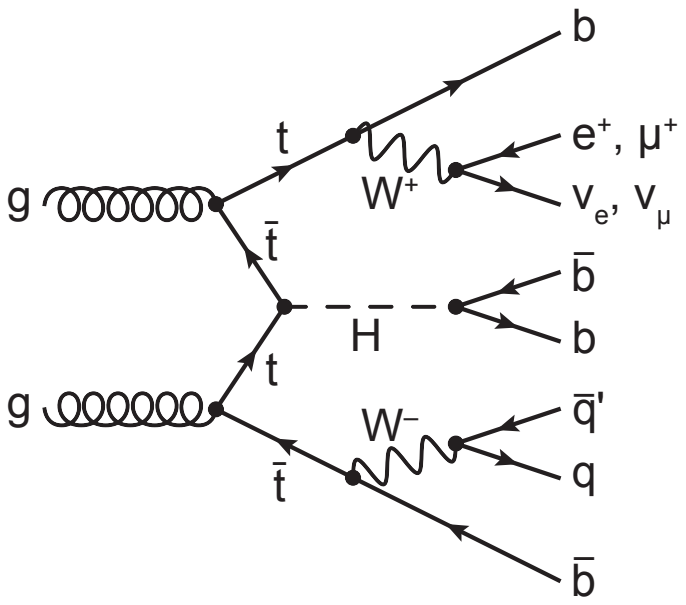


Figure 2.15: Leading-order Feynman diagram for the production of $t\bar{t}H$ with the subsequent $H \to b\bar{b}$ decay and semileptonic decay of the top quark-antiquark pair. Taken from Ref. [4].

2.6 The $t\bar{t}H(b\bar{b})$ process

To study the interaction between the Higgs boson and fermions, processes involving this interaction have to be measured. As was discussed, processes involving the top quark are especially suited for this since the top quark has the largest mass of all fermions and therefore the strongest interaction. The process studied in this thesis is the production of the Higgs boson in association with a top quark-antiquark pair with a subsequent decay of the Higgs boson into a bottom quark-antiquark pair ($t\bar{t}H(b\bar{b})$) and a semileptonic decay of the top quark-antiquark pair. This is motivated by the observation that the production of $t\bar{t}H$ has the largest production cross section of all Higgs boson processes involving the top quark and the Higgs boson decay into bottom quarks has the largest branching fraction. While the hadronic decay of the top quark-antiquark pair has the largest branching fraction, the existence of the lepton in the semileptonic decay allows for a purer event selection and has a branching fraction that is only a little bit smaller. A Feynman diagram of this process is shown in Fig. 2.15. The cross section for $t\bar{t}H(b\bar{b})$ production is given by the cross section of $t\bar{t}H$ production times the branching fraction for the $H \to b\bar{b}$ decay:

$$\sigma_{t\bar{t}H(b\bar{b})} = \sigma_{t\bar{t}H} \times \mathcal{B}_{H\to b\bar{b}} = 0.292^{+0.019}_{-0.029}(\text{scale})^{+0.013}_{-0.013}(\text{pdf} + \alpha_s)\,\text{pb}. \tag{2.79}$$

as listed in Ref. [92].

The main background for $t\bar{t}H(b\bar{b})$ analyses is $t\bar{t}$ production. Its cross section is about one order of magnitude higher than that of $t\bar{t}H(b\bar{b})$ production. Of special importance is $t\bar{t}+b\bar{b}$ production since it has the same final state as the signal process and therefore constitutes an irreducible background. To still be able to measure $t\bar{t}H(b\bar{b})$ with large significances, the signal is usually separated from the background with the help of separating observables (discriminants) constructed using multivariate analysis methods, as is discussed in Ch. 5 and Ch. 7 in the context of this thesis.

The process has been studied at both the CMS experiment and the ATLAS experiment. The feasibility of searches for $t\bar{t}H(b\bar{b})$ with the ATLAS and CMS experiments had been

studied even before the LHC began operations, for example in Refs. [123–126]. Analyses of $t\bar{t}H(b\bar{b})$ production have been performed in the semileptonic and dileptonic $t\bar{t}$ decay channels using $5.1\,\text{fb}^{-1}$ of data recorded at a center-of-mass energy of 7 TeV and $19.7\,\text{fb}^{-1}$ of data recorded at a center-of-mass energy of 8 TeV as described in Refs. [127, 128]. For the extraction of the signal, i.e. the separation of the signal events from the background events, artificial neural networks and boosted decision trees are used. The analyzed channels were combined with analyses targeting other Higgs boson decay channels to exploit correlations of uncertainties between them, resulting in an observed signal-strength modifier of

$$\hat{\mu}_{t\bar{t}H(b\bar{b})} = 0.7 \pm 1.9, \tag{2.80}$$

which is well compatible with a background-only hypothesis in which the $t\bar{t}H(b\bar{b})$ process does not exist. Therefore upper limits at the 95% confidence level[4] were calculated as $\mu_{t\bar{t}H(b\bar{b})} \leq 4.1$ while an upper limit of $\mu_{t\bar{t}H(b\bar{b})} \leq 3.5^{+1.5}_{-1.0}$ was expected for the background-only hypothesis. The data recorded at a center-of-mass energy of 8 TeV were also analyzed in the semileptonic, dileptonic and hadronic $t\bar{t}$ decay channels using a matrix-element method for the extraction of the signal, as described in Ref. [129], with a result of

$$\hat{\mu}_{t\bar{t}H(b\bar{b})} = 1.2^{+1.6}_{-1.5} \tag{2.81}$$

and an observed (expected) upper limit of $\mu_{t\bar{t}H(b\bar{b})} \leq 4.2$ $(3.3^{+1.6}_{-1.0})$ presenting therefore a more sensitive approach. An analysis using $20.3\,\text{fb}^{-1}$ of data recorded at a center-of-mass energy of 8 TeV in the semileptonic and dileptonic $t\bar{t}$ decay channels was performed by the ATLAS Collaboration using neural networks, with a matrix-element method being used to calculate input features for the neural networks, as described in Ref. [90]. This resulted in an observed signal-strength modifier of

$$\hat{\mu}_{t\bar{t}H(b\bar{b})} = 1.5 \pm 1.1 \tag{2.82}$$

and an observed (expected) upper limit of $\mu_{t\bar{t}H(b\bar{b})} \leq 3.4$ (2.2 ± 0.8). This analysis was combined with an analysis in the hadronic $t\bar{t}$ decay channel, using boosted decision trees for the signal extraction, and other Higgs boson decay channels, described in Ref. [130], resulting in

$$\hat{\mu}_{t\bar{t}H(b\bar{b})} = 1.4 \pm 1.0 \tag{2.83}$$

and an observed (expected) upper limit of $\mu_{t\bar{t}H(b\bar{b})} \leq 3.3$ $(2.1^{+0.9}_{-0.6})$. All presented analyses performed during LHC Run 1 were neither able to observe nor to exclude the existence of $t\bar{t}H(b\bar{b})$ production.

A search for $t\bar{t}H(b\bar{b})$ was also performed by the CMS Collaboration using the first $2.7\,\text{fb}^{-1}$ of data recorded in the year 2015 at a center-of-mass energy of 13 TeV. The analysis was performed in the semileptonic and dileptonic $t\bar{t}$ decay channels as described in Ref [131]. The signal extraction was performed using a combination of boosted decision trees and a matrix element method. Depending on the phase space, the separating observable was either obtained using a boosted decision tree with the output of the matrix element method used as input feature or the matrix element method was combined with the boosted decision trees to construct two-dimensional separating observables. This analysis also featured a special "boosted" phase space region enriched in events where the top quarks had high transverse momenta. In this boosted region, the top quarks can be more efficiently reconstructed than in other non-boosted regions, allowing for a more efficient event reconstruction and signal separation using boosted decision trees. The studies of this technique in preparation of this

[4] If not stated otherwise, quoted upper limits are always derived at the 95% confidence level.

analysis are described in Refs. [132–135]. The result of this analysis is a signal-strength modifier of

$$\hat{\mu}_{\mathrm{t\bar{t}H(b\bar{b})}} = -2.0 \pm 1.8 \tag{2.84}$$

and an observed (expected) upper limit of $\mu_{\mathrm{t\bar{t}H(b\bar{b})}} \leq 2.6$ $(3.6^{+1.6}_{-1.1})$. The signal-strength modifier was found to be negative and compatible with the background-only hypothesis. The negativity of the signal-strength modifier was attributed to statistical fluctuations.

The analyses in this thesis analyze data recorded in the years 2016 and 2017 at a center-of-mass energy of 13 TeV. Nevertheless, a preliminary analysis of the first $12.9\,\mathrm{fb}^{-1}$ of data recorded in the year 2016 was performed and is discussed in Ref. [136]. In this analysis the signal separation was achieved using a two-dimensional separating observable combining a matrix element method and boosted decision trees, as in the previous analysis. The analysis in the boosted region was omitted due to the required computational and calibration efforts and the low expected impact on the result. Preparatory studies for the analysis in the semileptonic channel are discussed in Refs. [137–139]. The result of this analysis is

$$\hat{\mu}_{\mathrm{t\bar{t}H(b\bar{b})}} = -0.19^{+0.80}_{-0.81} \tag{2.85}$$

with an observed (expected) upper limit of $\mu_{\mathrm{t\bar{t}H(b\bar{b})}} \leq 1.5$ $(1.7^{+0.7}_{-0.5})$ with a clearly improved increase in sensitivity compared to the previous analysis. The negativity of the signal-strength modifier was again attributed to statistical fluctuations.

Results of the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses performed by the CMS and ATLAS Collaborations with all data recorded in the years 2016 and 2017 will be discussed in Ch. 9.

3 The CMS Experiment at the LHC

To test how well physical theories such as the standard model describe nature and to measure quantities not predicted by the models it is necessary to conduct experiments. In high energy physics such experiments often are conducted at particle accelerators where particles of various types are accelerated to high energies and then brought to collision with either a fixed target or with other accelerated particles. The analysis of the $t\bar{t}H(b\bar{b})$ process presented in this thesis is performed using data, recorded with the Compact Muon Solenoid (CMS) experiment, of proton-proton collisions occurring at the Large Hadron Collider (LHC) at CERN. The LHC is currently the accelerator providing the highest center-of-mass energies and the only accelerator where the direct production of a Higgs boson in association with a top quark-antiquark pair is possible while the CMS experiment is a detector system designed to detect, reconstruct and measure all particles occurring in collisions at the LHC. The LHC is discussed in Sec. 3.1 of this chapter, followed by a description of the CMS experiment in Sec. 3.2.

3.1 The Large Hadron Collider

The LHC is a hadron storage ring and collider at the European Organization for Nuclear Research (CERN) near Geneva (Switzerland). It is situated below ground underneath the border region between France and Switzerland in a tunnel with about 27 km circumference, formerly housing the Large Electron Positron collider (LEP). In it protons and ions of heavy elements can be accelerated to the, as of yet, highest achieved energies. The LHC is described in detail in Ref. [140] and further information about CERN is given in Ref. [141]. Protons are first accelerated using a chain of pre-accelerators and then injected into the LHC where they are further accelerated to energies of up to 7 TeV using superconducting radio frequency cavities. Inside the LHC the protons travel in two spatially separated beam pipes in opposite directions. For technical reasons the protons do not travel in a continuous beam, but instead are separated into packets called bunches. Each bunch contains on the order of 10^{11} protons and up to 2808 bunches with a minimal bunch spacing of 25 ns can be filled into each LHC beam pipe. The beams of protons are forced to follow the ring-shape using superconducting dipole magnets and focused with superconducting quadrupole and higher-order magnets. After the desired energy is reached, the protons are not further accelerated and the radio frequency cavities only compensate the energy the protons lose from synchrotron radiation. Protons are not continually filled into the LHC, instead a single fill with the required number of bunches and the number of protons per bunch is

filled into the storage ring. These are then circulated and collided until the density and quality of the beam degrades below a threshold, usually after several hours, at which point the beams are discarded and the cycle begins again.

At four interaction points magnets are used to bend and focus the beams to bring them to collision, allowing interactions between protons of the two beams. The protons in both beams are accelerated to identical energies of up to 7 TeV resulting in center-of-mass energies of up to 14 TeV for these interactions. To measure the particles that are produced in these collisions large experiments are situated at these four interaction points. These are the ALICE experiment, described in Ref. [142], the LHCb experiment, described in Ref. [143] and the ATLAS and CMS experiments, described in Refs. [144, 145]. The ALICE experiment is specialized in the observation of collisions of heavy nuclei and the study of the strong interaction and the quark-gluon plasma. The detector of the LHCb experiment is designed to study rare B hadrons and CP-violation but has since extended its physics program to other measurements benefiting from its very good vertex reconstruction. The ATLAS and CMS experiments employ multipurpose detectors to study the interactions of the standard model and to search for new physics.

To allow for the precise study of the standard model and observation of rare processes a large number of proton-proton collisions is required. The total rate of proton-proton collisions depends on the total proton-proton cross section σ_{pp} and a quantity called *instantaneous luminosity* that measures how many proton-proton pairs per unit time come near enough to each other to meaningfully interact. This instantaneous luminosity is explained in Refs. [146, 147] and defined as:

$$L = \frac{fnN_1N_2}{4\pi\sigma_x\sigma_y}. \tag{3.1}$$

Here, f is the revolution frequency, n the number of bunches, N_1 and N_2 denote the number of protons per bunch and σ_x and σ_y are the spreads of the bunches perpendicular to the directions of flight at the interactions points. Additional effects such as those arising from the crossing angle of the beams and possible non-gaussian shape of the bunches are neglected in the above equation. The rate of proton-proton interactions is then given by

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \sigma_{\mathrm{pp}} \cdot L. \tag{3.2}$$

The instantaneous luminosity can be measured as is explained for example for the measurement by the CMS experiment in Ref. [148, 149] in which a precision of 2.5% is reached. The LHC is designed to provide instantaneous luminosities of up to $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for proton-proton collisions with protons of 7 TeV energy, although in practice for runs with 6.5 TeV protons about double that value was reached as stated in Ref. [150].

A measure of the total luminosity over a given time is often used to characterize the size of a sample of collision data and is given by the integrated luminosity

$$L_{\mathrm{int}} = \int L \,\mathrm{d}t, \tag{3.3}$$

usually measured in units of inverse femtobarn with $1\,\mathrm{fb}^{-1} = 10^{-39}\mathrm{cm}^{-2}$. The expected number of collisions with inelastic proton-proton scattering resulting in a specific process can then be calculated by

$$N = L \cdot \sigma \tag{3.4}$$

with the production cross section σ for the process.

The LHC first started operations in 2008 but an accident involving the destruction of several dipole magnets led to a delay for the start of the physics program until 2010. Between 2010

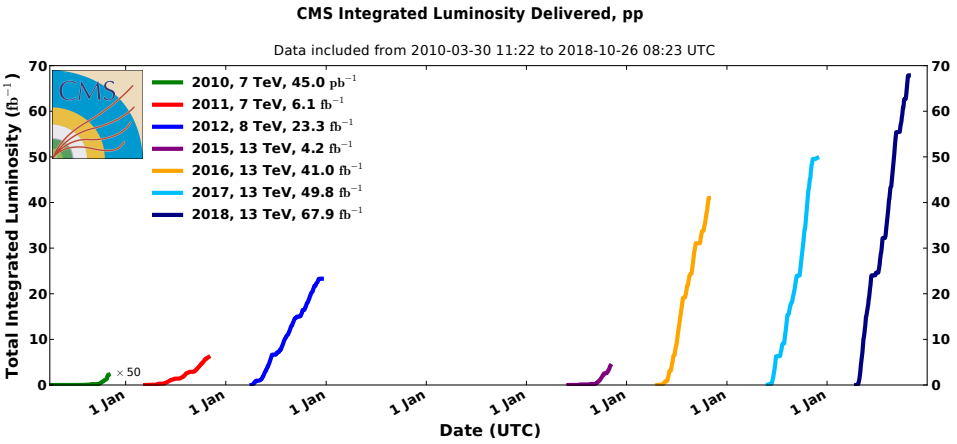


Figure 3.1: Integrated luminosities delivered to the CMS experiment during proton-proton runs. Taken from Ref. [150].

and 2012, a time-frame called *LHC Run 1*, the LHC delivered integrated luminosities of $6.1\,\mathrm{fb}^{-1}$ and $23.3\,\mathrm{fb}^{-1}$ at center-of-mass energies of 7 TeV and 8 TeV, respectively. After this successful first run the LHC was shut down for two years and received various improvements to allow higher energies. Operations continued in 2015, beginning LHC Run 2, which continued until 2018 and during which the LHC delivered an integrated luminosity of $158.6\,\mathrm{fb}^{-1}$ at a center-of-mass energy of 13 TeV and a higher instantaneous luminosity. The integrated luminosities for proton-proton runs at these center-of-mass energies are shown Fig. 3.1. The LHC also performed proton-proton runs with lower center-of-mass energies for development or commissioning purposes, as well as runs with proton-lead or lead-lead collisions. The integrated luminosities of all runs recorded by the CMS experiment can be found in Ref. [150].

After the successful completion of Run 2 the LHC was again shut down to make additional improvements for the upcoming Run 3 with a center-of-mass energy of 14 TeV which is planned to begin in 2021 and end in 2024 as stated in Ref. [151]. After Run 3 the LHC is scheduled to be upgraded into the High Luminosity LHC (HL-LHC), discussed for example in Ref. [152], with even higher instantaneous luminosities with the goal to accumulate $3000\,\mathrm{fb}^{-1}$ of proton-proton collisions at a center-of-mass energy of 14 TeV after several years of run-time.

3.2 The CMS Experiment

The Compact Muon Solenoid experiment is a multipurpose detector system designed and operated by the CMS Collaboration. A full description of the CMS experiment is given in Ref. [145]. It consists of several subdetectors that are layered around one of the interactions points of the LHC with the purpose of detecting and measuring all particles arising in the proton-proton interactions. An overview over the CMS detector as installed in the year 2018 is given in Fig. 3.2.

The whole detector is about 28.7 m long, has a diameter of about 15 m and weighs about 14,000 tons. The layered cylindrical design is motivated by the properties of the various particles that can arise from the collisions. As will be discussed in Ch. 6, not all particles are able to travel through the whole detector and are instead absorbed by the material of certain subdetectors. So subdetectors targeting different kinds of particles have to be situated in a specific order from the interaction point which lies in the center of the detector and to which protons are guided within beam pipes.

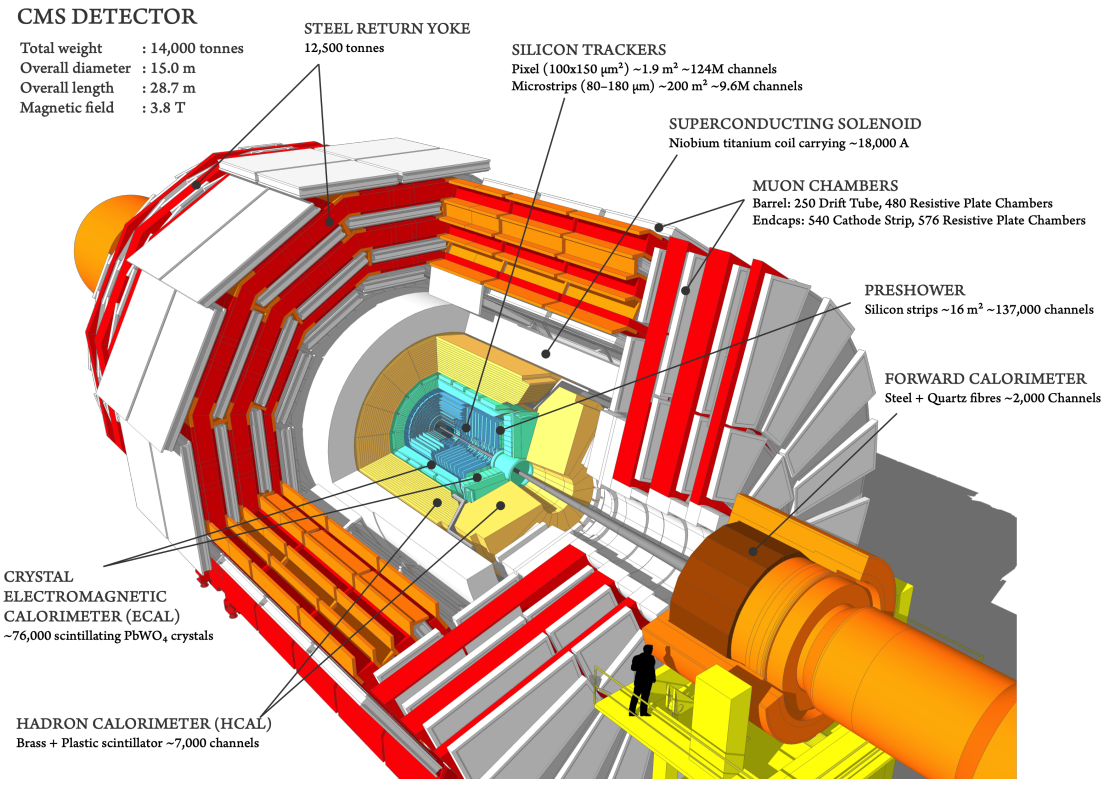


Figure 3.2: Overview over the CMS detector. The interaction point and the beam pipes are situated in the center and surrounded by the tracker system, the electromagnetic and hadron calorimeters, the solenoid magnet and finally the muon detector systems which are embedded in a iron return yoke guiding the magnetic field. Taken from Refs. [153, 154].

Going outward the first subdetector system is the tracking system whose purpose is the measurement of the trajectories of all charged particles occurring in the collisions. Surrounding this are the electromagnetic and hadron calorimeters that measure the energies of electromagnetically and hadronically interacting particles, respectively. Around all of these systems winds the coil of the superconducting solenoid magnet. It has a diameter of 6 m, a length of 12.5 m and is capable of producing a magnetic field of 3.8 T inside the coil. The purpose of the magnetic field is to bend electromagnetically charged particles via the Lorentz force, thereby allowing the calculation of their momenta from the curvature of their trajectories. The outermost subdetectors are the muon systems designed to identify and measure the momenta of muons. They are embedded within an iron support structure that also serves as flux-return yoke guiding the magnetic field outside of the solenoid magnet. A brief overview of these subdetectors is given in Secs. 3.2.2 to 3.2.5.

The nominal bunch spacing of 25 ns of the LHC corresponds to a bunch crossing frequency of 40 MHz. With current technology it is impossible to fully read-out and store the information of all channels of the CMS detector for each bunch crossing. Fortunately the cross sections of interesting processes are small compared to the total proton-proton cross section. Therefore, as is briefly discussed in Sec. 3.2.2, a two-tiered trigger system is used to identify and store only interesting bunch crossings.

Finally, this chapter concludes in Sec. 3.2.6 with a brief note about the processing of the data recorded with the CMS detector and used in this thesis. The reconstruction of particles detected with the CMS detector will be discussed in Ch. 6.

3.2.1 The coordinate system

The coordinate system used at the CMS experiment is a right-handed Cartesian coordinate system centered on the interaction point. The x-axis points from the interaction point towards the center of the LHC ring, the y-axis points upwards and the z-axis points in the direction of the counter-clockwise moving beam towards the Jura mountains near CERN. The orientation around the z-axis is measured with the azimutal angle ϕ with $\phi = 0$ pointing in the direction of the x-axis and the orientation around the x-axis is measured with the polar angle θ with $\theta = 0$ pointing in positive z direction. The two beams have equal and opposite momenta in the z-direction and the protons have negligible momenta in the x- and y-directions. The protons themselves are however composite particles made up of partons and the fractions of the proton momenta carried by the partons are not known. Therefore the center-of-mass of the partonic interactions can move along the z-axis with a non-zero momentum in the laboratory frame and only the transverse component of the momentum of detected particles can be measured. This quantity is denoted as p_T and defined as

$$p_\mathrm{T} = \sqrt{p_\mathrm{x}^2 + p_\mathrm{y}^2}. \tag{3.5}$$

To describe the polar angle of the particles with respect to the partonic center-of-mass frame a quantity called *pseudorapidity* η is used. It is related to the rapidity defined as

$$y = \frac{1}{2}\left(\frac{E + p_\mathrm{z}}{E - p_\mathrm{z}}\right) \tag{3.6}$$

which goes to zero for particles with $p_\mathrm{z} = 0$ and to $\pm\infty$ for particles with $p_\mathrm{T} \approx 0$. Furthermore, differences between the rapidities of two particles are invariant under Lorentz-boosts in the z-direction, making the rapidity a good quantity to measure angles between particles. For highly relativistic particles with $m \ll p$, such as most of those produced in the proton-proton collisions, the rapidity reduces to the pseudorapidity η defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{3.7}$$

thereby avoiding the immeasurable p_z momentum.

A measure often used for the angular separation between two particles is the quantity ΔR calculated from their differences in the azimutal angle and the pseudorapidity as

$$\Delta R = \sqrt{(\Delta\theta)^2 + (\Delta\eta)^2}. \tag{3.8}$$

3.2.2 The tracking system

The innermost subdetector is the tracking system with the purpose of measuring the *tracks*, i.e. the trajectories, of electromagnetically charged particles. Due to the strong magnetic field inside the solenoid, the tracks of the particles are bent. A measurement of the curvature of these tracks allows the reconstruction of the particle momenta. The precise measurement of the tracks also allows the reconstruction of the interaction and decay vertices of such particles, which can be used to identify different kinds of particles as will be discussed in Ch. 6. For the measurement of the tracks, each particle has to be detected multiple times along its trajectories, and the positions of these detections need to be measured precisely. The tracking system therefore consists of multiple layers of finely segmented silicon sensors installed inside a cylindrical shape around the interaction point. The detection of the particles is achieved by forming a p-n junction out of positively and negatively doped silicon to which a bias voltage is applied in reverse-bias direction

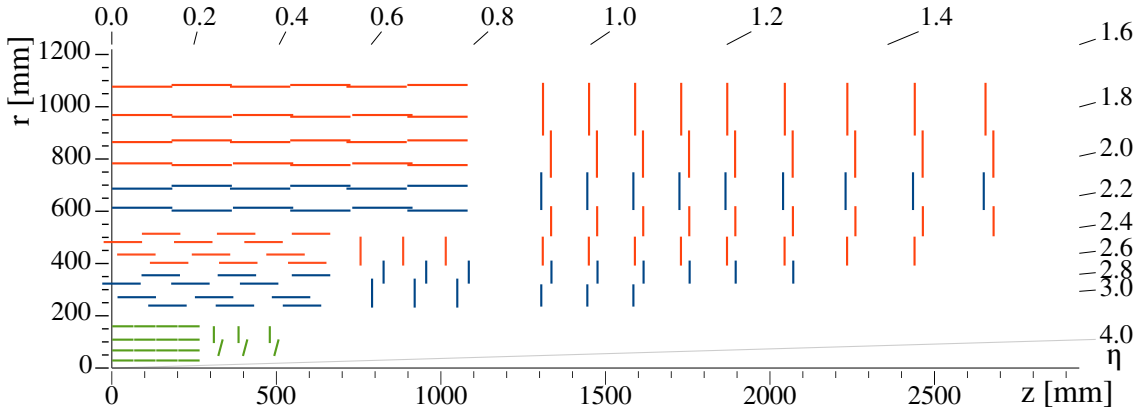


Figure 3.3: Layout of the CMS tracking system after the phase 1 upgrade. The interaction point is in the lower left corner. Closest to the interaction point are the silicon pixel detectors shown in green. Surrounding these are single- and double-sided silicon strip detectors shown in red and blue respectively. Taken from Ref. [158].

thereby creating a large depletion region. Charged particles traversing this region lead to the creation of electron-hole pairs, inducing a current, which is amplified and measured.

A sketch of the layout of the tracking system is shown in Fig. 3.3.

The innermost layers consist of silicon pixel detectors made up of pixels with a size of 100 μm by 150 μm. Due to the high flux of particles and therefore high occupancy of the pixels, each pixel is designed to be read-out individually allowing a precise determination of the detection in three dimensions. Up until the end of the year 2016, 66 million pixels were arranged in three cylindrical layers around the interaction point in the barrel section and two disks in each of the two endcap sections. Between 2016 and 2017 the so-called phase-1 upgrade was performed during which the pixel detectors were upgraded and partially replaced as described in Ref. [155]. The replacement of old pixel detectors was due to radiation damage caused by the high flux of energetic particles. In addition, additional pixel detectors were installed to improve the track reconstruction in light of increasing instantaneous luminosities. For this, an additional layer was inserted in the barrel section, moving the existing layers further outward. An additional layer was also inserted in each of the two endcaps extending the coverage of the pixel detector from $|\eta| < 2.4$ to $|\eta| < 2.5$. With this the total number of pixels was brought to 124 million.

Further away from the interaction point the flux of particles is lower and the tracking system consists of 15,148 silicon strip detector modules with about 9.6 million read-out channels in total. They are arranged in 10 cylindrical layers in the barrel section and 12 disks in the endcaps, as shown in Fig. 3.3, providing a coverage of $|\eta| < 2.5$. The strips in the barrel section are oriented to provide a precise measurement of the $r - \phi$ coordinate while those in the endcap modules provide a precise measurement of the $z - \phi$ direction. Several modules are double-sided with small relative angles between both sides, thereby providing a third component to the position of the particle.

A full description of the tracking system is given in Refs. [145, 156, 157]

3.2.3 The electromagnetic calorimeter

The energies of photons and electromagnetically charged particles are measured with the electromagnetic calorimeter (ECAL) system described in Refs. [145, 159]. The ECAL is a system of homogeneous calorimeters consisting of 75,848 transparent lead tungstate

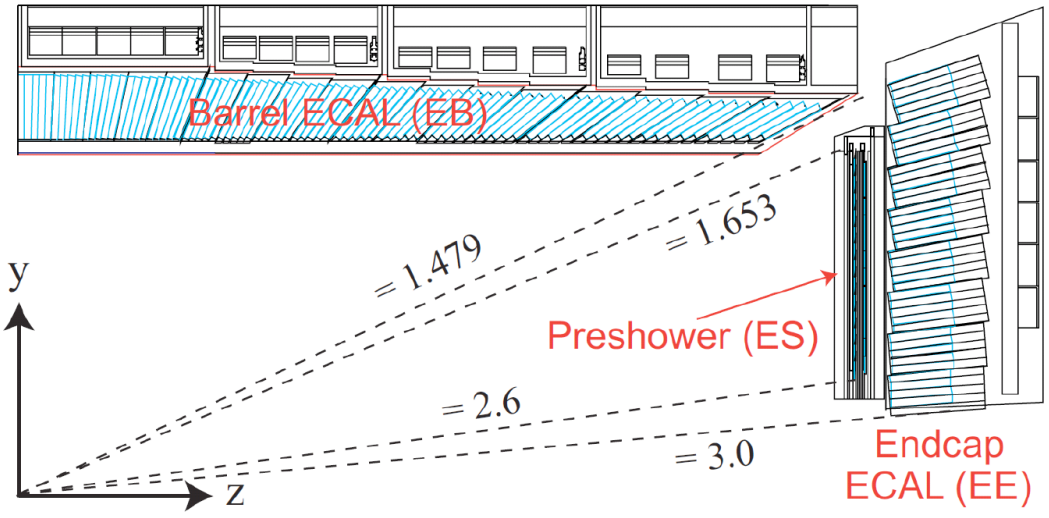


Figure 3.4: Layout of the CMS ECAL system consisting of lead tungstate crystals ($PBWO_4$) crystals. It is subdivided in a barrel (EB) and endcap (EE) detector. In the endcap region an addition preshower detector is installed. Taken from Ref. [160].

($PBWO_4$) crystals that are arranged around the tracking detector as depicted in Fig. 3.4. Electromagnetically charged particles and photons are absorbed in the crystal, thereby creating electromagnetic showers that result in scintillation light. Since the intensity of this light is proportional to the energy of the absorbed particle, a measurement of the light leads to a measurement of the particle energy. Using optical fibers, the scintillation light is led to photodetectors where it is measured. The ECAL is divided into a barrel section covering a region of $|\eta| < 1.479$ and endcap section covering a region of $1.479 < |\eta| < 3.0$. The crystals are segmented allowing the determination of the position of an energy measurement. The length of the crystals corresponds to $25.8X_0$ in the barrel and $24.7X_0$ in the endcap sections, where X_0, called radiation length, is the distance over which an electron loses all but $1/e$ of its energy. A large depth in radiation lengths is important to fully absorb the energy of the particles. In the region of $1.653 < |\eta| < 2.6$ a preshower detector is installed, consisting of a two-layer sampling calorimeter constructed from lead absorbers and silicon strip sensors. Its purpose is the detection of photons produced in the decays of neutral pions.

3.2.4 The hadron calorimeter

The energies of strongly interacting particles are measured with the hadron calorimeter (HCAL) system described in Refs. [145,161]. The HCAL is a system of sampling calorimeters consisting of alternating layers of brass absorber material and plastic scintillators. Strongly interacting particles cause hadronic showers in the brass absorber material, which extend outside of the brass layers into the scintillation layers. These showers contain particles that can induce electromagnetic showers leading to the production of measurable scintillation light in the plastic scintillators. The intensity of this light is again proportional to the energy of the absorbed strongly interacting particle. The layout of the HCAL system is shown in Fig. 3.5.

The hadron barrel (HB) and hadron endcap (HE) systems are located outside the ECAL and inside the coil of the solenoid magnet. The HB covers a range of $|\eta| < 1.3$ while the HE covers $1.3 < |\eta| < 3.0$ The total depths of the HB and HE in terms of the nuclear interaction length λ vary as a function of η between 5.82λ and 10.6λ. The ECAL crystals

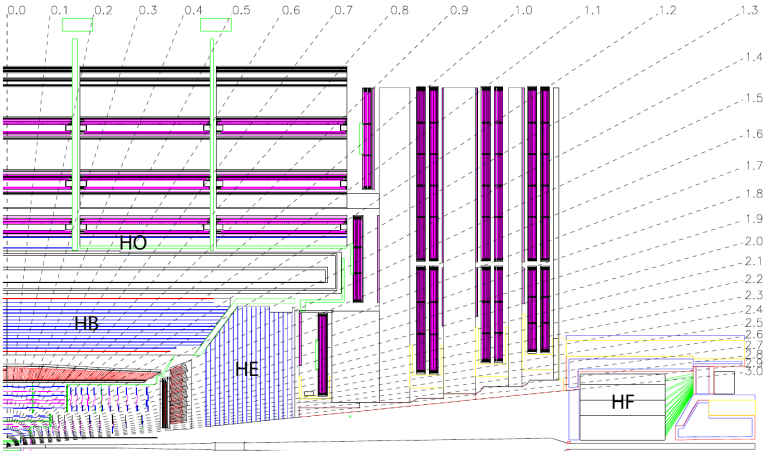


Figure 3.5: Layout of the CMS HCAL system with its subdetector systems: the hadron barrel (HB) and hadron endcap (HE) detectors inside the coil of the solenoid magnet as well as the hadron outer (HO) detector outside the coil of the solenoid magnet. In the very forward region an additional hadron forward (HF) detector is installed. Taken from Ref. [145].

in front of the HCAL add another 1.1λ. In the barrel region the combined depth of the ECAL and HB is not enough to fully stop all strongly interacting particles. Since the depth of the HB is constrained by the radius of the magnet coil, an additional hadronic outer (HO) calorimeter system is installed outside of the coil. The HB, HE and HO are segmented to allow for position determination of the measured energies. The HB, HE and HO systems are complemented by the hadron forward (HF) system covering the very foreward range of $3.0 < |\eta| < 5.0$ where the particle flux is very high. To withstand this the HF consists of steel absorber material interspersed with quartz fibers as scintillators.

3.2.5 The muon system

Muons with energies typical at the LHC are minimal ionizing particles and therefore only deposit small amounts of energy in detector materials. While most electromagnetically and strongly interacting particles are completely stopped by the calorimeters, such muons can escape the calorimeters and traverse further outward. Their detection is performed by the muon system described in Refs. [145, 162]. Since they are usually the only particles detected by the muon system, muons are easily identified. The muon system consists of multiple detectors that are installed inside the iron of the flux-return yoke and cover a range of $|\eta| < 2.4$. An overview over the muon system is shown in Fig. 3.6. Three types of detectors are used, drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). All three of these are gaseous detectors where muons are detected by the measurement of the charged particles they create when ionizing the gas. Their performance is discussed in Refs. [163, 164]. In the barrel region with $|\eta| < 1.2$ the rate of muons is low and the magnetic field is rather uniform. In this region DTs with good time and position resolutions are used. In the forward region the rate of muons is higher and the magnetic field more non-uniform. Here CSCs with an excellent position resolution and a slightly worse time resolution are used, covering a region of $0.9 < |\eta| < 2.4$. In both regions RPCs are used. Their position resolution is low compared to DTs and CSCs but they have a good time resolution.

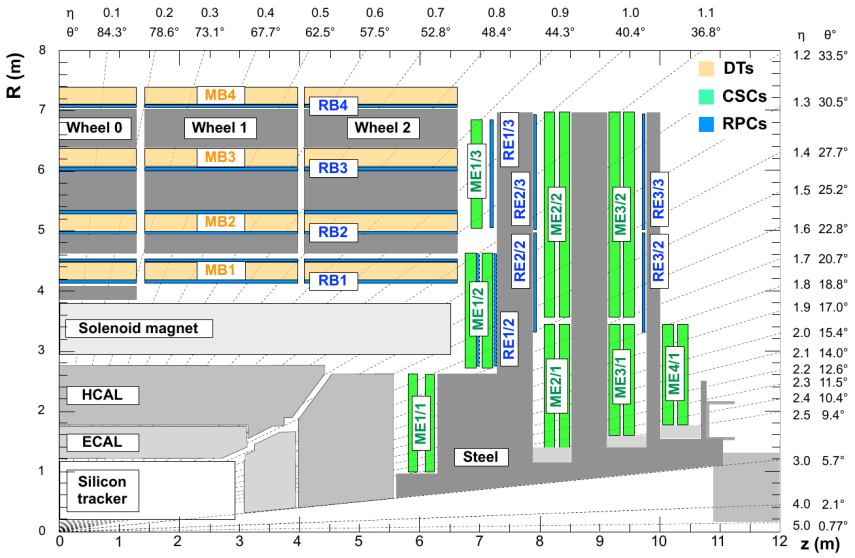


Figure 3.6: Layout of the CMS muon system consisting of drift tubes (DT) in the barrel region, cathode strip chambers (CSC) in the endcap region and resistive plate chambers (RPC) in both regions. Taken from Ref. [163].

3.2.6 The trigger system

Not only one but several inelastic proton-proton interaction events occur during a typical bunch crossing. For the nominal instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ and an inelastic proton-proton cross section of $\sigma_{\text{inelastic}} = 69.2\,\text{mb}$ (see for example Ref. [165]) an average of 17 inelastic proton-proton collision events occur per bunch crossing, although in practice the instantaneous luminosity and the number of these *pileup* interactions is even higher, as can be seen in Ref. [150]. As a result the amount of data produced by the CMS detector per bunch crossing is of the order of 1 MByte as discussed in Ref. [145]. With a bunch crossing frequency of 40 MHz this results in more data than could be realistically stored. Fortunately most bunch crossings do not result in interesting events [1]. To exploit this fact a very fast selection is implemented in form of a two-tiered trigger system that is described in detail in Refs. [145, 166–168].

The first stage is called L1 trigger and has to decide within 3.2 μs if an event should be discarded or kept for further analysis. During this time, the information of the event is buffered inside the detector components. To achieve this, the L1 trigger is implemented as parallelized as possible using custom-designed electronics such as field programmable gate arrays (FPGAs) or application-specific integrated circuits (ASICs). To minimize latency effects on decision time the components of the L1 trigger are partly installed inside the CMS detector and partly in a nearby service cavern. The trigger decision is based on a combination of data from the calorimeters and muon system. Muon detector hits are first compared to nearby muon hits to identify possible tracks and then combined in a global muon trigger system that collects all identified muon tracks and ranks them according to their reconstruction quality. At the same time energy clusters are calculated from nearby calorimeter cells and compared to threshold values. The global calorimeter trigger then combines all interesting energy clusters to calculate global observables and to identify

[1] Each bunch crossing usually only contains at most one interesting interaction event. For the remainder of this thesis the term *event* will denote all of the recorded data associated with a single bunch crossing while the term *pileup* will denote all proton-proton interactions in an event except the single interesting one which is usually that with the highest total energy imparted on its final-state particles.

candidates for jets, leptons and photons. Finally, the information is combined in the global trigger, which compares the identified objects to a list of 128 simple selection criteria.

If the event is kept the detector is read out and the full data of the event is sent to the high level trigger (HLT) which is a software trigger run on a nearby computing cluster. With the HLT the event is reconstructed using fast and optimized reconstruction methods, similar to those described in Ch. 6. This includes the clustering of jets and the reconstruction and identification of electron, muon, τ or photon candidates. To reduce the time until the HLT trigger decision is made, the reconstruction of the event is structured around so-called HLT trigger paths. These are predefined lists of processing steps of increasing complexity that implement both the reconstruction and the selection of objects as well as the event selection with the purpose of avoiding unnecessary computations. The list of HLT trigger paths is called the trigger menu and created to accommodate a large number of physics analyses performed at the CMS experiment.

3.2.7 Computing

The experiments at the LHC require a huge amount of computing capacity and data storage. The large amount of data recorded with the experiments have to be reconstructed and securely stored. In addition, even larger samples of simulated data have to be generated in computationally intensive processes. Both recorded and simulated data have to be stored and made available to researchers around the world. The physics analyses performed by these researchers require additional processing power and storage space. All this is achieved with the use of a worldwide network of computing centers called the Worldwide LHC Computing Grid (WLCG), which is described in Refs. [169, 170].

Both the reconstruction of the raw data as well as the creation of different data samples usable by analysts is performed using the CMS software (CMSSW) framework which is documented in Refs. [171, 172]. The CMSSW framework is also used for the first processing steps of the data used in the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis described in this thesis. Subsequent processing steps are performed using the ROOT framework, which is described in Ref. [173] and which was originally developed for the usage of data recorded by the LHC experiments, as well as the KERAS and Tensorflow packages, which are discussed in Refs. [174, 175]. Statistical analyses of the data, discussed in Ch. 4 and Ch. 9, are performed using ROOT as well as the RooStats and RooFit software packages, which are described in Refs. [176, 177].

4 Statistical Methods

By comparing the number of events observed in recorded data to that predicted by a physical model, inferences about the content of recorded data can be made. To make such inferences quantifiable and to reliably differentiate between different possible conclusions such comparisons are performed using rigorous statistical hypothesis tests.

In the context of searches for undiscovered physical processes, the tested hypotheses usually are the background-only hypothesis, consisting of the predictions made by the standard model without the new process, and a signal+background hypothesis for which the new process is added to the background. For both hypotheses a *test statistic* is constructed that quantifies, and increases with, the disagreement between the observed data and the data that would be expected if the hypothesis were true. Next, the *p-values* for the hypotheses are calculated, defined as the probability that, given the hypothesis is true, the found or an even larger test statistic value is observed. A hypothesis is rejected if this probability is below a predetermined *confidence level* which is usually denoted α and often chosen as $\alpha = 0.05$.

In searches for a new process, the presence of the process usually leads to an increased number of events compared to the prediction by the background-only hypothesis. Such searches therefore usually entail a selection of a suitable phase space and the application of signal enhancing methods with the goal to make the analysis as statistically significant as possible, i.e. to make the p-value for the background-only hypothesis as small as possible in the presence of the signal process. The significance of the disagreement between the observed data and the background-only hypothesis is usually presented in terms of the p-value for the background-only hypothesis.

For measurements of a physical parameter, such as a cross section, the decision to be made is not between a background-only and a signal+background hypothesis but instead one between many hypotheses: one for each possible value of the parameter.

A concise introduction to statistical hypothesis tests is given for example in Refs. [75,178] and statistical methods in the context of particle physics are discussed in Ref. [7]. All statistical calculations are performed using the CMS Higgs combination software, documented in Ref. [179], which implements a statistical framework that was developed by the ATLAS and CMS Collaborations and is described in Ref. [180]. The software package is based on the ROOT, RooStats and RooFit software packages that are described in Refs. [173, 176, 177].

Section 4.1 introduces the statistical model used for the $t\bar{t}H(b\bar{b})$ analysis. This is followed by discussions about the calculation of the significance of a possible signal in Sec. 4.2, the measurement of the signal in Sec. 4.3 and a short discussion about the calculation of expected results in Sec. 4.4.

4.1 Statistical model for the $t\bar{t}H(b\bar{b})$ analysis

Statistical analyses in high energy physics are usually performed counting the number of events in a specific phase space and comparing the number of events in recorded data to that expected from a specific model. Since the cross sections of relevant processes are small and the number of experiments is large (in most proton-proton collisions at the LHC nothing very interesting happens) the probability to observe n events in data, given an expected number of events λ, follows a Poisson distribution. Often one is not interested in the number of events in a single phase space region, but instead in the distribution of an observable, in which case the events are filled into a histogram with m bins i. The probability for the observation of the data is then given by the product of the Poisson probability distributions of all bins:

$$P(\boldsymbol{n}|\boldsymbol{\lambda}) = \prod_{i=1}^{m_{\text{bins}}} \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}. \tag{4.1}$$

with $\boldsymbol{n} = (n_1, ..., n_{m_{\text{bins}}})$ and $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_{m_{\text{bins}}})$ being the observed and expected numbers of events in each histogram bin. The expected number of events is the sum of different processes such as one signal process s and various background processes b_j. In the $t\bar{t}H(b\bar{b})$ analysis the measured quantity is the *signal-strength modifier* μ defined as the ratio of the observed cross section for $t\bar{t}H$ production to the cross section predicted by the standard model:

$$\mu = \frac{\sigma_{t\bar{t}H}}{\sigma_{t\bar{t}H,\text{SM}}}. \tag{4.2}$$

The expected number of events in bin i is then given by

$$\lambda_i = \mu \cdot s_i + \sum_{j}^{\text{backgrounds}} b_{ij} \tag{4.3}$$

where the signal-strength modifier is the same parameter for all bins (it is correlated between all bins) and scales the expected numbers of signal events in all bins simultaneously. The signal-strength modifier is now the only parameter defining the physics model and the probability $P(\boldsymbol{n}|\mu)$ can be interpreted as a likelihood function $L(\mu)$ for the parameter μ, which can be maximized with respect to μ to find the value $\hat{\mu}$ that best describes the observed data.

In reality, the predictions for the numbers of events s_i and b_{ij} are subject to systematic uncertainties, such as uncertainties on the calculated standard model cross sections or uncertainties on the acceptance and efficiency of the detector. To account for this, an additional continuous *nuisance parameter* θ is introduced for each uncertainty and the parameters become functions of the nuisance parameters: $s_i(\boldsymbol{\theta})$ and $b_{ij}(\boldsymbol{\theta})$. The nuisance parameters are usually only correlated between specific bins or only affect specific processes. For each nuisance parameter a systematic probability density function (p.d.f.) $p(\widetilde{\theta}|\theta)$, describing the probability to observe a value $\widetilde{\theta}$ if the true value is θ, is multiplied to the likelihood function $L(\mu)$. For the construction of the $p(\widetilde{\theta}|\theta)$ it is assumed that a systematic uncertainty arises from a real or imaginary auxiliary measurement with posterior p.d.f. $\rho(\theta|\widetilde{\theta})$ where θ is the true value of the nuisance parameter and $\widetilde{\theta}$ is the measured value, as

discussed in Ref. [180]. With this, the "measured" uncertainties can be related directly to the widths of the systematic probability density functions.

The type of systematic p.d.f. that is used depends on how the uncertainty affects the predicted distributions. Rate changing uncertainties affect only the normalization of a process and should not be able to change the normalization to negative values. For their p.d.f a log-normal distribution is used, as described in Refs. [180, 181], defined as

$$p(\widetilde{\theta}|\theta) = \frac{1}{\theta\sqrt{2\pi}\ln(\kappa)} \cdot \exp\left(-\frac{(\ln(\theta/\widetilde{\theta}))^2}{2(\ln(\kappa))^2}\right) \tag{4.4}$$

with a mode and mean of approximately 1 for $\widetilde{\theta} = 1$ and a standard deviation of $\ln(\kappa)$. The parameter κ can then be identified with a relative uncertainty as $\kappa = 1 + \delta\theta/\theta$. The probabilities for θ to be larger or smaller by a factor of κ^n are identical and follow gaussian tail probabilities, i.e. the $\pm 1\sigma$ interval around the mode contains 68% of the area under the curve. For uncertainties that affect the shape of a distribution a template morphing approach is used, as described in Refs. [181, 182]. For each nuisance parameter, two additional template histograms are created for which the nuisance parameter is varied up or down by one standard deviation. The effects on the numbers of events in affected bins as a function of the nuisance parameter θ are calculated by vertical interpolation between the two systematic template histograms. For this nuisance parameter a standard normal distribution is used as probability density function. If the normalizations of those histograms differ from the nominal histogram, the normalization is factorized and treated as an additional rate uncertainty with log-normal p.d.f. for the same nuisance parameter θ.

The complete likelihood function of the model with m bins and l nuisance parameters is then given by

$$L(\boldsymbol{n}|\mu, \boldsymbol{\theta}) = \prod_{i=1}^{m_{\text{bins}}} P(n_i|\mu, \boldsymbol{\theta}) \cdot \prod_{k=1}^{l_{\text{nuisances}}} p(\widetilde{\theta_k}|\theta_k). \tag{4.5}$$

4.2 Significances

To calculate the significance of an observed signal with respect to the background-only hypothesis, a hypothesis test is performed. For this, a test statistic q_μ for a given value of μ is defined as:

$$q_\mu = -2\ln\frac{L(\boldsymbol{n}|\mu, \hat{\boldsymbol{\theta}}_\mu)}{L(\boldsymbol{n}|\hat{\mu}, \hat{\boldsymbol{\theta}})}, \text{ with } 0 \leq \hat{\mu}. \tag{4.6}$$

Here $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are those parameter values that maximize the likelihood globally and $\hat{\boldsymbol{\theta}}_\mu$ are the estimates for $\boldsymbol{\theta}$ for which the likelihood is maximized given a fixed value of μ. For practical reasons, instead of a maximization, a minimization of the negative log-likelihood $-\ln L(\boldsymbol{n}|\mu, \boldsymbol{\theta})$ is performed, yielding the same results. The test statistic is defined as a likelihood-ratio since those provide the best statistical power for tests of simple hypotheses, as stated by the Neyman-Pearson lemma discussed in Ref. [183]. Large values of this test statistic correspond to a large disagreement between the observed data and the hypothesis defined by the choice of μ.

As discussed above, the p-value is defined as the probability to find a test statistic value at least as large as that observed in data given the hypothesis μ is true:

$$p = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|\mu, \hat{\boldsymbol{\theta}}_\mu^{\text{obs}})\mathrm{d}q_\mu. \tag{4.7}$$

The function $f(q_\mu|\mu, \hat{\boldsymbol{\theta}}_\mu^{\text{obs}})$ is the p.d.f. for the test statistic q_μ given the hypothesis μ, where $\hat{\boldsymbol{\theta}}_\mu^{\text{obs}}$ indicates that the hypothesis contains the maximum likelihood estimates of the nuisance parameters obtained from a fit with a fixed value of μ. In the case at hand, the tested hypothesis is the background-only hypothesis with $\mu = 0$ and the needed p.d.f. is $f(q_0|0, \hat{\boldsymbol{\theta}}_0^{\text{obs}})$. The construction of the p.d.f. requires the simulation of a large amount of pseudodata by sampling the post-fit probability density functions of the nuisance parameters for a background-only fit, which is computationally expensive. Fortunately, as discussed in Ref. [184], the p.d.f. can be approximated analytically in the limit of large numbers of events and the p-value can then be calculated as

$$p = 1 - \Phi(\sqrt{q_0}) \tag{4.8}$$

where Φ is the cumulative distribution function of a standard gaussian distribution.

By convention the p-value is translated to a significance Z in such a way that the $1 - p$ quantile of the standard normal distribution equals Z standard deviations (or $Z\sigma$) above the mean. The significance is then calculated from the p-value as

$$Z = \Phi^{-1}(1 - p). \tag{4.9}$$

where Φ^{-1} is the inverse of the cumulative distribution function of a standard gaussian distribution. In this scheme, a significance of 3σ corresponds to a p-value of about 1.35×10^{-3} and is by convention the threshold to claim evidence for a process, while an observation or discovery requires a higher threshold of 5σ or a p-value of 2.87×10^{-7}.

4.3 Parameter estimation

To obtain estimates for the parameters μ and $\boldsymbol{\theta}$ the predicted distributions are fitted to the data by maximizing the likelihood function $L(\boldsymbol{n}|\mu, \boldsymbol{\theta})$ with respect to all parameters μ and $\boldsymbol{\theta}$ at once.

To find the uncertainty on the estimated value of a parameter, the negative log-likelihood is scanned as a function of the parameter while all other parameters are profiled, meaning that for each scanned parameter value, the negative log-likelihood is minimized with respect to all other parameters, as discussed for example in Ref. [185]. This can be interpreted as a test of the hypothesis of the scanned parameter value against the hypothesis of the best-fit parameter hypothesis and would necessitate the calculation of the test statistic defined in Eq. 4.6 for each scanned parameter value. The test statistic as a function of the scanned parameter will usually have the approximate form of a parabola with a minimum and a value of zero at the best-fit value of the parameter. According to Wilk's theorem, introduced in Ref. [186], in the case at hand and in the limit of a large number of events, the probability distribution of the test statistic can be approximated by a χ^2 distribution with one degree of freedom, as discussed for example in Ref. [7]. The $S\sigma$ confidence intervals around the minimum of the test statistic are then given by the parameter values that increase the test statistic by S^2. Therefore, the $S\sigma$ confidence intervals for the profile likelihood estimate are given as those values closest to the minimum for which the negative log-likelihood takes on the values $L + S^2/2$. In practice only the calculation of the uncertainty on the signal-strength parameter is performed this way while the post-fit uncertainties on the nuisance parameters are estimated from the Hessian matrix at the global minimum. They are often constrained to values smaller than their pre-fit values since additional information about them is gained.

Since the negative log-likelihood is also minimized with respect to the nuisance parameters, they usually take on different values after the fit compared to their values before the fit.

The differences between the pre-fit and post-fit values normalized to the pre-fit uncertainty are called *pulls*. As will be discussed in Ch. 9, the $t\bar{t}H(b\bar{b})$ analysis not only contains bins enriched in signal events but also purposefully constructed bins enriched in various background processes. These serve as control regions with which nuisance parameters are measured and constrained. If these regions are sufficiently large compared to the signal regions, the simultaneous fit of all parameters leads to corrections of the background prediction, as well as a reduction of its uncertainties in the signal enriched bins.

4.4 Expected results

To avoid biases on the final results, the $t\bar{t}H(b\bar{b})$ analysis is designed in a "blinded" fashion, as are most analyses. That means that all design choices and optimization procedures are performed using simulated data only. To gauge the sensitivity that can be expected of the analysis, results are also calculated using simulated pseudodata. For this, one could in principle simulate many sets of pseudodata in which the signal and background predictions are smeared around their pre-fit uncertainties. However, as explained in Ref. [184], if the statistical analysis is performed with a large enough number of events, only one such set of pseudodata, that with the median values of all smeared nuisance parameters, needs to be generated for each signal hypothesis. This median set of pseudodata is called "Asimov" pseudodata and is introduced in Ref. [184]. Since the central values of the uncertainties correspond to their best-estimated values and the smearing of the nuisance parameters is performed based on the estimated uncertainties of the nuisance parameters, the Asimov pseudodata are the data with the central values of the parameters, i.e. the nominal predictions for the background and the signal hypothesis.

The expected best-fit estimate for the signal-strength modifier and its uncertainty are calculated as described above, only with the Asimov data instead of recorded data. The expected significance and expected upper limits [1] are calculated as the median of the distribution of results for all sets of pseudodata, which corresponds to the result with the Asimov data. Uncertainties on the expected significance and upper limits are calculated as the results at $S\sigma$ quantiles around the median, which can be calculated from the uncertainties on the smeared nuisance parameters, as described in Ref. [184].

The expected significances quoted in Ch. 9 are calculated with post-fit background-only pseudodata. This set of pseudodata is obtained by estimating the post-fit values and uncertainties of the nuisance parameters through a background-only fit to recorded data, and then constructing Asimov pseudodata using the post-fit values of the nuisance parameters.

[1] Upper limits are not calculated for the results in this thesis. However, several upper limits calculated in other analyses are quoted. They are calculated using the CL_S prescription as discussed in Ref. [180].

5 Multivariate analysis methods

To measure $t\bar{t}H(b\bar{b})$ with large precision it is necessary to select samples of events enriched in $t\bar{t}H(b\bar{b})$ events from data. As will be explained in Ch. 7, the first step for this is to select events based on the number of leptons, jets and other simple observables. However, the phase space thus defined also contains a large number of non-$t\bar{t}H(b\bar{b})$ background events. To further enrich the selected sample, events that likely originate from the signal process are separated from those resembling background processes using more sophisticated observables. In some analyses it is for example possible to achieve this by reconstructing the invariant mass of a searched-for particle and subsequent selection of only events around the peak of that mass. Unfortunately, since $t\bar{t}H(b\bar{b})$ events contain a large number of particles in the final state and the mass reconstructed from background events mimics that of the signal process, such a mass observable is not feasible for the separation between signal and background events. Instead, several observables, which themselves do not separate between signal and background that well, are combined into better suited observables using multivariate analysis (MVA) methods. The employed MVA methods achieve this by simultaneously evaluating all included observables and exploiting correlations between them in a manner not easily achievable by hand. The constructed observables then allow the classification of the events as belonging to signal or background processes. MVA methods are also used for the classification of jets according to the flavor of the parton whence they originate, as will be discussed in Sec. 6.6.

The term multivariate analysis methods encompasses a vast number of different methods to analyze and explore data. Commonly used methods for classification tasks include boosted decision trees, support vector machines, linear discriminant analysis, artificial neural networks and more. An introduction the MVA methods and a discussion of the important methods is given in for example Ref. [187].

As will be discussed in Sec. 7.5, ANNs are used for the $t\bar{t}H(b\bar{b})$ analysis in the single-lepton channel. In Sec. 5.1, the general concept of constructing MVAs for classification tasks is discussed with a focus on the application for the $t\bar{t}H(b\bar{b})$ analysis. An introduction to artificial neural networks is given in Sec. 5.2. The architecture and training procedure of the ANNs used in the $t\bar{t}H(b\bar{b})$ analysis will be further discussed in the same Sec. 7.5. Although they are not directly used for the results in this thesis, BDTs are a commonly used alternative, had been considered during the design of the analysis strategy and are mentioned at several points throughout this thesis. For this reason a brief introduction to the main idea behind them will be given in Sec. 5.3.

5.1 General Concept

Both boosted decision trees and artificial neural networks consist of a model (a complicated function) with which an often unknown target function is approximated from input data and an algorithmic prescription for the choice of the parameters of the model.

The target function can in principle take many forms. For classification tasks it is a function that maps a vector $\boldsymbol{x} = (x_1, ..., x_n)$ of n *input features* to an output vector $\boldsymbol{y} = (y_1, ..., y_m)$ of m *classes*, where the value of a component y_i can be interpreted as how likely it is that the input vector belongs to class i. In the $t\bar{t}H(b\bar{b})$ analysis the input features are observables of the events (or properties of jets) and the m output classes are signal and background processes (or the flavors of jets). Ideally, the model would classify an event with total certainty and only one component of $\boldsymbol{y}$ would take on a value of one while the other components would be zero. In practice, all components take on finite values. In such cases the classification is performed by assigning the event to that class for which the component takes on the highest value.

The parameters of the model are chosen using a supervised learning algorithm, a procedure called *training*. For this, the parameters are first set to initial values. Then the model is evaluated on a sample of events for which the true class $\hat{y}_i$ is known. This sample is called the *training sample* and is selected from simulated data. Next, a *loss function* (sometimes also called cost function), that quantifies the deviation between the predicted classes and the true classes over the whole sample, is defined. Such a loss function is for example given by the mean squared error calculated from the distances between the predicted and the true classes:

$$L = \frac{1}{n_{\text{events}}} \frac{1}{m_{\text{classes}}} \sum_{k}^{\text{events}} \sum_{j}^{\text{classes}} (y_{kj} - \hat{y}_{kj})^2. \tag{5.1}$$

The parameters of the model are then updated to minimize the loss function. After this update the procedure is repeated with a new *batch* of events. This is repeated until the loss function converges to a stable value or a maximum number of iterations is reached. After the training is complete, the performance of the model is evaluated on an independent *validation* sample.

A common issue encountered when training classification models is that the model can learn the statistical fluctuations occurring in the training sample. This is called overfitting (or *overtraining* or generalization error) and results in a better performance on the training sample than on other samples. This mostly occurs if the size of the training sample is small and the number of parameters is large, similar to how overfitting can occur when fitting a polynomial of high order to a small set of data points. If, due to statistical fluctuations, the multidimensional distribution of training data is now different from that of an independent sample, events in the second sample will be assigned wrong classes more often. Several techniques can be applied to avoid overfitting. In addition to making sure that the training sample is of sufficient size and the model not too large, overtraining can also be avoided by training in batches, adding regularization parameters to the loss function or not training the whole model at once. Such measures can on the other hand degrade the performance of the classifier more than necessary. The choice of the model and training procedure is therefore an important and difficult task. A related effect, called *domain adaption error*, occurs if the training data is systematically different from that to which the model is applied. This can be the case if the simulated samples used for the training do not reflect the real physics encountered in recorded data. This illustrates the importance of using well modeled simulated data for the training.

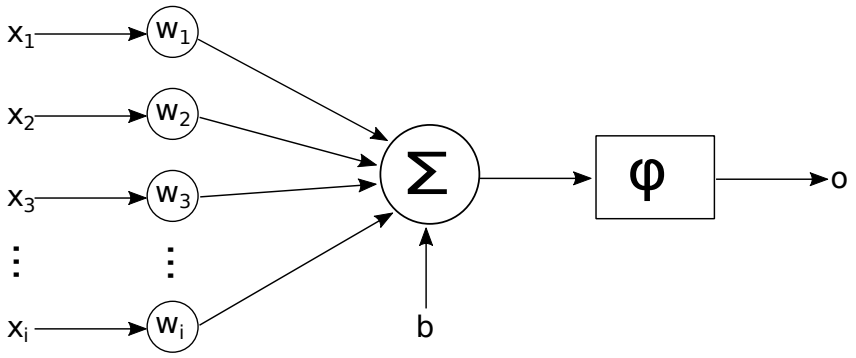


Figure 5.1: Illustration of an artificial neuron. The input features x_i are multiplied by weights w_i and summed. A bias weight b is added and the sum is passed through an activation function φ to produce the output o.

5.2 Artificial neural networks

Artificial neural networks are, as their name suggests, networks of artificial neurons modeled after biological neurons and brains. There are various types of ANNs depending on the functions of the neurons, the network structure and how information flows through the network. The networks used in the $t\bar{t}H(b\bar{b})$ analysis are feed-forward networks called multilayer perceptrons and are among the simplest network realizations.

The basic building block of the networks is a neuron. In a biological neuron, input signals arriving through multiple dendrites are summed and if that sum reaches a certain threshold an output signal is generated. This behavior is modeled in an artificial neuron, illustrated in Fig. 5.1. The components of the input feature vector $\boldsymbol{x}$ are first multiplied by weights to account for differences in the importance of the features and then summed. The threshold behaviour is modeled by adding a bias weight b and applying an activation function φ. The output value of the neuron is therefore given by

$$o = \varphi\left(b + \sum_{i}^{n_{\text{features}}} x_i w_i\right). \tag{5.2}$$

The weights w_i are free parameters of the model that are chosen by training of the model.

To create more complex models multiple layers of neurons are connected to create a network as illustrated in Fig. 5.2. The input features make up the input layer and are passed to all neurons of the first *hidden layer*. Their outputs are then connected to the following layer, and so on. The number of hidden layers and the number of neurons per layer are adjusted to suit the task at hand. Similarly, the activation function for the neurons has to be chosen. Commonly used ones for neurons in the hidden layer include the exponential linear unit (ELU), described in Ref. [188] and defined as

$$\varphi_{\text{ELU}}(x) = \begin{cases} x, \text{for } x \geq 0 \\ \alpha(e^x - 1), \text{for } x < 0, \end{cases} \tag{5.3}$$

with $\alpha > 0$ being an adjustable parameter, the rectified linear unit (ReLU), discussed in Ref [189] defined as

$$\varphi_{\text{ReLU}}(x) = \begin{cases} x, \text{for } x \geq 0 \\ 0, \text{for } x < 0, \end{cases} \tag{5.4}$$

and a function called leaky ReLU, described in Ref. [190] and defined as

$$\varphi_{\text{leakyReLU}}(x) = \begin{cases} x, \text{for } x \geq 0 \\ \beta x, \text{for } x < 0, \end{cases} \tag{5.5}$$

with an adjustable parameter β. As can be seen they differ in how they treat inputs of $x < 0$. Another common activation function is the sigmoid function defined as

$$\varphi_{\text{sigmoid}}(x) = \frac{1}{1 + e^{-x}}. \tag{5.6}$$

These activation functions are depicted in Fig. 5.3. For the relevant case of classification into multiple classes, each class is represented by one output neuron. The activation function used for these output neurons is the softmax function, discussed in Ref. [189]. For the output neuron o_i it is defined as

$$o_i = \frac{e^{z_i}}{\sum_j^{\text{outputs}} e^{z_j}} \tag{5.7}$$

where the sum is performed over all output neurons to ensure that the sum of all values predicted by the output neurons equals one.

For the training of the ANN, i.e. the choosing of the weights, the distributions of the input features are first transformed to have a mean value of zero and a standard deviation of one[1]. This is done to avoid large differences in the numerical values entering the neurons. Next the initial weights are randomly set to small values such as suggested in Refs. [191, 192]. The loss function to be minimized is the categorical cross entropy, which is discussed in Ref. [187] and defined as

$$L(\hat{\boldsymbol{o}}, \boldsymbol{o}) = -\frac{1}{N_{\text{classes}}} \sum_i^{N_{\text{classes}}} [\hat{o}_i \log(o_i) + (1 - \hat{o}_i) \log(1 - o_j)] \tag{5.8}$$

where $\hat{o}_i$ and o_i denote the true and predicted values for the class and the sum is performed over all classes. To avoid overtraining, an additional regularization term is added to the loss function. It is called L2-regularization, described in Ref. [193] and defined as

$$L_2 = \lambda \sum_i w_i^2 \tag{5.9}$$

where the sum is performed over all weights in the model and the parameter λ is an adjustable hyperparameter. This regularization term serves to suppress large weights of single neurons.

The training is performed iteratively with batches of events from the training sample, the batch size itself being an adjustable hyperparameter. After each batch the loss function is calculated and the weights of the model are updated using a gradient descent method:

$$w_{ij}^k \leftarrow w_{ij}^k - \kappa \frac{\partial L}{\partial w_{ij}^k} \tag{5.10}$$

where w_{ij}^k denotes the weight for the input i in neuron j of layer k and κ is a hyperparameter called learning rate. The derivative of the loss function with respect to a weight can be written as

$$\frac{\partial L}{\partial w_{ij}^k} = \frac{\partial L}{\partial o^{\text{N}}} \frac{\partial o^{\text{N}}}{\partial o^{\text{N}-1}} \ldots \frac{\partial o^{\text{k}}}{\partial w_{ij}^k} \tag{5.11}$$

[1] The same transformation is of course used during the application of the fully-trained ANN on unknown events.

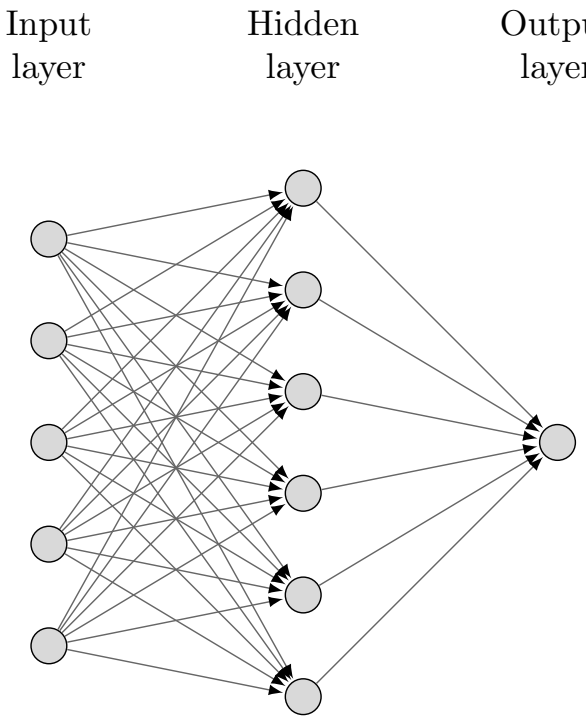


Figure 5.2: Illustration of an artificial neural network with one hidden layer. The input features in the input layer are all connected to the neurons of the hidden layer, which are all connected to the output layer. Taken from Ref. [121]

To increase the training speed, the derivatives are calculated iteratively using the back-propagation-of-errors algorithm discussed in Ref. [194]. To further reduce overtraining the *dropout* method, described in Ref. [195], is employed. With this method each neuron is deactivated for the current iteration with a probability that is an adjustable hyperparameter.

After each full pass over the training sample, called an *epoch*, the order of events in the training sample is randomized and the ANN is evaluated on an independent *test sample*. The training is performed until a maximum number of epochs is reached, the loss function is not reduced for several epochs or the loss function evaluated on the training and test samples diverges by more than a set threshold.

5.3 Boosted decision trees

Boosted decision trees (BDTs) are discussed in Refs. [187, 196] and their model consists of an ensemble of decision trees, where each decision tree is a sequence of binary decisions. Such a decision tree is illustrated in Fig. 5.4. BDTs are constructed by training an ensemble of single decision trees in such a way that each additional tree corrects the prediction made by the whole ensemble. The number of trees in the ensemble, the number of decisions per tree as well as several parameters controlling the training procedure are so-called hyperparameters and can be adjusted to optimize the performance and overtraining. For the analysis of data recorded in the year 2016, decision trees are trained to classify events either as $t\bar{t}H$ events or as $t\bar{t}$ events using the TMVA toolkit described in Ref [196, 197]. Since during the design of the analysis it was observed that the trained BDTs perform worse than ANNs, the analysis presented in this thesis uses only ANNs as discussed in Sec. 7.5. Therefore the BDTs are not further discussed here and further information about their

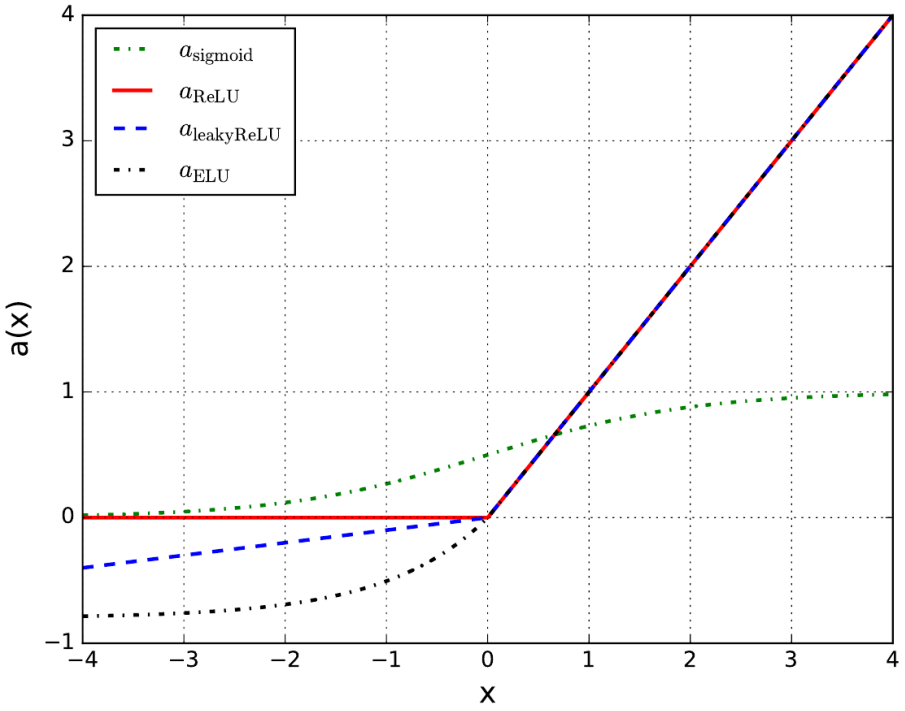


Figure 5.3: Different activation functions. Taken from Ref. [88]

general implementation in the context of a $t\bar{t}H(b\bar{b})$ analysis and optimization procedure is described in Ref. [134].

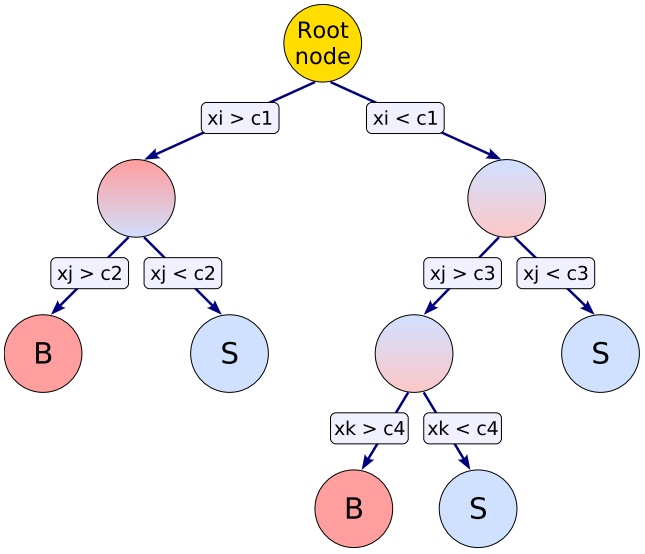


Figure 5.4: Illustration of a decision tree. Events are classified as either signal (S) or background (B) after a sequence of binary decisions based on four features $c_1, ..., c_4$. Taken from Ref. [196].

6 Analysis Objects

Most of the stable particles created in the proton-proton collisions within the CMS detector interact with the detector material leading to signals in the readout electronics. Analyses of proton-proton collisions usually require knowledge of the particles created in the scattering event and therefore require the reconstruction of the particles from the detected signals. In this chapter, the reconstruction and identification of the jets, leptons and other basic objects used in the analysis of $t\bar{t}H(b\bar{b})$ production is discussed.

Depending on their nature, different kinds of particles interact with different parts of the detector, as illustrated in Fig. 6.1. Neutral hadrons, for example, will not significantly interact with the tracker or electromagnetic calorimeters, but will deposit most of their energy into the hadronic calorimeters. Charged particles interact with the sensitive material in the tracker, leaving *hits* along their trajectories. Charged hadrons and electrons will then be stopped in the electromagnetic calorimeters, depositing their energy, whereas muons with sufficient energy will only deposit a small amount of energy into the calorimeters and will traverse through the whole detector to the muon detectors.
The reconstruction of particles is based on the measurements in the different subdetectors and begins with reconstructions of the *tracks* of charged particles, as will be discussed in Sec. 6.1 and the reconstruction of energy clusters in the calorimeters will be discussed in Sec. 6.2. The tracks and clusters are combined using the particle flow algorithm, which will be discussed in Sec. 6.3, to reconstruct candidates for neutral or charged hadrons, electrons, muons, and photons from the collections of hits and energy deposits.

The reconstruction and identification of photons and leptons will be described in Sec. 6.4.

As was discussed in Ch. 2, quarks and gluons created in the collision events undergo hadronization and form cone-shaped sprays of stable particles called *jets*. The reconstruction, identification and calibration of the jets used in this thesis will be discussed in Sec. 6.5. Jets originating from bottom quarks are of special interest for the analysis of $t\bar{t}H(b\bar{b})$ production. They can be identified using multivariate methods which will be discussed in Sec. 6.6.

Neutrinos or certain particles hypothesized by theories beyond the standard model can escape the detector without interacting with the sensitive material. This will lead to an imbalance in the energy deposited transverse to the beam axis which is called *missing transverse momentum*, which is reconstructed as described in Sec. 6.7.

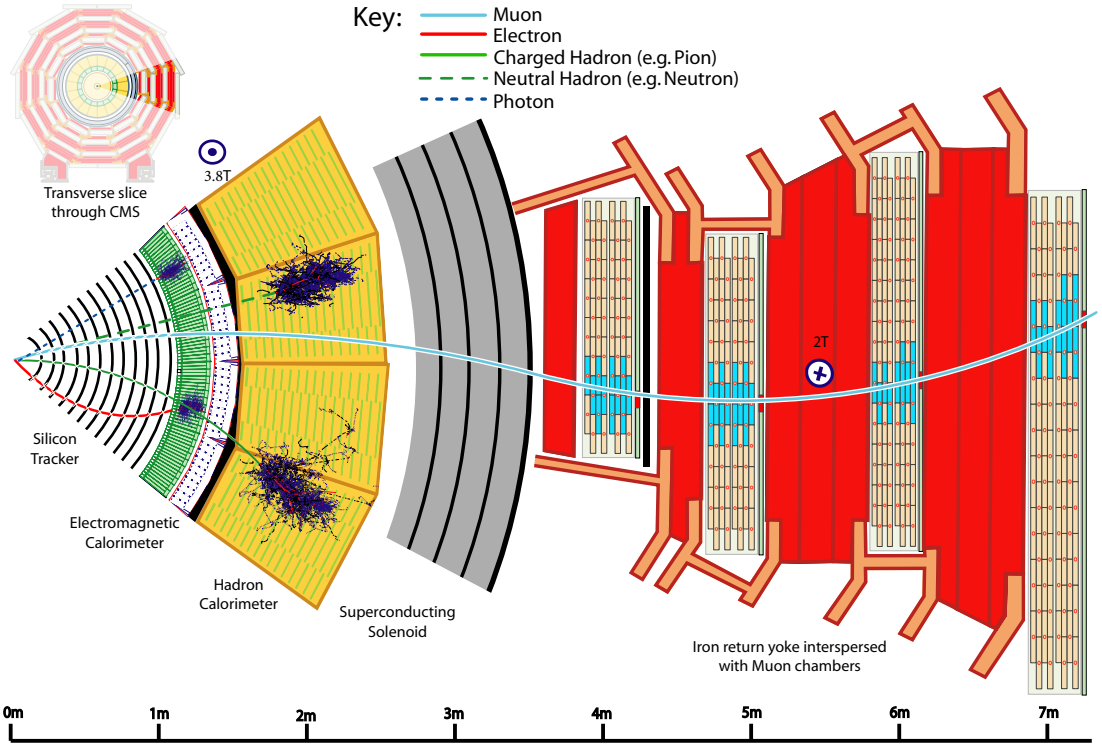


Figure 6.1: Slice of the CMS detector illustrating the behavior of different particles and the response of the various subdetectors. The trajectories of electrically charged particles are curved in the magnetic field of the CMS detector. Electrically charged particles interact with the silicon tracker, the muon detectors and the electromagnetic calorimeters. While charged hadrons, electrons and photons are stopped by the electromagnetic calorimeters, muons only deposit small amounts of energy and are able to traverse to the muon systems, aiding in their identification. Hadrons deposit energy in the hadronic calorimeters. Taken from Ref. [198].

To test the predictions made by theories such as the standard model, the data recorded with the CMS experiment is compared to simulated data. The generation process of these samples includes a simulation of the response of the detector up to the signals induced in the readout electronics, after which the simulated events can be reconstructed in the same way as is done for real data events. These simulations, however, are not perfect, causing not only differences in the event contents themselves, but also in the efficiencies for the reconstruction, identification and energy measurements of the particles and jets. The corrections applied to simulated samples to improve their description of the data will be discussed in Sec. 6.9.

6.1 Tracks and Vertices

When electrically charged particles pass through the CMS detector they can produce signals (*hits*) in the layers of the silicon tracker from which their *tracks* can be reconstructed. Highly energetic muons leaving the detector can also produce hits in the muon detectors. Due to the magnetic field, the tracks are curved with the curvature depending on the momentum, the mass and the charge of the particles. A reconstruction of the tracks therefore allows a measurement of the momenta and charges of the particles. Points where one or more

tracks originate (*vertices*) are of interest not only to identify tracks originating from the hard scattering interaction (*primary vertex*), but also for the identification of particles such as τ-leptons or B-hadrons, as will be discussed in Sec. 6.6. At current experimental conditions each bunch crossing can lead to thousands of hits and tracks. This makes the reconstruction of tracks with a high efficiency and low rate of fake tracks, which are tracks that are reconstructed from unrelated hits, a rather difficult and computationally intensive problem.

6.1.1 Hit reconstruction in the silicon tracker

The first step in reconstructing particle tracks is the reconstruction of hits in the silicon pixel tracker and the silicon strip tracker, which is described in detail Ref. [199]. Due to the fine segmentation of the tracker detectors, traversing particles will in general deposit charge clusters in neighboring pixels or strips. This effect is called charge sharing and can be exploited to improve the position resolution beyond the pixel or strip widths. The clusters themselves are usually sufficiently separated to identify them as belonging to one traversing particle. The reconstruction also takes into account the Lorentz drift of the charge caused by the magnetic field.

Hit reconstruction in the pixel detector

The positions of hits in the pixel detector are reconstructed from clusters of pixels with signals above a, for each pixel specific, noise threshold. A comparably fast algorithm reconstructs hits that are used for the seeding of the track finding algorithm and estimates their position from the geometric center of the cluster, the asymmetry of the charge distribution and the Lorentz drift of the created charges in the magnetic field. For the final track reconstruction, the hit position is estimated by fitting templates of cluster charge distributions to the observed charge distribution. The templates are derived using simulated tracks with various angles and also take into account changes to the charge collection and Lorentz drift due to radiation damage accumulated by the sensor material during its lifetime. The uncertainty on the hit position is estimated from the distribution of the difference between the reconstructed and actual hit position of simulated tracks.

Hit reconstruction in the strip detector

Similar to the treatment of the pixel detector, strip channels need to have read-out charges above adjustable thresholds to reduce noise. The clusters from which the hits are reconstructed are seeded by channels with a high signal-to-noise-ratio. Surrounding channels are added to these clusters if they pass a less strict signal-to-noise threshold. Hits in the strip detector are reconstructed from clusters of strips with high signal-to-noise ratios. The position of the cluster is calculated as the charge-weighted average of the strip channel positions and is corrected to account for Lorentz drift of the charge as well as a charge collection efficiency at the sensor backplanes. The uncertainty on the cluster position is estimated from the expected width of the cluster based on the incident angle.

6.1.2 Hit reconstruction in the muon detectors

The CMS detector contains three different types of gaseous detectors specialized in the detection of muons: drift tube chambers (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). The method of hit reconstruction differs for these three types of detectors and is explained in detail in Refs. [163, 164]. For the DTs the hit position is calculated from the position of the wires that registered a signal and the drift time of the electrons in the gas. The CSCs contain wires as well as finely segmented cathode strips.

The azimutal and radial position of the muon hit is calculated from the positions of the wires and strips that collected the induced charge. In RPCs, the produced electrons can be detected by multiple cathode strips and the hit position is calculated as the arithmetic mean of all affected strips. While the RPCs are single-layer chambers that can only register one hit per traversing muon, the CSCs and DTs are multi-layer detectors where muons can leave multiple hits. These are later used to reconstruct segments of muon tracks. For each detector element, the spatial resolution of the hits is calculated from the distributions of the residuals between registered hits and fitted track segments that were reconstructed with other detector elements. The spatial resolution is found to be in the range of 45 μm to 1.38 cm depending on the type and position of the detector.

6.1.3 Track finding

Once the hits are reconstructed, track finding algorithms are used to reconstruct the trajectories of charged hadrons, electrons and muons. Since electrons have a high probability to emit bremsstrahlung photons in the tracker and muons allow the inclusion of hits in the muon systems, their trajectories are reconstructed with different methods that will be briefly introduced in the following paragraphs.

Iterative Tracking

The track finding algorithm is based on an iterative application of a combinatorial Kalman filter as described in Refs [200–203]. In each iteration, tracks are reconstructed in four stages which are detailed in Refs. [198, 199].
In the first stage, track seeds are generated from small numbers of hits. The seeding requirements for each iteration are chosen to target specific kinds of particle trajectories. The first iteration, for example, targets prompt particles with high transverse momenta and requires hits in three consecutive layers of the inner pixel detector that are compatible with a high momentum track passing very near to the interaction point, while the sixth iteration targets very displaced tracks and requires seeds in two or three strip layers.
In the second stage, the initial track candidates are layer-wise extrapolated along plausible trajectories to search for additional compatible hits, taking into account Coulomb scattering, energy losses and the uncertainty on the hit position. For each compatible additional hit, a new track candidate is created. After this, the trajectories of the track candidates are updated and the search is extended to the next layer. In the third stage, the track candidates are refitted multiple times, taking into account hit uncertainties, multiple scattering, energy losses and the inhomogeneities of the magnetic field, while also removing possible hits that were wrongly assigned to the track candidate. This leads to a smoothing of the trajectory and the best possible estimates of the track parameters and uncertainties.
In the fourth and final stage, track candidates are discarded if they do not pass certain criteria such as the goodness of the fit or the number of associated hits. Any ambiguities in the assignment of hits to tracks are resolved by only keeping the best fitting track candidates.

To improve the track finding efficiency and to reduce the misreconstruction rate, especially for trajectories with low momentum, the combinatorial Kalman filter approach is performed multiple times. After each iteration, hits assigned to found tracks are removed from the collection of considered hits and the seeding requirements and the final track quality criteria are adjusted. The improvement to the track finding efficiency and reduction of the misreconstruction rate, due to the iterative approach is shown in Fig. 6.2.

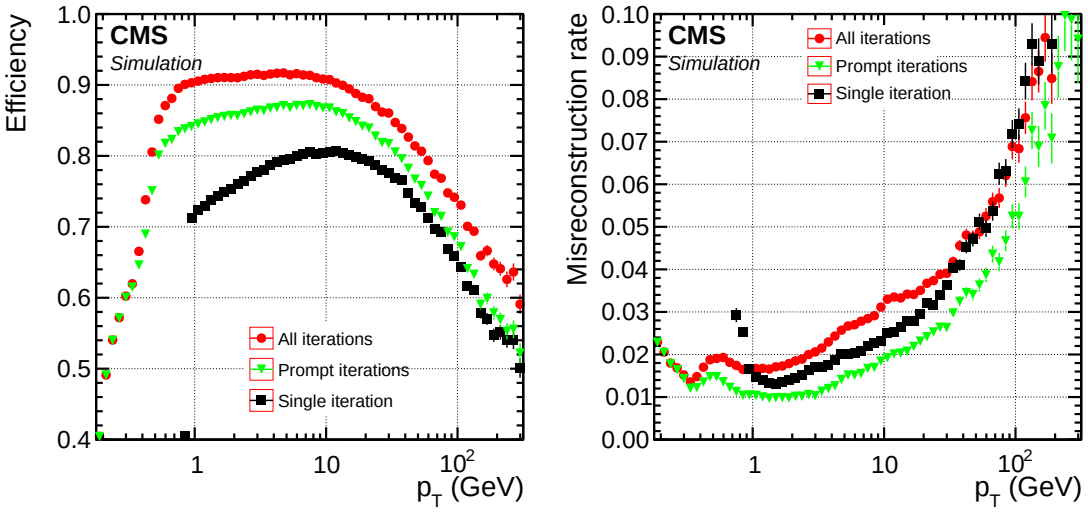


Figure 6.2: Efficiency (left) and misreconstruction rate (right) of the non-iterative approach (black squares), the iterative approach with only those iterations that include pixel hits in the seeds (green triangles) and the iterative approach with all iterations considered (red circles), as a function of the track p_T. The values were calculated for simulated charged hadrons in multijet events without pileup interactions. Only tracks with $|\eta| < 2.5$ are considered. Taken from Ref. [198].

Electron tracks

Electrons lose a large amount of energy through bremsstrahlung when traversing through the tracker material. Since this energy loss is non-gaussian, the Kalman filter approach is ill-suited for electron tracks which are instead reconstructed with a refined approach described in Refs. [198, 199]. With this approach, electron track seeds are generated with two methods. The first method targets isolated electrons with large bremsstrahlung losses and searches for seeds compatible with clusters of ECAL energy deposits of similar η but different ϕ coordinates under the assumption that these are caused by the bremsstrahlung photons and the electron itself. The second method is targeted at electrons where the emitted bremsstrahlung cannot be as easily identified, such as electrons inside of jets or very low momentum electrons. Here electron track seeds are generated from previously reconstructed tracks that either pass by a compatible ECAL energy deposit or have a bad track fit quality or missing hits which could be caused by bremsstrahlung emission. Starting from these seeds, electron candidate tracks are then found using the Kalman filter approach with relaxed conditions on the track quality. Finally, a track fit for the electron tracks is performed using the Gaussian Sum Filter approach that takes into account the expected energy loss of the electrons as described in Ref. [204].

Muon tracks

Muons generate hits not only in the silicon tracker but also in the muon detector systems. To combine both systems, muons are reconstructed in three ways, each using a Kalman filter approach for the track finding. *Standalone-muon tracks* are muon tracks that are reconstructed from only the hits in the muon systems. For *global-muon tracks* each standalone-muon track is matched to a track in the silicon tracker, and for *tracker-muon tracks* previously found tracks are extrapolated and matched to hits in the muon systems. Most muons are found to be reconstructed as global-muon tracks as well as tracker-muon tracks, and the combination of the tracker and the muon systems ensures a high purity

of reconstructed muon candidates as well as a good momentum resolution as discussed in Refs. [164, 198].

6.1.4 Primary-vertex reconstruction

Given the previously fitted tracks, the number, positions and associated uncertainties of the proton-proton interaction vertices in each event are reconstructed as discussed in Ref. [199]. For this, first all tracks compatible with being produced in the interaction region are selected. Then these tracks are clustered into spatially distinct interaction vertices using a deterministic annealing procedure, described in Ref. [205]. Finally the parameters of each clustered vertex are fitted using an adaptive vertex fitter, which is discussed in Ref. [206]. In subsequent steps the knowledge of the fitted vertices is used. For this only vertices that pass certain quality criteria are considered:

- the vertex has to be derived from tracks (in case no tracks are found the nominal interaction point is taken as vertex) and the vertex fit has to have a $\chi^2/n_{\text{dof}} > 0$,
- the number of degrees of freedom in the vertex fit has to be larger than 4,
- the distance from the nominal interaction point in the z-direction has to be smaller than 24 cm,
- the distance from the nominal interaction point in the $x - y$-direction (perpendicular to the beam axis) has to be smaller than 2 cm.

Since the process of interest in this analysis involves highly energetic and heavy particles, the vertex with the highest energy, determined by the scalar sum of the transverse momenta of all particles that can be matched to this vertex using their tracks, is denoted as the primary vertex and it is assumed that the interaction of interest originates from this vertex. All other vertices in the event are regarded as corresponding to pileup interactions.

6.2 Calorimeter clusters

Charged particles such as photons, electrons and charged hadrons are stopped or deposit energy in the ECAL while charged and neutral hadrons deposit their energy in the HCAL. By measuring the positions and energies of the deposits it is possible to infer the energy, direction and type of the corresponding particle. Often the energy is not deposited in a single calorimeter cell but in several adjacent cells. To collect the energy of adjacent cells and to distinguish between groups caused by different particles, a clustering algorithm is used, as described in Ref. [198]. The clustering is performed separately for the different calorimeter subdetectors: ECAL barrel and endcaps, HCAL barrel and endcaps, and the preshower detectors. Depositions in the hadronic forward calorimeter are not clustered, instead each cell is treated as a single cluster.

The clustering starts by identifying cells that have energy deposits above a certain threshold and above adjacent cells. Next, *topological clusters* are grown from these *cluster seeds* by adding adjacent cells. The position and energy of the particle causing a cluster is determined by fitting a gaussian energy profile to the energy deposits. In case a topological cluster includes multiple cluster seeds it is assumed that each cluster seed is caused by a particle and that the energy of each cell in the topological cluster is the sum of the energies of all particles. To resolve this, a gaussian-mixture model is used for the fit to the energy deposits, where instead of only one gaussian, a sum of multiple gaussians is used. The accurate measurement of the calorimeter cluster energies is important since they are for example the only way to detect and measure neutral hadrons in the events. Therefore the absolute and relative energy responses of the calorimeter clusters needs to be well-calibrated.

This calibration is performed using particles coming from test beam experiments, cosmic showers, radioactive sources, as well as simulated and recorded collision data as is discussed in Refs. [198, 207].

6.3 Particle Flow

As was discussed above, different types of particles interact with different subdetector systems. Charged hadrons for example will leave a track in the tracker, as well as energy deposits in the ECAL and HCAL, while neutral hadrons mostly deposit energy in the HCAL with only a small amount in the ECAL. Combining the information from different subdetectors therefore allows for an improved identification of particles. In addition, the combination of track information and calorimeter information leads to a better measurement of the particle momenta and energies. The combination of the information of the subdetectors is performed with the *particle flow* (PF) algorithm, which is discussed in detail in Ref. [198].

The algorithm links tracks and calorimeter clusters based on separation in the (η, ϕ) plane and classifies all recorded signals as either muon, electron, charged hadron, neutral hadron or photon candidates. For this, tracks are linked to calorimeter clusters by extrapolating the track into the calorimeters. Clusters in one calorimeter system are linked to clusters in another calorimeter system if the cluster in the system with the finer granularity lies within the cluster of the coarser system. In case multiple links are possible, only those with the smallest distance are chosen. Once a particle candidate is identified, all linked objects are removed for the reconstruction of the remaining candidates. This is repeated until all tracks and calorimeter clusters are assigned to a particle candidate.

Muon candidates are identified first based on the previously reconstructed muon tracks and linked ECAL clusters. Isolated muons are identified mainly due to the existence of a global-muon track. For non-isolated muons additional criteria are required for the muon tracks. Since muons only lose a small amount of energy in the calorimeters the muon-candidate momentum is mainly measured by the tracks. Next electrons and isolated photons are identified. Photons are identified by requiring the presence of an ECAL cluster, the absence of a track and that less than 10% of the ECAL energy is present in associated HCAL clusters. Photons also can undergo conversion to e^+e^- pairs in the tracker material. These photons are identified by the presence of two compatible tracks that usually start in outward lying tracker layers. Electrons require the presence of a track, an ECAL cluster and again that less than 10% of the ECAL energy is present in associated HCAL clusters. In addition, the PF algorithm adds compatible ECAL clusters and conversion photons, caused by bremsstrahlung, to the electron candidates. Electrons furthermore have to pass additional criteria on the track quality and track-ECAL distance to be identified as such. The remaining tracks and clusters give rise to charged hadrons, neutral hadrons and non-isolated photons. Inside the tracker acceptance, ECAL clusters with no linked tracks are identified as photons while HCAL clusters with no linked tracks are identified as neutral hadrons. The now remaining HCAL clusters are linked to one or more tracks and might also be linked to several ECAL clusters. Each of the tracks gives rise to a charged hadron and depending on the ratios of the energies of the tracks and the calorimeter clusters, additional neutral hadrons and photons may be identified. Outside of the tracker acceptance ($|\eta| > 2.5$) neutral and charged hadrons cannot be distinguished.
There, ECAL clusters without linked HCAL clusters are identified as photons. The energies of all identified particle candidates are calculated by an optimized approach from a mixture of the track momenta and calorimeter-cluster energies.

6.4 Lepton reconstruction and identification

Instead of directly using all electrons and muons identified by the PF algorithm, the analyses described in this thesis only select those that fulfill additional criteria designed to remove those candidates that are falsely reconstructed or do not come from the primary interaction of interest (such as leptons arising in the decays of hadrons). These additional identification criteria differ slightly for data recorded in the years 2016 or 2017 and are described in the following.

6.4.1 Electrons

All electrons used in this analysis are required to be reconstructed by the PF algorithm as electron candidates. The energies of electrons tend to be not fully accounted for in the ECAL clusters due to leakage of the shower into ECAL cells not connected to the cluster or leakage into other detector components. To correct this, a regression of the electron energy is performed using a multivariate analysis method as described in Ref. [208]. For electrons in simulated data the measured energy is slightly biased compared to data and the resolution of the energy measurement is slightly better. This is accounted for by scaling and smearing the measured momentum by a small amount.

Then, electrons are selected for the analysis if they have a transverse momentum of $p_\mathrm{T} > 30\,\mathrm{GeV}$ and a pseudorapidity of $|\eta| < 2.1(2.4)$ for data recorded in 2016 and 2017 respectively.

Furthermore, electrons have to fulfill additional requirements which are described below and explained in further detail in Refs. [4, 5, 208]. Electrons are discarded if they are detected in the transition region between the endcap and the barrel calorimeters at $1.4442 < |\eta| < 1.5560$. The electrons of interest in this analysis are expected to be isolated with the only hadronic activity in their vicinity coming from the underlying event and pileup. To ensure this, an effective-area-corrected isolation variable that measures the activity around the electron is calculated as

$$\mathrm{Iso_e} = \sum_{\Delta R<0.3} p_\mathrm{T}^\mathrm{CH} + \max\left(0, \sum_{\Delta R<0.3} p_\mathrm{T}^\mathrm{NH} + \sum_{\Delta R<0.3} p_\mathrm{T}^\mathrm{PH} - \rho A\right). \tag{6.1}$$

Here, p_T denotes the transverse momenta of different particles identified by the PF algorithm. The superscripts denote the considered particles being neutral hadrons (NH), charged particles (CH) or photons (PH). The parameter A stands for an effective area depending on the electron η and ρ for the energy density of neutral hadronic activity caused by pileup in the event. The last term therefore accounts for neutral hadronic activity in the event that is not accounted for by the previous terms. The sums are performed over all particles in a cone around the electron with radius $\Delta R < 0.3$. Then a relative isolation I_rel is defined as the ratio of the effective-area-corrected isolation Iso_e and the transverse momentum of the electron. Electrons recorded in 2016 are required to have a relative isolation of $I_\mathrm{rel}^\mathrm{e} < 0.06$ and for electrons recorded in 2017 the relative isolation has to be $I_\mathrm{rel}^\mathrm{e} < 0.036$ if the electron was detected in the barrel and $I_\mathrm{rel}^\mathrm{e} < 0.094$ if it was detected in the endcap region of the tracker. In addition, electrons are required to fulfill cuts on variables describing the angular separation between their track and the associated ECAL cluster ($\Delta\eta$, $\Delta\phi$), the lateral extension of the electromagnetic shower detected by 5×5 ECAL crystals centered around the peak of the energy distribution ($\sigma_{\eta\eta}$), the difference between the track momentum and the ECAL cluster energy ($|\frac{1}{E} - \frac{1}{p}|$), and the ratio of energy deposited in the HCAL and the ECAL (H/E). In all cases the electron track is allowed to have at most one hit missing in the inner tracker. The values of all selection criteria differ for electrons reconstructed in

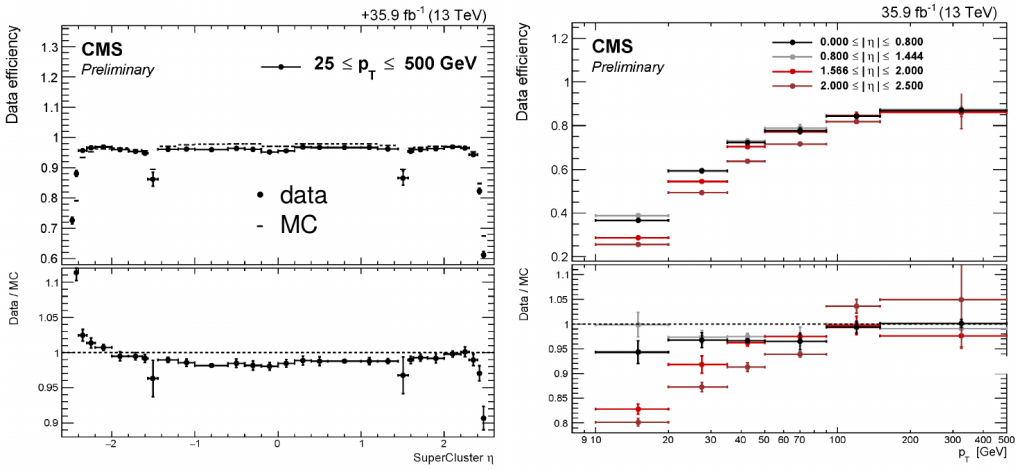


Figure 6.3: Efficiencies for (a) the reconstruction of electrons as a function of the η of the associated ECAL supercluster and (b) the identification of electrons as a function of the p_{T} and η of the electron. The top panels show the efficiencies in recorded data while the bottom panels show the ratio between the efficiencies in recorded and simulated data. The efficiencies are measured in data recorded in the year 2016. Taken from Ref. [209]

the barrel and endcap regions of the detector and are listed in Tab. 6.1. Different values of these selection criteria for the years 2016 and 2017 are motivated by changes to the detector, especially the upgrade of the silicon tracker, and the operating conditions. The measured efficiencies of the reconstruction and identification criteria of electrons are shown in Fig. 6.3 and Fig. 6.4 for data recorded in the years 2016 and 2017 respectively. They are measured using a tag-and-probe method as described in Ref. [209, 210]. As shown in the figures, the efficiencies depend on the p_{T} and η of the electrons and are different for recorded and simulated data. These sometimes large differences are corrected as described in Sec. 6.9.4.

Table 6.1: Criteria for the selection of electrons.

	2016		2017	
Variable	barrel ($\lvert\eta\rvert \leq 1.479$)	endcap ($\lvert\eta\rvert \geq 1.479$)	barrel ($\lvert\eta\rvert \leq 1.479$)	endcap ($\lvert\eta\rvert \geq 1.479$)
$p_{\mathrm{T}}>$ (GeV)	30	30	30	30
$\lvert\eta\rvert<$	2.1	2.1	2.4	2.4
$I^{\mathrm{e}}_{\mathrm{rel}}<$	0.06	0.06	0.036	0.094
$\Delta\eta<$	0.00308	0.00605	0.00353	0.00567
$\Delta\phi<$	0.0816	0.0394	0.0499	0.0165
$\sigma_{\eta\eta}<$	0.00998	0.0292	0.0104	0.0305
$\lvert\frac{1}{E}-\frac{1}{p}\rvert<$ (GeV^{-1})	0.0129	0.0129	0.0278	0.0158
$H/E<$	0.0414	0.0641	0.026	0.026

6.4.2 Muons

Possible misalignment of the detector and uncertainties on the knowledge about the magnetic field can lead to biases and differences in the momentum measurement of reconstructed

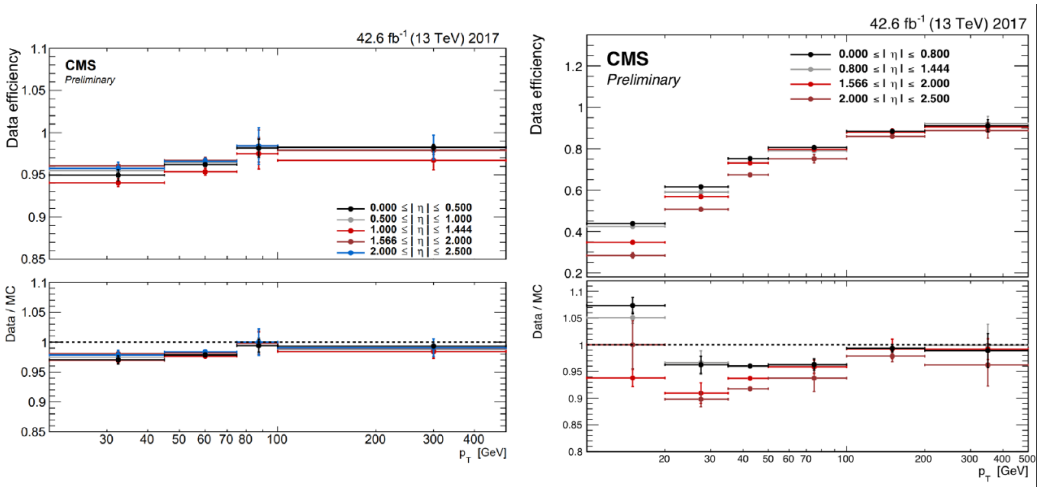


Figure 6.4: Efficiencies for (a) the reconstruction and (b) the identification of electrons recorded in the year 2017 as a function of the p_T and η of the electron. The top panels show the efficiencies in recorded data while the bottom panels show the ratio between the efficiencies in recorded and simulated data. Taken from Ref. [210]

muons in simulated and recorded data. These biases can be corrected using so-called *Rochester corrections*, discussed in Ref. [211], which effectively change the curvature of the muon tracks. For muons in simulated data the corrections also decrease the resolution of the momentum measurement. For the relevant muons in this analysis the corrections to the track curvature and therefore the p_T are smaller than 1% and the momentum resolution in simulated data is decreased by about 5%. For the muons in data recorded in the year 2016 such corrections were applied, whereas they had not yet been derived for muons in data recorded in the year 2017. The effect of these corrections on the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses in this thesis is however expected to be small.

Similar to electrons, additional requirements are posed for the selection of muons. All selected muons are required to be reconstructed by the PF algorithm as muon candidates and have a transverse momentum of $p_\mathrm{T} > 26(29)\,\mathrm{GeV}$ and a pseudorapidity of $|\eta| < 2.1(2.4)$ for data recorded in the years 2016 and 2017 respectively. Muons also have to pass a cut on a relative isolation variable $I^{\mu}_{\mathrm{rel}} = \mathrm{Iso}_{\mu}/p_\mathrm{T}$ where the isolation is defined as

$$\mathrm{Iso}_{\mu} = \sum_{\Delta R<0.4} p_\mathrm{T}^{\mathrm{CH}} + \max\left(0, \sum_{\Delta R<0.4} p_\mathrm{T}^{\mathrm{NH}} + \sum_{\Delta R<0.4} p_\mathrm{T}^{\mathrm{PH}} - \frac{1}{2}\sum_{\Delta R<0.4} p_\mathrm{T}^{\mathrm{PU}}\right). \tag{6.2}$$

Here, p_T denotes the transverse momenta of different particles identified by the PF algorithm and the superscripts denote the considered particles being neutral hadrons (NH), charged hadrons (CH), photons (PH) or charged hadrons identified as coming from pileup interactions (PU). The sums are performed over all particles in a cone around the muon with radius $\Delta R < 0.4$. Muons recorded in both years 2016 and 2017 are required to have relative isolation of $I^{\mu}_{\mathrm{rel}} < 0.15$. To further purify the sample of selected muons they have to fulfill the *tight muon ID* which is further explained in Ref. [164]. This ID requires that muons are reconstructed as both a tracker-muon and a global-muon, where the global-muon-track fit has a $\chi^2/n_{\mathrm{dof}} < 10$. Furthermore, the tracks have to have at least one hit in a pixel layer and at least 6 hits in any tracker layers. Finally, the impact parameter to the primary vertex has to be smaller than $d_{\mathrm{xy}} < 2\,\mathrm{mm}$ in the $x-y$ direction

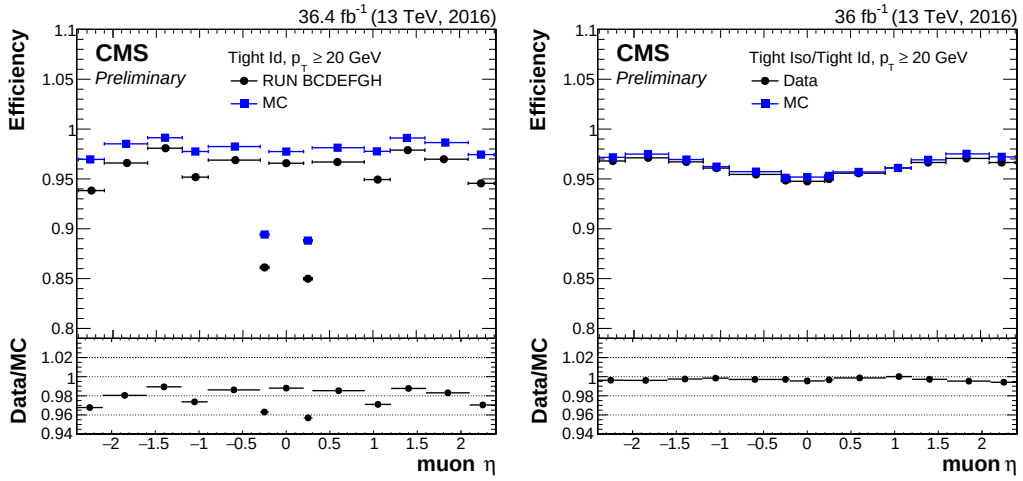


Figure 6.5: Efficiencies for (a) the identification of muons and (b) the isolation of the previously identified muons as a function of η of the muon. The efficiencies are shown for recorded and simulated data in the year 2016. The error bars indicate statistical uncertainties. Taken from Ref. [212]

and $d_z < 5\,\text{mm}$ in the z direction. Efficiencies of the isolation and identification criteria are measured using a tag-and-probe method (see Ref. [212, 213]) and shown for simulated and recorded data for the year 2016 in Fig. 6.5 and for the year 2017 in Fig. 6.6. The efficiencies depend on the η of the muons. Differences between recorded and observed data are visible, they are however smaller than for the efficiencies of the electrons shown in the previous section. The differences are corrected as described in Sec. 6.9.4.

6.5 Jets

Due to the properties of the strong interaction, quarks and gluons cannot be observed as free particles but only as color neutral bound states. For this reason, all quarks and gluons produced in collision events undergo a process called *parton showering* where a cascade of additional strongly interacting particles are radiated. Since only color neutral particles can be observed in nature all particles created in this parton shower then form neutral hadrons in a process called *hadronization*[1]. Most of the color neutral particles created this way are not stable and will further decay to stable particles. All resulting particles will move roughly in the same direction and the ensemble is called a *jet*. Due to momentum conservation the total energy, momentum and direction of these jets correspond to the energies, momenta and directions of the original quarks and gluons that were their cause, and with the detection of all these particles one could in principle reconstruct the original quark and gluon. This however is complicated by the fact that there are usually multiple jets in an event that might overlap, and that there usually is additional hadronic activity from pileup interactions. Therefore even if one detects all decay products, the assignment of these to individual jets is a difficult problem.
In experiments this problem is solved by clustering all decay products into jets using jet clustering algorithms. Usually experimental data is compared to theoretical predictions or simulated data. Since the constituent makeup of a jet is caused by a probabilistic process where the splitting of constituent partons in soft or collinear partons is rather

[1] A notable exception to this phenomenon is the top quark which almost instantly decays to a b-quark and a W-boson. These decay products however result in jets.

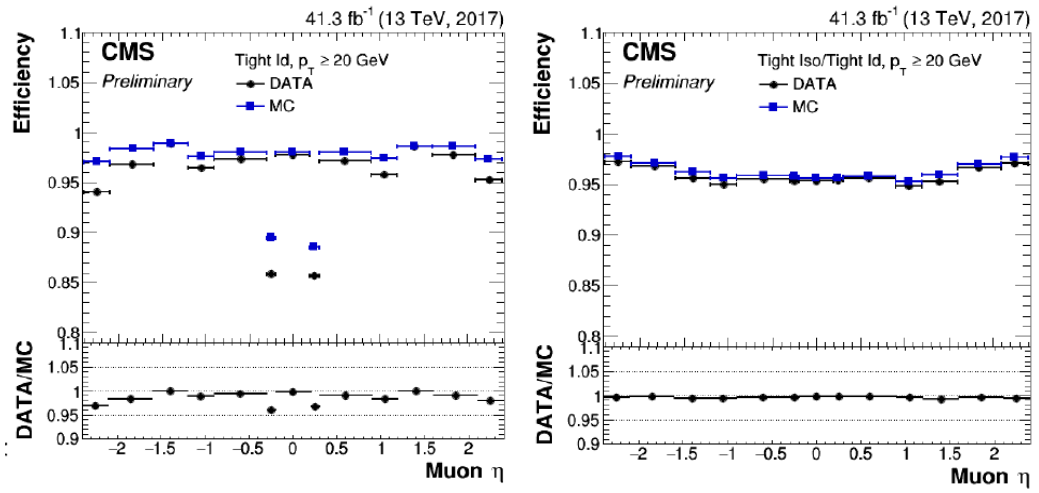


Figure 6.6: Efficiencies for (a) the identification of muons and (b) the isolation of the previously identified muons as a function of η of the muon. The efficiencies are shown for recorded and simulated data in the year 2017. The error bars indicate statistical uncertainties. Taken from Ref. [213]

ambiguous and hard to describe by theoretical calculations, comparisons between prediction and experiment are usually performed with jet observables. It is therefore of importance that the jet algorithms are insensitive to soft or collinear parton splittings, which is called infrared and collinear safety, respectively. There are many different jet algorithms that fulfill the requirement of infrared and collinear safety. Of special interest here are sequential recombination algorithms where, starting from a seed particle, additional particles are sequentially added to the jet until a cut-off requirement is fulfilled. Some of these algorithms have the advantageous property that the recombination steps often correspond to the splittings and decays inside the shower, which can aide in studies that analyze the jet substructure. The various sequential recombination algorithms differ mainly in the choice of the seed particle and in the order of the sequential recombinations, their computational complexity, as well as their suitability to different investigated particle decays and energy regimes. The result is that the different jet algorithms also result in jets with, for example, differently shaped cones. An overview over different jet clustering algorithms is given in Ref. [214].

6.5.1 The anti-k_{T} algorithm

The jets used for the analyses described in this thesis are clustered with the anti-k_{T} algorithm, described in Ref. [215], with a radius parameter of $R = 0.4$ using the FastJet software package, which is described in Ref. [216]. Such jets will be denoted as *AK4-jets* for the remainder of this thesis. The anti-k_{T} algorithm is a sequential recombination algorithm and is infrared and collinear safe. Starting with a collection of particles to be clustered the algorithm proceeds in three steps:

1. For each pair of input particles i and j with their respective transverse momenta $p_{\mathrm{T},i}$ and $p_{\mathrm{T},j}$ the particle-particle distance is calculated as

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2\alpha}, p_{\mathrm{T},j}^{2\alpha}\right) \frac{\Delta R_{ij}^2}{R^2}, \tag{6.3}$$

   where ΔR_{ij} is the distance of these particles in the (η, ϕ) plane. The parameter R denotes the choosable radius parameter that steers at which angular distances

particles are recombined and roughly corresponds to the radius of the resulting jets. The parameter α controls how the transverse momenta of the particles is taken into account in the recombination. For the anti-k_{T} algorithm it is set to $\alpha = -1$, while for example values of $\alpha = +1$ or $\alpha = 0$ lead to the k_{T} and Cambridge/Aachen algorithms. Also calculated is the so-called particle-beam distance for each input particle as:

$$d_{i\mathrm{B}} = p_{\mathrm{T},i}^{2\alpha}. \tag{6.4}$$

2. From all calculated distances d_{ij} and $d_{i\mathrm{B}}$ the smallest one is determined.

3. In case the minimal distance is found to be d_{ij}, the corresponding particles i and j are removed from the input collection. Their four-momenta are then added to create a single new particle that is added to the input collection. In case the minimal distance is a $d_{i\mathrm{B}}$ the particle i is called a jet and removed from the input collection. In either case the algorithm continues with step 1. and an updated list of input particles.

If there are no more particles in the input collection, all jets have been clustered and the algorithm terminates.

The jets used in this thesis are clustered from all particles previously found with the particle flow algorithm. To mitigate the effect of pileup a method called charged-hadron subtraction (CHS) is used and which is described in Refs. [165, 198]. For this, charged hadrons that likely originate from pileup interactions are identified by matching their tracks to pileup vertices. Those charged hadrons are then removed from the collection of input particles before the clustering algorithm is performed.

6.5.2 Jet energy calibration

After the clustering of a jet using the particle flow candidates measured in the detector, its energy should be equal to the energy of the underlying jet made up of the particles that ultimately originated in the hard scattering process. However, this is not necessarily the case. The response of the detector is not linear for jet constituents of different p_{T} and flavors and even varies from detector element to detector element. Particles originating from pileup interactions and detector noise can further influence the measured jet energy. Therefore, the jet energy scale (JES) of the measured jets is calibrated using a procedure comprised of multiple steps, which will be briefly described in the following and which is described in detail in Ref. [217, 218]. Detailed information about the procedures and their performance in LHC Run 2 are given in Refs. [219]. The calibration is applied by scaling the four-momenta of the jets by calibration factors that usually depend on their η and p_{T} values. For the determination of the correction factors the jets are first corrected using all correction factors derived in previous steps. In the following, the term *reco-level jet* denotes a jet clustered from constituents originating from the particle flow algorithm and therefore ultimately from measurements by the detector, whereas a *particle-level jet* is a jet that is clustered from the actual particles [2].

The first step of the calibration procedure is called *L1* and it has the purpose to remove the contribution to jet energy caused by detector noise and particles originating from pileup interactions. During the jet clustering procedure the application of the CHS procedure already removed charged hadrons that could be matched to pileup interactions. However, contributions from neutral hadrons still remain. In addition, there is a contribution of the so-called *out-of-time pileup* which is the misattribution of energy deposits from previous or subsequent bunch crossings to the current one and which is mainly caused by the finite

[2] Particle-level jets are obviously only available for simulated data.

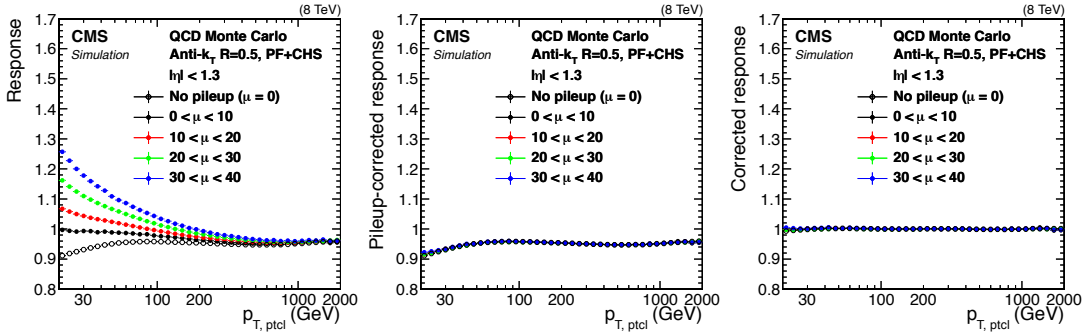
decay time of the signals in the calorimeters. To account for this, an offset depending on the p_T, η and area of the jet, as well as the *offset energy density* of the event is subtracted from the p_T of the jet. The offset energy density describes the energy per area not coming from the hard interaction and is determined for each event by adding a large number of particles with infinitesimal momentum called *ghosts* and then clustering jets. Most of the jets will then only contain such ghosts and their energy corresponds to the energy coming from pileup interactions. The offset energy density is then defined as the median of the energy per jet area of all jets, as is discussed in Ref. [220]. The correction factors are derived from simulated QCD multijet events by comparing the same event with and without the addition of simulated pileup interactions. Small differences between the measured energy offset densities in data recorded with the CMS detector and in simulated data remain and are determined by comparing measurements taken with a random trigger. Due to this, the applied L1 corrections slightly differ for recorded data and simulated data.

The second step is called *L2L3* [3] and accounts for the p_T dependence of the calorimeter response and tracking efficiency, as well as an observed non-uniformity of the response for different η values. The corrections are derived by matching reco-level jets in simulated QCD multijet events to their corresponding particle-level jets and comparing their transverse momenta. The L2L3 correction factors are parameterized in p_T and η and calculated as the ratio of the average p_T of all reco-level jets to the average p_T of all particle-level jets in this p_T and η bin.

The third calibration step is called *L2L3Residual* and accounts for any remaining differences between recorded and simulated data after the previous steps have been applied. It is only applied to recorded data and split into two parts. The first step corrects jets with large pseudorapidities relative to jets produced in the more central and better calibrated region of the detector. The correction factors are derived using dijet events in which one jet (called *tag jet*) has a pseudorapidity of $|\eta| < 1.3$ while the other jet (*probe jet*) is unconstrained in pseudorapidity. Since the events are dijet events, the transverse momenta of the two jets should balance and any imbalance could be caused by the jet energy scale. The correction factors are then calculated in such a way as to make the jet momenta equal when they are applied to the probe jet. The derivation also takes into account effects that lead to genuine momentum imbalances such as different resolutions of the energy measurements of the two jets and additional hadronic activity in the events. It further takes into account small dependencies of the correction factors on the average p_T of the jet pair. The correction factors are derived as a function of p_T and η of the probe jet. The second part corrects the absolute scale of the jet p_T in recorded data to that in simulated data by comparing the jet-energy response, determined by the difference of the measured jet p_T to a well reconstructed reference object using a p_T-balance method, between recorded and simulated data. For this, $Z(\rightarrow ee) + jet$, $Z(\rightarrow \mu\mu) + jet$ and $\gamma + jet$ events are used, in which the Z-boson or photon momenta can be reconstructed and used as the respective reference objects. The calibration of data recorded in the year 2016 additionally uses QCD multijet events in which a high momentum jet recoils against multiple jets. This correction is only derived for jets with $|\eta| < 1.3$.

The fourth and final calibration step is called *L5* and corrects differences between jets originating from particles of different flavors or gluons. Particles of different flavors produce jets with different radial energy distributions and particle contents. Since the previous correction levels are derived using samples with flavor compositions different than the flavor compositions present in specific phase spaces, additional correction factors are derived for specific flavor compositions and even jets of specific flavor (in case of simulated jets). For

[3] In the past these were two separate steps but the name was retained.



(a) (b) (c)

Figure 6.7: Average value of the ratio of the measured p_T of reco-level jets to the p_T of their matched particle-level jets for different levels of JES calibration: before any corrections (a), after the L1 corrections (b) and after all levels (c). The parameter μ denotes the average number of pileup interactions per bunch crossing. The jets are clustered in QCD multijet events, simulated with a center-of-mass energy of 8 TeV, and clustered with the anti-k_T algorithm with a radius parameter of $R = 0.5$. The applied jet energy corrections are those derived for data recorded in LHC run 1. Taken from Ref. [217].

the determination of these flavor-specific correction factors, samples with different flavor compositions are compared. The resulting correction factors are small and not used in this thesis. But uncertainties on the JES associated with the uncertainty of the flavor compositions of the samples used in the JES determination will be used in this thesis as will be discussed in chapter 8.

The effect of the JES calibration on the observed jet momenta is illustrated in Fig. 6.7, where the average ratio of the p_T of reco-level jets and the p_T of their matched particle-level jets is shown before any corrections are applied and after the jets are corrected only with the L1 correction or corrected with all JES levels.

All jets used in the analyses in this thesis are corrected using the L1 and L2L3 corrections and jets in recorded data are additionally corrected using the L2L3Residual corrections. Since the detector and run conditions change over time, the JES correction factors are determined separately for simulated data for the years 2016 and 2017 and in even smaller intervals for recorded data.

6.5.3 Jet Identification

Not all jets reconstructed and calibrated with the methods described above are used for the selection of events and further analyses in this thesis.

To remove jets that are clustered from noise, are badly reconstructed or are actually other objects that are reconstructed as jets, several quality criteria are required to be fulfilled. These criteria are chosen to only remove the unwanted jets and to retain almost all real and well reconstructed jets, as further discussed in Ref. [165]. They are listed in Table 6.2 and pose requirements on the number of hadronic and neutral constituents clustered in the jets as well as the energy fractions coming from neutral and charged constituents. The criteria differ slightly for the data recorded in 2016 and 2017, reflecting changes to the detector and operating conditions.

Although the effects of pileup interactions on the jets in an event are already mitigated by the previously discussed CHS method and L1 JES corrections, the possibility remains,

Table 6.2: Quality criteria for the selection of jets to be considered in the analyses, separately optimized for data recorded in the years 2016 and 2017.

Variable	2016	2017
Neutral hadron fraction	< 0.99	< 0.9
Charged hadron fraction	> 0	> 0
Neutral electromagnetic fraction	< 0.99	< 0.9
Charged electromagnetic fraction	< 0.99	-
Number of constituents	> 1	> 1
Charged particle multiplicity	> 0	> 0

and increases with the number of pileup interactions, that jets are clustered from constituents only coming from pileup interactions. Such jets are called *pileup jets*. To identify and subsequently reject such pileup jets various observables are combined into a single discriminant using boosted decision trees (BDTs) as described in detail in Ref. [165]. These observables describe the number of charged and neutral constituents, the fraction of energy associated with the primary vertex, the number of reconstructed vertices in the event, the radial distribution of energy as well as the shape of the jets. In non-pileup jets a large fraction of charged constituent energy should be associated to the primary vertex of the hard scattering process and constituents with large energy fractions should lie nearer to the jet axis as compared to pileup jets, for which the energy is more spread out in radial direction. A dedicated BDT is trained for data recorded in each year using simulated data. The need and effect of the BDT-based pileup-jet rejection is illustrated in Fig. 6.8. Shown are the ratios of all jets and only pileup jets to the number of jets originating from hard scattering as a function of the number of proton-proton interactions in an event, before and after the application of the BDT-based pileup-jet rejection. The study is performed using Drell-Yan events simulated for the conditions of the detector in the year 2018. Although these conditions differ from those for the analyses of data recorded in the years 2016 and 2017, the BDT-based rejection and its performance are similar.

Jets are also removed if a previously identified lepton is within a radius of 0.4 in the (η, ϕ) plane to the jet axis to avoid jets that include constituents that actually belong to the well reconstructed leptons.

Finally, jets are only selected if they have a transverse momentum of $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and a pseudorapidity of $|\eta| < 2.4$.

6.6 Identification of b jets

A large number of events recorded by the CMS detector contain the same number of jets as expected from $\mathrm{t\bar{t}H(b\bar{b})}$ production, but a significantly smaller fraction of events contain four b jets. With identification of such b jets, it is possible to predominantly select events containing b jets and thereby discarding a large number of background events. In addition, it is also possible to assign jets in selected events to whether they likely originate from the decays of the Higgs boson or the top quarks. The identification of b jets is performed using *b-tagging algorithms* and jets that are identified as such are called *b tagged*. Multiple b-tagging algorithms are used in the analyses of data recorded by the CMS experiment, two of which are used in the context of this thesis. The general principle of b tagging and

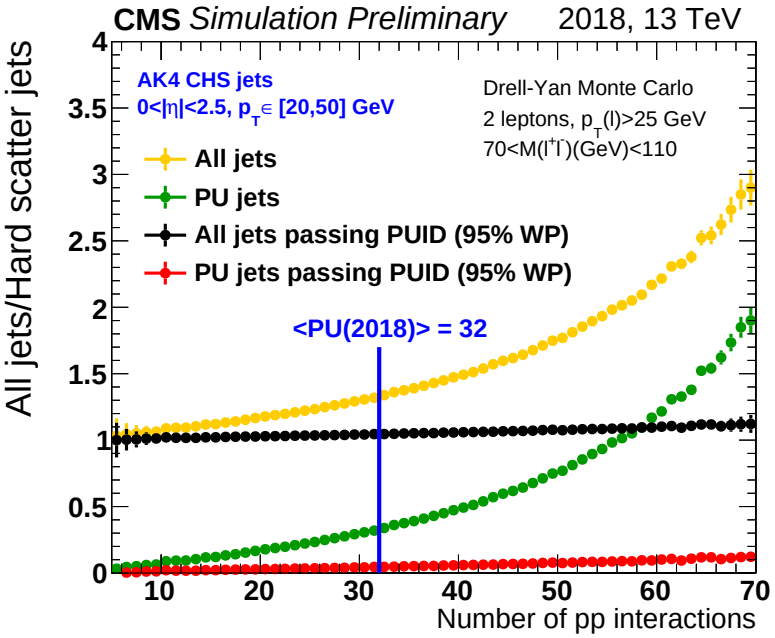


Figure 6.8: The ratios of the total number of jets with $|\eta| < 2.5$ and $p_T > 20\,\text{GeV}$ over number of jets form hard scattering interactions before (yellow) and after (black) applying the BDT-based pileup-jet rejection as a number of the proton-proton interactions in the event. Also shown are the ratios for only pileup jets before (green) and after (red) the rejection. The chosen working point corresponds to 95% efficiency. Taken from Ref. [221].

the algorithms used at the CMS experiment are discussed in the following. If not stated otherwise the information is based on the very detailed discussion of b tagging with the CMS detector found in Ref. [222]. In the following discussion it is helpful to distinguish between b jets and those originating from c quarks (*c jets*) or from lighter quarks or gluons (*udsg jets*). The identification of jets originating from top quarks is also possible with dedicated algorithms, but these algorithms differ in their working principle and will not be discussed here.

6.6.1 Principle of b tagging

The identification of b jets is based on several characteristics that set them apart from other jets and that are a result of the nature of the bottom quark. Since bottom quarks are lighter than top quarks and since flavor is conserved in QCD and QED interactions bottom quarks can only decay to an up or a charm quark and a W boson. These decays are described by the rather small off-diagonal elements of the CKM matrix (see Ch. 2) resulting in small transition amplitudes and a long lifetime of the bottom quark. Since it does not immediately decay, it hadronizes to a B meson which also has a long lifetime of around 1.5 ps, as listed in Ref. [7] and which can travel several millimeters inside the detector before it decays. This leads to the creation of a decay vertex that is spatially separated from the primary interaction vertex as illustrated in Fig. 6.9. The tracker of the CMS detector is designed to be able to reconstruct such *secondary vertices (SV)* that occur inside the tracker from *displaced tracks* that have a large impact parameter (IP) to the primary interaction vertex. Figure 6.10 shows an event recorded in the year 2016 at a

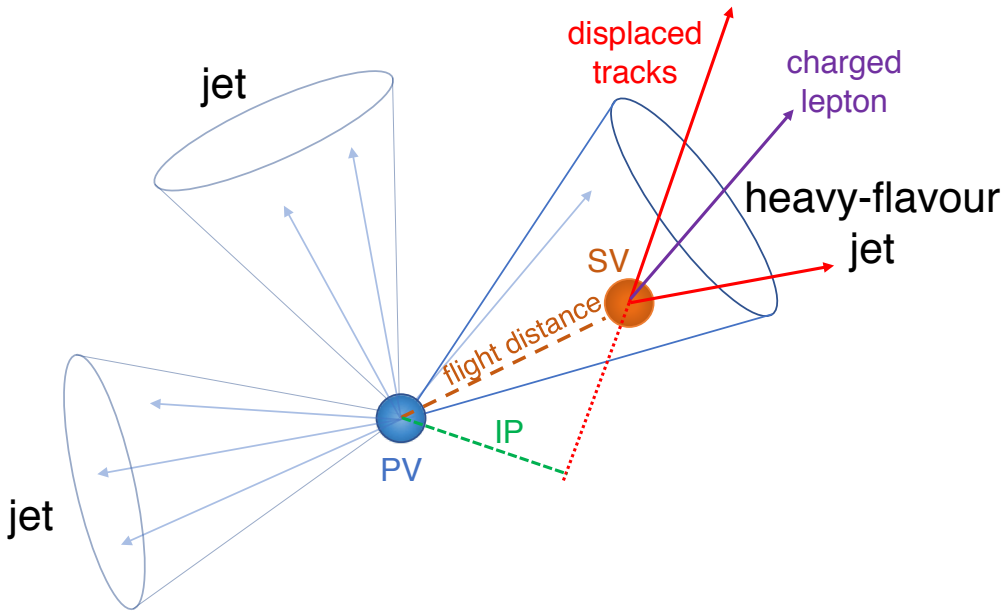


Figure 6.9: A secondary vertex occurring inside of a jet from the decay of a b or c hadron. The resulting charged-particle tracks are displaced with respect to the primary interaction vertex (PV) and therefore have a large impact parameter (IP) value. Taken from Ref. [222].

center-of-mass energy of 13 TeV containing three jets of which two include reconstructed secondary vertices. Both of these jets are b tagged using a b-tagging algorithm. Most of the other hadrons that occur inside of jets either have much longer or much shorter lifetimes leading to their almost immediate decay, for which the vertex cannot be distinguished from the primary vertex, or their survival until they are absorbed by the calorimeters. Similar to b jets, c jets with D mesons formed from c quarks also have a comparable lifetime of about 1 ps, as stated in Ref. [7]. The occurrence of a secondary vertex inside a jet is therefore a strong indication that the jet is a b or c jet. In addition, hadrons including b and c quarks have a relatively large mass compared to the masses of their decay products, which as a result have on average larger transverse momenta relative to the jet axis compared to the constituents of udsg jets. Such b and c jets therefore also have a different radial energy profile than udsg jets. Finally, decays of B and D mesons can include leptons in the final state, as listed in Ref. [7], therefore b and c jets can include *soft leptons* with momenta lower than the leptons expected from hard scattering interactions of interest. In summary, b jets have many characteristics that make them distinguishable from udsg jets. Since c jets share many of these characteristics they also can be distinguished from udsg jets but the distinction between b and c jets is more difficult, as will be seen later.

6.6.2 B-tagging algorithms

Observables describing the previously discussed jet characteristics can be used to distinguish between b, c and udsg jets. The combination of several observables using multivariate analysis methods however leads to much more powerful discriminants as is done for the b-tagging algorithms used at the CMS experiment. For the analyses in this thesis two different b-tagging algorithms are used. For the analyses of the data recorded in the year 2016 the *CSVv2* algorithm is used while for the analysis of the data recorded in the year 2017 the *DeepCSV* algorithm is used which was more developed at this point.

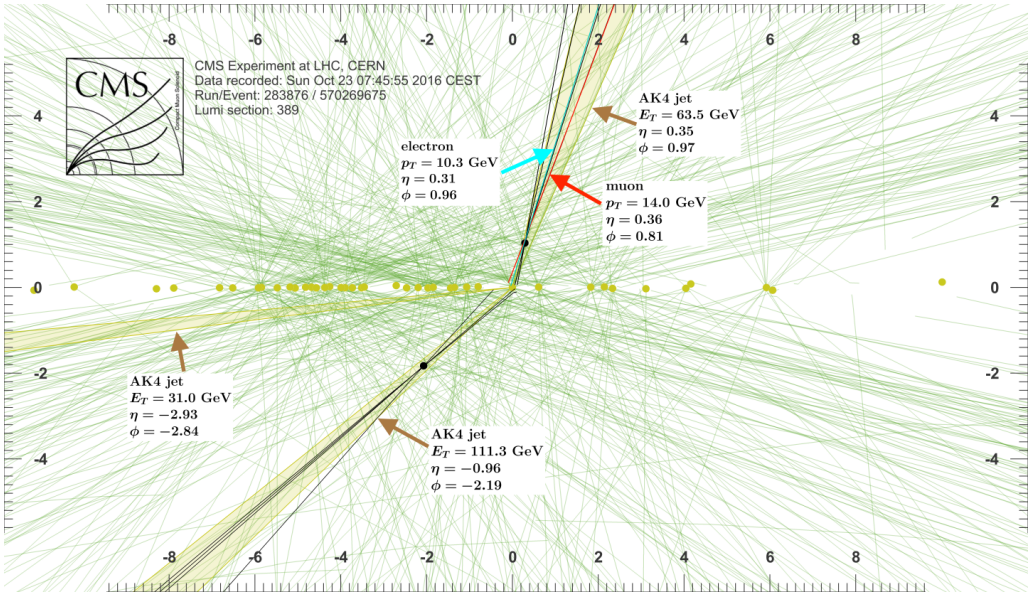


Figure 6.10: An event recorded in the year 2016 at a center-of-mass energy of 13 TeV containing three jets (shaded yellow) of which two include secondary vertices (black dots) with associated tracks (black lines). All reconstructed primary vertices (most of them from pileup interactions) are shown as yellow dots while the green lines indicate charge particle tracks with a $p_T > 0.5$ GeV. The dimensions are given in cm. Taken from Ref. [223].

Both of these algorithms use artificial neural networks (ANNs, see also Ch. 5) to distinguish between different jet flavors. The ANNs are trained on simulated data and both use the same jet observables as input features. The input features include variables describing the displacement to the primary vertex of the SV, the multiplicity, mass and jet energy fraction of the SV, the impact parameters of the tracks to the primary and secondary vertices, the momenta and energy ratios of the tracks and lastly the p_T and η of the jets themselves. For the training of the ANNs, simulated data is used for which the true flavor of a jet is defined using a ghost association method, which is described in Ref. [220]. For this, the jets are not only clustered from the constituents reconstructed by the particle flow algorithm. Instead the collection of clusterable constituents also include the b and c hadrons (available since this is simulated data with truth information) before they decay. To avoid any biases in the clustering order, the momenta of the b and c hadrons are scaled to a very small number. As a result the clustered jets are the same as if they had not included these so-called *ghosts* but the jets now include the ghost hadrons as their constituents. True b jets are now defined as those jets containing at least one b hadron, c jets as those with no b hadrons but at least one c hadron and udsg jets as those without neither b nor c hadrons. The training of the ANNs is also performed in a way as to not introduce biases depending on the p_T or η of the jets.

The CSVv2 algorithm uses binary ANNs trained to distinguish between b jets and other jets. The resulting discriminant distribution of the CSVv2 algorithm is shown in Fig. 6.11 for a sample of simulated $t\bar{t}$ events generated using the conditions for data taken in the year 2016. As expected, b jets can be well separated from udsg jets but less so from c jets.

The DeepCSV algorithm uses a more sophisticated ANN and is designed to not only distinguish between b jets and udsg jets, but also between them and c jets as well as jets containing more than one b or c hadrons (bb and cc jets). To achieve this it contains more layers and more neurons as well as five outputs, one for each target jet class. The five output

discriminants are constructed in such a way as to give the probability for a jet to belong to the respective class, with all of them summing to one. For the $t\bar{t}H(b\bar{b})$ analyses the b jet and bb jet discriminant outputs are summed to a single one. The output discriminant distributions are shown in Fig. 6.12 for a sample of simulated $t\bar{t}$ events generated using the conditions for data taken in the year 2016. As can be seen the b and c jet outputs serve to identify b and c jets and the udsg output identifies udsg jets as intended.

Fig. 6.13 shows the receiver operating characteristic (ROC) curve of the two b-tagging algorithms evaluated with simulated $t\bar{t}$ events generated using the conditions for data taken in the year 2016. It shows the misidentification probability for c and udsg jets and the identification efficiency for b jets for a fictive selection cut on the b-tagging discriminant output. It also shows the same curves for various other b-tagging algorithms which are described in Ref. [222] and not further discussed here. As can be seen, for a given b jet efficiency of for example 60%, the DeepCSV tagger has a smaller misidentification probability for both udsg jets and c jets than the CSVv2 tagger (not to be confused with the CSVv2(AVR) tagger). The misidentification probabilities for c jets are however worse than for udsg jets. As discussed in Ch. 3, the CMS detector was upgraded in the year 2017 including an improvement to the silicon tracker. This improvement leads to an improvement in tracking efficiency which in turn leads to an improvement in the b-tagging performance. This is shown for simulated $t\bar{t}$ events in Fig. 6.14 where the DeepCSV ROC curve is compared for the phase 0 detector as in the year 2016 and for the phase 1 detector in the year 2017. Also shown is the ROC curve of the DeepFlavor b-tagging algorithm, which is an even more sophisticated method that was developed after the analyses in this thesis had been finished.

For the analyses in this thesis a jet is declared as b tagged if the respective discriminant value exceeds a value defined by a mistag probability for udsg jets of approximately 1%. For the CSVv2 tagger used for the analysis of the data recorded in the year 2016 this is at a discriminant value of 0.8484 and corresponds to an efficiency for tagging b (c) jets of $\approx 65\%(10\%)$. For the DeepCSV tagger used for the analysis of the data recorded in the year 2017 the same udsg misidentification probability is at a discriminant value of 0.4941 for the sum of the b and bb jet ANN outputs. This corresponds to an efficiency for tagging b (c) jets of $\approx 76\%(18\%)$. The chosen working points for the b-tagging algorithms will also be called *medium working points* in the remainder of this thesis.

6.7 Missing transverse momentum

Particles that originate from the hard scattering interaction but leave the detector without being detected, such as the neutrinos from leptonic top-quark decays, produce an imbalance in the total transverse momentum with respect to all detected particles of the interaction. This imbalance can be quantified by a quantity called *missing transverse momentum* (also called MET or $\vec{\not{E}}$) and is defined in Ref. [198]. It is calculated as the component transverse to the beam axis of the vectorial sum of the transverse momenta of all particles reconstructed with the particle flow algorithm

$$\vec{\not{E}} = -\sum_{i=0}^{N_{\text{particles}}} \vec{p}_{\text{T},i}. \tag{6.5}$$

The corrections to the JES discussed above are propagated to the missing transverse momentum. For this, all particles contained in jets are first removed from the $\vec{\not{E}}$ sum and then the four momenta of the corrected jets are added again. It is also observed that the missing transverse momentum has a sinusoidal modulation in ϕ, possibly due to misalignment of the detector. An additional correction is applied to the missing transverse momentum vector to remove this modulation.

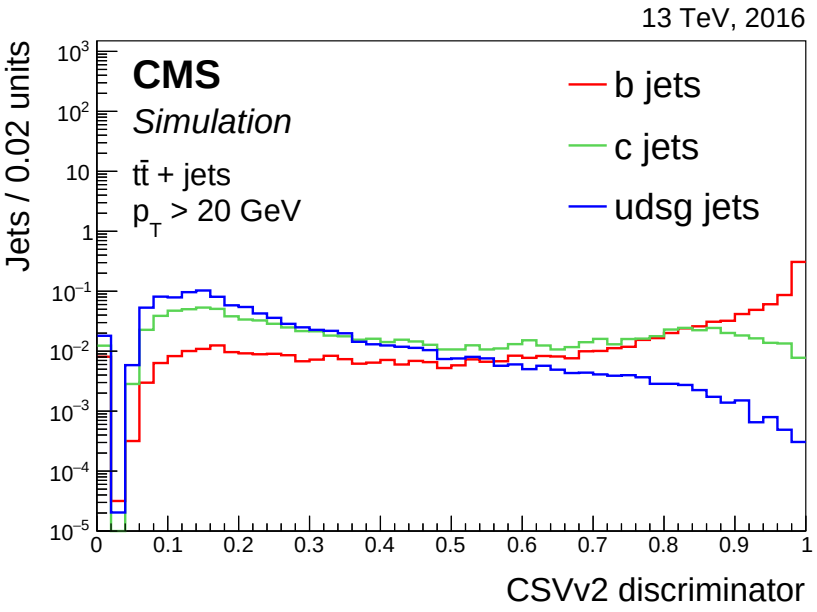


Figure 6.11: Distribution of the CSVv2 discriminant values for jets of different flavors in simulated $t\bar{t}$ events. The distributions for the different flavors are normalized to unit area to make shape comparisons easier. All jets without a selected track and secondary vertex are assigned a negative discriminant value and filled into the first bin. Taken from Ref. [222].

6.8 Triggers

As discussed in Ch. 3, events during data taking are only recorded by the CMS detector if they are accepted by at least one trigger path. During the generation of simulated data, the trigger system and its decisions are also simulated. Any trigger selections can therefore be applied to recorded as well as simulated data. There are many different triggers paths, many of them overlapping, designed to target specific events and suiting various kinds of analyses. The $t\bar{t}H(b\bar{b})$ analyses discussed in this thesis target events with semileptonically decaying $t\bar{t}$ pairs. Therefore, the triggers are chosen to select events containing one electron or muon. Although there are many other processes that contain one lepton, they can later be suppressed by additional requirements on the jets and their b-tagging discriminant values. Trigger paths can be *prescaled*, meaning that only a set fraction of events accepted by a trigger are recorded. For triggers targeting standard model processes with large cross sections this is a reasonable approach to reduce the number of recorded events to levels manageable by the computing system. However, the cross section for $t\bar{t}H(b\bar{b})$ production is small and therefore none of the chosen triggers is prescaled.

Events recorded in the year 2016 need to be accepted by a single-muon trigger requiring the presence of an isolated muon with a transverse momentum of $p_T > 24\,\text{GeV}$ or a single-electron trigger requiring the presence of a well-identified and isolated electron with a transverse momentum of $p_T > 27\,\text{GeV}$. The efficiencies of both triggers as function of the lepton p_T is shown in Fig. 6.15. The efficiencies are measured using a tag-and-probe method with $Z \to ee(\mu\mu)$ events where the tag lepton has to fulfill strict identification criteria. As can be seen the efficiencies increase with the p_T of the lepton until they reach a plateau. In the case of the single-muon trigger it can also be seen that the efficiencies differ for simulated (denoted by MC in the figure) and recorded data. Such differences are mitigated by reweighting the simulated data as described in Sec. 6.9.5. The offline selection

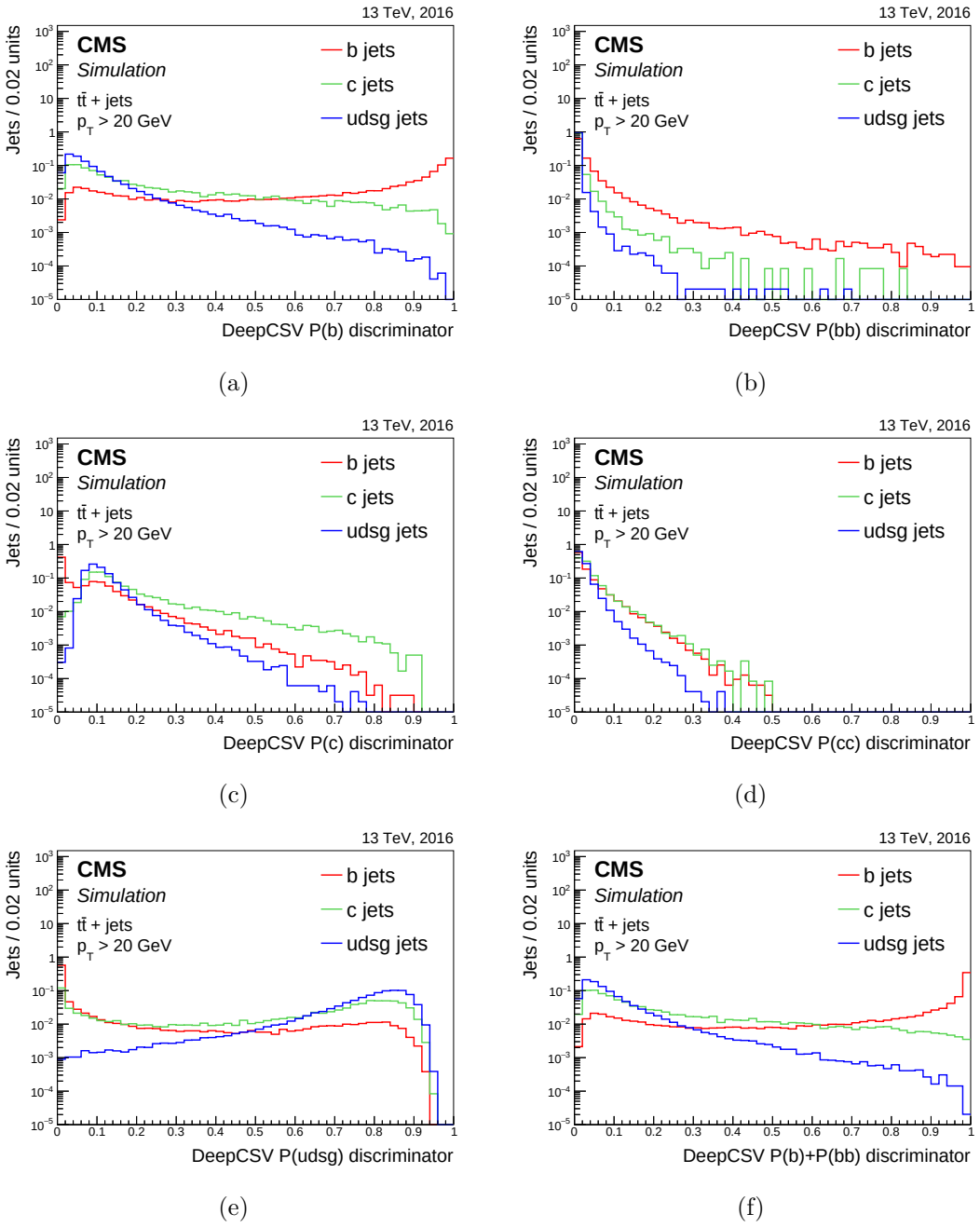


Figure 6.12: Distributions of the discriminant values of the DeepCSV algorithm. The distributions (a)–(e) are the output values of the ANN targeting (a) b jets, (b) bb jets, (c) c jets, (d) cc jets and (e) udsg jets while the last distribution (f) shows the sum of the b and bb jet outputs. All jets without a selected track and secondary vertex are assigned a negative discriminant value and filled into the first bin. Taken from Ref. [222].

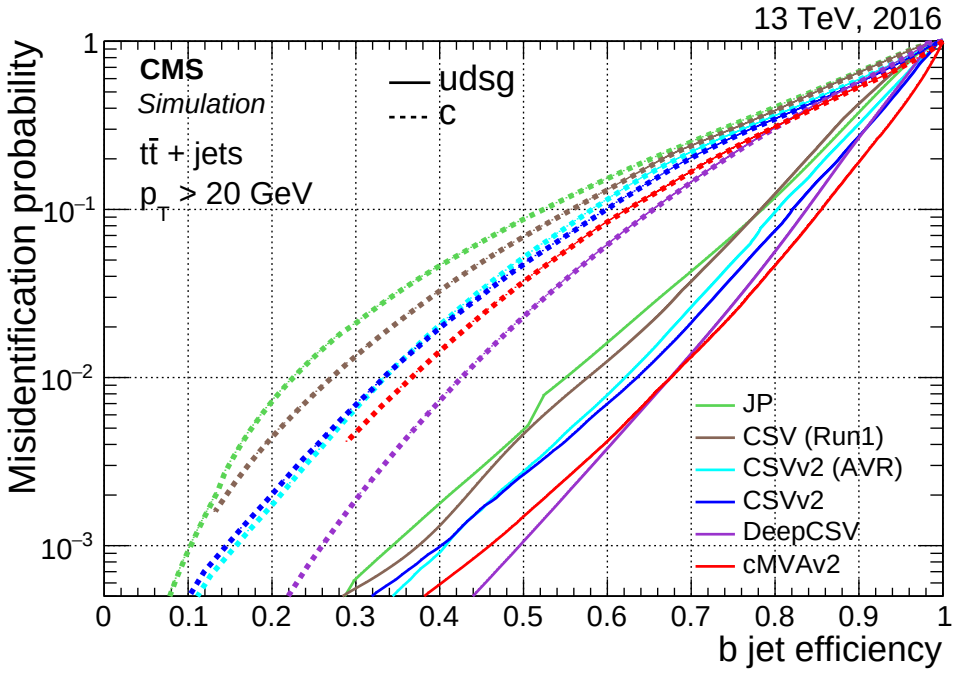


Figure 6.13: Receiver operating characteristic curve for various b-tagging algorithms evaluated with simulated $t\bar{t}$ events generated with the conditions of the detector in the year 2016. Taken from Ref. [222].

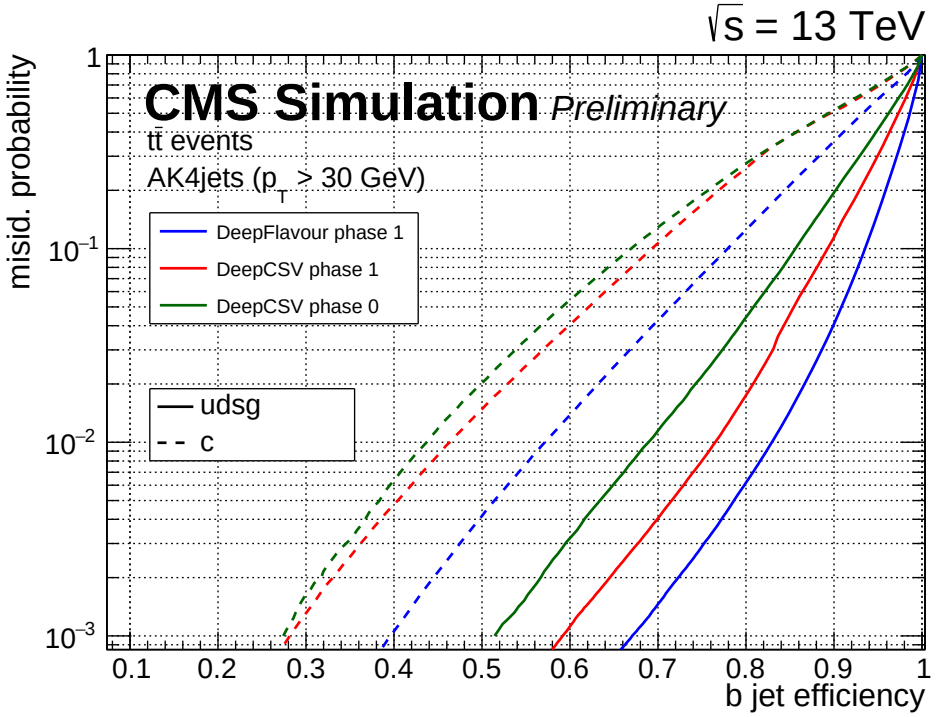
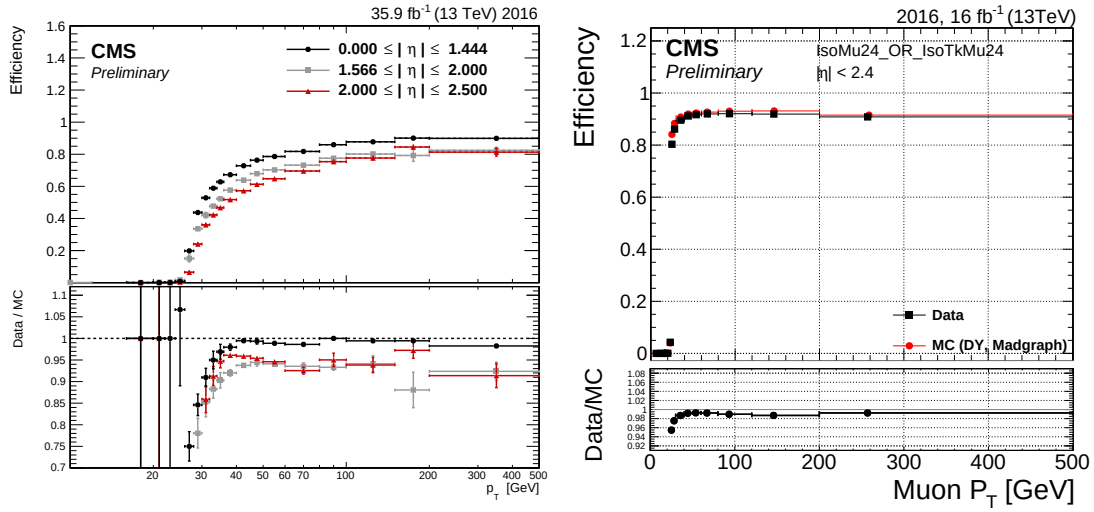


Figure 6.14: Receiver operating characteristic (ROC) curve for the DeepCSV algorithm evaluated with simulated $t\bar{t}$ events generated for the detector conditions in the years 3016 (phase 0) and 2017 (phase 1). Also shown is the ROC curve of the DeepFlavor algorithm for phase1. Taken from Ref. [224].



(a) (b)

Figure 6.15: Efficiencies as a function of lepton p_T for the triggers used for the analysis of data recorded in the year 2016 for: (a) the single-electron trigger [226] and (b) the single-muon trigger. Taken from Ref. [225].

cuts on the lepton p_T of 30(26) GeV for the electrons and muons respectively lie within the region of rising efficiency. Additional information can be found in Ref. [225, 226].

For events recorded in the year 2017 different triggers are available. The triggers targeting single electron events are a simple single-electron trigger requiring the presence of one isolated and well-identified electron with $p_T > 35\,\text{GeV}$ or an electron-jet trigger requiring one isolated and well-identified electron with $p_T > 28\,\text{GeV}$ and $|\eta| < 2.1$ as well as $H_T > 150\,\text{GeV}$ (H_T denoting the scalar sum of the transverse momenta of all jets with $p_T > 30\,\text{GeV}$, $|\eta| < 2.5$) as discussed in Ref. [227]. The simple single-electron trigger with a requirement of $p_T > 35\,\text{GeV}$ has the lowest p_T threshold of all un-prescaled single-electron triggers. Its exclusive usage would decrease the number of selected $t\bar{t}H$ events by a large amount. To still reach events with lower electron p_T the simple trigger is complemented by the electron-jet trigger with an additional requirement on the p_T of the jets in the event that is easily fulfilled by the energetic jets expected in $t\bar{t}H(b\bar{b})$ production.

The triggers targeting single muon events require the presence of one isolated muon with $p_T > 24\,\text{GeV}$ and $|\eta| < 2.1$ or one isolated muon with $p_T > 27\,\text{GeV}$. Both triggers were chosen because one of them was prescaled for a part of the data taking period.

6.9 Corrections for simulated data

At the time the simulated data are generated, the condition of the CMS detector during data taking might not be known with sufficient precision and might even change over time. Other effects might be very difficult to model by the simulation software. As a result simulated and recorded data differ in various ways, several of which can have large impacts on the $t\bar{t}H(b\bar{b})$ analyses. To reduce the differences between recorded and simulated data, additional correction steps are performed on simulated data. These are usually implemented as event weights that are individually calculated for each event, thereby changing the fraction the event contributes to the prediction of the simulated sample. The corrections applied for the $t\bar{t}H(b\bar{b})$ analyses in this thesis are discussed in the following.

6.9.1 JES and JER

After properly applying the JES corrections described above, the measured energies and momenta of jets in recorded and simulated data should be comparable. It is however observed that the resolution with which the jet energy is measured (JER) is worse in recorded data than simulated data. The JER is measured in recorded and simulated dijet events using an asymmetry method, as described in Ref. [217, 218]. For this, an asymmetry in measured jet p_{T} is defined as

$$\mathcal{A} = \frac{p_{\mathrm{T},1} - p_{\mathrm{T},2}}{p_{\mathrm{T},1} + p_{\mathrm{T},2}} \tag{6.6}$$

where $p_{\mathrm{T},1}$ and $p_{\mathrm{T},2}$ denote the p_{T} of the two jets. If both jets have similar η values the width of the distribution of $\mathcal{A}$ is correlated to the energy resolution of a single jet. To account for the observed differences the jet energy resolution in simulated data is decreased by a smearing procedure discussed in Ref. [228]. In case a particle-level jet can be matched to the reco-level jet that should be corrected, the jet four momentum is scaled by a factor of

$$c_{\mathrm{JER}} = 1 + (s_{\mathrm{JER}}(\eta^{\mathrm{reco}}) - 1)\frac{p_{\mathrm{T}}^{\mathrm{reco}} - p_{\mathrm{T}}^{\mathrm{particle}}}{p_{\mathrm{T}}^{\mathrm{reco}}} \tag{6.7}$$

where the $p_{\mathrm{T}}^{\mathrm{reco}}$ and $p_{\mathrm{T}}^{\mathrm{particle}}$ stand for the p_{T} of the reco-level and particle-level jet respectively and the $s_{\mathrm{JER}}(\eta^{\mathrm{reco}})$ is a scale factor that depends on the η of the reco-level jet. The scale factors $s_{\mathrm{JER}}(\eta^{\mathrm{reco}})$ are derived from the previously measured JER in recorded and simulated data as well as the *true* JER in simulated data defined as the distribution of the ratio of reco-level to particle-level jet p_{T}. In case that no particle-level jet can be matched to the reco-level jet, the reco-level jet is scaled by an amount randomly generated based on the measured JER in recorded and simulated data. For this the scaling factor is calculated as

$$c_{\mathrm{JER}} = 1 + \mathcal{N}(0, \sigma_{\mathrm{JER}}(p_{\mathrm{T}}^{\mathrm{reco}}, \eta^{\mathrm{reco}}, \rho))\sqrt{\max(s_{\mathrm{JER}}^2(\eta^{\mathrm{reco}}) - 1)} \tag{6.8}$$

where the $p_{\mathrm{T}}^{\mathrm{reco}}$, η^{reco} and $s_{\mathrm{JER}}(\eta^{\mathrm{reco}})$ are the same as above and $\sigma_{\mathrm{JER}}(p_{\mathrm{T}}^{\mathrm{reco}}, \eta_{\mathrm{T}}^{\mathrm{reco}}, \rho)$ is the JER measured for simulated jets parameterized as a function of jet p_{T}, η and the hadronic activity coming from pileup interaction in the event. The $\mathcal{N}(0, \sigma)$ denotes a randomly generated number with a Gaussian distribution around the mean 0 and the width σ. In both cases the resulting scale factor is truncated at zero to avoid negative jet energies. The determined JER scale factors $\sigma_{\mathrm{JER}}(p_{\mathrm{T}}^{\mathrm{reco}}, \eta_{\mathrm{T}}^{\mathrm{reco}}, \rho)$ are shown in Fig. 6.16 for data recorded in the years 2016, 2017 and 2018. As can be seen, the scale factors tend to be larger than one. As a result the JER correction has the effect to increase the differences between reco-level and particle-level transverse momenta.

6.9.2 b-tagging scale factors

The b-tagging algorithms employed in this thesis use various observables as inputs, several of which might not be modeled well by the generators used to create simulated data. Indeed, it is observed that jets with a given p_{T}, η and flavor have a different probability to be assigned a specific b-tagging discriminant value in recorded than in simulated data for both the CSVv2 and DeepCSV algorithms. This results in a difference in b-tagging efficiency as well as a difference in the shape of the discriminant distributions. Since both of these are important for the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses, simulated events are reweighted according to their p_{T}, η, b-tagging discriminant value and flavor, where the flavor of a simulated jet is defined using a ghost association procedure described in Sec. 6.6.2. The procedure for the derivation of these b-tagging scale factors is similar for the CSVv2 and DeepCSV b-tagging algorithms, is described in detail in Ref. [222] and will be sketched for the example of the CSVv2 case in the following.

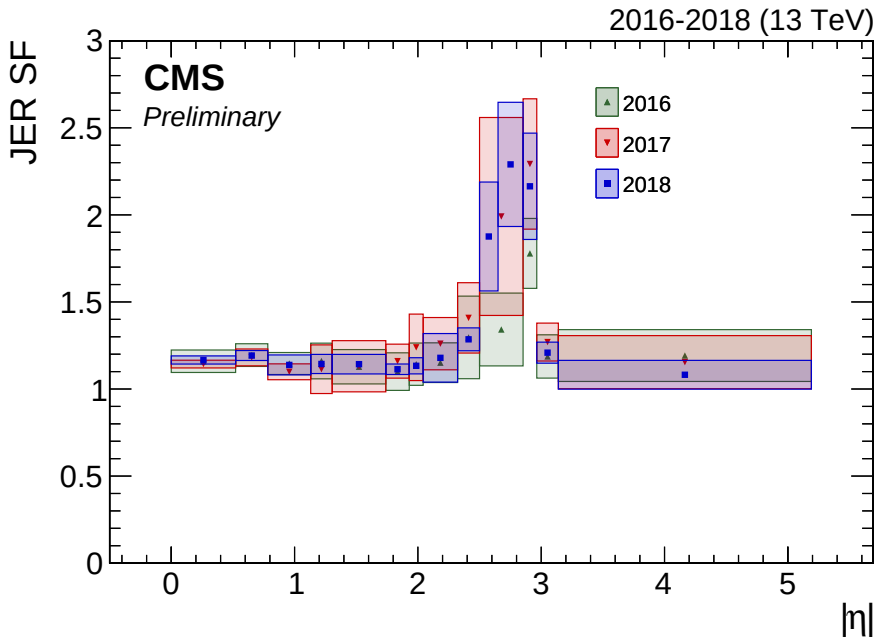


Figure 6.16: Measured JER scale factor as a function of η for data recorded in the years 2016-2018. The markers indicate the central values while the shaded areas indicate the corresponding total uncertainties. Taken from Ref. [219].

The scale factors are derived for udsg jets and b jets in phase space regions enriched in these jets using a tag-and-probe approach. No scale factors are derived for c jets due to the difficulty of defining a suitable phase space region. For the b-jet scale factors, a sample enriched in dileptonically decaying $\mathrm{t\bar{t}}$ events with exactly two jets is selected. One of the jets in the event is called the *tag jet* and needs to pass the medium working point of the CSVv2 tagger (see Sec. 6.6.2). The other jet is the *probe jet*. After this selection the selected sample consists to 87% of $\mathrm{t\bar{t}}$, 6% of single top and 7% of Z+jets events. For the udsg-jet scale factors, a sample enriched in Z+jets events is selected in which the Z boson decays to two leptons and with exactly two jets. Here, the tag jet is not allowed to pass the *loose working* point, which has a b-tagging efficiency lower than the medium working point. The selected sample consists of 99.9% Z+jets events. The same selections are applied to simulated events. The b-tagging discriminant distributions of the probe jets in simulated data are normalized to the ones observed in recorded data.

Then the scale factors are derived in bins of p_T and η as well as the b-tagging discriminant value of the probe jets. In the b jet region, the contribution of udsg-flavor jets is estimated from simulated data, removed from the simulated data and subtracted from the recorded data. During this calculation true c jets are treated as true b jets. The analogue operation is performed for the udsg region where now the contributions from b and c jets are removed. The scale factors for each bin of p_T, η and b-tagging discriminant value are then calculated such as to obtain the same number of events in simulated and recorded data. For the b jets this can be formulated as

$$S_\mathrm{b}(p_\mathrm{T}, \eta, \mathrm{CSVv2}) = \frac{D_\mathrm{b}(p_\mathrm{T}, \eta, \mathrm{CSVv2}) - M_\mathrm{udsg}(p_\mathrm{T}, \eta, \mathrm{CSVv2})}{M_\mathrm{b}(p_\mathrm{T}, \eta, \mathrm{CSVv2})} \tag{6.9}$$

where the D and M stand for the bin content in recorded and simulated data respectively and the subscripts udsg or b indicate the flavor of the probe jet. Due to the subtraction of b jets, the scale factors derived for udsg jets will be influenced by the scale factors for b

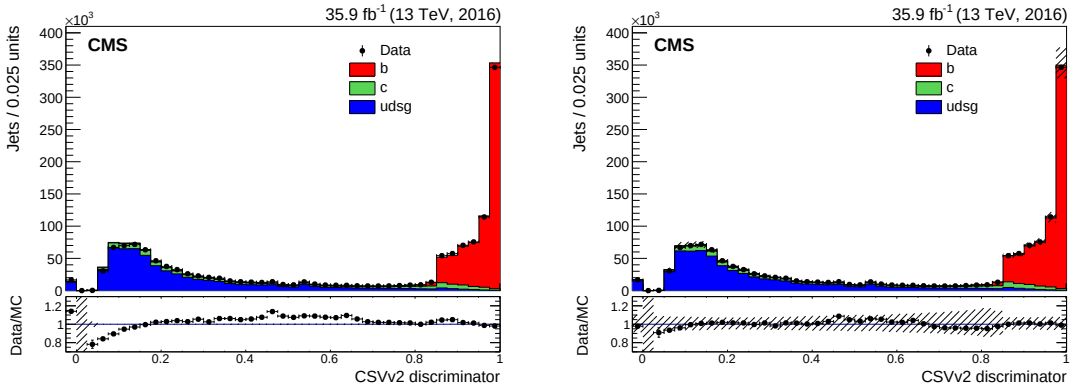


Figure 6.17: Distribution of the CSVv2 discriminant values before (left) and after (right) the application of the b-tagging scale factors. The distribution is shown for a sample of semileptonically decaying $t\bar{t}$ events with exactly four jets of which two have to pass the medium working point of the CSVv2 algorithm. The hatched band shows the statistical uncertainty (left), and the total uncertainty (right) in the measured scale factors. The simulation is normalized to the total number of data events. The bin below 0 contains the jets with a default discriminant value. Taken from Ref. [222].

jets and vice versa. To account for this the procedure is performed iteratively with the first iteration not taking into account the scale factors. For subsequent iterations the scale factors of the previous iteration are applied to the probe jets before the subtraction. The iterations are performed until the resulting scale factors converge. To mitigate the effect of statistical fluctuations on the result, the final scale factors are parameterized as function of the b-tagging discriminant value using a 6th order polynomial for the udsg-jet scale factors and a smooth interpolation for the b-jet scale factors. The procedure is affected by several uncertainties related to the simulation-based event estimation and the parameterization. These are of great importance in the $t\bar{t}H(b\bar{b})$ analyses and will be further discussed in Ch. 8.

For the application of the scale factors to simulated data, a scale factor is retrieved for each jet in an event, all scale factors are multiplied and the product is applied as a weight to the event. For this purpose c jets have a scale factor of unity. The effect of applying scale factor is shown for the CSVv2 b-tagging algorithm in Fig. 6.17 where it can be seen that they improve the agreement between simulated and recorded data.

6.9.3 Pileup reweighting

The number of pileup interactions in a bunch crossing depends on the total inelastic proton-proton cross section and the instantaneous luminosity, which in turn depends on the conditions of the beam during data taking and usually changes during the data taking period. For proton-proton collisions inside the CMS detector the inelastic proton-proton cross section is found to be $\sigma_{\text{inelastic}} = 69.2\,\text{mb}$ in Ref. [165]. Since pileup introduces additional particles into an event and can be mitigated only to a certain extent, the amount of pileup in an event has an influence on most observables. It is therefore important that the number of pileup interactions is modeled well in simulated data. The samples of simulated data for the years 2016 and 2017 had however been generated before the data in the respective years was fully recorded and the number of simultaneous proton-proton interactions could not be realistically simulated. To correct the simulated data, the distribution of pileup

interactions is reweighted to match that observed in recorded data. In simulated data this distribution is known since the number of pileup interactions is stored for each event. Since not every pileup interaction will be reconstructed, the distribution in recorded data has to be estimated from the instantaneous luminosity measured for each *luminosity section* of the data taking period. The average number of pileup interactions μ per bunch crossing in a luminosity section with instantaneous luminosity $\mathcal{L}$ is given by

$$\mu = \sigma_{\text{inelastic}} \frac{L}{f} \tag{6.10}$$

where f is the number of bunch crossings per unit time (the instantaneous luminosity is also measured by unit time). The number of pileup interactions for an event in this bunch crossing will then be distributed around the average μ according to a Poisson distribution. This way the distribution of all events can be constructed.

6.9.4 Lepton scale factors

The reconstruction and identification efficiencies for electrons and muons are found to be different for recorded and simulated events, as is shown in Sec. 6.4 and in more detail in Refs. [209, 210, 212, 213]. To alleviate the disagreement, scale factors are derived and applied to simulated data. First, the efficiency of the quantity that is to be corrected is measured in bins of the p_{T} and η of the lepton (or in some cases the η of the supercluster associated to an electron) using tag-and-probe methods on a sample of recorded or simulated $\text{Z} \to \text{ee}(\mu\mu)$ events. The observed differences between recorded and simulated data are taken as scale factors that are then used as weights for simulated events. For muons recorded in the years 2016 and 2017 the scale factors correct the modeling of the identification and the isolation criteria. In addition, for muons recorded in the year 2016 scale factors are also applied for the tracking efficiency. These tracking efficiency scale factors were not yet available at the date of the publication of the analysis of the data recorded in the year 2017. For electrons, scale factors correcting the modeling of the reconstruction and the identification are applied for data recorded in both years.

For the electrons (muons) relevant for this thesis the scale factors take on values of up to 10%(5%). The associated uncertainties are smaller and on the order of a few percent.

6.9.5 Trigger scale factors

As can be seen in Fig. 6.15b, the efficiency of the single-muon trigger used for the analysis of data taken in the year 2016 differs between recorded and simulated data. Similar differences are found for the other triggers in both years. To reduce the disagreement between simulated and recorded data, scale factors are derived that are used to reweight the simulated events.

For the single-electron trigger used in the year 2016, the efficiencies for recorded and simulated data are measured using a tag-and-probe method on a sample of $\text{Z} \to \text{ee}(\mu\mu)$ events. The tag-electron is required to be well reconstructed and the efficiencies are measured for the probe-electron. The scale factors are then calculated in bins of electron p_{T} and η as the ratio of the trigger efficiencies measured in recorded and simulated data. For the electron triggers used in the year 2017, the scale factors are derived for the combination of both triggers (using a logical OR operator) using a reference-trigger method. Here, the trigger efficiencies are calculated with respect to an uncorrelated reference trigger. In this case the reference trigger is a combination of single-muon triggers requiring isolated muons with a $p_{\text{T}} > 24\,\text{GeV}$. The efficiencies are measured using a sample enriched in dileptonically decaying $\text{t}\bar{\text{t}}$ events that are accepted by the single-muon triggers and that contain one muon

and one electron, both of which have to be well-identified and isolated. The efficiency of the probe trigger is then defined as

$$\mathrm{Eff}_{\mathrm{probe}} = \frac{n(\mathrm{ref}, \mathrm{probe})}{n(\mathrm{ref})} \tag{6.11}$$

where $n(\mathrm{ref})$ denotes the number of events passing the reference trigger while $n(\mathrm{ref}, \mathrm{probe})$ denotes the number of events passing the probe trigger as well as the reference trigger. The scale factors are again calculated in bins of electron p_{T} and η as the ratio of the trigger efficiencies measured in recorded and simulated data.

For the single-muon triggers the scale factors are derived for both years using a tag-and-probe method, described in Ref. [164], analogous to the one used for the single-electron trigger in the year 2016.

For events used in the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses the scale factors can change the number of predicted events by up to 20% for the single-electron triggers and up to 10% for the single-muon triggers, showing the importance of this correction to simulated data. The associated uncertainties of the scale factors are, however, of the order of 1-3%.

6.9.6 L1 prefiring scale factors

The L1 trigger system is explained in Ref. [166] and is designed decentralized and hierarchically to minimize the time needed for its decisions. Trigger decisions are first made by local groups of nearby detector components based on simple criteria. A local ECAL trigger, for example, could *fire* based on the energy sum of several nearby ECAL crystals. If such a local trigger is fired the information is passed to the next level which combines the information of multiple local groups and then passes it to the global trigger that makes the final decision.

Degradation of ECAL crystals in the foreward region of the detector, caused by the accumulation of radiation damage over time, resulted in an issue regarding the association of ECAL deposits to the correct bunch crossing. For some events large ECAL deposits in the region between $2.0 < \eta < 3.0$ were associated with the previous instead of the current bunch crossing. As a result the L1 trigger could *prefire* one bunch crossing early. The L1 trigger is, however, not allowed to accept events based on the same local groups in two consecutive bunch crossings. This led to the effect that affected events effectively vetoed themselves during data taking. Although the effect already occurred during data taking in the year 2016 it was only reported during the analysis of the data recorded in the year 2017. This effect is therefore not taken into account in the analysis of data recorded in the year 2016 and also not present in any simulated data for either years. To account for the effect in the analysis of data recorded in the year 2017, simulated events are reweighted according to their probability to self-veto. The correction is implemented in the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis as scale factors that decrease the number of predicted events in the analysis categories (see Ch. 7). The scale factors are calculated for all background processes together and separately for the signal process. Differences between the signal and the background scale factors are found to be very small. The effect of the scaling is a decrease of about 2-3% in the predicted number of events.

7 Analysis strategy

In order to search for $t\bar{t}H$ production, proton-proton collision data that have been recorded with the CMS detector in the years 2016 and 2017 during LHC Run 2 at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, are analyzed. Originally, the data taken in 2016 were analyzed first and independently from the data taken in 2017, which were analyzed a year later, resulting in three different publications of the results, which were presented in Refs. [4, 5, 229]. The overall strategies of all analyses are very similar but still differ in various details. For one, the experimental conditions during both data taking periods differed due to upgrades to the CMS detector. In addition, the analysis methods applied in the analysis of the 2016 data had since then undergone further development and refinement.

These measurements of $t\bar{t}H(b\bar{b})$ production include all decay modes of top quark-antiquark pairs. Although the work performed in this thesis concerns the single-lepton (SL) channel, where top quark-antiquark pairs decay semileptonically, the analysis of this channel was studied with the whole $t\bar{t}H(b\bar{b})$ analysis in mind and in close collaboration with the analyses of the channels targeting the other $t\bar{t}$ decay modes (dileptonically (DL) or fully-hadronically (FH) decaying top quark-antiquark pairs). The general analysis strategy is similar for the analyses of 2016 and 2017 data and is discussed in Sec. 7.1. It is also similar for all three $t\bar{t}$ decay channels, but will only be described in detail for the analyses in the SL channel. The samples of simulated and recorded data that are used for the analyses in SL channel are discussed in Sec. 7.2. Not all events recorded with the CMS detector are of interest for a $t\bar{t}H(b\bar{b})$ analysis. The selection criteria of events included in the analyses are discussed in Sec. 7.3. Samples of simulated data are used in these analyses to train artificial neural networks (ANNs) to discriminate between signal and background processes and to predict the expected contributions and kinematic distributions of signal and background processes in a given phase space. To this end it is of importance to verify whether the simulated data describe the real data reasonably well. Such comparisons are performed on a rather general level in Sec. 7.4. The design of the ANNs that are used to discriminate between signal and background processes, the observables used as their input features and the validation of the description of these input features by simulated data is described in Sec. 7.5. The distributions of the outputs of these ANNs will later be used as templates in a binned maximum likelihood fit to extract the signal contribution in data (see Ch. 4). These templates are shown in Sec. 7.6. Following this discussion of the single-lepton channel, the analyses of the dileptonic and fully-hadronic channels are briefly summarized in Sec. 7.7.

7.1 General analysis strategy

For the relevant center-of-mass energy of 13 TeV the predicted production $t\bar{t}H$ cross section $\sigma(t\bar{t}H) = 507^{+35}_{-50}$ fb, as calculated in Ref. [92], is much smaller than the total inelastic proton-proton cross section of about 80 mb, which is given in Refs. [230, 231]. Therefore, the first step in all analyses is to select from the recorded data the events that are compatible with $t\bar{t}H$ production and a subsequent decay of the Higgs boson into a bottom quark-antiquark pair.

Events are required to fulfill several selection criteria designed to construct signal regions enriched in signal events or control regions enriched in certain background processes. First, during data taking, events need to be accepted by triggers that require the presence of either one or two electrons and/or muons, several jets, a certain amount of total transverse energy or combinations thereof. Next, the recorded events are reconstructed using more sophisticated methods than allowed by the time-sensitive online-reconstruction used for the trigger system. During this step, jets are clustered and b-jets are identified, as discussed in Ch. 6. Events are further selected based on the multiplicities and kinematics of leptons, jets and b-tagged jets they contain. These selection criteria are motivated by the number of final-state objects one would expect from $t\bar{t}H(b\bar{b})$ production and the top quark decay modes. For example, events in the dilepton channel should contain exactly two well reconstructed leptons with high transverse momenta and a minimum of missing transverse momentum to account for the neutrinos of the W-boson decays, as well as four b-tagged jets (two from the decaying top quarks and two from the decaying Higgs boson). However, these requirements are generally relaxed to account for acceptance or efficiency effects. All event selection criteria are chosen in such a way as to ensure that events can be unambiguously selected into only one analysis channel. To further suppress background contributions, additional channel specific selection criteria are applied depending on the channel. This will be discussed for the SL channel later in this chapter.

After the event selection, the samples of data still mostly consist of various background processes, although with a larger contribution of $t\bar{t}H$ events than before the selection. The dominant background in all channels stems from the production of top quark-antiquark pairs with additional jets (referred to as $t\bar{t}$+jets). Of special interest is the case where the top quark-antiquark pair is produced with two additional b jets. Since the particles in the final state of this process are the same as in the $t\bar{t}H(b\bar{b})$ process and the properties of the resulting jets are very similar to those found in $t\bar{t}H(b\bar{b})$, this particular background process cannot be easily discarded by selection criteria and is an irreducible background to these analyses. In the FH channel, a large background contribution consists of QCD multijet production, whereas in the SL and DL channels this process is effectively suppressed by the requirement that the events contain isolated leptons. The remaining background originates from the production of: single top quarks with additional jets (referred to as single t), W bosons or $Z/\gamma^\star$ bosons with additional jets (referred to as W+jets and Z+jets, or commonly referred to as V+jets), top quark-antiquark pairs in association with a W or Z boson ($t\bar{t}$+W and $t\bar{t}$+Z or commonly referred to $t\bar{t}$+V), and pairs of vector bosons (WW, ZZ, WZ or commonly referred to as diboson).

The $t\bar{t}H(b\bar{b})$ signal contribution in the selected samples is still very small compared to the background processes. For the measurement of the $t\bar{t}H$ signal strength with a high significance it is, however, necessary to find regions of phase space in which the contribution of $t\bar{t}H$ production is sufficiently large compared to the total number of events. In all three channels this is achieved using multivariate analysis methods that combine multiple weakly separating observables into a single one that better distinguishes between the signal and background processes. In the analyses of the SL channel, artificial neural networks (ANNs)

are trained to distinguish between $t\bar{t}H(b\bar{b})$ and several classes of $t\bar{t}$ events, where the latter are defined by the flavors of the additional jets that do not originate from the top quark decays. Other backgrounds are not considered, since they tend to behave less signal-like than $t\bar{t}$ events and can be well separated using ANNs trained against $t\bar{t}$ events. It has proven beneficial to train separate ANNs depending on the multiplicity of jets in the events, since different jet multiplicities allow for different observables to be exploited in the training. The trained ANNs are then evaluated on recorded and simulated data alike, resulting in an unambiguous classification for each event to either be a $t\bar{t}H(b\bar{b})$ event or one of the classes of $t\bar{t}$ events. For each predicted class and jet multiplicity, distributions are constructed by filling the ANN output values of the events assigned to the class into a histogram. The distributions of each class are enriched in events of the targeted class and each distribution in turn has regions that are even more enriched in the specific class. The histograms of the distributions are created separately for recorded data and simulated data for the $t\bar{t}H$ and the various background processes. Finally, the histograms of simulated data are jointly fit to the corresponding histograms of recorded data using a maximum likelihood fit, as will be further explained in Ch. 9.

7.2 Recorded and simulated data

In the following the samples of recorded and simulated data used for the analysis in the SL channel are discussed.

7.2.1 Recorded data

The analyses of $t\bar{t}H(b\bar{b})$ are performed with data recorded in 2016 and 2017, corresponding to integrated luminosities of $35.9\,\text{fb}^{-1}$ and $41.5\,\text{fb}^{-1}$, respectively. During the operation of the CMS experiment temporary malfunctions of individual detector components can occur. Events recorded at a time in which such a malfunction was registered are discarded.

7.2.2 Simulated samples

Several samples of simulated data are utilized for the prediction of signal and background contributions to the selected data, for the study of reconstruction or selection efficiencies and for the training of the ANNs. All simulated samples are created using Monte Carlo event generators (see Sec. 2.3) with either leading-order (LO) or next-to-leading-order (NLO) accuracy depending on the simulated process. Where applicable the mass of the Higgs boson is set to 125 GeV and the mass of the top quark to 172.5 GeV. For the samples used in the analyses of the data recorded in the year 2016 (2017) the proton structure is described by the NNPDF3.0 (NNPDF3.1) parton density functions in the 5FS, described in Ref. [53], and the parton shower, the hadronization and the underlying event are simulated using Pythia (v.8.200) (Pythia (v.8.2.3.0)), which is described in Ref. [71]. The simulated properties of the underlying event are adjusted to describe data using empirically derived "tunes". The samples used in the analysis of the data recorded in the year 2016 are simulated using the CUETP8M2T4 tune, described in Ref. [232], in case of the $t\bar{t}H$, $t\bar{t}$ and single top samples and the CUETP8M1 tune, described in Ref. [233] for all other processes. For the samples used in the analysis of the data recorded in the year 2016 the CP5 tune, discussed in Ref. [234], is used. Additional proton-proton interactions in the same bunch crossing (pileup) are simulated by adding minimum-bias events using the appropriate version of the Pythia generator. The event simulation is finalized with a detailed simulation of the interactions of the particles with the CMS detector based on Geant4 (v.9.4), described in Ref. [74], which allows simulated events to be analyzed in the same way as events recorded in the CMS detector. The simulated samples are generated

independently for the analyses of the 2016 and 2017 data, with the simulation of pileup and the detector reflecting the conditions in the respective data taking periods.

The simulation of $\mathrm{t\bar{t}H}$ production with all decay channels of the Higgs boson is performed at NLO accuracy with the POWHEG (v.2) event generator, which is described in Refs. [65–67, 235]. The $\mathrm{t\bar{t}}$+jets background and the contributions from t- and tW-channel production to the single top background are simulated at NLO using POWHEG (v.2) as is discussed in Ref. [236, 237]. The $\mathrm{t\bar{t}}$+W and $\mathrm{t\bar{t}}$+Z backgrounds and the s-channel contribution to the single top background are simulated at NLO accuracy using MADGRAPH5_aMC@NLO (v.2.2.2), described in Ref. [64]. The W+jets and Z+jets processes are generated at LO accuracy using MADGRAPH5_aMC@NLO (v.2.2.2) or MADGRAPH5_aMC@NLO (v.2.4.2) with the MLM prescription to match the matrix-element final state to the parton shower, as discussed in Ref. [60]. For the $\mathrm{t\bar{t}}$+W simulation the matrix-element final state is matched to the parton shower using the FxFx scheme, which is discussed in Ref. [61]. Events for the WW, ZZ and WZ processes are simulated at LO accuracy using PYTHIA (v.8.200) or PYTHIA (v.8.2.3.0) for the analyses of the data recorded in the years 2016 and 2017 respectively.

All simulated samples are normalized to the expected number of events for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ or $41.5\,\mathrm{fb}^{-1}$ according to the production cross sections and branching fractions predicted by the standard model for the center-of-mass energy of 13 TeV. These predicted cross sections are calculated to at least NLO accuracy with the mass of the top quark set to 172.5 GeV and the mass of the Higgs boson set to 125 GeV where appropriate. The calculations are accompanied by uncertainties related to the choice of the renormalization and factorization scales, the parton density functions and the strong coupling constants. As will be discussed in Ch. 8, these uncertainties will be considered during the fits of simulated data to recorded data. The normalization is implemented by weighting each simulated event with the predicted cross section divided by the total number of simulated events in the sample before any selection is applied.

The $\mathrm{t\bar{t}H}$ sample is normalized to the predicted production cross section of $\sigma_{\mathrm{t\bar{t}H}} = 0.5071^{+0.0291}_{-0.0462}(\mathrm{scale})^{+0.0181}_{-0.0181}(\mathrm{pdf}+\alpha_\mathrm{s})\,\mathrm{pb}$ for a Higgs boson mass of 125.0 GeV, calculated at NLO accuracy in QCD and including electroweak corrections as provided in Ref. [92]. The sample is split into subsamples based on the Higgs boson decays of the simulated events. The normalization of the $\mathrm{t\bar{t}H(b\bar{b})}$ part is adjusted to account for the $\mathrm{H}\to\mathrm{b\bar{b}}$ branching fraction $\mathcal{B}_{\mathrm{H}\to\mathrm{b\bar{b}}}$ predicted by the standard model as given in Ref. [92]. The branching fractions of the other decay channels are predicted by the generators of the sample, there sum is however scaled to $1-\mathcal{B}_{\mathrm{H}\to\mathrm{b\bar{b}}}$. Only a very small amount of $\mathrm{t\bar{t}H}$ events with Higgs boson decays into particles other than bottom quarks will survive the event selection. Therefore, the possible mismodelling caused by this approach should have negligible effects on the analysis and the result.

The predicted cross section for $\mathrm{t\bar{t}}$ production is calculated at NNLO accuracy with resummation of soft gluon terms to next-to-next-to-leading-logarithmic accuracy as $\sigma_{\mathrm{t\bar{t}}} = 831.76^{+19.77}_{-29.20}(\mathrm{scale})^{+35.06}_{-35.06}(\mathrm{pdf}+\alpha_\mathrm{s})\,\mathrm{pb}$, as described in Refs. [79–85]. The events in the simulated $\mathrm{t\bar{t}}$ sample are normalized according to this production cross section and the branching fractions of the top quark given in Ref. [7]. To allow for a more sophisticated treatment of $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$ production in the analysis, the $\mathrm{t\bar{t}}$ sample is further divided into subsamples according to the flavor of the additional jets not originating from the decay of the top quark-antiquark pairs. For this consideration, the flavor of a jet is defined using a ghost hadron procedure previously described in Sec. 6.6.2. The $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$ sample consists of events containing two or more additional b jets. In case events contain only one additional b jet, but that jet contains two or more b hadrons they are sorted into a $\mathrm{t\bar{t}}$+2b sample

instead. This can occur for example when the bottom quarks originate from the splitting of a gluon with a large momentum and are merged into one jet. The $t\bar{t}$+b sample consists of events with exactly one additional b jet containing exactly one b hadron, which can be caused by the other b hadron being produced outside of the acceptance of the detector. Events containing no additional b jets, but containing one or more c jets are labeled as $t\bar{t}$+$c\bar{c}$ events and all other events are labeled as $t\bar{t}$ + light flavor events.

The predicted cross sections for t- and s-channel single top production are calculated at NLO accuracy as described in Refs. [107–109], and for the tW-channel at approximate NNLO accuracy, as described in Refs. [109–111]. For the prediction of both the $t\bar{t}$+W and the $t\bar{t}$+Z processes, two samples are simulated. The background from $t\bar{t}$+W and $t\bar{t}$+Z is estimated using samples of events where the W or Z bosons decay into quarks or leptons. The samples are normalized according to production cross sections calculated at NLO accuracy as provided in Ref. [238] and the branching fractions for the W- or Z-boson decays given in Ref. [7]. For the W+jets and Z+jets processes only events with W-boson decays to a lepton and a neutrino or Z-boson decays to leptons are considered since the other decay modes can be neglected due to the requirement that selected events contain a lepton. The samples of W+jets and Z+jets events are normalized to cross sections calculated at NNLO accuracy using the FEWZ (v.3.1) program as described in Refs. [239, 240]. The cross section for the WW background is calculated at NNLO accuracy and those for the WZ and ZZ backgrounds are calculated at NLO accuracy using the MCSM framework as described in Refs. [241–244]. Contributions from QCD multijet production and other processes are very small and are neglected for the further analyses. All cross section values are listed in Tab. 7.1.

Finally, all samples of simulated data are corrected to account for known differences between simulated data and recorded data as described in Sec. 6.9.

7.3 Event selection

All events considered in the analyses have to fulfill several selection criteria. First the events have to be accepted by at least one of the several triggers discussed in Sec. 6.8 and which require the events to contain one electron or muon with high transverse momentum. Next, events have to pass a series of filters that reject events with badly reconstructed particle flow candidates, events in which known noise patterns occurred or events in which there are particles originating from interactions of the beam with the beampipe or remaining gas (so called *beam-halo interactions*). Accepted events have to contain a well reconstructed primary vertex as described in Sec. 6.1.4.

All events have to further pass a selection designed to only retain events compatible with $t\bar{t}H(b\bar{b})$ while rejecting most other events. For this, events have to contain exactly one well-reconstructed electron or muon, as defined in Sec. 6.4, with a sufficiently large transverse energy to account for the lepton expected by the semileptonically decaying top quark-antiquark pair. Events in which the $t\bar{t}$ decay produces a τ lepton that in turn decays into an electron or muon are also accepted. For the analysis of data recorded in the year 2016, events have to contain either one electron with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$ or a muon with $p_T > 26\,\text{GeV}$ and $|\eta| < 2.1$. They are not allowed to contain any additional electrons or muons with $p_T > 15\,\text{GeV}$ and $|\eta| < 2.4$. Events recorded in the year 2017 have to contain either one electron with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ or a muon with $p_T > 29\,\text{GeV}$ and $|\eta| < 2.4$ and similarly no additional leptons with $p_T > 15\,\text{GeV}$ and $|\eta| < 2.4$. The selection thresholds differ between both years due to an adjustment to the trigger thresholds related to a higher instantaneous luminosity during the year 2017. To account for the presence of the neutrino, events recorded in both years are required to have a missing transverse

Table 7.1: Calculated cross sections and associated uncertainties used for the normalization of the samples of simulated data.

Process	Cross section (pb)	Uncertainty (pb)	Reference
$t\bar{t}H$	0.507	$^{+0.029}_{-0.046}$(scale)$^{+0.018}_{-0.018}$(pdf+α_s)	[92]
$t\bar{t}$+jets	831.76	$^{+19.77}_{-29.20}$(scale)$^{+35.06}_{-35.06}$(pdf+α_s)	[79–85]
single (anti-)top s-channel, W $\to$ lν	3.36	$^{+0.09}_{-0.08}$(scale)$^{+0.09}_{-0.09}$(pdf+α_s)	[7, 107–109]
single top t-channel	136.02	$^{+4.09}_{-2.92}$(scale)$^{+3.52}_{-3.52}$(pdf+α_s)	[107–109]
single anti-top t-channel	80.95	$^{+2.53}_{-1.71}$(scale)$^{+3.18}_{-3.18}$(pdf+α_s)	[107–109]
single top tW-channel	35.85	$^{+0.90}_{-0.90}$(scale)$^{+1.70}_{-1.70}$(pdf+α_s)	[109–111]
single anti-top tW-channel	35.85	$^{+0.90}_{-0.90}$(scale)$^{+1.70}_{-1.70}$(pdf+α_s)	[109–111]
$t\bar{t}$+W (W $\to l\nu$)	0.210	$^{+13\%}_{-11.6\%}$(scale)$^{+1.7\%}_{-1.3\%}$(pdf+α_s)	[7, 238]
$t\bar{t}$+W (W $\to qq$)	0.435	$^{+13\%}_{-11.6\%}$(scale)$^{+1.7\%}_{-1.3\%}$(pdf+α_s)	[7, 238]
$t\bar{t}$+Z (Z $\to ll/\nu\nu$)	0.263	$^{+10.3\%}_{-11.7\%}$(scale)$^{+2.0\%}_{-2.5\%}$(pdf+α_s)	[7, 238]
$t\bar{t}$+Z (Z $\to qq$)	0.611	$^{+10.3\%}_{-11.7\%}$(scale)$^{+2.0\%}_{-2.5\%}$(pdf+α_s)	[7, 238]
W+jets (W $\to l\nu$)	61526.7	$^{+497.1}_{-264.6}$(scale)$\pm$2312.7(pdf+α_s)	[239, 240]
Z+jets (Z $\to ll$, $m_{ll} > 50$ GeV),	6077.22	$\pm$121.5(scale) $\pm$ 14.78(pdf)	[239, 240]
WW	118.7	$^{+2.5\%}_{-2.2\%}$	[241]
WZ ($m_{ll} > 12$ GeV)	65.54	$\pm$2.37(scale) $\pm$2.88(pdf)	[242–244]
ZZ ($m_{ll} > 12$ GeV)	15.83	$\pm$0.51(scale) $\pm$0.59(pdf)	[242–244]

momentum of at least 20 GeV (see Sec. 6.7). Events recorded in both years have to contain at least four well-identified jets with $p_T > 30$ GeV and $|\eta| < 2.4$, at least two of which have to be b tagged at the medium working point. The expected number of jets in $t\bar{t}H(b\bar{b})$ events with H $\to b\bar{b}$ and semileptonically decaying top quark-antiquark pairs is six, of which four should be b jets. To allow for some of those jets to fail their selection criteria or to fall out of the detector acceptance, as well as to allow for the construction of a larger control region enriched in $t\bar{t}$ events, the analysis also accepts events with less jets.

The phase space defined by this *baseline selection* serves as a control region containing mostly $t\bar{t}$ + light flavor events with which the general modeling of the recorded data by simulated data is studied. The actual analyses of the $t\bar{t}H(b\bar{b})$ process are performed on a smaller phase space requiring an additional b-tagged jet and in which the $t\bar{t}$ + light flavor contribution is much smaller. The selection criteria for the jets and b tagging are further discussed in Sec. 6.5 and Sec. 6.6. All selection criteria are summarized in Tab. 7.2.

Although this event selection is designed to select $t\bar{t}H(b\bar{b})$ events, the selected sample also contains a small amount of $t\bar{t}H$ events with Higgs decays to particles other than bottom quarks. If not stated otherwise, all observables denoted by $t\bar{t}H$ always contain the targeted $t\bar{t}H(b\bar{b})$ events as well as these small amount of $t\bar{t}H$(non-$b\bar{b}$) events.

7.4 Quality of simulated data

To study the quality of the description of the recorded data by the simulated data the baseline selection, requiring at least four jets of which two have to be b tagged, is used.

Table 7.2: Event selection criteria for the analysis of events recorded in the years 2016 and 2017.

	2016	2017
Event filters	pass	pass
Number of primary vertices	1	1
$\vec{\not{E}}$ (GeV)	> 20	> 20
Number of leptons	1	1
p_{T} of electron (muon) (GeV)	$> 30(26)$	$> 30(29)$
$\lvert\eta\rvert$ of electron (muon)	$< 2.1(2.1)$	$< 2.4(2.4)$
p_{T} of additional electrons (muons) (GeV)	$< 15(15)$	$< 15(15)$
$\lvert\eta\rvert$ of additional electrons (muons)	$> 2.4(2.4)$	$> 2.4(2.4)$
Number of jets	> 4	> 4
p_{T} of jets (GeV)	> 30	> 30
$\lvert\eta\rvert$ of jets	< 2.4	< 2.4
Number of b-tagged jets (baseline)	> 2	> 2
Number of b-tagged jets (analysis)	> 3	> 3

The numbers of events selected from recorded data and the predicted numbers of events of the various simulated processes are shown in Tab. 7.3 and Tab. 7.4 for data recorded in the years 2016 and 2017, respectively. The total number of selected events is predicted reasonably well by the simulation. The largest contribution stems from the $\mathrm{t\bar{t}}$ + light flavor, $\mathrm{t\bar{t}}$+$\mathrm{c\bar{c}}$ and single t processes while $\mathrm{t\bar{t}H}$ has only a very small contribution.

The observed and predicted multiplicities of jets and b-tagged jets are shown for both years in Fig. 7.1. The distributions predicted from simulated data describe the distributions observed in recorded data within their uncertainties. As can be seen, many events contain additional jets not coming from $\mathrm{t\bar{t}}$ or $\mathrm{t\bar{t}H}$ systems, several of which are even btagged. They also show that $\mathrm{t\bar{t}}$ + light flavor events tend to have a small number of b-tagged jets while the contribution from $\mathrm{t\bar{t}}$ + heavy flavor and $\mathrm{t\bar{t}H}$ increases with the number of b-tagged jets. No single bin of these distributions, or even the combination of jet and b-tagged jets multiplicity selection, contains a large enough fraction of $\mathrm{t\bar{t}H}$ events to extract the signal yield from observed data with a high significance. As will be discussed in the next section, regions with even larger signal-to-background ratios are constructed using ANNs, based on a multitude of observables. Most of these observables describe the momenta of jets and leptons, the angles between them and b-tagging discriminants of the jets. Distributions of the momenta and pseudorapidities of the jets and the momentum of the lepton in events are shown in Fig. 7.2 while the missing transverse momentum, the scalar sum of the transverse momenta of all jets and the average $\Delta\eta$ between jets are shown in Fig. 7.3. Jets and leptons in $\mathrm{t\bar{t}H}$ events tend to have higher transverse momenta and lie more centrally in the detector. In addition, the jets in $\mathrm{t\bar{t}H}$ events tend to be more spread out, the events are more isotropic. In the distributions the simulated data models the recorded data within its uncertainties. Figure 7.4 shows the distributions of the b-tagging discriminant outputs of the jets. While for the analysis of data recorded in the year 2016 the CSVv2 algorithm is used, the analysis of the data recorded in the year 2017 uses the DeepCSV algorithm as

Table 7.3: Number of events observed in data recorded in the year 2016 and predicted by corresponding simulations after the baseline selection requiring one lepton and at least 4 jets of which at least 2 have to be b tagged. The quoted uncertainties are statistical uncertainties only.

Process	2016
$t\bar{t}$ + light flavor	463658 ± 174
$t\bar{t}$+$c\bar{c}$	76012 ± 70
$t\bar{t}$+b	22416 ± 38
$t\bar{t}$+2b	9052 ± 24
$t\bar{t}$+$b\bar{b}$	10897 ± 27
single t	25215 ± 166
V+jets	12309 ± 58
$t\bar{t}$+V	2456.5 ± 12
WW/WZ/ZZ	449 ± 14
Total background	622466 ± 263
$t\bar{t}$H	1232 ± 2
Data	610556

Table 7.4: Number of events observed in data recorded in the year 2017 and predicted by corresponding simulations after the coarse selection requiring one lepton and at least 4 jets of which at least 2 have to be b tagged. The quoted uncertainties correspond to statistical and systematic uncertainties (excluding the additional $t\bar{t}$ + heavy flavor rate uncertainty) and are further discussed in Ch. 8.

Process	2017
$t\bar{t}$ + light flavor	718341 ± 83944
$t\bar{t}$+$c\bar{c}$	96581 ± 13795
$t\bar{t}$+b	27222 ± 3749
$t\bar{t}$+2b	10537 ± 2206
$t\bar{t}$+$b\bar{b}$	12770 ± 2050
single t	38170 ± 3720
V+jets	14491 ± 1754
$t\bar{t}$+V	2963 ± 286
WW/WZ/ZZ	503 ± 61
Total background	921576 ± 97714
$t\bar{t}$H	1747 ± 167
Data	923936

described in Sec. 6.6. Since $t\bar{t}H(b\bar{b})$ and $t\bar{t}+b\bar{b}$ events tend to have many b jets this is an important observable to distinguish such events from $t\bar{t}$ + light flavor and other background processes. The distributions are well modeled in the relevant region of high b-tagging discriminant output values. Distributions of other quantities can be found in Appendix A.

7.5 Multivariate analysis

To further increase the sensitivity to $t\bar{t}H$, a multivariate analysis with ANNs is performed. For this, only events with at least three b-tagged jets are considered. Events with only two b-tagged jets are enriched in $t\bar{t}$ + light flavor and discarded for the remainder of the analyses. A general overview over ANNs and an explanation of the terms is given in Ch. 5 and here only the specifics relevant to the analyses will be discussed.

During the development of the analyses, various strategies had been tested and compared such as different ANN designs, the usage of Boosted Decision trees, the usage of single discriminating variables or combinations thereof. Several of these studies are discussed in Refs. [137, 245–248]. The presented strategy was chosen based on the expected results of the full $t\bar{t}H$ analyses.

7.5.1 Design of ANNs

The ANNs are designed to classify events as one of six processes: $t\bar{t}H$, $t\bar{t}$ + light flavor, $t\bar{t}+c\bar{c}$, $t\bar{t}+b\bar{b}$, $t\bar{t}$+2b or $t\bar{t}$+b. Compared to a classification into either $t\bar{t}H$ or background events, this *multiclass* approach has the benefit that it results in six regions, each of them enriched in one of these six processes. Since a perfect separation between the processes is not feasible, the resulting regions enriched in $t\bar{t}H$ will have contributions of all background process. Each of the five background processes is in turn affected by different dominant systematic uncertainties. With the construction of regions dedicated to each process the relevant uncertainties can be better constrained during the combined fit to data (discussed in Ch. 9), leading to smaller uncertainties in the $t\bar{t}H$ region. To achieve this sixfold separation multiclass ANNs are used, with each ANN having six output nodes: one for each process (henceforth labeled as $t\bar{t}H$-node, $t\bar{t}$ + light flavor-node etc.). The loss function that is minimized during the ANN training is the categorical cross-entropy discussed in Sec. 5.2. As a result, each event that is evaluated with a given ANN gets assigned six output values which can be interpreted as the probability of the event to belong to each class. Since each event should only be used once in the analysis, each event gets assigned to the class for which the output node takes on the highest value. This is illustrated in Fig. 7.5. Events are only filled in the distribution to which they have the highest probability to belong to and the value of an event in the distribution is another measure that they belong to the corresponding class. In a $t\bar{t}H$-node distribution, for example, all events have the highest probabilities to belong to the $t\bar{t}H$ class instead of the other classes, and those events that are very $t\bar{t}H$-like accumulate on the right-hand side of the distribution, while those that are less $t\bar{t}H$-like accumulate on the left-hand side. The distributions of the node output values of all events that are assigned to the respective classes serve as discriminant distributions that will be fit to data to extract the $t\bar{t}H$ signal strength. This approach therefore results in six discriminant distributions per ANN. The resulting distributions are shown in Sec. 7.6.

The ANNs are trained using simulated data for which the true class of an event is known, as described in Sec. 5.2. For the training only $t\bar{t}H(b\bar{b})$ and $t\bar{t}$ events are considered, omitting the other background processes and signal events where the Higgs boson decays to particles other than pairs of bottom quarks. To minimize effects from overtraining these samples are only used for this purpose and are not considered for the further analyses and the extraction of the $t\bar{t}H$ signal strength. The samples are chosen to be sufficiently large as

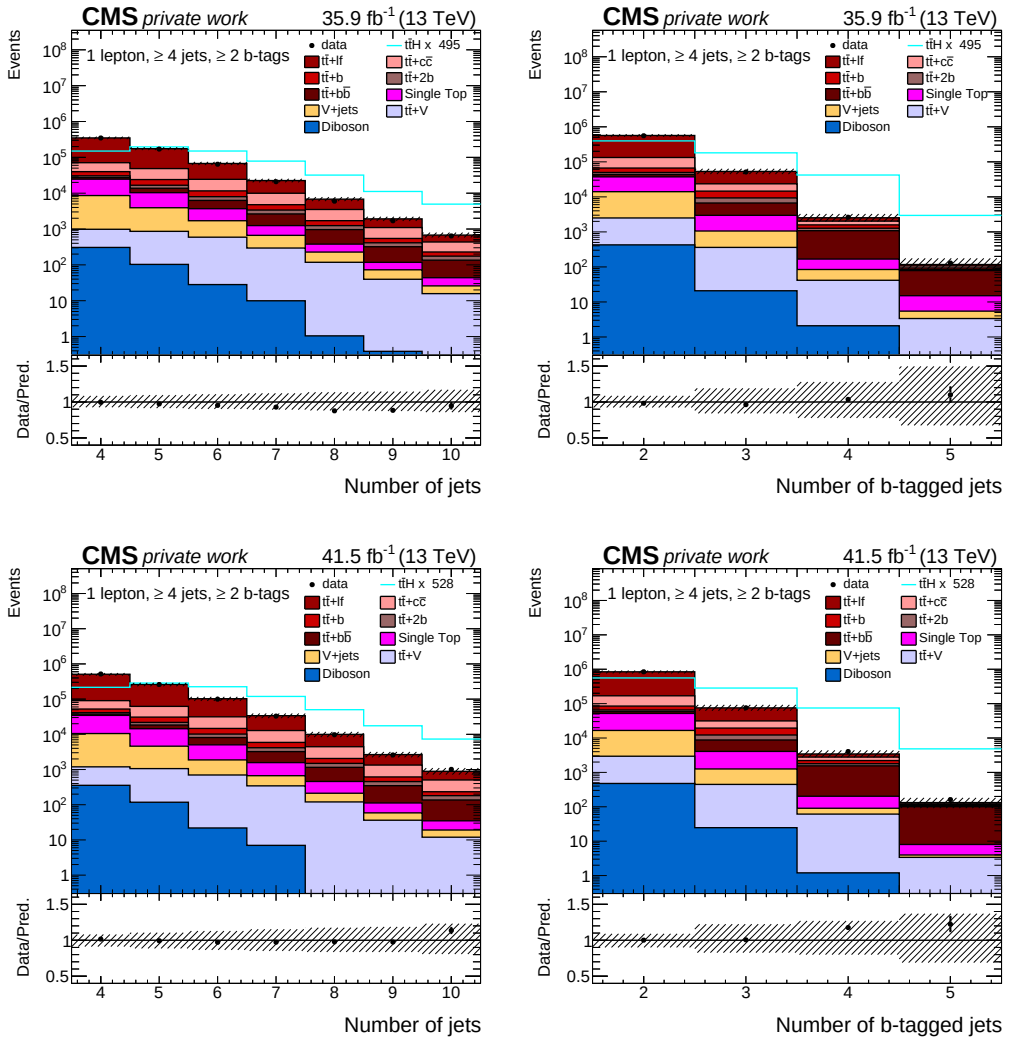


Figure 7.1: Multiplicities of jets (left) and b-tagged jets (right) in data recorded in the years 2016 (top) and 2017 (bottom) after the baseline selection. The distributions observed in recorded data are represented by the black markers. The expected background contributions are shown as stacks of filled histograms. The expected distributions for the $t\bar{t}H$ process as blue line which is additionally scaled to the integral of the background for better visibility. For the 2016 distributions the last bins contain all events with 10 (5) or more jets (b-tagged jets). The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding the additional 50% rate uncertainty on $t\bar{t}$ + heavy flavor (see Ch. 8). The bottom panels show the ratio of data to the background prediction. Taken from [4, 5].

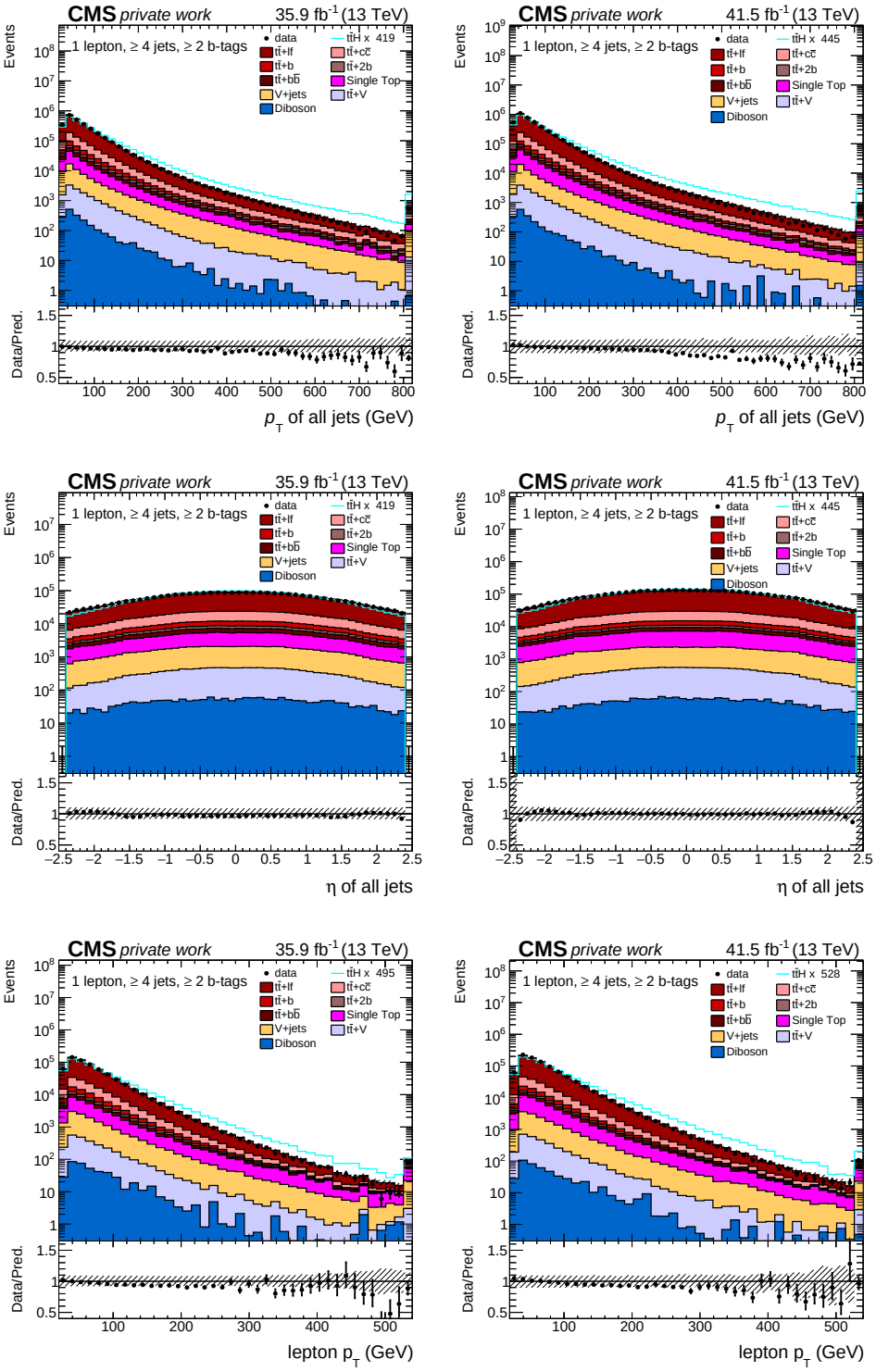


Figure 7.2: Transverse momenta (top) and pseudorapidities (middle) of jets as well as the transverse momenta of leptons (bottom) in data recorded in the years 2016 (left) and 2017 (right) after the baseline selection. The distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $t\bar{t}H$ process as blue line which is additionally scaled to the integral of the background for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding those affecting only the rate. The bottom panels show the ratio of data to the background prediction.

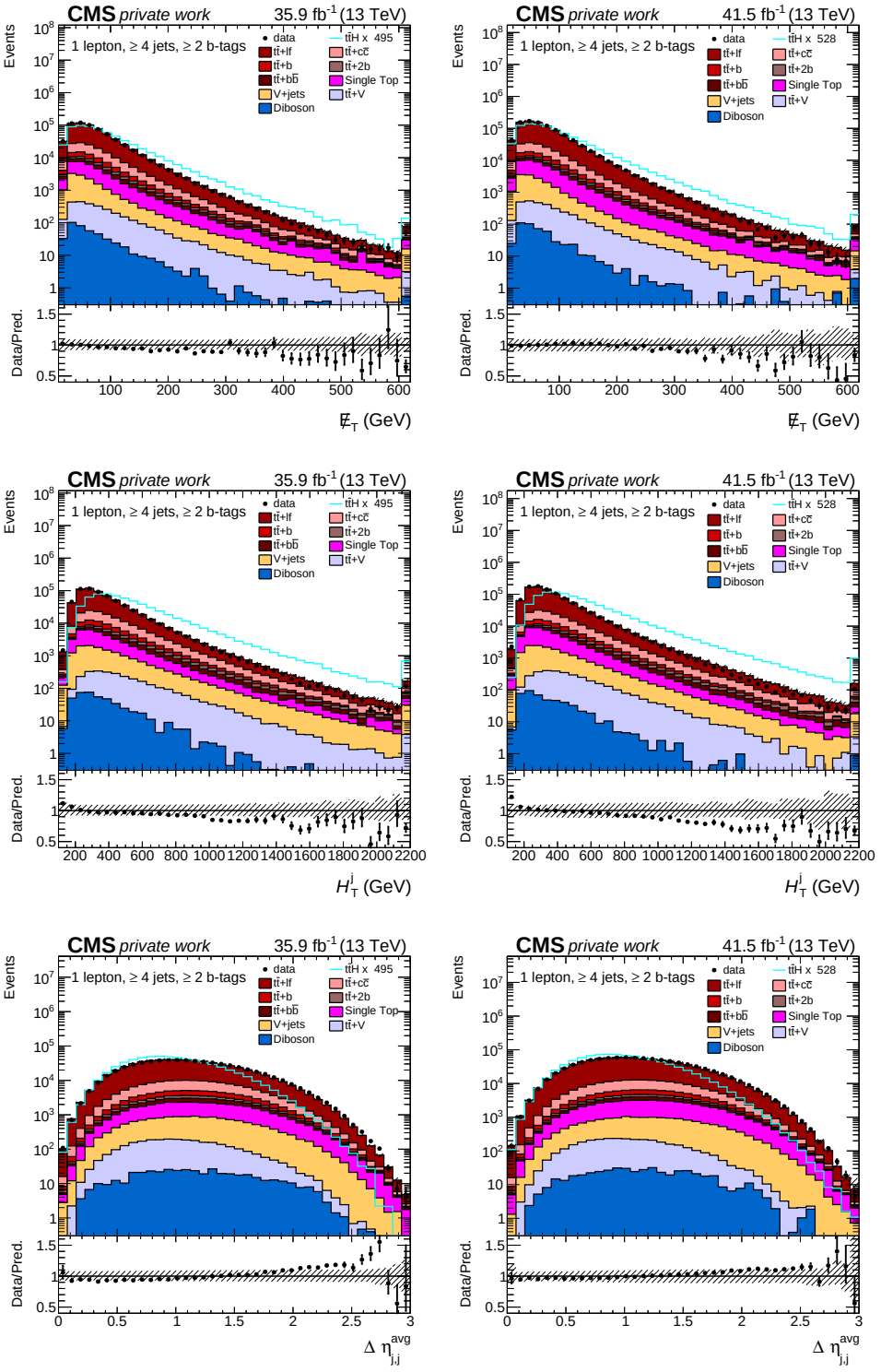


Figure 7.3: Missing transverse momentum (top), the scalar sum of the transverse momenta of all jets (middle) and the average $\Delta\eta$ between two jets (bottom) in data recorded in the years 2016 (left) and 2017 (right) after the baseline selection ($(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$). The distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $\mathrm{t\bar{t}H}$ process as blue line, which is additionally scaled to the integral of the background for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding those affecting only the rate. The bottom panels show the ratio of data to the background prediction.

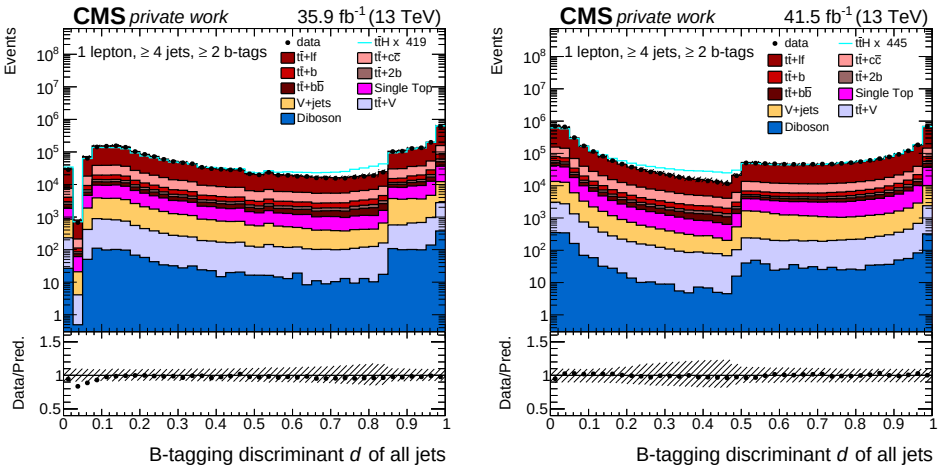


Figure 7.4: B-tagging discriminant value for jets in data recorded in the years 2016 (left) and 2017 (right) after the baseline selection. The analysis of the data recorded in 2016 uses the CSVv2 algorithm while for the data recorded in 2017 the DeepCSV algorithm is used. The distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $t\bar{t}H$ process as blue line which is additionally scaled to the integral of the background for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding those affecting only the rate. The bottom panels show the ratio of data to the background prediction.

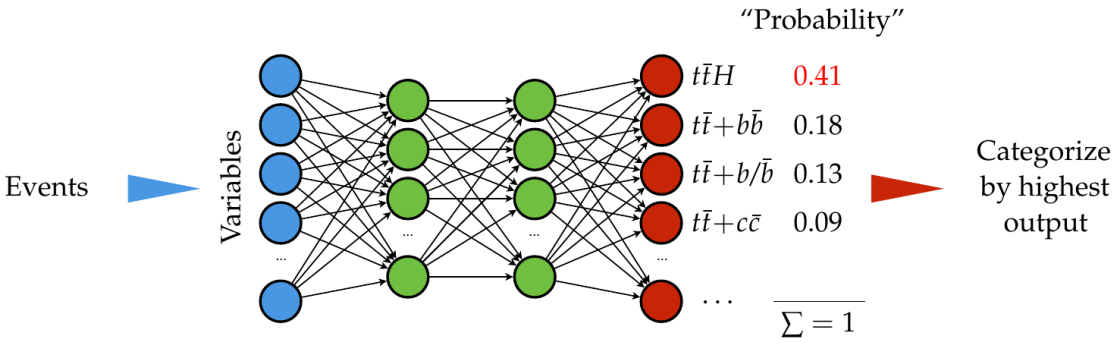


Figure 7.5: Illustration of the ANN strategy. Events are classified by the ANN based on their input features (Variables). The events are assigned six values that correspond to the probabilities to belong to the $t\bar{t}H$ or $t\bar{t}$ classes. They are subsequently classified as belonging to the class for which this output takes on the highest value. Taken from Ref. [248].

to mitigate overtraining of certain classes due to a too small number of available events. During the training, the samples are further split into three parts, the training sample, for which the loss function is minimized, the validation sample, on which the result is evaluated after each training epoch, and the test sample, which is used to control the early stopping of the procedure. The events of each class that are used for training are reweighted to sum to unity, therefore assigning equal importance to each class during the ANN training.

Separate ANNs are trained for the analyses of data recorded in the years 2016 and 2017. In addition, to accommodate differences in event properties and reconstructable observables, the events are split into three categories based on the number of jets: ($4\,\text{jets}, \geq 3\,\text{b-tags}$), ($5\,\text{jets}, \geq 3\,\text{b-tags}$) and ($\geq 6\,\text{jets}, \geq 3\,\text{b-tags}$). For each of these categories a dedicated ANN is trained. The input features used for the different ANNs are discussed in the following section.

For the analysis of the data recorded in the year 2016, the ANNs are implemented using the TensorFlow package [175]. The training of each ANN is performed in two stages. In the first stage an ANN is trained to recognize whether a given event contains b quarks from the decays of the top quarks, the W bosons or additional b quarks. In the second stage, the network is extended by additional layers, with the first of these layers being connected to all layers of the first stage as well as the input features. The ANN is then trained to distinguish between the six processes as described above. For the hidden layers the ELU function is used as activation function while the softmax function is used as activation function in the final classification layer. To avoid overtraining, dropout and L2 regularization are used. As a minimization algorithm the ADAM optimizer with an initial learning rate of 10^{-4} is chosen, which is described in Ref. [249]. The training is performed using batches of 5000 events over a maximum of 500 epochs. Beginning after 50 epochs the training is stopped early if the performance on the test and validation samples differs by more than 2%. The architecture and hyperparameters of all three networks are optimized for the performance on the test sample and the chosen values are summarized in Tab. 7.5. The design of the ANN and studies motivating the two-stage approach are discussed in further detail in Ref. [248].

For the analysis of the data recorded in the year 2017 a simpler design with additional hidden layers instead of the two-stage approach is chosen. The ANNs are implemented using the Keras and Tensorflow packages which are discussed in Refs. [174, 175]. The activation functions are again ELU for the hidden layers and softmax for the classification layer. To avoid overtraining, dropout and L2 regularization are used and the chosen minimizer is the ADAM algorithm with an initial learning rate of 10^{-4}. The training is performed in batches of 5000 events with identical early stopping criteria as for the 2016 ANNs. The architecture and chosen hyperparameters are summarized in Tab. 7.6.

7.5.2 Input variables and validation

The ANNs achieve their separation power by combining the information of many different weakly separating input features and the correlations between them. Therefore, even observables that on a first glance seem very similar between the classes can aid in the separation if they are included in an ANN. The input features used for the ANNs in the $t\bar{t}H(b\bar{b})$ analyses are chosen based on a two stage approach.

In the first step, a list of all observables that should differ between $t\bar{t}H$ and the background processes is created. This includes, for example, observables describing angles between objects, masses and momenta of reconstructed objects such as candidates for the Higgs boson and variables such as the scalar sum of the transverse momenta of all jets (compare Fig 7.3). Then, other observables that are not directly separating but that can aid in

Table 7.5: Architecture and hyperparameters of the ANNs in the analysis categories with $(4\,\text{jets}, \geq 3\,\text{b-tags})$, $(5\,\text{jets}, \geq 3\,\text{b-tags})$ and $(\geq 6\,\text{jets}, \geq 3\,\text{b-tags})$ used for the analysis of the data recorded in the year 2016.

	$(4\,\text{jets}, \geq 3\,\text{b-tags})$	$(5\,\text{jets}, \geq 3\,\text{b-tags})$	$(\geq 6\,\text{jets}, \geq 3\,\text{b-tags})$
hidden layers (stage 1)	2	2	2
hidden layers (stage 2)	2	1	2
nodes per hidden layer (stage 1)	100	100	100
nodes per hidden layer (stage 2)	100	100	100
loss function	cross-entropy	cross-entropy	cross-entropy
dropout percentage	0.7	0.7	0.7
L2 regularisation	10^{-5}	10^{-5}	10^{-5}
batch size	5000	5000	5000
optimizer	Adam(10^{-4})	Adam(10^{-4})	Adam(10^{-4})
activation function (hidden)	ELU	ELU	ELU
activation function (classification)	softmax	softmax	softmax
early stopping percentage	2%	2%	2%
early stopping minimal epochs	50	50	50
number of input variables	36	44	42

Table 7.6: Architecture and hyperparameters of the ANNs in the analysis categories with $(4\,\text{jets}, \geq 3\,\text{b-tags})$, $(5\,\text{jets}, \geq 3\,\text{b-tags})$ and $(\geq 6\,\text{jets}, \geq 3\,\text{b-tags})$ used for the analysis of the data recorded in the year 2017.

	$(4\,\text{jets}, \geq 3\,\text{b-tags})$	$(5\,\text{jets}, \geq 3\,\text{b-tags})$	$(\geq 6\,\text{jets}, \geq 3\,\text{b-tags})$
hidden layers	3	3	3
nodes per hidden layer	100	100	100
loss function	cross-entropy	cross-entropy	cross-entropy
dropout percentage	0.5	0.5	0.5
L2 regularisation	10^{-5}	10^{-5}	10^{-5}
batch size	5000	5000	5000
optimizer	Adam(10^{-4})	Adam(10^{-4})	Adam(10^{-4})
activation function (hidden)	ELU	ELU	ELU
activation function (classification)	softmax	softmax	softmax
early stopping percentage	2%	2%	2%
early stopping minimal epochs	50	50	50
number of input variables	14	15	11

combination with already included observables are added. For the analysis of the data recorded in the year 2017, a preselection of input features is performed. For this, the ANNs are trained with all chosen features. Then for each feature the total sum of the magnitudes of the weights that are connected to the feature in the first layer is calculated. This is interpreted as a value of the importance of the feature, with more important features leading to larger weights. Only the 20 features with the largest sum of weights are considered further. It was observed that the performance of the ANNs does not decrease by a large amount when removing the other features deemed less important by the method.

In the second step, all observables that are not sufficiently well described by simulated data are discarded. This is necessary since the ANNs pick up on even small features of the training data, which – if not correctly modeled – can lead to large differences in the ANN response to simulated and recorded data.

A multitude of observables are considered, and ultimately chosen, as input features and a discussion of every single one of them would be outside the scope of this thesis. However, they can be grouped into a small number of classes that follow similar considerations:

- p_T and mass variables:
  Due to the large mass of the Higgs boson, objects originating from the $\mathrm{t\bar{t}H}$ process tend to have higher transverse momenta compared to those originating in $\mathrm{t\bar{t}}$ production or other backgrounds. To reflect this, the input features include the p_T and η of the four jets with the highest transverse momenta, the p_T and η of the lepton and the missing transverse momentum (to account for the neutrino). Several input features are constructed from the combination of such observables such as the scalar sum of the transverse momenta of all (b-tagged) jets $H_\mathrm{T}^{(b)}$ and the ratio of $H_\mathrm{T}^{(b)}$ to the total energy of all jets or the combined p_T of the pair of (b-tagged) jets closest to each other in ΔR.
  Several input features are constructed to account for the masses of the top quarks, the Higgs boson, the W bosons or the whole $\mathrm{t\bar{t}H}$ system. These include the total invariant mass of all jets, the lepton and the missing transverse energy, as well as the invariant mass of the trijet system with the highest transverse energy. In addition, the average invariant mass of all (b-tagged) jet pairs is included. The average invariant mass of all b-tagged jet pairs in ($4\,\text{jets}, \geq 3\,\text{b-tags}$) events is shown in Fig. 7.6a.
  The invariant mass of the pair of the two b-tagged jets that are closest to each other and for the pair for which the invariant mass is closest to the Higgs boson mass is also considered. Another input feature stems from a reconstruction of the $\mathrm{t\bar{t}H}$ system. For this, the $\mathrm{t\bar{t}}$ system is reconstructed by finding the jet-to-particle assignment that minimizes a χ^2 cost function. If there are two unassigned b-tagged jets they are reconstructed as the Higgs boson and their invariant mass is taken as an input feature.

- angular variables:
  Several input features describe the minimal, maximal and average angles ΔR and $\Delta\eta$ between (b-tagged) jets or the jets and the lepton. Especially the angles between b-tagged jets are of interest since the angle between b jets coming from a Higgs boson tends to be smaller than for those coming from a gluon with a comparable momentum. In addition, the angles between non-tagged jets, b-tagged jets and the lepton are used since those particles tend to be closer to each other when originating from an energetic top quark. The distribution of the ΔR between the pair of b-tagged jets closest to each other in ($\geq 6\,\text{jets}, \geq 3\,\text{b-tags}$) events is shown in Fig. 7.6b.

- b-tagging:
  Since $\mathrm{t\bar{t}H(b\bar{b})}$ events contain more b jets than most background events several b-

tagging related input features are chosen. These include the b-tagging discriminant values of the four jets with the highest transverse momenta, the four highest b-tagging discriminant values of all jets, as well as the average discriminant value of all (b-tagged) jets. This is complemented by a b-tagging likelihood ratio that is designed to discriminate between events with four b jets and events with only two b jets and that is further described in Ref. [129]. The b-tagging likelihood ratio is shown for (≥ 6 jets, ≥ 3 b-tags) events in Fig. 7.6c.

- event shape:
  These variables describe the overall shape of the events. The sphericity S, transverse sphericity S_T and aplanarity A are calculated from the sphericity tensor as described in Ref. [250]. It is a 3×3 matrix with eigenvalues λ_1, λ_2 and λ_3 with $\lambda_1 > \lambda_2 > \lambda_3$ and is defined as:
  $$S_{\alpha\beta} = \frac{\sum_\mathrm{jets} p^i_\alpha p^i_\beta}{\sum_\mathrm{jets} |p^i|^2}, \tag{7.1}$$
  where the sum is performed over all jets i and α and β denote the x, y and z components of the jet momenta. The sphericity S and aplanarity A are calculated, as described in Ref. [251], as $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ and $A = \frac{3}{2}\lambda_3$ respectively. In addition, a transverse sphericity S_T is defined as $\frac{2\lambda_2}{\lambda_2+\lambda_1}$ as described in Ref. [252]. As their name suggests they describe how spherical or planar a given event is. The sphericity has a value of one for isotropic events and the aplanarity has a value of zero for planar events. The sphericity is exemplarily shown for (5 jets, ≥ 3 b-tags) events in Fig. 7.6d. The transverse sphericity describes how spherical or "pencil-shaped" an event is and has a value of zero for pencil-shaped events. Due to the large mass of the $\mathrm{t\bar{t}H}$ system and the large number of resulting jets, $\mathrm{t\bar{t}H}$ events tend to be more spherical and aplanar compared to $\mathrm{t\bar{t}}$ events.
  The observable C is defined as the ratio of the scalar sum of the transverse momenta p_T of all jets to the sum of the energies E of all jets:
  $$C = \frac{\sum_\mathrm{jets} p^i_\mathrm{T}}{\sum_\mathrm{jets} E^i}. \tag{7.2}$$
  The first Fox-Wolfram moments H_0, H_1, H_2, H_3 and H_4 are used as possible input features. They are defined in Refs. [253, 254] and describe on how many axis the energy in an event flows.

- Matrix element method:
  An important and strongly discriminating feature is the so-called matrix-element-method (MEM) discriminant which takes into account the full leading-order parton-level dynamics of the $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$ processes. A detailed description of the method is given in Refs. [4, 129]. The discriminating feature, simply called MEM in this thesis, is defined as the ratio of probability density values for a hypothesis $\mathcal{H}$ of the event being a $\mathrm{t\bar{t}H(b\bar{b})}$ or $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$ event:
  $$\mathrm{MEM} = \frac{p(\vec{y}|\mathrm{t\bar{t}H(b\bar{b})})}{p(\vec{y}|\mathrm{t\bar{t}H(b\bar{b})}) + 0.1 p(\vec{y}|\mathrm{t\bar{t}}+\mathrm{b\bar{b}})} \tag{7.3}$$
  where the $\vec{y}$ are the four vectors and b-tagging discriminant values of the objects in the event and the numerical factor of 0.1 was found empirically well suited for the discrimination task at hand.
  The probability density values $p(\vec{y}|\mathcal{H})$ are calculated as the sum of cross sections for different assignments of observed jets to the final-state particles of the hypothesized process $\mathcal{H}$
  $$p(\vec{y}|\mathcal{H}) = \sum_\mathrm{permutations} \sigma_{\mathrm{pp}\to\mathcal{H}\to\mathrm{f}} \tag{7.4}$$

The cross sections are calculated based on the leading-order matrix element of the hypothesized process, where the production via gluon-gluon fusion is assumed and where the partonic scattering amplitude is provided by OpenLoops [68, 69]. Particles in the hypothesized final state are assigned to jets or leptons in the event, with the restriction that the four b quarks have to be assigned to the four most b-jet-like jets (as found using the b-tagging likelihood ratio mentioned above). For each possible assignment, the cross section is calculated, integrating over unassigned particles. For the assignment of the particles to jets it is assumed that the jet four-vector corresponds to the particle four-vector within the jet energy resolution, over which an integration is performed. The probability density value $p(\vec{y}|\mathcal{H})$ is then the squared sum of the squared scattering amplitudes of all assignments. The MEM discriminant is itself an observable that strongly separates between $t\bar{t}H(b\bar{b})$ and $t\bar{t}+b\bar{b}$ events. It is shown for events recorded in the year 2017 in Fig. 7.7.

All input features have to be sufficiently well modeled to be used in the ANN training. This is ensured by a statistical validation procedure. For each input feature and for each pair of input features, the respective one-dimensional or two-dimensional distributions are constructed from simulated and recorded data. Then, for each distribution a goodness-of-fit test is performed by fitting the distributions of simulated data to those of recorded data. The goodness-of-fit tests are based on the saturated-model approach, which is described in Ref. [255] and implemented in the CMS Higgs combination software as described in Ref. [256]. For the fits the full fit model with all uncertainties as described in Ch. 8 is used and the signal strength modifier is fixed to one. For the calculation of the p-values, the distributions of the test statistics are constructed using pseudodata sampled from the post-fit probability density functions of the nuisance parameters. If the p-value of any of the fits involving a specific variable is below 0.05 the variable is discarded. The validation is performed separately for each year and for each of the categories defined by the multiplicities of jets and b-tagged jets.

The final list of all chosen input features used in the ANN training is given in Tables 7.7 and 7.8 and their distributions are shown in Appendix B.

7.6 Final discriminants

The discriminant distributions that result from the ANN trainings are depicted in Appendix E. The discriminant distributions for events with at least six jets recorded in the year 2017 are also shown in Fig. 7.8. Each of the distributions defined by an ANN node is compared to the other distributions, enriched in the process that the node targeted. In the $t\bar{t}H$-nodes the discrimination between $t\bar{t}H$ and the background processes is clearly visible. The discrimination between $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ is worse than the discrimination between $t\bar{t}H$ and the other background processes. The distributions of simulated data describe the recorded data within their uncertainties. The numbers of events that are assigned to each node are tabulated in Appendix E.

7.7 Summary of analyses in the DL and FH channels

The general strategies of the analyses in the DL and FH channels are similar to the one for SL channel. However, owing to the differences in the targeted final states, each of these analyses presents unique challenges that lead to differences compared to the SL channel analyses. The analyses of data recorded in the years 2016 and 2017 in the DL channel are described in more detail in Refs. [4, 5] and the ones in the FH channel in Refs. [5, 229]. Nevertheless, since the results presented in Ch. 9 are based on the combination of the SL channel with the other two channels, the analyses will be briefly summarized in the following.

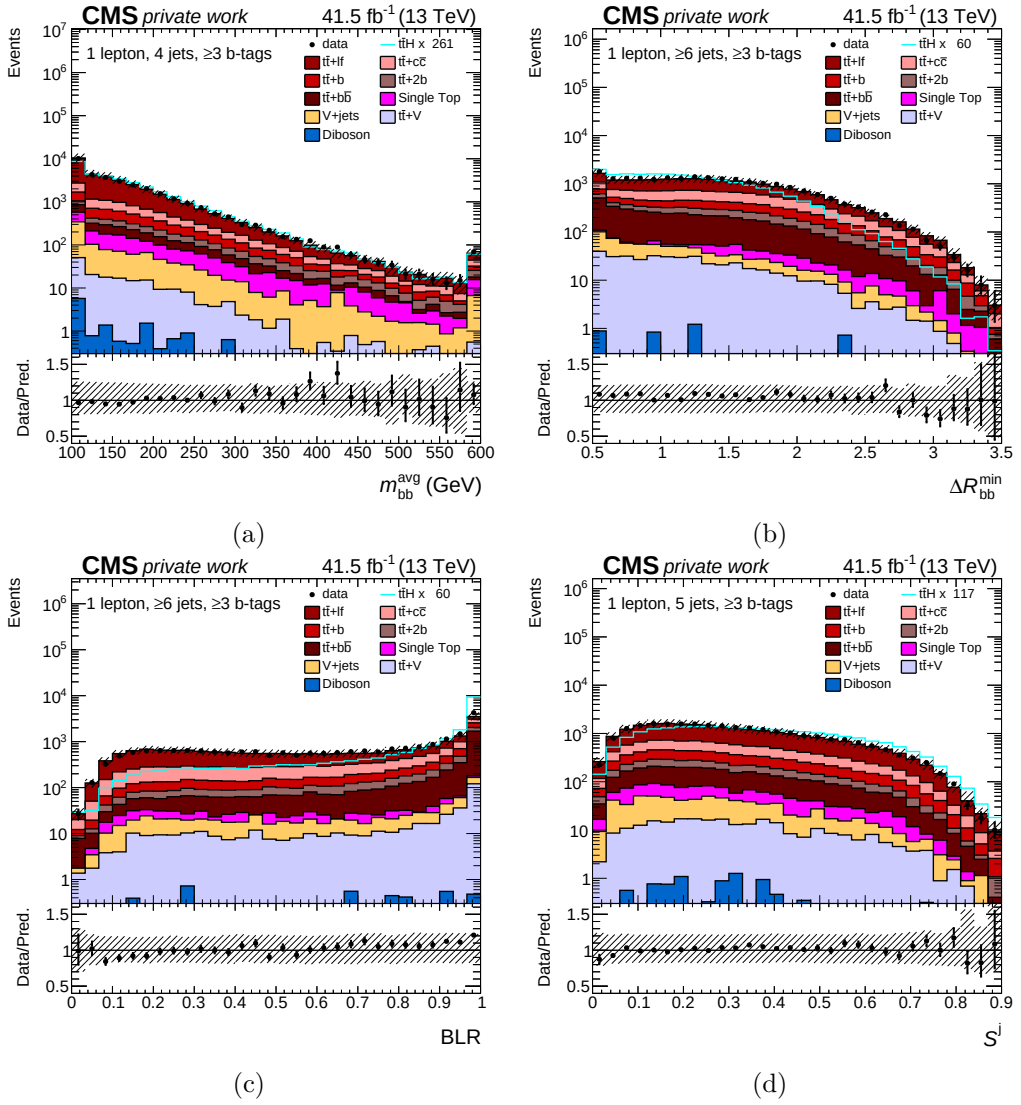


Figure 7.6: Exemplary input features used in the ANNs for the analysis of data recorded in the year 2017. Shown are (a) the average dijet mass of b-tagged jets in (4 jets, $\geq$ 3 b-tags) events, (b) the ΔR between the pair of b-tagged jets closest to each other in ($\geq$ 6 jets, $\geq$ 3 b-tags) events, (c) the b-tagging likelihood ratio in ($\geq$ 6 jets, $\geq$ 3 b-tags) events and (d) the sphericity in (5 jets, $\geq$ 3 b-tags) events. The distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $t\bar{t}H$ process as blue line which is additionally scaled to the integral of the background for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding those affecting only the rate. The bottom panels show the ratio of data to the background prediction.

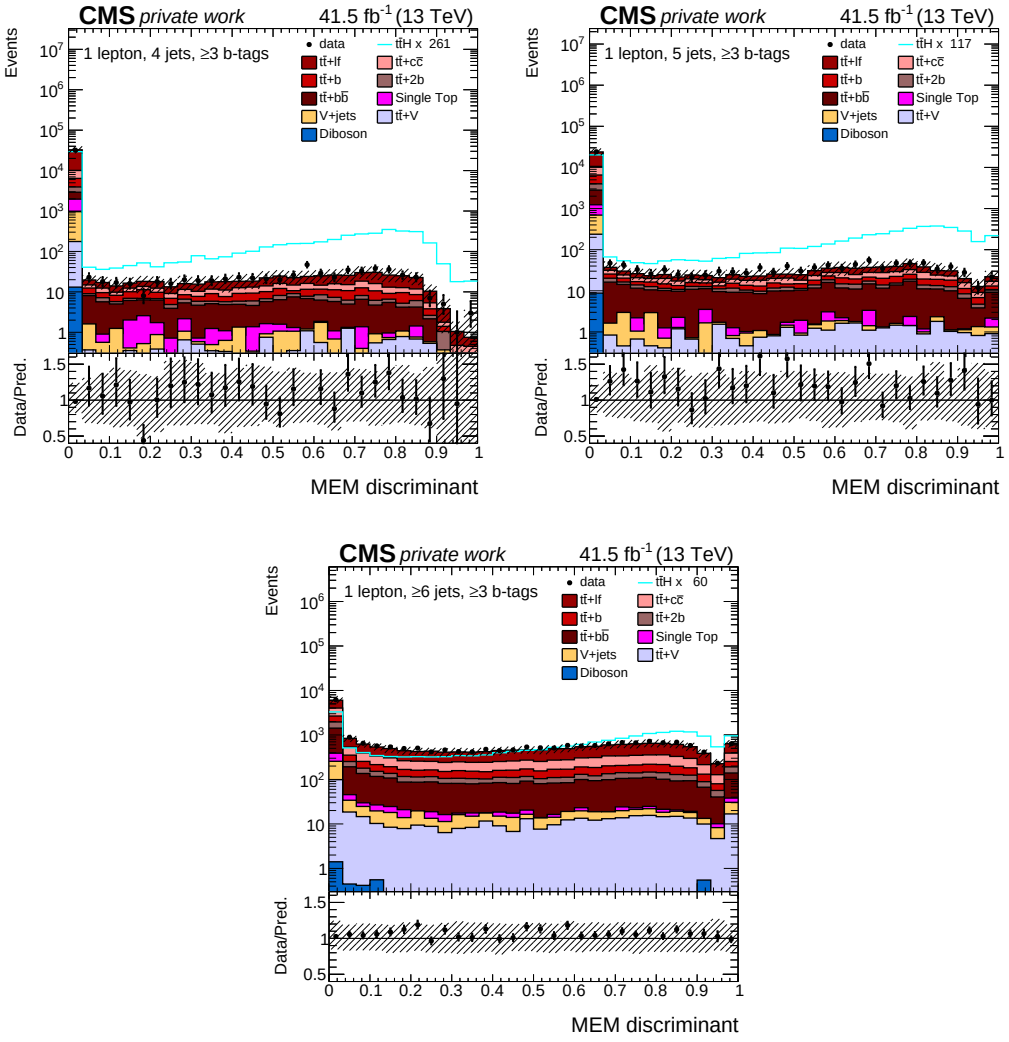


Figure 7.7: Matrix element method (MEM) discriminant in the $(4\,\text{jets}, \geq 3\,\text{b-tags})$, $(5\,\text{jets}, \geq 3\,\text{b-tags})$ and $(\geq 6\,\text{jets}, \geq 3\,\text{b-tags})$ categories for events recorded in the year 2017. The distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $\mathrm{t\bar{t}H}$ process as blue line which is additionally scaled to the integral of the background for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding those affecting only the rate. The bottom panels show the ratio of data to the background prediction.

Table 7.7: Input variables used for the training of the ANNs. Variables used in a specific analysis category are denoted by a "+" and unused variables by a "−". (Continued in Tab. 7.8.)

		2016			2017		
Variable	Definition	(4 jets, ≥ 3 b-tags)	(5 jets, ≥ 3 b-tags)	(≥ 6 jets, ≥ 3 b-tags)	(4 jets, ≥ 3 b-tags)	(5 jets, ≥ 3 b-tags)	(≥ 6 jets, ≥ 3 b-tags)
p_{T}(jet 1)	p_{T} of the highest-p_{T} jet	+	+	-	-	+	-
η(jet 1)	η of the highest-p_{T} jet	-	+	+	-	-	-
d(jet 1)	b-tagging discriminant of the highest-p_{T} jet	+	+	+	-	-	-
p_{T}(jet 2)	p_{T} of the second highest-p_{T} jet	-	+	-	-	-	-
η(jet 2)	η of the second highest-p_{T} jet	+	+	+	-	-	-
d(jet 2)	b-tagging discriminant of the second highest-p_{T} jet	+	+	+	-	-	-
p_{T}(jet 3)	p_{T} of the third highest-p_{T} jet	-	+	-	-	+	-
η(jet 3)	η of the third highest-p_{T} jet	+	+	+	-	-	-
d(jet 3)	b-tagging discriminant of the third highest-p_{T} jet	+	+	+	-	-	-
p_{T}(jet 4)	p_{T} of the fourth highest-p_{T} jet	+	+	-	-	-	-
η(jet 4)	η of the fourth highest-p_{T} jet	+	+	+	-	-	-
d(jet 4)	b-tagging discriminant of the fourth highest-p_{T} jet	+	-	+	+	-	-
p_{T}(lep)	p_{T} of the lepton	-	+	+	-	-	-
η(lep)	η of the lepton	+	-	+	-	-	-
$d_{\mathrm{j}}^{\mathrm{avg}}$	average b-tagging discriminant value of all jets	+	+	+	+	+	+
$d_{\mathrm{b}}^{\mathrm{avg}}$	average b-tagging discriminant value of b-tagged jets	+	+	+	+	+	+
$\frac{1}{N}\sum_b \left(d - d_{\mathrm{b}}^{\mathrm{avg}}\right)^2$	mean squared deviation of the b-tagging discriminant values of b-tagged jets to the average b-tagging discriminant value of all b-tagged jets	+	+	+	+	-	+
$d_{\mathrm{j}}^{\mathrm{max}}$	maximal b-tagging discriminant value of all jets	+	+	+	-	-	-
$d_{\mathrm{b}}^{\mathrm{max}}$	maximal b-tagging discriminant value of b-tagged jets	+	+	+	-	-	-
$d_{\mathrm{j}}^{\mathrm{min}}$	minimal b-tagging discriminant value of all jets	+	+	+	-	-	-
$d_{\mathrm{b}}^{\mathrm{min}}$	minimal b-tagging discriminant value of b-tagged jets	+	+	+	+	+	-
d_2	second highest b-tagging discriminant value of all jets	+	+	+	+	+	+
N_{b}(tight)	number of b-tagged jets at the tight b-tagging working point with 0.1% mistag rate	+	+	+	+	+	-
BLR	likelihood ratio discriminating between 4 b quark jets and 2 b quark jets events	+	+	+	+	-	+
$\mathrm{BLR}^{\mathrm{trans}}$	transformed BLR defined as ln[BLR/(1.0 − BLR)]	+	+	+	+	-	+

Table 7.8: Input variables used for the training of the ANNs. Variables used in a specific analysis category are denoted by a "+" and unused variables by a "−". (Continued from Tab 7.7.)

		2016			2017		
Variable	Definition	(4 jets, ≥ 3 b-tags)	(5 jets, ≥ 3 b-tags)	(≥ 6 jets, ≥ 3 b-tags)	(4 jets, ≥ 3 b-tags)	(5 jets, ≥ 3 b-tags)	(≥ 6 jets, ≥ 3 b-tags)
$\Delta R^{\text{min}}_{\text{j,j}}$	minimal ΔR between any two jets	+	+	+	-	-	-
$\Delta R^{\text{min}}_{\text{b,b}}$	minimal ΔR between any two b-tagged jets	+	+	+	-	-	+
$\Delta R^{\text{max}}_{\text{j,j}}$	largest ΔR between any two jets	-	+	-	-	+	-
$\Delta R^{\text{max}}_{\text{b,b}}$	largest ΔR between any two b-tagged jets	-	-	+	-	-	-
$\Delta\eta^{\text{avg}}_{\text{b,b}}$	average $\Delta\eta$ between b-tagged jets	-	-	+	-	-	-
$\Delta R^{\text{avg}}_{\text{b,b}}$	average ΔR between b-tagged jets	-	+	+	-	+	+
$\Delta R^{\text{min}}_{\text{lep,j}}$	minimal ΔR between lepton and any jet	+	+	-	-	-	-
$\Delta R^{\text{min}}_{\text{lep,b}}$	minimal ΔR between lepton and any b-tagged jet	-	+	+	-	+	-
$m^{\text{min}\Delta R}_{\text{lep,b}}$	mass of lepton and b-tagged jet closest in ΔR	+	+	+	-	-	+
$m^{\text{min}\Delta R}_{\text{b,b}}$	mass of b-tagged jets closest in ΔR	+	+	+	-	+	-
$m^{\text{avg}}_{b,b}$	average mass of all pairs of b-tagged jets	-	-	-	+	-	-
$m^{\text{avg}}_{\text{j}}$	average mass of all jets	+	+	+	-	-	+
$(m^2)^{\text{avg}}_{\text{b}}$	average squared mass of all b-tagged jets	+	-	+	-	-	-
$m^{\text{closest to 125}}_{\text{b,b}}$	mass of pair of b-tagged jets closest to 125GeV	-	+	+	-	-	-
MEM	matrix element method discriminant	+	+	+	+	+	+
H^{j}_{T}	scalar sum of jet p_{T}	-	+	-	-	-	-
H^{b}_{T}	scalar sum of p_{T} of all b-tagged jets	+	+	+	+	+	+
C^{j}	H^{j}_{T} divided by the sum of the energies of all jets	-	-	+	-	-	-
C^{b}	H^{b}_{T} divided by the sum of the energies of all b-tagged jets	-	-	+	-	-	-
A^{j}	$\frac{3}{2}\lambda_3$ where λ_i are the eigenvalues of the momentum tensor built with jets [250, 251]	-	+	+	-	-	-
A^{b}	$\frac{3}{2}\lambda_3$ where λ_i are the eigenvalues of the momentum tensor built with b-tagged jets [250, 251]	+	+	+	-	-	-
S^{j}	$\frac{3}{2}(\lambda_2+\lambda_3)$ where λ_i are the eigenvalues of the momentum tensor built with jets [250, 251]	+	+	+	+	+	-
S^{b}	$\frac{3}{2}(\lambda_2+\lambda_3)$ where λ_i are the eigenvalues of the momentum tensor built with b-tagged jets [250, 251]	-	+	+	-	+	-
S^{j}_{T}	$\frac{2\lambda_2}{\lambda_2+\lambda_1}$ where λ_i are the eigenvalues of the momentum tensor built with jets [250, 252]	+	+	+	+	-	-
S^{b}_{T}	$\frac{2\lambda_2}{\lambda_2+\lambda_1}$ where λ_i are the eigenvalues of the momentum tensor built with b-tagged jets [250, 252]	+	+	+	-	-	-
H_2	second Fox–Wolfram moment [253, 254]	-	+	-	-	-	-
H_3	third Fox–Wolfram moment [253, 254]	+	+	-	-	-	-
H_4	fourth Fox–Wolfram moment [253, 254]	+	-	+	-	-	-

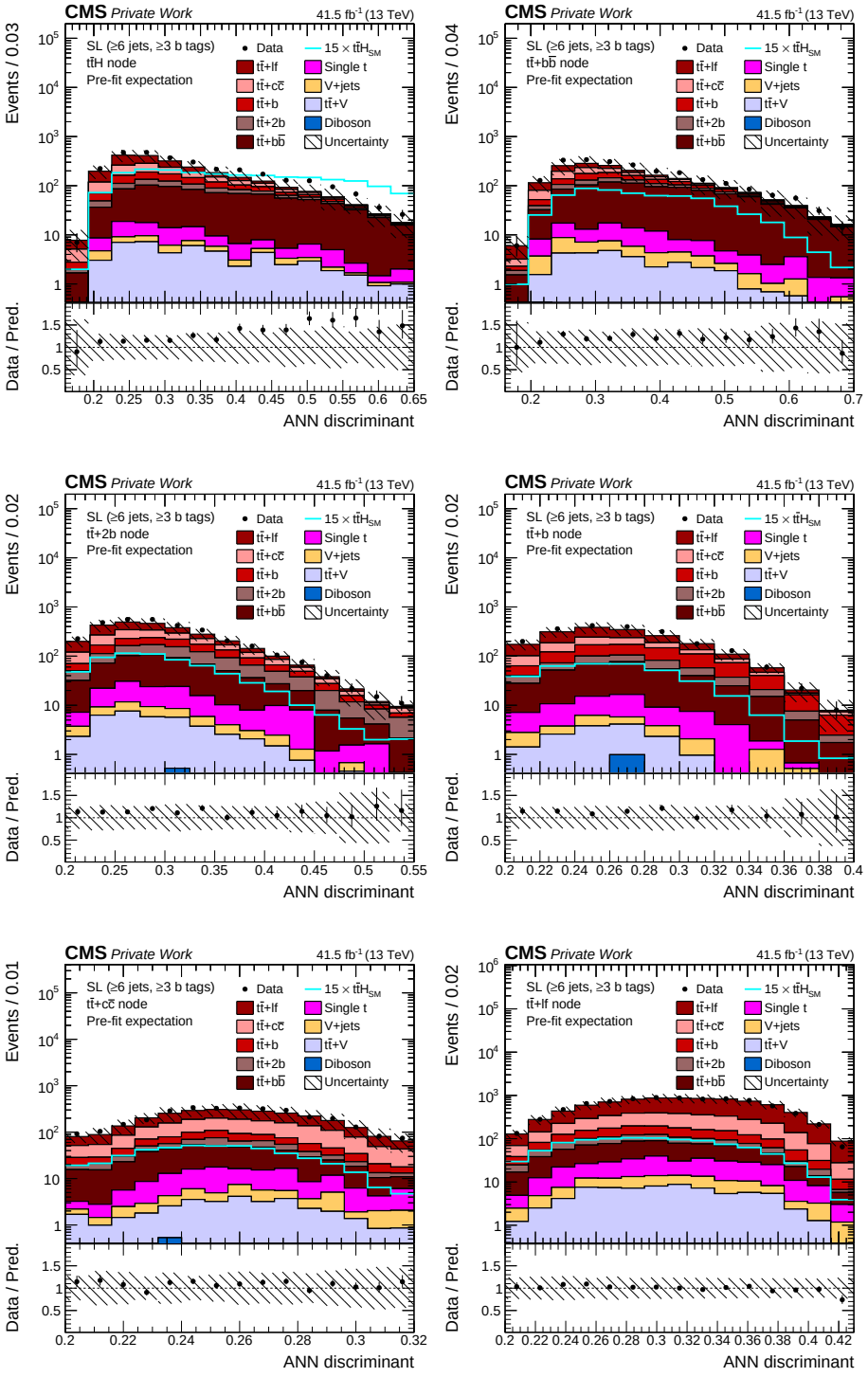


Figure 7.8: ANN discriminant distributions for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2017 before the fit to data. The distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $\mathrm{t\bar{t}H}$ process as blue lines and which are additionally scaled by a factor of 15 for better visibility. The hatched uncertainty bands show the total uncertainty. The bottom panels show the ratio of data to the background prediction.

Analyses in the DL channel

In the DL channel, selected events have to be accepted by triggers that require the presence of either two electrons, two muons or one electron and one muon. Next, events have to contain exactly two well-reconstructed and isolated leptons with large transverse momenta, the invariant mass of which has to be sufficiently different from the Z-boson mass in case of two same-flavor leptons. In addition, events recorded in the year 2016 are required to contain at least 4 jets of which at least three are b tagged, while events recorded in the year 2017 are required to contain at least three jets of which at least two have to be b tagged. Although the requirement of the second lepton leads to a sample with a higher signal purity it also leads to a much smaller number of selected events compared to the SL channel. This results in a larger impact of the statistical uncertainties and also to smaller samples of events available for the training of the multivariate methods [1].

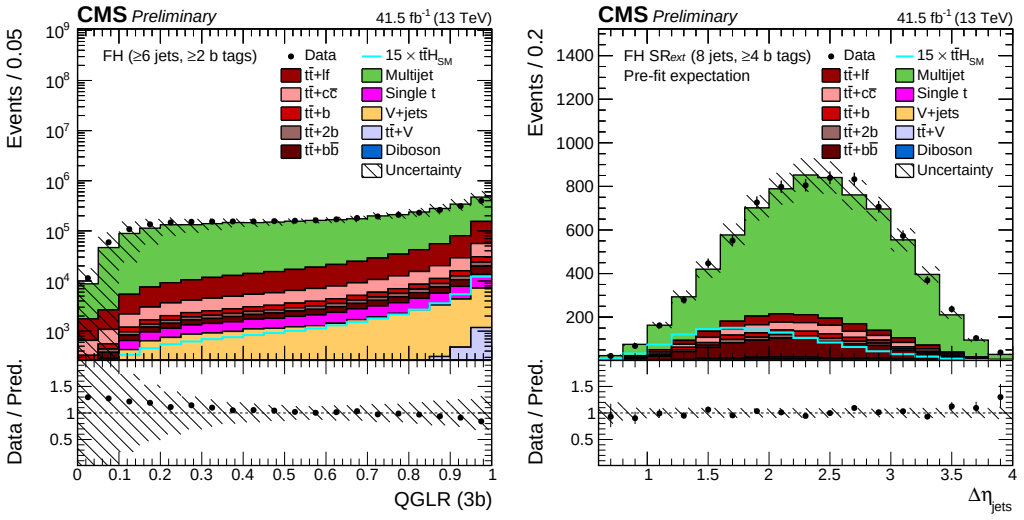
For the analysis of the data recorded in the year 2016, boosted decision trees (BDTs) are trained to separate between $t\bar{t}H$ and all of $t\bar{t}$. For events with exactly 3 b-tagged jets, the final discriminant that is fitted to data is the output of the BDT, while events with more than three b-tagged jets are sub-categorized by a BDT into a signal-like and a background-like sub-category in which the MEM is used as final discriminant. For the analysis of data recorded in the year 2017, selected events are first sorted into five categories according to their jet and b-tagged jet multiplicities. Then in each category a BDT is trained so separate $t\bar{t}H$ from $t\bar{t}$ events, with the MEM discriminant used as input feature for the BDTs. This approach was chosen from several studied approaches based on the expected results. ANNs were observed to not perform better than BDTs while being more susceptible to overtraining.

Analyses in the FH channel

For events to be selected into the FH channel, they have to be accepted by a combination of triggers that target events with a large amount of H_{T} and at least four jets, of which several should to be b tagged. Events are not allowed to contain any well-reconstructed leptons that could come from the decays of top quarks. In addition, all selected events are required to contain at least seven jets of which at least three have to be b tagged. To only select events that likely contain top quarks, at least one pair of untagged jets has to have an invariant mass compatible with the W-boson mass.

The fully-hadronic $t\bar{t}$ decay has the largest branching fraction of the three decay channels, leading to a large number of selected $t\bar{t}H$ events. Unfortunately the requirement that no leptons are to be present in the events also leads to a large background from QCD multijet production. To reduce the QCD multijet background, a quark-gluon-likelihood-ratio (QGLR) discriminant is constructed that distinguishes between events containing light-flavor jets originating from W-boson decays and those where the light-flavor jets originate from gluons. It is based on a quark-gluon-likelihood function that calculates the likelihood for a specific jet to either be a quark or a gluon jet based on the differences in particle multiplicity, collimation and fragmentation between quark and gluon jets [257, 258]. The QGLR is then constructed as the ratio of quark-gluon-likelihood functions under the hypothesis that N light flavor jets in the event originate from quarks or gluons, where N depends on the number of jets and b-tagged jets in the event. It is shown for events recorded in the year 2017 and with at least six jets of which at least two are b-tagged jets in Fig. 7.9a. For events to be considered in the analysis, their QGLR value has to exceed 0.5.

[1] The smaller sample size available for training can in principal be avoided by creating more simulated events, which however is computationally costly.



(a) (b)

Figure 7.9: Observables used for the rejection of QCD multijet events in the FH channel. Shown are (a) the QGLR calculated for events with at least six jets of which at least two are b tagged and not including the three b-tagged jets with the highest b-tagging discriminant values in the calculation, and (b) maximal $\Delta\eta$ of each jet to any other jet in the event, averaged over all jets in the event, calculated for events with eight jets of which at least 4 are b tagged. The distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $t\bar{t}H$ process as blue line which is additionally scaled by a factor of 15 for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding the additional 50% rate uncertainty on $t\bar{t}$ + heavy flavor discussed in Ch. 8. The bottom panels show the ratio of data to the background prediction. Taken from Ref. [5].

For the analysis of data recorded in the year 2017, events from QCD multijet production are further suppressed by an additional selection on the maximal $\Delta\eta$ of each jet to any other jet in the event, averaged over all jets in the event. The distribution of this quantity is shown in Fig. 7.9b where it can be seen that in QCD multijet events the jets tend to be separated further from each other in η than in $t\bar{t}H$ events. After this selection steps the dominant backgrounds are a still sizable number of events from QCD multijet production followed by $t\bar{t}$ production.

Next, the selected events are categorized according to their jet and b-tagged jet multiplicities and in each category the MEM discriminant is used as final discriminant. Although the MEM is constructed to distinguish between $t\bar{t}H(b\bar{b})$ and $t\bar{t}$+$b\bar{b}$, it was found to also be able to separate $t\bar{t}H(b\bar{b})$ from $t\bar{t}$ + light flavor and QCD multijet production.

With the exception of the QCD multijet background, the background contribution is predicted from simulated data. QCD multijet production is however notoriously difficult to predict using simulated data. Instead the MEM distributions for QCD multijet production are predicted from control regions defined by exactly two b-tagged jets at the medium working point and at least three b-tagged jets at a looser working point. The normalization of the QCD multijet distributions is left unconstrained in the final fit to data.

8 Systematic Uncertainties

The analyses discussed in this thesis are based on experimental measurements and are therefore subject to systematic uncertainties. Their correct estimation and treatment are crucial for the validity of the final result.
As was discussed in Ch. 4, in the binned maximum likelihood approach that is used to fit the predicted ANN distributions to those derived on data, sources of systematic uncertainties are treated as constrained nuisance parameters with normal or log-normal probability density functions corresponding to the uncertainty with which they are known.

In this scheme, systematic uncertainties are implemented either as *rate* or as *shape* uncertainty. Rate uncertainties are used for systematic uncertainty sources that affect the overall predicted number of events of one or more processes, such as cross sections or integrated luminosities. Uncertainty values for rate uncertainties discussed in this chapter correspond to one standard deviation estimates and are mapped to the one standard deviation quantiles of the probability density functions of their respective nuisance parameters.
Shape uncertainties are used for systematic uncertainties that depend on the properties of the individual events and affect each bin of the ANN distributions differently. They are implemented using additional template histograms that represent the ANN shapes if the uncertainty source in question would be one standard deviation larger or smaller than for the nominal prediction. These templates are then mapped to the one standard deviation quantiles of the probability density functions of their respective nuisance parameters. The distributions of simulated data can then be adjusted by the fit by interpolating between the nominal distributions and the uncertainty templates. In most cases the additional template histograms are constructed using event weights that change the predicted rate of specific events, whereas the uncertainties affecting the jet energy calibration affect the momenta of the jets themselves, prompting a complete reanalysis of the data.

All systematic uncertainties can be classified into either experimental uncertainties that reflect the uncertainties associated with the measurement of the events in the CMS detector or modeling uncertainties associated with the simulation of the simulated data. A list of all systematic uncertainties is given in Tab. 8.1. The individual sources of experimental and modeling uncertainties are discussed in Sec. 8.1 and Sec. 8.2 respectively. Several systematic uncertainty sources are shared between the analyses of the data recorded in the years 2016 and 2017. In such cases the systematic uncertainties are fully correlated when fitting both analyses together, as discussed in Sec. 8.3.

Table 8.1: Systematic uncertainties, their type and which processes are affected by them.

Source	Type	Remarks
Integrated luminosity	rate	Signal and all backgrounds
Lepton identification/isolation	shape	Signal and all backgrounds
Trigger efficiency	shape	Signal and all backgrounds
Trigger prefiring correction	rate	Signal and all backgrounds
Pileup	shape	Signal and all backgrounds
Jet energy scale	shape	Signal and all backgrounds
Jet energy resolution	shape	Signal and all backgrounds
b-tag HF	shape	Signal and all backgrounds
b-tag HF linear	shape	Signal and all backgrounds
b-tag HF quadratic	shape	Signal and all backgrounds
b-tag LF	shape	Signal and all backgrounds
b-tag LF linear	shape	Signal and all backgrounds
b-tag LF quadratic	shape	Signal and all backgrounds
b-tag C linear	shape	Signal and all backgrounds
b-tag C quadratic	shape	Signal and all backgrounds
Renorm./fact. scales ($t\bar{t}H$)	rate	Scale uncertainty of NLO $t\bar{t}H$ prediction
Renorm./fact. scales ($t\bar{t}$)	rate	Scale uncertainty of NNLO $t\bar{t}$ prediction
$t\bar{t}$ + heavy flavor cross sections	rate	Additional 50% rate uncertainty of $t\bar{t}$ + heavy flavor predictions
Renorm./fact. scales (t)	rate	Scale uncertainty of NLO single t prediction
Renorm./fact. scales (V)	rate	Scale uncertainty of NNLO W+jet and Z+jet prediction
Renorm./fact. scales (VV)	rate	Scale uncertainty of NLO diboson prediction
PDF (gg)	rate	PDF uncertainty for gg initiated processes except $t\bar{t}H$
PDF (gg $t\bar{t}H$)	rate	PDF uncertainty for $t\bar{t}H$
PDF ($q\bar{q}$)	rate	PDF uncertainty of $q\bar{q}$ initiated processes ($t\bar{t}$+W,W,Z)
PDF (qg)	rate	PDF uncertainty of qg initiated processes (single t)
PDF shape variations ($t\bar{t}H$, $t\bar{t}$)	shape	Based on the NNPDF variations, same for $t\bar{t}H$ and $t\bar{t}$ subprocesses
μ_R scale ($t\bar{t}$)	shape	Renormalization scale uncertainty of the $t\bar{t}$ matrix element generator (POWHEG), correlated between $t\bar{t}$ subprocesses
μ_F scale ($t\bar{t}$)	shape	Factorization scale uncertainty of the $t\bar{t}$ matrix element generator (POWHEG), correlated between $t\bar{t}$ subprocesses
PS scale: ISR ($t\bar{t}$)	shape	Initial-state radiation uncertainty of the PS for $t\bar{t}$ simulation, uncorrelated between $t\bar{t}$ subprocesses
PS scale: FSR ($t\bar{t}$)	shape	Final-state radiation uncertainty of the PS for $t\bar{t}$ simulation , uncorrelated between $t\bar{t}$ subprocesses
ME-PS matching ($t\bar{t}$)	rate	NLO ME to PS matching $t\bar{t}$ simulation, uncorrelated between $t\bar{t}$ subprocesses
Underlying event ($t\bar{t}$)	rate	Underlying event for $t\bar{t}$ simulation, uncorrelated between $t\bar{t}$ subprocesses
size of simulated samples	shape	Statistical uncertainty of the signal and background prediction due to the limited sample size

8.1 Experimental uncertainties

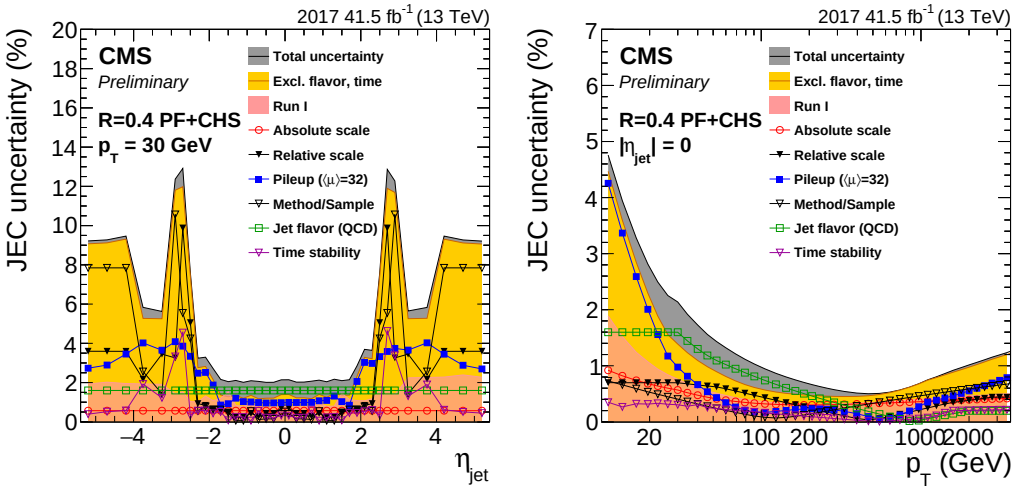
Jet energy calibration

The jet energy scale and resolution corrections, described in Sections 6.5.2 and 6.9.1, are derived using a plethora of samples of recorded and simulated data. Hence, the calibration corrections are subject to several uncertainties related to the size and purity of the selected samples, the reconstruction efficiencies of the jets and the modeling of the simulated data. For the analyses described in this thesis, one uncertainty on the jet energy resolution and 26 uncertainties on the jet energy scale are considered. All of these are implemented as shape changing uncertainties. For the creation of the corresponding template histograms all samples of simulated data are reprocessed with modified JER smearing or JES scaling factors. Therefore the momenta of all jets in a given event are changed with respect to the nominal JER and JES corrections. These changes are then propagated through the event selection, the calculation of the input features, including the MEM discriminant, and finally the ANNs.

For the JER uncertainty, the smearing of the simulated jets described in Sec. 6.9.1 is performed with different JER factors corresponding to the uncertainty of the resolution measured in data. Only a single JER uncertainty is used, meaning the uncertainties for different p_{T} and η regions are fully correlated.

The total JES uncertainty is split into 26 different uncertainties that either affect different η regions or correspond to different uncertainty sources. The usage of the different uncertainties is motivated by the observation that different uncertainties affect the analysis regions and processes differently. Since a full discussion of these uncertainties would exceed the scope of this thesis while a short description would not be illuminating, only their names will be listed in the following. A full description and discussion of each uncertainty is given in Ref. [218]. The sequences of letters "BB", "EC1", "EC2" and "HF" that can be found inside square brackets in the names of several uncertainties denote that the uncertainty affects one of several η regions: "BB" for $|\eta| < 1.3$, "EC1" for $1.3 < |\eta| < 2.5$, "EC2" for $2.5 < |\eta| < 3.0$ and "HF" for $|\eta| > 3.0$. While selected jets in the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses are required to have $\eta < 2.4$, the uncertainties for the regions with larger η can still affect the missing transverse momentum. The 26 JES uncertainties are:

- AbsoluteScale
- AbsoluteStat
- AbsoluteMPFbias
- Fragmentation
- SinglePionECAL
- SinglePionHCAL
- TimePtEta
- RelativeJER[EC1][EC2][HF]
- RelativeFSR
- RelativePt[BB][EC1][EC2][HF]
- RelativeBal
- RelativeStatFSR
- RelativeStat[EC][HF]



(a) (b)

Figure 8.1: JES uncertainties by source group and total uncertainty, calculated from the individual uncertainty sources by quadratic addition, as a function of jet η (left) and jet p_{T} (right) for data recorded in the year 2017. Also shown is the total uncertainty of Run 1 without the time and flavor dependent components. Taken from Ref. [219].

- PileUpDataMC
- PileUpPtRef
- PileUpPt[BB][EC1][EC2][HF]
- FlavorQCD

The magnitudes of the uncertainties as a function of p_{T} and η of jets are shown in Fig. 8.1. Here several uncertainties, such as for example all uncertainties related to the absolute scale part of the L2L3Residual correction, are combined by quadratic addition. Comparable figures for data recorded in the year 2016 are given in Ref. [219]. As can be seen, for the jets selected in this thesis the total uncertainty due to JES is about 1–2%.

b-tagging scale factors

As was discussed in Sec. 6.9.2, simulated data are corrected to account for an observed difference in the efficiency of the b-tagging algorithm compared to recorded data. The scale factors used for this correction are parameterized as functions of the p_{T}, η, b-tagging discriminant values and the true flavor of jets in the simulated events. They are derived based on samples enriched in b jets or udsg jets, as discussed in Sec. 6.9.2, and therefore affected by systematic uncertainties as described in Ref. [222]. These uncertainties are implemented in the form of additional scale factors which will be described in the following. Since they depend on the properties of individual jets in the events and will therefore change the shape of kinematic observables, their effects on the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses are taken into account by creating additional template histograms using these additional scale factors instead of the central ones.

During the derivation of the scale factors, contributions of udsg (b) jets to the b-tagging discriminant distributions for b (udsg) jets are estimated and subtracted using simulated data and are therefore subject to the uncertainties of these simulated data. To account for this, two systematic uncertainties are introduced, representing variations of the contamination

of either udsg jets in the b jet sample (*b-tag HF*) or of b jets in the udsg enriched sample (*b-tag LF*). They are estimated by increasing or decreasing the contamination by 20% and refitting the scale factors. Therefore, in this scheme, a $\pm 1\sigma$ variation of the corresponding nuisance parameter corresponds to a $\pm 20\%$ change in the predicted fraction of jets of the contaminating flavor. By construction, these uncertainties have the largest effect on the scale factors of the flavors of their respective enriched regions (e.g. b-tag HF on b jets). However, due to the iterative fitting procedure they also have a smaller effect on jets of the other flavor class.

The statistical uncertainty caused by the limited size of the selected samples is considered using four uncertainties, two each for the udsg jet and b jet scale factors. They account for statistical fluctuations that would either tilt the b-tagging discriminant distribution or change the number of events in the middle of the distribution compared to its edges. For their estimation a linear function $f_1(x) = 1 - 2x$ and a quadratic function $f_2(x) = 1 - 6x(1 - x)$ are introduced, parameterizing the tilting and bending of the discriminant distribution respectively. The initial scale factors are varied according to the statistical fluctuations in the bins of the discriminant distribution and the values of these functions. Finally, the scale factors are refitted resulting in two additional pairs of scale factors for each flavor class that envelop the statistical fluctuations around the nominal scale factors (*b-tag LF linear, b-tag LF quadratic, b-tag HF linear* and *b-tag HF quadratic*).

Since the b-tagging scale factors are derived as a function of the jet p_{T}, they naturally depend on the jet energy scale calibration. To account for this, additional scale factors are derived for each variation of the jet energy calibration discussed in the previous section. Whenever a JEC uncertainty is considered by variation of a jet four-momentum, for example during the creation of the ANN discriminant histograms for the JEC uncertainties, the corresponding b-tagging scale factor is applied instead of the central b-tagging scale factor.

The uncertainties affecting the b jet and udsg jet scale flavors are derived independently and treated as uncorrelated. Since for the c jets the central scale factor values are taken to be unity, additional larger uncertainties are estimated for these jets. For this, all uncertainties for b jets are doubled and added in quadrature. Then, the scale factors for c jets are calculated using a linear and a quadratic function similar to the procedure for the statistical uncertainties described above resulting in two additional uncorrelated uncertainties: *b-tag C linear* and *b-tag C quadratic*. However, instead of the statistical uncertainty in a given bin, the total uncertainty of the bin as calculated from the b-jet scale factors is used.

Lepton and trigger scale factors

As described in Sections 6.9.4 and 6.9.5, the scale factors for the correction of the lepton isolation, the lepton identification efficiencies as well as the trigger are calculated using tag-and-probe methods. They are affected by several uncertainties related to the sizes of the samples, the uncertainty on their selection efficiencies and the modeling of the simulated data as discussed in Refs. [164, 208]. For leptons used in the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses, the total uncertainties on each individual scale factor type are on the order of 1–2% and have a negligible effect on the result. Therefore, uncertainties of all electron and muon scale factors are treated as fully correlated for each lepton but uncorrelated for electrons and muons. To this end, for each individual scale factor its systematic and statistical uncertainties are added in quadrature. Then the scale factors are simultaneously varied in the same direction and template histograms are created where the varied scale factors are applied instead of the central ones. This results in one uncertainty for electrons (*electron identification*) and one for muons (*muon identification*).

The total uncertainties on the trigger scale factors are similarly small and found to be on the order of 1–3%. They are taken to be uncorrelated between the electron and muon

triggers, resulting in two uncertainties (*electron trigger* and *muon trigger*) and implemented using additional template histograms.

Since the scale factors of each event depend on the kinematics of the leptons, these uncertainties can in principle lead to changes in the shapes of observables. It is however found that the ANNs are rather insensitive to small changes in the lepton kinematics and therefore the shape changes are very small.

Pileup reweighting

The scale factors used to reweight the distribution of pileup interactions are calculated based on the inelastic proton-proton cross section as described in Sec. 6.9.3. The uncertainty on this cross section is estimated to be 4.6% and taken into account by recalculating the pileup scale factors with a cross section varied by this amount. The corresponding uncertainty is denoted simply as *pileup* and implemented by creating addition template histograms with the varied scale factors instead of the central ones.

L1 prefiring scale factors

For the analysis of data recorded in the year 2017 so-called L1 prefiring corrections are applied by decreasing the expected number of events based on the probabilities that events self-veto, as described in Sec. 6.9.6. The corresponding uncertainty is denoted *L1 prefire* and is implemented as a rate changing uncertainty. It is estimated by recalculating the change in the event yields with the probabilities varied within their uncertainties as measured on recorded data. The uncertainty is found to change the number of events by about 0.5–1% depending on the process and the analysis category.

Integrated luminosity measurement

All samples of simulated data are normalized to integrated luminosities of $35.9\,\mathrm{fb}^{-1}$ (2016) or $41.5\,\mathrm{fb}^{-1}$ (2017). The integrated luminosity is measured during the data taking as described in Refs. [148, 259] and is therefore subject to systematic and statistical uncertainties. For the year 2016 the total uncertainty on integrated luminosity is found to be 2.5% while for the year 2017 it is found to be 2.3%. This uncertainty is implemented as a rate changing uncertainty.

Size of simulated samples

The template histograms for simulated data are created by taking individual simulated events, evaluating their ANN outputs, reweighting them using all appropriate scale factors and then filling them into the respective bins of the histograms. Since the bins are filled with individual events and the production of an event that could be filled into a given bin is a probabilistic process, the number of unweighted events in each bin follows a Poisson distribution as for events in recorded data. For unweighted events (such as those in recorded data) the statistical uncertainty of a given bin is the square root of the variance, with the variance being equal to the expected value, which, for this purpose, is given by the bin content itself (the bin content is the expected number of events as given by the simulation):

$$\sigma_{\text{unweighted}} = \sqrt{\sum_i 1}. \tag{8.1}$$

Here the sum is performed over the events i in the bin. In case the events i are weighted by weights w_i the uncertainty caused by the statistical fluctuations of the simulated sample is instead given by

$$\sigma_{\text{weighted}} = \sqrt{\sum_i \mathrm{Var}(1 \cdot w_i)} = \sqrt{\sum_i w_i^2 \cdot \mathrm{Var}(1)} = \sqrt{\sum_i w_i^2} \tag{8.2}$$

with Var(x) being the variance of x. This reflects that the bin content predicted by a large number of events with small weights has a smaller uncertainty than the same bin content predicted by a small number of events with large weights. Since the number of accessible resources is finite, the number of simulated events is also finite and this statistical uncertainty of the simulated samples has to be taken into account. This uncertainty is especially important when the number of simulated events in a histogram bin is small and even more so if the cross section of the simulated process is large and only a small fraction of the events pass the event selection, as is for example the case for V+jets events.

In the $t\bar{t}H(b\bar{b})$ analyses this uncertainty is implemented based on the Barlow-Beeston method, which is introduced in Ref. [260]. For each bin in each ANN distribution and for each simulated process, an additional uncertainty is introduced that corresponds to the statistical uncertainty of this process in this bin. The corresponding nuisance parameters are constrained using a Poisson distribution with the nominal number of events as mean. These uncertainties are implemented by duplicating the nominal template histograms twice and increasing or decreasing the bin contents by the statistical uncertainties calculated with Eq. 8.2. This leads to a very large number of additional uncertainties: 4465 and 6460 for the analyses of the data recorded in the years 2016 and 2017 respectively.
Since for many bins these uncertainties are negligible and such a large number of uncertainties leads to instabilities of the fit and excessive runtimes, a simplification method is used that merges many of the uncertainties as described in Refs. [182, 261]. For each bin the total number of events n_{tot}, their total uncertainty σ_{tot} and the effective number of events $n_{\text{effective}} = n_{\text{tot}}^2/\sigma_{\text{tot}}^2$ are calculated. If the effective number of events is above an adjustable threshold, a combined uncertainty is created for that bin with a gaussian prior for the nuisance parameter. Otherwise the standard Barlow-Beeston method is retained. Several thresholds were tested and it was found that a threshold of zero (all bins use combined uncertainties) greatly reduces the runtime of the fit while changes to the result were negligible.

8.2 Modeling uncertainties

Cross section uncertainties

All samples of simulated data are normalized to the standard model cross sections of the described processes, as was discussed in Ch. 7. Uncertainties on these cross sections stem from uncertainties regarding the used parton density functions, the value of the strong coupling constant α_s, as well as the choices of the renormalization and factorization scales, as was discussed in Ch. 2. The uncertainties on the renormalization and factorization scales are treated fully correlated, as are those of the parton density functions and the strong coupling constant. They are listed together with the central values in Tab. 7.1. The procedures with which each uncertainty is estimated can be found in the relevant references given in Sec. 7.2. If the same uncertainty source affects multiple processes, the uncertainty is treated as fully correlated among the relevant processes. Such is the case for, for example, the renormalization and factorization scales of the $t\bar{t}$, $t\bar{t}$+Z and $t\bar{t}$+W processes. With regards to the parton-density-function and α_s uncertainties, processes are classified according to their dominant initial states (gg, qg or $q\bar{q}$) and the uncertainties of processes with similar dominant initial states are treated as fully correlated. The uncertainties affecting the $t\bar{t}H$ signal process are not correlated to any other processes. The resulting correlation scheme and the size of the uncertainties as a percentage of the predicted number of events are shown in Tab. 8.2.

Table 8.2: Uncertainties on the cross sections used for the normalization of the simulated samples. The values are gives as percentage of the nominal cross section values. Uncertainties in the same column affect multiple processes and are treated as fully correlated.

Process	pdf + α_s				Renorm./fact. scales				
	$gg_{\mathrm{t\bar{t}H}}$	gg	$q\bar{q}$	qg	$t\bar{t}$	t	V	VV	$t\bar{t}H$
$\mathrm{t\bar{t}H}$	3.6%	-	-	-	-	-	-	-	-9.2%/+5.8%
$\mathrm{t\bar{t}}$+jets	-	4%	-	-	-4%/+2%	-	-	-	-
$\mathrm{t\bar{t}}$+W	-	-	2%	-	-12%/+13%	-	-	-	-
$\mathrm{t\bar{t}}$+Z	-	3%	-	-	-12%/+10%	-	-	-	-
single t	-	-	-	3%	-	-2%/+3%	-	-	-
V+jets	-	-	4%	-	-	-	1%	-	-
WW/WZ/ZZ	-	-	2%	-	-	-	-	2%	-

Renormalization and factorization scale uncertainties

The choice of the renormalization scale μ_{R} and factorization scale μ_{F} used by the matrix-element generator during the generation of the simulated samples is rather arbitrary. To reflect this, uncertainties are introduced where both scales are increased or decreased independently by factors of 0.5 or 2.0. These variations are implemented with weights that are generated during the matrix-element level simulation step. Additional template histograms are created for these uncertainties by multiplying the overall weight of each event by the corresponding uncertainty weight. Since the overall change of the normalization caused by variation of these scales is already accounted for by the cross section uncertainties described in the paragraph above, the uncertainty histograms are normalized to the integrals of the nominal histograms, leaving only the changes of the shapes of the observables. The renormalization and factorization scale uncertainties are only applied to the $\mathrm{t\bar{t}}$ background and treated as fully correlated between the five $\mathrm{t\bar{t}}$ subprocesses. For all other processes the change to the shape of the observables is neglected.

Initial- and final-state radiation during the parton shower

The simulation of initial- (ISR) and final-state (FSR) radiation during the parton shower simulation, discussed in Ch. 2, is sensitive to the strong coupling constant α_{S}. Uncertainties on α_{S} are taken into account by additional shape changing uncertainties where renormalization scales used in the calculation of the ISR and FSR are varied up or down by a factor of two, as described in Ref. [262]. The uncertainties on the ISR and FSR are treated as separate uncorrelated uncertainties. Furthermore, they are only taken into account for the five $\mathrm{t\bar{t}}$ subprocesses and the uncertainties for each subprocess are treated as uncorrelated, motivated by the observation that the additional heavy flavor jets predominantly stem from additional radiation.

For the analysis of the data recorded in the year 2016 the template histograms are created using additional simulated samples for which the value of the renormalization scale for either the ISR or FSR simulation is varied. The size of these dedicated samples is much smaller than the size of the nominal samples, leading to large statistical fluctuations that can overshadow the actual effect of the uncertainty in the bins of the template histograms. In several cases this leads to bins where both the up and down variation change the predicted number of events in the same direction. To avoid this issue the uncertainties are instead estimated conservatively by the change of the predicted number of events as a

function of the jet multiplicity. For each considered jet multiplicity class and process the change of the predicted number of events is compared to the statistical uncertainty of this prediction. In case the up and down variations change the number of events in opposite direction, the larger value of either the predicted change or its statistical uncertainty is used as uncertainty. If the variations change the number of events in the same direction, the larger change is compared to its statistical uncertainty, where again the larger value of either the change itself or its statistical uncertainty is taken as final uncertainty for this direction. For the opposite direction, the statistical uncertainty of the corresponding change in the number of events is used. As a result, the number of events for each jet multiplicity class is assigned a well defined two-sided uncertainty. The jet multiplicity dependent uncertainties are finally correlated between all jet multiplicities to account for migration effects since changes in the ISR or FSR can lead to changes in the number of jets.

For the analysis of the data recorded in the year 2017, improvements of the simulation procedure allowed for the implementation of event weights reflecting the ISR and FSR uncertainties, similar to the μ_R and μ_F scales discussed above. For this analysis the additional template histograms are created using these event weights.

ME-PS matching

During the simulation of events, real-emission radiation can originate from the matrix element as well as the parton shower simulation. When interfacing a matrix element simulation with a parton shower simulation, care has to be taken that the overall high-p_T radiation matches observable data. The POWHEG generator, described in Refs. [65–67], matches the matrix element calculation and the parton shower simulation using a scheme in which real-emission radiation is dampened depending on its transverse momentum. The parameter that controls this dampening is called h_damp and effectively regulates high-p_T radiation.

For the simulated samples of $\mathrm{t\bar{t}}$ events the choice of this parameter is taken into account by an additional uncertainty, uncorrelated for each $\mathrm{t\bar{t}}$ subprocess. During the simulation the nominal parameter is set to $h_\mathrm{damp} = 1.58 m_\mathrm{t}$ where m_t denotes the mass of the top quark. For the estimation of the uncertainty additional samples are generated with values of $h_\mathrm{damp} = 1.0 m_\mathrm{t}$ and $h_\mathrm{damp} = 2.24 m_\mathrm{t}$. For both the analyses of the data recorded in the year 2016 and the year 2017 the sizes of the additional samples are limited, leading to sizable statistical uncertainties on the predicted number of events in the bins of the corresponding template histograms. For the analysis of data recorded in 2016, h_damp is treated with the same approach as for the ISR and FSR uncertainties. In the analysis of data recorded in 2017, the statistical uncertainties of the uncertainty histograms are even larger. In this case the ME-PS matching uncertainty is estimated independently for each ANN node by averaging the changes of the numbers of predicted events caused by the up and down variations of the h_damp parameter. These averages are then taken as symmetric rate uncertainties for each ANN node, changing the predicted numbers of events by about 5-10%. The uncertainties of all ANN nodes are correlated among identical processes.

Underlying-event tune

The effects on the simulated samples due to uncertainties of the underlying-event tune are estimated using additional samples in which the parameters of the tune are varied from their nominal values. The sizes of the additional samples are again limited and the uncertainties are implemented in the same way as those of the h_damp parameter. The uncertainty is only applied to the prediction of the $\mathrm{t\bar{t}}$ subprocesses and treated as uncorrelated among them.

A more detailed description of the tune and the corresponding uncertainties can be found in Ref. [232].

PDF uncertainties

The uncertainties of the PDF sets used for the generation of the simulated samples are taken into account using event weights. These weights are calculated on a per-event basis from additional replica PDF sets that are provided with the nominal PDF sets, as discussed in Refs. [53, 263]. First, a distribution of the central PDF weight and all 100 replica PDF weights is created. By construction, the median of this distribution corresponds to the central PDF weight. The weights corresponding to the up and down variations are then calculated as the 68% quantiles around the median. The resulting event weights are used to construct additional template histograms. Since the effect of the PDF uncertainties on the normalization is already considered by the cross section uncertainties described above, the resulting template histograms are normalized to the integrals of the corresponding nominal histograms. The PDF uncertainties are only considered for the $t\bar{t}H$ and $t\bar{t}$ subprocesses and are correlated among these processes.

$t\bar{t}$ + heavy flavor modeling

As was discussed before, $t\bar{t}$ + heavy flavor production, especially $t\bar{t}+b\bar{b}$, is an irreducible and in many ANN nodes dominant background process for this analysis, and since the prediction of the $t\bar{t}$ subprocesses is taken from simulated data their correct description is of utmost importance. The production of $t\bar{t}+b\bar{b}$ was discussed in Sec. 2.4.1 where it was found that the description is challenging and that currently available simulations tend to underestimate the production cross section.

To study the effect of the description of kinematic observables on the $t\bar{t}H(b\bar{b})$ analysis, different predictions are compared. For this, shapes of the ANN discriminant distributions and of important input features are evaluated using samples of simulated data, generated with different MC generators. As was discussed in Sec. 7.2 of the previous chapter, the sample of $t\bar{t}+b\bar{b}$ events used in the analysis is obtained from a simulation of $t\bar{t}$+jets production with Powheg+Pythia in the 5-flavor scheme (5FS). This is compared to dedicated samples of $t\bar{t}+b\bar{b}$ events that were created in the 4-flavor scheme (4FS) using Sherpa, discussed in Ref. [70], interfaced with OpenLoops, discussed in Ref. [69], in the case of the 2016 data analysis, as well as Powheg interfaced with OpenLoops (4FS), Powheg interfaced with Helac (5FS), which is discussed in Ref. [264], and MadGraph5_aMC@NLO (5FS) in the case of the 2017 data analysis. It is found that the shapes predicted by the nominal Powheg+Pythia simulation are compatible with those of the alternative simulations within statistical uncertainties.

Inclusive cross sections for $t\bar{t}+b\bar{b}$ production at a center-of-mass energy of 13 TeV have been previously measured with uncertainties ranging from 13% to 36%, as discussed in Refs. [102–105] and in Sec. 2.4. In these studies it was found that the 5FS $t\bar{t}$+jets simulation generated with Powheg+Pythia predicted a smaller $t\bar{t}+b\bar{b}$ cross section than the one measured in recorded data, although the predicted cross section values were still compatible with the measured values within their uncertainties. Theoretical calculations with NLO accuracy at center-of-mass energies of 8 TeV and 13 TeV found uncertainties of 20% to 40% for the $t\bar{t}+b\bar{b}$ cross section, as described in Refs. [91, 93–98].

Since these results show that the shape of the $t\bar{t}$ + heavy flavor processes is rather well described, whereas the $t\bar{t}+b\bar{b}$ cross section has large uncertainties, additional normalization uncertainties are introduced in the $t\bar{t}H$ analyses. For this, each of the four $t\bar{t}$ + heavy flavor

processes is assigned an additional 50% rate uncertainty, in addition to the overall cross section uncertainties affecting all $t\bar{t}$ subprocesses.

To verify that the full uncertainty model including these additional $t\bar{t}$ + heavy flavor uncertainties is able to cover possible mismodelling of the $t\bar{t}$+$b\bar{b}$ process without the introduction of large biases in the fit results, the predicted ANN distributions are fitted together to simulated toy data. These tests are performed separately for the analyses of data recorded in the years 2016 and 2017. In one test case, the simulated toy data are constructed by replacing the $t\bar{t}$+ $\geq$ 1b parts (in case of the 2016 data analysis) or the $t\bar{t}$+$b\bar{b}$ part (in case of the 2017 data analysis) with any of the available alternative predictions. The $t\bar{t}$H contribution to the toy data is either set to 0 (no signal) or to 1 (SM signal). In all cases the fit model is able to recover the injected signal strength with only small biases that are fully covered by the uncertainty. In another test performed for the analysis of the data recorded in the year 2016, the nominal $t\bar{t}$+$b\bar{b}$ shape is scaled by a factor of 1.3, corresponding to a larger than predicted $t\bar{t}$+$b\bar{b}$ cross section as found in the discussed measurements. The $t\bar{t}$H contribution to the toy data is set to 1 (SM signal). In this case, the fit model is able to recover the injected signal strength within 5%, while for multiple repetitions of this test using toy data sampled from the modified toy data the mean bias was found to be 2%. It is also found that a reduction of the additional $t\bar{t}$ + heavy flavor uncertainties to 35% does not reduce the uncertainty on the measured signal strength while increasing the bias on the recovered signal strength to 7%. An increase of the additional $t\bar{t}$ + heavy flavor uncertainties to 100% on the other hand reduces the bias to only 2% while also increasing the uncertainty on the signal strength by 5%. The chosen value of 50% for the additional $t\bar{t}$ + heavy flavor uncertainties is taken as an acceptable compromise between accuracy and precision of the measurement. Several of the discussed studies and additional ones are discussed in greater detail in Ref. [265].

Finally, other measurements of $t\bar{t}$H($b\bar{b}$) introduce additional uncertainties constructed from the differences between different MC generators, for example in Ref. [266], instead of the additional $t\bar{t}$ + heavy flavor uncertainties used in this thesis. This approach is not chosen in this thesis since in such a setup, pulls away from the central values of the corresponding nuisance parameters are not clear to interpret, whereas pulls of the additional $t\bar{t}$ + heavy flavor nuisance parameters can be interpreted as changes to the rates of the $t\bar{t}$ + heavy flavor processes.

8.3 Correlation of 2016 and 2017 uncertainties

Almost all systematic uncertainties are considered for both the analyses of the data recorded in the years 2016 and 2017 and several of them share the same source for both analyses. In these cases, the uncertainties are treated as fully correlated for the combined analysis of the data of both years. The correlated uncertainties are listed in Tab. 8.3. The uncertainties on the inclusive cross sections are treated as fully correlated since the same calculations are used for the normalization of the samples for both years. The pileup uncertainty stems from the uncertainty of the inelastic proton-proton cross section which is also identical for both years. The shape uncertainties due to the choice of the renormalization and factorization scales used for the simulation of the $t\bar{t}$ samples is correlated since the same matrix-element generator is used with identical settings. Several of the JES uncertainties are correlated between the years following the recommendation of the jet energy calibration group of the CMS collaboration. All other uncertainties are treated as fully uncorrelated since they were either derived using different samples or affect different generators or algorithms.

Table 8.3: Correlation of systematic uncertainties in the combination of the 2016 and 2017 data analyses.

Correlated uncertainties
Renorm./fact. scales, PDF
Pileup
Renorm./fact. scales of $t\bar{t}$ simulation
JES: AbsoluteMPFbias
JES: AbsoluteScale
JES: FlavorQCD
JES: Fragmentation
JES: PileUpDataMC
JES: PileUpPt[BB][EC1][EC2][HF]
JES: PileUpPtRef
JES: RelativeFSR
JES: RelativeJERHF
JES: RelativePt[BB][HF]
JES: RelativeBal
JES: SinglePionECAL
JES: SinglePionHCAL

9 Results and Interpretation

To measure the fraction of $t\bar{t}H(b\bar{b})$ events in the selected data, the ANN discriminant distributions of simulated data are fitted to those of recorded data using a binned profile likelihood method (see Ch. 4). Although the analysis phase space and ANNs are specifically constructed to select $t\bar{t}H(b\bar{b})$ events, a very small number of $t\bar{t}H$ events with Higgs boson decays to particles other than bottom quarks are also selected. For the final results this $t\bar{t}H(\text{non-}b\bar{b})$ contribution will not be further distinguished from the $t\bar{t}H(b\bar{b})$ contribution, instead the results will be calculated for the joint $t\bar{t}H$ process assuming standard model couplings for branching fractions of the Higgs boson. The predicted ANN discriminant distributions for signal and background processes are initially normalized to their respective standard model cross sections. Systematic uncertainties affecting the normalization of the processes or the shapes of the distributions are implemented as constrained nuisance parameters with normal or log-normal probability density functions as discussed in Ch. 8. The normalization of the $t\bar{t}H$ process is additionally scaled by an unconstrained parameter called *signal-strength modifier* and denoted μ. Initially this parameter is set to a value of 1 corresponding to the standard model cross section. During the fit to data the signal-strength modifier and the nuisance parameters are varied simultaneously to maximize the likelihood function. The best-fit value of μ can then be interpreted as the ratio of the measured $t\bar{t}H$ cross section with respect to the standard model $t\bar{t}H$ cross section: $\mu = \sigma/\sigma_{\text{SM}}$. As discussed in Ch. 4, the uncertainties on μ are estimated as the closest values of $\mu \pm \Delta\mu$ for which the profile likelihood test statistic increases by a value of 1 with respect to the global minimum of the test statistic.

Results of the $t\bar{t}H(b\bar{b})$ analyses in the single-lepton (SL) channel are studied in Sec. 9.1. The combination with the analyses in the fully-hadronic (FH) and dilepton (DL) channels are discussed in Sec. 9.2. As was discussed before, the analysis presented in this thesis was developed in close collaboration with members of the CMS Collaboration. The main results presented for the SL channel and the combinations with the DL and FH channels have been previously published in Refs. [4, 5], in which however a less detailed discussion of the SL channel results has been given.

The analysis of $t\bar{t}H(b\bar{b})$ production in the SL channel has also included in combinations performed by the CMS Collaboration in which analyses targeting different Higgs boson production and decay modes are jointly fit to data. These combinations and their results will be discussed in Sec. 9.3. Finally, measurements of $t\bar{t}H(b\bar{b})$ production and the coupling

of the Higgs boson to the top quark, to which the $t\bar{t}H(b\bar{b})$ analyses in this thesis contribute, are compared to similar measurements in Sec. 9.4.

9.1 Results in the single-lepton channel

The joint fit of the ANN discriminant distributions for the analysis of data recorded in the year 2016 results in a best-fit signal-strength modifier of

$$\hat{\mu}_{\mathrm{SL},2016} = 0.84^{+0.52}_{-0.50}(\mathrm{tot.}) \left[^{+0.27}_{-0.27}(\mathrm{stat.})^{+0.44}_{-0.42}(\mathrm{syst.})\right] \tag{9.1}$$

where the statistical uncertainty is calculated by freezing all systematic nuisance parameters to their respective post-fit values and the systematic uncertainties is calculated by subtracting the statistical component quadratically from the total uncertainty. In comparison, for a fit to pseudodata with a standard model $t\bar{t}H$ cross section the expected signal-strength modifier and total uncertainty is $\hat{\mu}_{\mathrm{SL},2016,\mathrm{exp.}} = 1.00^{+0.59}_{-0.56}$. This corresponds to an observed (expected) significance of 1.7σ (1.9σ) above the background-only hypothesis.

For the analysis of the data recorded in the year 2017 the resulting best-fit signal-strength modifier is found to be

$$\hat{\mu}_{\mathrm{SL},2017} = 1.84^{+0.62}_{-0.56}(\mathrm{tot.}) \left[^{+0.26}_{-0.26}(\mathrm{stat.})^{+0.56}_{-0.50}(\mathrm{syst.})\right] \tag{9.2}$$

with an expected signal-strength modifier and total uncertainty of $\hat{\mu}_{\mathrm{SL},2017,\mathrm{exp.}} = 1.00^{+0.45}_{-0.43}$. This corresponds to an observed (expected) significance of 3.3σ (1.9σ).

For the combination of both analyses, correlating the nuisance parameters as described in Ch. 8, the resulting signal-strength modifier is

$$\hat{\mu}_{\mathrm{SL}} = 1.22^{+0.41}_{-0.37}(\mathrm{tot.}) \left[^{+0.19}_{-0.19}(\mathrm{stat.})^{+0.36}_{-0.32}(\mathrm{syst.})\right] \tag{9.3}$$

with an expected signal-strength modifier and total uncertainty of $\hat{\mu}_{\mathrm{SL},2017,\mathrm{exp.}} = 1.00^{+0.37}_{-0.34}$, corresponding to an observed (expected) significance of 3.3σ (2.7σ). The fit results are summarized together with the results of the individual analysis categories in Fig. 9.1. The results for the combined fits are all compatible with the standard model.

Figure 9.2 shows the distribution of the ANN output in the $t\bar{t}H$ node for events with at least 6 jets in data recorded in the year 2017 before the fit to data (Fig. 9.2a), after a signal+background fit of only this distribution (Fig. 9.2b), after background-only (Fig. 9.2c) and signal+background (Fig. 9.2d) fits of all distributions simultaneously to 2017 data, and a joint signal+background fit of all distributions to 2016 and 2017 data (Fig. 9.2e).
The addition of the $t\bar{t}H$ signal, shown as blue line before and as blue histogram after the fit to data, improves the fit of simulated data to recorded data, as can be seen when comparing the distributions in Fig. 9.2a , in Fig. 9.2c and Fig. 9.2d. In addition, the total uncertainty on the background before the fit in Fig. 9.2a is greatly reduced after the fit to data recorded in the year 2017 in Fig. 9.2d and is even further reduced after the combined fit in Fig. 9.2e. This reduction of the total uncertainty is due to constraints of the nuisance parameters which are mainly caused by the addition of the background enriched distributions, as can be seen by comparison to the fit of only the $t\bar{t}H$ node distribution in Fig. 9.2b. All ANN distributions before and after the fits to data can be found in Appendix C. The numbers of simulated and recorded events before and after the joint fit to data are given in Appendix E.

To investigate how different classes of uncertainties contribute to the total uncertainty, all nuisance parameters are grouped into three distinct classes: experimental uncertainties, theory uncertainties and uncertainties related to the size of the simulated samples.

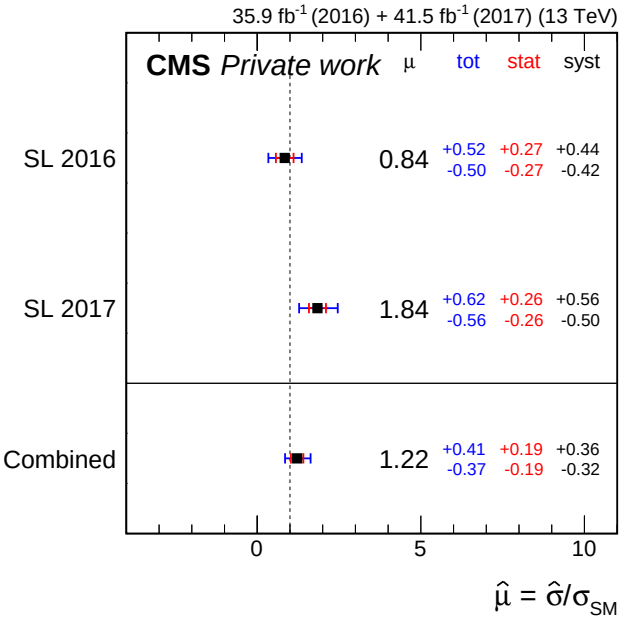


Figure 9.1: Best-fit values of the signal-strength modifier μ for the analyses of data recorded in the years 2016 and 2017. The total uncertainties (outer error bar) at the 68% confidence level and the statistical uncertainties (inner error bar) are also shown.

Experimental uncertainties related to either b tagging or the jet energy calibration are further grouped into sub-classes as are theory uncertainties related to the $t\bar{t}H$ modeling or the $t\bar{t}$ + heavy flavor modeling. The contribution of all uncertainty classes is estimated by fixing the nuisance parameters of that class to their post-fit values, repeating the fit to data and then quadratically subtracting the resulting uncertainty of that fit from the total uncertainty of the fit with all nuisance parameters included. The statistical uncertainty is calculated by freezing all nuisance parameters to their post-fit values. The breakdown of the contributions is shown in Tab. 9.1 for the joint fit to 2016 and 2017 data. The uncertainties of the classes do not add to the total uncertainty due to correlations between uncertainties from different classes. The total systematic uncertainty $\sigma_{\text{syst}} = ^{+0.36}_{-0.32}$ is larger than the statistical uncertainty $\sigma_{\text{stat}} = \pm 0.19$, with the total uncertainty being $\sigma_{\text{syst}} = ^{+0.41}_{-0.37}$. The largest contribution to the systematic uncertainty stems from the modeling of the signal and background processes with $\sigma_{\text{theory}} = ^{+0.28}_{-0.24}$, while the experimental uncertainties and the uncertainty due to the limited size of the simulated samples have smaller effects with $\sigma_{\text{exp.}} = \pm 0.14$ and $\sigma_{\text{sample size}} = ^{+0.13}_{-0.14}$, respectively. The uncertainties only affecting the $t\bar{t}$ + heavy flavor modeling, consisting of the additional 50%, normalization uncertainties discussed in Sec. 8.2 and parton shower related uncertainties, have with $\sigma_{t\bar{t}+\text{heavy flavor mod.}} = \pm 0.20$ the largest contribution of all modeling uncertainties. The uncertainties affecting the other processes have minor effects. Another large contribution stems from the uncertainty on the cross section for $t\bar{t}H$ production with $\sigma_{\text{signal mod.}} = ^{+0.16}_{-0.07}$. The largest contributions to the experimental uncertainties are from uncertainties related to b tagging, with $\sigma_{\text{b tag}} = ^{+0.1}_{-0.11}$, and the jet energy calibration with $\sigma_{\text{JEC}} = ^{+0.08}_{-0.09}$ while the other experimental uncertainties, such as those related to the triggers or the leptons, have smaller contributions. The effects of the different sources of uncertainties follow a similar pattern as was expected by a fit to simulated data. The results for the fits to data recorded either in the year 2016 or 2016 can be found in Appendix D.1 and show similar behavior.

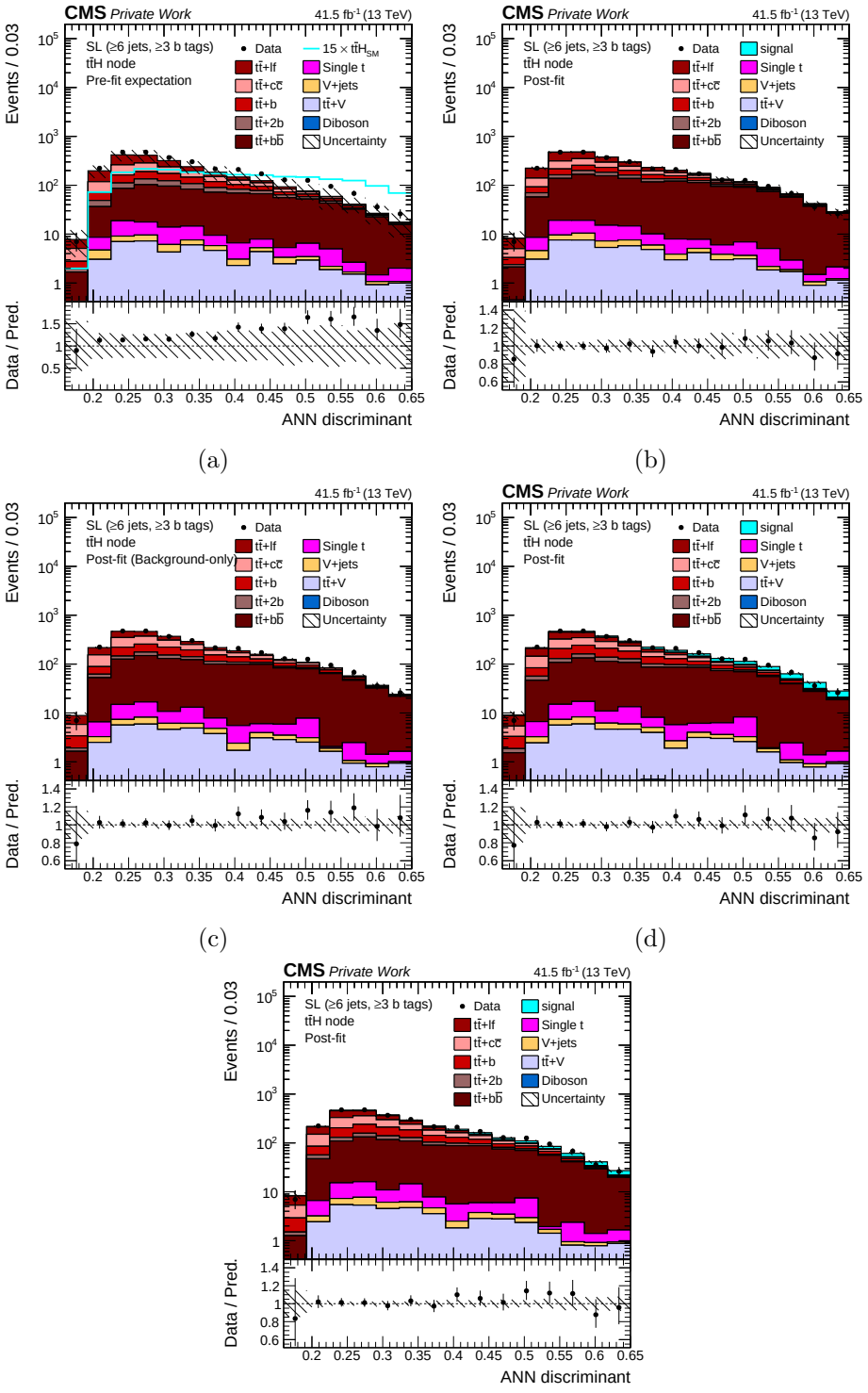


Figure 9.2: ANN discriminant distribution of the $t\bar{t}H$ node in the (≥ 6 jets, ≥ 3 b-tags) category of 2017 data analysis: (a) before the fit to data , (b) after a signal+background (s+b) fit of only this node, (c) after a background-only (b-only) fit of all ANN nodes to 2017 data, (d) after a s+b fit of all ANN nodes to 2017 data, and (e) after a s+b fit of all ANN nodes to 2016 and 2017 data. Recorded data are shown as black markers, the expected background as stacks of histograms and the $t\bar{t}H$ process as blue line (histogram) before (after) the fit to data. The $t\bar{t}H$ process is additionally scaled by a factor of 15 before the fit. The hatched uncertainty bands show the total uncertainty before and after the fit to data.

Table 9.1: Contributions of different classes of uncertainties to the total uncertainty for the joint fit to the data recorded in the years 2016 and 2017. The expected value is calculated for a standard model signal+background prediction. The quoted uncertainties $\Delta\mu$ are obtained by fixing the uncertainties in the class to their post-fit values, repeating the fit, and quadratically subtracting the obtained uncertainty from the uncertainty of the full fit.

Uncertainty source	$\pm\Delta\mu$ (observed)	$\pm\Delta\mu$ (expected)
Total experimental	+0.14/-0.14	+0.12/-0.11
b tagging	+0.1/-0.11	+0.07/-0.06
jet energy scale	+0.08/-0.09	+0.09/-0.08
Total theory	+0.28/-0.24	+0.24/-0.21
signal modeling	+0.16/-0.07	+0.14/-0.05
parton shower (PS)	+0.11/-0.11	+0.1/-0.1
$t\bar{t}$ + heavy flavor model (50% norm. + PS)	+0.2/-0.2	+0.17/-0.17
$t\bar{t}$ + heavy flavor add. 50% norm.	+0.12/-0.12	+0.1/-0.11
Total theory & experimental	+0.34/-0.31	+0.31/-0.28
Size of simulated samples	+0.13/-0.14	+0.11/-0.11
Total systematic	+0.36/-0.32	+0.32/-0.29
Statistical	+0.19/-0.19	+0.18/-0.18
Total	+0.41/-0.37	+0.37/-0.34

The behavior of the nuisance parameters can be further investigated by looking at their post-fit values, their constraints and their correlations. The normalized difference of the post-fit value of a nuisance parameter and its pre-fit value is called *pull* and is calculated as:

$$\text{pull} = \frac{\theta_{\text{post-fit}} - \theta_{\text{pre-fit}}}{\sigma_{\text{pre-fit}}}. \tag{9.4}$$

The constraint of a nuisance parameter is defined as the ratio of its post-fit uncertainty and its pre-fit uncertainty:

$$\text{constraint} = \frac{\sigma_{\text{post-fit}}}{\sigma_{\text{pre-fit}}}. \tag{9.5}$$

The pulls and constraints of all nuisance parameters, except those related to the size of the simulated samples, are shown in Fig. 9.3 for the joint fit to data recorded in the years 2016 and 2017. The pulls and constraints for the fits to data recorded in the years 2016 and 2017 and for the joint fit are shown in Appendix D.2 in groupings for better visibility.

Nearly all nuisance parameters are slightly constrained and pulled away from their initial values by at most one standard deviation in terms of their post-fit uncertainty. Large pulls and constraints can be observed for nuisance parameters of the b-tagging uncertainties and additional $t\bar{t}$ + heavy flavor normalization uncertainties, which is not unexpected since the $t\bar{t}H(b\bar{b})$ analyses are especially sensitive to b tagging and $t\bar{t}$ + heavy flavor processes.

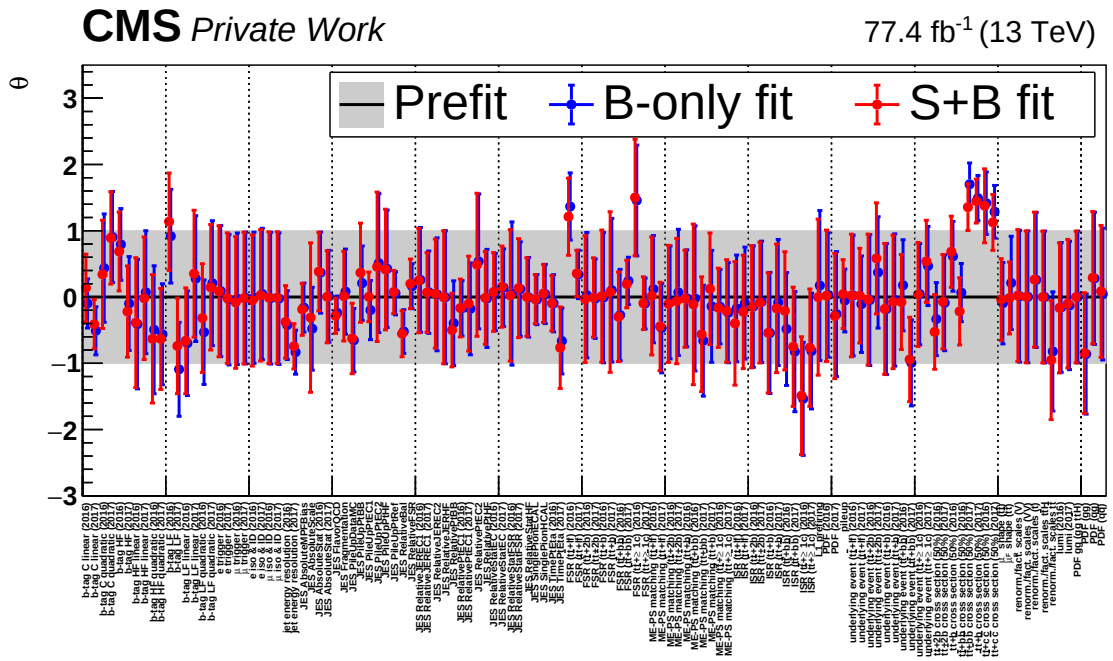


Figure 9.3: Pulls and constraints of the nuisances parameters for the combined fit to data recorded in the years 2016 and 2017. For better visibility the nuisance parameters related to the statistical uncertainty of the simulated samples are omitted.

Large pulls can also be observed for nuisance parameters related to the conservatively estimated uncertainties on the initial- and final-state radiation as well as the underlying event.
The nuisance parameters of the additional $t\bar{t}$ + heavy flavor normalization uncertainties are separately shown in Fig. 9.4. Those nuisance parameters of the additional $t\bar{t}$ + heavy flavor normalization uncertainties are all pulled upwards, reflecting larger than predicted numbers of $t\bar{t}$ + heavy flavor events in the data. This could be related to the observation that previous measurements of $t\bar{t}$+$b\bar{b}$ production consistently measured more $t\bar{t}$+$b\bar{b}$ events than predicted by the POWHEG+PYTHIA simulation, as was discussed in Sec. 8.2. Since $t\bar{t}$+$b\bar{b}$ production is an irreducible background for $t\bar{t}H(b\bar{b})$ and the shapes of the $t\bar{t}$+$b\bar{b}$ process look somewhat similar to the shapes of the signal process, large correlation coefficients could be expected between nuisance parameters controlling the $t\bar{t}$ + heavy flavor modeling and the signal-strength modifier. This in turn could lead to the nuisance parameters hiding signal or the signal-strength modifier being used to account for shortcomings of the $t\bar{t}$ + heavy flavor modeling.

For further investigation, the constrained nuisance parameter for the $t\bar{t}$+$b\bar{b}$ normalization is replaced by an unconstrained parameter $\mu_{t\bar{t}+b\bar{b}}$. The result of this now 2-dimensional fit is shown in Fig. 9.5 in the form of the distribution of the profile likelihood test statistic. As can be seen, the $t\bar{t}$+$b\bar{b}$ signal strength is fitted to a value of about $\mu_{t\bar{t}+b\bar{b}} \approx 1.48^{+0.19}_{-0.21}$ which is consistent with the results of the dedicated $t\bar{t}$+$b\bar{b}$ measurements discussed in Sec. 2.4.1. The $t\bar{t}H$ and $t\bar{t}$+$b\bar{b}$ signal-strength modifiers are only weakly anti-correlated with a correlation coefficient of $c = -0.14$. Similar tests are performed for the $t\bar{t}$+2b, $t\bar{t}$+b processes as well as for the sum of the $t\bar{t}$+$b\bar{b}$, $t\bar{t}$+2b and $t\bar{t}$+b processes. The corresponding figures are shown in Appendix D.3. This suggests that the pulls of the $t\bar{t}$ + heavy flavor

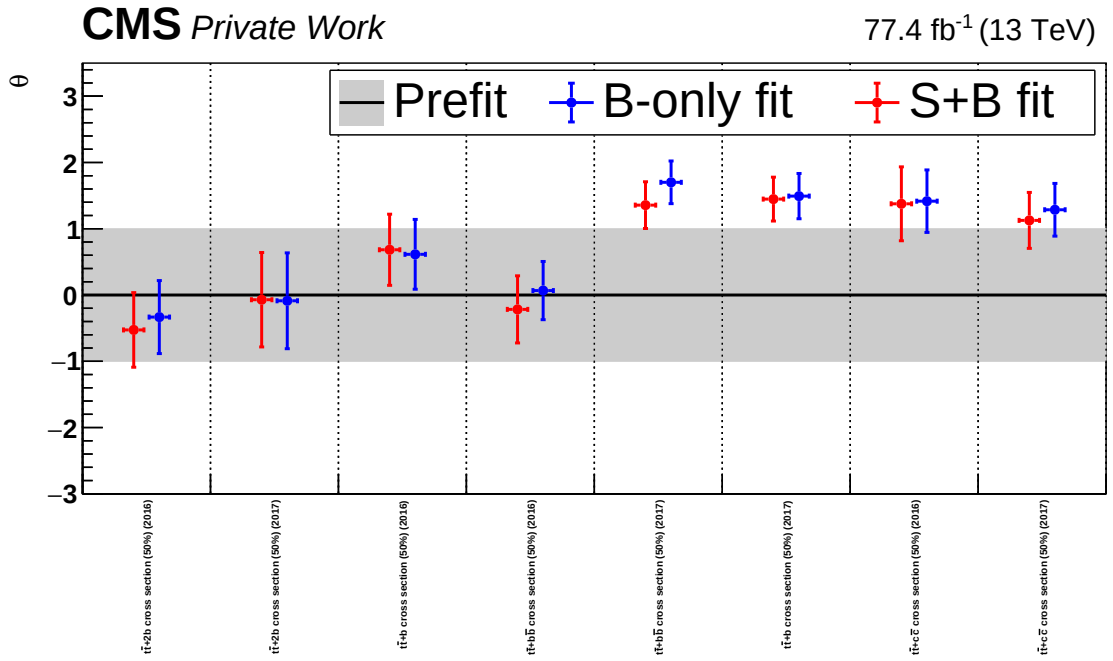


Figure 9.4: Pulls and constraints of the nuisances parameters related to the normalization of the $t\bar{t}$ + heavy flavor processes for the joint fit to data recorded in the years 2016 and 2017.

normalization are due to an increase of $t\bar{t}$ + heavy flavor production with respect to the predicted values

Returning to the fit with a single signal-strength modifier $\mu_{t\bar{t}H}$, correlation coefficients between the signal strength modifier and several nuisance parameters are calculated from the covariance matrix of the fit and are shown in Fig. 9.6. Since the number of nuisance parameters in the fit is high, only a selected view are shown. Those are the twenty nuisance parameters with the strongest correlations to the signal-strength modifier and the twenty nuisance parameters with the strongest correlations among themselves. The signal-strength modifier μ is only moderately correlated to the nuisance parameters. The strongest correlations can be observed between nuisance parameters related to the $t\bar{t}$ + heavy flavor normalization, the parton shower simulation and the renormalization and factorization scales. This can be understood by the observation that, due to the conservative approach for their construction – explained in Ch. 8– the parton shower and scale uncertainties have only small effects on the shapes of the distributions and somewhat large effects on the normalization. As discussed in Sec. 8.2, the limitation due to the sample size is less severe for the analysis of the 2017 data. The correlation coefficients for the individual fits to data recorded in the years 2016 and 2017 are shown in Appendix D.4.

The presented studies suggest that the $t\bar{t}$ + heavy flavor background is modeled by the corresponding nuisance parameters and that they do not introduce a bias to the fitted signal strength.

Another measure of correlation between the signal-strength modifier and a nuisance parameter is given by the *impact* $\Delta\hat{\mu}$. It is calculated as the difference between the nominal best-fit value $\hat{\mu}$ and the value obtained when freezing the nuisance parameter to its post-fit value plus or minus its post-fit uncertainty. It therefore shows how much the result would change if the central value of the nuisance parameter would be larger

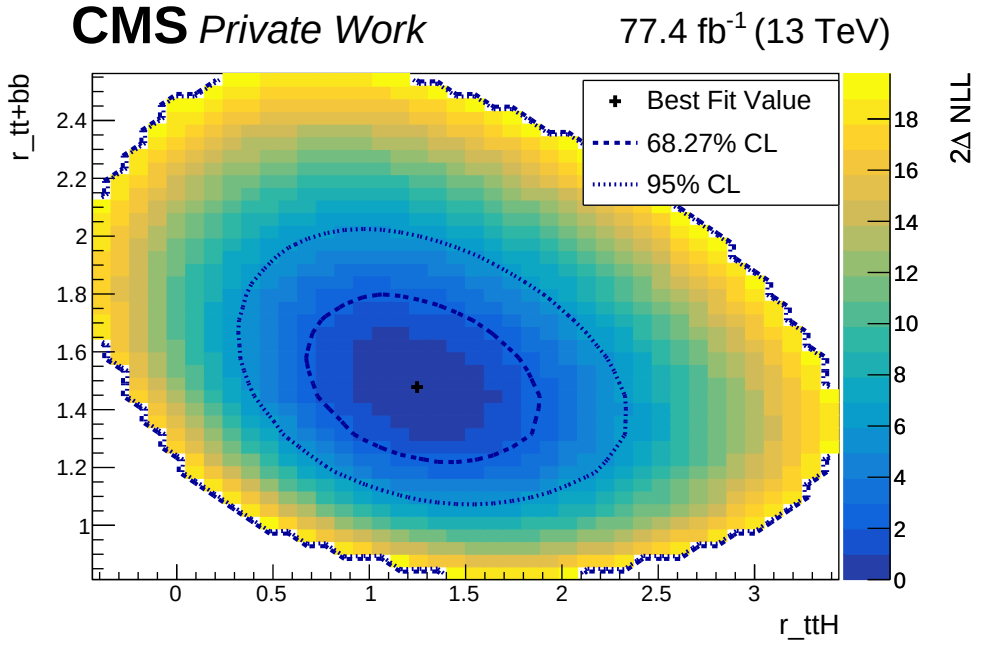


Figure 9.5: Profile likelihood test statistic values for a fit with two signal-strength modifiers: one for $t\bar{t}H$ and one for $t\bar{t}+b\bar{b}$ production. The 50% normalization uncertainty for $t\bar{t}+b\bar{b}$ production is removed for this purpose. All other nuisance parameters are profiled.

or smaller by one standard deviation. The impacts of the 20 nuisance parameters with the largest impacts for the combined analysis of data recorded in both years are shown in Fig. 9.7. Impact parameters for the other nuisance parameters and the individual analyses are shown in the Appendix. D.5. The nuisance parameter with the largest impact is the renormalization and factorization scale uncertainty on the $t\bar{t}H$ cross section, followed by the additional $t\bar{t}+b\bar{b}$ normalization uncertainty. Fixing either of these to their upper or lower uncertainty bounds changes the resulting signal-strength modifier by up to 0.12 or about 10%. Other important uncertainties are related to the other $t\bar{t}$ + heavy flavor normalizations, the modeling of the $t\bar{t}$ background (normalization and parton shower), b tagging of udsg jets and c jets, as well as the statistical uncertainty of the simulated events in the $t\bar{t}H$ ANN nodes.

In summary, trends in the nuisance parameters are well understood and their impacts, while not the detailed order, are found as to be expected.

9.2 Combination with other ttH(bb) channels

While the previous results for the analyses in the single-lepton channel represent the main result of this thesis, the analyses were developed in close collaboration with the analyses targeting dileptonically (DL) and fully hadronically (FH) decaying top quark-antiquark pairs, which are summarized in Sec. 7.7 and discussed in detail in Refs. [4, 5, 229]. The ANN distributions of the SL analyses are combined with the corresponding discriminant distributions of these other analyses in combined fits to reduce the statistical uncertainty and to exploit correlations between nuisance parameters shared between the analyses. All uncertainties affecting all three channels are treated as fully correlated for this purpose. The results of the separate analyses for the years 2016 and 2017 in the three channels

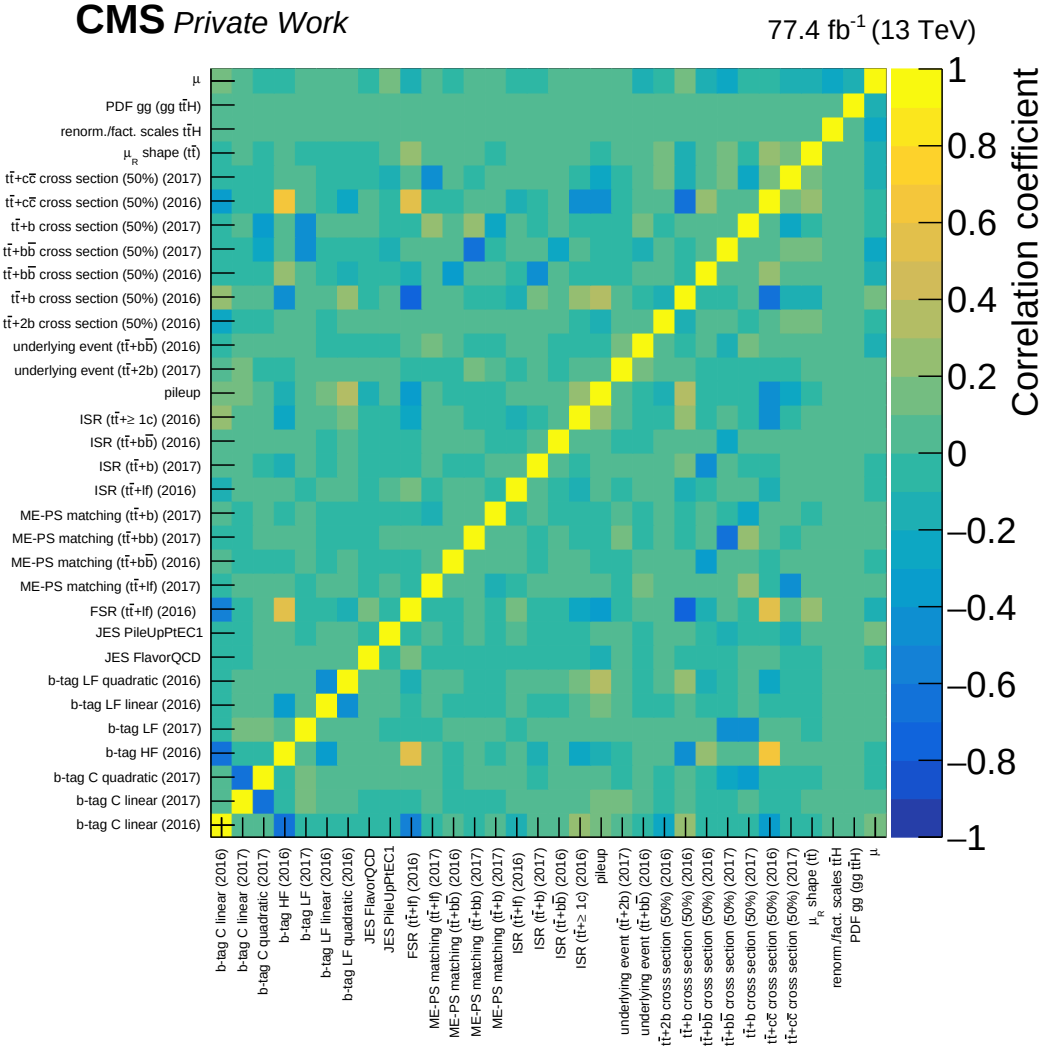


Figure 9.6: Correlation coefficients of the signal-strength modifier and several nuisance parameters for the combined fit to data recorded in the years 2016 and 2017.

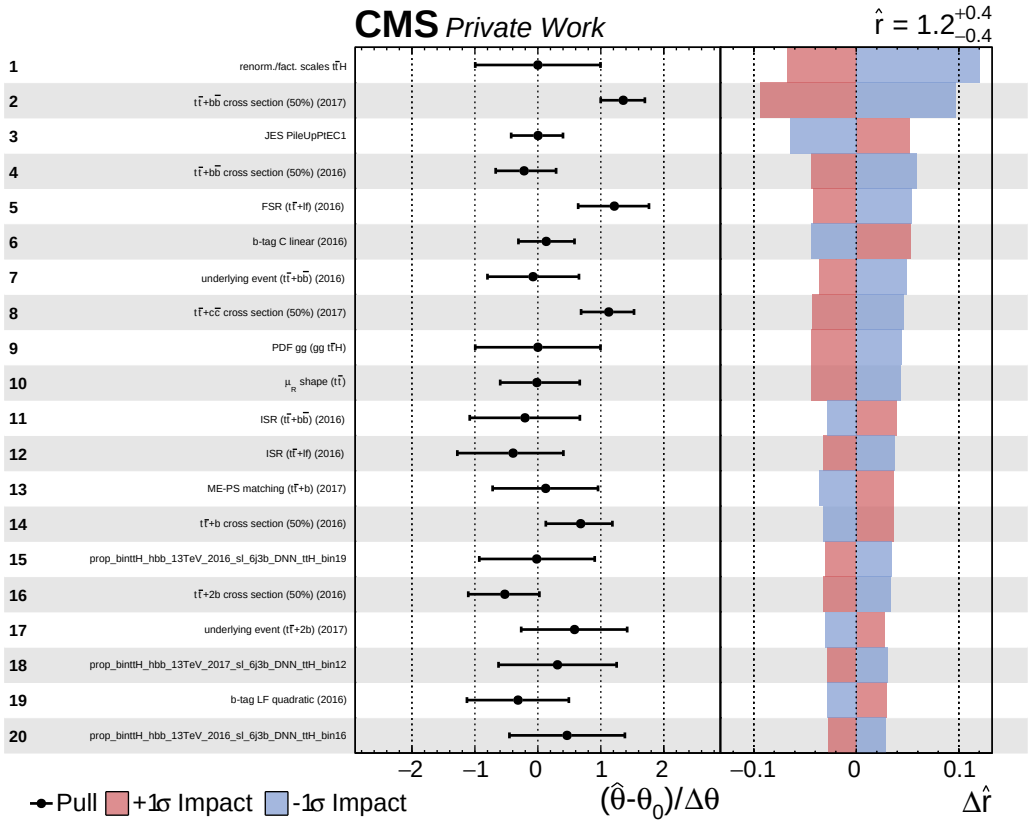


Figure 9.7: Impacts of the 20 nuisance parameters with the largest impacts on the signal-strength modifier for the combined fit to the data recorded in the years 2016 and 2017. The impacts of the nuisance parameters are shown on the right hand side of the figure while their pulls and constraints are shown in the middle column.

are summarized in Fig. 9.8a while combinations of these analyses are shown in Fig. 9.8b. The combination of all three analyses for data recorded in the year 2016 results in a signal-strength modifier of

$$\hat{\mu}_{2016} = 0.85^{+0.43}_{-0.41}(\text{tot.}) \left[{}^{+0.22}_{-0.22}(\text{stat.}) {}^{+0.37}_{-0.35}(\text{syst.}) \right] \tag{9.6}$$

with an expected signal-strength modifier and total uncertainty of $\hat{\mu}_{2016,\text{exp.}} = 1.00^{+0.50}_{-0.47}$, corresponding to an observed (expected) significance of 2.0σ (2.4σ) above the background-only hypothesis. The combination of the analyses for the data recorded in the year 2017 results in

$$\hat{\mu}_{2017} = 1.49^{+0.44}_{-0.40}(\text{tot.}) \left[{}^{+0.21}_{-0.20}(\text{stat.}) {}^{+0.39}_{-0.35}(\text{syst.}) \right] \tag{9.7}$$

with an expected result of $\hat{\mu}_{2017,\text{exp.}} = 1.00^{+0.38}_{-0.36}$. This corresponds to an observed (expected) significance of 3.7σ (2.6σ) above the background-only hypothesis. Finally, the combination of all analyses for both years results in

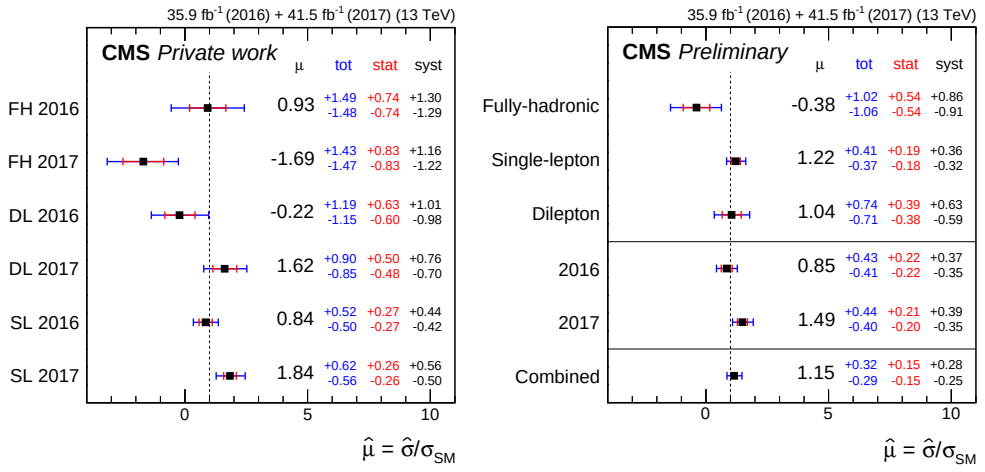
$$\hat{\mu}_{2016+2017} = 1.15^{+0.32}_{-0.29}(\text{tot.}) \left[{}^{+0.15}_{-0.15}(\text{stat.}) {}^{+0.28}_{-0.25}(\text{syst.}) \right]. \tag{9.8}$$

For this case the expected result is $\hat{\mu}_{2016+2017\text{exp.}} = 1.00^{+0.31}_{-0.29}$ and the observed (expected) significance is found to be 3.9σ (3.5σ) above the background-only hypothesis, thereby establishing evidence for $\text{t}\bar{\text{t}}\text{H}$ production in the $\text{H} \to \text{b}\bar{\text{b}}$ decay channel. The results are again compatible with the standard model Higgs boson.

Comparing the statistical and systematic uncertainties of the individual channels with those of the combination reveals that not only the statistical but also the systematic components are reduced. By combining the analyses, additional phase space regions are added through which nuisance parameters can be better constrained. Affected by this are, for example, nuisance parameters related to the jet energy calibration since the FH channel contains large numbers of jets and is sensitive to their modeling. Uncertainties related to b tagging, on the other hand, can be well measured by the SL and DL channels since they contain purer samples of b jets than the FH analysis.

9.3 Combination with other analyses by the CMS Collaboration

The analyses described in this thesis have been included in several combination studies performed by the CMS Collaboration, in which several analyses of Higgs boson production are jointly fit to data to consistently measure the couplings of the Higgs boson. The strength of the Higgs boson decay to bottom quarks can be measured in analyses targeting this decay such as, for example, $\text{t}\bar{\text{t}}\text{H}(\text{b}\bar{\text{b}})$ or analyses of the associated production with vector bosons and subsequent $\text{H} \to \text{b}\bar{\text{b}}$ decay. If independently analyzed, their results will however strongly depend on the assumed coupling of the Higgs boson in the production process since they always measure a process involving the couplings of both the production and the decay interaction, in the case of $\text{t}\bar{\text{t}}\text{H}(\text{b}\bar{\text{b}})$ for example the top-quark and bottom-quark Yukawa couplings. By simultaneously measuring the production process in a joint fit with an analysis targeting this production but a different decay, for example $\text{t}\bar{\text{t}}\text{H}, \text{H} \to \tau^+\tau^-$ this dependency can be reduced. The analyses discussed in the following combine multiple analyses of different Higgs boson production and decay channels in a multidimensional joint fit to data in which all coupling modifiers are treated as independent unconstrained parameters. They are parameterized either as signal-strength modifiers and branching fractions or as coupling-strength modifiers in the κ framework, which was introduced in Ch. 2 and is discussed in detail in Refs. [92, 120]. The ANN distributions predicted for



(a) (b)

Figure 9.8: Best-fit values of the signal-strength modifier μ for the analyses in the single-lepton (SL), dilepton (DL) and fully-hadronic (FH) channels of data recorded in the years 2016 and 2017. Shown are the results for the individual analyses (a) and for combinations of the channels (b). The total uncertainties (outer error bar) at the 68% confidence level and the statistical uncertainties (inner error bar) are also shown. The figure of the combined result (b) was taken from Ref. [5].

$t\bar{t}H(b\bar{b})$ process are for example scaled by two modifiers, one for the top-Higgs coupling and one for the bottom-Higgs coupling. For this purpose, the $t\bar{t}H(b\bar{b})$ analyses provide separate template histograms for $t\bar{t}H$ production with different Higgs boson decay channels. In the joint fits to data, nuisance parameters are correlated between different analysis channels if appropriate and all included analyses are performed on distinct non-overlapping phase space regions.

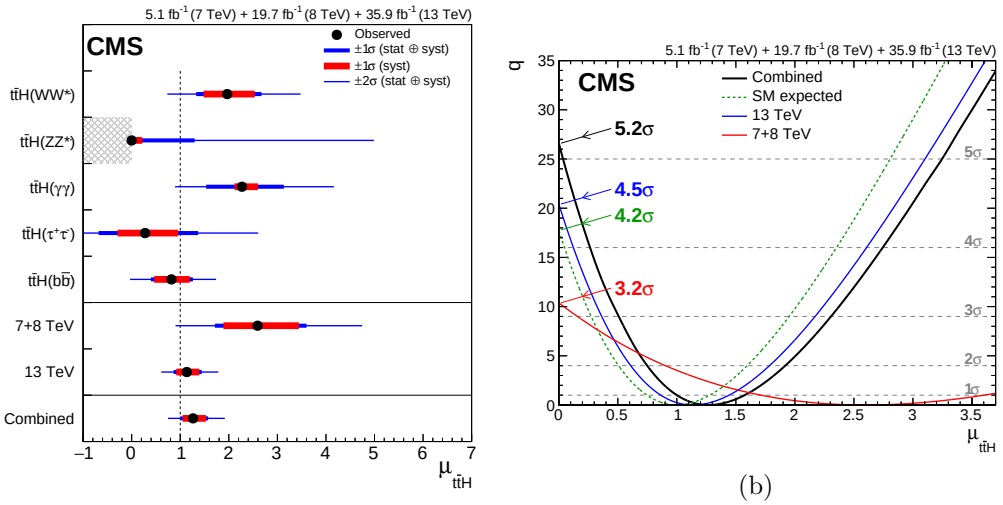
9.3.1 $t\bar{t}H$ combination

To measure the production of $t\bar{t}H$ independent from assumed Higgs boson branching fractions, the analyses in the SL, DL and FH channels of data recorded in the year 2016 have been combined with analyses targeting the decay of the Higgs boson to two photons ($H \to \gamma\gamma$), the decay of the Higgs boson to τ leptons or vector bosons ($H \to \tau^+\tau^-, WW^*, ZZ^*$) as well as corresponding analyses performed earlier at center-of-mass energies of 7 TeV and 8 TeV, as described in Ref. [267]. The combination increases the sensitivity of individual analyses through the reduction of the statistical uncertainty and the correlation of systematic uncertainties.

This combination results in a best-fit value of the signal-strength modifier for $t\bar{t}H$ production of

$$\hat{\mu}_{t\bar{t}H} = 1.26^{+0.31}_{-0.26}(\text{tot.})\left[^{+0.16}_{-0.16}(\text{stat.})^{+0.17}_{-0.15}(\text{exp.})^{+0.14}_{-0.13}(\text{theo. bkg.})^{+0.15}_{-0.07}(\text{theo. sig.})\right] \quad (9.9)$$

compatible with the standard model Higgs boson. The total systematic uncertainty is split into experimental and theory uncertainties for the modeling of the signal and background processes. The addition of other channels sensitive to $t\bar{t}H$ clearly increases the sensitivity to $t\bar{t}H$ production compared to the results measured in the SL channel (Eq. 9.1) and the combined analysis (Eq. 9.6) of data recorded in the year 2016. This gain in sensitivity is due to multiple factors. For one, the branching fraction of $H \to b\bar{b}$ is only about



(a) (b)

Figure 9.9: (a) Best-fit values of the signal-strength modifiers of the individual channels and the combined results for a combination of different $t\bar{t}H$ analyses performed at center-of-mass energies of 7 TeV, 8 TeV and 13 TeV. The uncertainties corresponding to one and two standard deviations are shown as blue and red bands. (b) Profile likelihood test statistic q as a function of $\mu_{t\bar{t}H}$ for all analysis channels. Taken from Ref. [267].

0.58 and the number of $t\bar{t}H(\text{non-}b\bar{b})$ events selected in the $t\bar{t}H(b\bar{b})$ analyses is minute. The combination with the other analyses greatly increases the acceptance for $t\bar{t}H$ events. The appropriate correlation of uncertainties between the analyses leads to reductions of systematic uncertainties since the different analyses provide different control regions for the measurement of the corresponding nuisance parameters. In addition, the reliance on the description of the $t\bar{t}$ + heavy flavor background and the associated large uncertainties is reduced.

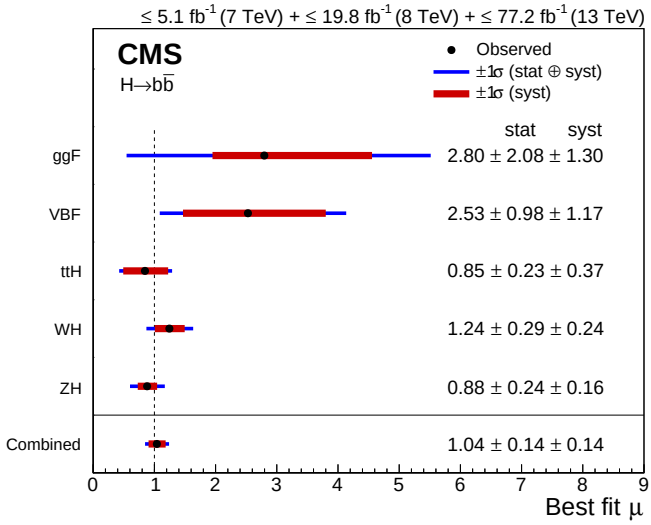
The result corresponds to an observed (expected) significance of 5.2σ (4.2σ) above the background-only hypothesis and therefore by convention establishes the observation of $t\bar{t}H$ production by the CMS collaboration. The results of the different channels included in this combination are shown in Fig. 9.9a and the distribution of the profile likelihood test statistic is shown in Fig. 9.9b.

The $t\bar{t}H(b\bar{b})$ analyses of data recorded in the years 2016 and 2017 have also been included in an analysis studying the coupling constants of the Higgs boson, which is described in Ref. [122]. Several analyses performed with data recorded in the years 2016, 2017 and 2018 at a center-of-mass energy of 13 TeV and with integrated luminosities of up to 137fb^{-1} are combined. The included analyses target a multitude of different production and decay channels of the Higgs boson. This includes analyses targeting $t\bar{t}H$ production as well as production of a Higgs boson via gluon-gluon fusion, vector boson fusion or production in association with a vector boson. Targeted decay channels include $H \to b\bar{b}$, $H \to \gamma\gamma$, $H \to \tau^+\tau^-$, $H \to WW^*$, $H \to ZZ^*$ and $H \to \mu^+\mu^+$. The combined fit results in a signal-strength modifier of

$$\hat{\mu}_{t\bar{t}H} = 1.14^{+0.21}_{-0.20}(\text{tot.})\left[^{+0.13}_{-0.13}(\text{stat.})^{+0.17}_{-0.15}(\text{syst.})\right] \tag{9.10}$$

which is an improvement compared to the previous combination (Eq. 9.9). The combination also results in a best-fit value for the coupling of the Higgs boson to the top quark κ_t

$$\hat{\kappa}_t = 1.01^{+0.11}_{-0.11}(\text{tot.})\left[^{+0.06}_{-0.06}(\text{stat.})^{+0.09}_{-0.08}(\text{syst.})\right] \tag{9.11}$$



(a)

Figure 9.10: Best-fit value of the signal-strength modifier for the $H \to b\bar{b}$ process for the five individual production modes considered, as well as the overall combined result. The total and systematic uncertainties are indicated by blue and red uncertainty bands. Taken from Ref. [269].

expressed in terms of the κ-framework, which is discussed in Ch. 2 and Ref. [268]. For the standard model κ_t takes on a value of unity, making the result therefore compatible.

9.3.2 $H \to b\bar{b}$ combination

The analyses in the SL and DL channels of data recorded in the year 2016 have also been combined with other analyses targeting the $H \to b\bar{b}$ decay channel resulting in the observation of the $H \to b\bar{b}$ process, as described in Ref. [269]. Besides the SL and DL $t\bar{t}H(b\bar{b})$ analyses, this combination includes their corresponding pendants performed on data recorded at center-of-mass energies of 7 TeV and 8 TeV as well as analyses targeting production of a Higgs boson via gluon-gluon fusion, vector boson fusion or in association with a vector boson, which have been performed at center-of-mass energies of 7 TeV, 8 TeV and 13 TeV (years 2016 and 2017). The combination results in a signal-strength modifier for the $H \to b\bar{b}$ process of

$$\hat{\mu}_{H\to b\bar{b}} = 1.04^{+0.20}_{-0.20}(\text{tot.})\,[\,\pm 0.14(\text{stat.})\ \ \pm 0.14(\text{syst.})] \tag{9.12}$$

corresponding to an observed (expected) significance of 5.6σ (5.5σ) above the background-only hypothesis. The results for all included analysis channels are summarized in Fig. 9.10. The most sensitive contributing channels are the WH and ZH channels followed by the $t\bar{t}H$ channel.

The coupling constant measurement described in the previous Sec. 9.3.1 and Ref. [122] also results in an estimate for the $H \to b\bar{b}$ signal-strength modifier with a best-fit value of

$$\hat{\mu}_{H\to b\bar{b}} = 1.11^{+0.20}_{-0.19}(\text{tot.})\left[\pm 0.13(\text{stat.})^{+0.16}_{-0.15}(\text{syst.})\right]. \tag{9.13}$$

While still compatible with the previous $H \to b\bar{b}$ result (and the standard model) it is noticeable that the result of this combination has the same sensitivity as the direct $H \to b\bar{b}$ measurement, which included more channels and data, for example data recorded in both the years 2016 and 2017 instead of only the year 2016. The direct measurement does,

however, not include data recorded at 7 TeV and 8 TeV for the analyses in the VH channels, which in the previous $H \to b\bar{b}$ combination increased the observed (expected) significances for only the VH channel from 4.4σ (4.2σ) to 4.8σ (4.9σ). Together with the observation that the VH channels are important for the sensitivity to $H \to b\bar{b}$ this could explain the similar sensitivities.

9.4 Comparison with other analyses

The $t\bar{t}H(b\bar{b})$ results of this thesis are further compared to other measurements of $t\bar{t}H(b\bar{b})$ production.
The coupling constant measurement by the CMS Collaboration described in Ref. [122] included the $t\bar{t}H(b\bar{b})$ analyses described in this thesis in its combined fit and measured the signal-strength modifier for $t\bar{t}H(b\bar{b})$ production as

$$\hat{\mu}_{t\bar{t}H(b\bar{b}),\mathrm{CMS}} = 1.13^{+0.33}_{-0.30}(\mathrm{tot.})\left[\pm 0.16(\mathrm{stat.})^{+0.29}_{-0.25}(\mathrm{syst.})\right]. \tag{9.14}$$

The small difference compared to the result of the direct $t\bar{t}H(b\bar{b})$ measurement presented above (Eq. 9.8) is due to the combination with analyses targeting either $t\bar{t}H(\text{non-}b\bar{b})$ or $H \to b\bar{b}$ decays with other production channels. As was discussed, in this study the couplings to the bottom quark and the top quark are both measured simultaneously. Therefore the measurement of the these couplings in other production and decay channels influences the result for $t\bar{t}H(b\bar{b})$ production.

The $t\bar{t}H(b\bar{b})$ signal strength has also been measured by the ATLAS Collaboration in the SL and DL channels using data recorded in the year 2016 at a center-of-mass energy of 13 TeV with an integrated luminosity of 36.1 fb^{-1}, as described in Ref. [266], resulting in a signal-strength modifier of

$$\hat{\mu}_{t\bar{t}H(b\bar{b}),\mathrm{ATLAS}} = 0.84^{+0.64}_{-0.61}(\mathrm{tot.})\left[\pm 0.29\ (\mathrm{stat.})^{+0.57}_{-0.54}(\mathrm{syst.})\right]. \tag{9.15}$$

In this analysis, the signal is separated from the background using boosted decision trees (BDTs) instead of ANNs as in this thesis. Input features for the BDTs include a discriminant calculated with a matrix element method and observables calculated from a BDT based reconstruction of the $t\bar{t}H$ system. In the fit model of the $t\bar{t}H(b\bar{b})$ analysis of this thesis, the $t\bar{t}+$heavy flavor background is modeled by assigning very large normalization uncertainties to each of four $t\bar{t}+b\bar{b}$, $t\bar{t}+b$ $t\bar{t}+2b$ and $t\bar{t}+c\bar{c}$ subprocesses. In the analysis by the ATLAS Collaboration, the background is modeled by introducing unconstrained parameters for the normalization of the $t\bar{t}+c\bar{c}$ and $t\bar{t}+\geq 1b$ processes and additional nuisance parameters that can change the shapes of these processes are derived from comparison between different simulations, thereby allowing the nuisance parameters to morph to shape predicted from the nominal simulation to that of another simulation. The relative fractions of $t\bar{t}+b\bar{b}$, $t\bar{t}+b$, $t\bar{t}+2b$ and $t\bar{t}+\geq 3b$ events within the $t\bar{t}+\geq 1b$ process are controlled by nuisance parameters derived through variation of the renormalization and factorization scales of the nominal simulation and differences between PDF sets.

This analysis had then been combined with measurements of other Higgs boson production and decay channels, including $tH(b\bar{b})$, performed on data recorded in the years 2016, 2017 and 2018, described in Refs. [270, 271], resulting in a value for the signal-strength modifier for the combined $t\bar{t}H(b\bar{b}) + tH(b\bar{b})$ production of

$$\hat{\mu}_{t\bar{t}H(b\bar{b})+tH(b\bar{b}),\mathrm{ATLAS}} = 0.79^{+0.60}_{-0.59}(\mathrm{tot.})\left[\pm 0.29\ (\mathrm{stat.})^{+0.52}_{-0.51}(\mathrm{syst.})\right]. \tag{9.16}$$

Since $t\bar{t}H(b\bar{b}) + tH(b\bar{b})$ production involves two processes, it cannot be directly compared to a measurement of $t\bar{t}H(b\bar{b})$ production alone. However, both processes are sensitive

to the top-quark Yukawa coupling, the tH($\mathrm{b\bar{b}}$) process even to the sign. In the case of the standard model coupling structure, a deviation from the predicted magnitude of the coupling constant should affect the production cross sections of both processes in a related way. In practice, analyses of tH($\mathrm{b\bar{b}}$) and $\mathrm{t\bar{t}H(b\bar{b})}$ often feature a sizable acceptance for the respective other process.

As previously discussed, the $\mathrm{t\bar{t}H(b\bar{b})}$ analyses had been combined with other Higgs boson measurements, described in Ref. [122], yielding a measurement of the coupling of the Higgs boson to the top quark κ_t of

$$\hat{\kappa}_{\mathrm{t;CMS,t\bar{t}H}} = 1.01^{+0.11}_{-0.11}(\mathrm{tot.}) \left[^{+0.06}_{-0.06}(\mathrm{stat.})^{+0.09}_{-0.08}(\mathrm{syst.})\right]. \tag{9.17}$$

For this measurement the tH process was fixed to its standard model prediction since no dedicated analysis categories were included. Furthermore, it was assumed that no contributions from physics beyond the standard model exist, and the loop contributions occurring in ggH, $\mathrm{H} \to \mathrm{gg}$ and $\mathrm{H} \to \gamma\gamma$ were treated as effective vertices parameterized by effective coupling strength modifiers. A comparable analysis with similar assumptions by the ATLAS Collaboration is given in Ref. [271] where a coupling constant of

$$\hat{\kappa}_{\mathrm{t;ATLAS,t\bar{t}H+tH}} = 1.0^{+0.12}_{-0.12} \tag{9.18}$$

was found. In this analysis, the tH process is included and the sign of κ_t is allowed to be negative. A negative sign is, however, excluded with an observed (expected) significance of 2.9σ (2.7σ) standard deviations. Another measurement of κ_t was provided by an analysis by the CMS Collaboration performed on data recorded in the year 2016 targeting tH production, as described in Ref. [272]. Here, the coupling of the Higgs boson to vector bosons was fixed to its standard model value $\kappa_\mathrm{V} = 1$ and contribution from $\mathrm{t\bar{t}H}$ events were considered. As can be seen in Fig. 9.11, the profile likelihood test statistic has two minima approximately corresponding to values $\kappa_{\mathrm{t;CMS,t\bar{t}H+tH,neg.}} = -0.7^{+0.2}_{-0.2}$ and $\kappa_{\mathrm{t;CMS,t\bar{t}H+tH,pos.}} = 1.5^{+0.6}_{-0.5}$. The positive value is favored over the negative value by about 1.5 standard deviations.

Recently, a novel method to access the top-Higgs Yukawa coupling constant y_t was proposed based on the observation that near the threshold for production of top quark-antiquark pairs, the cross section for $\mathrm{t\bar{t}}$ production is sensitive to higher order corrections involving the coupling of the Higgs boson to top quarks, and therefore the Yukawa coupling constant. This is further discussed in Refs. [273–275].
The "traditional" coupling constant measurements in the κ framework have the disadvantage that they need to consider assumptions about the coupling of the Higgs boson to all other fermions and vector bosons, which for the $\mathrm{t\bar{t}}$ production studied in this method only contribute via even higher order corrections [1]. Analyses based on this approach have been carried out by the CMS Collaboration in the form of double-differential cross section measurements of the $\mathrm{t\bar{t}}$ cross section as a function of the invariant $\mathrm{t\bar{t}}$ mass and the pseudorapidity between the top quark-antiquark pair, as described in Refs. [277, 278]. For better comparability, with other analyses the measured quantity is defined as the ratio of the observed coupling constant to its standard model value of $Y_\mathrm{t} = y_\mathrm{t}/y_{\mathrm{t,SM}}$ and is therefore equal to κ_t. In the first analysis performed on data recorded in the year 2016 the coupling constant modifier was found to be $Y_\mathrm{t} = 1.07^{+0.34}_{-0.43}$, while in the second analysis performed on data recorded in the years 2016, 2017 and 2018 it was found to be

$$Y_\mathrm{t} = 1.16^{+0.24}_{-0.35}. \tag{9.19}$$

[1] Limits on y_t independent of the other couplings can also be derived by analyzing the production of four top quarks, to which $\mathrm{t\bar{t}H}, \mathrm{H} \to \mathrm{t\bar{t}}$ is a contribution, as is done in Ref. [276]. This a rare process with a very small cross section, which can however be increased by physics beyond the standard model.

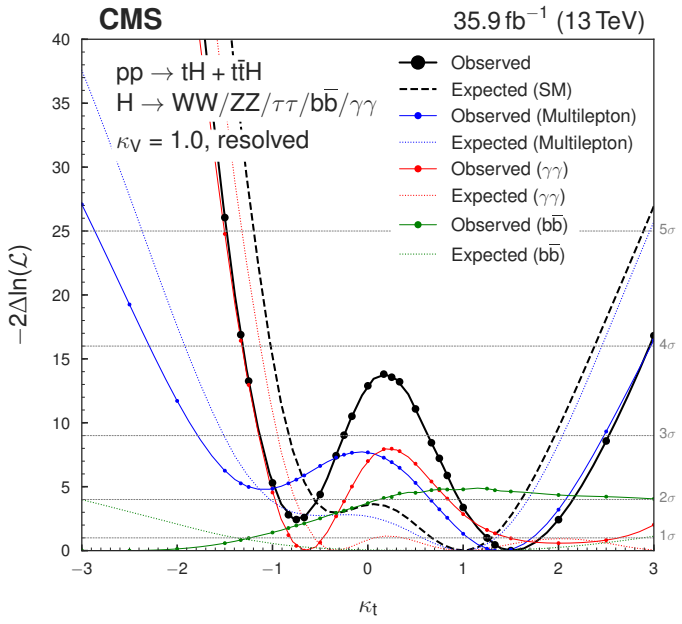


Figure 9.11: Scan of the profile likelihood test statistic as a function of κ_t with a fixed $\kappa_V = 1$. The observed test statistic is shown as solid black line. Also shown are observed test statistics of individual contributing channels and values expected from simulation. Taken from Ref. [272].

While these results are not as sensitive as those obtained by the coupling measurements in Higgs boson production and decays, they are compatible and more importantly derived using an independent method, serving as important cross check. In summary, several measurements of $t\bar{t}H(b\bar{b})$ production and of the coupling of the Higgs boson to the top quark have been performed, all showing compatibility with the standard model value of the top-Higgs Yukawa coupling.

10 The future of the $t\bar{t}H(b\bar{b})$ analysis

Analyses of the data recorded during LHC Run 2, i.e. in the years 2016, 2017 and 2018, have led to the observation of the $t\bar{t}H$ production and $H \to b\bar{b}$ decay processes. In addition, measurements of the signal-strength modifier for the $t\bar{t}H(b\bar{b})$ process and the Yukawa coupling constant between the Higgs boson and the top quark have been performed. The precision of these measurements is limited by the amount of available data, the analysis methods employed, as well as experimental and theory uncertainties. As was seen, the measured values for the Yukawa coupling constant and the $t\bar{t}H(b\bar{b})$ signal-strength modifier are well compatible with the standard model predictions, thereby reducing the parameter space of models of physics beyond the standard model. In the continuing search for physics beyond the standard model it is therefore now of interest to further increase the precision of such measurements of the Higgs sector of the SM.

To achieve this goal, the LHC and accompanying experiments will be upgraded to reach higher center-of-mass energies and higher instantaneous luminosities. In the first step, currently underway, the center-of-mass energy will be increased to 14 TeV. The recording of data is planned to resume in 2022 with the goal to record about $300\,\text{fb}^{-1}$ of data. Subsequently, the LHC will be upgraded again to the *High Luminosity LHC* (HL-LHC), reaching an instantaneous luminosity of $5 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, as described in Ref. [279]. It is planned for the HL-LHC to resume operations in the year 2027 with the ultimate goal to accumulate $4000\,\text{fb}^{-1}$ of data after several years of run-time. The increased instantaneous luminosity presents a challenging experimental environment with up to 200 proton-proton collisions per bunch crossing. This necessitates extensive upgrades to the detectors of the experiments, which for the CMS detector are described in Refs. [280–284]. Studies of the object performance for the future CMS detector are discussed in Ref. [285].

The prospects of Higgs physics analyses at the HL-LHC have been studied extensively, as detailed in Ref. [286] [1]. This includes extrapolations of existing analyses to higher integrated luminosities taking into account expected improvements to systematic uncertainties. In Sec. 10.1, such projections for the measurement of $t\bar{t}H(b\bar{b})$ production and the top-Higgs Yukawa coupling based on the $t\bar{t}H(b\bar{b})$ analyses in the SL and DL channels of data recorded in the year 2016 are discussed.

Sensitivity to $t\bar{t}H(b\bar{b})$ production cannot only be gained through increased integrated luminosities, but also through improvements to the analysis strategy itself. After the

[1] The report also includes studies of Higgs physics at a proposed 27 TeV hadron collider called the High Energy LHC.

publication of the results of the $t\bar{t}H(b\bar{b})$ analyses in Refs. [4, 5] the analysis of the data recorded in the year 2018 began. In addition, the data recorded in the years 2016 and 2017 is currently being re-analyzed to not only benefit from improvements to the reconstruction and calibration of the physics objects, but to also capitalize on such improvements to the analysis strategy. Several studies on possible improvements are discussed in Sec. 10.2.

10.1 Projection studies

In the following, projection studies of analysis of $t\bar{t}H(b\bar{b})$ production are discussed. In Sec. 10.1.1 the methodology of these studies is described. This is followed by a discussion of the obtained results in Sec. 10.1.2.

10.1.1 Method

Since the HL-LHC will have a center-of-mass energy of 14 TeV and the CMS detector will have different properties any prediction of attainable results would ideally encompass simulation of collisions at a center-of-mass energy of 14 TeV recorded with a simulation of the future CMS detector, followed by a complete analysis of this data. The production of such data would however be prohibitively time consuming. Therefore the projections of the $t\bar{t}H(b\bar{b})$ analysis to higher integrated luminosities are instead performed by scaling the template histograms of kinematic distributions of the simulated data of the existing analysis of the data recorded in the year 2016 (Ref. [4]), originally derived for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$, to higher integrated luminosities. In addition, several assumptions are made on the evolution of the systematic uncertainties to account for the changes of the CMS detector. For each studied value of integrated luminosity, pseudodata is constructed by summing the scaled background distributions and additionally adding the equally scaled signal distribution for a standard model value of $\mu_{t\bar{t}H(b\bar{b})} = 1$. The scaled template histograms are then fitted to the pseudodata and the uncertainty of the resulting signal-strength modifier is interpreted as a measure of the sensitivity. With this procedure the statistical component of the total uncertainty is already taken into account and is expected to scale with $1/\sqrt{R}$ where R is the ratio of the targeted integrated luminosity divided by $35.9\,\mathrm{fb}^{-1}$. Results are calculated for two separate scenarios. In the first one, called "Run 2 systematic uncertainties" or "S1", all systematic uncertainties remain as they were for the original $t\bar{t}H(b\bar{b})$ analysis. In the second scenario, called "YR18 systematic uncertainties" or "S2", the systematic uncertainties are reduced based on studies and recommendations by theoreticians and the ATLAS and CMS Collaborations. Theory uncertainties are reduced by a factor of two motivated by the assumption that higher order calculations will be feasible in the future, as discussed in Ref. [286], and by an expected improvement to the measurement of parton density functions with HL-LHC data, as discussed in Ref. [287]. The treatment of experimental uncertainties is based on studies performed by relevant detector and object experts of the CMS Collaboration, taking into account the future CMS detector, run conditions and expected improvements to reconstruction methods, as discussed in Refs. [285, 286, 288]. Based on this, all systematic uncertainties are scaled with $1/\sqrt{R}$ down to a minimum value motivated by the assumption that most experimental uncertainties are derived using recorded data. The minimal values of the experimental uncertainties are given in Tab. 10.1. As a further simplification it is assumed that the size of all simulated samples will be sufficiently large and corresponding uncertainties are not considered for the projection studies.

Table 10.1: Sources of systematic uncertainty, their size during LHC Run 2 and the floor to which they are scaled in the projection scenario S2. Taken from Ref. [288].

Source	Component	Run 2 uncertainty	Projection minimum uncertainty
Muon ID		1–2%	0.5%
Electron ID		1–2%	0.5%
Photon ID		0.5–2%	0.25–1%
Hadronic tau ID		6%	2.5%
Jet energy scale	Absolute	0.5%	0.1–0.2%
	Relative	0.1–3%	0.1–0.5%
	Pileup	0–2%	Same as Run 2
	Method and sample	0.5–5%	No limit
	Jet flavour	1.5%	0.75%
	Time stability	0.2%	No limit
Jet energy res.		Varies with p_T and η	Half of Run 2
MET scale		Varies with analysis selection	Half of Run 2
b-Tagging	b-/c-jets (syst.)	Varies with p_T and η	Same as Run 2
	light mis-tag (syst.)	Varies with p_T and η	Same as Run 2
	b-/c-jets (stat.)	Varies with p_T and η	No limit
	light mis-tag (stat.)	Varies with p_T and η	No limit
Integrated lumi.		2.5%	1%

10.1.2 Result

Projection of the $t\bar{t}H(b\bar{b})$ measurement

The results for the projection of the uncertainty of the $t\bar{t}H(b\bar{b})$ measurement are shown for both scenarios in Fig. 10.1. Shown are the total uncertainty as well as the contribution of following sources and groups:

1. **Stat**: statistical uncertainty;
2. **SigTh**: theory uncertainties related to the normalization and simulation of the predicted $t\bar{t}H$ signal events;
3. **BkgTh**: theory uncertainties related to the normalization and simulation of the predicted background events. This includes the additional $t\bar{t}$ + heavy flavor 50% normalization uncertainties;
    a) **Add. tt+HF XS**: additional 50% normalization uncertainties on the $t\bar{t}$ + heavy flavor processes, previously discussed in Ch. 8.2;
4. **Exp**: experimental uncertainties;
    a) **Luminosity**: uncertainty on the measurement of the integrated luminosity;
    b) **B tagging**: uncertainties related to b tagging;
    c) **JES**: uncertainties related to the jet energy scale.

All other uncertainties, i.e. uncertainties related to identification and reconstruction of leptons, the trigger efficiencies and the jet energy resolution, have only a minor effect on the result and are not shown as separate groups. The contribution of the uncertainty groups in percent relative to the fitted signal-strength modifier $\mu_{t\bar{t}H(b\bar{b})} = 1$ are shown in Tab. 10.2. For the scenario S1 the total uncertainty is reduced from 0.49% for an integrated luminosity of 35.9 fb^{-1} to 0.11% for an integrated luminosity of 3000 fb^{-1} and for scenario S2 it is reduced from 0.46% to 0.07% for the same integrated luminosities, or about 10% in both cases [2]. The difference between the results for scenarios S1 and S2 for the integrated

[2] The result for an integrated luminosity of 35.9 fb^{-1} and scenario S1 slightly differs from that presented in Ch. 9 due to the removal of the uncertainties related to the size of the simulated samples.

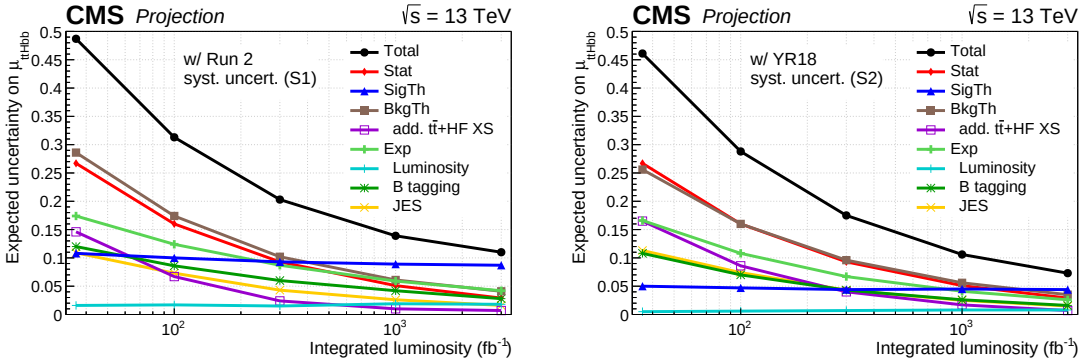
luminosity of $35.9\,\mathrm{fb}^{-1}$ is caused by the reduction of the theory uncertainties by a factor of two in scenario S2 [3]. As can be seen in Fig. 10.1 the statistical uncertainty decreases with increasing integrated luminosity as expected. The contributions of the experimental uncertainties and the uncertainty due to the modeling of the background similarly decrease with increasing integrated luminosity due to the ability of the fit to better constrain the related nuisance parameters. The uncertainty related to JES behaves similar for both scenarios, indicating that the reduced JES components do not have a large effect on the final result. The uncertainty on the integrated luminosity is decreased between scenarios S1 and S2 but does, as expected, not further decrease with rising integrated luminosities since the analysis lacks the power to constrain this already small uncertainty. In the previous chapter it was discussed that the $t\bar{t}H(b\bar{b})$ analysis is well able to constrain b-tagging uncertainties due to the importance of b tagging in the analysis, which is also visible here. Furthermore, the contribution of the b-tagging uncertainties decreases between scenarios S1 and S2, showing that the $t\bar{t}H(b\bar{b})$ analysis can benefit from improvements to the precision of b tagging.

The analysis is also highly sensitive to the modeling of the background, especially of the $t\bar{t}$ + heavy flavor processes. While the background modeling is the dominant uncertainty for the integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ it can be strongly constrained for higher integrated luminosities. With an integrated luminosity of $3000\,\mathrm{fb}^{-1}$ the uncertainty related to the $t\bar{t}$ + heavy flavor modeling is constrained to less than 1% in both scenarios. This indicates that the result is sensitive to even minute mismodelling of the $t\bar{t}$ + heavy flavor processes and that the current approach with the additional 50% normalization uncertainties will be insufficient. With an integrated luminosity of $300\,\mathrm{fb}^{-1}$ the $t\bar{t}$ + heavy flavor uncertainties are already constrained to several percent, motivating the introduction of such an improved model already at this point. Motivated by, amongst other things, this projection study, a new background model is being developed by the CMS Collaboration for an analysis of the data recorded in the year 2018, and reanalysis of the data recorded in the years 2016 and 2017 as well as in preparation for LHC Run 3. By construction, the uncertainty related to the theory modeling is halved going from scenario S1 to scenario S2 and its constraints do not strongly increase with integrated luminosity. This is expected since the uncertainties are directly correlated to the unconstrained signal-strength modifier. The signal model uncertainties become dominant uncertainties at an integrated luminosity of $300\,\mathrm{fb}^{-1}$ in scenario S1 and at an integrated luminosity of $3000\,\mathrm{fb}^{-1}$ in scenario S2. Therefore, with higher integrated luminosity, measurements of $t\bar{t}H(b\bar{b})$ become limited by the prediction of the signal process and could be improved with more precise predictions.

With the presented results it can be roughly estimated that for both scenarios, the $t\bar{t}H(b\bar{b})$ process could be observed with about five to six standard deviations above the background-only hypothesis with an integrated luminosity of $300\,\mathrm{fb}^{-1}$.

This study does not take into account improvements to the analysis methodology itself. As discussed in this thesis, the analysis of the data recorded in the year 2017 is performed with different ANNs, an improved b-tagging algorithm and improvements to the uncertainty model, namely to the uncertainties related to the initial- and final-state radiation simulation. The expected result for the combination of the SL and DL channels of the analysis of data recorded in the year 2017 with an integrated luminosity of $41.5\,\mathrm{fb}^{-1}$ is $\hat{\mu}_{\mathrm{SL+DL,2017,exp.}} = 1.00^{+0.40}_{-0.37}$. The uncertainty of this result is smaller than the uncertainties of about 0.46 and 0.44 projected for scenarios S1 and S2, as can be seen in Fig. 10.1a. For the combination of the SL and DL analysis of data recorded in the years 2016 and 2017, corresponding to a total integrated luminosity of $77.4\,\mathrm{fb}^{-1}$, the expected result is $\hat{\mu}_{\mathrm{SL+DL,2016+2017,exp.}} = 1.00^{+0.33}_{-0.30}$

[3] The differences of the experimental uncertainties are caused by effective correlations of the nuisance parameters.



(a) (b)

Figure 10.1: Expected uncertainties on the $t\bar{t}H(b\bar{b})$ signal-strength modifier $\mu_{t\bar{t}H(b\bar{b})}$ as a function of the integrated luminosity for: (a) uncertainty scenario S1 with "Run 2 systematic uncertainties" and (b) uncertainty scenario S2 with "YR18 systematic uncertainties". Shown are the total uncertainty in black and different groups of uncertainties. Taken from Ref. [288].

and is also below the value predicted by the projection. This consideration illustrates that additional sensitivity can be gained not only by an increase in the size of the analyzed sample but also by an improvement of the analysis methods. How the observed improvement ultimately translates to the precision with an integrated luminosity of 3000 fb^{-1} is not studied since the projection study was performed before the analysis of the data recorded in the year 2017 was performed. In addition, it is likely that additional improvements to methodology will be developed for the analysis of the full HL-LHC data.

Table 10.2: Contribution of different groups of uncertainties to the total expected uncertainty on the $t\bar{t}H(b\bar{b})$ signal-strength modifier $\mu_{t\bar{t}H(b\bar{b})}$ at different integrated luminosities for uncertainty scenarios S1 (Run 2 systematic uncertainties) and S2 (YR18 systematic uncertainties). The uncertainties are quoted in percent relative to $\mu_{t\bar{t}H(b\bar{b})} = 1$. Taken from Ref. [288].

	S1			S2		
Source	35.9 fb^{-1}	300 fb^{-1}	3000 fb^{-1}	35.9 fb^{-1}	300 fb^{-1}	3000 fb^{-1}
Total	48.7	20.4	11.1	46.1	17.6	7.3
Stat	26.7	9.3	2.9	26.7	9.3	2.9
SigTh	10.8	9.3	8.7	5.0	4.5	4.4
BkgTh	28.6	10.3	4.1	25.6	9.6	3.5
Add. tt+HF XS	14.6	2.6	0.8	16.5	4.1	0.7
Exp	17.4	8.7	4.2	16.6	6.7	2.6
Luminosity	1.6	1.8	1.8	0.5	0.7	0.8
B tagging	12.0	6.1	2.8	10.8	4.4	1.6
JES	10.9	4.5	1.6	11.3	4.4	1.6

Projection of the Higgs combination

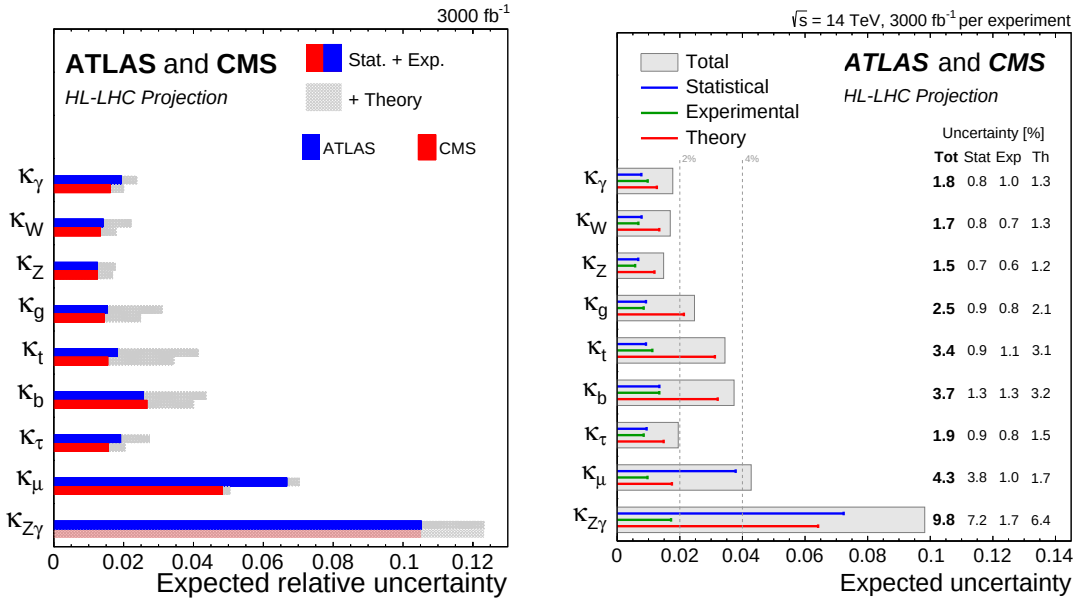
The SL, DL and FH channel $t\bar{t}H(b\bar{b})$ analyses of the data recorded in the year 2016 have been included by the CMS Collaboration in a combination with other analyses of data recorded in the year 2016 targeting other Higgs production and decay channels. The projection of the combination is described in Ref. [288] and is based on the combination

described in Ref. [289]. It was performed in a similar manner as the combination previously discussed in Sec. 9.3. A noteworthy detail of the combination is that the nuisance parameters of the additional 50% uncertainties on $t\bar{t}$ + heavy flavor production were not allowed to be constrained by more than a factor of 2 (3) for the scenarios S1 (S2) to reflect the expected theory improvement. This is therefore a more conservative approach than chosen for the $t\bar{t}H(b\bar{b})$ projection. The combination results in an estimated uncertainty on the signal-strength modifier for $t\bar{t}H$ production of 14.7% (10.7%) for an integrated luminosity $300\,\text{fb}^{-1}$ and scenario S1 (S2). For an integrated luminosity of $3000\,\text{fb}^{-1}$ the uncertainty is reduced to 9.9% for scenario S1 and 6.2% for scenario S2. The combination also predicts precisions of 8% and 6% for the measurement of κ_t in the scenarios S1 and S2 with an integrated luminosity $300\,\text{fb}^{-1}$ and 5.5% and 3.5% with an integrated luminosity of $3000\,\text{fb}^{-1}$. Results for the signal-strength modifiers of other production and decay channels as well as other κ-framework coupling parameters can be found in Ref. [289].

This combination by the CMS Collaboration has been further combined with a similar combination by the ATLAS Collaboration to obtain estimates of the ultimate precision reachable by the HL-LHC program, as described in Ref. [286]. In the CMS-only combination the uncertainty of one parameter of interest was derived with a multi-dimensional fit of all observables at once, whereby all other unconstrained parameters were profiled. The combination of the CMS and ATLAS results were however calculated one parameter at a time using the BLUE method described in Ref. [290]. Taking into account correlations between parameters was found to underestimate the uncertainties with respect to a future scenario in which the ATLAS and CMS Collaborations will have agreed on a more consistent treatment of theory uncertainties. Therefore a more conservative approach was chosen in which such correlations were not taken into account. The experimental and statistical components of the uncertainty of a parameter of interest were treated as completely uncorrelated between both experiments, while the signal and background theory uncertainties were treated as fully correlated. The results by the ATLAS Collaboration were calculated for a center-of-mass energy of 14 TeV, while those of the CMS Collaboration were calculated for a center-of-mass energy of 13 TeV. The effect of this on the combined results was found to be small.

For an integrated luminosity of $3000\,\text{fb}^{-1}$ and scenario S2 the combination finds an estimated precision for the $t\bar{t}H(b\bar{b})$ cross section of 10.7% dominated by theory uncertainties. This result interestingly is worse than that obtained by the projection of the SL and DL channels described above. This difference could be caused by several factors. The result by the ATLAS Collaboration (14.2%) is slightly worse than the standalone result of the CMS Collaboration (10.8%). In addition the result of the CMS Collaboration included the analysis in the FH channel and employs a different treatment of nuisance parameters such as those for the $t\bar{t}$ + heavy flavor normalization. For the inclusive $t\bar{t}H$ production cross section the combination finds a precision of 4.3%, again dominated by theory uncertainties. Finally, results are also calculated for the parameters of the κ-framework and shown in Fig. 10.2. The predicted precision of the measurement of κ_t for an integrated luminosity of $3000\,\text{fb}^{-1}$ and scenario S2 is found to be 3.4%, which is dominated by theory uncertainties. Since at this point the theory uncertainty clearly dominates, the combination of the ATLAS and CMS measurements does not lead to a significant increase of sensitivity.

In summary, for a realistic scenario and an amount of data recorded at the HL-LHC corresponding to an integrated luminosity of $3000\,\text{fb}^{-1}$, the $t\bar{t}H(b\bar{b})$ cross section is predicted to be measurable with a precision of about 10%, while the coupling of the Higgs boson to the top quark is predicted to be measurable with a precision of about 3.5%.



(a) (b)

Figure 10.2: Expected uncertainties on the parameters of the κ-framework for scenario S2 with "YR18 systematic uncertainties" and an integrated luminosity of 3000 fb^{-1}. (a) Results by the ATLAS and CMS Collaborations. (b) Results of the combination of the individual results. Taken from Ref. [286].

10.2 Improvements to the $t\bar{t}H(b\bar{b})$ analysis

Multiple studies on possible improvements of future $t\bar{t}H(b\bar{b})$ analyses in the single-lepton channel have been conducted by students of the Karlsruhe Institute of Technology (KIT) and will be briefly discussed in the following. These studies focused on the description of $t\bar{t}$+heavy flavor production, the handling of systematic uncertainties in the statistical model or possible gains to sensitivity through the adoption of more sophisticated multivariate analysis methods or exploration of new phase space regions. In addition, the measurement of $t\bar{t}H(b\bar{b})$ production in the framework of simplified template cross sections, described in Ref. [92], was explored.

10.2.1 $t\bar{t}+b\bar{b}$ modeling

The description of the background due to $t\bar{t}$ + heavy flavor production is of utmost importance to the analysis of $t\bar{t}H(b\bar{b})$ production. In Sec. 2.4 it was discussed, that the description of $t\bar{t}+b\bar{b}$ production should be more accurate using 4FS simulations, in which the additional b jets also originate from the matrix element calculation, that previous measurements of $t\bar{t}+b\bar{b}$ production consistently observed a $t\bar{t}+b\bar{b}$ cross section larger than predicted, and that the uncertainties of measurements as well as theoretical calculations are sizable. As was discussed in Chs. 7 and 8, in the current analyses this background is modeled using simulated events of a 5FS Powheg+Pythia simulation of $t\bar{t}$+jets production, normalized to a NNLO+NNLL production cross section, in which the additional b jets are described by the parton shower simulation. Large normalization uncertainties are introduced for the $t\bar{t}$ + heavy flavor background subprocesses and the parton shower related uncertainties are decorrelated between the subprocesses. It was found that the shapes of the ANN distributions predicted by the current simulation are compatible with predictions made by

other simulations and that this uncertainty treatment is sufficient for the analyses of the data recorded in the years 2016 and 2017. The projection study discussed in the previous section, however, indicates that, for a higher integrated luminosity, this approach might not be flexible enough to account for a mismodelling of the $t\bar{t}$ + heavy flavor background. An improvement to the $t\bar{t}$ + heavy flavor modeling is studied in Ref. [101] in which the $t\bar{t}$+$b\bar{b}$, $t\bar{t}$+2b and $t\bar{t}$+b events in the 5FS sample are replaced by events obtained using a 4FS POWHEG simulation of $t\bar{t}$+$b\bar{b}$ production. The total normalization of the replaced events is kept identical so that this merging procedure only changes the relative fractions of the three subprocesses and distributions of kinematic observables. It is found that this procedure does not change the expected sensitivity of the $t\bar{t}H(b\bar{b})$ analysis while the impact of parton shower related uncertainties is decreased and the impact of scale uncertainties on $t\bar{t}$ + heavy flavor production are increased. This study demonstrates the feasibility of improving the background modeling by using a 4FS simulation for the $t\bar{t}$ + heavy flavor description while retaining a 5FS simulation for the remaining $t\bar{t}$+jets background. This procedure should therefore be adopted by future $t\bar{t}H(b\bar{b})$ analyses.

Since ANNs combine kinematic observables in a highly non-trivial way, differences between the predictions made by different simulations can lead to different ANN discriminant distributions. An ANN trained with, for example, $t\bar{t}$+$b\bar{b}$ events generated with a 5FS simulation could have a different response to $t\bar{t}$+$b\bar{b}$ events in recorded data if those were better described by a 4FS simulation. This could lead to a biased measurement of the signal-strength modifier. In Refs. [291, 292] a method called *adversarial neural networks* is studied to construct ANNs that perform similar on different $t\bar{t}$+$b\bar{b}$ predictions. For this, an additional "adversary" ANN is introduced to distinguish between different $t\bar{t}$+$b\bar{b}$ predictions, in the tested cases the nominal 5FS simulation and an alternative 4FS simulation, based on the output of the classification ANN. If the adversary ANN is able to distinguish between the simulations, the classification ANN produces different results for the two simulations. The training of both ANNs is performed alternatingly and the loss function of the classification ANN is penalized depending on ability of the adversary ANN to predict the $t\bar{t}$+$b\bar{b}$ simulation.
In Ref. [291] it is shown that the adversarial approach leads to more similar distributions for the two $t\bar{t}$+$b\bar{b}$ predictions. For a multiclass ANN, similar to the one used in the $t\bar{t}H(b\bar{b})$ analysis, this study is, however, not able to demonstrate a reduction of the bias on the signal-strength modifier when the predictions made with one $t\bar{t}$+$b\bar{b}$ simulation are fit to pseudodata created with another $t\bar{t}$+$b\bar{b}$ simulation.
In Ref. [292] a simpler binary classification ANN is used to distinguish between $t\bar{t}H$ and $t\bar{t}$+$b\bar{b}$ and it is shown that the adversarial approach leads to a small reduction of the bias on the signal-strength modifier. While the adversarial approach is interesting and well motivated, it requires additional studies before it can be effectively used in $t\bar{t}H(b\bar{b})$ analyses.

10.2.2 ANNs and uncertainties

While the output of the ANNs in this thesis can be interpreted as the probability of an event to belong to a given class, there is no measure as to how certain the calculated output value is, i.e. how large the confidence interval around the output value is. A repetition of the training with different seeds for the random number generators will result in slightly different output values. This randomness could be considered a systematic uncertainty of the method. The output values should also be less certain if the considered events are from a phase space region with a small representation in the training sample. In Refs. [293, 294] a method called *Bayesian neural networks* (BNNs), with which the uncertainty on the output values can be quantified, is studied in the context of a $t\bar{t}H(b\bar{b})$ analysis. In BNNs

each trainable parameter is replaced by probability density function (p.d.f.) with trainable mean and standard deviation, that reflects the uncertainty on each trainable parameter. The output value of a given event is then calculated by sampling a single value from each p.d.f. The central value of the prediction for an event and its associated uncertainty are then obtained as the mean and standard deviation of a distribution of many output values. In Ref. [293] it is shown that this approach leads to uncertainties that reflect the randomness of the model, and that these uncertainties also increase when the training data is different from the data on which the BNN is applied. It is also be possible to train the BNNs with additional events with systematic variations. The application of BNNs for the analysis of $t\bar{t}H(b\bar{b})$ production is interesting, especially since the uncertainties reflect differences in the training and application samples and the possible inclusion of uncertainty variations in the training sample, and should be further studied. How the resulting uncertainties can be used in the $t\bar{t}H(b\bar{b})$ analysis also requires further investigation. Since an ANN is just a function for the calculation of a new observable, a naive addition of this uncertainty to the statistical model might not be correct. It could, however, be useful during the optimization of the analysis strategy.

10.2.3 Improvements to signal and background separation

Several studies have been performed to improve the separation between signal and background events and thereby increase the sensitivity to $t\bar{t}H(b\bar{b})$ production.

Sufficiently large ANNs should in principle be able to separate between signal and background using only simple kinematic observables. The training of such networks is however technically difficult. When using not too complicated and easier to train ANNs, as in the current analysis, the ability to separate between signal and background strongly depends on the chosen input features. A reconstruction of the Higgs boson, the top quarks or even the whole $t\bar{t}H$ system leads to powerful input features [4]. The classic method to reconstruct such particles is the χ^2 method in which different permutations of jets, leptons and the missing transverse momentum are assigned to the targeted final-state particles and the best permutation is chosen based on the observed and expected masses of the involved particles. In Ref. [295] an ANN based reconstruction of the $t\bar{t}$ system is studied. Similar to the χ^2 method, jets, leptons and the missing transverse energy are assigned to the particles expected from a semileptonic $t\bar{t}$ decay. The best permutation is, however, chosen based on an ANN with observables of the reconstructed system used as input features. It is found that the reconstruction efficiency in $t\bar{t}$ events can be improved compared to the χ^2 method by about 17% from an efficiency of 27% to an efficiency of 32%. A similar approach for the reconstruction of the Higgs boson is studied in Ref. [296] where it is found that the reconstruction efficiency in $t\bar{t}H$ events increases by about 100% from about 17% to about 33%. An ANN based reconstruction of the $t\bar{t}H$ system still has to be studied as does the effect of this reconstruction on the full $t\bar{t}H(b\bar{b})$ analysis.

An entirely different approach for the improvement of the multivariate methods is studied in Refs. [297, 298], in whicIn the following, projection studies of analysis of ttH(bb) production are discussed. In Sec. 10.1.1 the methodology of these studies is described. This is followed by a discussion of the obtained results in Sec. 10.1.2.h the issue of the trainability of ANNs using only basic input features is solved using convolutional neural networks (CNNs), which are discussed for example in Ref. [189]. CNNs are usually employed in image processing and based on the idea that the neurons in the first layers of an ANN do not need to consider all pixels of an image (the input features) at once, but only nearby ones and that the learned

[4] The input features of the current $t\bar{t}H(b\bar{b})$ analysis include for example several observables related to the invariant mass of pairs of bottom quarks, i.e. candidates for a Higgs boson.

correlations between nearby pixels, called *filters*, can be useful for different parts of the image. This idea results in the implementation of multiple filters that learn to extract different specific features from a limited set of pixels, for example the detection of an edge in an area of 8×8 pixels, which are then passed on to subsequent layers of the network. This idea can be used in the analysis of events recorded with a particle detector, as discussed for example in Ref. [299] in the context of jet identification. The application of CNNs in the $t\bar{t}H(b\bar{b})$ analysis is explored by interpreting the jets measured in the detector as a picture in the $\eta - \phi$ plane with the momentum or the b-tagging discriminant value being the intensity of a color. Different variations of CNNs with multiple filters and subsequent feed-forward ANNs are trained to separate between $t\bar{t}H$ and $t\bar{t}$ events. The studies find that, for the purpose of the separation between $t\bar{t}H$ and $t\bar{t}$ events, simple feed-forward ANNs perform equally or better to the tested CNNs. Several possible improvements are however suggested in Refs. [297, 298], which could be investigated in further studies.

10.2.4 Analysis in the boosted region

Final-state particles originating from a $t\bar{t}H$ interaction tend to have slightly higher transverse momenta than those originating from a $t\bar{t}$ process. Therefore, by selecting events containing jets with large transverse momenta, a phase space region is constructed that is relativIn the following, projection studies of analysis of ttH(bb) production are discussed. In Sec. 10.1.1 the methodology of these studies is described. This is followed by a discussion of the obtained results in Sec. 10.1.2.ely enriched in $t\bar{t}H$ events. If the required transverse momenta are sufficiently high, the decay products of the top quarks and the Higgs boson are merged into *fat jets* which can be identified and tagged using dedicated substructure algorithms. An analysis in this "boosted" region had been previously performed with data recorded in the year 2016 as described in Refs. [131, 132]. It was found that the resulting phase space contained only a small number of events and that the resulting increase in sensitivity was small compared to the large amount of additional time consumed by the calibration and optimization of the analysis in this region. An analysis in the boosted region might, however, become more interesting with increased integrated luminosity and a higher center-of-mass energy in LHC Run 3. In a study described in Ref. [300], simulated $t\bar{t}H$ and $t\bar{t}$, produced at a center-of-mass energy of 13 TeV, are analyzed. The difference of this study to the previous studies lies in the usage of a different kind of fat jet that is already clustered and calibrated by the CMS Collaboration. It is found that, using optimized jet identification criteria, a signal-to-background ratio of about 8% is reached while in about 18% of the selected events the hadronically decaying top quark and the Higgs boson are identified correctly, thereby demonstrating the feasibility of the boosted region with this new fat jets. While the analysis of the data recorded in the full LHC Run 2 might already benefit of the analysis in the boosted region, it should be definitely considered for analyses during LHC Run 3.

10.2.5 Simplified template cross sections

To further study the Higgs boson it is of interest to measure differential cross sections of various production modes. The ATLAS and CMS Collaborations, as well as theoreticians, agreed to conduct and interpret such measurements in the framework of simplified template cross sections (STXS), which are discussed in Ref. [92]. In differential cross section measurements, the predicted distribution of an observable, for example the reconstructed Higgs boson p_{T}, is usually fit to recorded data and then unfolded to particle or parton level. This can be difficult for measurements of processes such as $t\bar{t}H$ production, which cannot easily reconstruct such distributions and require the excessive use of MVAs to extract the signal, for which the discriminant bins cannot be easily extrapolated to the full phase space

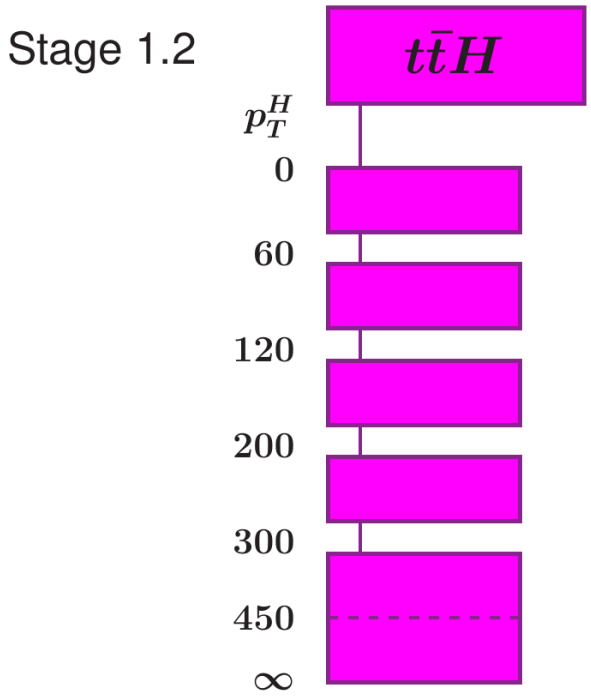


Figure 10.3: The five STXS bins in p_{T} of the Higgs boson at the generator level to be used for a measurement of the differential $\mathrm{t\bar{t}H}$ cross section within the framework of simplified template cross sections. The dashed line in the last bin symbolizes an optional additional bin. Taken from Ref. [301].

or unfolded in a way useful to theoreticians. To still be able to measure unfolded and extrapolated cross sections, the STXS framework is used. In this framework, multiple bins of an observable, for example the Higgs boson p_{T}, are defined on a generator level. The $\mathrm{t\bar{t}H}$ sample is then split into subsamples according to these STXS bins and each sub-sample is treated as separate process with its own unconstrained parameter in the fit of the MVA discriminant distributions to data. This allows the use of the advanced MVA methods while measuring the $\mathrm{t\bar{t}H}$ cross section in different bins of Higgs boson p_{T} at the generator level. The implementation of the STXS framework in the context of the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis is studied in Ref. [301]. Multiple definitions of Higgs boson p_{T} STXS bins and different methods to construct ANNs resulting in discriminant bins enriched in these STXS bins are studied with the goal to optimize the sensitivity to all STXS bins.
These studies served as input to discussions by the LHC Higgs cross section working group (see Ref. [302]) which settled on a STXS binning with five bins in p_{T} of the Higgs boson at the generator level: 0–60 GeV; 60–120 GeV; 120–200 GeV; 200–300 GeV; 300–∞ GeV with an optional additional bin above 450 GeV. This binning scheme is depicted in Fig. 10.3. In Ref. [301] it is found that, for this binning, the best results for all STXS bins are reached by using a combination of two ANNs. The first ANN is a multiclass ANN trained to separate between $\mathrm{t\bar{t}H}$ signal events and background events, as in the current $\mathrm{t\bar{t}H(b\bar{b})}$ analysis, and the second ANN is trained to further separate all events classified as $\mathrm{t\bar{t}H}$ events into the five STXS bins. With this two-ANN approach, the cross sections of the STXS bins are expected to be measured with a precision between 100% and 200%, depending on the STXS bin, with 41.5fb^{-1} of data recorded at a center-of-mass energy of 13 TeV. The studied approach should be used in the analysis of the data recorded in LHC Run 2.

10.3 Summary

It was seen that, with the integrated luminosities available at the HL-LHC, $t\bar{t}H(b\bar{b})$ production and the Yukawa coupling could be measured with precisions of a few percent. An analysis with higher integrated luminosities, however, requires a new and improved model for the prediction of the $t\bar{t}$ + heavy flavor background. Several studies on such an improved background model, as well as general improvements to the analysis strategy have been conducted by students at the Karlsruhe Institute of Technology (KIT). The implementation of several studied improvements is viable and suggests improvements already in the near-term future for analyses with lower integrated luminosities, such as those of the full LHC Run 2 or LHC Run 3.

11 Conclusion

After many decades of experimental searches for the Higgs boson, the last missing piece of the standard model, it was finally discovered in 2012 by the ATLAS and CMS experiments at the LHC. In the standard model bosons, quarks and leptons acquire their masses through the Higgs mechanism. The interaction of the Higgs boson with a quark or lepton is a Yukawa interaction with a coupling strength proportional to the mass of the fermion. Although the Higgs boson itself was discovered, the Yukawa interactions still have to be scrutinized. And although there have been no new particles discovered yet, many open issues remain, such as the nature of dark matter or a unification of the standard model with gravity. Many theories for physics beyond the standard model have been proposed to solve such issues and several of these predict different Yukawa coupling constants or mechanisms for the fermion masses.

Prime candidates for a study of the Yukawa interactions are processes involving top quarks, since the high mass of the top quarks leads to high interaction strengths and cross sections. The process with the highest predicted cross section suitable for direct measurements is the production of a Higgs boson in association with a top quark-antiquark pair ($\mathrm{t\bar{t}H}$). Other suitable processes to study are those involving bottom quarks or τ leptons since those are the fermions with the next highest masses. In the case of the bottom quark, the coupling constant is accessible through the measurement of the branching fraction for $\mathrm{H} \to \mathrm{b\bar{b}}$ decays or the associated production $\mathrm{b\bar{b}H}$, the latter of which is challenging to analyze.

In this thesis, a measurement of $\mathrm{t\bar{t}H(b\bar{b})}$ production with semileptonic $\mathrm{t\bar{t}}$ decays was presented. The analysis was performed with data of proton-proton collisions at a center-of-mass energy of 13 TeV recorded by the CMS experiment at the LHC in the years 2016 and 2017, corresponding to an integrated luminosity of $77.4\,\mathrm{fb}^{-1}$. The data was analyzed using artificial neural networks to suppress backgrounds and construct phase space regions enriched in signal. By comparing the prediction of the standard model to the observed data a signal-strength modifier of $\hat{\mu}_{\mathrm{SL}} = \hat{\sigma}_{\mathrm{SL}}/\sigma_{\mathrm{SL,SM}} = 1.22^{+0.41}_{-0.37}(\mathrm{tot.})\left[{}^{+0.19}_{-0.19}(\mathrm{stat.}){}^{+0.36}_{-0.32}(\mathrm{syst.})\right]$ is found. This value is compatible with the standard model prediction for $\mathrm{t\bar{t}H(b\bar{b})}$ production and corresponds to an observed (expected) significance of 3.3σ (2.7σ) above the background-only hypothesis and therefore evidence for $\mathrm{t\bar{t}H(b\bar{b})}$ production. The measurement is limited by systematic uncertainties, the most important being those related to the prediction of an irreducible background consisting of $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$ production. To further increase the sensitivity, the analysis was combined with corresponding analyses targeting

dileptonic and fully hadronic $t\bar{t}$ decays. This combination found a signal-strength modifier of $\hat{\mu}_{\mathrm{SL+DL+FH}} = 1.15^{+0.32}_{-0.29}(\mathrm{tot.})\left[^{+0.15}_{-0.15}(\mathrm{stat.})^{+0.28}_{-0.25}(\mathrm{syst.})\right]$, corresponding to an observed (expected) significance of 3.9σ (3.5σ) above the background-only hypothesis.

The analysis of $t\bar{t}H(b\bar{b})$ production presented in this thesis has been included in combinations with other analyses of the $H \to b\bar{b}$ or $t\bar{t}H$ processes performed by the CMS Collaboration. The result of this was the observation of the $H \to b\bar{b}$ decay with an observed (expected) significance of 5.6σ (5.5σ) above the background-only hypothesis, as well as the observation of $t\bar{t}H$ production with an observed (expected) significance of 5.2σ (4.2σ) above the background-only hypothesis. In both cases the signal-strength modifiers were compatible with the standard model within one standard deviation.

While the production of $t\bar{t}H$ was observed it is still of interest to observe $t\bar{t}H(b\bar{b})$ production independently. The LHC is expected to continue operations in 2021 with a center-of-mass energy of 14 TeV, starting LHC Run 3. In the year 2027, the LHC is planned to be upgraded to the High-Luminosity LHC and an integrated luminosity of $3000\,\mathrm{fb}^{-1}$ is expected to be recorded after about a decade of runtime. To assess the potential for future $t\bar{t}H(b\bar{b})$ analyses, projection studies were performed. Based on the SL and DL channel analyses of the data recorded in the year 2016, expected results were calculated for different values of integrated luminosities and possible improvements in the description of background processes and experimental uncertainties. It was found that the uncertainty of this $t\bar{t}H(b\bar{b})$ measurement could be expected to be reduced to about 10% with an integrated luminosity of $3000\,\mathrm{fb}^{-1}$. At this point the dominant uncertainties will be related to the prediction of the signal process itself. The now dominant background uncertainties could be constrained to 1%, making the analysis highly sensitive to mismodelling of the background. It was also found that with an integrated luminosity of only $300\,\mathrm{fb}^{-1}$ the $t\bar{t}H(b\bar{b})$ analyses should be able to observe the process with a significance of 5σ, well within the reach of LHC Run 3. These projections did however take into account neither the analysis in the hadronic channel, nor the higher cross sections at 14 TeV or possible improvements to the analysis methods.

In light of these results, several studies performed by students at the Karlsruhe Institute of Technology (KIT) were discussed, showing promising results regarding improvements to the background modeling, the multivariate analysis methods, the analysis of $t\bar{t}H$ with events with high transverse momenta and the interpretation of the results. Several of these improvements will be incorporated in the upcoming analysis of the full data recorded during LHC Run 2. Instead of just measuring a signal-strength, the results of the analysis will be calculated in terms of the simplified template cross section framework, described in Ref. [92], thereby reducing the dependence on the signal model uncertainties and allowing the measurement of cross sections differential in the p_{T} of the Higgs boson.

The modeling of $t\bar{t}$ + heavy flavor production will be improved by using a 4FS scheme $t\bar{t}$+$b\bar{b}$ simulation for its description instead of the 5FS $t\bar{t}$+jets simulation that is currently used. In the future, it could also be explored to replace the simulation with a data driven approach where the background is predicted from side-band regions. The definition of such a side-band region with sufficient amounts of data and purity is however challenging since $t\bar{t}$+$b\bar{b}$ has the same final state as $t\bar{t}H(b\bar{b})$ and is kinematically rather similar.

The studied improvements to the multivariate analysis methods require further research and will not be implemented for the upcoming analysis. The field of artificial intelligence is, however, advancing with significant speed and further research could lead to improvements to b tagging, object calibration, the event reconstruction and the separation of the signal and background processes.

A higher center-of-mass energy in LHC Run 3 could also make an analysis targeting a "boosted" phase space in which the particles have higher transverse momenta more interesting.

Several proposals for future hadron or lepton colliders have been made and with both collider technologies $t\bar{t}H(b\bar{b})$ could be produced, given sufficient energy. A lepton collider would have the advantage that the energy of the initial state could be set and that the final state would be much cleaner due to less hadronic activity.

After the analysis of the full data recorded in LHC Run 2, even more potential for the analysis of the interaction between the top quark and the Higgs boson will be provided by the upcoming LHC Run 3 with a center-of-mass energy of 14 TeV and in the long term with the High Luminosity LHC.

Bibliography

[1] ATLAS Collaboration, "Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[2] CMS Collaboration, "Observation of a New boson at a Mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[3] CMS Collaboration, "A new boson with a bass of 125 GeV observed with the CMS experiment at the Large Hadron Collider", *Science* **338** (2012) 1569–1575, doi:10.1126/science.1230816.

[4] CMS Collaboration, "Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with leptonic $\mathrm{t\bar{t}}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **03** (2019) 026, doi:10.1007/JHEP03(2019)026, arXiv:1804.03682.

[5] CMS Collaboration, "Measurement of $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel in $41.5\,\mathrm{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$", CMS Physics Analysis Summary CMS-PAS-HIG-18-030, 2019. URL:https://cds.cern.ch/record/2675023.

[6] D. J. Griffiths, "Introduction to elementary particles". Wiley, New York, NY, 1987. ISBN 0-471-60386-4 ; 0-471-61544-7.

[7] Particle Data Group, P. A. Zyla et al., "Review of particle physics", *Prog. Theor. Exp. Phys.* **2020** (2020) 083C01, doi:10.1093/ptep/ptaa104.

[8] Y. Fukuda et al., "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562.

[9] Q. R. Ahmad et al., "Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory", *Phys. Rev. Lett.* **87** (Jul, 2001) 071301, doi:10.1103/PhysRevLett.87.071301.

[10] KATRIN Collaboration, "Improved upper limit on the neutrino mass from a direct kinematic method by katrin", *Phys. Rev. Lett.* **123** (Nov, 2019) 221802, doi:10.1103/PhysRevLett.123.221802.

[11] M. Gell-Mann, "Symmetries of baryons and mesons", *Phys. Rev.* **125** (1962) 1067–1084, doi:10.1103/PhysRev.125.1067.

[12] M. Gell-Mann, "A schematic model of baryons and mesons ", *Physics Letters* **8** (1964), no. 3, 214 – 215, doi:http://dx.doi.org/10.1016/S0031-9163(64)92001-3.

[13] J. D. Bjorken and E. A. Paschos, "Inelastic electron-proton and γ-proton scattering and the structure of the nucleon", *Phys. Rev.* **185** (Sep, 1969) 1975–1982, doi:10.1103/PhysRev.185.1975.

[14] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry", *Phys. Rev. D* **2** (Oct, 1970) 1285–1292, doi:10.1103/PhysRevD.2.1285.

[15] J. Bjørken and S. Glashow, "Elementary particles and SU(4)", *Physics Letters* **11** (1964), no. 3, 255 – 257, doi:http://dx.doi.org/10.1016/0031-9163(64)90433-0.

[16] J. E. Augustin et al., "Discovery of a narrow resonance in e^+e^- annihilation", *Phys. Rev. Lett.* **33** (Dec, 1974) 1406–1408, doi:10.1103/PhysRevLett.33.1406.

[17] J. J. Aubert et al., "experimental observation of a heavy particle J", *Phys. Rev. Lett.* **33** (Dec, 1974) 1404–1406, doi:10.1103/PhysRevLett.33.1404.

[18] M. L. Perl et al., "Evidence for anomalous lepton production in $e^+ - e^-$ annihilation", *Phys. Rev. Lett.* **35** (Dec, 1975) 1489–1492, doi:10.1103/PhysRevLett.35.1489.

[19] S. W. Herb et al., "Observation of a dimuon resonance at 9.5 GeV in 400-GeV proton-nucleus collisions", *Phys. Rev. Lett.* **39** (Aug, 1977) 252–255, doi:10.1103/PhysRevLett.39.252.

[20] F. Abe et al., "Observation of top quark production in $\overline{p}p$ collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (Apr, 1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626.

[21] S. Abachi et al., "Observation of the top quark", *Phys. Rev. Lett.* **74** (Apr, 1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632.

[22] K. Kodama et al., "Observation of tau neutrino interactions", *Phys. Lett.* **B504** (2001) 218–224, doi:10.1016/S0370-2693(01)00307-0, arXiv:hep-ex/0012035.

[23] S. Tomonaga, "On a relativistically invariant formulation of the quantum theory of wave fields", *Progress of Theoretical Physics* **1** (1946), no. 2, 27–42, doi:10.1143/PTP.1.27.

[24] R. P. Feynman, "The theory of positrons", *Phys. Rev.* **76** (Sep, 1949) 749–759, doi:10.1103/PhysRev.76.749.

[25] J. Schwinger, "Quantum electrodynamics. I. A covariant formulation", *Phys. Rev.* **74** (Nov, 1948) 1439–1461, doi:10.1103/PhysRev.74.1439.

[26] C. S. Wu et al., "experimental test of parity conservation in beta decay", *Phys. Rev.* **105** (Feb, 1957) 1413–1415, doi:10.1103/PhysRev.105.1413.

[27] N. Cabibbo, "Unitary symmetry and leptonic decays", *Phys. Rev. Lett.* **10** (Jun, 1963) 531–533, doi:10.1103/PhysRevLett.10.531.

[28] M. Kobayashi and T. Maskawa, "CP violation in the renormalizable theory of weak interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.

[29] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π Decay of the K_2^0 Meson", *Phys. Rev. Lett.* **13** (Jul, 1964) 138–140, doi:10.1103/PhysRevLett.13.138.

[30] V. Fanti et al., "A new measurement of direct CP violation in two pion decays of the neutral kaon ", *Physics Letters B* **465** (1999), no. 1–4, 335 – 348, doi:http://dx.doi.org/10.1016/S0370-2693(99)01030-8.

[31] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** (1961), no. 4, 579 – 588, doi:http://dx.doi.org/10.1016/0029-5582(61)90469-2.

[32] A. Salam and J. Ward, "Electromagnetic and weak interactions", *Physics Letters* **13** (1964), no. 2, 168 – 171, doi:http://dx.doi.org/10.1016/0031-9163(64)90711-5.

[33] S. Weinberg, "A model of leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[34] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[35] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[36] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[37] H. Politzer, "Reliable perturbative results for strong interactions?", *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[38] D. J. Gross and F. Wilczek, "Ultraviolet behavior of nonabelian gauge theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[39] K. G. Wilson, "Confinement of quarks", *Phys. Rev. D* **10** (Oct, 1974) 2445–2459, doi:10.1103/PhysRevD.10.2445.

[40] F. Mandl and G. Shaw, "Quantum field theory". Wiley, Chichester, 2. ed. edition, 2010. ISBN 978-0-471-49683-0; 978-0-471-49684-7.

[41] G. 'tHooft, "Renormalization of massless Yang-Mills fields ", *Nuclear Physics B* **33** (1971), no. 1, 173 – 199, doi:http://dx.doi.org/10.1016/0550-3213(71)90395-6.

[42] G. 'tHooft, "Renormalizable Lagrangians for massive Yang-Mills fields ", *Nuclear Physics B* **35** (1971), no. 1, 167 – 188, doi:http://dx.doi.org/10.1016/0550-3213(71)90139-8.

[43] W. de Boer, "The discovery of the Higgs boson with the CMS detector and its implications for supersymmetry and cosmology", in *Time and Matter 2013, Venice, Italy*. 2013. arXiv:1309.0721.

[44] J. Goldstone, A. Salam, and S. Weinberg, "Broken symmetries", *Phys. Rev.* **127** (Aug, 1962) 965–970, doi:10.1103/PhysRev.127.965.

[45] S. L. Glashow, "The renormalizability of vector meson interactions", *Nucl. Phys.* **10** (1959) 107–117, doi:10.1016/0029-5582(59)90196-8.

[46] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[47] H. Yukawa, "On the interaction of elementary particles I", *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48–57, doi:10.1143/PTPS.1.1.

[48] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory". Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.

[49] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641–653. URL:http://www.jetp.ac.ru/cgi-bin/dn/e_046_04_0641.pdf.

[50] V. Gribov and L. Lipatov, "Deep inelastic ep scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450. [Yad. Fiz.15,781(1972)].

[51] G. Altarelli and G. Parisi, "Asymptotic freedom in parton language", *Nucl. Phys. B* **126** (1977) 298–318, doi:10.1016/0550-3213(77)90384-4.

[52] J. Gao, L. Harland-Lang, and J. Rojo, "The structure of the proton in the LHC precision era", *Phys. Rept.* **742** (2018) 1–121, doi:10.1016/j.physrep.2018.03.002, arXiv:1709.04922.

[53] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[54] P. Skands, "Introduction to QCD", in *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pp. 63–124. 2017. arXiv:1207.2389. doi:10.1142/9789814525220_0008.

[55] V. Sudakov, "Vertex parts at very high-energies in quantum electrodynamics", *Sov. Phys. JETP* **3** (1956) 65–71. [Zh.Eksp.Teor.Fiz. 30 (1956) 87-95]. URL:http://www.jetp.ac.ru/cgi-bin/dn/e_003_01_0065.pdf.

[56] F. James, "Monte Carlo theory and practice", *Rept. Prog. Phys.* **43** (1980) 1145, doi:10.1088/0034-4885/43/9/002.

[57] F. A. Berends, R. Kleiss, and S. Jadach, "Monte Carlo Simulation of radiative corrections to the processes $e^+e^- \to \mu^+\mu^-$ and $e^+e^- \to q\bar{q}$ in the Z0 region", *Comput. Phys. Commun.* **29** (1983) 185–200, doi:10.1016/0010-4655(83)90073-5.

[58] S. Höche, "Introduction to parton-shower event generators", in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pp. 235–295. 2015. arXiv:1411.4085. doi:10.1142/9789814678766_0005.

[59] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[60] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[61] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[62] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton fragmentation and string dynamics", *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.

[63] B. Webber, "A QCD model for jet fragmentation including soft gluon interference", *Nucl. Phys. B* **238** (1984) 492–528, doi:10.1016/0550-3213(84)90333-X.

[64] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[65] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *Journal of High Energy Physics* **2004** (nov, 2004) 040–040, doi:10.1088/1126-6708/2004/11/040.

[66] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *Journal of High Energy Physics* **2007** (nov, 2007) 070–070, doi:10.1088/1126-6708/2007/11/070.

[67] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing nlo calculations in shower Monte Carlo programs: the POWHEG BOX", *Journal of High Energy Physics* **2010** (Jun, 2010) 43, doi:10.1007/JHEP06(2010)043.

[68] F. Buccioni et al., "OpenLoops 2", *Eur. Phys. J. C* **79** (2019), no. 10, 866, doi:10.1140/epjc/s10052-019-7306-2, arXiv:1907.13071.

[69] F. Cascioli, P. Maierhofer, and S. Pozzorini, "Scattering Amplitudes with Open Loops", *Phys. Rev. Lett.* **108** (2012) 111601, doi:10.1103/PhysRevLett.108.111601, arXiv:1111.5206.

[70] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.

[71] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[72] J. Bellm et al., "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J. C* **76** (2016), no. 4, 196, doi:10.1140/epjc/s10052-016-4018-8, arXiv:1512.01178.

[73] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[74] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[75] G. Cowan, "Statistical data analysis". Oxford University Press, 1998. ISBN 978-0-19-850156-5.

[76] M. Dobbs et al., "Les Houches guidebook to Monte Carlo generators for hadron collider physics", in *3rd Les Houches Workshop on Physics at TeV Colliders*, pp. 411–459. 3, 2004. arXiv:hep-ph/0403045.

[77] ATLAS Collaboration, "Measurements of top-quark pair spin correlations in the $e\mu$ channel at $\sqrt{s} = 13$ TeV using pp collisions in the ATLAS detector", *Eur. Phys. J. C* **80** (2020), no. 8, 754, doi:10.1140/epjc/s10052-020-8181-6, arXiv:1903.07570.

[78] CMS Collaboration, "Measurement of the top quark polarization and $\mathrm{t\bar{t}}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **100** (2019), no. 7, 072002, doi:10.1103/PhysRevD.100.072002, arXiv:1907.03729.

[79] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders Through $O(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[80] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[81] M. Botje et al., "The PDF4LHC Working Group interim recommendations", (2011). arXiv:1101.0538.

[82] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections", *Eur. Phys. J.* **C64** (2009) 653–680, doi:10.1140/epjc/s10052-009-1164-2, arXiv:0905.3531.

[83] J. Gao et al., "CT10 next-to-next-to-leading order global analysis of QCD", *Phys. Rev.* **D89** (2014), no. 3, 033009, doi:10.1103/PhysRevD.89.033009, arXiv:1302.6246.

[84] R. D. Ball et al., "Parton distributions with LHC data", *Nucl. Phys. B* **867** (2013) 244–289, doi:10.1016/j.nuclphysb.2012.10.003, arXiv:1207.1303.

[85] The ATLAS Collaboration, The CMS Collaboration, "ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program". accessed on 14.02.2020. URL:https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO.

[86] CMS Collaboration, "Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons", *JHEP* **02** (2019) 149, doi:10.1007/JHEP02(2019)149, arXiv:1811.06625.

[87] CMS Collaboration, "Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **97** (2018), no. 11, 112003, doi:10.1103/PhysRevD.97.112003, arXiv:1803.08856.

[88] K. M. Flöh, "Investigation of the Top Quark Yukawa Coupling in Higgs boson Production in Association with Top Quarks at 13 TeV with the CMS experiment". Doctoral dissertation, Karlsruher Institut für Technologie (KIT), 2020. doi:10.5445/IR/1000105375.

[89] LHC Top physics Working Group, "LHCTopWG summary plots". accessed on 26.02.2020. URL:https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots.

[90] 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 $\sqrt{s} = 8$ TeV with the ATLAS detector", *Eur. Phys. J. C* **75** (2015), no. 7, 349, doi:10.1140/epjc/s10052-015-3543-1, arXiv:1503.05066.

[91] T. Ježo, J. M. Lindert, N. Moretti, and S. Pozzorini, "New NLOPS predictions for $t\bar{t} + b$ -jet production at the LHC", *Eur. Phys. J. C* **78** (2018), no. 6, 502, doi:10.1140/epjc/s10052-018-5956-0, arXiv:1802.00426.

[92] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector", *CERN* (2016) doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[93] F. Cascioli et al., "NLO matching for $tt\bar{b}\bar{b}$ production with massive b-quarks", *Phys. Lett. B* **734** (2014) 210–214, doi:10.1016/j.physletb.2014.05.040, arXiv:1309.5912.

[94] G. Bevilacqua, M. Garzelli, and A. Kardos, "$tt\bar{b}\bar{b}$ hadroproduction with massive bottom quarks with PowHel", arXiv:1709.06915.

[95] M. Garzelli, A. Kardos, and Z. Trócsányi, "Hadroproduction of $t\bar{t}b\bar{b}$ final states at LHC: predictions at NLO accuracy matched with Parton Shower", *JHEP* **03** (2015) 083, doi:10.1007/JHEP03(2015)083, arXiv:1408.0266.

[96] G. Bevilacqua et al., "Assault on the NLO Wishlist: $pp \to t\bar{t}b\bar{b}$", *JHEP* **09** (2009) 109, doi:10.1088/1126-6708/2009/09/109, arXiv:0907.4723.

[97] A. Bredenstein, A. Denner, S. Dittmaier, and S. Pozzorini, "NLO QCD corrections to $pp \to t\bar{t}b\bar{b} + X$ at the LHC", *Phys. Rev. Lett.* **103** (2009) 012002, doi:10.1103/PhysRevLett.103.012002, arXiv:0905.0110.

[98] F. Buccioni, S. Kallweit, S. Pozzorini, and M. F. Zoller, "NLO QCD predictions for $t\bar{t}b\bar{b}$ production in association with a light jet at the LHC", *JHEP* **12** (2019) 015, doi:10.1007/JHEP12(2019)015, arXiv:1907.13624.

[99] M. A. Harrendorf, "Measurement of the $t\bar{t}+ \geq 1$ jet cross section using novel multivariate analysis techniques at the CMS experiment". Doctoral dissertation, Karlsruher Institut für Technologie (KIT), 2018. doi:10.5445/IR/1000084837.

[100] A. Saibel, "Study of tt+bb modelling and uncertainties of Monte Carlo generators for the ttH(bb)-analysis at the CMS experiment". Master's thesis, Karlsruher Institut für Technologie (KIT), 2017. IEKP-KA/2017-11. URL:https://publish.etp.kit.edu/record/21290.

[101] M. M. Horzela, "Merging of tt and tt+bb simulations for an improved background modeling for ttH(bb) measurements". Master's thesis, Karlsruher Institut für Technologie (KIT), 2019. ETP-KA/2019-13. URL:https://publish.etp.kit.edu/record/21675.

[102] CMS Collaboration, "Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **776** (2018) 355–378, doi:10.1016/j.physletb.2017.11.043, arXiv:1705.10141.

[103] CMS Collaboration, "Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **803** (2020) 135285, doi:10.1016/j.physletb.2020.135285, arXiv:1909.05306.

[104] CMS Collaboration, "Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2020) 125, doi:10.1007/JHEP07(2020)125, arXiv:2003.06467.

[105] ATLAS Collaboration, "Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **04** (2019) 046, doi:10.1007/JHEP04(2019)046, arXiv:1811.12113.

[106] CMS Collaboration, CMS, "CMS Top Quark Physics Summary Figures". accessed on 26.02.2020. URL:https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryFigures.

[107] M. Aliev et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034–1046, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.

[108] P. Kant 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** (2015) 74–89, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

[109] The ATLAS Collaboration, The CMS Collaboration, "NLO single-top channel cross sections. ATLAS – CMS recommended predictions for single-top cross sections using the Hathor v2.1 program.". accessed on 13.02.2020. URL:https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec.

[110] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W^-orH$^-$", *Phys. Rev.* **D82** (2010) 054018, doi:10.1103/PhysRevD.82.054018, arXiv:1005.4451.

[111] N. Kidonakis, "Top quark production", in *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013*, pp. 139–168. 2014. arXiv:1311.0283. doi:10.3204/DESY-PROC-2013-03/Kidonakis.

[112] LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, OPAL, "Search for the standard model Higgs boson at LEP", *Phys. Lett. B* **565** (2003) 61–75, doi:10.1016/S0370-2693(03)00614-2, arXiv:hep-ex/0306033.

[113] CDF Collaboration, "Combination fo Searches for the Higgs boson Using the Full CDF Data Set", *Phys. Rev. D* **88** (2013), no. 5, 052013, doi:10.1103/PhysRevD.88.052013, arXiv:1301.6668.

[114] D0 Collaboration, "Combined search for the Higgs boson with the D0 experiment", *Phys. Rev. D* **88** (2013), no. 5, 052011, doi:10.1103/PhysRevD.88.052011, arXiv:1303.0823.

[115] S. Dimopoulos and H. Georgi, "Softly broken supersymmetry and SU(5)", *Nuclear Physics B* **193** (1981), no. 1, 150 – 162, doi:https://doi.org/10.1016/0550-3213(81)90522-8.

[116] E. Witten, "Dynamical breaking of supersymmetry", *Nuclear Physics B* **188** (1981), no. 3, 513 – 554, doi:https://doi.org/10.1016/0550-3213(81)90006-7.

[117] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[118] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, "The hierarchy problem and new dimensions at a millimeter", *Phys. Lett. B* **429** (1998) 263–272, doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.

[119] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, "Electroweak symmetry breaking from dimensional deconstruction", *Phys. Lett. B* **513** (2001) 232–240, doi:10.1016/S0370-2693(01)00741-9, arXiv:hep-ph/0105239.

[120] LHC Higgs Cross Section Working Group, "LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle", arXiv:1209.0040.

[121] N. Faltermann, "Single Top Quark Production at 13 TeV with the CMS experiment: from Rediscovery to Search for Rare Channels and Determination of Higgs boson Couplings". Doctoral dissertation, Karlsruher Institut für Technologie (KIT), 2018. doi:10.5445/IR/1000085709.

[122] CMS Collaboration, "Combined Higgs boson production and decay measurements with up to 137 fb-1 of proton-proton collision data at sqrts = 13 TeV", CMS Physics Analysis Summary CMS-PAS-HIG-19-005, 2020. URL:https://cds.cern.ch/record/2706103.

[123] V. Drollinger, "Reconstruction and analysis methods for searches of Higgs bosons in the decay mode $H^0 \to b\bar{b}$ at hadron colliders". Doctoral dissertation, Karlsruher Institut für Technologie (KIT), 2001. URL:https://publish.etp.kit.edu/record/21613.

[124] V. Drollinger, T. Müller, and D. Denegri, "Searching for Higgs bosons in association with top quark pairs in the $H^0 \to b\bar{b}$ decay mode", arXiv:hep-ph/0111312. IEKP-KA-2001-23, CMS-NOTE-2001-054.

[125] A. Schmidt, "Search for $H \to b\bar{b}$ in association with a $t\bar{t}$ pair in proton-proton collisions at $\sqrt{14} = 14\,\text{TeV}$". Doctoral dissertation, Karlsruher Institut für Technologie (KIT), 2006. URL:https://publish.etp.kit.edu/record/21643.

[126] J. Cammin, "Study of a light standard model Higgs boson in the $t\bar{t}H$ channel with ATLAS at LHC and decay-mode independent searches for neutral Higgs bosons with OPAL at LEP". Doctoral dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn, 2004. URL:http://hdl.handle.net/20.500.11811/2029.

[127] CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087, arXiv:1408.1682. [Erratum: JHEP 10, 106 (2014)].

[128] CMS Collaboration, "Search for the standard model Higgs boson produced in association with a top-quark pair in *pp* collisions at the LHC", *JHEP* **05** (2013) 145, doi:10.1007/JHEP05(2013)145, arXiv:1303.0763.

[129] CMS Collaboration, "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. C* **75** (2015), no. 6, 251, doi:10.1140/epjc/s10052-015-3454-1, arXiv:1502.02485.

[130] ATLAS Collaboration, "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* **05** (2016) 160, doi:10.1007/JHEP05(2016)160, arXiv:1604.03812.

[131] CMS Collaboration, "Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with $\sqrt{s} = 13$ TeV pp collisions at the CMS experiment", CMS Physics Analysis Summary CMS-PAS-HIG-16-004, 2016. URL:https://cds.cern.ch/record/2139578.

[132] S. D. Williamson, "Search for Higgs-boson production in association with a top-quark pair in the boosted regime with the CMS experiment". Doctoral dissertation, Karlsruher Institut für Technologie (KIT), 2016. doi:10.5445/IR/1000068213.

[133] T. Pfotzer, "Performance of jet-substructure techniques in Higgs-boson production in association with $t\bar{t}$ and a study of first LHC Run-II data at CMS". Master's thesis, Karlsruher Institut für Technologie (KIT), 2015. IEKP-KA/2015-18. URL:https://publish.etp.kit.edu/record/21236.

[134] K. El Morabit, "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's thesis, Karlsruher Institut für Technologie (KIT), 2015. URL:https://publish.etp.kit.edu/record/21342.

[135] H. Mildner, "Measurements of the top-Higgs coupling with the CMS experiment". Doctoral dissertation, Karlsruher Institut für Technologie (KIT), 2016. doi:10.5445/IR/1000065207.

[136] CMS Collaboration, "Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with 2016 pp collision data at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-HIG-16-038, 2016. URL:https://cds.cern.ch/record/2231510.

[137] M. Waßmer, "Spin-correlation studies in a search for Higgs-boson production in association with top quarks at CMS". Master's thesis, Karlsruher Institut für Technologie (KIT), 2017. IEKP-KA/2017-06. URL:https://publish.etp.kit.edu/record/21410.

[138] K. Schweiger, "Improvement of the b-jet energy measurement in a search for Higgs-boson production in association with top quarks at CMS using regression techniques". Master's thesis, Karlsruher Institut für Technologie (KIT), 2016. IEKP-KA/2016-20. URL:https://publish.etp.kit.edu/record/21326.

[139] F. Riese, "Boosted-jet reconstruction methods in a search for Higgs-boson production in association with a top-quark-antiquark pair at the CMS experiment". Master's thesis, Karlsruher Institut für Technologie (KIT), 2017. IEKP-KA/2017-08. URL:https://publish.etp.kit.edu/record/21292.

[140] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[141] CERN, "About CERN". accessed on 21.09.2020. URL:https://home.cern/about.

[142] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[143] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[144] ATLAS Collaboration, "The ATLAS experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. 437 p, doi:10.1088/1748-0221/3/08/S08003. Also published by CERN Geneva in 2010.

[145] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[146] B. Muratori, "Luminosity and luminous region calculations for the LHC", Technical Report LHC-PROJECT-NOTE-301, Sep, 2002. URL:https://cds.cern.ch/record/691967.

[147] D. Brandt, ed., "Intermediate accelerator physics. Proceedings, CERN Accelerator School, Zeuthen, Germany, September 15-26, 2003", CERN Yellow Reports: School Proceedings. (4, 2006). doi:10.5170/CERN-2006-002.

[148] CMS Collaboration, "CMS luminosity measurements for the 2016 data taking period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. URL:https://cds.cern.ch/record/2257069.

[149] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019. URL:https://cds.cern.ch/record/2676164.

[150] CMS Collaboration, "CMS luminosity – public results". accessed on 21.09.2020. URL:https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[151] CERN, "LHC long-term schedule". accessed on 21.09.2020. URL:https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm.

[152] L. Rossi and O. Brüning, "The High Luminosity Large Hadron Collider: the new machine for illuminating the mysteries of Universe". Advanced series on directions in high energy physics. World Scientific, Hackensack, NJ, 2015. doi:10.1142/9581.

[153] T. Sakuma and T. McCauley, "Detector and event visualization with sketchup at the CMS experiment", *J. Phys. Conf. Ser.* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032, arXiv:1311.4942.

[154] CMS Collaboration, "3D models of the CMS detector and events in SketchUp". accessed on 9.10.2020. URL:https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS.

[155] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade", Technical Design Report CMS CERN-LHCC-2012-016, CMS-TDR-011, 2012. doi:10.2172/1151650.

[156] CMS Collaboration, "The CMS tracker system project: Technical Design Report", Technical Design Report CMS CERN-LHCC-98-006, 1997. URL:https://cds.cern.ch/record/368412.

[157] CMS Collaboration, "The CMS tracker: addendum to the Technical Design Report", Technical Design Report CMS CERN-LHCC-2000-016, 2000. URL:https://cds.cern.ch/record/490194.

[158] CMS Collaboration, "CMS tracker detector performance results". accessed on 21.09.2020. URL:https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK.

[159] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", Technical Design Report CMS CERN-LHCC-97-033, 1997. URL:https://cds.cern.ch/record/349375.

[160] CMS Collaboration, "The CMS ECAL performance with examples", *JINST* **9** (2014) C02008, doi:10.1088/1748-0221/9/02/C02008.

[161] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". Number CERN-LHCC-97-031. CERN, Geneva, 1997.

[162] CMS Collaboration, "The CMS muon project: Technical Design Report", Technical Design Report CMS CERN-LHCC-97-032, 1997. URL:http://cds.cern.ch/record/343814.

[163] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC", *JINST* **8** (2013) P11002, doi:10.1088/1748-0221/8/11/P11002, arXiv:1306.6905.

[164] A. Sirunyan et al., "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of Instrumentation* **13** (jun, 2018) P06015–P06015, doi:10.1088/1748-0221/13/06/p06015.

[165] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020), no. 09, P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.

[166] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[167] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems", Technical Design Report CMS CERN-LHCC-2000-038, 2000. URL:http://cds.cern.ch/record/706847.

[168] CMS Collaboration, "CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project", Technical Design Report CMS CERN-LHCC-2002-026, 2002. URL:http://cds.cern.ch/record/578006.

[169] I. Bird et al., "Update of the computing models of the WLCG and the LHC experiments", Technical Report CERN-LHCC-2014-014. LCG-TDR-002, Apr, 2014. URL:https://cds.cern.ch/record/1695401.

[170] K. Bos et al., "LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)", Technical Design Report LCG CERN-LHCC-2005-024, 2005. URL:https://cds.cern.ch/record/840543.

[171] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software", Technical Design Report CMS CERN-LHCC-2006-001, 2006. There is an error on cover due to a technical problem for some items. URL:https://cds.cern.ch/record/922757.

[172] CMS Collaboration, "The CMS Offline SW Guide". accessed on 23.09.2020. URL:https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuide.

[173] I. Antcheva et al., "ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization", *Comput. Phys. Commun.* **180** (2009) 2499, doi:10.1016/j.cpc.2009.08.005, arXiv:1508.07749.

[174] F. Chollet et al., "Keras". https://keras.io, 2015.

[175] M. Abadi et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems", arXiv:1603.04467.

[176] L. Moneta et al., "The RooStats project", in *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. arXiv:1009.1003. PoS(ACAT2010)057.

[177] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", in *Proceedings of the 13th International Conference for Computing in High-Energy and Nuclear Physics (CHEP03)*. 2003. arXiv:physics/0306116. [eConf C0303241, MOLT007].

[178] G. Bohm and G. Zech, "Introduction to statistics and data analysis for physicists". DESY, Hamburg, 2014. doi:10.3204/DESY-BOOK/statistics, ISBN 978-3-935702-88-1.

[179] CMS Collaboration, "Combine Manual". accessed on 21.09.2020. URL:http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/.

[180] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, Aug, 2011. URL:https://cds.cern.ch/record/1379837.

[181] CMS Collaboration, "Combine Manual: Binned shape analysis". accessed on 21.09.2020. URL:http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part2/settinguptheanalysis/#binned-shape-analysis.

[182] J. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", in *PHYSTAT 2011*, pp. 115–120. 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.

[183] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **231** (1933), no. 694-706, 289–337, doi:10.1098/rsta.1933.0009.

[184] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[185] S. A. Murphy and A. W. V. D. Vaart, "On profile likelihood", *Journal of the American Statistical Association* **95** (2000), no. 450, 449–465, doi:10.1080/01621459.2000.10474219.

[186] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses", *Ann. Math. Statist.* **9** (03, 1938) 60–62, doi:10.1214/aoms/1177732360.

[187] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, "The elements of statistical learning : data mining, inference, and prediction". Springer series in statistics. Springer, New York, NY, 2. ed., corr. at 10. print. edition, 2013. ISBN 978-0-387-84857-0.

[188] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)", in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Y. Bengio and Y. LeCun, eds. 2016. arXiv:1511.07289.

[189] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016. ISBN 9780262035613. URL:http://www.deeplearningbook.org.

[190] A. L. Maas, "Rectifier nonlinearities improve neural network acoustic models", in *Proc. ICML*. 2013.

[191] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification", in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034. 2015. arXiv:1502.01852.

[192] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", in *Proceedings of Machine Learning Research*, Y. W. Teh and M. Titterington, eds., volume 9 of *Proceedings of Machine Learning Research*, pp. 249–256. JMLR Workshop and Conference Proceedings, Chia Laguna Resort, Sardinia, Italy, 13–15 May, 2010. URL:http://proceedings.mlr.press/v9/glorot10a.html.

[193] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance", in *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, p. 78. Association for Computing Machinery, New York, NY, USA, 2004. doi:10.1145/1015330.1015435.

[194] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors". MIT Press, Cambridge, MA, USA, 1988. ISBN 0262010976.

[195] N. Srivastava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research* **15** (2014), no. 56, 1929–1958. URL:http://jmlr.org/papers/v15/srivastava14a.html.

[196] A. Hocker et al., "TMVA - Toolkit for Multivariate Data Analysis with ROOT: users guide. TMVA - Toolkit for Multivariate Data Analysis", technical report, Mar, 2007. arXiv:physics/0703039, TMVA-v4 Users Guide: 135 pages, 19 figures, numerous code examples and references.

[197] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt, "TMVA, the toolkit for multivariate data analysis with ROOT", in *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*, p. 40. 2007. [PoS(ACAT)040]. doi:10.22323/1.050.0040.

[198] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (Jun, 2017) P10003. 82 p, doi:10.1088/1748-0221/12/10/P10003.

[199] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[200] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track reconstruction in the CMS tracker", Technical Report CMS-NOTE-2006-041, Dec, 2006. URL:https://cds.cern.ch/record/934067.

[201] P. Billoir, "Progressive track recognition with a kalman-like fitting procedure", *Computer Physics Communications* **57** (1989), no. 1, 390 – 394, doi:https://doi.org/10.1016/0010-4655(89)90249-X.

[202] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the kalman filtering method", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **294** (1990), no. 1, 219 – 228, doi:https://doi.org/10.1016/0168-9002(90)91835-Y.

[203] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **395** (1997), no. 2, 169 – 184, doi:https://doi.org/10.1016/S0168-9002(97)00705-5.

[204] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (jul, 2005) N9–N20, doi:10.1088/0954-3899/31/9/n01.

[205] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *Proceedings of the IEEE* **86** (1998), no. 11, 2210–2239, doi:10.1109/5.726788.

[206] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", Technical Report CMS-NOTE-2007-008, Mar, 2007. URL:https://cds.cern.ch/record/1027031.

[207] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) 9009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016.

[208] CMS Collaboration, "Performance of electron reconstruction and aelection with the CMS Detector in proton-proton collisions at $\sqrt{\text{s}} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[209] CMS Collaboration, "Electron and photon performance in CMS with the full 2016 data sample.", CMS Detector performance note, Mar, 2017. URL:https://cds.cern.ch/record/2255497.

[210] CMS Collaboration, "Electron and photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 conference", CMS Detector performance note, May, 2018. URL:https://cds.cern.ch/record/2320638.

[211] A. Bodek et al., "Extracting muon momentum scale corrections for hadron collider experiments", *Eur. Phys. J. C* **72** (2012) 2194, doi:10.1140/epjc/s10052-012-2194-8, arXiv:1208.3710.

[212] CMS Collaboration, "Muon identification and isolation efficiency on full 2016 dataset", CMS Detector performance note, Mar, 2017. URL:https://cds.cern.ch/record/2257968.

[213] CMS Collaboration, "Muon identification and isolation efficiencies with 2017 and 2018 data", CMS Detector performance note, Jul, 2018. URL:https://cds.cern.ch/record/2629364.

[214] G. P. Salam, "Towards jetography", *Eur. Phys. J. C* **67** (2010) 637–686, doi:10.1140/epjc/s10052-010-1314-6, arXiv:0906.1833.

[215] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-ktjet clustering algorithm", *Journal of High Energy Physics* **2008** (apr, 2008) 063–063, doi:10.1088/1126-6708/2008/04/063.

[216] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[217] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[218] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[219] CMS Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018", CMS Detector performance note, Apr, 2020. URL:http://cds.cern.ch/record/2715872.

[220] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[221] CMS Collaboration, "Performance of the pile up jet identification in CMS for Run 2", CMS Detector performance note, Apr, 2020. URL:https://cds.cern.ch/record/2715906.

[222] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[223] CMS Collaboration, "Event displays with heavy flavour jets from 2016 CMS data", CMS Detector performance note, Aug, 2017. URL:http://cds.cern.ch/record/2281816.

[224] CMS Collaboration, "Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector", CMS Detector performance note, Jun, 2018. URL:https://cds.cern.ch/record/2627468.

[225] CMS Collaboration, "Muon HLT performance on 2016 data", CMS Detector performance note, Dec, 2017. URL:https://cds.cern.ch/record/2297529.

[226] CMS Collaboration, "Single and double electron trigger efficiencies using the full Run 2 dataset", CMS Detector performance note, Mar, 2020. accessed on 26.06.2020. URL:https://twiki.cern.ch/twiki/bin/view/CMSPublic/EgammaFullRun2Data.

[227] CMS Collaboration, "Performance of top triggers in Run 2", CMS Detector performance note, Aug, 2019. URL:https://cds.cern.ch/record/2687014.

[228] CMS Collaboration, "Jet energy resolution recommendations". accessed on 18.06.2020, only accessible to CMS members. URL:https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution.

[229] CMS Collaboration, "Search for $\mathrm{t\bar{t}H}$ production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **06** (2018) 101, doi:10.1007/JHEP06(2018)101, arXiv:1803.06986.

[230] TOTEM Collaboration, "First measurement of elastic, inelastic and total cross-section at $\sqrt{s} = 13$ TeV by TOTEM and overview of cross-section data at LHC energies", *Eur. Phys. J.* **C79** (2019), no. 2, 103, doi:10.1140/epjc/s10052-019-6567-0, arXiv:1712.06153.

[231] ATLAS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector at the LHC", *Phys. Rev. Lett.* **117** (2016), no. 18, 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.

[232] CMS Collaboration, "Investigations of the impact of the parton shower tuning in PYTHIA 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016. URL:https://cds.cern.ch/record/2235192.

[233] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[234] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020), no. 1, 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[235] H. B. Hartanto, B. Jäger, L. Reina, and D. Wackeroth, "Higgs boson production in association with top quarks in the POWHEG BOX", *Phys. Rev. D* **91** (May, 2015) 094003, doi:10.1103/PhysRevD.91.094003.

[236] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, doi:10.1007/JHEP02(2010)011, 10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: JHEP02,011(2010)].

[237] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J.* **C71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[238] F. Maltoni, D. Pagani, and I. Tsinikos, "Associated production of a top-quark pair with vector bosons at NLO in QCD: impact on $t\bar{t}H$ searches at the LHC", *JHEP* **02** (2016) 113, doi:10.1007/JHEP02(2016)113, arXiv:1507.05640.

[239] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.

[240] C. Collaboration, "Standard model cross sections for CMS at 13 TeV". accessed on 10.07.2020, only accessible to CMS members. URL:https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV.

[241] T. Gehrmann et al., "W^+W^- production at hadron colliders in next-to-next-to-leading order QCD", *Phys. Rev. Lett.* **113** (2014), no. 21, 212001, doi:10.1103/PhysRevLett.113.212001, arXiv:1408.5243.

[242] J. M. Campbell and R. Ellis, "An Update on vector boson pair production at hadron colliders", *Phys. Rev. D* **60** (1999) 113006, doi:10.1103/PhysRevD.60.113006, arXiv:hep-ph/9905386.

[243] J. M. Campbell, R. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[244] J. M. Campbell, R. K. Ellis, and W. T. Giele, "A Multi-Threaded Version of MCFM", *Eur. Phys. J. C* **75** (2015), no. 6, 246, doi:10.1140/epjc/s10052-015-3461-2, arXiv:1503.06182.

[245] M. Welsch, "Trennung von Signal und Untergrund mit TensorFlow in einer Suche nach ttH-Produktion am CMS-Experiment". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2017. IEKP-BACHELOR-KA/2017-03. URL:https://publish.etp.kit.edu/record/21284.

[246] J. van der Linden, "MVA-Studien mit binären neuronalen Netzen in der Suche nach ttH-Produktion am CMS experiment". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2017. IEKP-BACHELOR-KA/2017-14. URL:https://publish.etp.kit.edu/record/21428.

[247] L. Hilser, "Multivariate Klassifizierung mit künstlichen neuronalen Netzen zur Trennung von Unterprozessen der Top-Quark-Antiquark-Paar-Produktion". Master's thesis, Karlsruher Institut für Technologie (KIT), 2018. ETP-KA/2018-06. URL:https://publish.etp.kit.edu/record/21386.

[248] M. Rieger, "Search for Higgs boson production in association with top quarks and decaying into bottom quarks using deep learning techniques with the CMS experiment". Doctoral dissertation, RWTH Aachen University, Aachen, 2019. doi:10.18154/RWTH-2019-06415.

[249] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, eds. 2015. arXiv:1412.6980.

[250] J. D. Bjorken and S. J. Brodsky, "Statistical model for electron-positron annihilation into hadrons", *Phys. Rev. D* **1** (Mar, 1970) 1416–1420, doi:10.1103/PhysRevD.1.1416.

[251] ALEPH Collaboration, "Studies of quantum chromodynamics with the ALEPH detector", *Phys. Rept.* **294** (1998) 1–165, doi:10.1016/S0370-1573(97)00045-8.

[252] ALICE Collaboration, "Transverse sphericity of primary charged particles in minimum bias proton-proton collisions at $\sqrt{s} = 0.9$, 2.76 and 7 TeV", *Eur. Phys. J. C* **72** (2012) 2124, doi:10.1140/epjc/s10052-012-2124-9, arXiv:1205.3963.

[253] G. C. Fox and S. Wolfram, "Event shapes in e^+e^- annihilation", *Nuclear Physics B* **149** (1979), no. 3, 413 – 496, doi:https://doi.org/10.1016/0550-3213(79)90003-8.

[254] G. C. Fox and S. Wolfram, "Observables for the analysis of avent shapes in e^+e^- annihilation and other processes", *Phys. Rev. Lett.* **41** (Dec, 1978) 1581–1585, doi:10.1103/PhysRevLett.41.1581.

[255] R. Cousins, "Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms". accessed on 02.07.2021. URL:http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[256] CMS Collaboration, "Combine Manual: Goodness of fit tests". accessed on 02.07.2021. URL:http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part3/commonstatsmethods.html#goodness-of-fit-tests.

[257] CMS Collaboration, "Performance of quark/gluon discrimination in 8 TeV pp data", CMS Physics Analysis Summary CMS-PAS-JME-13-002, 2013. URL:https://cds.cern.ch/record/1599732.

[258] CMS Collaboration, "Performance of quark/gluon discrimination in 13 TeV data", CMS Detector performance note, Nov, 2016. URL:https://cds.cern.ch/record/2234117.

[259] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018. URL:http://cds.cern.ch/record/2621960.

[260] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219–228, doi:10.1016/0010-4655(93)90005-W.

[261] CMS Collaboration, "Combine Manual: Automatic MC statistical uncertainties". accessed on 22.07.2020. URL:http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part2/bin-wise-stats.

[262] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C* **74** (2014), no. 8, 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[263] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J. C* **75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.

[264] C. Papadopoulos and M. Worek, "HELAC - A Monte Carlo generator for multi-jet processes", in *14th International Workshop on Deep Inelastic Scattering*, pp. 507–510. 6, 2006. arXiv:hep-ph/0606320. doi:10.1142/9789812706706_0116.

[265] P. Keicher, "Model stability studies in a search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ channel at CMS". Master's thesis, Karlsruher Institut für Technologie (KIT), 2018. ETP-KA/2018-08. URL:https://publish.etp.kit.edu/record/21907.

[266] ATLAS Collaboration, "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D* **97** (2018), no. 7, 072016, doi:10.1103/PhysRevD.97.072016, arXiv:1712.08895.

[267] CMS Collaboration, "Observation of $t\bar{t}H$ production", *Phys. Rev. Lett.* **120** (2018), no. 23, 231801, doi:10.1103/PhysRevLett.120.231801, arXiv:1804.02610.

[268] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", doi:10.5170/CERN-2013-004, arXiv:1307.1347.

[269] CMS Collaboration, "Observation of Higgs boson decay to bottom quarks", *Phys. Rev. Lett.* **121** (2018), no. 12, 121801, doi:10.1103/PhysRevLett.121.121801, arXiv:1808.08242.

[270] ATLAS Collaboration, "Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment", *Phys. Rev. D* **101** (2020), no. 1, 012002, doi:10.1103/PhysRevD.101.012002, arXiv:1909.02845.

[271] ATLAS Collaboration, "A combination of measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment", Technical Report ATLAS-CONF-2020-027, Aug, 2020. URL:http://cds.cern.ch/record/2725733.

[272] CMS Collaboration, "Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **99** (2019), no. 9, 092005, doi:10.1103/PhysRevD.99.092005, arXiv:1811.09696.

[273] J. Kühn, A. Scharf, and P. Uwer, "Weak interactions in top-quark pair production at hadron colliders: An update", *Phys. Rev. D* **91** (2015), no. 1, 014020, doi:10.1103/PhysRevD.91.014020, arXiv:1305.5773.

[274] M. J. Strassler and M. E. Peskin, "Threshold production of heavy top quarks: Qcd and the higgs boson", *Phys. Rev. D* **43** (Mar, 1991) 1500–1514, doi:10.1103/PhysRevD.43.1500.

[275] M. Beneke, A. Maier, J. Piclum, and T. Rauh, "Higgs effects in top anti-top production near threshold in e^+e^- annihilation", *Nucl. Phys. B* **899** (2015) 180–193, doi:10.1016/j.nuclphysb.2015.07.034, arXiv:1506.06865.

[276] CMS Collaboration, "Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **80** (2020), no. 2, 75, doi:10.1140/epjc/s10052-019-7593-7, arXiv:1908.06463.

[277] CMS Collaboration, "Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the lepton+jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **100** (2019), no. 7, 072007, doi:10.1103/PhysRevD.100.072007, arXiv:1907.01590.

[278] CMS Collaboration, "Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the dilepton final state in proton-proton collisions at $\sqrt{s} =$ 13 TeV", arXiv:2009.07123. Submitted to Phys. Rev. D.

[279] G. Apollinari et al., "High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report". CERN Yellow Reports: Monographs. CERN, Geneva, 2015. doi:10.5170/CERN-2015-005.

[280] CMS Collaboration, "Technical Proposal for the Phase-II Upgrade of the CMS Detector", Technical Report CMS-TDR-15-02, Jun, 2015. URL:https://cds.cern.ch/record/2020886.

[281] CMS Collaboration, "The phase-2 upgrade of the CMS tracker", Technical Report CMS-TDR-014, 6, 2017. URL:https://cds.cern.ch/record/2272264.

[282] CMS Collaboration, "The phase-2 upgrade of the CMS barrel calorimeters", Technical Report CMS-TDR-015, Sep, 2017. URL:https://cds.cern.ch/record/2283187.

[283] CMS Collaboration, "The Phase-2 upgrade of the CMS endcap calorimeter", Technical Report CMS-TDR-019, Nov, 2017. URL:https://cds.cern.ch/record/2293646.

[284] CMS Collaboration, "The Phase-2 upgrade of the CMS muon detectors", Technical Report CMS-TDR-016, Sep, 2017. URL:https://cds.cern.ch/record/2283189.

[285] CMS Collaboration, "Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC", Technical Report CMS-NOTE-2018-006, Dec, 2018. URL:https://cds.cern.ch/record/2650976.

[286] M. Cepeda et al., "Report from Working Group 2: Higgs physics at the HL-LHC and HE-LHC", *CERN Yellow Rep. Monogr.* **7** (Dec, 2018) 221–584. 364 p, doi:10.23731/CYRM-2019-007.221.

[287] R. Abdul Khalek et al., "Towards ultimate parton distributions at the High-Luminosity LHC", *Eur. Phys. J. C* **78** (2018), no. 11, 962, doi:10.1140/epjc/s10052-018-6448-y, arXiv:1810.03639.

[288] CMS Collaboration, "Sensitivity projections for Higgs boson properties measurements at the HL-LHC", CMS Physics Analysis Summary CMS-PAS-FTR-18-011, 11, 2018. URL:https://cds.cern.ch/record/2647699.

[289] CMS Collaboration, "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **79** (2019), no. 5, 421, doi:10.1140/epjc/s10052-019-6909-y, arXiv:1809.10733.

[290] A. Valassi, "Combining correlated measurements of several different physical quantities", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **500** (2003), no. 1, 391 – 405, doi:https://doi.org/10.1016/S0168-9002(03)00329-2. NIMA Vol 500.

[291] J. Schindler, "Reducing background modeling uncertainties in the CMS analysis of the associated production of a Higgs boson and a top quark-antiquark pair with adversarial neural networks". Master's thesis, Karlsruher Institut für Technologie (KIT), 2019. ETP-KA/2019-12. URL:https://publish.etp.kit.edu/record/21672.

[292] S. Ehnle, "Uncertainty studies in a measurement of $\text{t}\bar{\text{t}}\text{H}$ production in the $\text{H} \to \text{b}\bar{\text{b}}$ Channel at CMS". Master's thesis, Karlsruher Institut für Technologie (KIT), 2020. ETP-KA/2020-16. URL:https://publish.etp.kit.edu/record/21972.

[293] N. Shadskiy, "Treating uncertainties with Bayesian Neural Networks in the measurement of ttH(bb) production". Master's thesis, Karlsruher Institut für Technologie (KIT), 2020. ETP-KA/2020-18. URL:https://publish.etp.kit.edu/record/21982.

[294] Y.-Y. C. Cung, "Untersuchung von Optimierungsstrategien für Bayesian Neural Networks im Rahmen der ttH(bb)-Analyse am CMS experiment am CERN". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2020.

ETP-BACHELOR-KA/2020-10.
URL:https://publish.etp.kit.edu/record/21999.

[295] J. von den Driesch, "Rekonstruktion des Top-Quark-Antiquark-Systems mit neuronalen Netzen am CMS experiment". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2019. ETP-BACHELOR-KA/2019-14.
URL:https://publish.etp.kit.edu/record/21915.

[296] M. Heim, "Rekonstruktion des Higgs-bosons in assoziierter Produktion mit einem Top-Quark-Antiquark-Paar mithilfe Neuronaler Netze bei CMS". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2020. ETP-BACHELOR-KA/2020-03.
URL:https://publish.etp.kit.edu/record/21946.

[297] Y. Seyffert, "Untersuchung von Convolutional Neural Networks im Rahmen einer ttH-Analyse am CMS experiment". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2019. ETP-BACHELOR-KA/2019-16.
URL:https://publish.etp.kit.edu/record/21747.

[298] J. C. Vautz, "Weiterführende Studien zu Convolutional Neural Networks im Rahmen einer ttH-Analyse am CMS experiment". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2020. ETP-BACHELOR-KA/2020-09.
URL:https://publish.etp.kit.edu/record/21997.

[299] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, "Deep-learning top taggers or the end of QCD?", *JHEP* **05** (2017) 006, doi:10.1007/JHEP05(2017)006, arXiv:1701.08784.

[300] L. Weidemann, "Rekonstruktion des Higgs-boson-Zerfalls in Assoziation mit einem Top-Quark-Antiquark-Paar im Boosted Regime am CMS experiment". Bachelor's thesis, Karlsruher Institut für Technologie (KIT), 2018.
ETP-BACHELOR-KA/2018-12.
URL:https://publish.etp.kit.edu/record/21538.

[301] P. Krämer, "Measurement of $t\bar{t}H$ Simplified Template Cross Sections in the $H \to b\bar{b}$ channel at the CMS experiment". Masters's thesis, Karlsruher Institut für Technologie (KIT), 2020. ETP-KA/2020-04.
URL:https://publish.etp.kit.edu/record/21939.

[302] LHC Higgs Cross Section Working Group, "LHC Higgs Cross Section Working Group". accessed on 9.10.2020.
URL:https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

Appendix

A Basic observables

In this appendix, distributions of basic observables are shown for events with one lepton, (≥ 4 jets and ≥ 2 b-tags) in data recorded in the year 2016 in Sec. A.1 and for data recorded in the year 2017 in Sec. A.2.
In all figures, the distributions observed in recorded data are shown as black markers. The expected background contributions before the fit to data are shown as stacks of histograms and the $t\bar{t}H$ process as blue lines which are additionally scaled to the integral of the background for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding those affecting only the rate. The bottom panels show the ratio of data to the background prediction.

A.1 Basic observables in data recorded in the year 2016

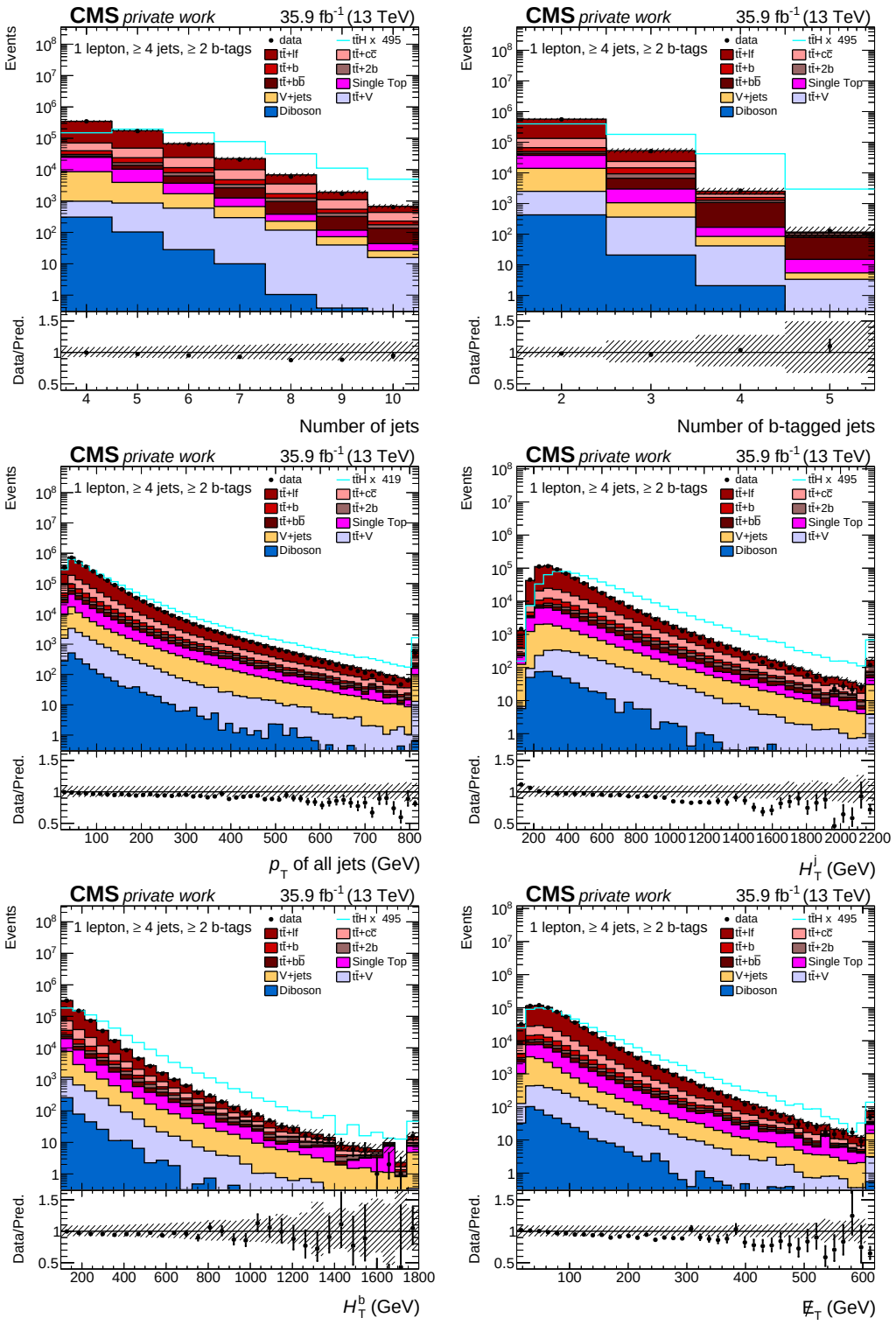


Figure A.1: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2016. A description of the figures is given at the beginning of Sec. A.

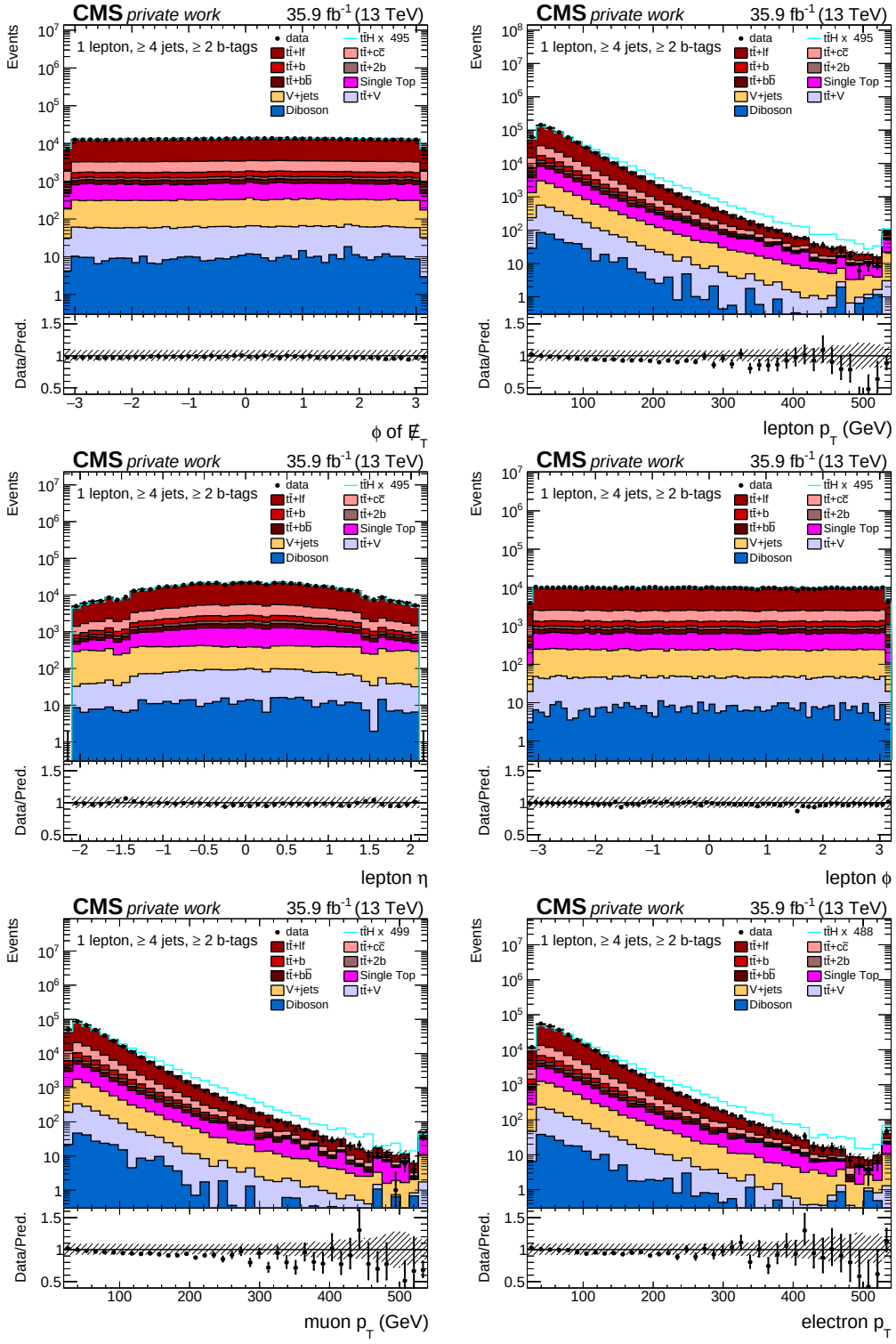


Figure A.2: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2016. A description of the figures is given at the beginning of Sec. A.

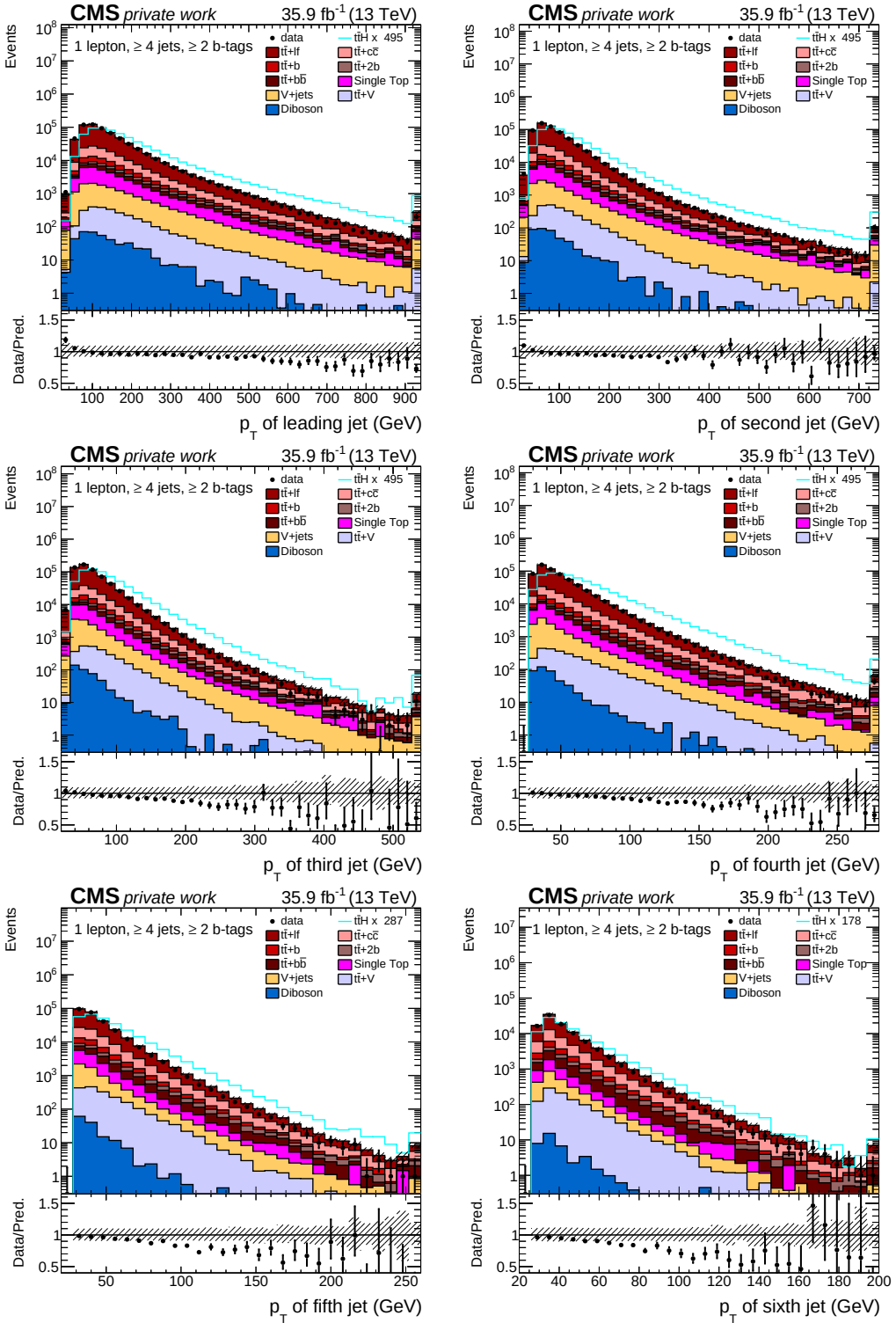


Figure A.3: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2016. A description of the figures is given at the beginning of Sec. A.

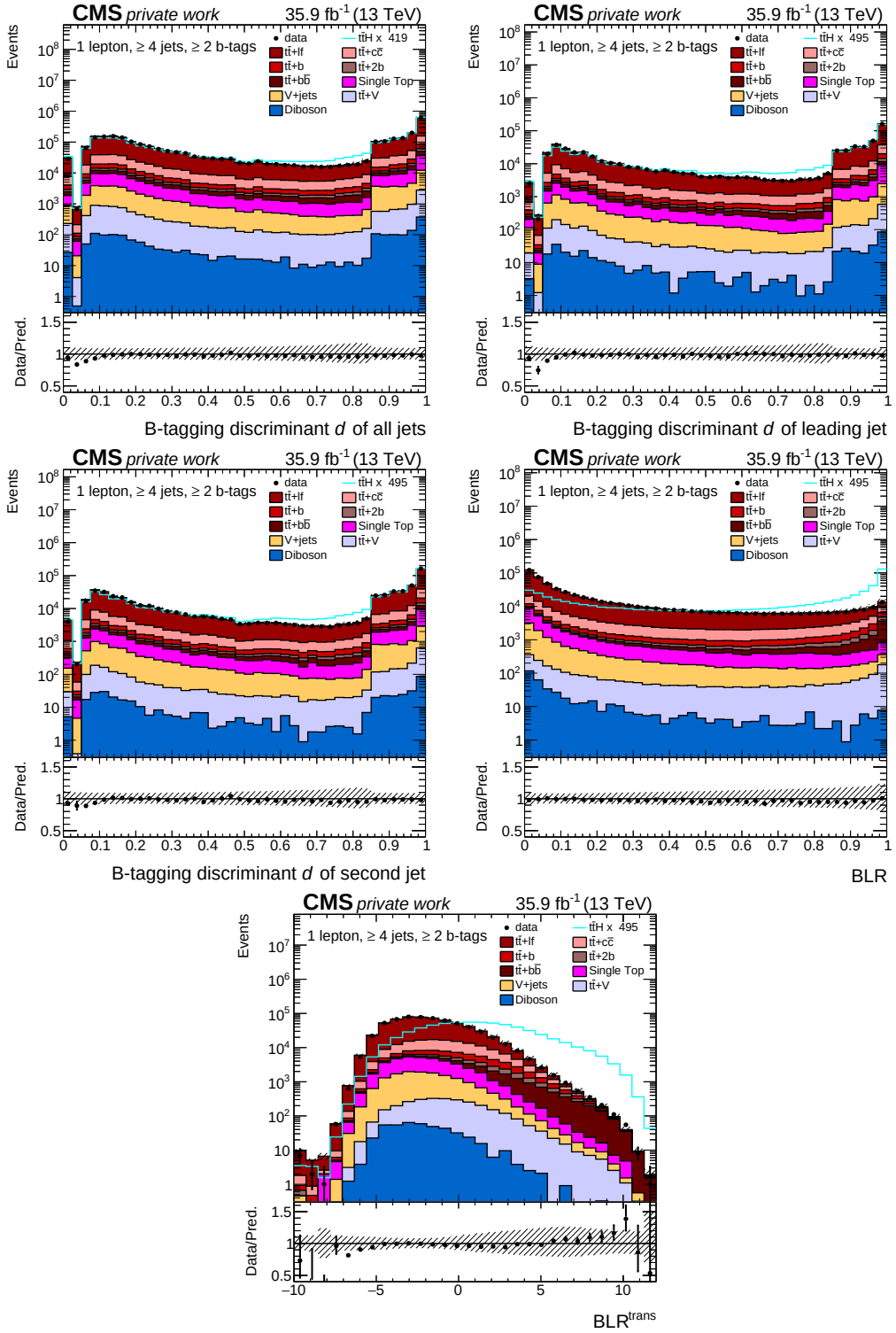


Figure A.4: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2016. A description of the figures is given at the beginning of Sec. A.

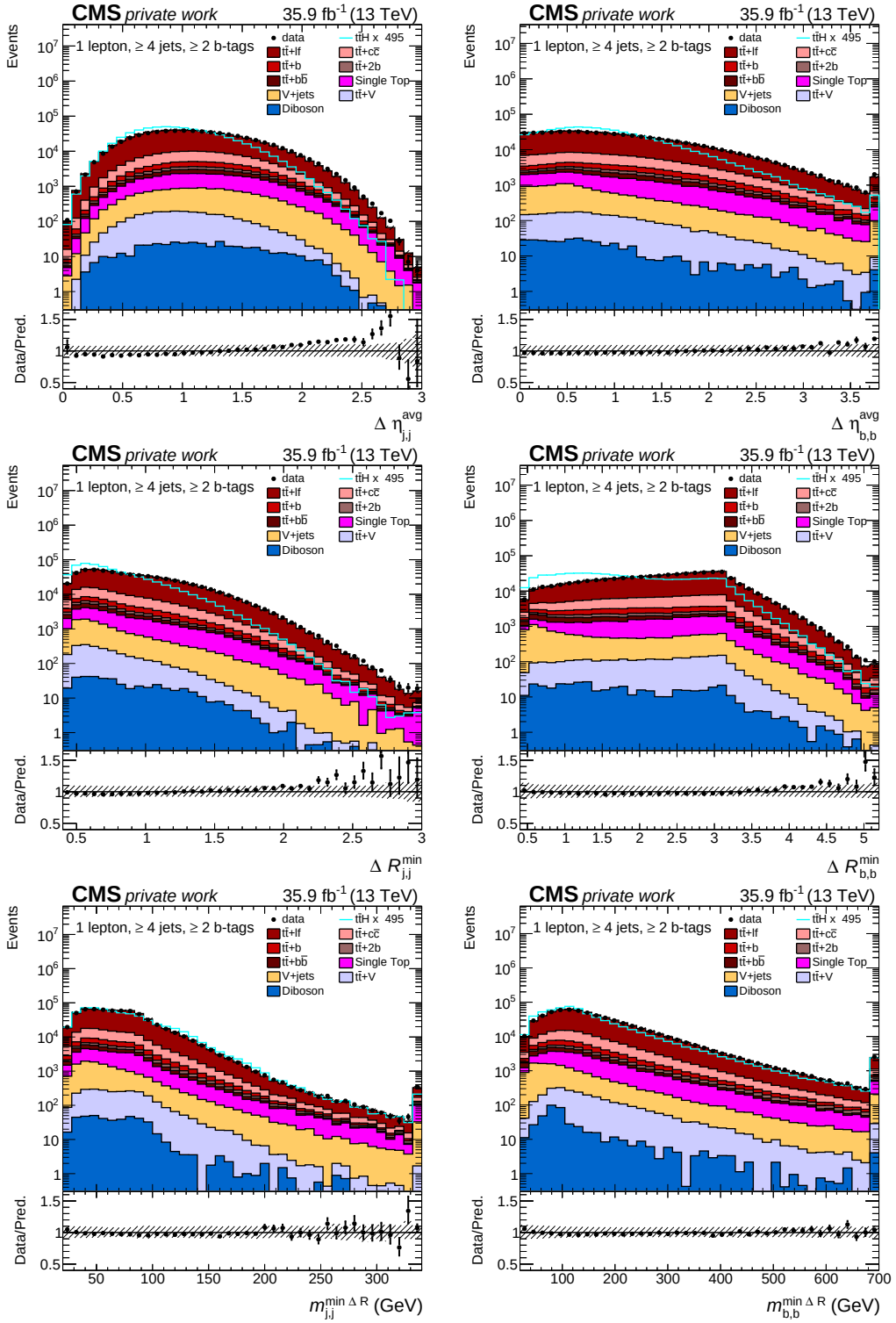


Figure A.5: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2016. A description of the figures is given at the beginning of Sec. A.

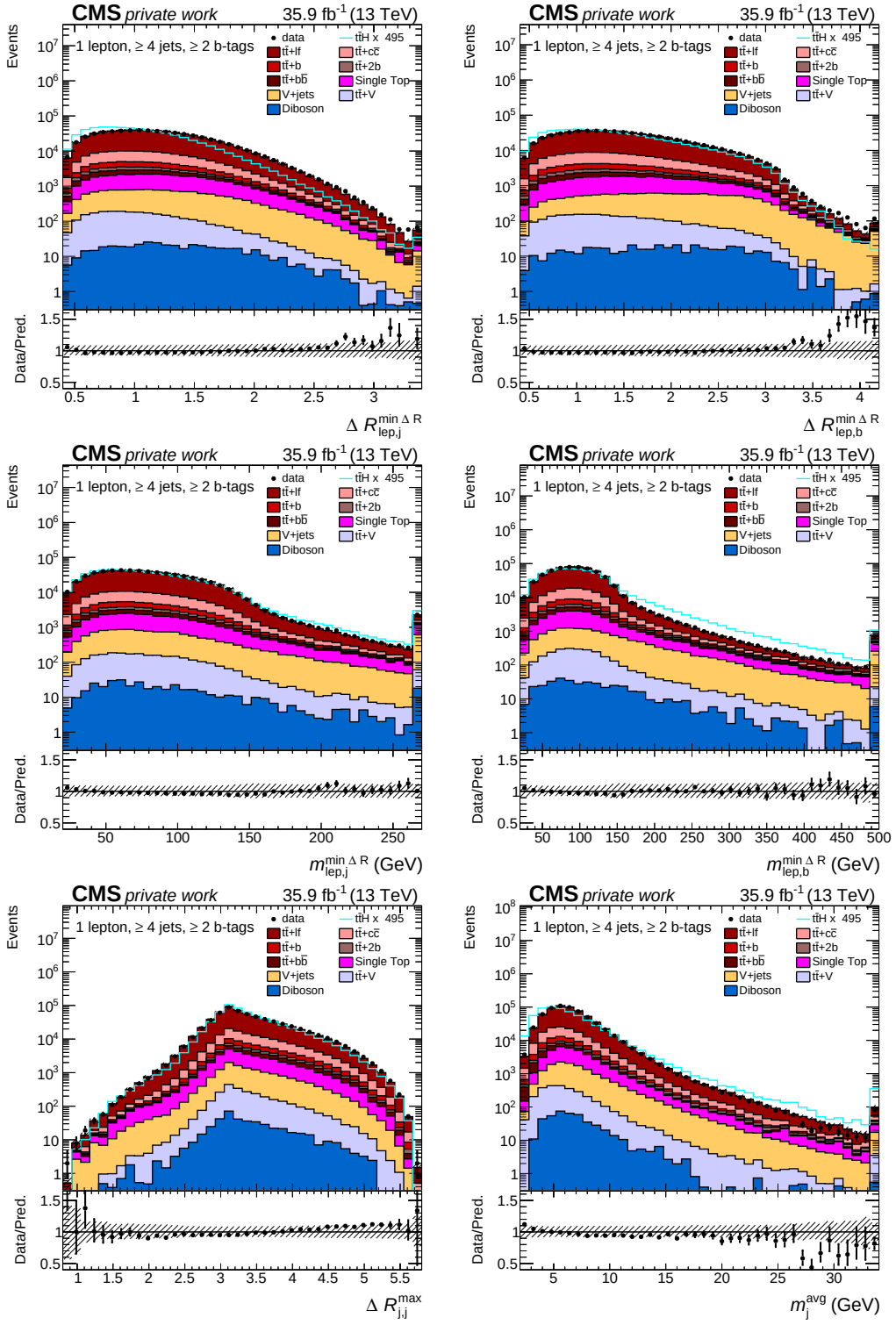


Figure A.6: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2016. A description of the figures is given at the beginning of Sec. A.

A.2 Basic observables in data recorded in the year 2017

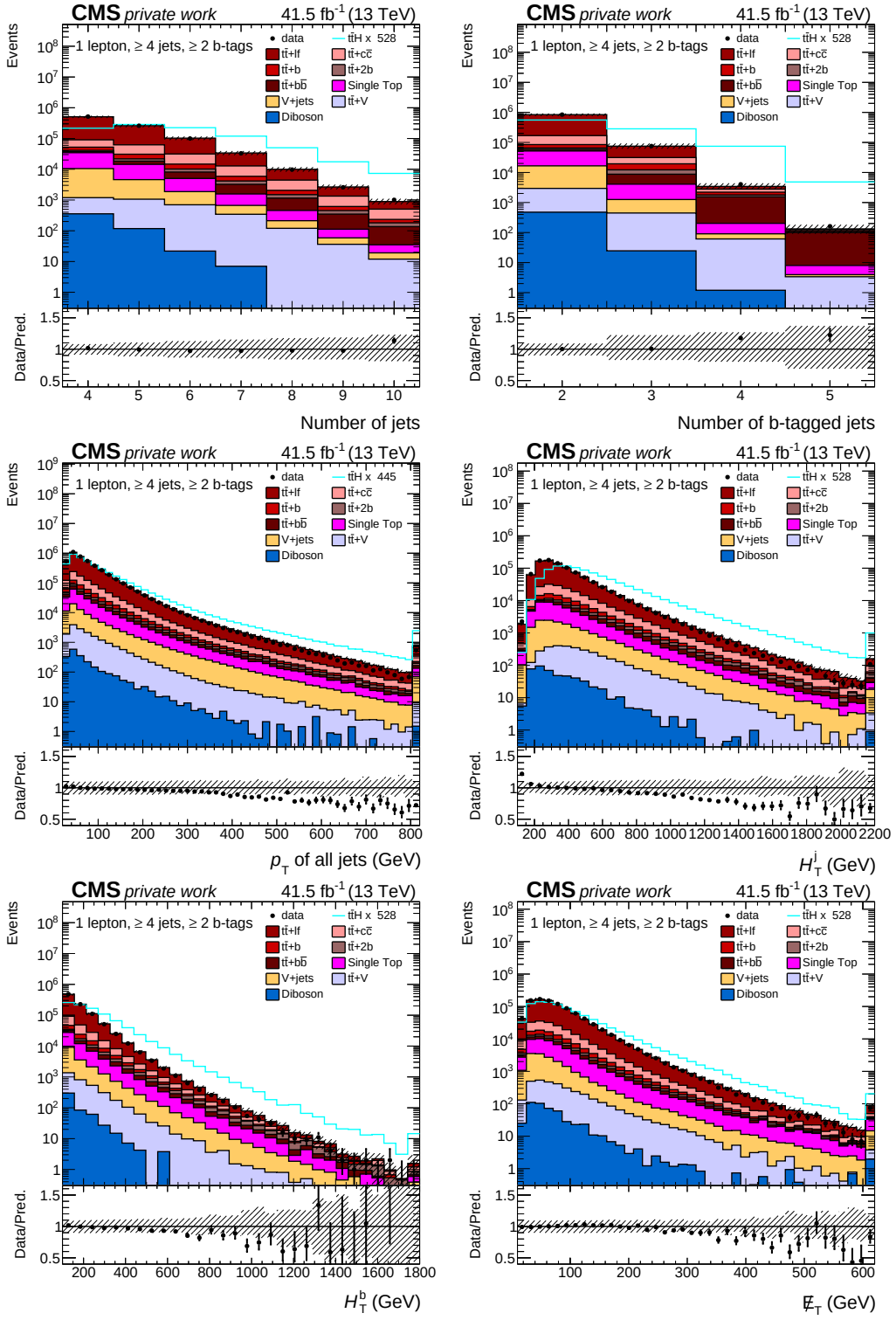


Figure A.7: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2017. A description of the figures is given at the beginning of Sec. A.

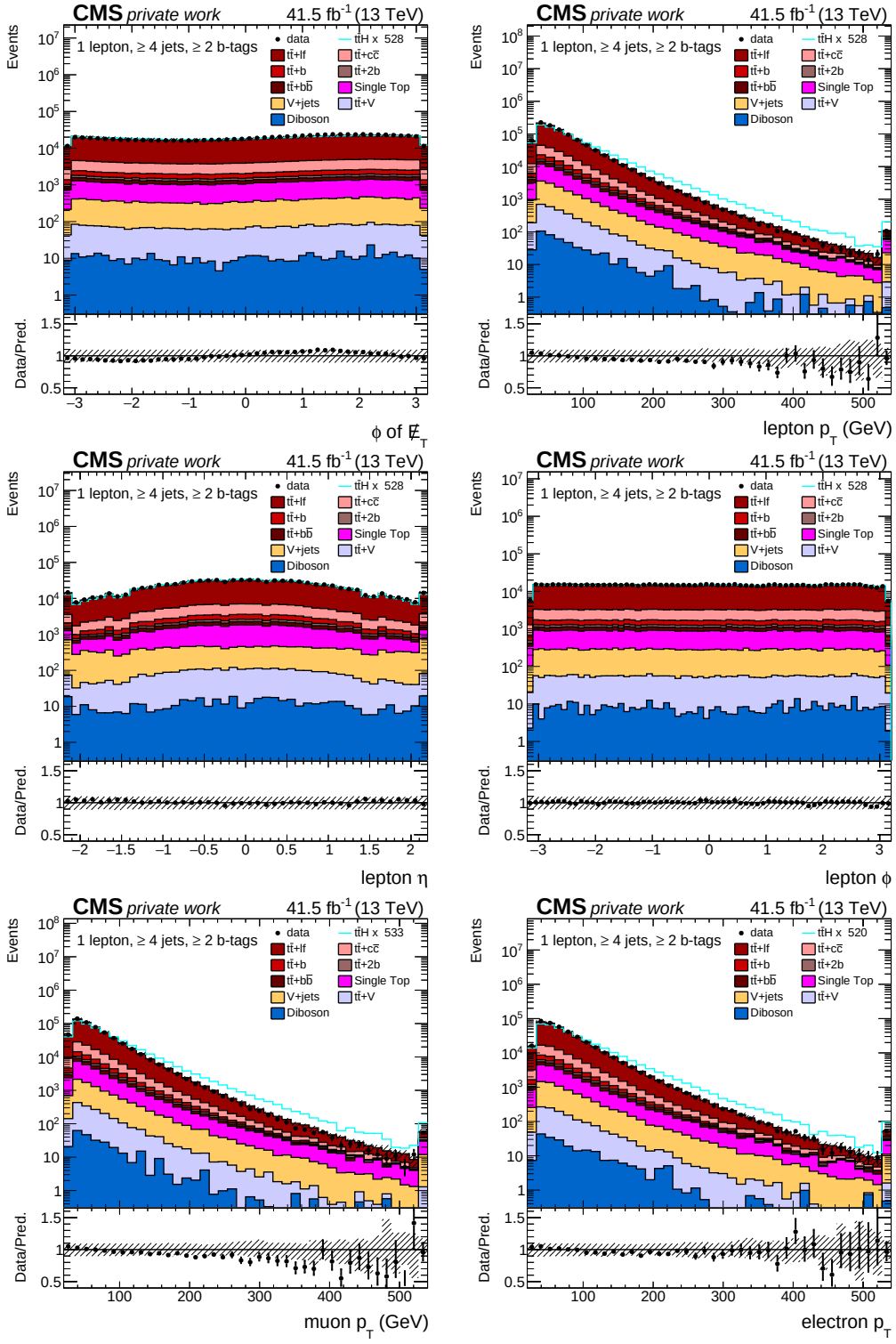


Figure A.8: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2017. A description of the figures is given at the beginning of Sec. A.

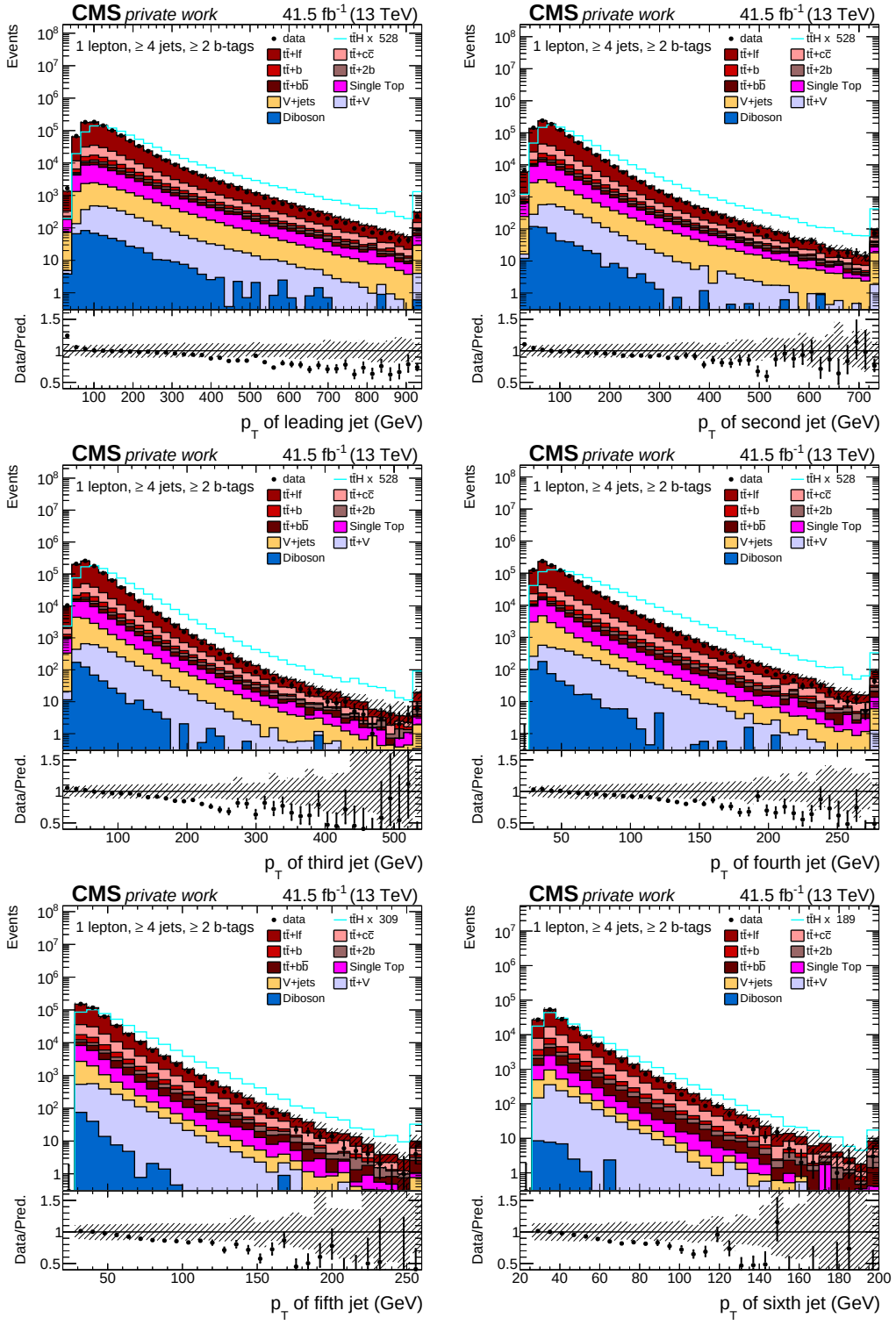


Figure A.9: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2017. A description of the figures is given at the beginning of Sec. A.

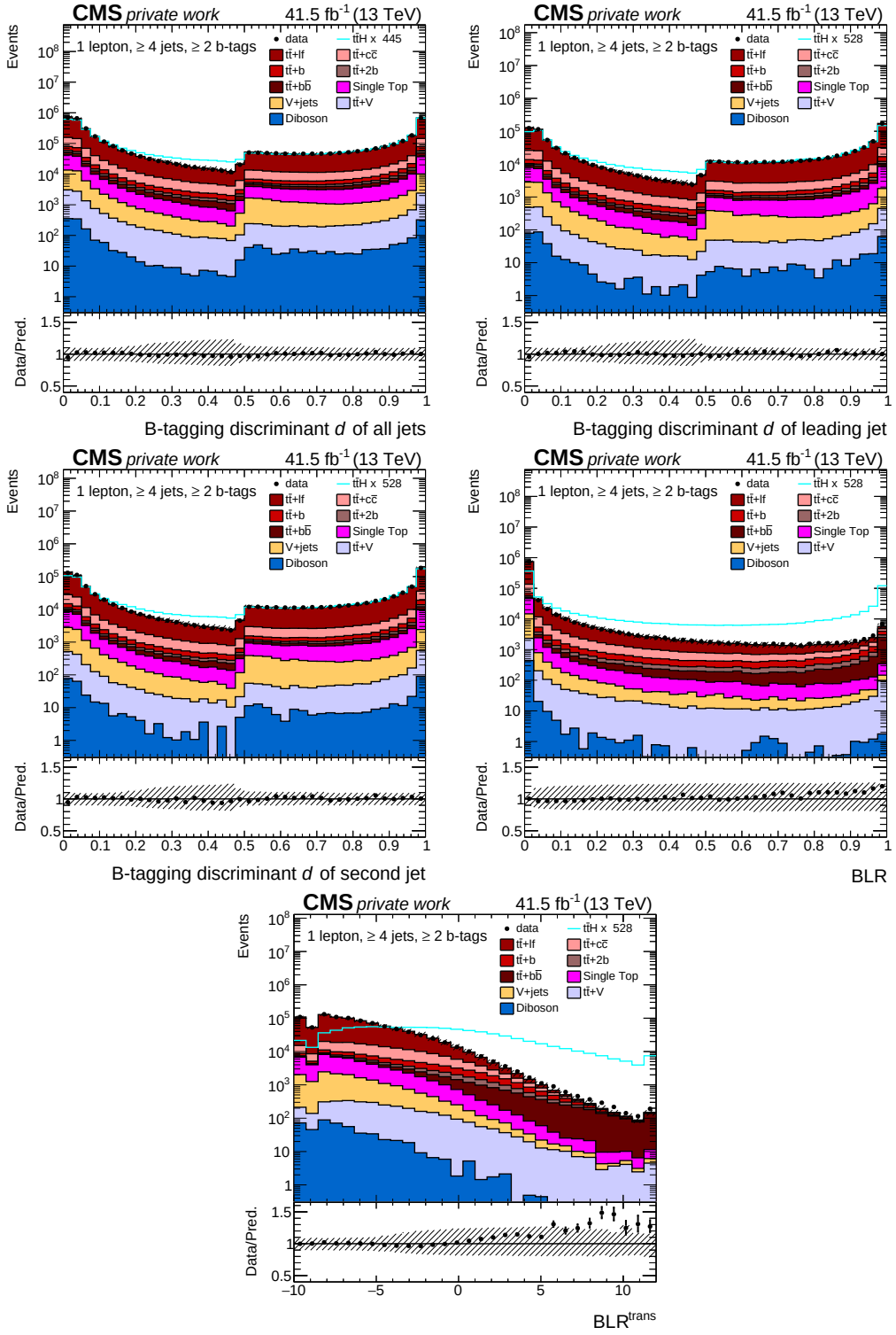


Figure A.10: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2017. A description of the figures is given at the beginning of Sec. A.

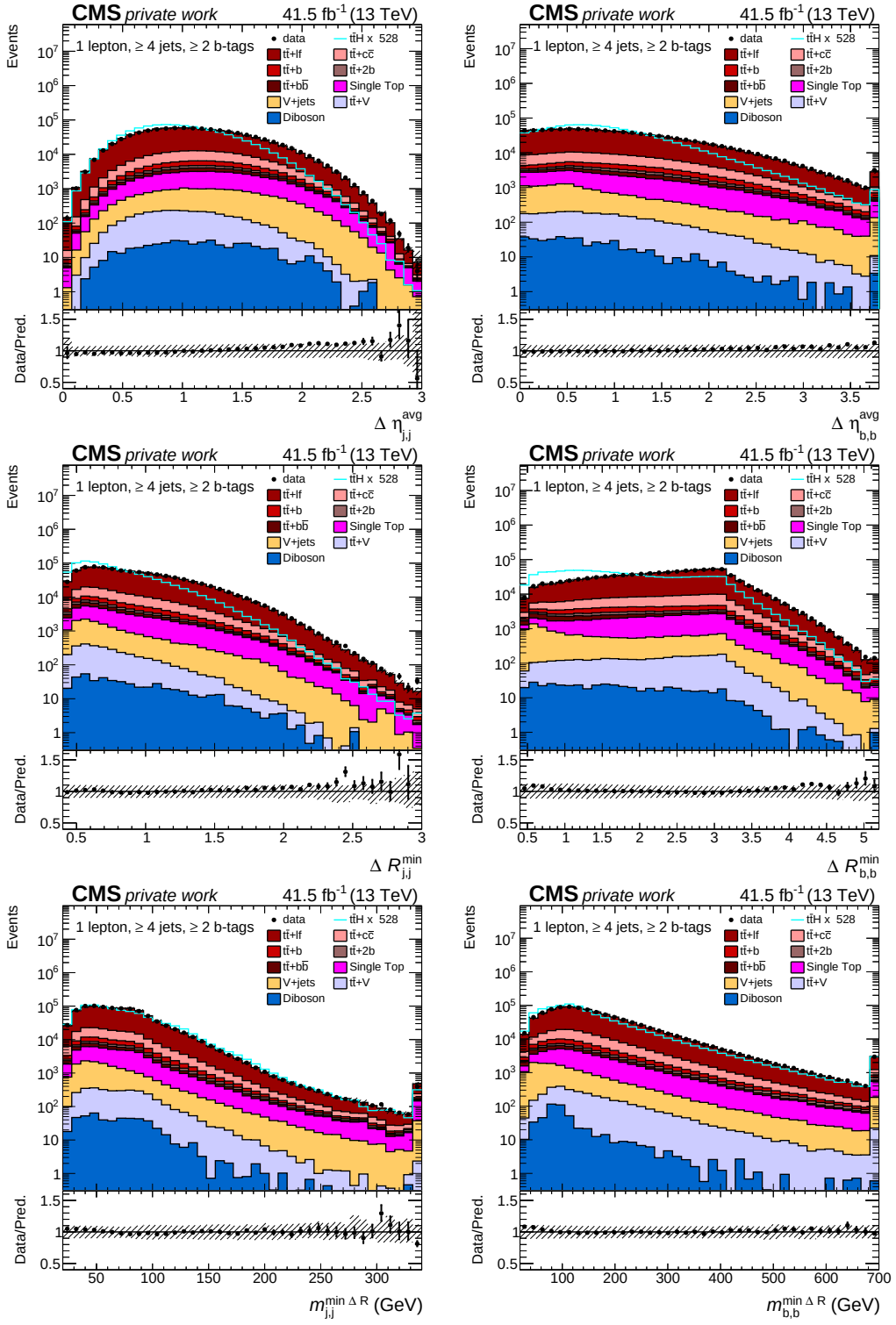


Figure A.11: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2017. A description of the figures is given at the beginning of Sec. A.

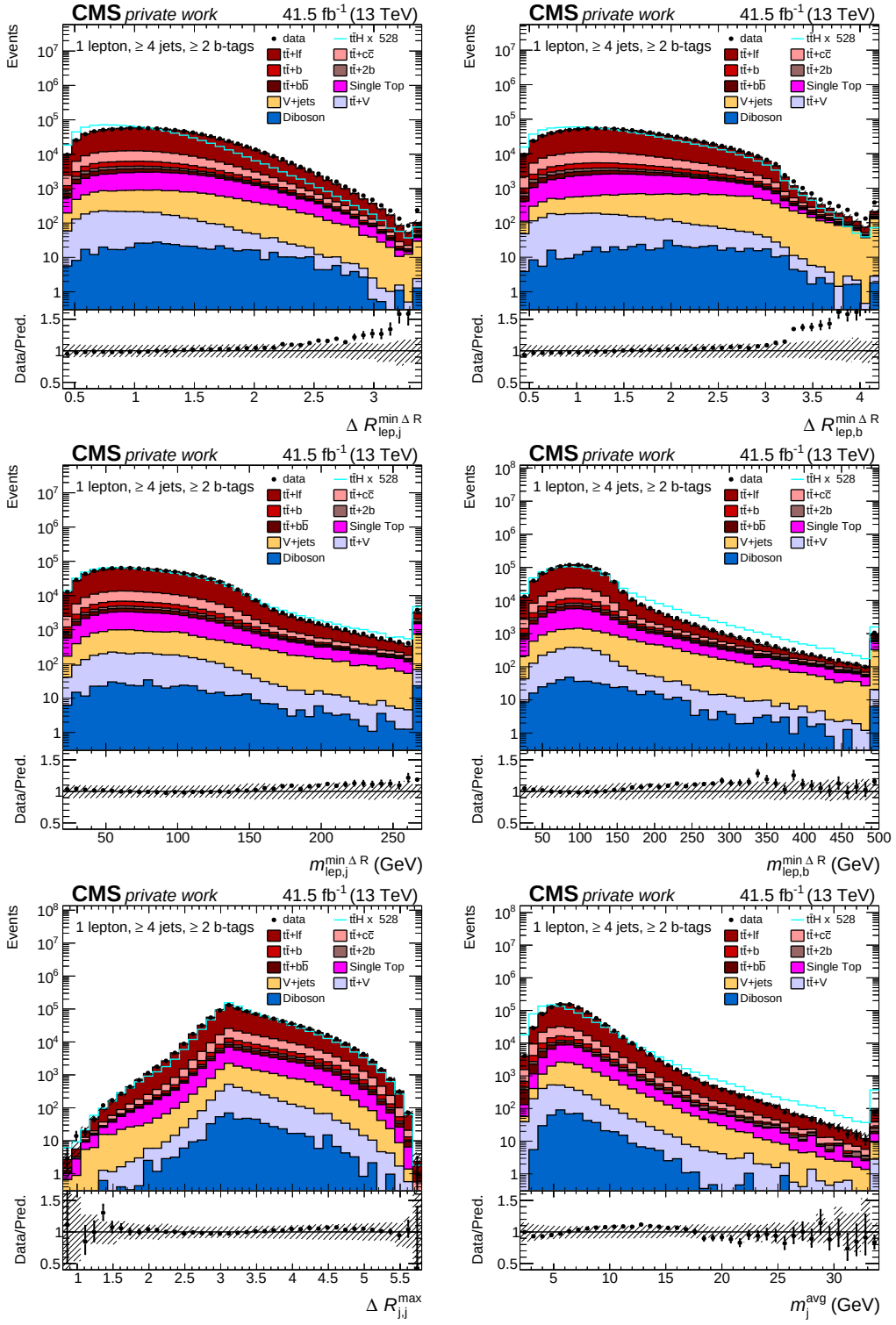


Figure A.12: Basic observables for events with $(1\,\mathrm{l}, \geq 4\,\mathrm{jets}, \geq 2\,\mathrm{b\text{-}tags})$ in data recorded in the year 2017. A description of the figures is given at the beginning of Sec. A.

B Input features of the ANNs

The observables used as input features for the ANNs are shown for the analysis of data recorded in the year 2016 in Sec. B.1 and for the analysis of data recorded in the year 2017 in Sec. B.2. In all figures, the distributions observed in recorded data are shown as black markers, the expected background contributions as stacks of histograms and the $t\bar{t}H$ process as blue line which is additionally scaled to the integral of the background for better visibility. The hatched uncertainty bands contain statistical and systematic uncertainties added in quadrature, excluding those affecting only the rate. The bottom panels show the ratio of data to the background prediction. Definitions of all observables are given Tables 7.7 and 7.8 in Sec. 7.5.2.

B.1 Input features used in the analysis of data recorded in 2016

Input features for the ANN in the $(4\,\text{jets}, \geq 3\,\text{b-tags})$ category

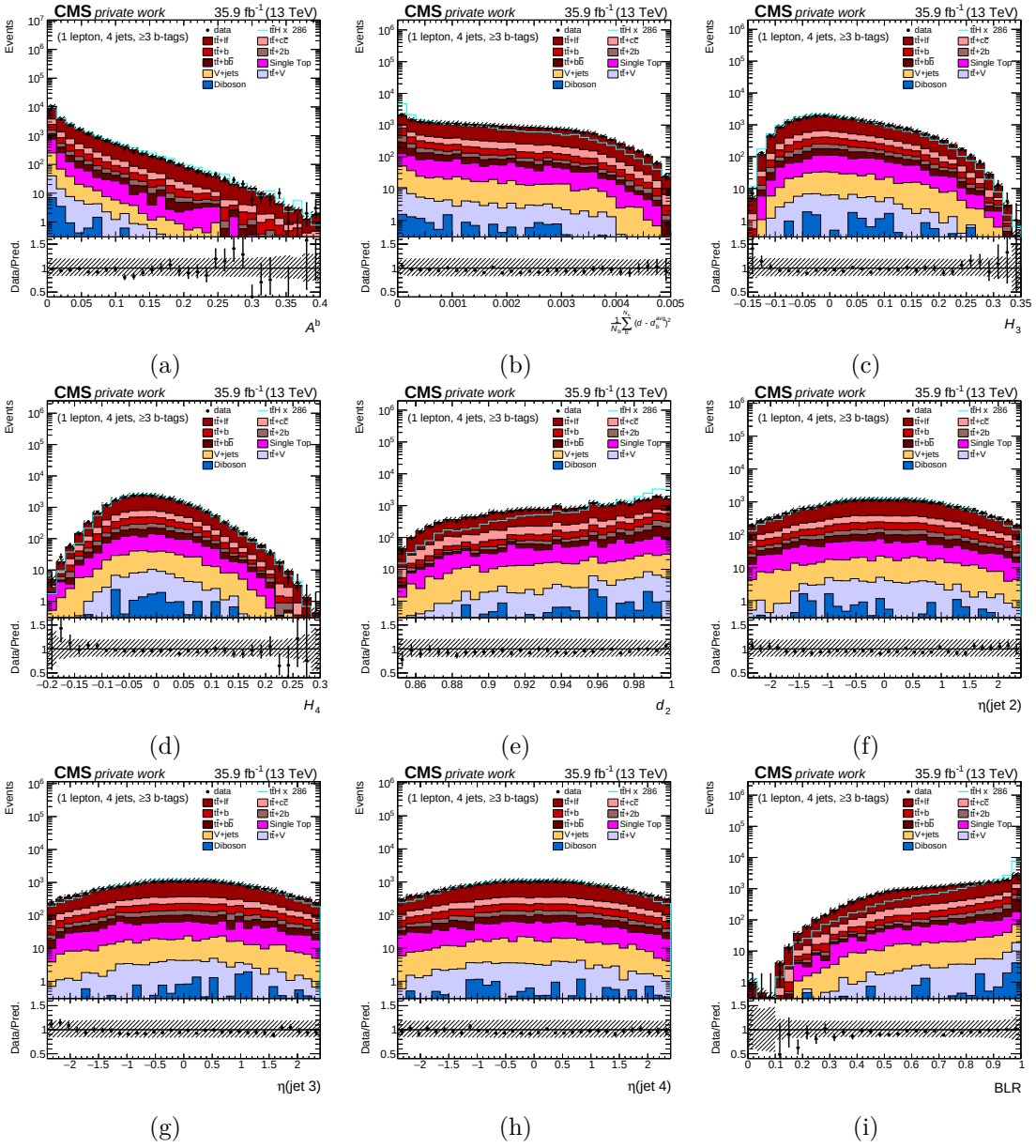


Figure B.13: Observables used as input features for the ANNs for $(4\,\text{jets}, \geq 3\,\text{b-tags})$ events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

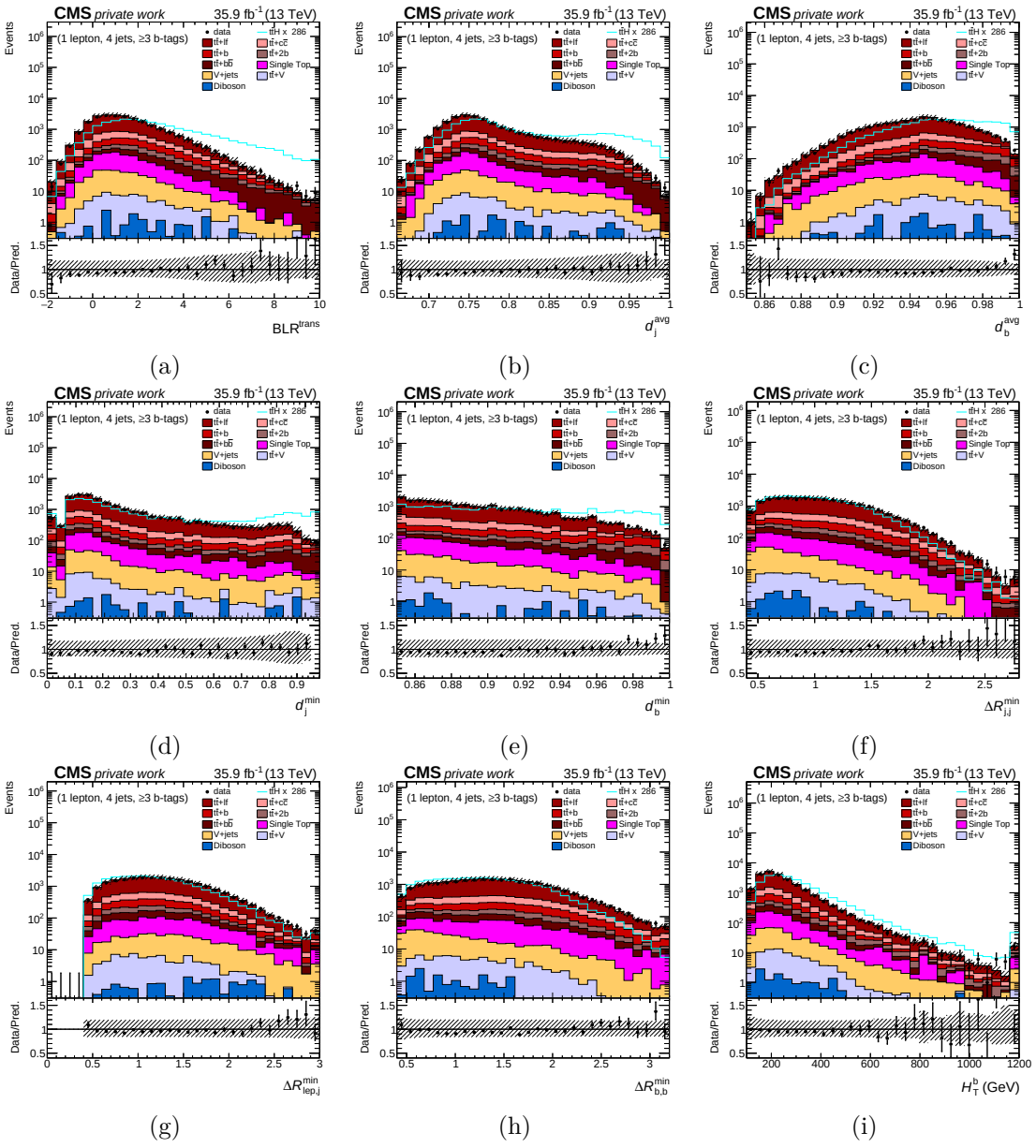


Figure B.14: Observables used as input features for the ANNs for (4 jets, $\geq$ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

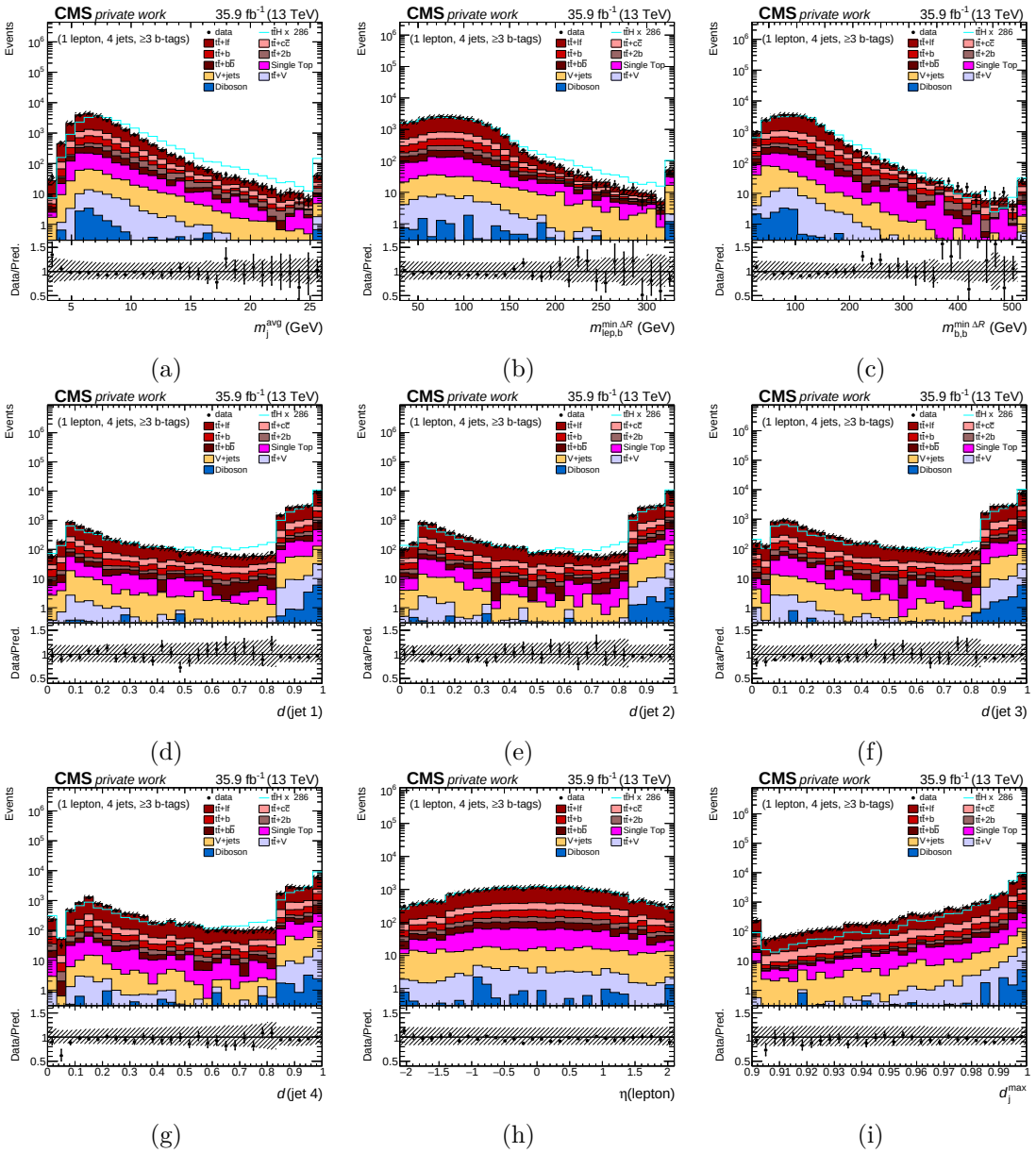
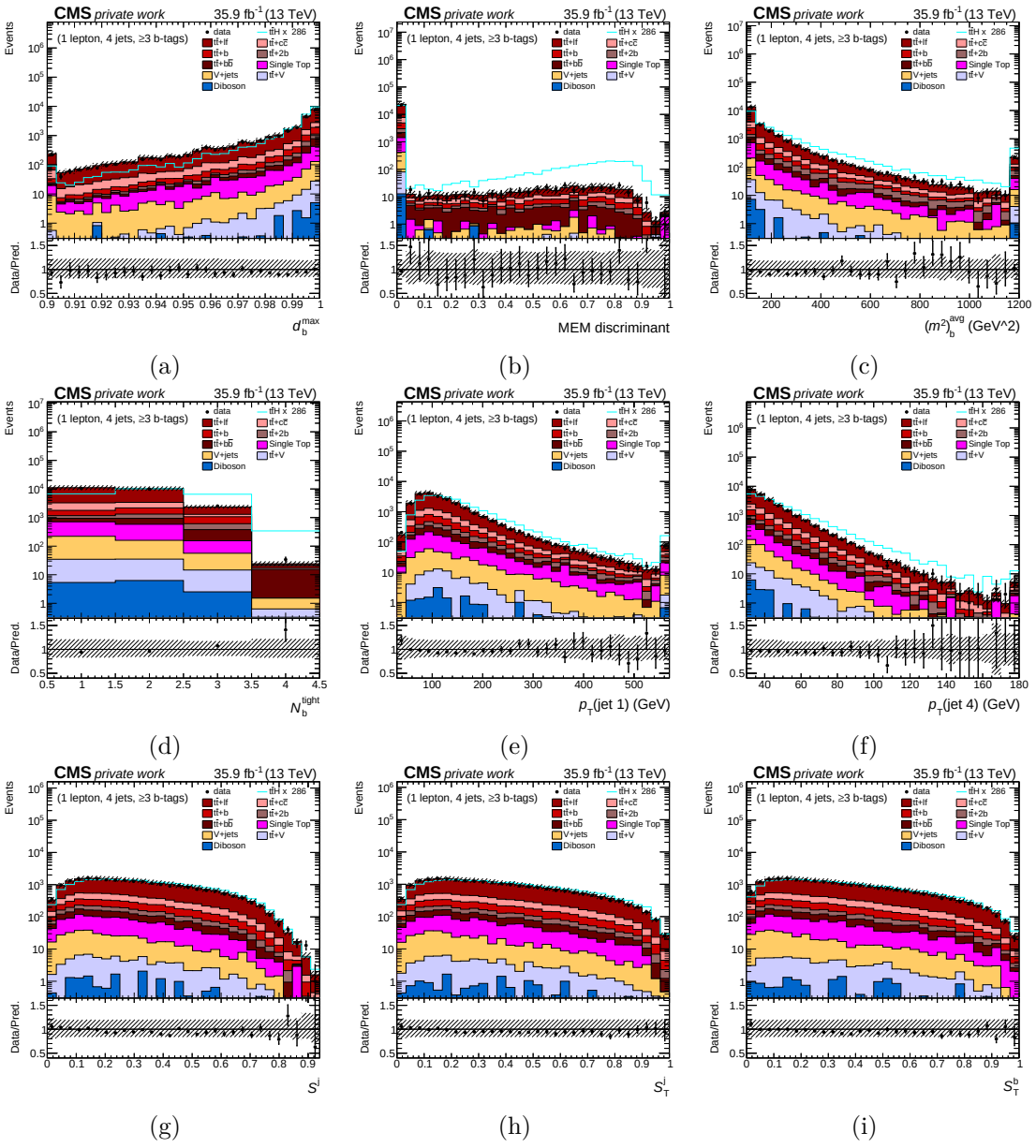


Figure B.15: Observables used as input features for the ANNs for (4 jets, ≥ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure B.16: Observables used as input features for the ANNs for (4 jets, $\geq$ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

Input features for the ANN in the $(5\,\text{jets}, \geq 3\,\text{b-tags})$ category

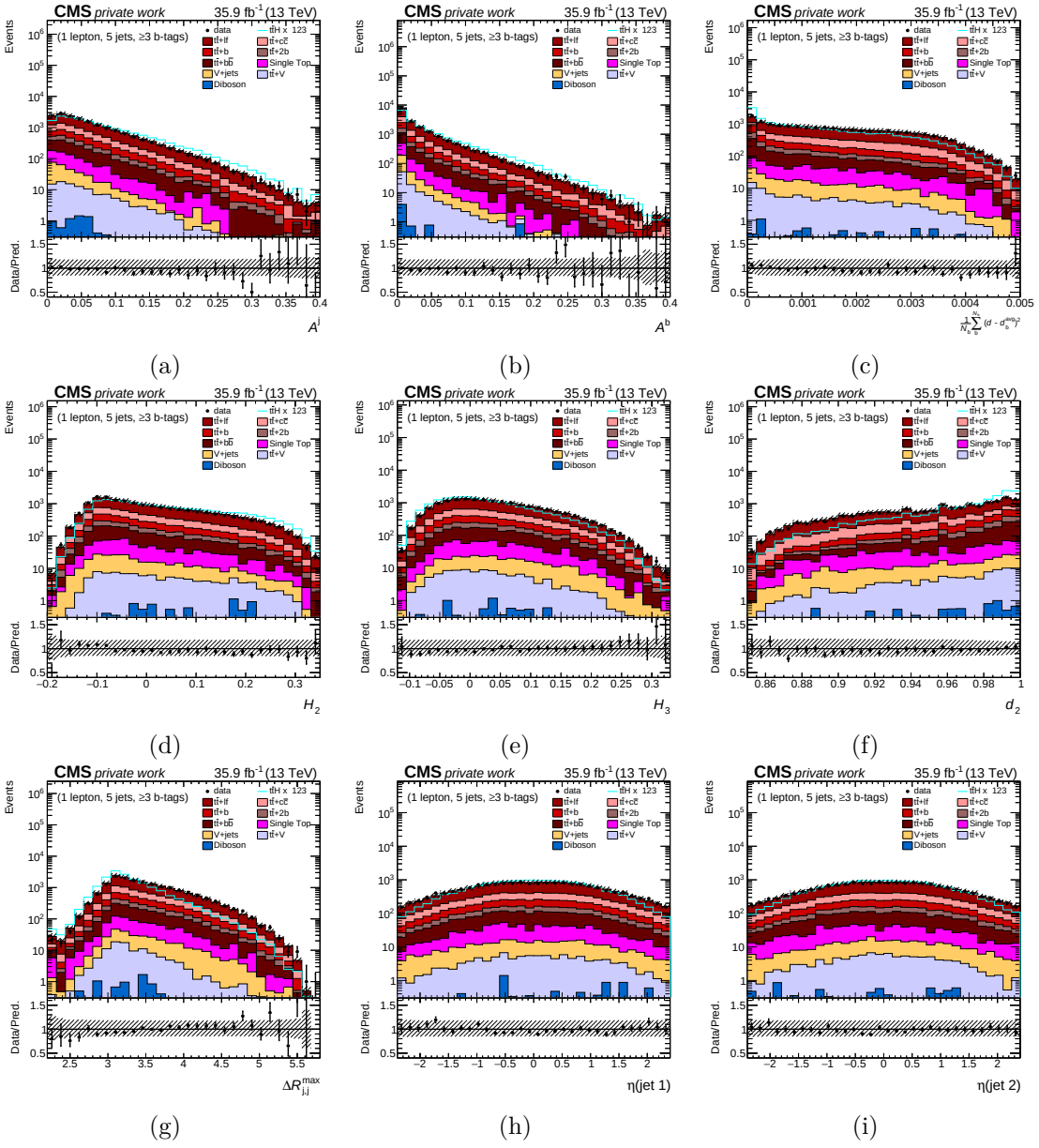


Figure B.17: Observables used as input features for the ANNs for $(5\,\text{jets}, \geq 3\,\text{b-tags})$ events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

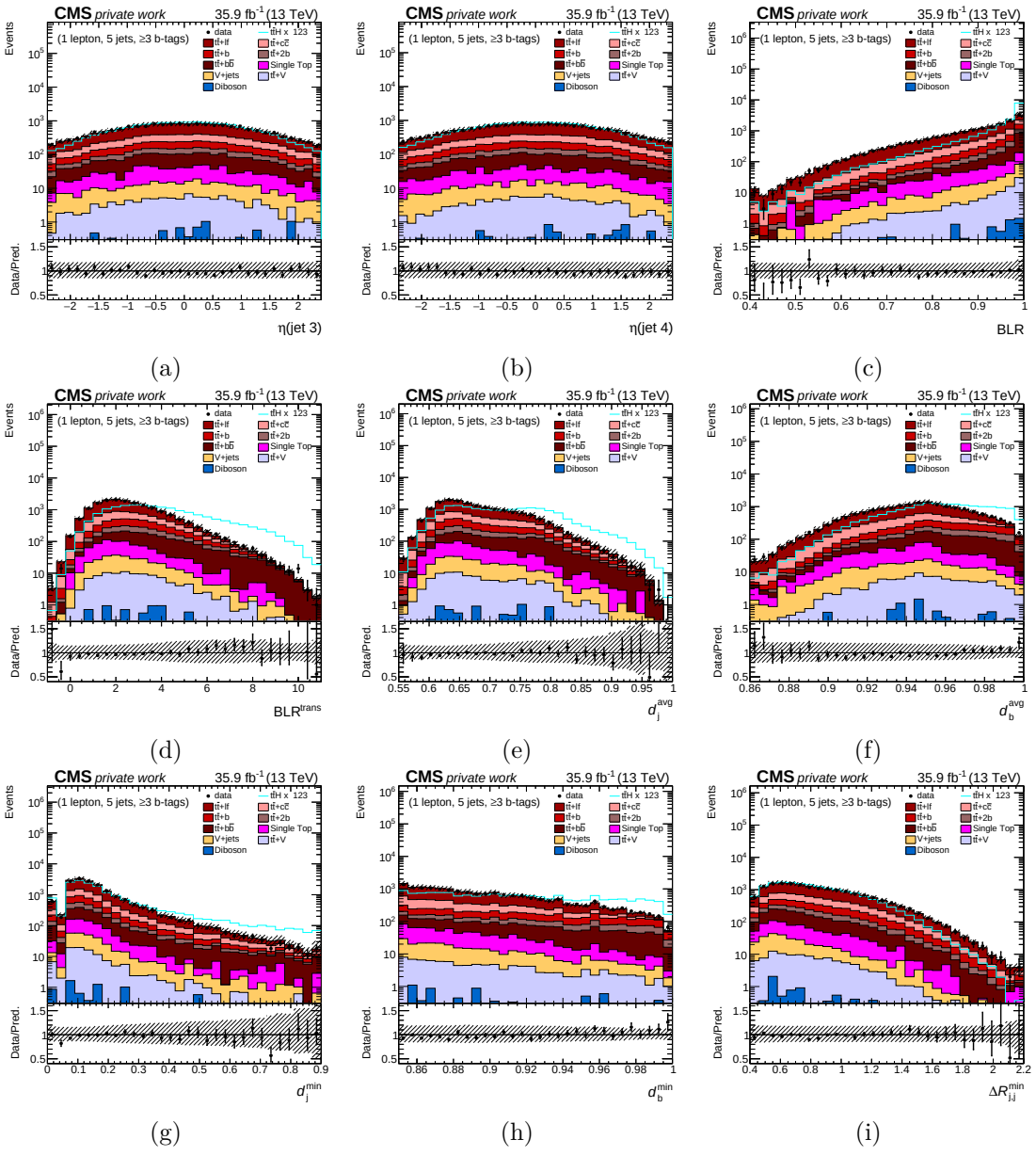


Figure B.18: Observables used as input features for the ANNs for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

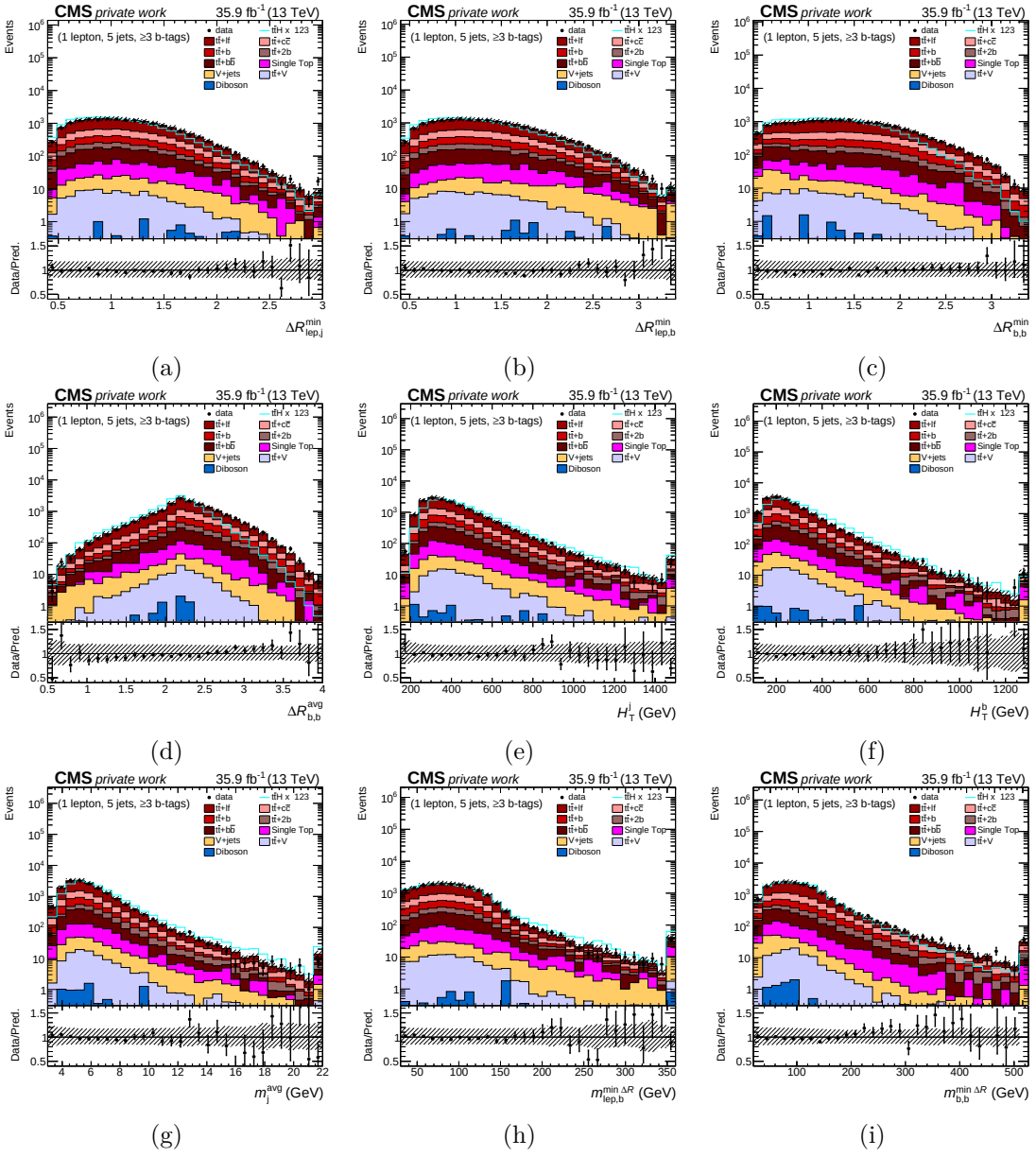
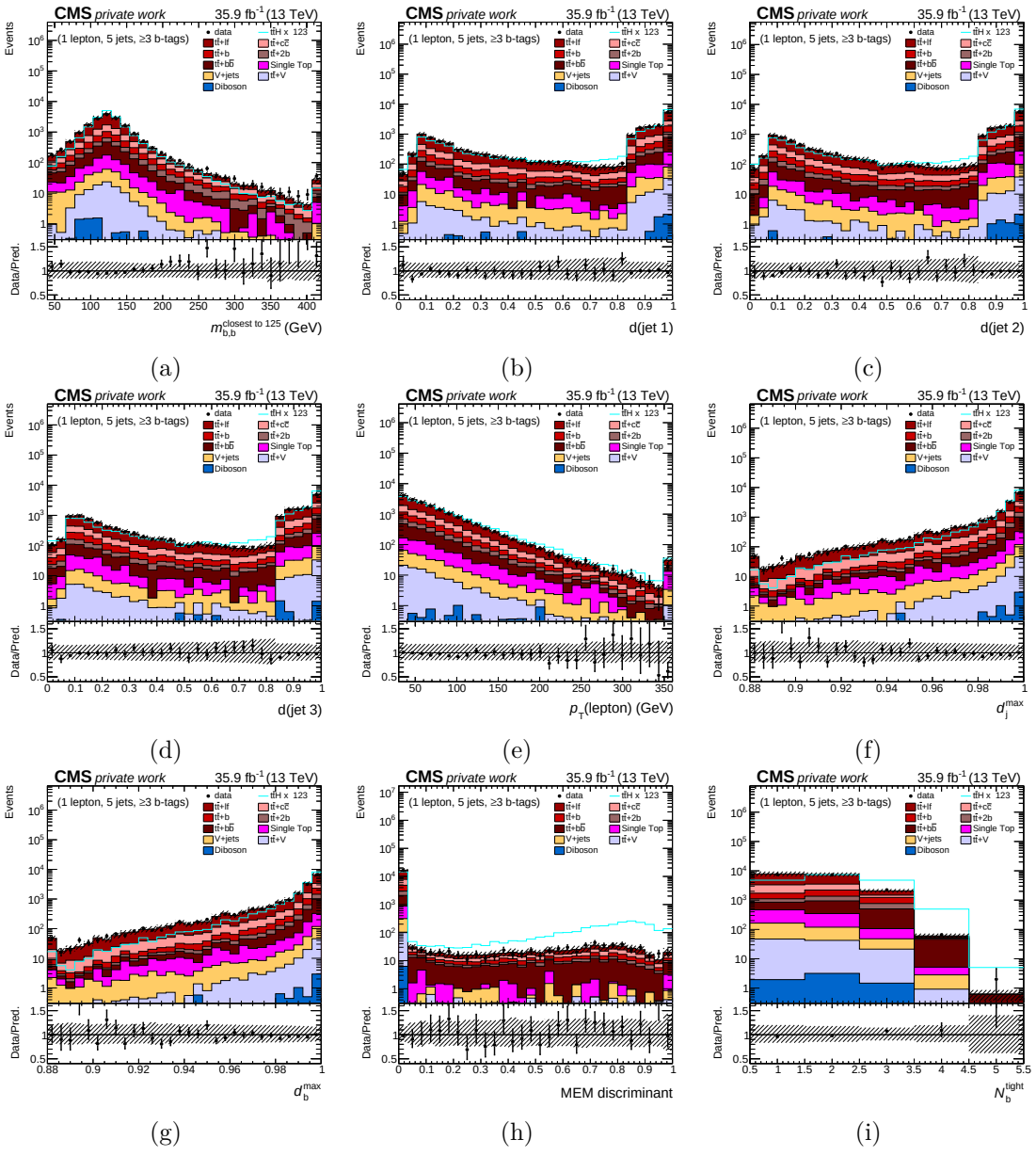


Figure B.19: Observables used as input features for the ANNs for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure B.20: Observables used as input features for the ANNs for (5 jets, $\geq$ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

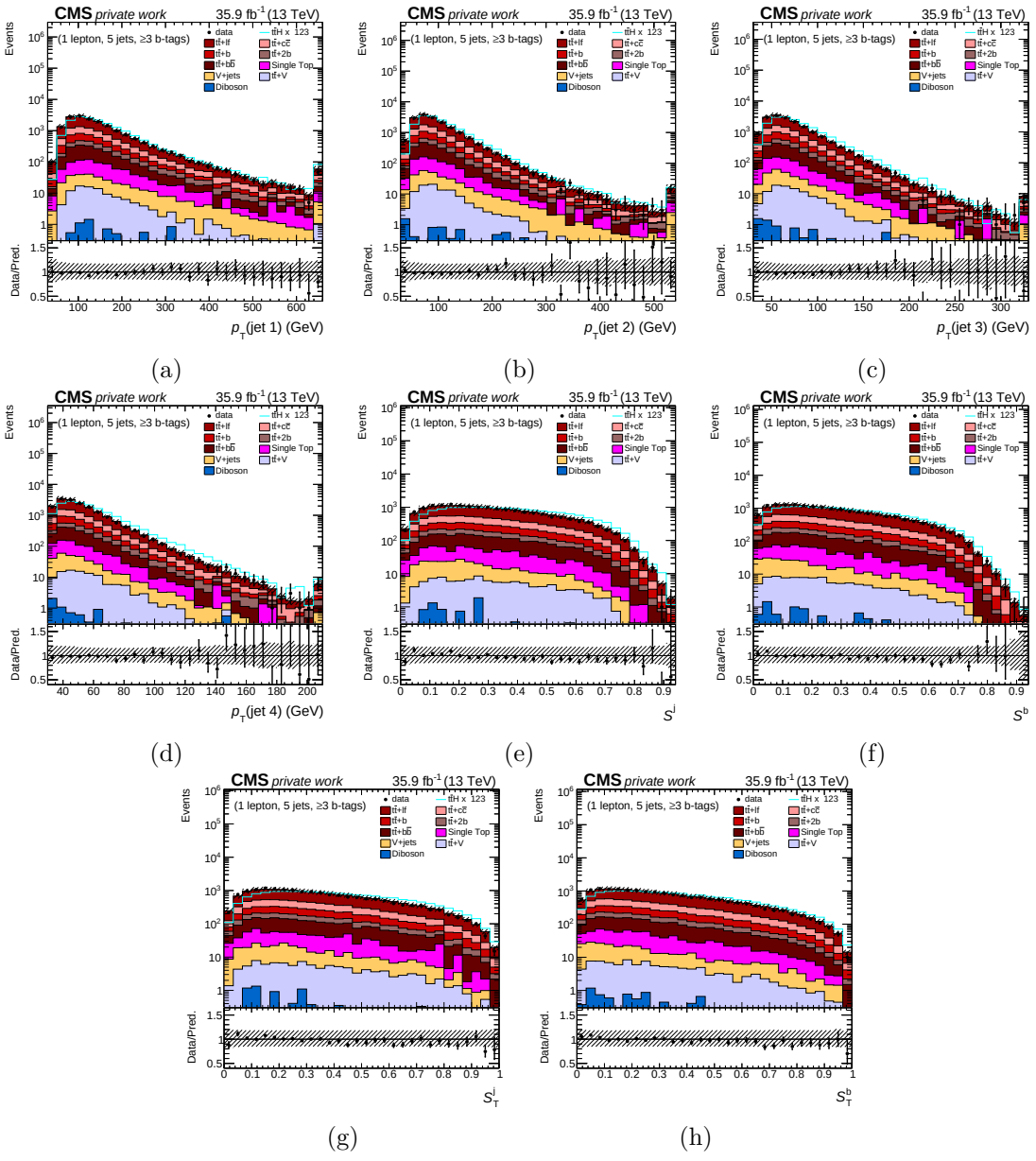


Figure B.21: Observables used as input features for the ANNs for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

Input features for the ANN in the (≥ 6 jets, ≥ 3 b-tags) category

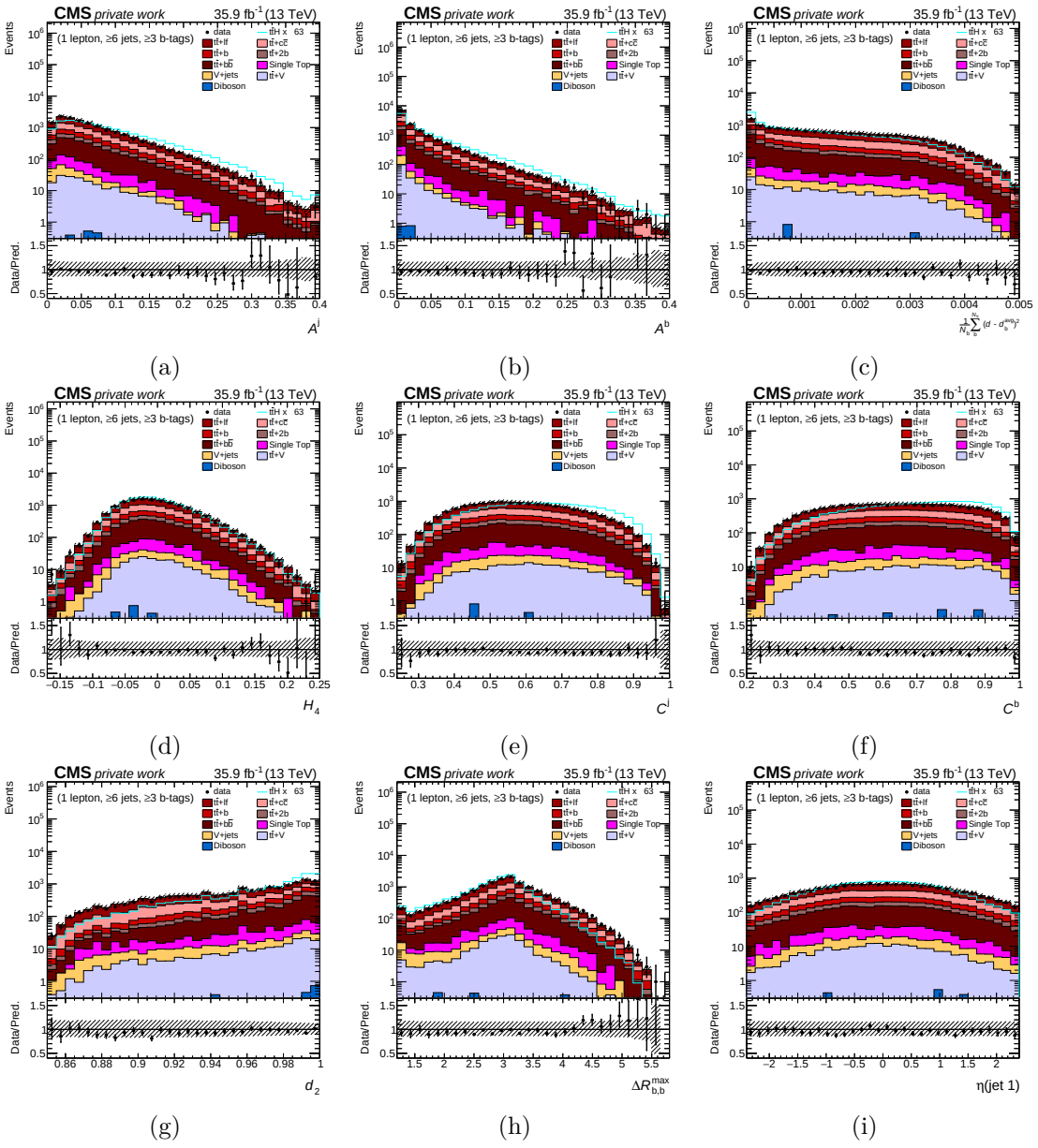
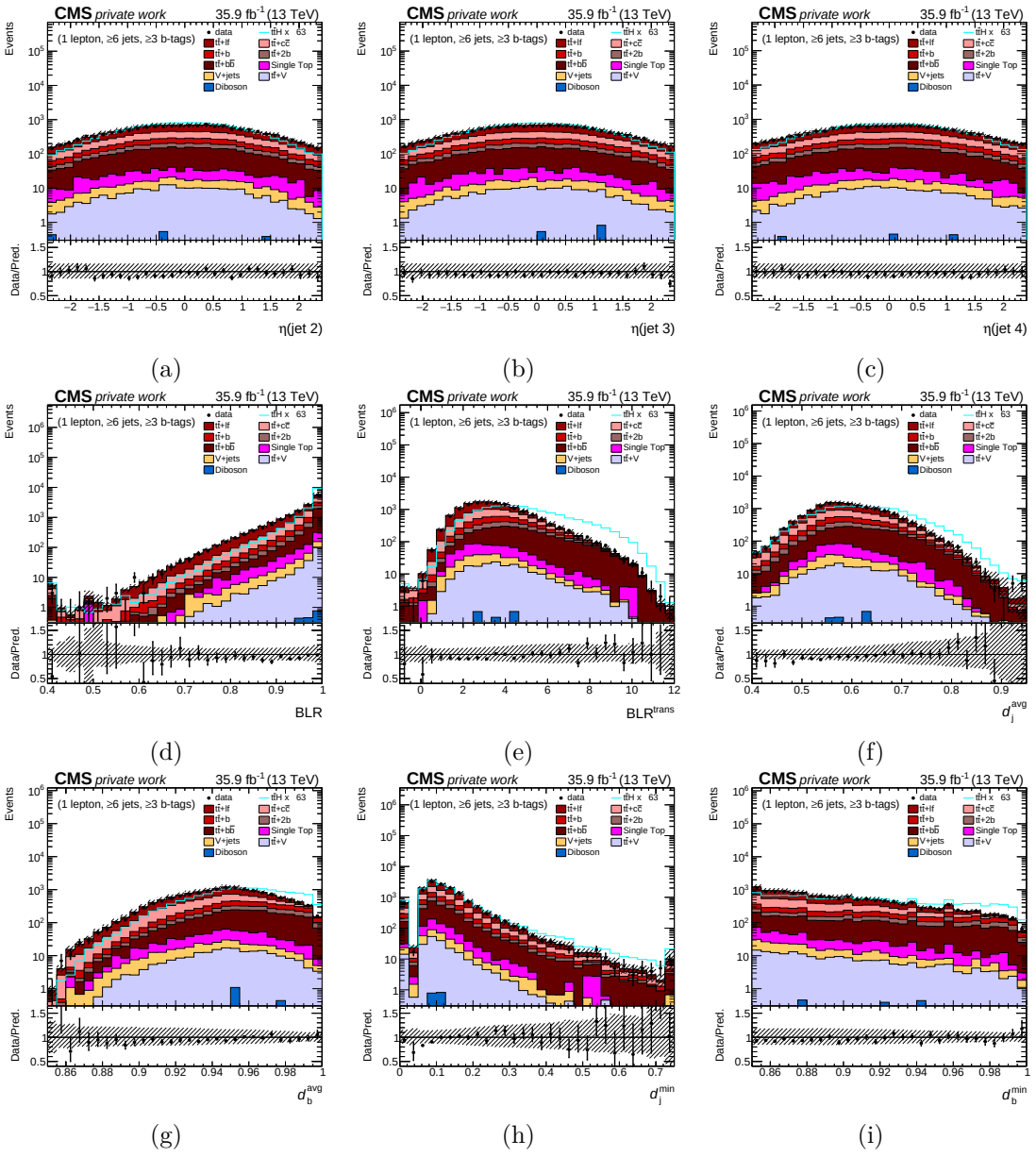


Figure B.22: Observables used as input features for the ANNs for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure B.23: Observables used as input features for the ANNs for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

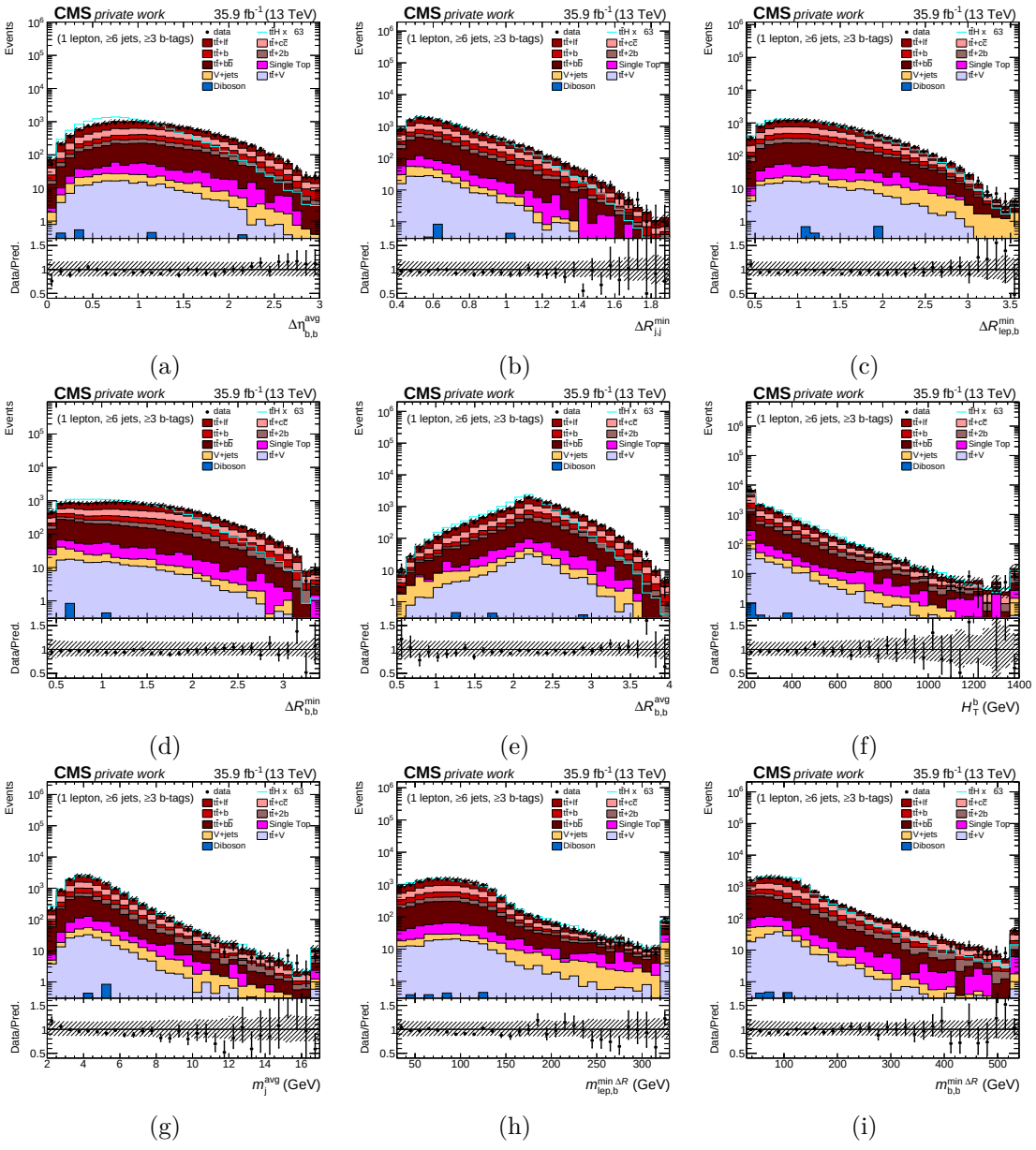


Figure B.24: Observables used as input features for the ANNs for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

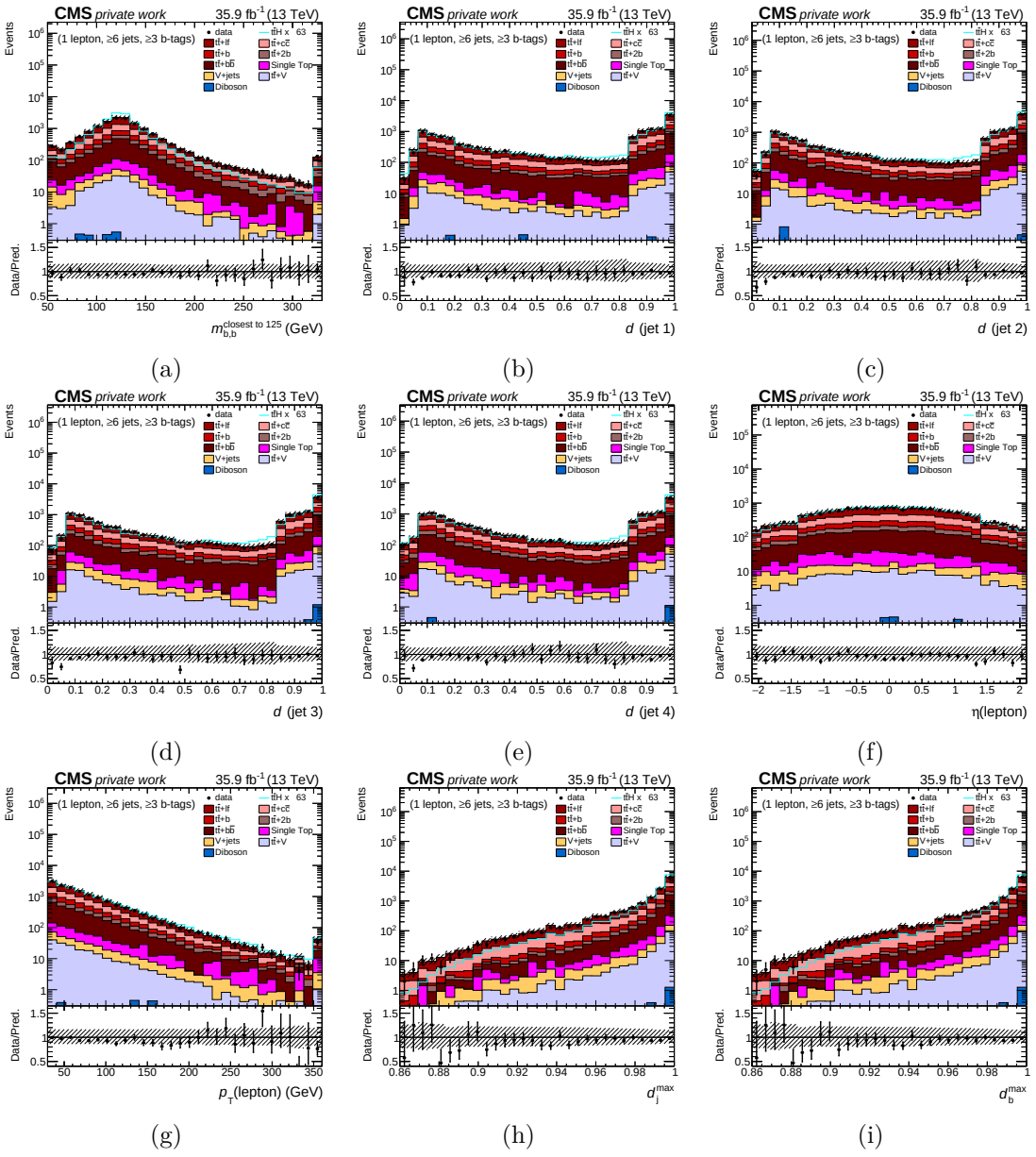


Figure B.25: Observables used as input features for the ANNs for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

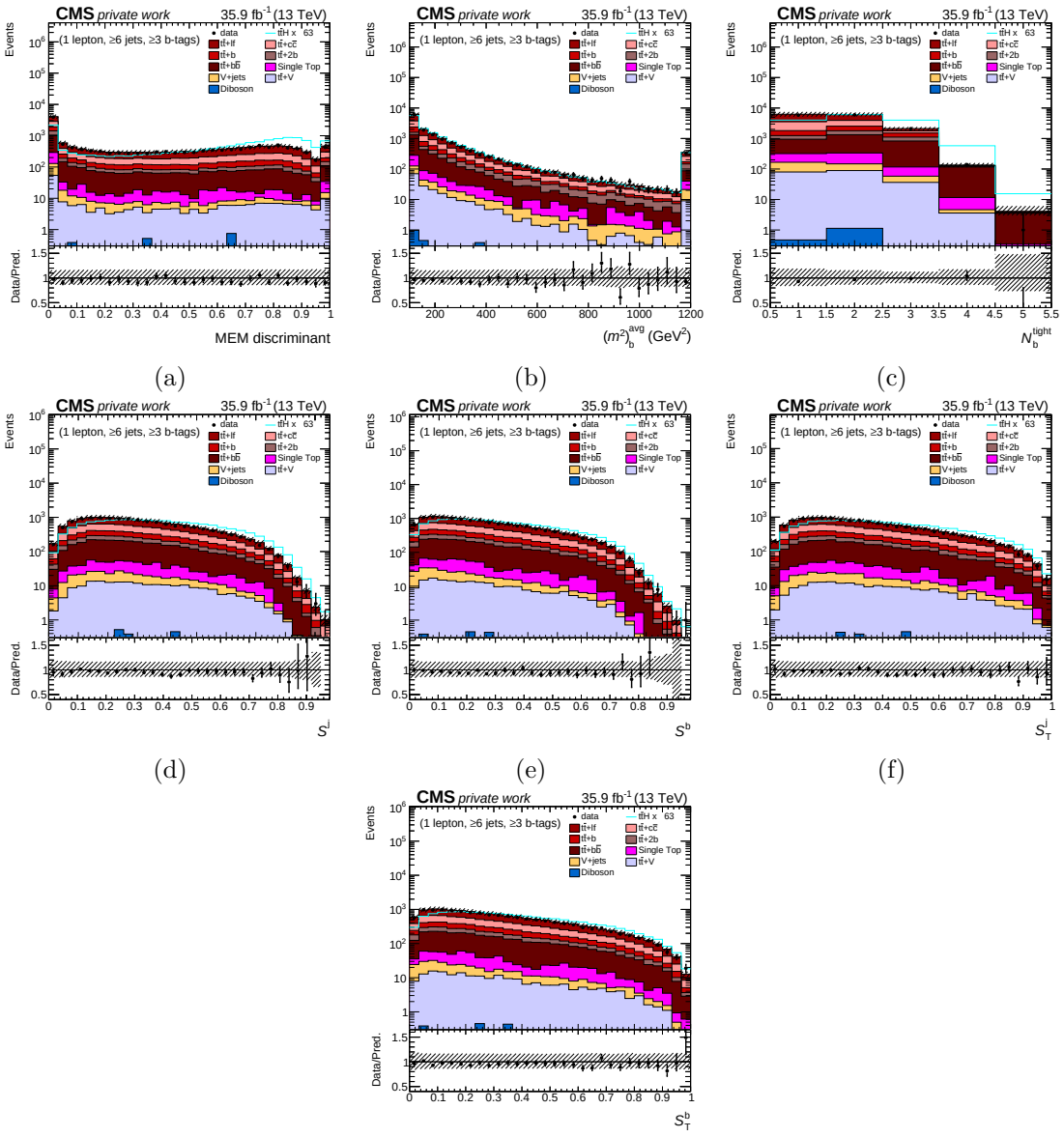


Figure B.26: Observables used as input features for the ANNs for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016. A description of the figures is given at the beginning of Sec. B.

B.2 Input features used in the analysis of data recorded in 2017

Input features for the ANN in the $(4\,\text{jets}, \geq 3\,\text{b-tags})$ category

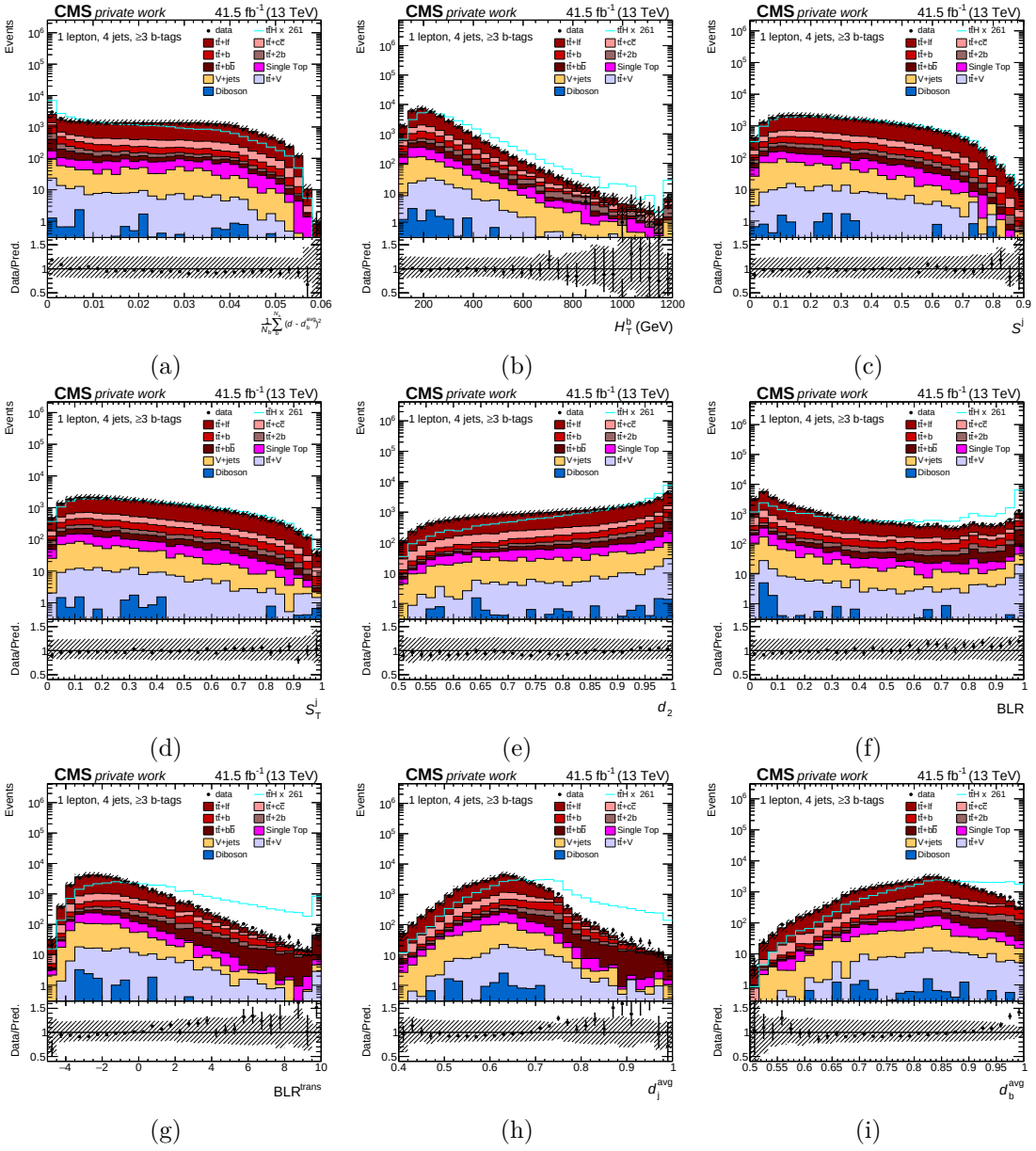


Figure B.27: Observables used as input features for the ANNs for $(4\,\text{jets}, \geq 3\,\text{b-tags})$ events recorded in the year 2017. A description of the figures is given at the beginning of Sec. B.

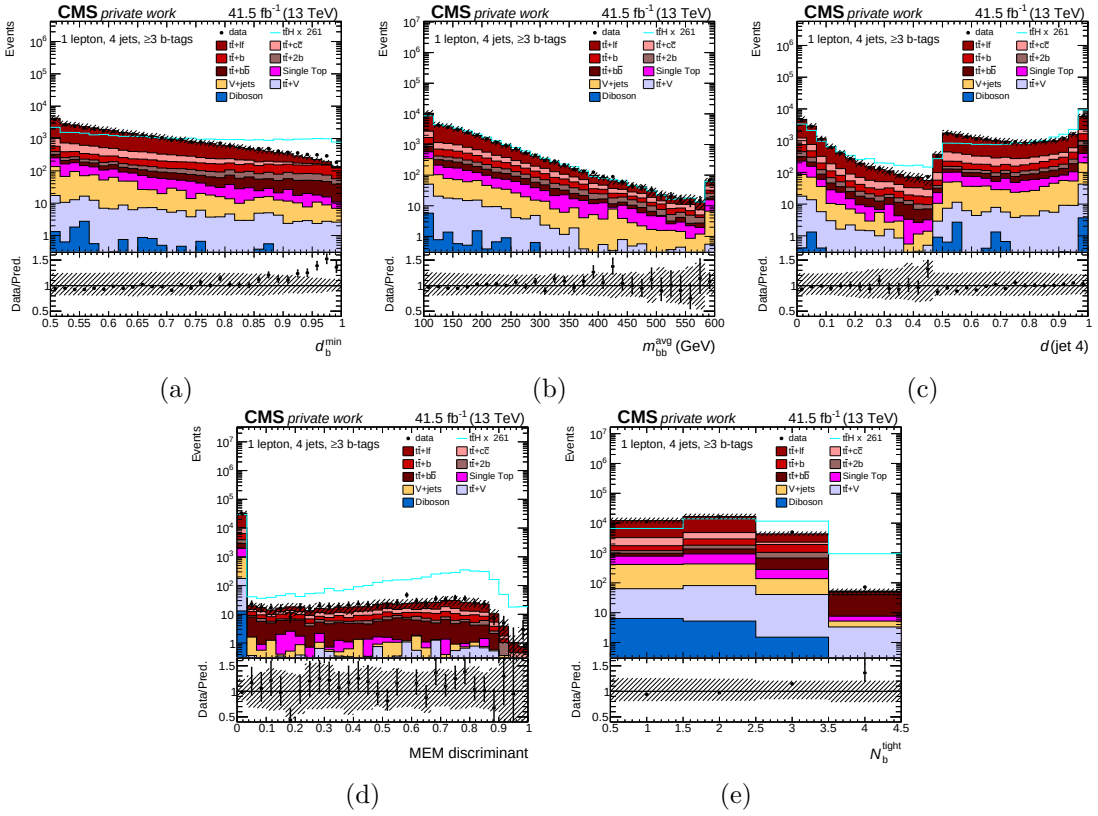


Figure B.28: Observables used as input features for the ANNs for (4 jets, $\geq$ 3 b-tags) events recorded in the year 2017. A description of the figures is given at the beginning of Sec. B.

Input features for the ANN in the $(5\,\text{jets}, \geq 3\,\text{b-tags})$ category

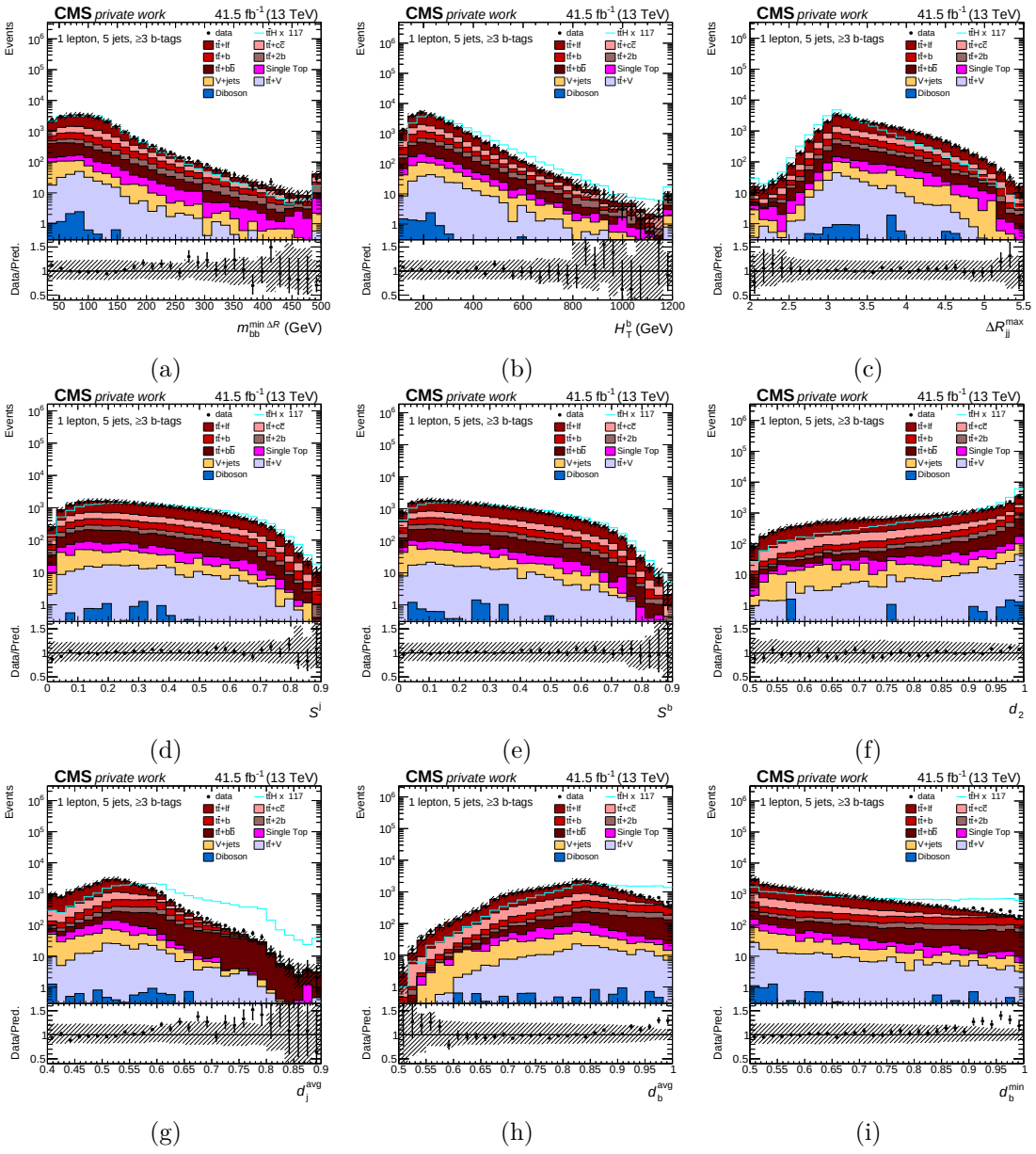


Figure B.29: Observables used as input features for the ANNs for $(5\,\text{jets}, \geq 3\,\text{b-tags})$ events recorded in the year 2017. A description of the figures is given at the beginning of Sec. B.

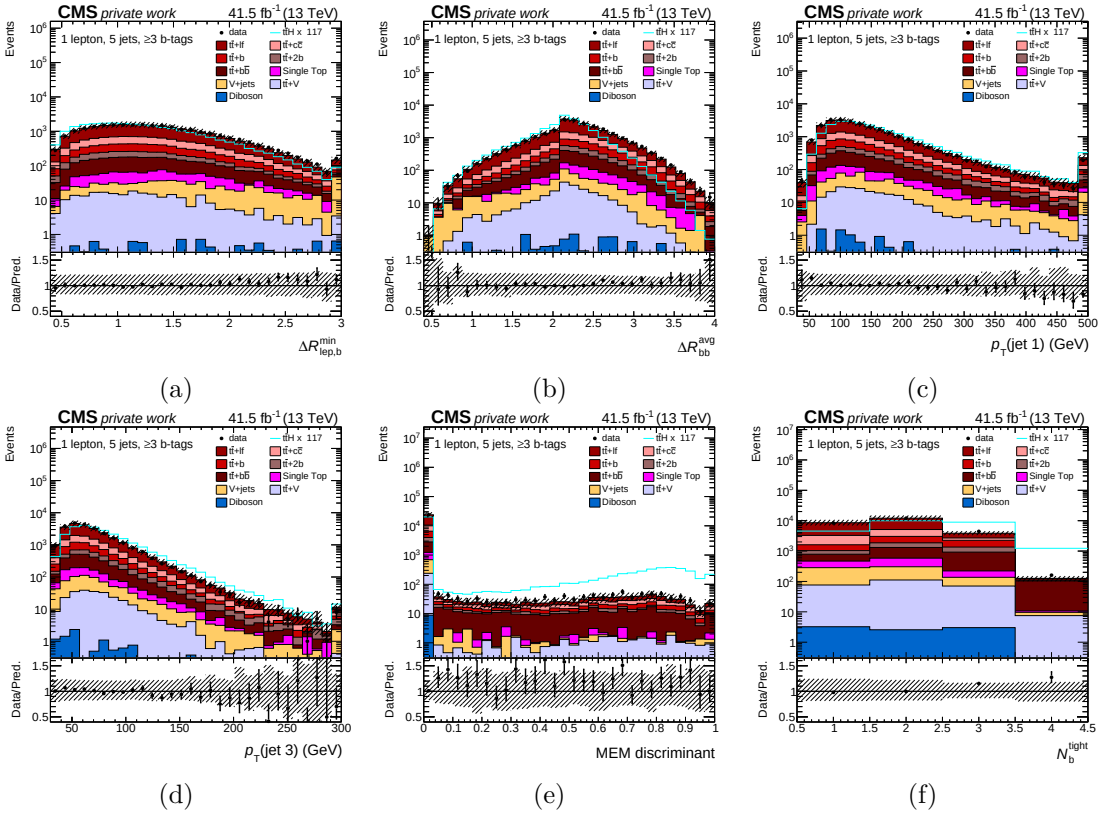


Figure B.30: Observables used as input features for the ANNs for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2017. A description of the figures is given at the beginning of Sec. B.

Input features for the ANN in the (≥ 6 jets, ≥ 3 b-tags) category

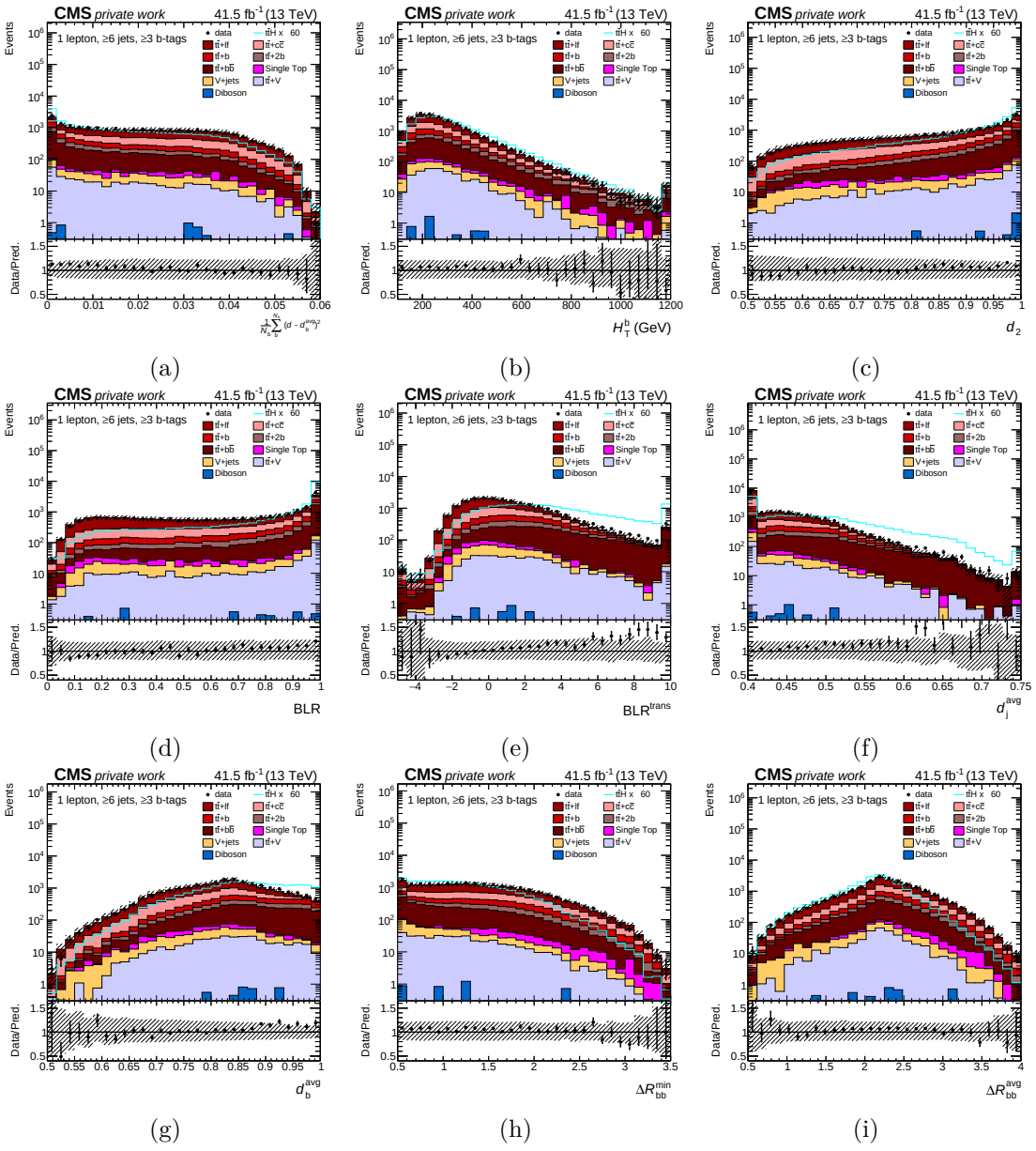
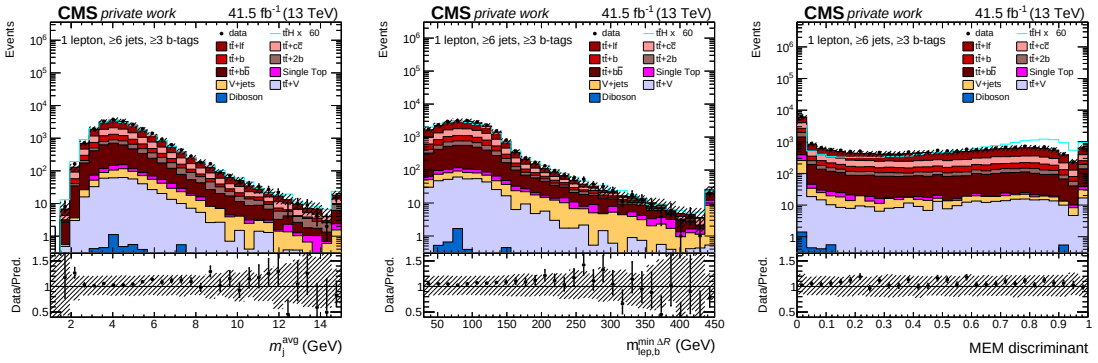


Figure B.31: Observables used as input features for the ANNs for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2017. A description of the figures is given at the beginning of Sec. B.



(a) (b) (c)

Figure B.32: Observables used as input features for the ANNs for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2017. A description of the figures is given at the beginning of Sec. B.

C ANN discriminant distributions

In this section the ANN discriminant distributions before and after the fit to data are shown.
The distributions for the analyses of data recorded the years 2016 and 2017 are shown before the fit to data in Sec. C.1 and Sec. C.2, respectively.

The distributions after the fit to data recorded in 2016 are shown in Sec. C.3 and those after the fit to data recorded in 2017 in Sec. C.4.

The distributions after the joint fit to data recorded in both years are shown in Sec. C.5.

In all figures, the distributions observed in recorded data are shown as black markers, the background contributions as stacks of histograms and the $t\bar{t}H$ process as blue line. In the figures before the fit to data, the background processes are scaled to their standard model prediction and the $t\bar{t}H$ process is scaled by a factor of 15 for better visibility. After the fit to data, the background and the $t\bar{t}H$ processes are scaled according to the fit results. The hatched uncertainty bands show the total uncertainty before or after the fit. The bottom panels show the ratio of data to the sum of the background processes.

C.1 Distributions for the analysis of 2016 data before the fit to data

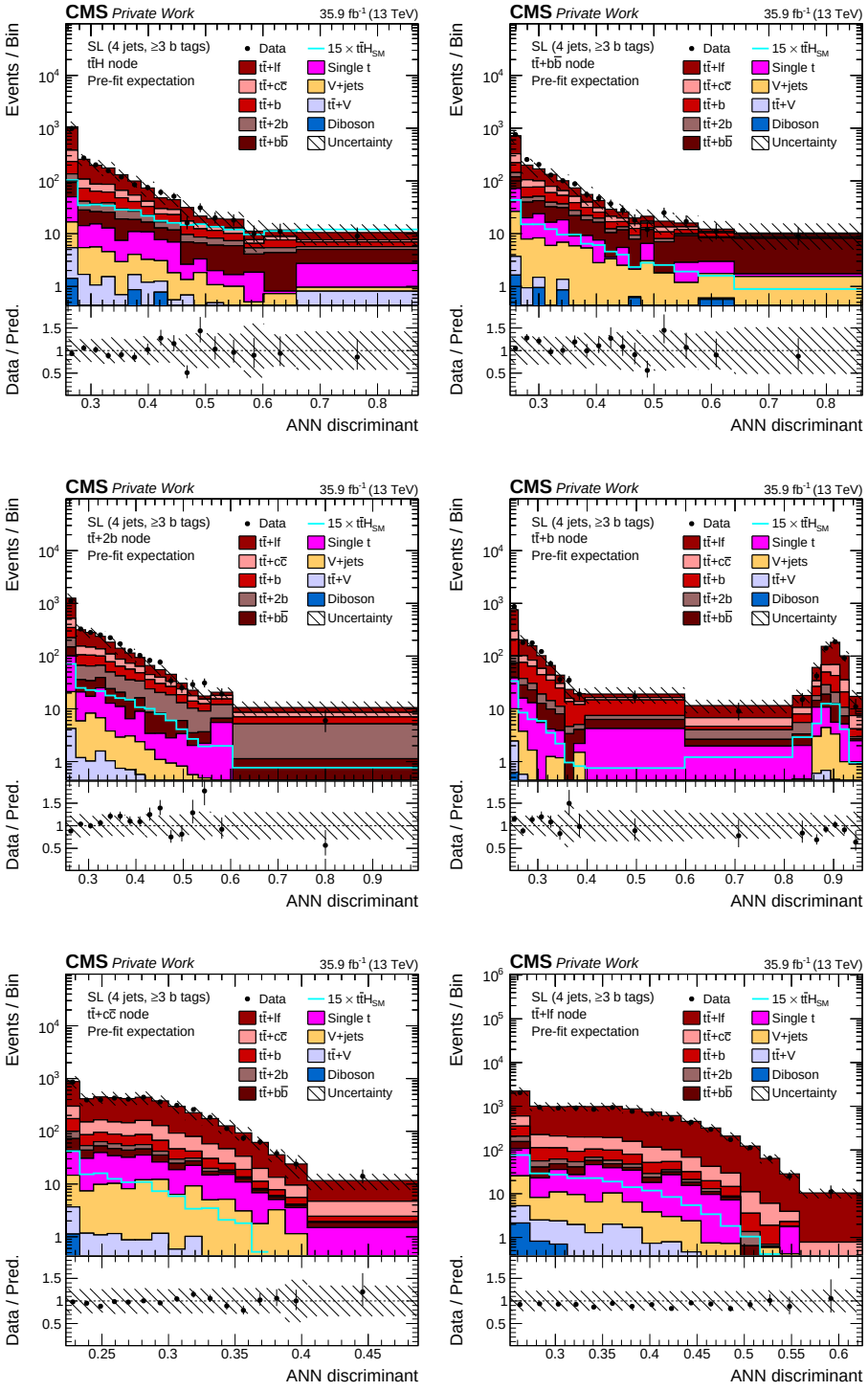


Figure C.33: ANN discriminant distributions for (4 jets, ≥ 3 b-tags) events recorded in the year 2016 before the fit to data. A description of the figures is given at the beginning of Sec. C.

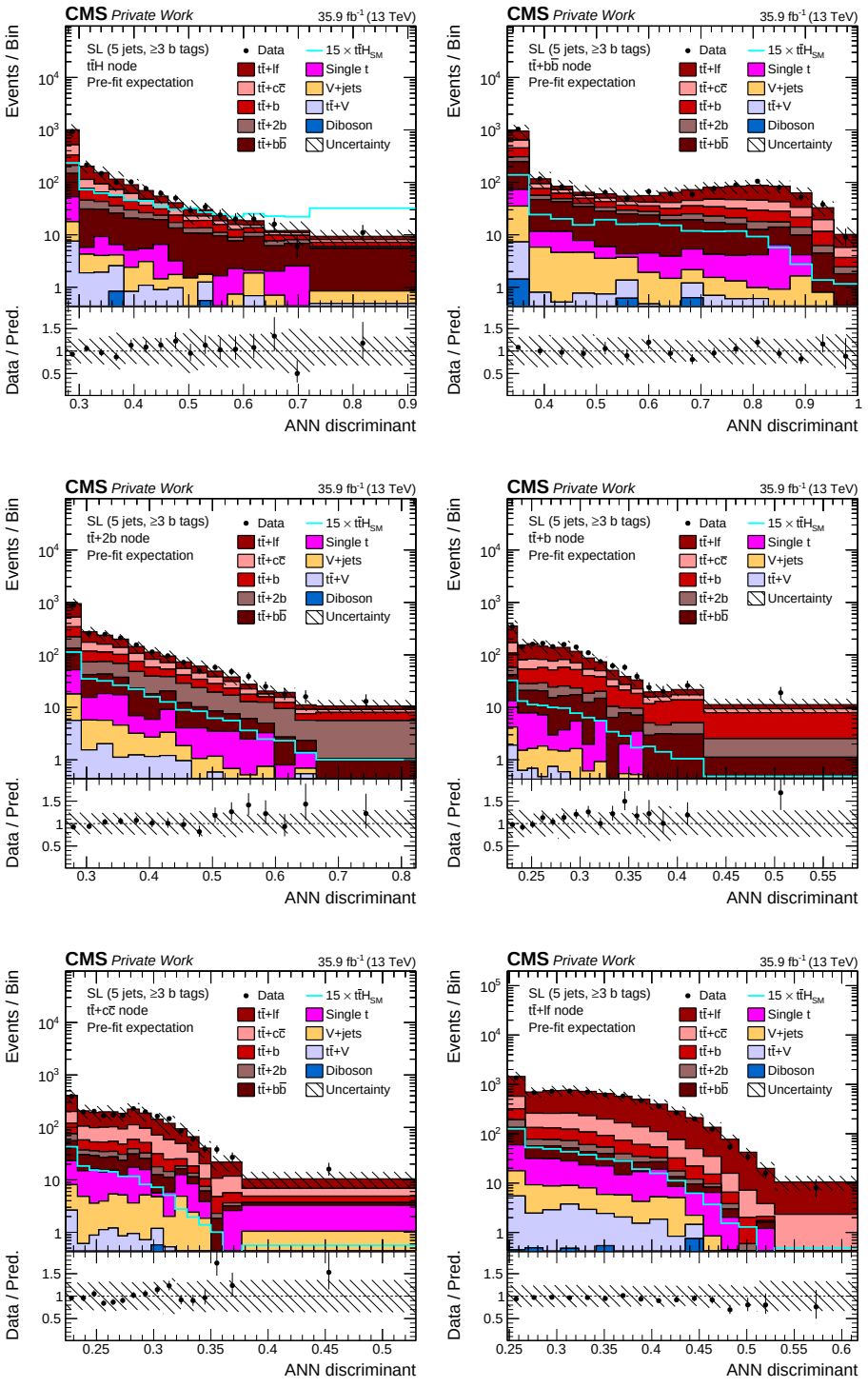


Figure C.34: ANN discriminant distributions for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2016 before the fit to data. A description of the figures is given at the beginning of Sec. C.

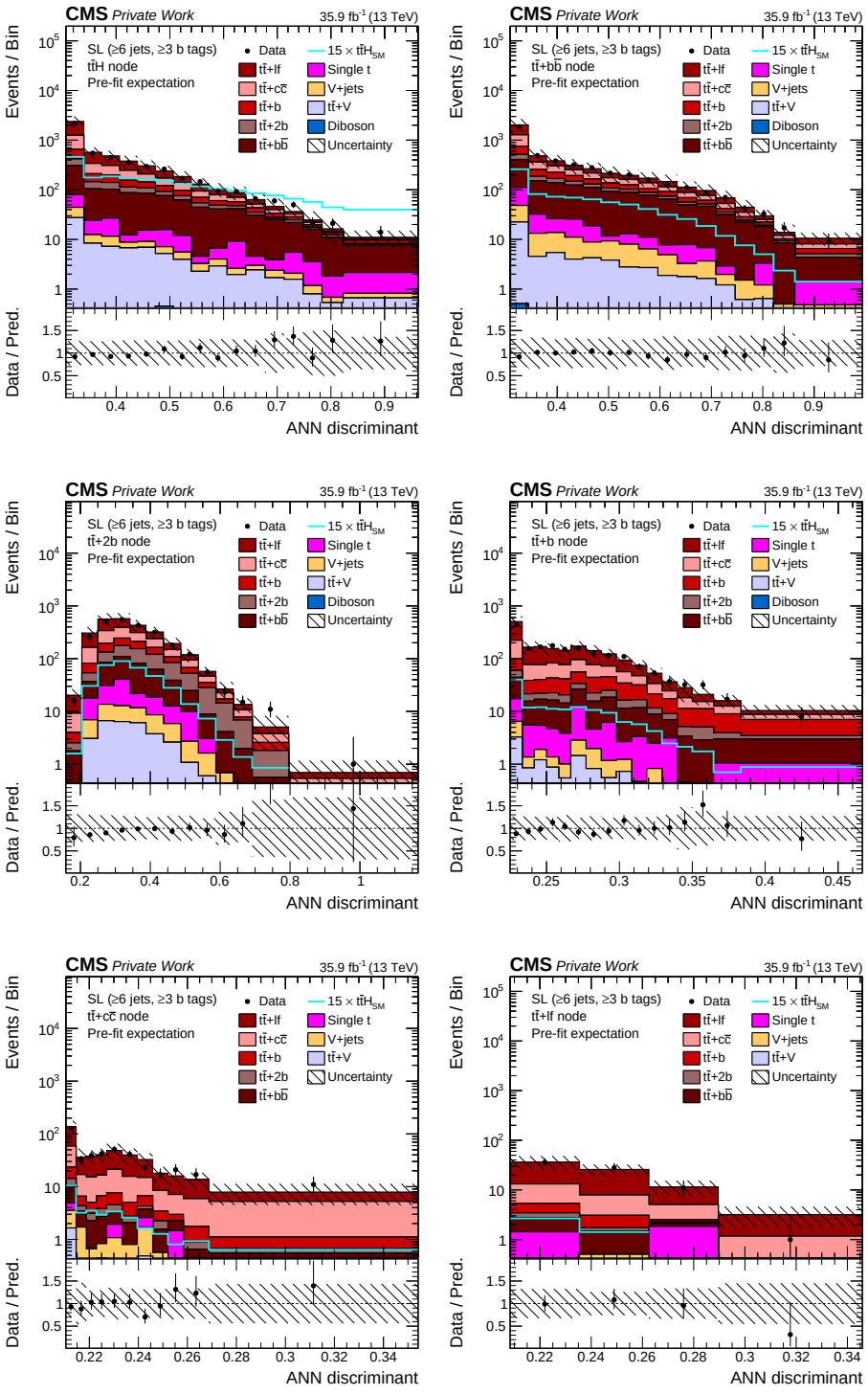


Figure C.35: ANN discriminant distributions for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016 before the fit to data. A description of the figures is given at the beginning of Sec. C.

C.2 Distributions for the analysis of 2017 data before the fit to data

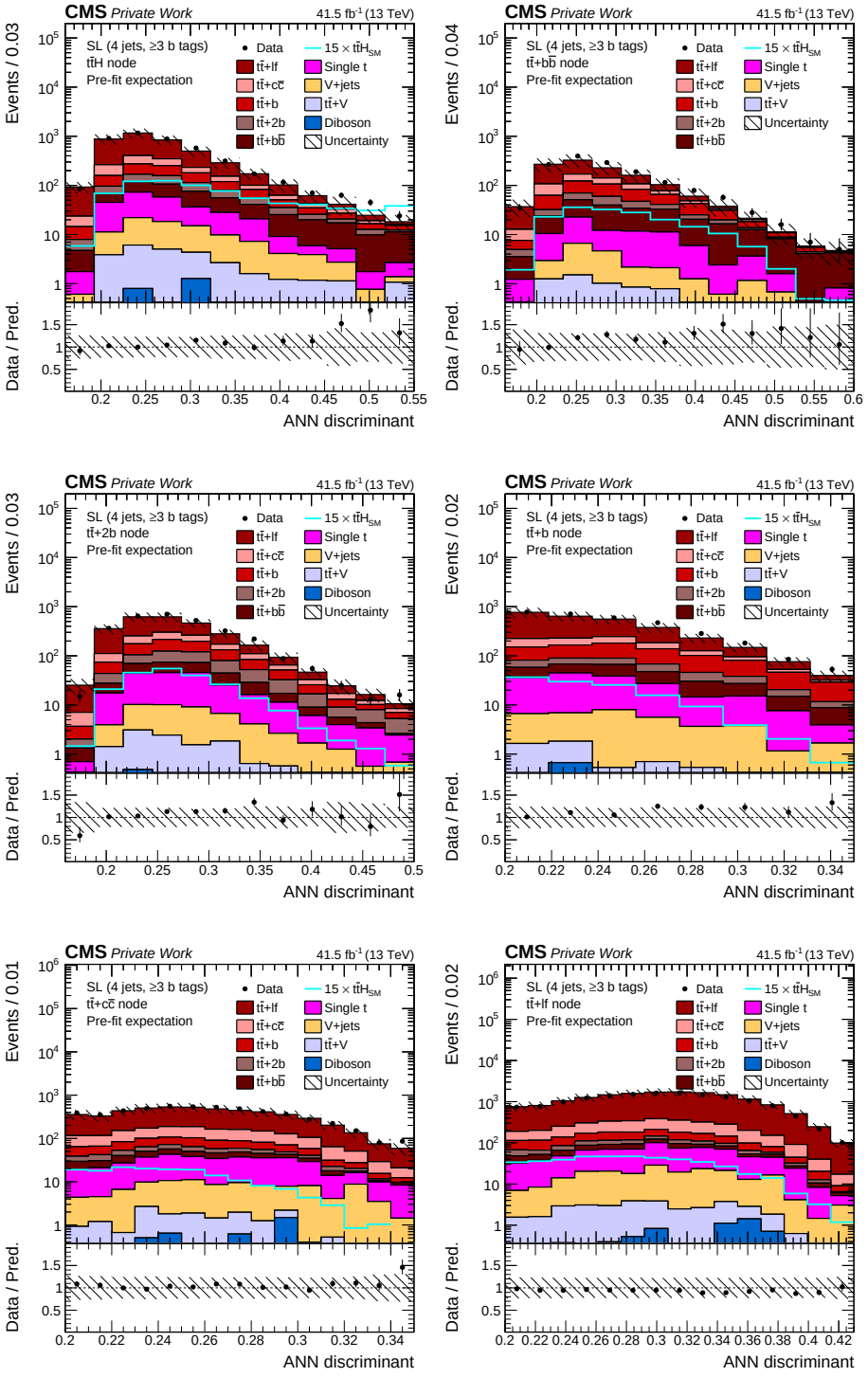


Figure C.36: ANN discriminant distributions for (4 jets, ≥ 3 b-tags) events recorded in the year 2017 before the fit to data. A description of the figures is given at the beginning of Sec. C.

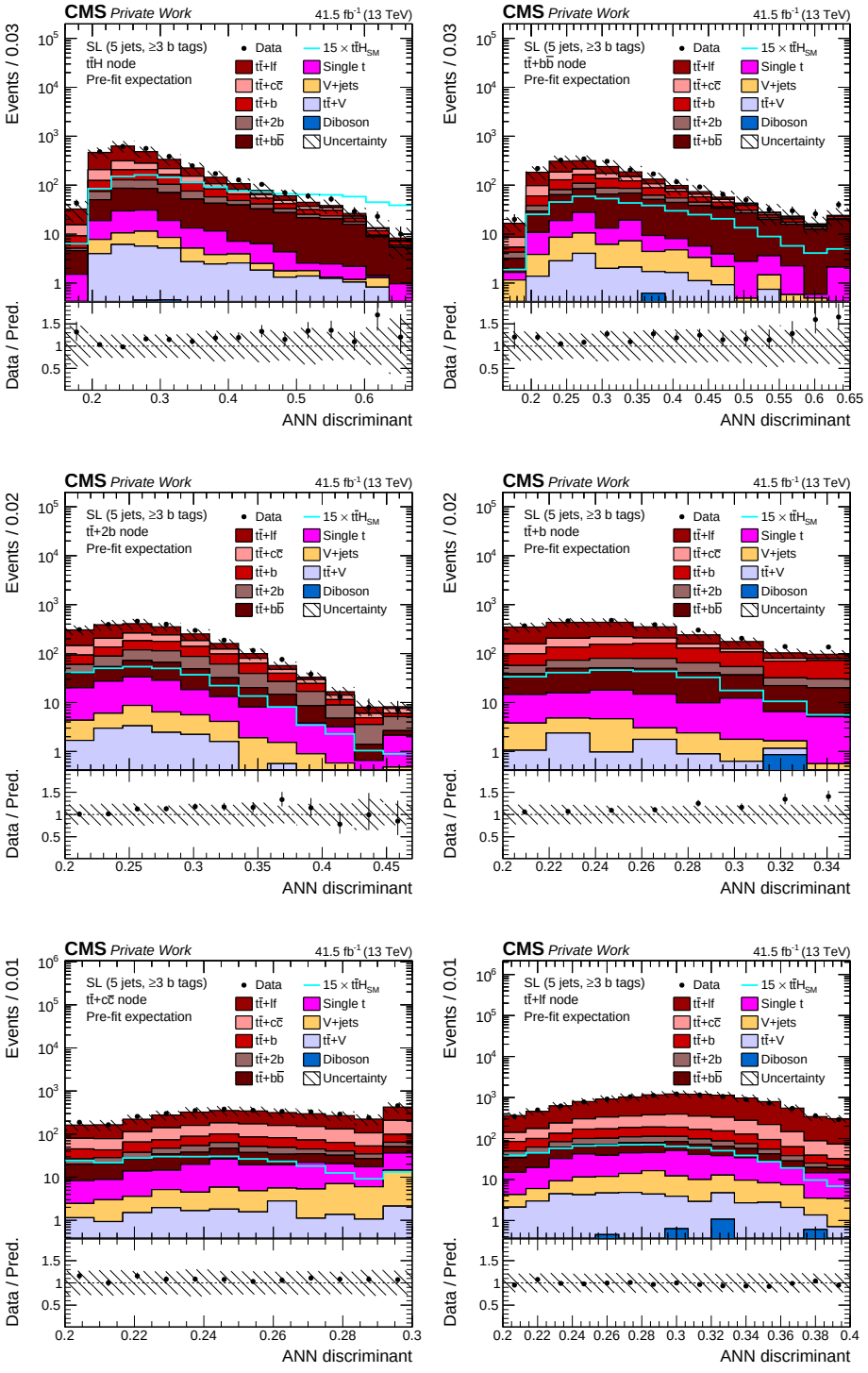


Figure C.37: ANN discriminant distributions for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2017 before the fit to data. A description of the figures is given at the beginning of Sec. C.

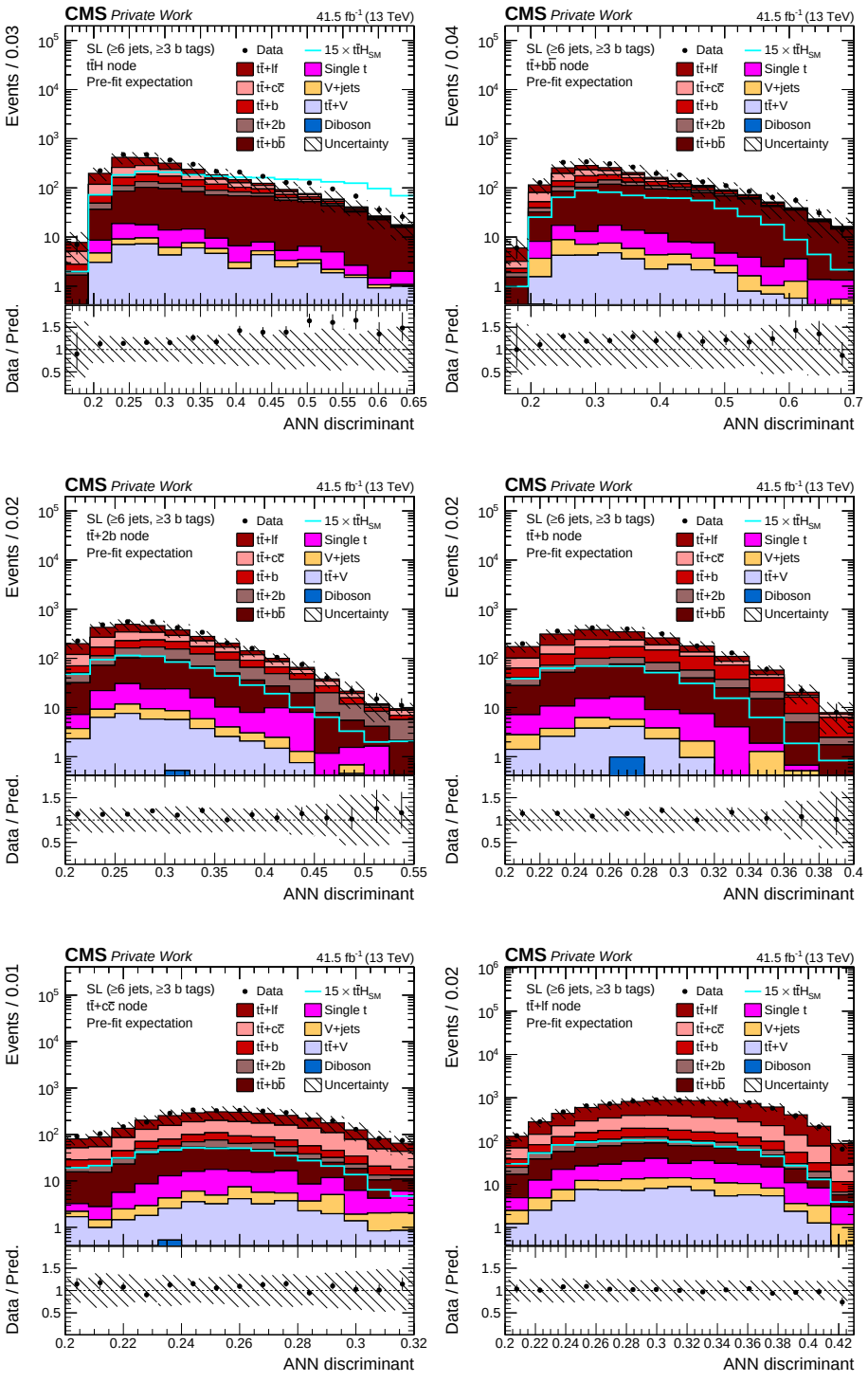


Figure C.38: ANN discriminant distributions for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2017 before the fit to data. A description of the figures is given at the beginning of Sec. C.

C.3 Distributions for the analysis of 2016 data after the fit to 2016 data

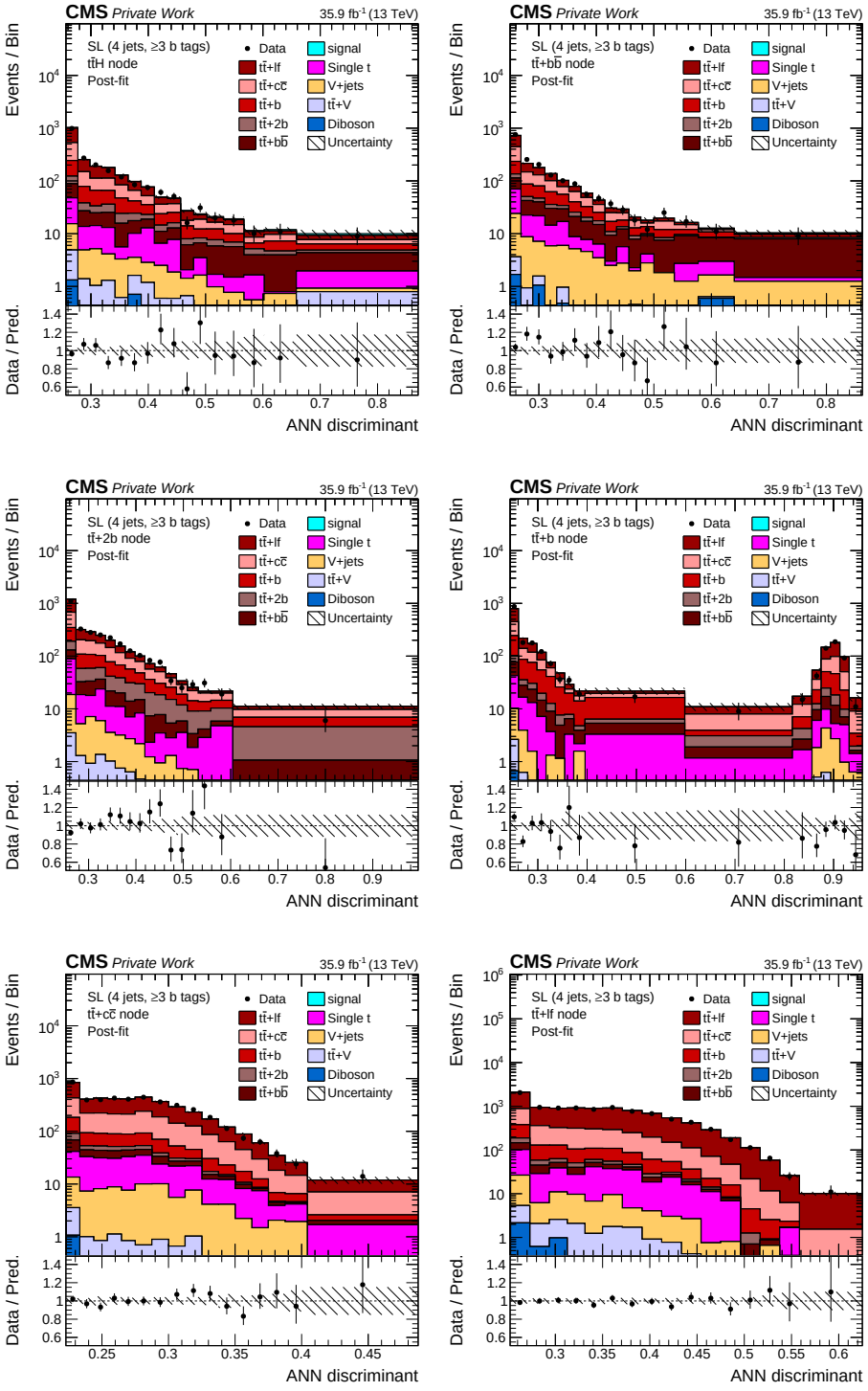


Figure C.39: ANN discriminant distributions for ($4\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2016 after the fit to data recorded in the year 2016. A description of the figures is given at the beginning of Sec. C.

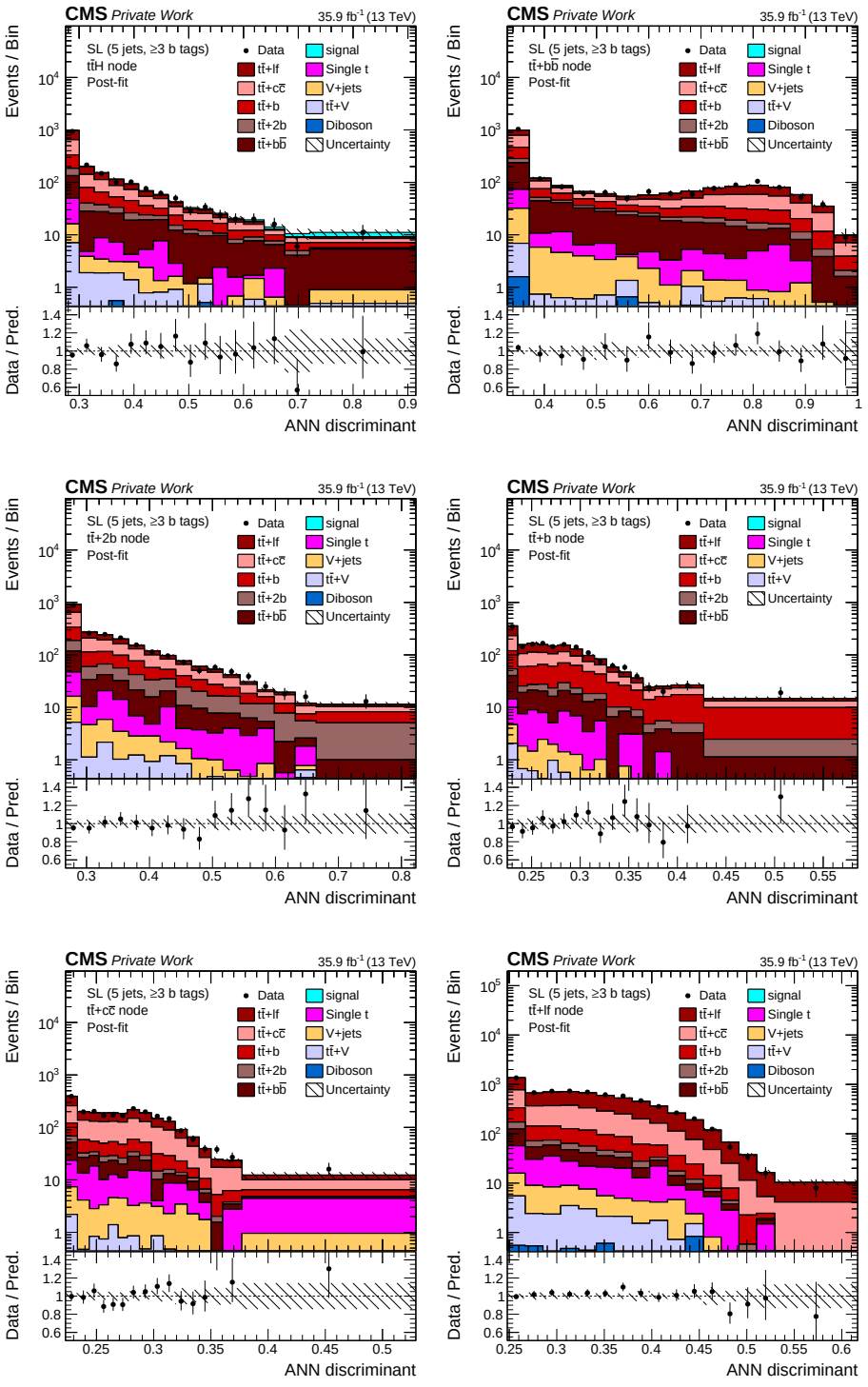


Figure C.40: ANN discriminant distributions for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2016 after the fit to data recorded in the year 2016. A description of the figures is given at the beginning of Sec. C.

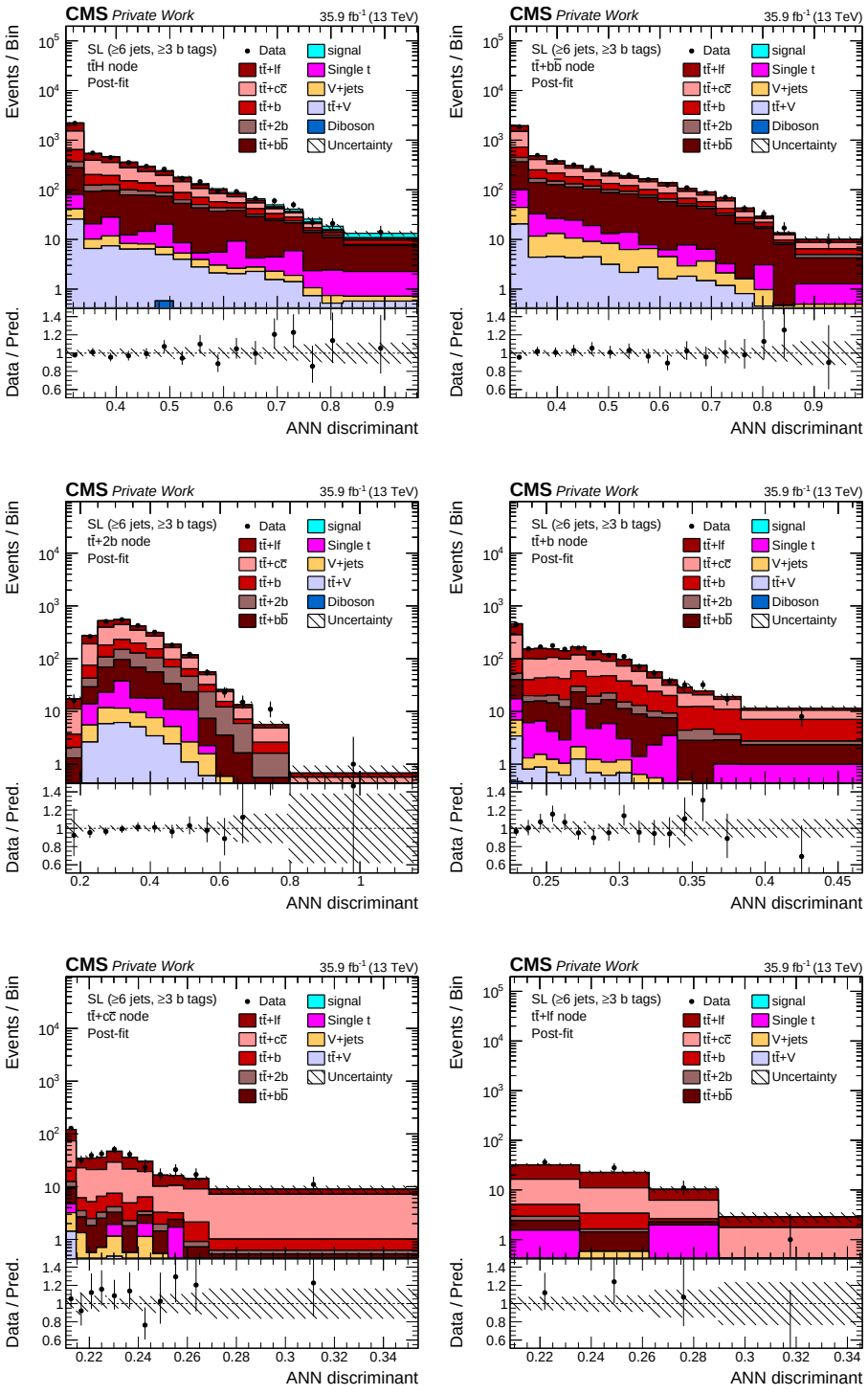


Figure C.41: ANN discriminant distributions for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016 after the fit to data recorded in the year 2016. A description of the figures is given at the beginning of Sec. C.

C.4 Distributions for the analysis of 2017 data after the fit to 2017 data

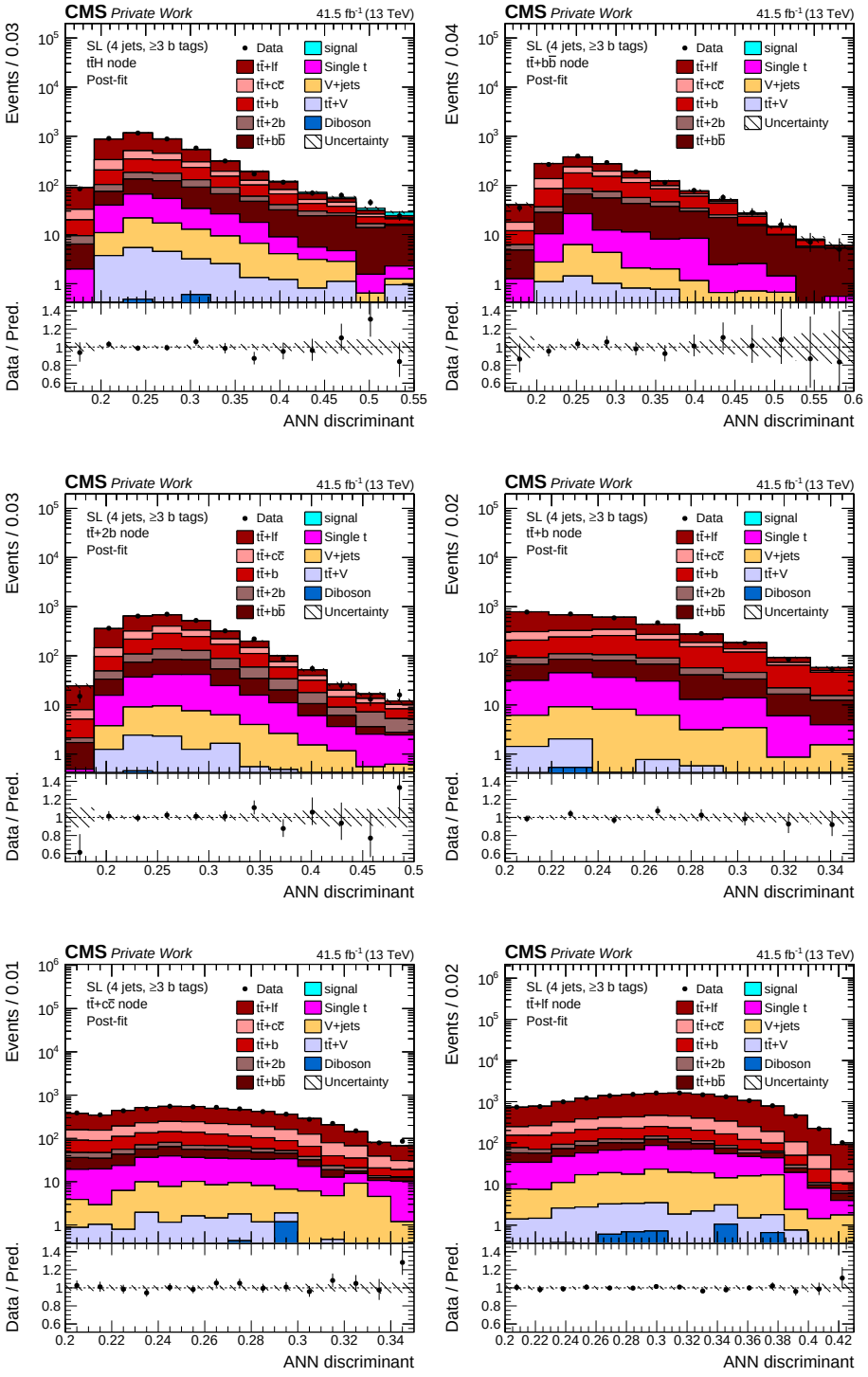


Figure C.42: ANN discriminant distributions for ($4\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2017 after the fit to data recorded in the year 2017. A description of the figures is given at the beginning of Sec. C.

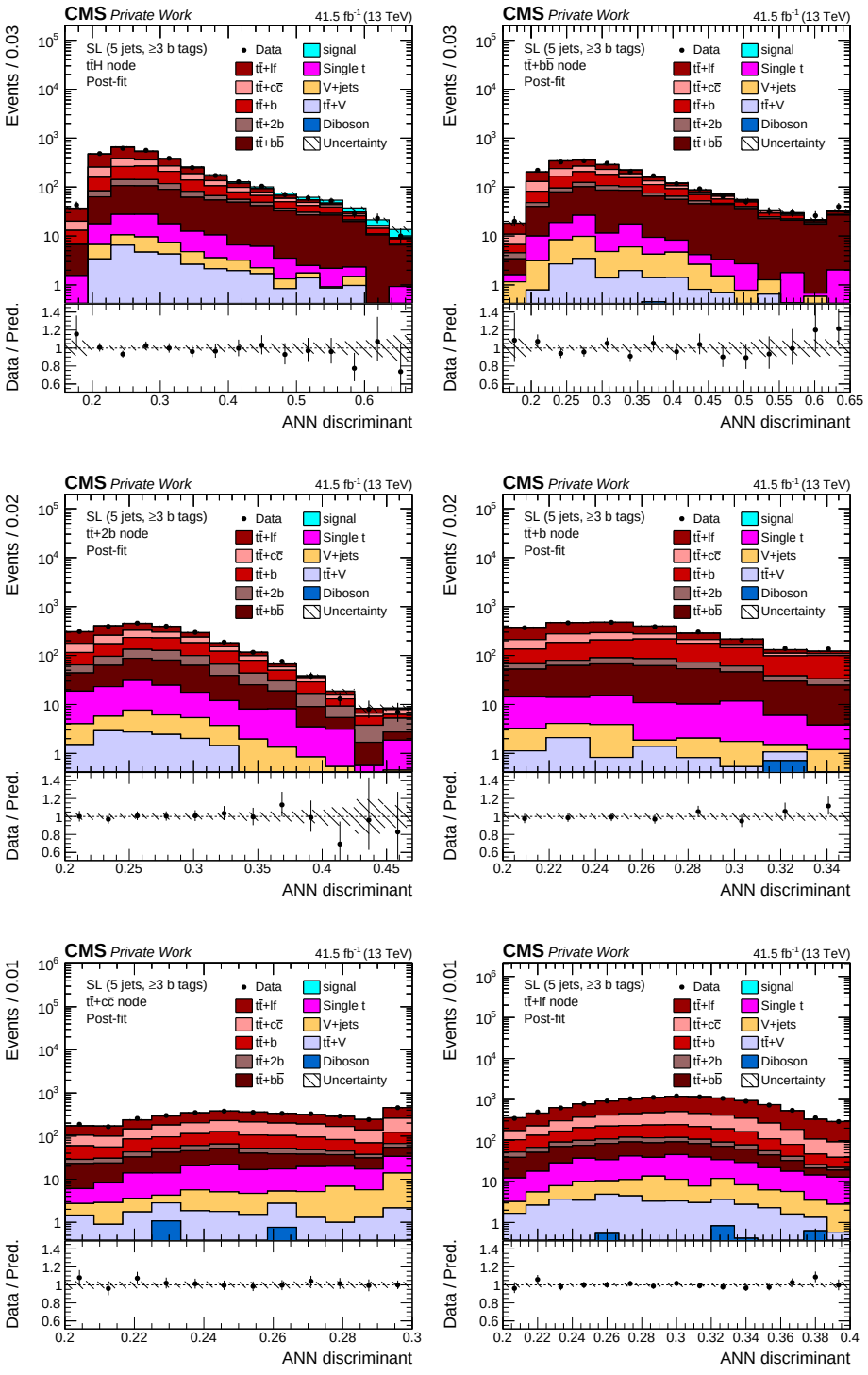


Figure C.43: ANN discriminant distributions for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2017 after the fit to data recorded in the year 2017. A description of the figures is given at the beginning of Sec. C.

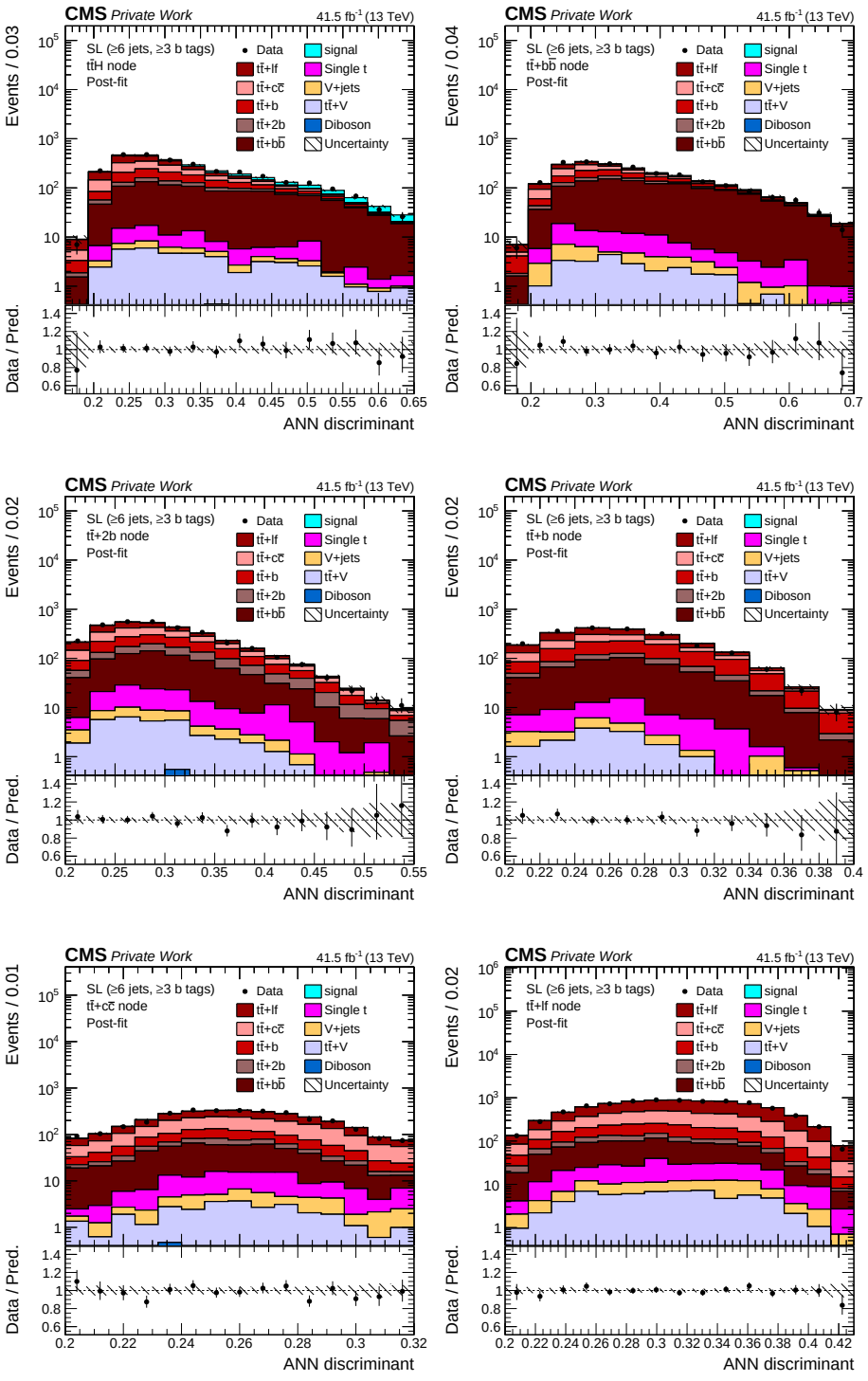


Figure C.44: ANN discriminant distributions for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2017 after the fit to data recorded in the year 2017. A description of the figures is given at the beginning of Sec. C.

C.5 ANN Distributions after the joint fit to 2016 and 2017 data

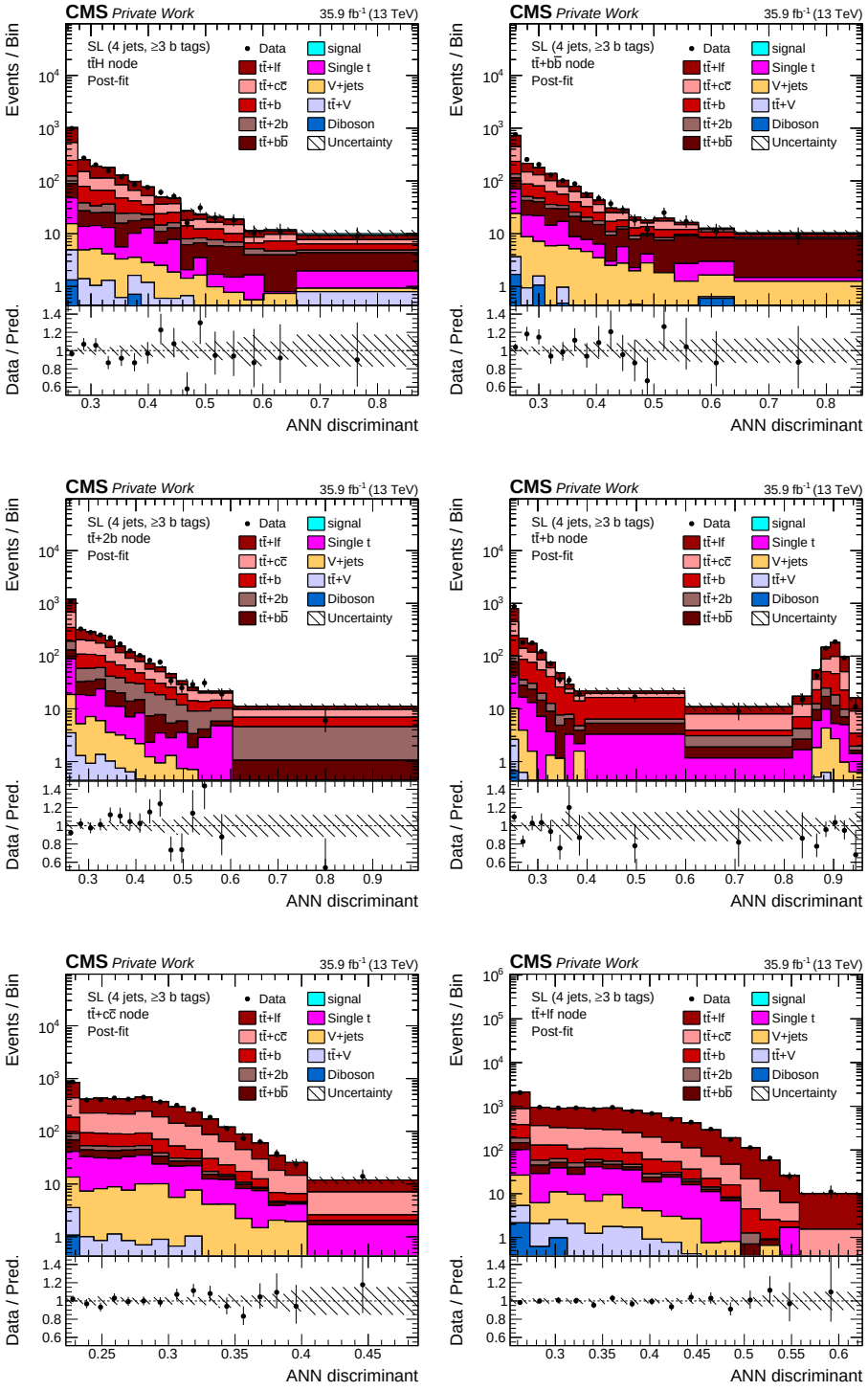


Figure C.45: ANN discriminant distributions for (4 jets, ≥ 3 b-tags) events recorded in the year 2016 after the joint fit to data recorded in the years 2016 and 2017. A description of the figures is given at the beginning of Sec. C.

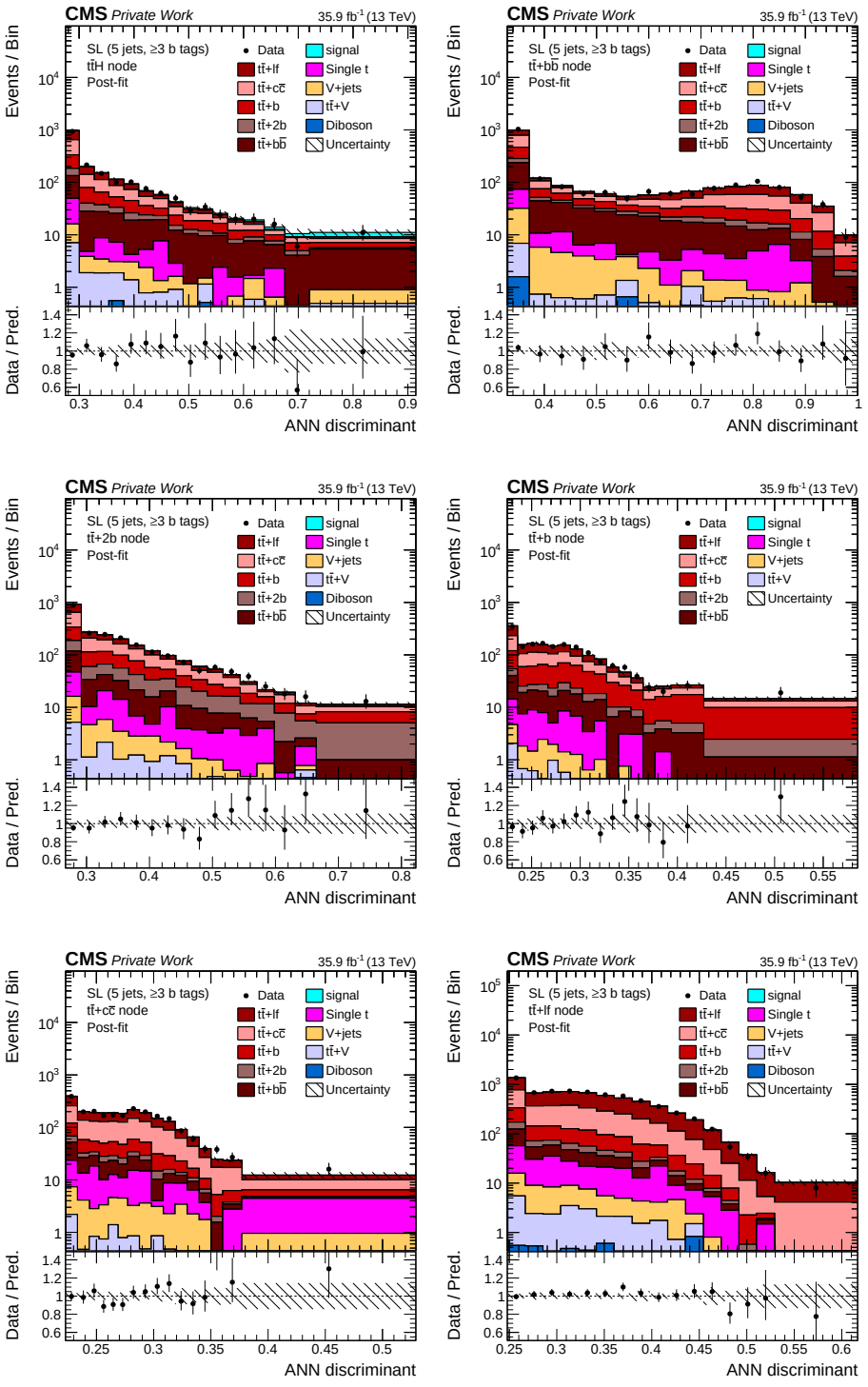


Figure C.46: ANN discriminant distributions for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2016 after the joint fit to data recorded in the years 2016 and 2017. A description of the figures is given at the beginning of Sec. C.

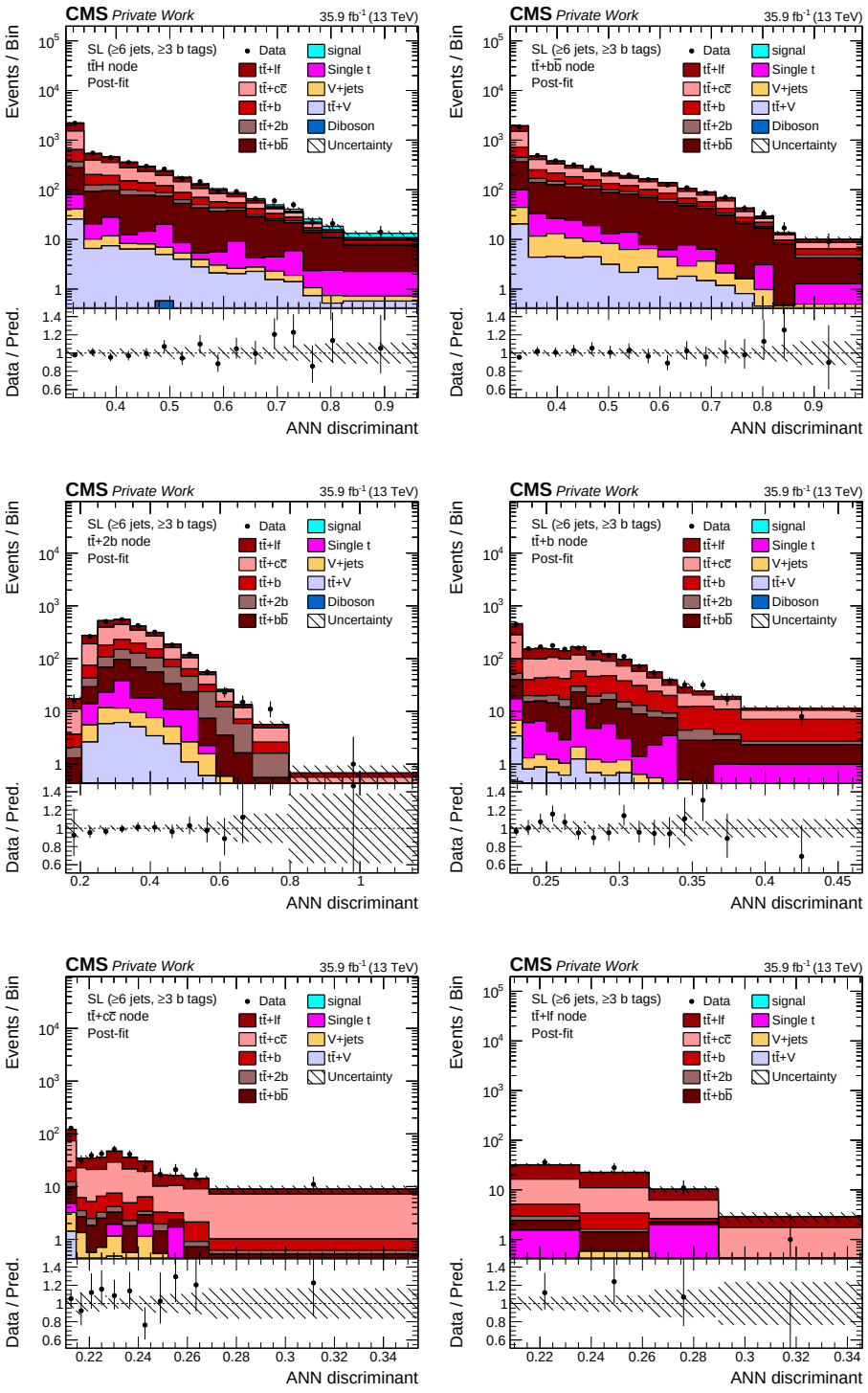


Figure C.47: ANN discriminant distributions for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2016 after the joint fit to data recorded in the years 2016 and 2017. A description of the figures is given at the beginning of Sec. C.

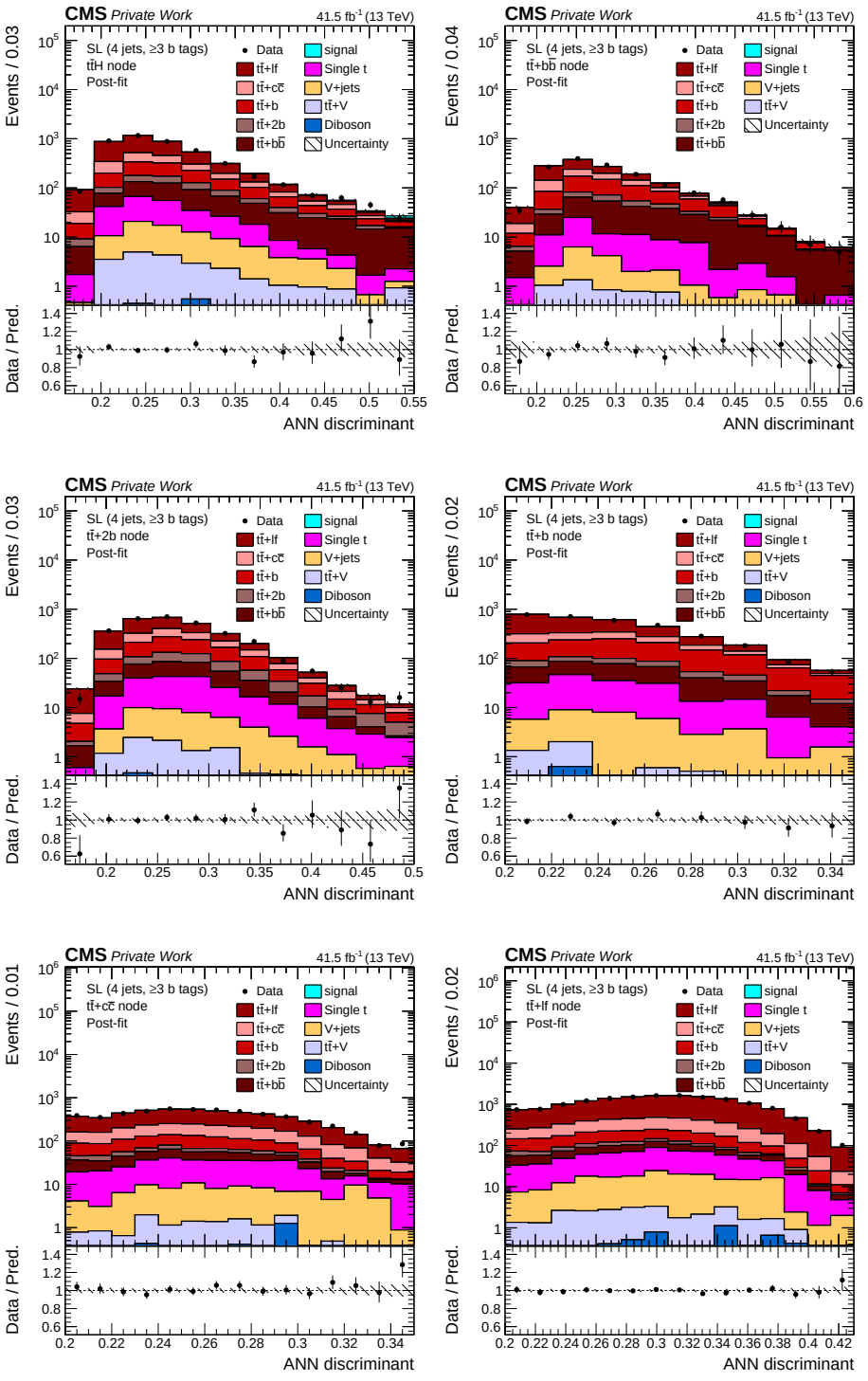


Figure C.48: ANN discriminant distributions for ($4\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2017 after the joint fit to data recorded in the years 2016 and 2017. A description of the figures is given at the beginning of Sec. C.

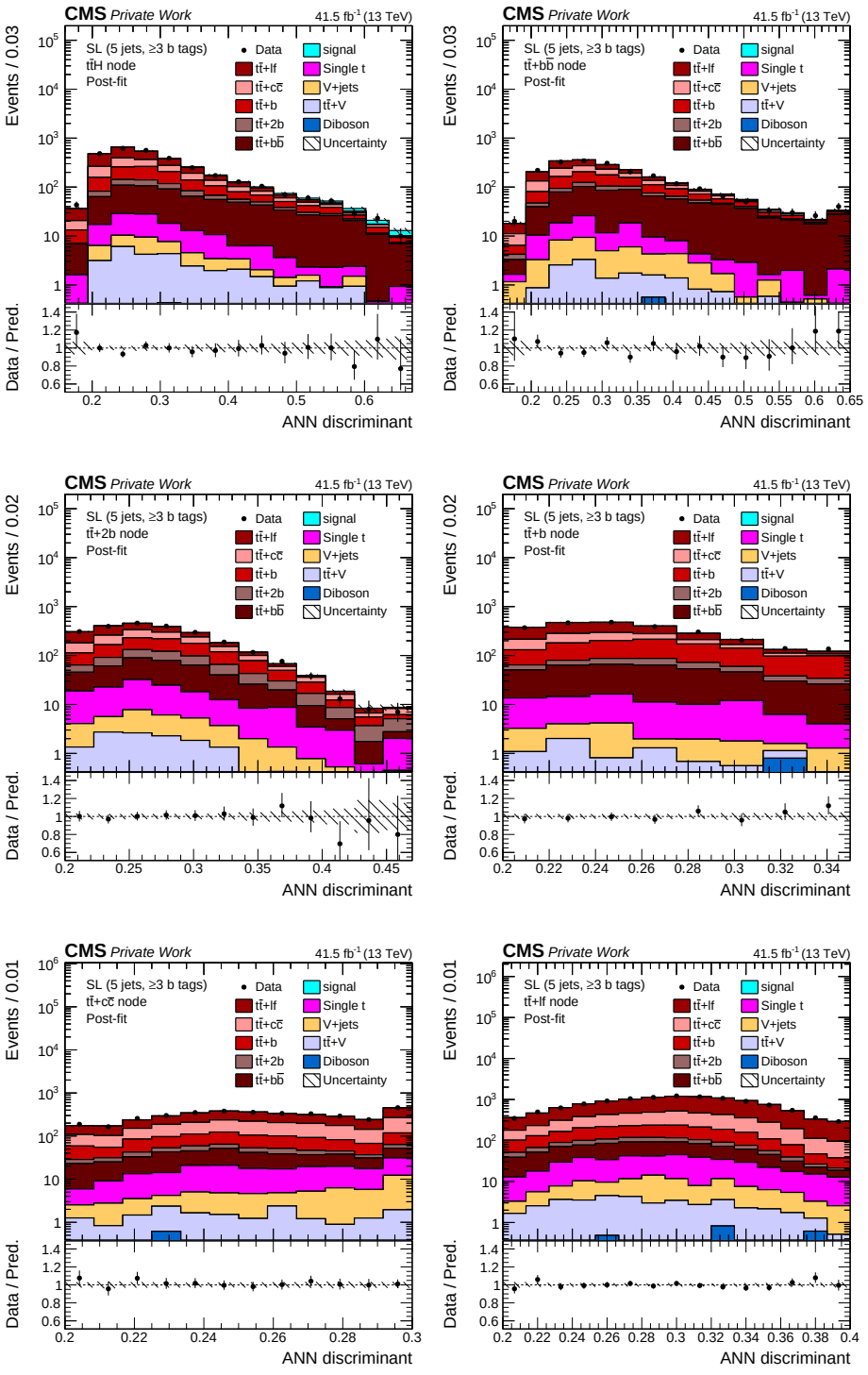


Figure C.49: ANN discriminant distributions for ($5\,\text{jets}, \geq 3\,\text{b-tags}$) events recorded in the year 2017 after the joint fit to data recorded in the years 2016 and 2017. A description of the figures is given at the beginning of Sec. C.

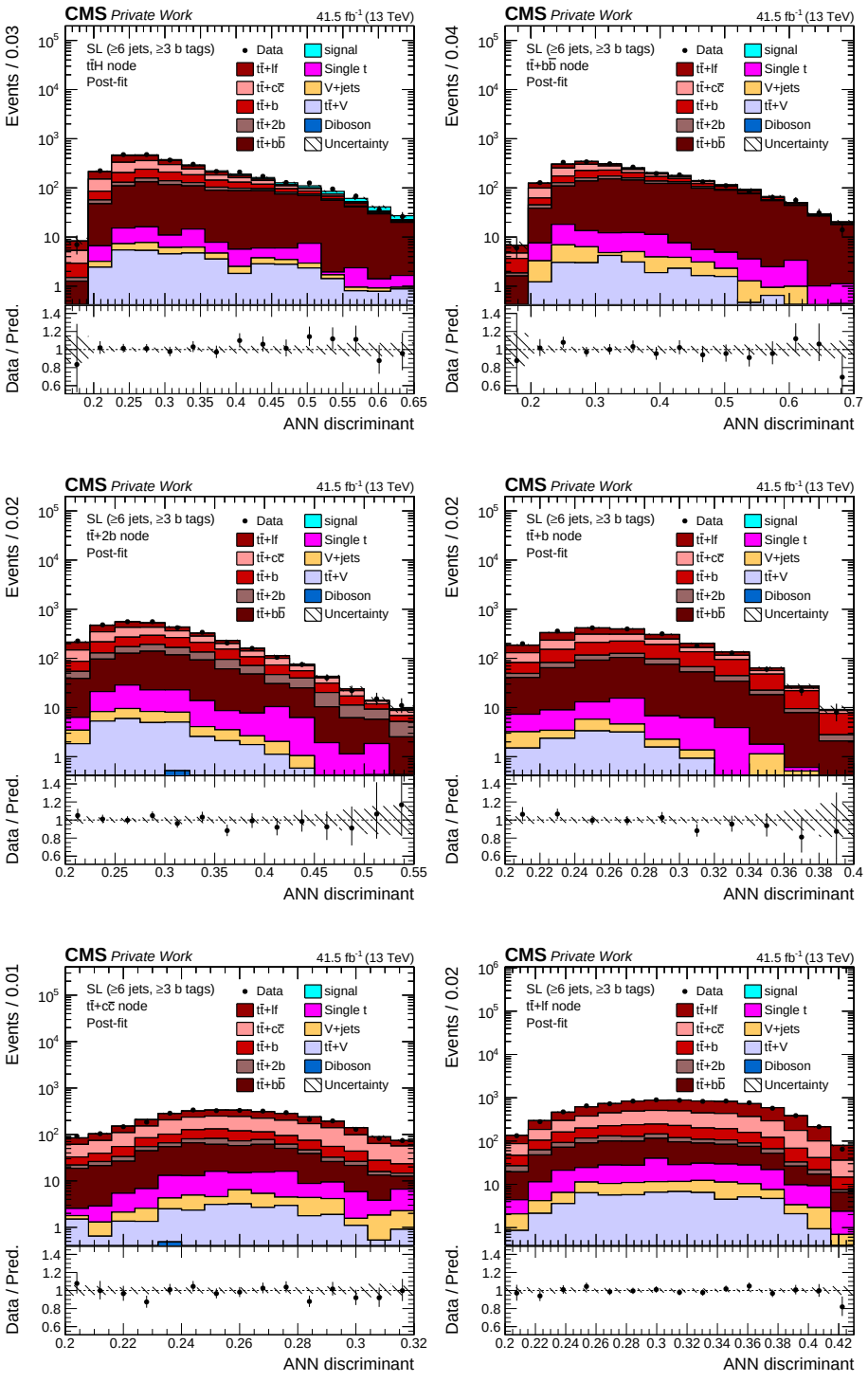


Figure C.50: ANN discriminant distributions for (≥ 6 jets, ≥ 3 b-tags) events recorded in the year 2017 after the joint fit to data recorded in the years 2016 and 2017. A description of the figures is given at the beginning of Sec. C.

D Additional fit results

In this appendix further results of the profile likelihood fits, discussed in Ch. 9, are presented for fits to data recorded in the years 2016, 2017 and jointly for both years. In Sec. D.1 the contributions of different classes of uncertainties to the total uncertainty are shown.
The pulls and constraints of the nuisance parameters are shown in Sec. D.2. Test statistic distributions for two-dimensional fits with $t\bar{t}H$ and a $t\bar{t}$ + heavy flavor process as separate unconstrained signal-strength modifiers are shown in Sec. D.3. Correlation coefficients between parameters of the fits are shown in Sec. D.4 and impact parameters are shown in Sec. D.5. Event yields in the ANN nodes before and after the joint fit to data recorded in both years are presented in Sec. E.

D.1 Uncertainty breakdowns

In this section, the contribution of different classes of uncertainties are shown for the fit to data recorded in the year 2016 in Tab. D.1 and for the fit to data recorded in the year 2017 in Tab. D.2. The results for the joint fit to data are discussed in Ch. 9.

Table D.1: Contributions of different classes of uncertainties to the total uncertainty for the fit to the data recorded in the year 2016. The expected value is calculated for a standard model signal+background prediction. The quoted uncertainties $\Delta\mu$ are obtained by fixing the uncertainties in the class to their post-fit values, repeating the fit, and quadratically subtracting the obtained uncertainty from the uncertainty of the full fit.

Uncertainty source	$\pm\Delta\mu$ (observed)	$\pm\Delta\mu$ (expected)
Total experimental	+0.19/-0.17	+0.23/-0.21
b tagging	+0.16/-0.14	+0.16/-0.16
jet energy scale	+0.09/-0.06	+0.14/-0.12
Total theory	+0.33/-0.31	+0.37/-0.34
signal modeling	+0.14/-0.04	+0.15/-0.02
parton shower	+0.14/-0.06	+0.14/-0.12
$t\bar{t}$ + heavy flavor model	+0.24/-0.24	+0.23/-0.25
Size of simulated samples	+0.19/-0.18	+0.19/-0.19
Total systematic	+0.43/-0.4	+0.5/-0.46
Statistical	+0.27/-0.27	+0.3/-0.29
Total	+0.52/-0.49	+0.59/-0.56

Table D.2: Contributions of different classes of uncertainties to the total uncertainty for the fit to the data recorded in the year 2017. The expected value is calculated for a standard model signal+background prediction. The quoted uncertainties $\Delta\mu$ are obtained by fixing the uncertainties in the class to their post-fit values, repeating the fit, and quadratically subtracting the obtained uncertainty from the uncertainty of the full fit.

Uncertainty source	$\pm\Delta\mu$ (observed)	$\pm\Delta\mu$ (expected)
Total experimental	+0.24/-0.22	+0.13/-0.12
b tagging	+0.05/-0.09	+0.06/-0.04
jet energy scale	+0.21/-0.19	+0.1/-0.1
Total theory	+0.41/-0.36	+0.28/-0.26
signal modeling	+0.24/-0.11	+0.14/-0.05
parton shower	+0.17/-0.21	+0.14/-0.13
$\mathrm{t\bar{t}}$ + heavy flavor model	+0.3/-0.31	+0.22/-0.23
Size of simulated samples	+0.18/-0.19	+0.13/-0.13
Total systematic	+0.52/-0.46	+0.37/-0.34
Statistical	+0.26/-0.26	+0.22/-0.22
Total	+0.62/-0.56	+0.45/-0.43

D.2 Nuisance parameter pulls and constraints

The pulls and constraints of the nuisance parameters in the fits to data recorded in either the years 2016 or the year 2017 are shown in Sec. D.2 and Sec. D.2, respectively. Pulls and constraints in the joint fit to data recorded in both years are shown in Sec. D.2 For better visibility the nuisance parameters are grouped into different classes and the nuisance parameters related to the statistical uncertainty of the simulated samples are omitted.

Nuisance parameter pulls and constraints in the fit to data recorded in the year 2016

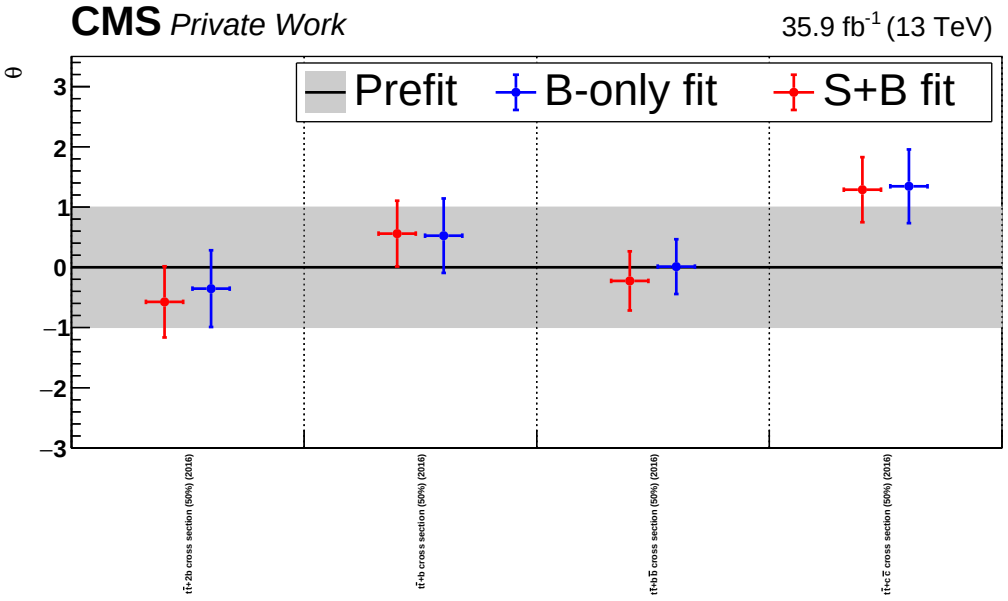


Figure D.51: Pulls and constraints of the nuisance parameters related to $t\bar{t}$ + heavy flavor production in the fit to data recorded in the year 2016.

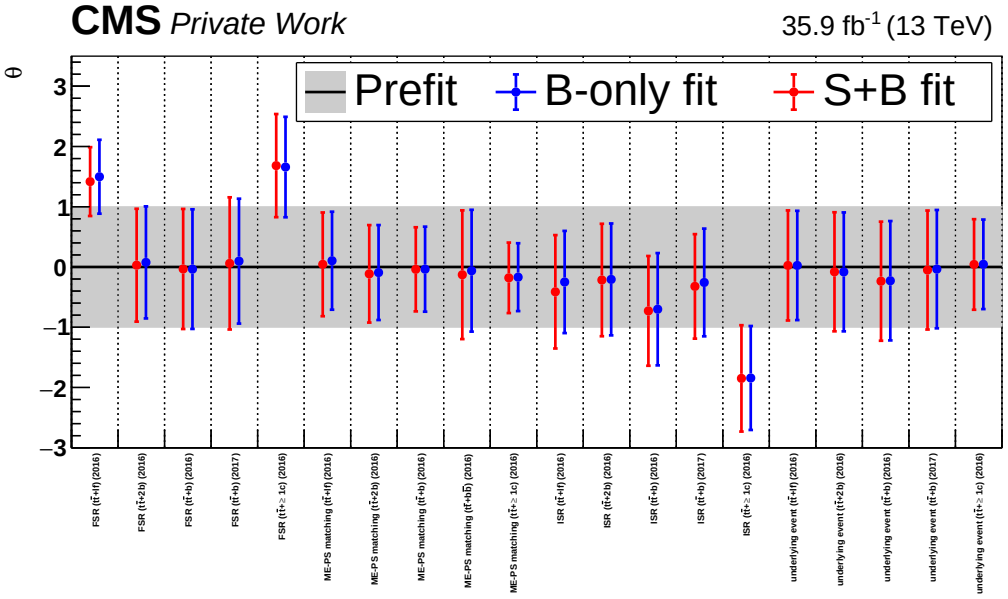


Figure D.52: Pulls and constraints of the nuisance parameters related to the parton shower simulation in the fit to data recorded in the year 2016.

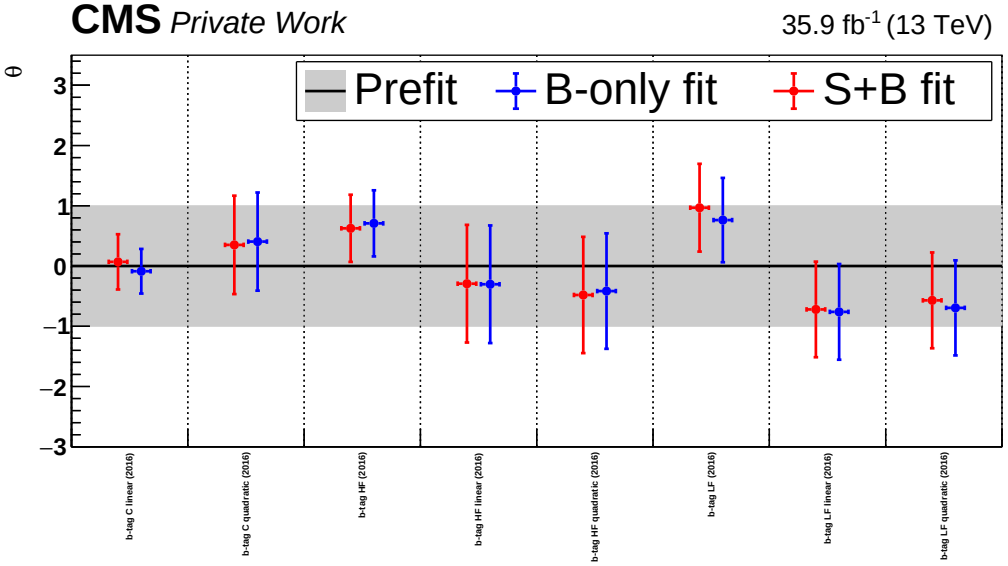


Figure D.53: Pulls and constraints of the nuisance parameters related to b-tagging in the fit to data recorded in the year 2016.

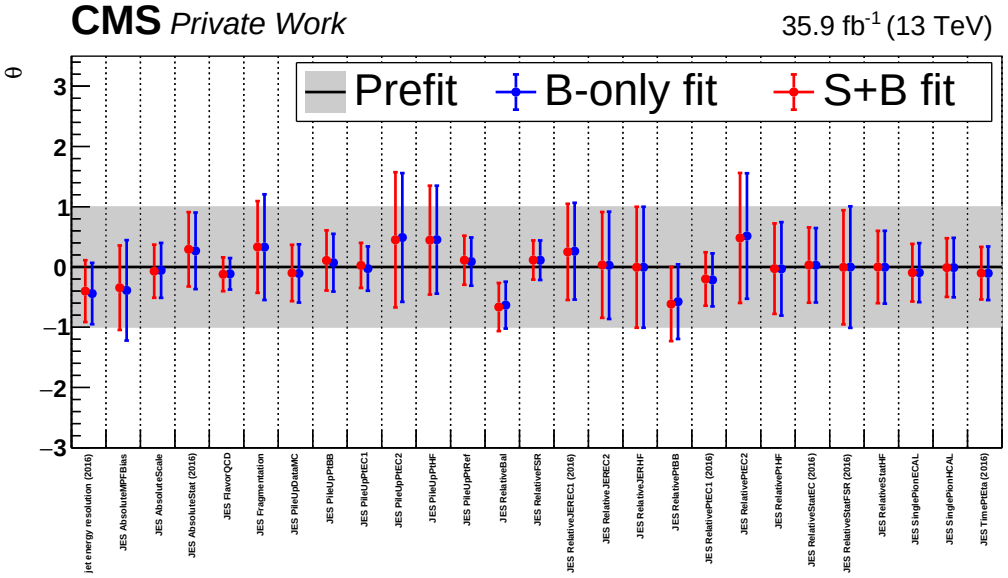


Figure D.54: Pulls and constraints of the nuisance parameters related to the jet energy calibration (JEC) in the fit to data recorded in the year 2016.

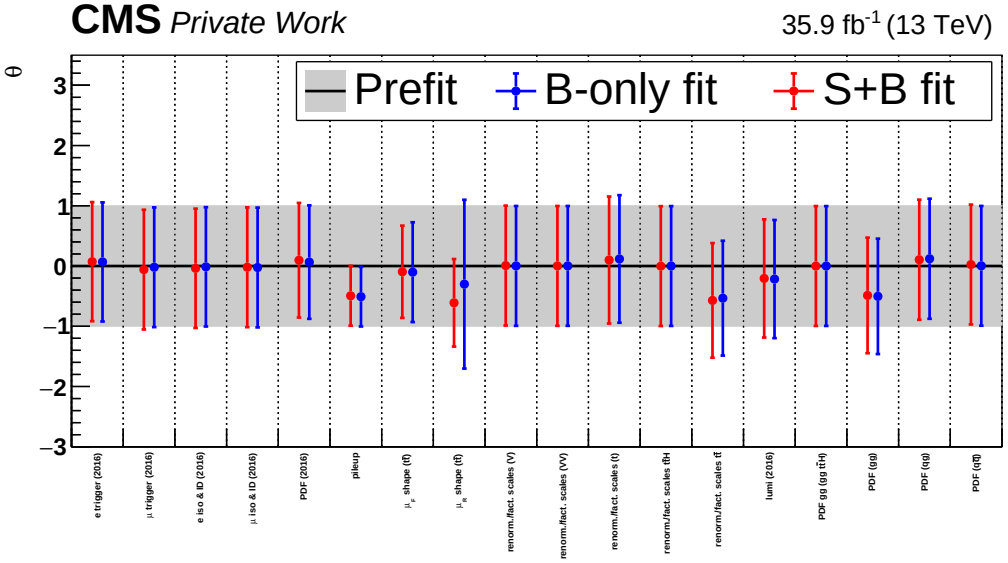


Figure D.55: Pulls and constraints of the remaining nuisance parameters in the fit to data recorded in the year 2016.

Nuisance parameter pulls and constraints in the fit to data recorded in the year 2017

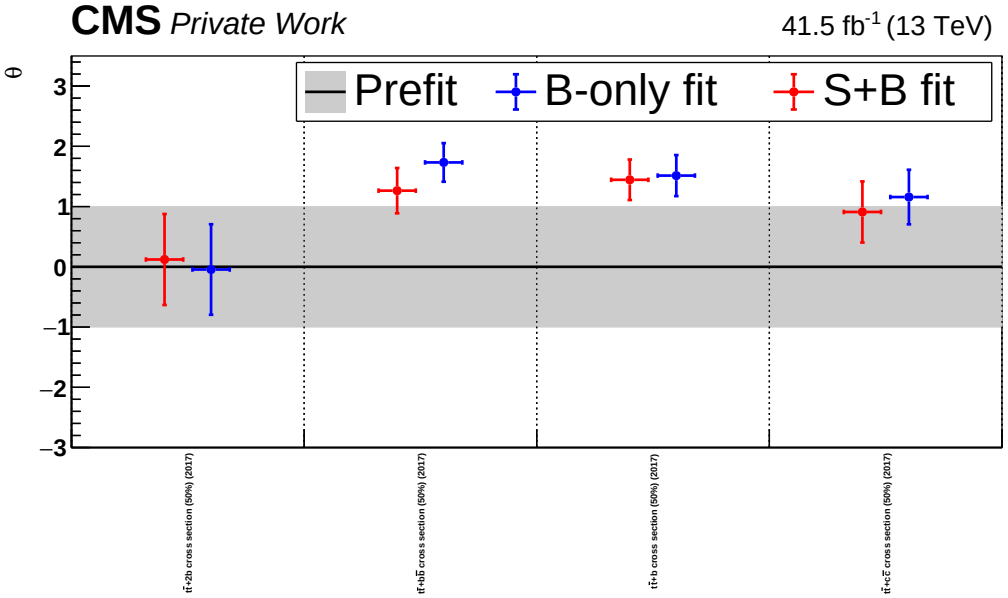


Figure D.56: Pulls and constraints of the nuisance parameters related to $\mathrm{t\bar{t}}$ + heavy flavor production in the fit to data recorded in the year 2017.

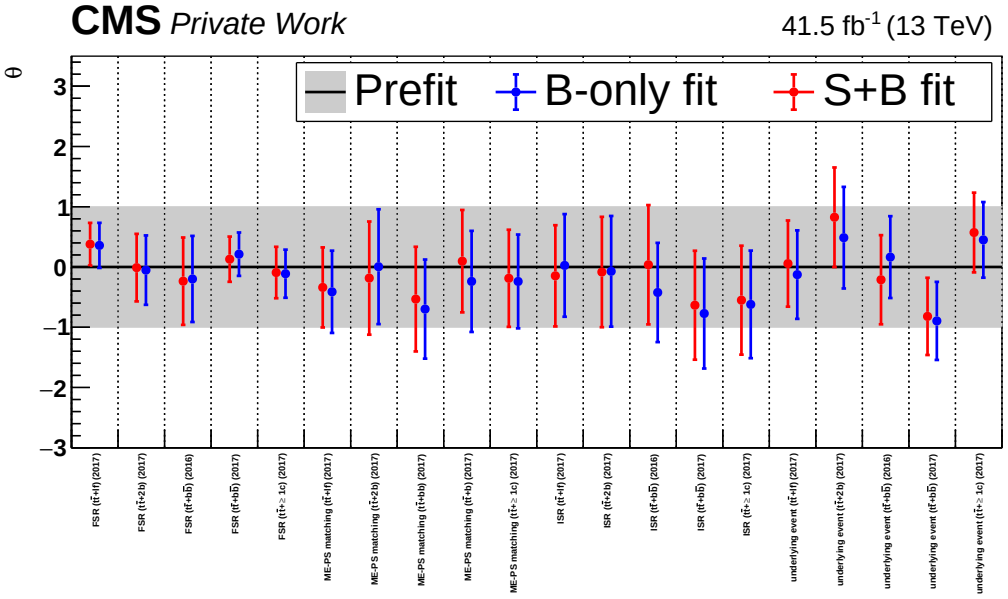


Figure D.57: Pulls and constraints of the nuisance parameters related to the parton shower simulation in the fit to data recorded in the year 2017.

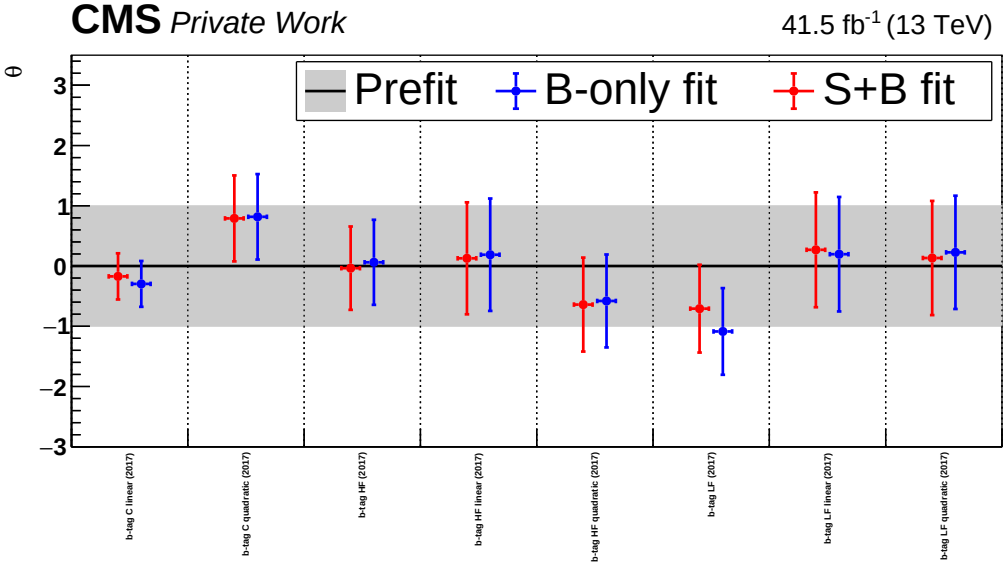


Figure D.58: Pulls and constraints of the nuisance parameters related to b-tagging in the fit to data recorded in the year 2017.

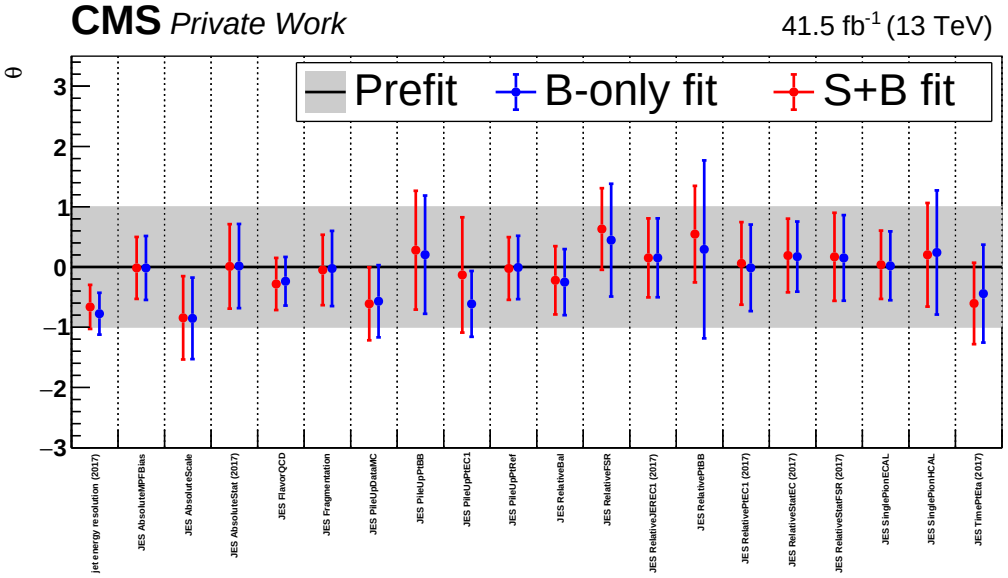


Figure D.59: Pulls and constraints of the nuisance parameters related to the jet energy calibration (JEC) in the fit to data recorded in the year 2017.

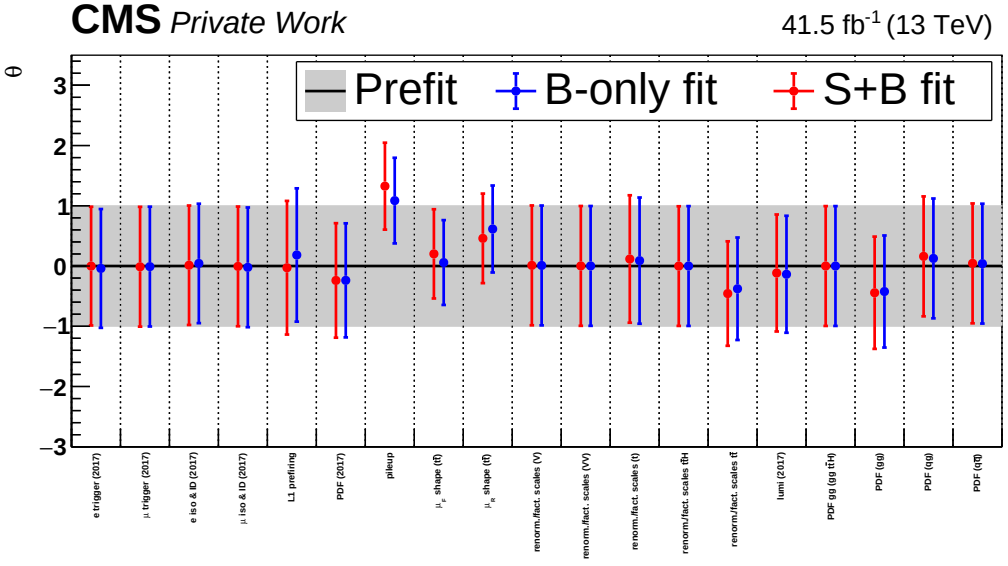


Figure D.60: Pulls and constraints of the remaining nuisance parameters in the fit to data recorded in the year 2017.

Nuisance parameters pulls and constraints in the joint fit to data recorded in the years 2016 and 2017

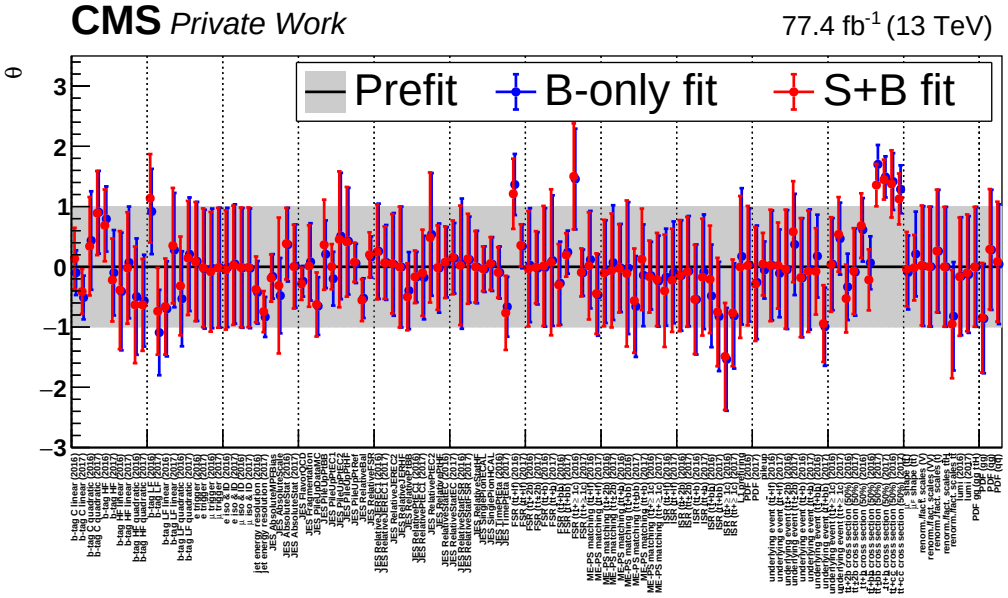


Figure D.61: Pulls and constraints of the nuisance parameters in the joint fit to data recorded in the years 2016 and 2017.

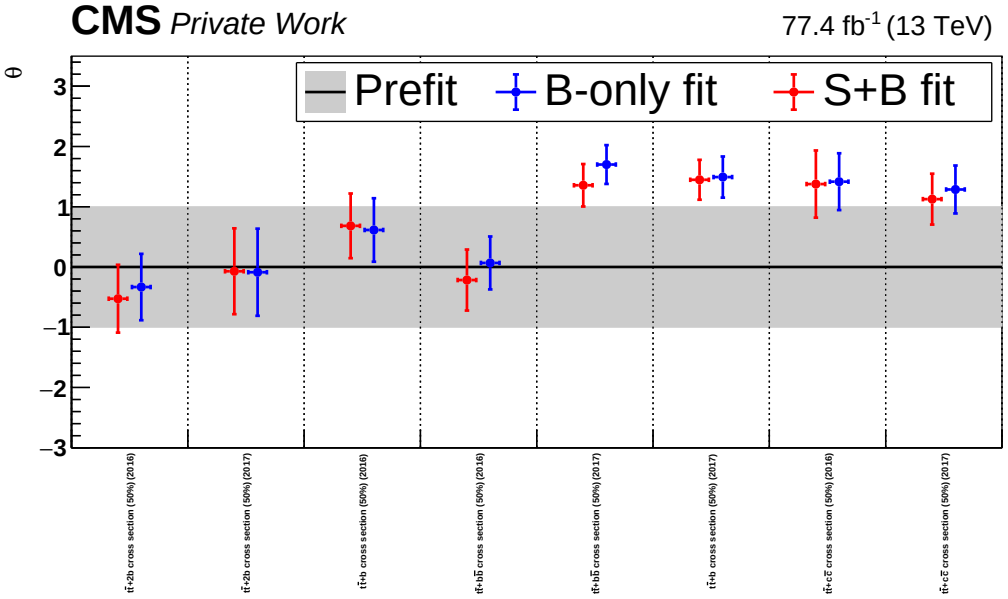


Figure D.62: Pulls and constraints of the nuisance parameters related to $t\bar{t}$ + heavy flavor production in the joint fit to data recorded in the years 2016 and 2017.

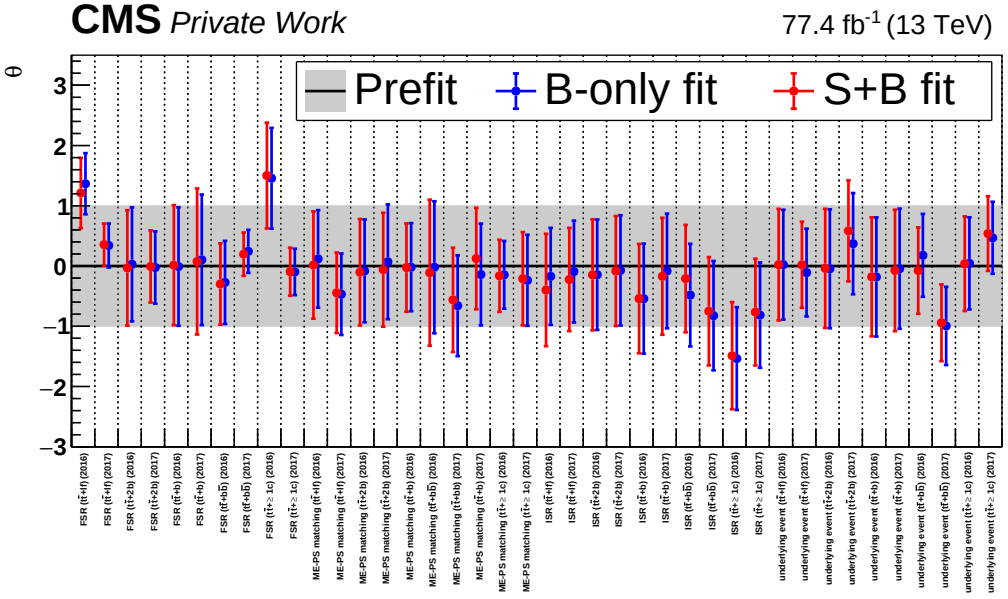


Figure D.63: Pulls and constraints of the nuisance parameters related to the parton shower simulation in the joint fit to data recorded in the years 2016 and 2017.

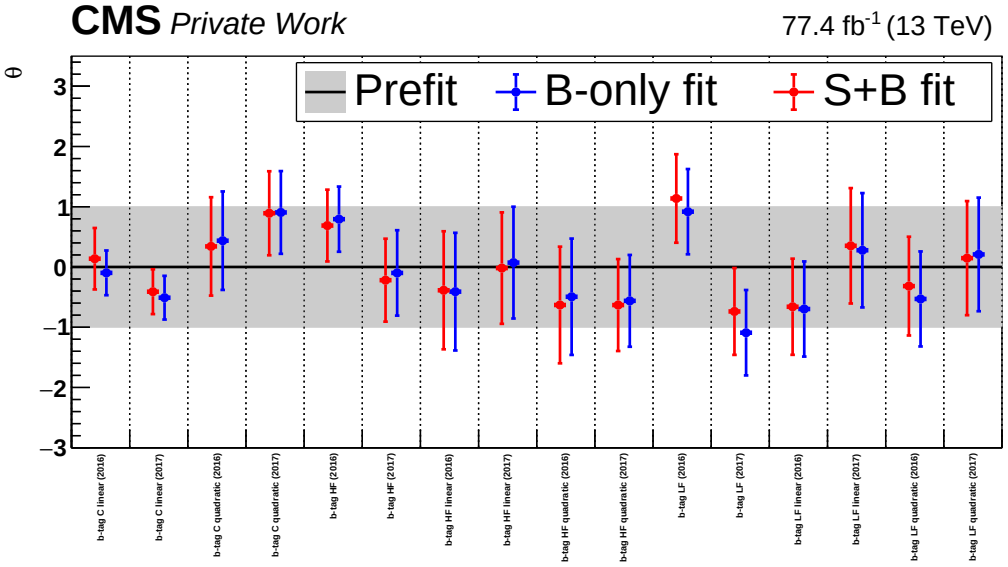


Figure D.64: Pulls and constraints of the nuisance parameters related to b-tagging in the joint fit to data recorded in the years 2016 and 2017.

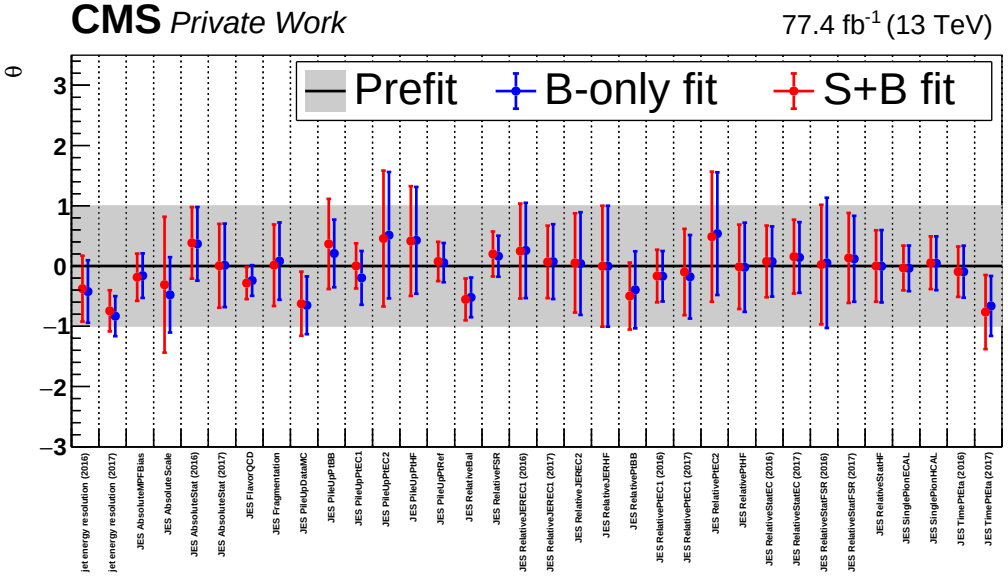


Figure D.65: Pulls and constraints of the nuisance parameters related to the jet energy calibration (JEC) in the joint fit to data recorded in the years 2016 and 2017.

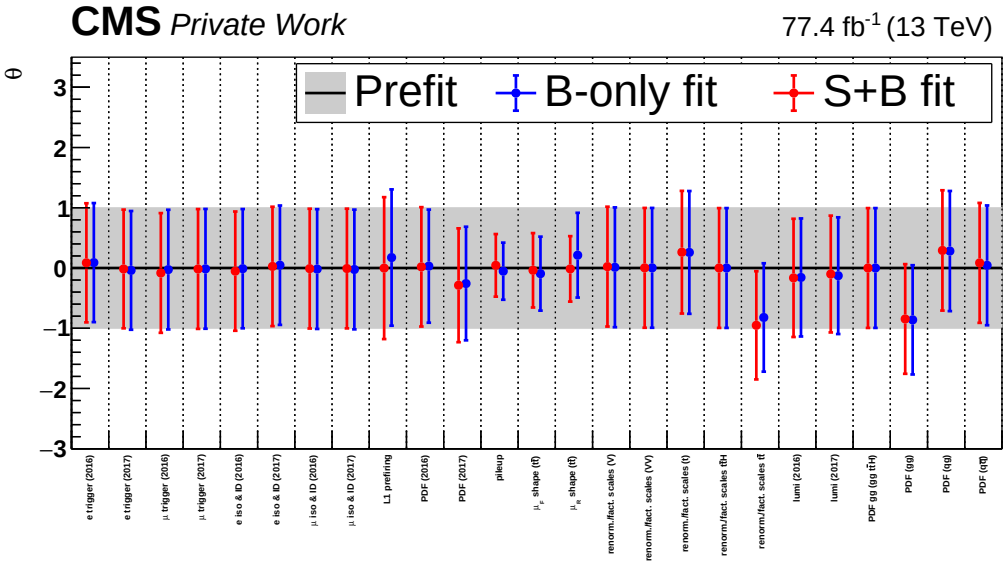


Figure D.66: Pulls and constraints of the remaining nuisance parameters in the joint fit to data recorded in the years 2016 and 2017.

D.3 2D scans

In the following, test statistic distributions are shown for fits with two signal-strength modifiers, one for $t\bar{t}H$ production and one for either $t\bar{t}+b\bar{b}$ production, $t\bar{t}+b$ production, $t\bar{t}+2b$ production or all three $t\bar{t}+\geq 1b$ processes together. The two-dimensional fit with a signal-strength modifier for $t\bar{t}+b\bar{b}$ production is further discussed in Ch. 9.

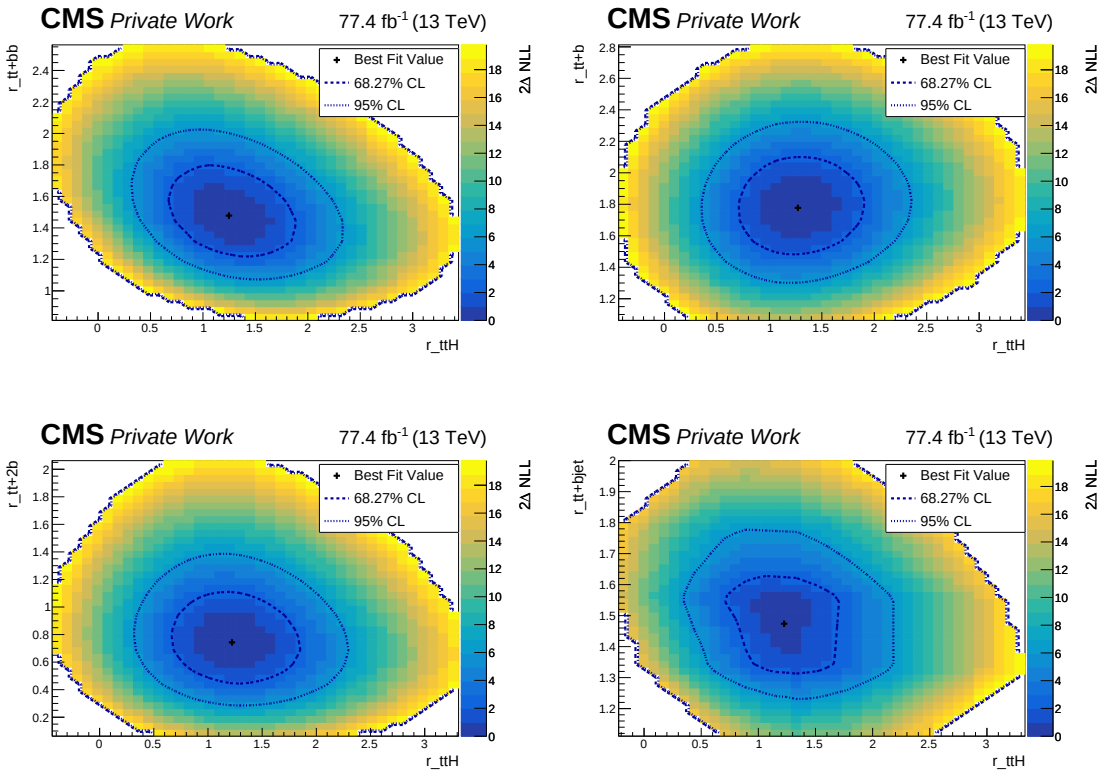


Figure D.67: Profile likelihood test statistic values for a fit with two signal-strength modifiers, one for $t\bar{t}H$ and one for either (a) $t\bar{t}+b\bar{b}$ production, (b) $t\bar{t}+b$ production, (c) $t\bar{t}+2b$ production or (d) the sum of all three processes. The additional 50% normalization uncertainty for $t\bar{t}+$heavy flavor is removed for the studied process. All other nuisance parameters are profiled.

D.4 Nuisance parameter correlations

Correlations between the signal-strength modifier and several nuisance parameters are shown in the following for the individual fits to data recorded in the years 2016 and 2017. The result for the joint fit is presented in Ch. 9. The shown nuisance parameters are those with the twenty strongest correlations to the signal-strength modifier or the twenty nuisance parameters with the strongest correlations among themselves.

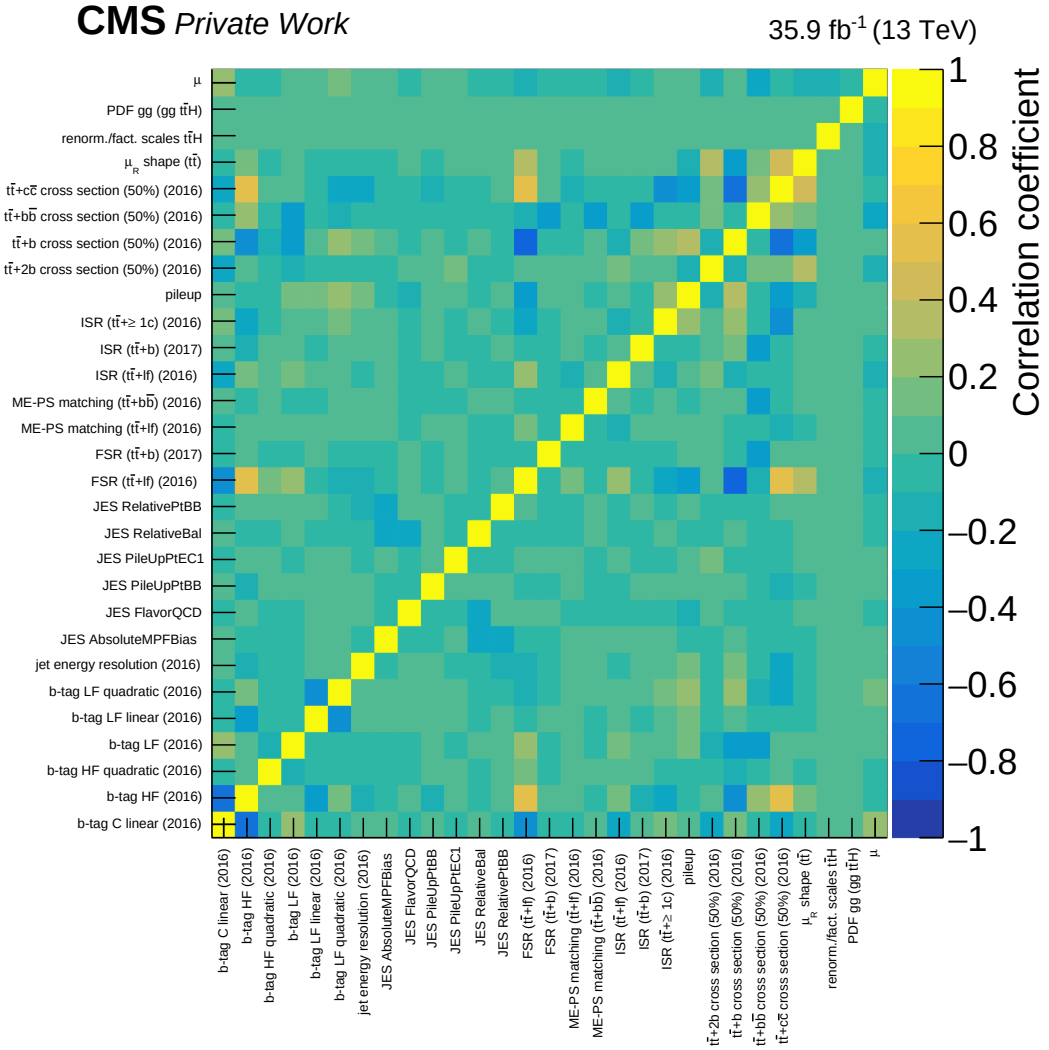


Figure D.68: Correlation coefficients of the signal-strength modifier and several nuisance parameters for the fit to data recorded in 2016.

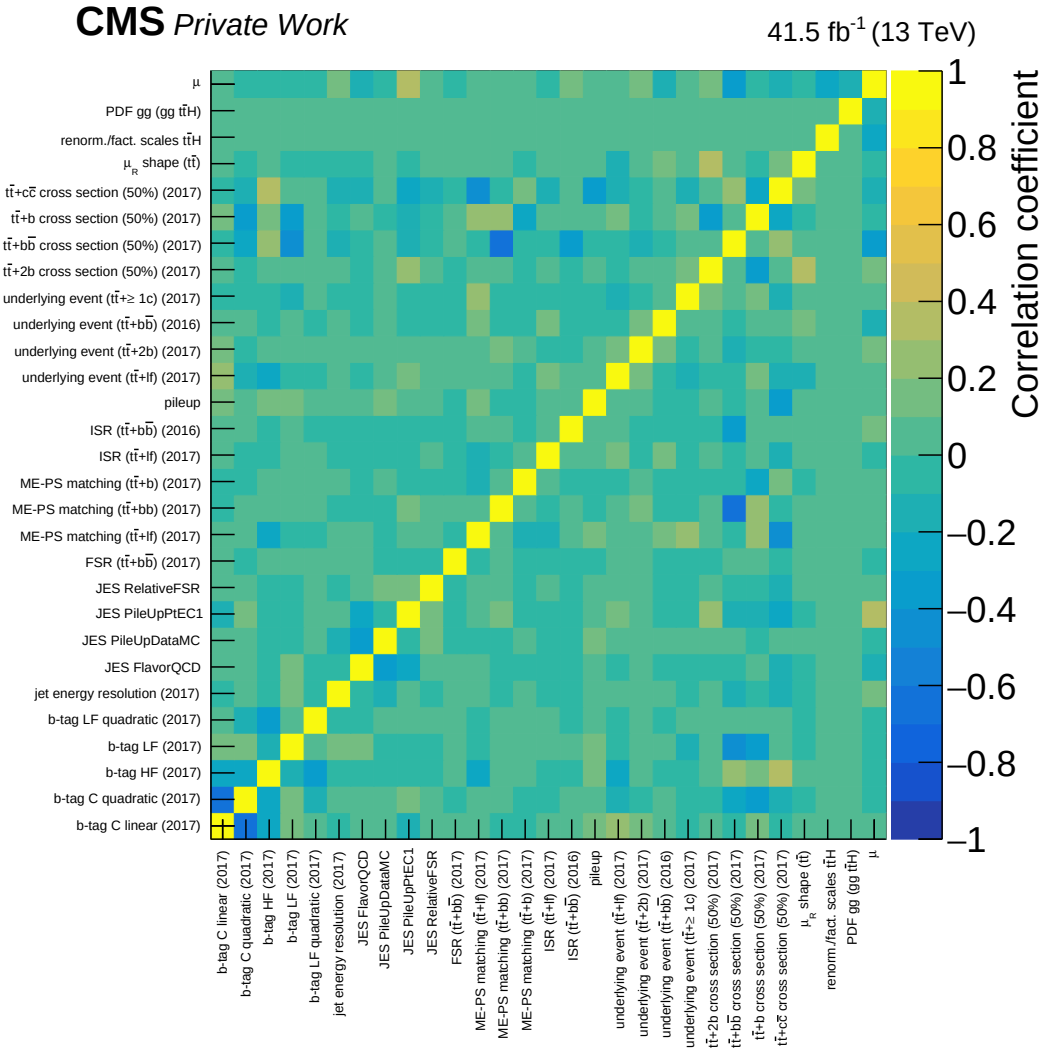


Figure D.69: Correlation coefficients of the signal-strength modifier and several nuisance parameters for the fit to data recorded in 2017.

D.5 Impact parameters

Shown are the impact parameters for the nuisance parameters with highest impact parameters, for the fits to data recorded in 2016 or 2017. The impact parameter is calculated as the difference between the nominal best-fit value $\hat{\mu}$ and the value obtained when freezing the nuisance parameter to its post-fit value plus or minus its post-fit uncertainty.

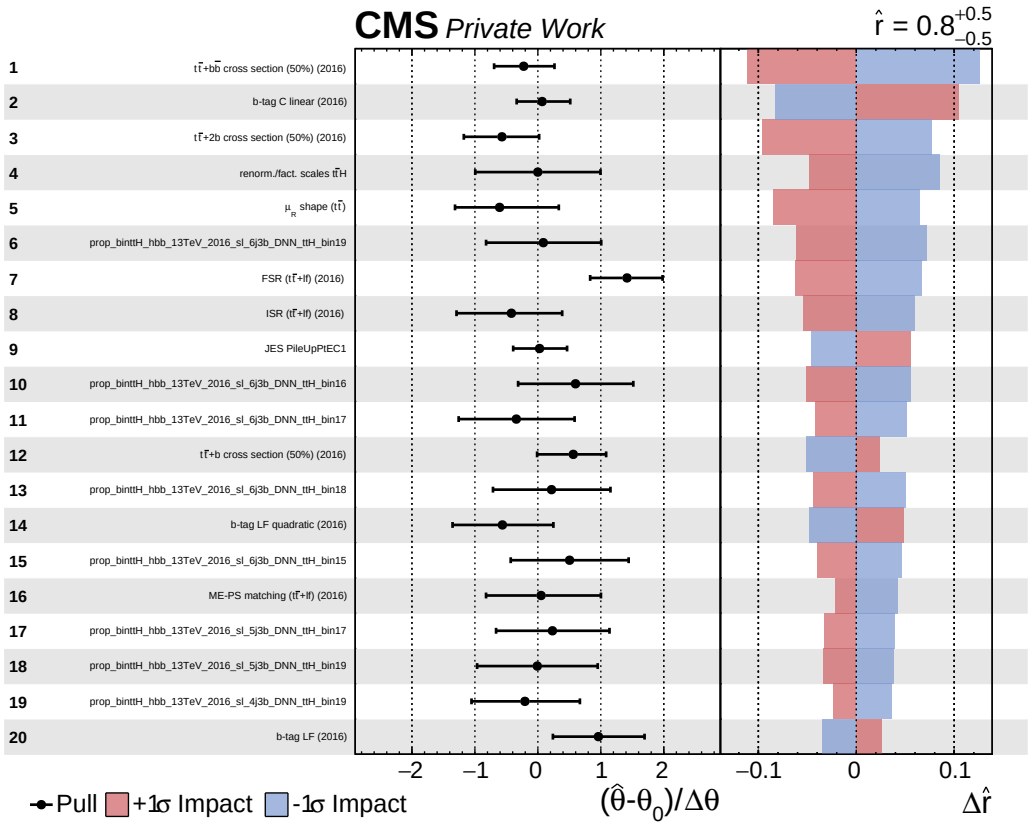


Figure D.70: Impacts of the 20 nuisance parameters with the largest impacts on the signal-strength modifier for the fit to data recorded in the year 2016 . The impacts of the nuisance parameters are shown on the right hand side of the figure while their pulls and constraints are shown in the middle column.

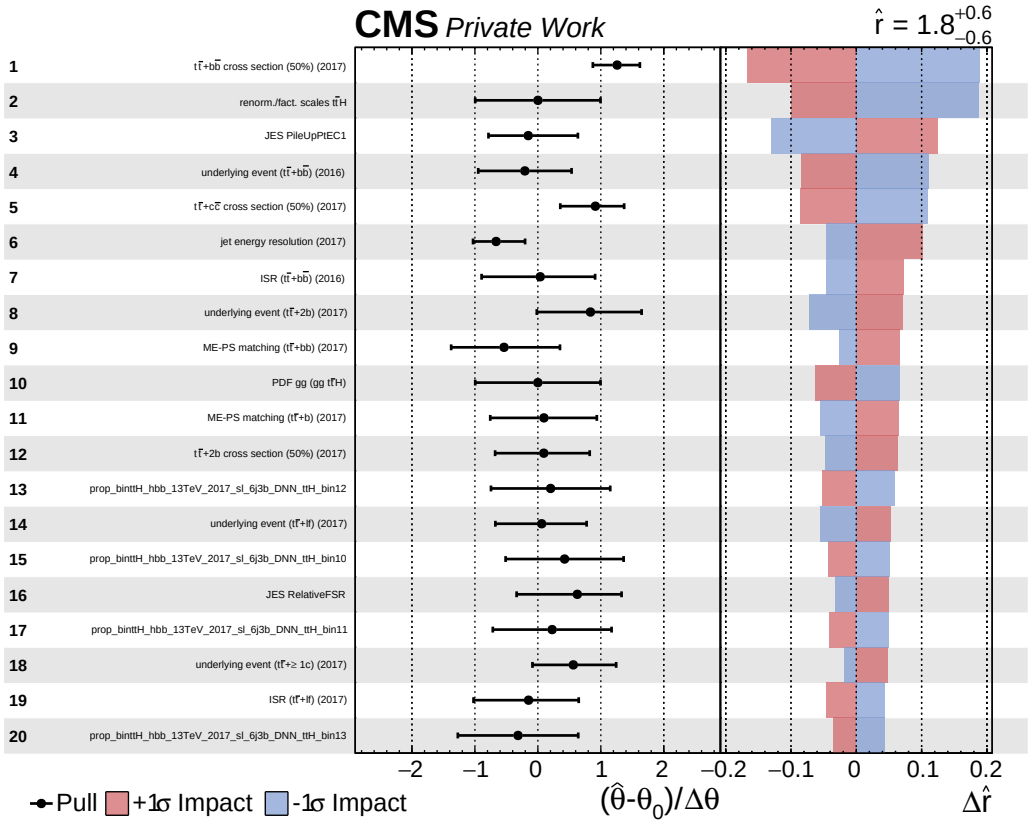


Figure D.71: Impacts of the 20 nuisance parameters with the largest impacts on the signal-strength modifier for the fit to data recorded in the year 2017. The impacts of the nuisance parameters are shown on the right hand side of the figure while their pulls and constraints are shown in the middle column.

E Event yields

In this section, the numbers of simulated and recorded events are shown before and after the joint fit to data recorded in the years 2016 and 2017, in each category defined by the multiplicities of jets and b-tagged jets and the prediction of the ANNs.

Table E.3: Numbers of simulated and recorded events recorded in 2016 in the six subclasses as defined by the ANN prediction in the ($4\,\text{jets}, \geq 3\,\text{b-tags}$) category before and after the joint fit to data recorded in the years 2016 and 2017. The quoted uncertainty is the total uncertainty, including statistical and systematic components.

	pre-fit (post-fit) yields											
Process	$t\bar{t}H$ node		$t\bar{t}+b\bar{b}$ node		$t\bar{t}+2b$ node		$t\bar{t}+b$ node		$t\bar{t}+c\bar{c}$ node		$t\bar{t}+$ lf node	
$t\bar{t}+$ light flavor	1249	(890)	727	(507)	1401	(1000)	1035	(763)	2909	(2085)	8463	(6447)
$t\bar{t}+c\bar{c}$	298	(571)	232	(440)	428	(837)	251	(499)	686	(1320)	1022	(2063)
$t\bar{t}+b$	253	(327)	215	(283)	370	(488)	326	(443)	308	(398)	469	(630)
$t\bar{t}+2b$	124	(98)	77	(64)	317	(258)	90	(75)	100	(81)	134	(111)
$t\bar{t}+b\bar{b}$	139	(131)	191	(183)	149	(133)	105	(98)	119	(109)	133	(123)
single t	96	(97)	117	(107)	167	(161)	93	(96)	231	(228)	304	(303)
V+jets	37	(36)	76	(72)	48	(44)	27	(26)	97	(83)	69	(66)
$t\bar{t}+V$	13	(12)	6	(5)	12	(10)	6	(5)	10	(9)	16	(15)
WW/WZ/ZZ	4	(3)	5	(4)	0.9	(0.9)	0.6	(0.7)	2	(2)	4	(4)
Total bkg.	2213	(2164)	1645	(1664)	2892	(2932)	1935	(2006)	4462	(4315)	10614	(9763)
$\pm$ tot unc.	$\pm$508	($\pm$56)	$\pm$415	($\pm$50)	$\pm$588	($\pm$81)	$\pm$402	($\pm$64)	$\pm$1051	($\pm$119)	$\pm$2359	($\pm$220)
$t\bar{t}H$	27	(36)	9	(12)	16	(21)	7	(9)	9	(12)	16	(22)
$\pm$ tot unc.	$\pm$4	($\pm$11)	$\pm$1	($\pm$4)	$\pm$2	($\pm$6)	$\pm$1.0	($\pm$3)	$\pm$1	($\pm$4)	$\pm$2	($\pm$7)
Data	2125		1793		2896		2027		4366		9693	

Table E.4: Numbers of simulated and recorded events recorded in 2016 in the six subclasses as defined by the ANN prediction in the ($5\,\text{jets}, \geq 3\,\text{b-tags}$) category before and after the joint fit to data recorded in the years 2016 and 2017. The quoted uncertainty is the total uncertainty, including statistical and systematic components.

	pre-fit (post-fit) yields					
Process	$t\bar{t}H$ node	$t\bar{t}+b\bar{b}$ node	$t\bar{t}+2b$ node	$t\bar{t}+b$ node	$t\bar{t}+c\bar{c}$ node	$t\bar{t}+\text{lf}$ node
$t\bar{t}+\text{light flavor}$	785 (508)	647 (406)	830 (525)	683 (474)	1148 (736)	4903 (3360)
$t\bar{t}+c\bar{c}$	336 (557)	341 (574)	445 (764)	264 (471)	552 (913)	1207 (2108)
$t\bar{t}+b$	257 (324)	290 (365)	355 (450)	321 (440)	219 (276)	494 (635)
$t\bar{t}+2b$	136 (107)	128 (101)	324 (254)	89 (76)	85 (67)	184 (147)
$t\bar{t}+b\bar{b}$	266 (237)	410 (375)	224 (194)	150 (136)	144 (124)	228 (201)
single t	62 (63)	82 (83)	98 (96)	45 (59)	114 (112)	189 (193)
V+jets	25 (22)	54 (50)	34 (30)	11 (12)	46 (38)	54 (49)
$t\bar{t}+V$	20 (18)	14 (12)	17 (15)	7 (6)	11 (9)	25 (22)
WW/WZ/ZZ	1 (1)	3 (3)	0.4 (0.4)	< 0.1 (< 0.1)	0.6 (0.3)	3 (4)
Total bkg.	1889 (1837)	1969 (1968)	2326 (2328)	1570 (1673)	2320 (2275)	7287 (6718)
$\pm$ tot unc.	$\pm$459 ($\pm$50)	$\pm$485 ($\pm$67)	$\pm$489 ($\pm$66)	$\pm$334 ($\pm$45)	$\pm$597 ($\pm$80)	$\pm$1655 ($\pm$160)
$t\bar{t}H$	53 (69)	21 (28)	20 (25)	8 (11)	11 (13)	28 (37)
$\pm$ tot unc.	$\pm$7 ($\pm$20)	$\pm$3 ($\pm$9)	$\pm$2 ($\pm$7)	$\pm$1 ($\pm$3)	$\pm$1 ($\pm$4)	$\pm$3 ($\pm$11)
Data	1848	2040	2299	1690	2302	6918

Table E.5: Numbers of simulated and recorded events recorded in 2016 in the six subclasses as defined by the ANN prediction in the ($\geq 6\,\text{jets}, \geq 3\,\text{b-tags}$) category before and after the joint fit to data recorded in the years 2016 and 2017. The quoted uncertainty is the total uncertainty, including statistical and systematic components.

	pre-fit (post-fit) yields					
Process	$t\bar{t}H$ node	$t\bar{t}+b\bar{b}$ node	$t\bar{t}+2b$ node	$t\bar{t}+b$ node	$t\bar{t}+c\bar{c}$ node	$t\bar{t}+\text{lf}$ node
$t\bar{t}+\text{light flavor}$	1982 (1194)	1280 (751)	852 (492)	916 (578)	243 (147)	50 (32)
$t\bar{t}+c\bar{c}$	1150 (1685)	998 (1451)	636 (944)	444 (682)	115 (170)	16 (24)
$t\bar{t}+b$	549 (654)	575 (686)	314 (377)	253 (314)	28 (32)	4 (4)
$t\bar{t}+2b$	306 (235)	282 (213)	372 (289)	78 (63)	10 (7)	1 (0.8)
$t\bar{t}+b\bar{b}$	834 (719)	1156 (1007)	299 (245)	145 (122)	17 (14)	3 (2)
single t	110 (118)	146 (143)	92 (79)	53 (51)	4 (4)	3 (3)
V+jets	38 (36)	78 (73)	34 (29)	10 (8)	7 (6)	0.6 (0.6)
$t\bar{t}+V$	80 (70)	58 (50)	31 (26)	11 (10)	4 (4)	0.4 (0.3)
WW/WZ/ZZ	0.9 (1)	0.5 (0.4)	0.4 (0.4)	0.4 (0.5)	< 0.1 (< 0.1)	< 0.1 (< 0.1)
Total bkg.	5049 (4713)	4575 (4375)	2629 (2482)	1911 (1828)	429 (385)	77 (67)
$\pm$ tot unc.	$\pm$1216 ($\pm$149)	$\pm$1156 ($\pm$122)	$\pm$603 ($\pm$74)	$\pm$422 ($\pm$56)	$\pm$107 ($\pm$12)	$\pm$18 ($\pm$2)
$t\bar{t}H$	142 (180)	53 (67)	24 (29)	10 (12)	2 (3)	0.3 (0.4)
$\pm$ tot unc.	$\pm$19 ($\pm$52)	$\pm$8 ($\pm$20)	$\pm$3 ($\pm$9)	$\pm$1 ($\pm$3)	$\pm$0.2 ($\pm$0.8)	$\pm < 0.1$ ($\pm$0.1)
Data	4822	4400	2484	1852	422	76

Table E.6: Numbers of simulated and recorded events recorded in 2017 in the six subclasses as defined by the ANN prediction in the ($4\,\text{jets}, \geq 3\,\text{b-tags}$) category before and after the joint fit to data recorded in the years 2016 and 2017. The quoted uncertainty is the total uncertainty, including statistical and systematic components.

	pre-fit (post-fit) yields					
Process	$t\bar{t}H$ node	$t\bar{t}+b\bar{b}$ node	$t\bar{t}+2b$ node	$t\bar{t}+b$ node	$t\bar{t}+c\bar{c}$ node	$t\bar{t}$ + lf node
$t\bar{t}$ + light flavor	2495 (2133)	529 (459)	1378 (1191)	1654 (1418)	3354 (2845)	12897 (10876)
$t\bar{t}+c\bar{c}$	473 (665)	178 (242)	361 (527)	293 (438)	798 (1112)	1599 (2198)
$t\bar{t}+b$	439 (676)	229 (373)	392 (594)	450 (752)	341 (580)	656 (1023)
$t\bar{t}+2b$	230 (206)	79 (73)	240 (215)	124 (117)	129 (122)	225 (199)
$t\bar{t}+b\bar{b}$	267 (393)	167 (237)	136 (203)	134 (220)	124 (201)	207 (313)
single t	190 (178)	61 (64)	171 (163)	150 (145)	297 (290)	554 (531)
V+jets	68 (65)	16 (15)	39 (38)	32 (32)	90 (86)	184 (168)
$t\bar{t}+V$	26 (22)	7 (6)	12 (10)	5 (5)	14 (12)	29 (24)
WW/WZ/ZZ	2 (2)	< 0.1 (< 0.1)	0.5 (0.5)	0.9 (0.8)	3 (2)	6 (5)
Total bkg.	4191 (4339)	1267 (1467)	2728 (2940)	2841 (3128)	5151 (5250)	16358 (15337)
$\pm$ tot unc.	$\pm$978 ($\pm$61)	$\pm$338 ($\pm$32)	$\pm$555 ($\pm$54)	$\pm$612 ($\pm$56)	$\pm$1211 ($\pm$77)	$\pm$3765 ($\pm$191)
$t\bar{t}H$	49 (58)	12 (14)	15 (17)	8 (10)	11 (13)	29 (35)
$\pm$ tot unc.	$\pm$6 ($\pm$17)	$\pm$2 ($\pm$4)	$\pm$2 ($\pm$5)	$\pm$1 ($\pm$3)	$\pm$1 ($\pm$4)	$\pm$3 ($\pm$10)
Data	4415	1485	2990	3157	5355	15289

Table E.7: Numbers of simulated and recorded events recorded in 2017 in the six subclasses as defined by the ANN prediction in the ($5\,\text{jets}, \geq 3\,\text{b-tags}$) category before and after the joint fit to data recorded in the years 2016 and 2017. The quoted uncertainty is the total uncertainty, including statistical and systematic components.

	pre-fit (post-fit) yields					
Process	$t\bar{t}H$ node	$t\bar{t}+b\bar{b}$ node	$t\bar{t}+2b$ node	$t\bar{t}+b$ node	$t\bar{t}+c\bar{c}$ node	$t\bar{t}$ + lf node
$t\bar{t}$ + light flavor	1099 (910)	488 (401)	749 (590)	959 (806)	1700 (1382)	8105 (6659)
$t\bar{t}+c\bar{c}$	432 (567)	261 (364)	365 (486)	302 (430)	807 (1100)	1893 (2582)
$t\bar{t}+b$	393 (590)	296 (428)	348 (524)	460 (712)	330 (520)	771 (1163)
$t\bar{t}+2b$	212 (170)	128 (112)	264 (235)	162 (123)	139 (117)	300 (249)
$t\bar{t}+b\bar{b}$	411 (573)	448 (621)	198 (288)	211 (310)	176 (251)	349 (502)
single t	91 (88)	74 (72)	125 (117)	74 (68)	159 (145)	337 (313)
V+jets	25 (23)	36 (34)	25 (24)	14 (12)	49 (44)	92 (78)
$t\bar{t}+V$	35 (29)	18 (14)	16 (14)	8 (7)	19 (17)	45 (38)
WW/WZ/ZZ	2 (1)	2 (2)	< 0.1 (< 0.1)	1 (1)	0.4 (1)	4 (3)
Total bkg.	2700 (2953)	1750 (2047)	2090 (2278)	2190 (2470)	3378 (3576)	11896 (11586)
$\pm$ tot unc.	$\pm$646 ($\pm$55)	$\pm$459 ($\pm$41)	$\pm$421 ($\pm$50)	$\pm$447 ($\pm$44)	$\pm$858 ($\pm$67)	$\pm$2611 ($\pm$146)
$t\bar{t}H$	83 (95)	25 (29)	19 (22)	15 (18)	18 (21)	46 (55)
$\pm$ tot unc.	$\pm$10 ($\pm$32)	$\pm$4 ($\pm$9)	$\pm$2 ($\pm$7)	$\pm$2 ($\pm$5)	$\pm$2 ($\pm$6)	$\pm$5 ($\pm$17)
Data	2993	2043	2302	2482	3651	11600

Table E.8: Numbers of simulated and recorded events recorded in 2017 in the six subclasses as defined by the ANN prediction in the (≥ 6 jets, ≥ 3 b-tags) category before and after the joint fit to data recorded in the years 2016 and 2017. The quoted uncertainty is the total uncertainty, including statistical and systematic components.

	pre-fit (post-fit) yields					
Process	$t\bar{t}H$ node	$t\bar{t}+b\bar{b}$ node	$t\bar{t}+2b$ node	$t\bar{t}+b$ node	$t\bar{t}+c\bar{c}$ node	$t\bar{t}$ + lf node
$t\bar{t}$ + light flavor	581 (454)	232 (177)	742 (589)	608 (466)	1077 (849)	4782 (3737)
$t\bar{t}+c\bar{c}$	448 (583)	256 (337)	596 (787)	326 (429)	908 (1164)	1889 (2496)
$t\bar{t}+b$	313 (447)	239 (334)	435 (655)	394 (552)	266 (413)	662 (988)
$t\bar{t}+2b$	185 (148)	149 (126)	435 (344)	156 (125)	156 (137)	303 (264)
$t\bar{t}+b\bar{b}$	714 (963)	845 (1133)	462 (647)	295 (428)	318 (434)	470 (674)
single t	56 (51)	62 (61)	104 (102)	46 (41)	94 (89)	209 (199)
V+jets	15 (13)	23 (20)	22 (20)	11 (9)	24 (24)	57 (55)
$t\bar{t}+V$	49 (39)	29 (23)	38 (31)	15 (13)	33 (27)	74 (61)
WW/WZ/ZZ	0.2 (0.2)	0.7 (0.6)	0.5 (0.5)	1.0 (0.3)	0.9 (0.9)	0.5 (0.3)
Total bkg.	2362 (2699)	1835 (2210)	2834 (3176)	1853 (2063)	2877 (3138)	8447 (8474)
$\pm$ tot unc.	$\pm$632 ($\pm$56)	$\pm$552 ($\pm$56)	$\pm$624 ($\pm$78)	$\pm$407 ($\pm$46)	$\pm$846 ($\pm$69)	$\pm$1871 ($\pm$119)
$t\bar{t}H$	139 (155)	41 (46)	42 (48)	23 (27)	31 (36)	65 (76)
$\pm$ tot unc.	$\pm$18 ($\pm$48)	$\pm$6 ($\pm$14)	$\pm$5 ($\pm$15)	$\pm$3 ($\pm$8)	$\pm$5 ($\pm$11)	$\pm$7 ($\pm$22)
Data	2930	2253	3215	2087	3114	8538

Danksagung

Zunächst möchte ich mich bei Prof. Dr. Ulrich Husemann für die Ermöglichung meiner Promotion bedanken. Außerdem bedanke ich mich für das schnelle und gründliche Korrekturlesen, die Sommerschulen und Konferenzen die ich besuchen durfte, die Aufnahme in seine großartige Arbeitsgruppe, die gute Zusammenarbeit und die sehr gute und lehrreiche Betreuung und Unterstützung, sowohl während dieser Promotion als auch schon während meiner Bachelor- und Masterarbeit. Bei Ihm hatte ich das Privileg an Veröffentlichungen zu aktuellen Themen der Teilchenphysik mitzuwirken und die Möglichkeit frei an weniger brisanten Fragestellungen zu forschen und mich persönlich weiterzuentwickeln.

Des Weiteren möchte ich mich bei Prof. Dr. Thomas Müller für die Übernahme des Korreferats und seine Unterstützung, sowohl bei dieser Promotion als auch bei meiner Masterarbeit, bedanken.

Ich möchte mich auch bei Dr. Matthias Schröder für die gute Zusammenarbeit, die Hilfe bei meiner Arbeit und den guten Rat bedanken.

Außerdem bedanke ich mich bei Philip Keicher, Simon Kudella, Jan van der Linden, Andrej Saibel, Korbinian Schweiger, Michael Waßmer, Sebastian Wieland, sowie allen anderen derzeitigen und ehemaligen Mitgliedern meiner Arbeitsgruppe für die tolle Zeit und die Hilfe, ohne die meine Arbeit sicherlich nicht möglich gewesen wäre. Mein besonderer dank gilt auch Marco Alexander Harrendorf, Hannes Mildner und Shawn Williamson, die mir in den ersten Jahren quasi alles praktische beigebracht haben. Ich bedanke mich auch bei allen anderen Mitglieder des Instituts für experimentelle Teilchenphysik für die gute Atmosphäre und die Unterstützung. Ich möchte mich auch besonders bei Frau Fellner, Frau Bräunling, Frau Gering und Frau Chen, der Admingruppe des ETPs sowie allen anderen die durch die Bewältigung der Bürokratie und die Bereitstellung der Infrastruktur meine Arbeit überhaupt erst ermöglicht haben, bedanken.

I would also like to thank all members of the CMS Collaboration that contributed to this work, especially the members of the HIG group and María Aldaya Martín.

Zu guter Letzt möchte ich mich bei meiner Familie dafür bedanken, dass sie mich immer unterstützt und mir das Studium und diese Promotion ermöglicht haben.