Karlsruher Institut für Technologie

IEKP-KA/2017-02

Suche nach Higgs-Boson-Produktion in Assoziation mit einem Top-Quark als Test der Higgs-Boson-Kopplungen bei 13 TeV mit dem CMS-Experiment

Denise Müller

Fakultät für Physik
Karlsruher Institut für Technologie (KIT)

MASTERARBEIT

Referent: Prof. Dr. Thomas Müller
Korreferent: Prof. Dr. Ulrich Husemann

Institut für Experimentelle Kernphysik

18. Januar 2017

Akzeptiert vom ersten Referenten der Masterarbeit.
Karlsruhe, den 18. Januar 2017

. . . . . .
(**Prof. Dr. Thomas Müller**)

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, den 18. Januar 2017

. . . . . .

(**Denise Müller**)

Deutsche Zusammenfassung

Das Higgs-Boson war das letzte gesuchte Elementarteilchen, das vom Standardmodell (SM) der Teilchenphysik vorhergesagt wurde. Die ersten Theorien des Higgs-Mechanismus entstanden bereits vor rund 50 Jahren, jedoch waren mehrere Experimente an unterschiedlichen Teilchenbeschleunigern bis zur Entdeckung des Higgs-Bosons erforderlich. Die Entdeckung dieses Bosons erfolgte im Jahr 2012 am Large Hadron Collider (LHC) in Genf mit Hilfe der beiden Mehrzweck-Detektoren CMS und ATLAS [1, 2]. Der LHC ist ein unterirdischer Ringbeschleuniger, in welchem Protonen mit einer hohen Energie beschleunigt und zur Kollision gebracht werden. An den Kollisionspunkten befinden sich Teilchendetektoren, die die Zerfallsprodukte der Kollision vermessen. Die auf diese Weise gewonnenen Daten werden anschließend für die Rekonstruktion von Ereignissen, bestehend aus physikalischen Objekten wie z. B. Myonen, verwendet. Um zu bestätigen, dass es sich bei dem gefundenen Boson tatsächlich um das vorhergesagte Higgs-Boson handelt, sind präzise Messungen der Eigenschaften dieses Bosons notwendig. Die bisherigen Ergebnisse weisen zwar darauf hin, dass es sich bei dem entdeckten Boson um das Higgs-Boson handelt, jedoch können kleinste Abweichungen vom Standardmodell ein Hinweis auf neue Physik jenseits des Standardmodells sein.

Eine Eigenschaft, die sich hinsichtlich solcher Abweichungen untersuchen lässt, ist die Kopplung des Higgs-Bosons an Top-Quarks sowie an massive Eichbosonen. Diese beiden Kopplungen lassen sich direkt durch die Produktion eines Higgs-Bosons in Assoziation mit einem einzelnen Top-Quark (tH) analysieren. Diese Arbeit befasst sich daher mit der Suche nach diesem bisher nicht entdeckten Produktionskanal in Abhängigkeit der Higgs-Boson-Kopplungen. Für die Suche wird der gesamte im Jahr 2015 bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV vom CMS-Detektor aufgenommene Datensatz, welcher einer integrierten Luminosität von $L_{\text{int}} = 2{,}3\ \text{fb}^{-1}$ entspricht, analysiert. Die Analyse fokussiert sich auf den Zerfall des Higgs-Bosons in ein Bottom-Quark-Paar sowie auf den leptonischen Zerfall des Top-Quarks. Somit besteht der Endzustand des gesuchten Prozesses neben den beiden Bottom-Quarks des Higgs-Bosons aus einem weiteren Bottom-Quark, einem geladenen Lepton sowie einem Neutrino.

Diese Zusammensetzung des Endzustandes erlaubt bereits eine signifikante Reduktion des Untergrundes, allerdings liegt noch immer ein geringes Signal-zu-Untergrund-Verhältnis vor.

Die Arbeit selbst ist in zwei Analysen unterteilt: Eine Referenzanalyse, welche auf der publizierten Arbeit der CMS-Kollaboration [3] basiert, sowie einer Analyse, welche das Ziel hat, die Referenzanalyse zu optimieren und die Ergebnisse zu verbessern.

Im Laufe der Arbeit werden multivariate Analysemethoden eingesetzt, wie eine Ereignisrekonstruktion sowie eine Ereignisklassifikation mit Hilfe von Entscheidungsbäumen. Die Ereignisrekonstruktion gestaltet sich hierbei schwierig, da die Rekonstruktion von Higgs-Boson- sowie von Top-Quark-Kandidaten eine korrekte Identifizierung und Zuordnung von mindestens vier sogenannten Jets erfordert. Ein Jet bezeichnet hierbei einen gebündelten Teilchenstrahl, welcher durch die Hadronisierung einzelner Quarks oder Gluonen erzeugt wird. Aufgrund der vielen Möglichkeiten, die Jets den entsprechenden Quarks des Endzustandes zuzuordnen, werden für die Ereignisrekonstruktion Entscheidungsbäume verwendet, welche mit Hilfe von mehreren Ja- oder Nein-Entscheidungen bestimmen, welche Jet-Quark-Zuordnung korrekt und welche falsch ist. Da die semileptonische Top-Quark-Paarproduktion einen dominierenden Untergrund darstellt, erfolgt die Rekonstruktion nicht nur unter der Hypothese eines Signalereignisses, sondern auch unter der Hypothese, dass es sich bei dem Ereignis um ein semileptonisch zerfallendes Top-Quark-Paar handelt. Nach der Ereignisrekonstruktion erfolgt mit Hilfe von weiteren Entscheidungsbäumen eine abschließende Klassifikation der Ereignisse als signal- oder untergrundartig. Die Ausgabevariable der Ereignisklassifikation, welche kontinuierlich verteilt ist, dient als Grundlage für die Anpassung der simulierten Daten an den experimentell gemessenen Datensatz. Aus dieser Anpassung werden anschließend obere Grenzen auf den Produktionswirkungsquerschnitt des tH-Prozesses in Abhängigkeit von insgesamt 51 unterschiedlichen betrachteten Higgs-Boson-Kopplungen berechnet.

Die zweite Analyse unterscheidet sich von der oben beschriebenen Referenzanalyse dadurch, dass eine zusätzliche Klassifikation bezüglich der Top-Quark-Paarproduktion durchgeführt wird. Diese hat das Ziel, mit Hilfe von Entscheidungsbäumen zu bestimmen, ob die zusätzlich im Anfangs- oder Endzustand dieses Prozesses abgestrahlten Jets ihren Ursprung in Bottom-Quarks, Charm-Quarks oder leichteren Quarks haben. Für diese sogenannte Flavor-

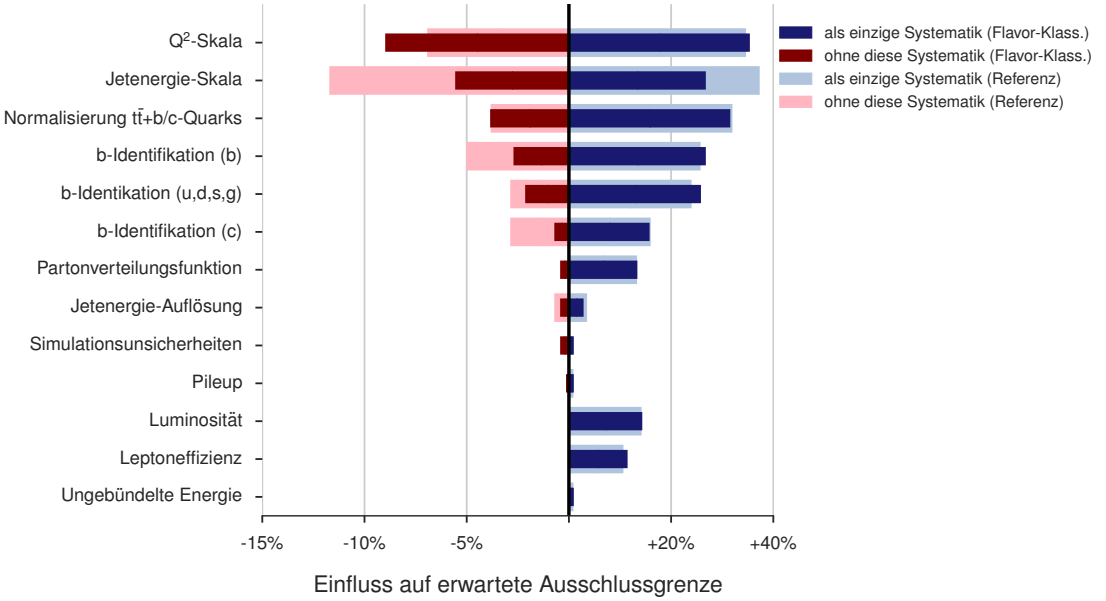


Abbildung I: Einfluss der systematischen Unsicherheiten auf die erwartete Ausschlussgrenze. Der Einfluss ist dargestellt für die Referenzanalyse (helle Balken) sowie für die mittels Flavor-Klassifikation optimierte Analyse (dunkle Balken). Die Auswirkungen auf die Ausschlussgrenze wurden zum einen durch das Entfernen einzelner systematischer Unsicherheiten bestimmt (rot) und zum anderen berechnet, wenn nur eine bestimmte systematische Unsicherheit berücksichtigt wird (blau).

Klassifikation, welche in einer dileptonischen Kontrollregion durchgeführt wird, werden sowohl „b tagging“- als auch „c tagging“-Algorithmen eingesetzt. Der „c tagging“-Algorithmus basiert auf den für die Identifizierung von b-Jets verwendeten „Combined Secondary Vertex“ (CSV)-Algorithmus und ist in der Lage, Jets, welche von Charm-Quarks stammen (c-Jets), von b-Jets sowie von Jets, welche von leichten Quarks erzeugt wurden, zu unterscheiden. Die Ausgangsvariable der Flavor-Klassifikation wird bei der Anpassung der simulierten an die gemessenen Ereignisse mit berücksichtigt, aus welcher wie bereits bei der Referenzanalyse beschrieben Ausschlussgrenzen auf die tH-Produktion bestimmt werden.

In Abbildung I ist der Einfluss der systematischen Unsicherheiten auf die erwartete Ausschlussgrenze sowohl für die Referenzanalyse als auch für die mittels Flavor-Klassifikation optimierte Analyse dargestellt. Eine deutliche Reduktion der systematischen Unsicherheiten auf die Jetenergieskala sowie auf die Identifikation von b-Jets ist erzielt worden, welche sich positiv auf die Ausschlussgrenze auswirkt. Die Verbesserung der Ausschlussgrenzen ist in Tabelle I für den

Tabelle I: Ausschlussgrenzen in einem 95 % Konfidenzintervall auf die Produktion eines Higgs-Bosons in Assoziation mit einem Top-Quark. Angegeben sind die Ergebnisse für die Referenzanalyse sowie für die optimierte Analyse für zwei ausgewählte Higgs-Boson-Kopplungen.

Analyse	Erwartete Ausschlussgrenze	Verbesserung ggü. Referenzanalyse in %	Beobachtete Ausschlussgrenze
Standardmodell-Kopplung (SM)			
Referenz	$99{,}5 \times \sigma_{\mathrm{SM}}$	-	$116{,}1 \times \sigma_{\mathrm{SM}}$
Flavor-Klassifikation	$91{,}3 \times \sigma_{\mathrm{SM}}$	8,2	$102{,}1 \times \sigma_{\mathrm{SM}}$
Invertierte Top-Quark-Kopplung (ITC)			
Referenz	$6{,}5 \times \sigma_{\mathrm{ITC}}$	-	$6{,}2 \times \sigma_{\mathrm{ITC}}$
Flavor-Klassifikation	$5{,}5 \times \sigma_{\mathrm{ITC}}$	15,4	$4{,}7 \times \sigma_{\mathrm{ITC}}$

Fall einer vom Standardmodell vorhergesagten Higgs-Boson-Kopplung sowie für den Fall einer invertierten Kopplung des Higgs-Bosons an das Top-Quark zusammengefasst. Diese besondere Kopplung, welche sich nicht durch den Betrag, aber durch das Vorzeichen von der Standardmodell-Kopplung unterscheidet, wird als „inverted top coupling“ (ITC) bezeichnet und wurde bereits bei einer Schwerpunktsenergie von 8 TeV näher untersucht [4].

Die Referenzanalyse schließt im Fall der Standardmodell-Kopplung des Higgs-Bosons die tH-Produktion für einen Wirkungsquerschnitt von $99{,}5 \times \sigma_{\mathrm{SM}}$ aus, während für das ITC-Szenario eine deutlich geringere Ausschlussgrenze von $6{,}5 \times \sigma_{\mathrm{ITC}}$ bestimmt wurde. Diese kommt durch den deutlich erhöhten Produktionswirkungsquerschnitt aufgrund der invertierten Top-Quark-Kopplung zustande. Mit Hilfe der zusätzlichen Flavor-Klassifikation lässt sich die Ausschlussgrenze im Fall des Standardmodells auf $91{,}3 \times \sigma_{\mathrm{SM}}$ reduzieren, was einer Verbesserung von 8,2 % entspricht. Im Fall des ITC-Szenarios wurde die erwartete obere Grenze um 15,4 % auf einen Wert von $5{,}5 \times \sigma_{\mathrm{ITC}}$ gesenkt.

Die Ergebnisse zeigen, dass sich der tH-Prozess für beide betrachteten Higgs-Boson-Kopplungen bisher nicht ausschließen lässt. Dies wird sich, wie eine in dieser Arbeit durchgeführten Studie zeigt, durch die Aufnahme weiterer Daten ändern, sodass in naher Zukunft zumindest der mögliche Ausschluss der tH-Produktion unter der Annahme einer invertierten Kopplung des Higgs-Bosons an das Top-Quark in Reichweite rückt. Die Produktion eines Higgs-Bosons in Assoziation mit einem Top-Quark wird somit einen entscheidenden Beitrag zum Verständnis der Natur der Higgs-Boson-Kopplungen liefern können.

Karlsruher Institut für Technologie

IEKP-KA/2017-02

Search for Higgs Boson Production in Association with a Single Top Quark as a Test of Higgs Boson Couplings at 13 TeV with the CMS Experiment

Denise Müller

Department of Physics
Karlsruhe Institute of Technology (KIT)

MASTER THESIS

Referee: Prof. Dr. Thomas Müller
Co-Referee: Prof. Dr. Ulrich Husemann

Institut für Experimentelle Kernphysik

January 18, 2017

Introduction

The Higgs boson has been the last missing particle predicted by the Standard Model (SM) of particle physics. First theories on the Higgs mechanism arose around 50 years ago, but several particle colliders were required until the boson has been finally discovered by the CMS and ATLAS collaborations at the Large Hadron Collider (LHC) in 2012 [1, 2]. The discovery of the Higgs boson is essential for the SM, as it proves the existence of a scalar field, the so-called Higgs field, that gives particles their masses.

In order to confirm that the discovered boson with a mass of 125 GeV is indeed the Higgs boson predicted by the SM, several of its properties need to be measured. Small deviations from SM predictions could be an indication of physics beyond the Standard Model (BSM). For instance, such deviations can occur for the Higgs boson couplings to top quarks and to vector bosons. Hence, further studies on this topic are required.

The Higgs boson couplings can be directly accessed by studying the production of a single top quark produced in association with a Higgs boson (tH) as this process is not only highly sensitive to the magnitude of the couplings, but also to their relative sign. The SM predicts only a very small cross section, whereas anomalous couplings can increase it significantly.

This thesis documents the direct search for the tH process under the assumption of different Higgs boson couplings and the Higgs boson decaying into a bottom quark pair. Boosted decision trees are employed for the event reconstruction and classification into signal and background events. Furthermore, an additional classification of the dominating background, top quark pair production, is performed. This classification is based on the flavor of the additional jets which are radiated in the initial or final state of the top quark pair production. Upper limits on the cross section of the tH process are determined by using the 2015 data set, recorded by the CMS experiment at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with an integrated luminosity of $2.3\,\text{fb}^{-1}$.

Chapter 1 introduces the theoretical principles of the Standard Model, the Higgs mechanism as well as the physics of the Higgs boson and the top quark. The LHC accelerator complex and the CMS experiment are described in detail in Chapter 2. Chapter 3 describes the simulation of the events using Monte Carlo methods and the reconstruction of physical objects. The applied boosted decision trees and statistical methods are explained in Chapter 4. Chapter 5 addresses the actual analysis, which is divided in two sub analyses: one reference analysis and one analysis with an additional classification of top quark pair production. Furthermore, a projection to a higher luminosity is presented in this chapter. Finally, a summary and conclusion of this thesis are given in Chapter 6.

Contents

1. Theory

The Standard Model (SM) of particle physics is a fundamental theory which describes the properties of all known elementary particles as well as their interactions. This quantum field theory, developed throughout the 1960s and 1970s, predicted many new particles, such as the top quark, which have been discovered afterwards. The last missing piece of the Standard Model, the Higgs boson, has been experimentally verified in 2012 at the Large Hadron Collider (LHC) in Geneva [1, 2]. It was first theoretically introduced in 1964 [5–7] and is part of the so-called Higgs mechanism. This mechanism is essential since the Higgs field causes the electroweak symmetry breaking which is responsible for the mass of the fundamental particles.

This chapter gives an overview of the particles and their interactions in the Standard Model and shortly introduces theoretical concepts of modern particle physics. In the second part of this chapter, the properties of the Higgs boson and of the top quark are explained. In all formulas, natural units are used for simplification, i. e. $\hbar = c = 1$.

1.1. Standard Model

All observed elementary particles are influenced by only four fundamental forces: electromagnetic, strong and weak interaction, and gravity. The first three forces are described by the Standard Model. In case of gravity, no quantum field theory exists till this day. Since the particle interactions only occur at small energy scales, gravity is negligible on a microscopic scale. On a macroscopic scale, the gravitational forces are described by the theory of General Relativity [8]. An overview of the three interactions of the Standard Model, as well as the properties of their gauge bosons, is given in Table 1.1.

The Standard Model describes each force with a relativistic quantum field theory and includes every observed elementary particle. It is a gauge theory which combines the interactions with the symmetry group $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$.

Table 1.1.: The three interactions of particles in the Standard Model. The properties of the gauge bosons for the electromagnetic, strong and weak force are listed. All values are taken from [9].

Force	couples to	Gauge boson	Mass (GeV)
electromagnetic	electric charge	photon (γ)	0
strong	color charge	8 gluons (g)	0
weak	weak charge	$W^{\pm}$	80.385
		Z^0	91.188

The $SU(3)_C$ gauge group represents the strong interaction and is described by the theory of quantum chromodynamics (QCD) [10, 11]. The gauge bosons of the strong force, the gluons, are massless and carry a color charge and an anticolor charge. Three types of color charge exist: red, green and blue. The gluons couple to all particles which carry a color charge. Since the gluons themselves carry color and anticolor charges, they are able to interact among themselves. These circumstances require eight different gluons for this complex symmetry group instead of simply one gauge boson.

The theory of the electromagnetic force is called quantum electrodynamics (QED) [12–14]. In this theory, the photon is the gauge boson which couples to particles with electric charge. In contrast to QCD, the massless photons cannot interact among each other because they do not carry electric charge themselves.

It is not possible to describe the weak interaction itself by a quantum field theory, but the weak and the electromagnetic force can be combined under the electroweak theory described by the Glashow-Weinberg-Salam model [15–17]. This theory is represented by the symmetry $SU(2)_L \otimes U(1)_Y$ with the massive gauge bosons $W^{\pm}$ and Z^0. The generators of the $SU(2)_L$ and $U(1)_Y$ group are the weak isospin and the hypercharge, respectively. The fact that only the gauge bosons of the weak interaction are massive is an important issue which is resolved by the introduction of the Higgs mechanism. This mechanism will be explained later in this chapter.

The particles of the Standard Model are classified according to their spin. An overview of this classification is given in Figure 1.1. Particles with a spin of $1/2$

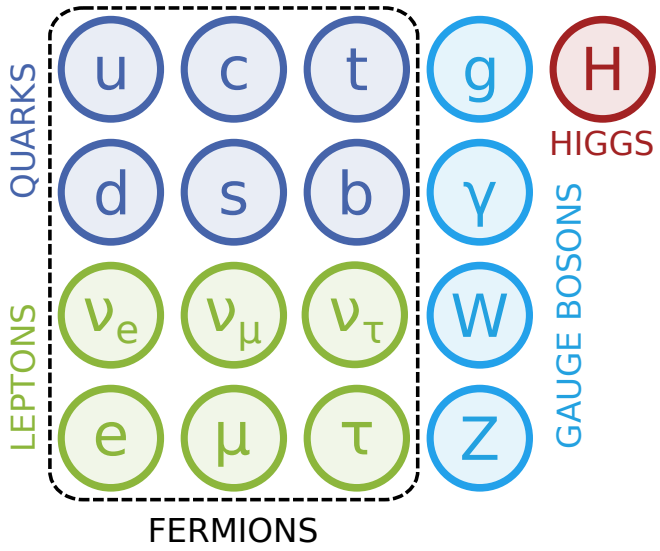


Figure 1.1.: The Standard Model of particle physics. The Standard Model consists of two different classes of elementary particles: Fermions (i. e. quarks and leptons) and bosons (i. e. gauge bosons and the Higgs boson).

are called fermions, while the gauge bosons of the three interactions always possess spin 1. In contrast to the other bosons of the Standard Model, the Higgs boson is a scalar boson with a spin of 0.

Fermions are grouped in six leptons and in six quarks. For every fermion an anti-particle with the same mass and spin exists. In contrast to the particles, anti-particles possess inverted electric charge, color and weak isospin. The leptons and the quarks can both be grouped according to their generation and by their 3^{rd} component of the weak isospin. The properties of the fermions are given in Table 1.2.

Since quarks possess electric charge, color charge and weak isospin, they are the only particles that interact with all known forces of the Standard Model. Quarks are divided according to their electric charge into three up-type and three down-type quarks. The up-type quarks (u, c, t) have an electric charge of $+2/3\,e$, the down-type quarks (d, s, b) of $-1/3\,e$. Each generation consists of one up-type quark and one down-type quark. It is not possible to observe free quarks due to the so-called confinement of QCD since the strength of the

Table 1.2.: Properties of fermions. Fermions are divided into quarks and leptons. There are six quarks and six leptons in total. In contrast to leptons, quarks have an additional color charge. The fermions are ordered by their generation and by their 3rd component of the weak isospin. All values are taken from [9].

Fermions	Generation			Electric	3rd comp.
	1	2	3	charge (e)	of isospin
Quarks	up (u)	charm (c)	top (t)	$+\frac{2}{3}$	$+\frac{1}{2}$
	down (d)	strange (s)	bottom (b)	$-\frac{1}{3}$	$-\frac{1}{2}$
Leptons	ν_e	ν_μ	ν_τ	0	$+\frac{1}{2}$
	e	μ	τ	-1	$-\frac{1}{2}$

gluon field increases with longer distances. Therefore, quarks have to appear in groups of three quarks or three anti-quarks (baryons), or in groups of one quark and one anti-quark (mesons).

Similar to quarks, the leptons can also be divided according to their electric charge. The three charged leptons (e, μ, τ) possess an electric charge of $-1\,e$, while the three neutrinos (ν_e, ν_μ, ν_τ), which are massless in the Standard Model, carry no electric charge. However, recent experiments prove that neutrinos have at least small masses, as they are able to oscillate between the three flavors [18]. Each charged lepton constitutes, together with a neutrino of the same flavor, one generation. Since the leptons do not possess color charge, they cannot interact via the strong force. Charged leptons are able to interact via weak and electromagnetic force, while neutrinos can only couple to the gauge bosons of the weak interaction.

The fundamental bosons of the Standard Model are classified by their spin. The vector bosons with spin 1 are the gauge bosons that mediate interactions between the particles, while the scalar boson with spin 0 is called Higgs boson, which is the excitation of the Higgs field.

There are twelve gauge bosons in total: the photon for the electromagnetic interaction, the eight gluons for the strong interaction and the three bosons $W^\pm$ and Z^0 for the weak interaction. Since the ranges of the interactions are inversely proportional to the masses of the gauge bosons, the range of the electromagnetic force is infinite due to the massless photon. However, the range

of the massless gluons is restricted to $\mathcal{O}(1\,\mathrm{fm})$ due to their self-interactions. The gauge bosons of the weak force are massive, 80.4 GeV in case of $W^{\pm}$ and 91.2 GeV in case of Z^0, which results in a short range of the weak interaction of only $\mathcal{O}(10^{-3}\,\mathrm{fm})$.

Furthermore, it has to be mentioned that the Standard Model cannot explain all observed phenomena in nature. Examples are, as already stated above, the neutrino oscillations, as well as the observation of dark matter and dark energy in the universe, and the so-called hierarchy problem. An overview of the outstanding questions of the Standard Model can be found in [19]. These issues could be potentially solved by physics beyond the Standard Model (BSM) (see e. g. [20]). However, there is no evidence found for BSM theories yet.

1.2. Spontaneous Symmetry Breaking and Higgs Mechanism

In the following section, the theory behind the unification of the weak and the electromagnetic force as well as the crucial role of the Higgs mechanism for the masses of the weak gauge bosons and of the fermions are presented. Details concerning the Lagrangian formalism, the gauge invariance and the Higgs couplings to bosons and fermions can be found, for example, in [21–23].

Without the Higgs field, it is difficult to explain why the weak gauge bosons are massive since mass terms are forbidden in the Lagrangian by gauge invariance. These particles obtain their masses by the spontaneous symmetry breaking caused by the Higgs mechanism. In this mechanism, a new field ϕ, called the Higgs field, which corresponds to a SU(2) doublet of two complex scalar fields, is introduced:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} . \tag{1.1}$$

This leads to the following Lagrangian, where $V(\phi)$ denotes the Higgs potential:

$$\mathcal{L}_{\mathrm{Higgs}} = (\partial^\mu \phi)^\dagger (\partial_\mu \phi) - V(\phi) = (\partial^\mu \phi)^\dagger (\partial_\mu \phi) + \mu^2 (\phi^\dagger \phi) - \lambda (\phi^\dagger \phi)^2 . \tag{1.2}$$

The form of the Higgs potential depends on the sign of μ^2. For $\mu^2 > 0$, only one minimum at $\phi = 0$ occurs, while for $\mu^2 < 0$ two minima exist in one dimension

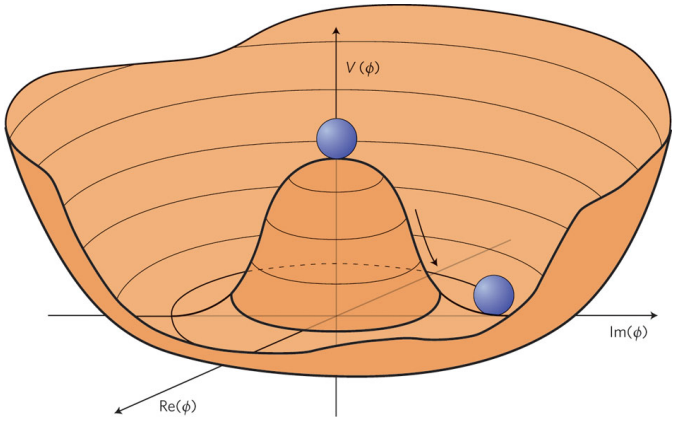


Figure 1.2.: The two-dimensional Higgs potential. The real and the imaginary parts of the Higgs field for the case $\mu^2 < 0$ are shown. In the two-dimensional case, not only two minima, but an infinite number of minima located at a circle with radius $|\phi| = v$ exist. Due to its unique shape, the Higgs potential is often simply called "Mexican hat potential". The figure is taken from [24].

and infinite minima in two dimensions, respectively. For the second case the minima are given by:

$$\phi^2 = \frac{-\mu^2}{\lambda} =: v^2 \,, \tag{1.3}$$

where $v = \sqrt{-\mu^2/\lambda}$ corresponds to the vacuum expectation value of 246 GeV. Due to the infinite number of minima, the U(1) symmetry is spontaneously broken since one particular minimum has to be chosen. This ground state is not invariant under U(1) transformations anymore. If only one distinct minimum exists (i. e. $\mu^2 > 0$), the ground state is invariant under U(1) symmetry. The general form of the two-dimensional Higgs potential is illustrated in Figure 1.2.

The charged component ϕ^+ of the Higgs field given in Equation 1.1 has to be set to zero, otherwise the electric charge would not be conserved anymore. The neutral component ϕ^0 can be rewritten with v and with a scalar field component $h(x)$:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \,. \tag{1.4}$$

In order to preserve local gauge invariance, the covariant derivative needs to

be used in the Lagrangian

$$\mathcal{L}_{\text{Higgs}} = (D^{\mu}\phi)^{\dagger}(D_{\mu}\phi) - V(\phi) \ . \tag{1.5}$$

Here, the covariant derivative D_{μ} is given by

$$D_{\mu} = \partial_{\mu} - \frac{\mathrm{i}}{2}g'B_{\mu} - \frac{\mathrm{i}}{2}gW_{\mu}^{j}\sigma^{j} \ . \tag{1.6}$$

In this notation, the parameter g' corresponds to the coupling strength of the hypercharge, and the parameter g to the coupling strength of the weak gauge field. The three gauge fields for the weak interaction are given by W_{μ}^{j} (with $j = 1, 2, 3$), while B_{μ} is the representation of the gauge field for the U(1) hypercharge. The three Pauli matrices are denoted by σ^{j}.

By introducing a weak mixing angle θ_{W}, called the Weinberg angle, one receives the gauge bosons from a rotation of the gauge fields:

$$A = B\cos\theta_{\text{W}} + W^{3}\sin\theta_{\text{W}} \ , \tag{1.7}$$

$$Z = -B\sin\theta_{\text{W}} + W^{3}\cos\theta_{\text{W}} \ , \tag{1.8}$$

$$W^{\pm} = \frac{1}{\sqrt{2}}(W^{1} \mp \mathrm{i}W^{2}) \ , \tag{1.9}$$

with

$$\cos\theta_{\text{W}} := \frac{g}{\sqrt{g^{2} + g'^{2}}} \ . \tag{1.10}$$

The experimentally measured value is $\sin^{2}\theta_{\text{W}} = 0.23129 \pm 0.00005$ [9].

The resulting masses of the gauge bosons can be derived from the mass terms of the Lagrangian by using the definition of the Higgs field in Equation 1.4. One receives:

$$M_{\gamma} = 0 \ , \tag{1.11}$$

$$M_{\text{Z}} = \frac{v}{2}\sqrt{g^{2} + g'^{2}} \ , \tag{1.12}$$

$$M_{\text{W}} = \frac{vg}{2} = M_{\text{Z}}\cos\theta_{\text{W}} \ , \tag{1.13}$$

where the photon γ corresponds to the gauge field A. The scalar field component $h(x)$ provides the mass of the Higgs boson:

$$m_{\mathrm{H}} = \sqrt{-2\mu^2}\,. \tag{1.14}$$

The Higgs boson mass is experimentally determined as $125.09 \pm 0.24\,\mathrm{GeV}$ [9].

With the electroweak symmetry breaking explained above, only the masses of the gauge bosons can be derived. The mass of the fermions can be explained by the Yukawa coupling of the Higgs boson to fermions. This coupling corresponds to an interaction between the Higgs field ϕ and the fermion fields ψ. The masses of quarks and charged leptons are given by the mass term of the following Lagrangian:

$$\mathcal{L} = -m(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R)\,, \tag{1.15}$$

where ψ_L denotes the SU(2) doublet and ψ_R the SU(2) singlet. It is to be noted that this Lagrangian is not gauge invariant. In case of an electron, one can write a gauge invariant Lagrangian by using Equation 1.4 as follows:

$$\mathcal{L}_{\mathrm{Yukawa}} = -y_{\mathrm{e}}\left[\begin{pmatrix}\overline{\nu}_{\mathrm{e}} & \overline{e}\end{pmatrix}_L\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}e_R + \overline{e}_R\begin{pmatrix}\tilde{\phi}^+ & \tilde{\phi}^0\end{pmatrix}\begin{pmatrix}\nu_{\mathrm{e}}\\ e\end{pmatrix}_L\right] \tag{1.16}$$

$$= -\frac{y_{\mathrm{e}}}{\sqrt{2}}(v+h)(\overline{e}_Le_R + \overline{e}_Re_L)\,, \tag{1.17}$$

where y_{e} corresponds to the Yukawa coupling of the Higgs boson to the electron. By comparison with Equation 1.15, the mass of an electron is defined by

$$m_{\mathrm{e}} = \frac{y_{\mathrm{e}}v}{\sqrt{2}}\,. \tag{1.18}$$

The masses of muons and tauons are received the same way. In case of quarks the same mass definition can be obtained via a slightly different derivation. Since there exist no right-handed neutrinos according to the Standard Model, neutrinos cannot receive their mass via Yukawa coupling.

1.3. Higgs Boson Properties

As the Higgs boson has been discovered just a few years ago, not all of its properties that are derived by the Standard Model are experimentally confirmed.

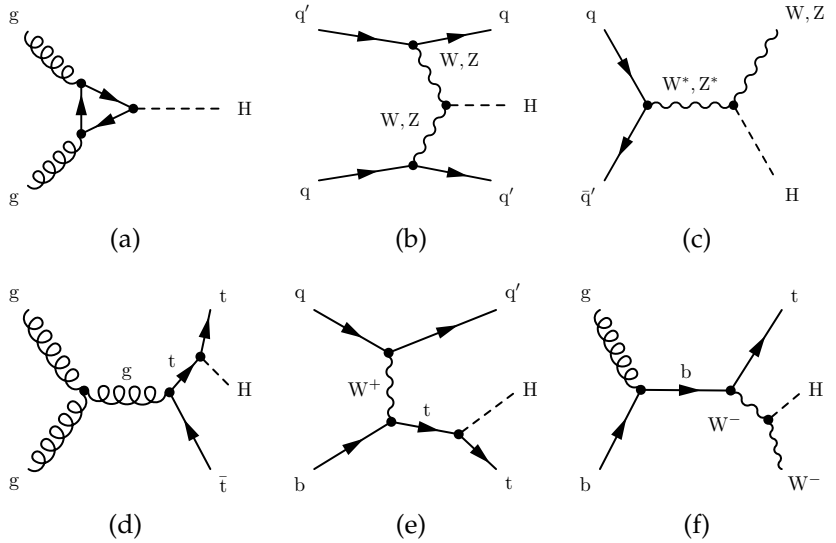


Figure 1.3.: Higgs boson production modes. The Higgs boson can be produced via gluon-gluon fusion (a), via vector boson fusion (b), in association with a vector boson ("Higgs-strahlung") (c), in association with a top quark pair (d) or in association with a single top quark in the *t* channel (e) and in association with a W boson, respectively (f).

One example, which will be explained later in this chapter and which will be a crucial aspect throughout this thesis, is the value of the Higgs boson coupling to the top quark.

1.3.1. Production Modes

At the LHC, the Higgs boson is produced by gluons or by valence and sea quarks of the two colliding protons. In Figure 1.3, the Feynman diagrams of the four most important production modes as well as the two production modes investigated in this thesis are shown.

The most dominant production mode is the gluon-gluon fusion. In this channel, the Higgs boson is produced by two gluons. As the Higgs boson can only couple to massive particles and as the gluons are massless, a loop mostly consisting of virtual top quarks is necessary.

In the second production channel, called vector boson fusion (VBF), two quarks each emit a massive vector boson, either $W^{\pm}$ or Z^0. Afterwards, these two vector bosons merge into a Higgs boson.

The production in association with a weak vector boson is the third most frequent channel, which is also referred to as "Higgsstrahlung". In this mode, a quark and an antiquark merge into a weak vector boson. This boson then radiates a Higgs boson.

The fourth most important production mode is the Higgs boson production in association with a top quark pair. In case of the most frequently production mode at the LHC, two gluons produce a top quark pair and either the top quark or the antitop quark emits the Higgs boson afterwards. This process enables the measurement of the absolute value of the Higgs boson coupling to the top quark, y_t.

The production of a Higgs boson in association with a single top quark (tH) is a rare production mode, but crucial for this thesis. In this mode, the single top quark is produced either via the t channel (called tHq) or in association with a W boson (referred to as tHW). In both cases, the Higgs boson can be emitted from the top quark or from the W boson. In contrast to the Higgs boson production with a top quark pair, this production channel is sensitive to the sign of y_t [25]. This is caused by the two possibilities for the emission of the Higgs boson. For tHq and for tHW each, two Feynman diagrams exist that differ by the emission of the Higgs boson. The amplitudes of these two Feynman diagrams interfere and the main term of the resulting amplitude is

$$\mathcal{A} \propto (\kappa_t - \kappa_V) \,, \tag{1.19}$$

where $\kappa_t = y_t / y_t^{SM}$ is the ratio of the actual Higgs boson coupling to the top quark compared to the Standard Model prediction. Analogously, the ratio of the Higgs boson couplings to vector bosons is denoted by $\kappa_V = g_V / g_V^{SM}$.

According to Equation 1.19, the cross sections of the tHq and the tHW process depend on the values for κ_t and κ_V. In case of the Standard Model (i. e. $\kappa_t = \kappa_V = +1$), one receives a destructive interference which leads to a small predicted cross section at a center-of-mass energy of 13 TeV [26, 27]:

$$\sigma_{tHq}^{SM} = 71\,\text{fb} \,, \tag{1.20}$$

$$\sigma_{tHW}^{SM} = 16\,\text{fb} \,. \tag{1.21}$$

If one considers, for instance, the case of an inverted Higgs top coupling param-

eter $\kappa_t = -1$, a constructive interference occurs which leads to a significantly higher cross section for the tH processes [26, 27]:

$$\sigma_{tHq}^{ITC} = 739\,\text{fb}\ , \tag{1.22}$$

$$\sigma_{tHW}^{ITC} = 147\,\text{fb}\ , \tag{1.23}$$

where the acronym ITC denotes the case of the so-called inverted top coupling scenario (i. e. $\kappa_t = -1$, $\kappa_V = +1$).

1.3.2. Decay Modes

Several possible decay channels exist for a Higgs boson with a mass of 125 GeV. Figure 1.4 provides an overview of these decay modes. The most common decay channel is the decay of the Higgs boson into a bottom quark pair. The Higgs boson cannot decay into a top quark pair, although it has the highest coupling strength to the top quark. This decay is kinematically not allowed as the sum of the top quark pair mass is higher than the Higgs boson mass. The decay into a pair of photons, called the "golden channel", as well as the decay into one real and one virtual Z^0 boson, played a crucial role in the discovery of the Higgs boson. Despite the small branching ratio, these two decays provide a signal that can be relatively clearly separated from the background.

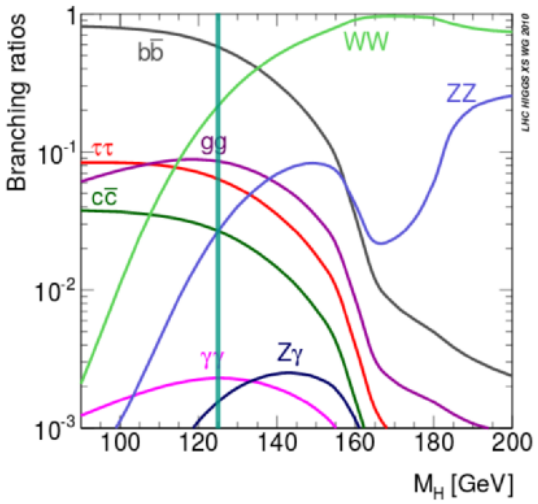


Figure 1.4.: Branching ratio of the Higgs boson. For the Higgs boson mass of $m_{\mathrm{H}} \approx 125\,\mathrm{GeV}$ (marked by the vertical line in the figure), the dominating decay channels are $\mathrm{H} \to \mathrm{b\bar{b}}$, $\mathrm{H} \to \mathrm{WW}^*$ and $\mathrm{H} \to \mathrm{gg}$. The Figure is taken in revised form from [28].

1.4. Top Quark Properties

The top quark is the heaviest of the six quarks and was discovered at the Tevatron in 1995 [29, 30]. Due to its high mass of $173.34 \pm 0.76\,\mathrm{GeV}$ [31], it cannot form bound states. For this thesis, it is essential to know the possible production and decay modes of the top quark as it provides the highest coupling to the Higgs boson and as it can be produced in association with a Higgs boson. In addition, the production of top quark pairs and single top quarks is a non-negligible background process for tH production.

1.4.1. Top Quark Pair Production

The production of a top quark pair via the strong interaction is the dominant production mode for top quarks. Figure 1.5 presents the four relevant Feynman diagrams at leading order for $\mathrm{t\bar{t}}$ production. In the first three Feynman diagrams, two gluons produce the top quark pair, while in the fourth diagram one quark and one antiquark annihilate into a virtual gluon that splits into a top quark and an antitop quark. As the antiquark in the initial state can only originate from sea quarks, this production mode is rare at the LHC. The dominating production mode is therefore the production via two gluons in the initial state. These two gluons can either annihilate into a virtual gluon (s channel) or can

scatter via the exchange of a top quark (t and u channel). The difference between the t and u channel is given by the switched final state.

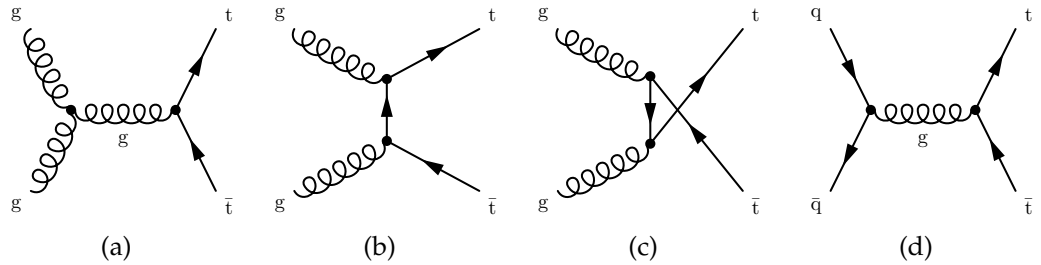


Figure 1.5.: Top quark pair production modes. The top quark pair can be produced via gluon-gluon fusion in the s channel (a), t channel (b) or u channel (c), or via quark-antiquark annihilation (d).

1.4.2. Single Top Quark Production

As single top quarks can only produced via the electroweak interaction, they are more rarely produced than top quark pairs. Nevertheless, the single top production is important for Standard Model measurements as this process allows e. g. the direct measurement of the Cabibbo–Kobayashi–Maskawa (CKM) matrix element V_{tb} which plays an essential role for the branching ratios. In Figure 1.6, the leading order Feynman diagrams are shown.

In the production via t channel, a quark and a bottom quark interchange a virtual W boson, causing both initial quarks to change the flavor. Another possibility for single top quark production is the associated tW production, where a gluon and a b quark produce either a space like top quark or a time like bottom quark. The final state is in both cases given by a top quark and a W boson. In the last production mode via s channel, a quark and an antiquark annihilate into a W boson, which produces a top quark and an antibottom quark. As this production channel requires an antiquark, it is an infrequently observed process at the LHC.

1.4.3. Decay Modes

As the V_{tb} element is given by ≈ 1, the top quark almost always decays into a bottom quark and a W boson. Decays into quarks of another generation are

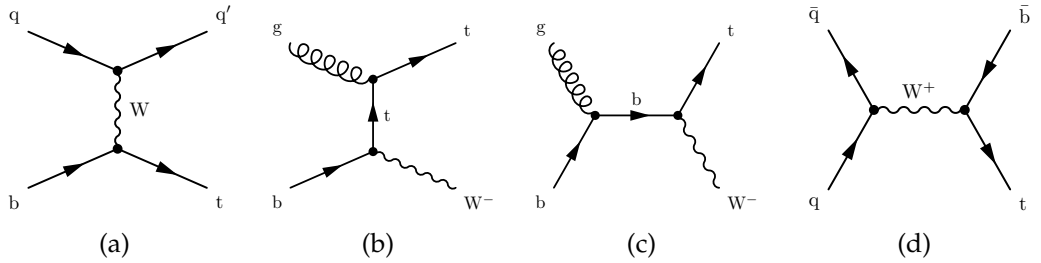


Figure 1.6.: Single top quark production modes. The single top quark can be produced via the *t* channel (a), in association with a W boson (b)(c) or via the *s* channel (d).

hence suppressed. The produced W boson can further decay either into a quark-antiquark pair or into a charged lepton and a neutrino. Therefore, the top quark decay can be classified according to the final state of the W boson.

For the decay of a top quark pair one receives three possible decay modes: full-hadronic, full-leptonic and semi-leptonic decay. In the full-hadronic channel, both W bosons decay into a quark-antiquark pair. Due to confinement, these quarks form boundary states. This process is called hadronization and gives the decay channel its name. If the two W bosons decay into a charged lepton and a neutrino, the process is called full-leptonic decay, while in the semi-leptonic channel, one W boson decays leptonically and the other hadronically.

2. The Compact Muon Solenoid Experiment at the Large Hadron Collider

In contrast to protons and neutrons, which consist of up and down quarks, and electrons, one cannot observe the other particles of the Standard Model in everyday life. In order to study the properties of these fundamental particles, such as the Higgs boson and the top quark, a suitable experimental setup, consisting of an accelerator and a detector system, is needed that provides sufficiently high energies to produce heavy particles.

In this chapter, a brief overview of the Large Hadron Collider as well as of one of its detectors, the Compact Muon Solenoid experiment, is given.

2.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [32, 33] is located at the European Organization for Nuclear Research (CERN[1]) in the area of Geneva, Switzerland. It is the last element of the CERN accelerator complex, which is shown in Figure 2.1. The LHC consists mainly of a ring of accelerating structures and superconducting magnets with a circumference of 27 km and can be operated in two different modes, providing proton-proton collisions, which is the main mode, or heavy ion collisions.

The accelerator tunnel, which is built around 100 m beneath the ground, is made of eight arcs that each contain 154 dipole magnets. These magnets are cooled down with liquid helium to a temperature of 1.9 K, thus they are able to provide a maximum magnetic field of 8.33 T. In addition to the dipole magnets, which bend the two beams, there are also 392 quadrupole magnets that focus

[1] **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire

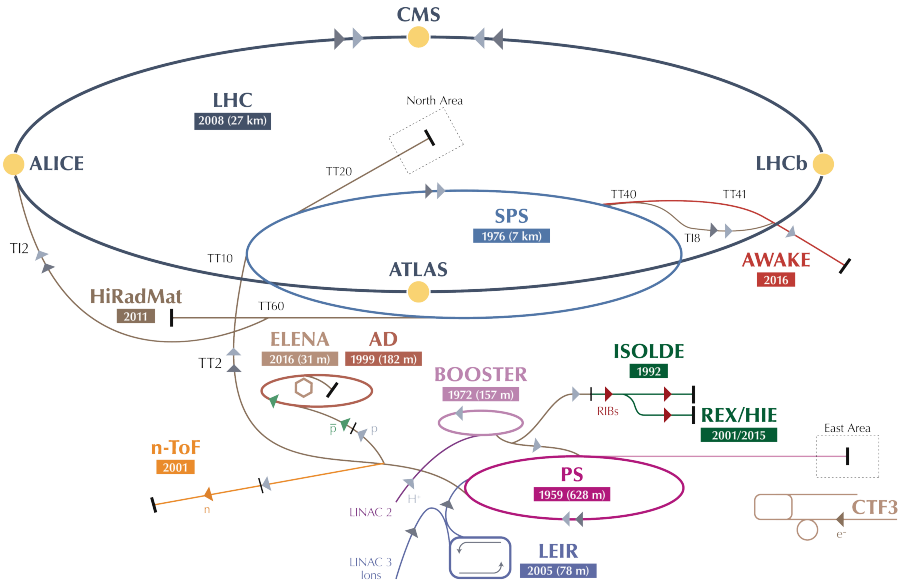


Figure 2.1.: The CERN accelerator complex. The LHC with its four experiments CMS, ATLAS, ALICE and LHCb as well as several pre-accelerators are illustrated. The two beams, which are accelerated in the LHC up to a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$, are injected eventually by the Super Proton Synchrotron (SPS). The figure is taken from [34].

each beam. Furthermore, other magnet types such as sextupoles and octupoles are used, resulting in a total amount of 9593 magnets. In order to accelerate the protons and to reach high collision energies, eight cavities per beam are needed, which are placed between the eight arcs. For a proton-proton collision two beams are accelerated in opposite directions, hence two pipes in total are needed. With this setup it is possible to accelerate two beams that each contain 2808 bunches, which consist of 10^{11} protons per bunch.

The beams are brought into collision at four points in the collider ring. At these points, the four main detectors of the LHC are installed (see also Figure 2.1): **C**ompact **M**uon **S**olenoid (CMS), **A** **T**oroidal **L**HC **A**pparatu**S** (ATLAS) as well as **A** **L**arge **I**on **C**ollider **E**xperiment (ALICE) and **L**arge **H**adron **C**ollider **b**eauty (LHCb). The first two detectors are used for general-purpose experiments, while the latter two are used for special purposes, i. e. for the study of heavy ions and bottom quarks, respectively.

Since the first beam was injected in 2008, the energy has been increased gradually. In the Run I of the LHC, data has been taken between March 2010 and December 2012 at two different center-of-mass energies: $\sqrt{s} = 7\,\text{TeV}$ (2010 –

2011), i. e. 3.5 TeV per proton beam, as well as $\sqrt{s} = 8$ TeV (2012), corresponding to a per-beam energy of 4 TeV. In order to reach the designated center-of-mass energy of 14 TeV, the LHC was shut down between 2013 and 2015 for the implementation of several improvements. In June 2015, the Run II of the LHC has started with a center-of-mass energy of 13 TeV. It is planned that the Run II will accelerate the two beams up to 7 TeV each and will record data until 2018 [35].

A very important measure of the performance of an accelerator is the so-called instantaneous luminosity L. For a collider ring this parameter is defined as follows:

$$L = \frac{n_\mathrm{B} N_a N_b}{4\pi\sigma_x\sigma_y} \cdot f \ , \tag{2.1}$$

where n_B denotes the number of bunches, $N_{a,b}$ the number of protons per bunch, f the revolution frequency of a single bunch and $4\pi\sigma_x\sigma_y$ the area of the bunches perpendicular to the beam axis. The design luminosity of the LHC is $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which has been achieved first in June 2016 [36]. Using L and the total cross section σ of a process, one can determine the number of interactions:

$$N = \int \sigma L \,\mathrm{d}t = \sigma L_\mathrm{int} \ , \tag{2.2}$$

with the integrated luminosity L_int that measures the amount of collected data.

2.2. The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) detector [37] is a multi-purpose detector and is located near Cessy, France. With a length of 21 m, a diameter of 15 m and with a weight of 14 000 t [38], the CMS detector is half as voluminous, but twice as heavy than the competing ATLAS detector. The CMS experiment consists of several sub-detectors (see Figure 2.2) which will be explained in the following sections. The majority of these sub-elements is built within a superconducting solenoid which was initially designed to provide a magnetic field of 4 T. In order to increase the life span of the magnet, the operating field is scaled down to 3.8 T [39].

The conventional coordinate system of the CMS detector is defined as follows: The x-axis points radially inward towards the center of the LHC, the y-axis vertically upwards. The z-axis points along the beam direction so that a right-handed coordinate system is formed. Additionally, two angles can be defined,

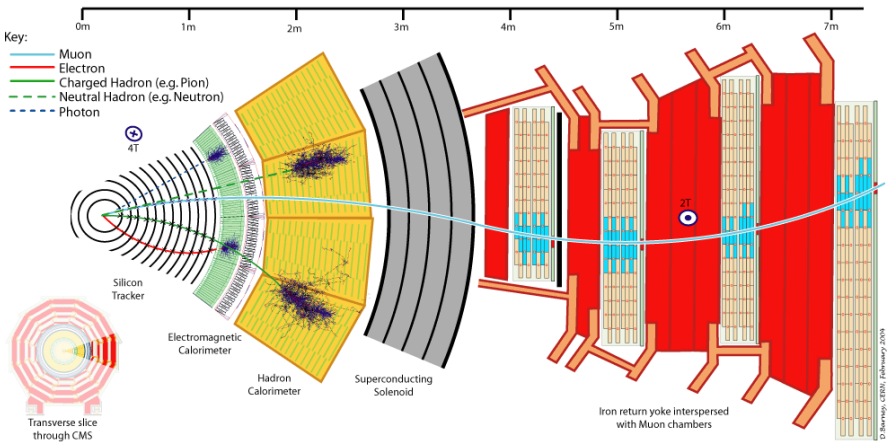


Figure 2.2.: Transverse slice of the CMS detector. The components of the CMS detector are shown. Different particles interact with certain parts of the detector. Except the muon system, all sub-detectors are built in the superconducting solenoid. The figure is taken from [40].

namely the azimuthal angle φ in the x-y plane and the polar angle θ. Another observable, which is frequently used in particle physics, is the rapidity y:

$$y = \frac{1}{2} \ln \left(\frac{E - p_z}{E + p_z} \right) , \tag{2.3}$$

where E corresponds to the energy and p_z to the z-component of the momentum of a particle. As a result of its definition, the rapidity is Lorentz invariant under Lorentz boosts along the z-axis. In order to be only dependent on the angle θ of a particle's momentum vector, one introduces the even more commonly used variable called pseudorapidity η, which is given by

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] . \tag{2.4}$$

For massless particles the rapidity and the pseudorapidity are identical. The possible values for η vary between 0 (i. e. a particle that flies perpendicular to the beam axis) and $+\infty$ ($-\infty$) for a particle that flies alongside the beam axis. This corresponds to an angle ranging from $\theta = \pi/2$ to $\theta = 0$ ($\theta = \pi$). Another important kinematic variable to identify particles is the transverse momentum p_T:

$$p_T = \sqrt{p_x^2 + p_y^2} , \tag{2.5}$$

where $p_{x,y}$ denotes the momentum of a particle along the x-axis and y-axis, respectively. The transverse momentum is Lorentz invariant under Lorentz boosts along the z-axis.

2.2.1. Silicon Tracker

The tracking system [41–43] is built tightly around the collision point of the two beams and measures the tracks of all charged particles which are produced in a collision. Since it is the innermost layer of the CMS detector, it has to be robust against radiation to ensure a long life span. Hence, the whole tracking system is made of semiconducting silicon detector material. Charged particles pass through the silicon modules and create electrons and holes. These electron-hole pairs can be eventually measured as a current. The tracker system itself consists of two layers, namely the inner silicon pixel detector and the outer silicon strip detector. With this setup it is ensured that the hit position of particles are determined within an accuracy of 10 µm.

The pixel detector consists of three barrel-shaped layers with the radii 4 cm, 7 cm and 11 cm as well as two disk layers on each side. These three layers contain in total 65 million pixels, mostly having a size of 100 µm $\times$ 150 µm. As the pixels are distributed in multiple layers, they can provide a three-dimensional spatial resolution of the particle.

After the pixel detector, the particles pass the strip detector, which consists of ten layers and reaches out to a radius of approximately 130 cm. The layers are classified into four inner barrel layers with two inner endcaps and into six outer barrel layers that are closed off by two outer endcaps. In total, this part of the tracker system consists of 15 200 modules containing 10 million detector strips. These modules are highly sensitive and consist not only of sensors, but also of readout electronics. In order to minimize the damage caused by radiation, the strip detector is cooled down to $-20\,^{\circ}\mathrm{C}$.

2.2.2. Electromagnetic Calorimeter

The tracker system is enclosed by the electromagnetic calorimeter (ECAL) [44–46] that measures the energy of all electromagnetically interacting particles, i. e. electrons, photons and neutral pions. The ECAL consists of a barrel section with 61 000 lead tungstate ($PbWO_4$) crystals as well as two endcaps containing 15 000 crystals. $PbWO_4$ crystals can perform as an absorber and as a scintillator at the

same time since they have a high density of $8.3\,\mathrm{g\,cm^{-3}}$ and are transparent due to their short radiation length of 0.9 cm. If a particle traverses the scintillator, it emits photons (scintillation). The intensity of these photons is then roughly proportional to the initial energy of the incoming particle.

Electromagnetically interacting particles pass the ECAL and interact with it via electron-positron pair production, Compton effect or "bremsstrahlung". This results into an electromagnetic shower consisting of lower-energetic particles which are absorbed by the crystals. For a calorimeter of length x the energy of an electron or positron that is left after passing through it is given by

$$E(x) = E_0 \cdot \mathrm{e}^{-\frac{x}{X_0}} \,, \tag{2.6}$$

where E_0 denotes the initial energy of the particle and X_0 the radiation length, which is a characteristic of the material. As the size of the $PbWO_4$ crystals in the barrel is 2.2 cm × 2.2 cm × 23 cm and in the endcap 3 cm × 3 cm × 22 cm, the effective length is given by $\approx 25\,X_0$. This means that every incoming electron, positron and neutral pion should be absorbed entirely by the ECAL. In order to identify electrically uncharged pions that decay into a pair of low-energetic photons, a preshower detector is installed in front of the ECAL.

2.2.3. Hadron Calorimeter

The hadron calorimeter (HCAL) [47, 48] absorbs and measures the energy of all incoming hadrons, i. e. particles that consist of quarks. Additionally, it provides an indirect measurement of neutrinos. In contrast to the ECAL, the HCAL is a sampling calorimeter, as it is made of alternating layers of non-magnetic absorber material (brass) and of plastic scintillators. Since brass has a short nuclear interaction length of 16.4 cm, the HCAL needs to be around 1.2 m thick in order to absorb all hadrons. The HCAL consists of a hadron barrel, hadron endcaps as well as an outer hadron calorimeter and forward calorimeters. This setup ensures that a range of up to $|\eta| = 5$ is covered which enables the measurement of jets in forward direction.

Similarly to the ECAL, the incoming particles cause a hadron shower by interacting with the absorber layers via inelastic scattering. In this process, secondary lower-energetic hadrons are produced. As the number of particles in a hadron shower is typically lower than in an electromagnetic shower, the energy resolution of the HCAL is not as good as of the ECAL.

2.2.4. Muon System

As indicated by its name, the CMS detector has an outstanding muon system [49, 50]. It was specifically constructed to detect important processes, such as one of the most important channels for the Higgs boson discovery, $H \rightarrow ZZ^* \rightarrow 4\mu$. In contrast to most other particles, muons are able to pass through several meters of iron without any interaction since they lose less energy than electrons. This is caused by the fact that the amount of bremsstrahlung is inversely proportional to the particle mass.

The long range of muons explains why the muon system is the outer most part of the CMS detector. It is built within the iron return yoke, where the field is reversed and weakened to a strength of 2 T. The muon system consists of three different types of detectors: 250 muon drift tubes, 540 cathode strip chambers and 610 resistive plate chambers. The drift tubes are located only in the barrel, while the cathode strip chambers form the two endcaps. These two detector types measure the position of muons, whereas the resistive plate chambers, which are located in the barrel and in the endcaps, are used as a fast trigger.

2.2.5. Trigger System

The trigger system [51, 52] plays a crucial role in selecting the physically relevant events from the bulk of the collision events as it is not possible to store all collisions that occur at Run II of the LHC with a rate of 40 MHz. In order to select and to store events of potential physics interest at the CMS experiment, the trigger system consists of two trigger levels, namely the Level-1 trigger (L1) and the High Level Trigger (HLT).

The L1 trigger is able to reduce the data rate to no more than 100 kHz and is implemented in custom hardware. It selects events that contain candidate objects, based on the information from the calorimeters and the muon system. The HLT is implemented in software and further refines the purity of the physics objects. Hence, an average event rate of 1 kHz is selected by the HLT to be stored offline. By prescaling the number of events passing the selection criteria of specific algorithms, it is possible to adjust the output rate of the whole trigger system.

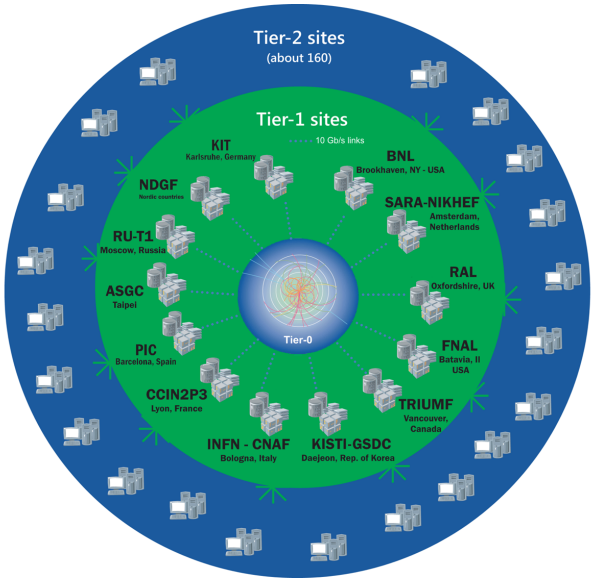


Figure 2.3.: The Worldwide LHC Computing Grid (WLCG). The Tier-1 center at the Karlsruhe Institute of Technology (KIT) in Karlsruhe, GridKa, is the only Tier-1 site in Germany. The figure is taken from [57].

2.2.6. Computing

Despite the drastic reduction of the collision events by the trigger system, the CMS detector still provides a huge amount of event data which needs to be stored, distributed and analyzed. These issues can be solved by employing a distributed data storage and computing infrastructure at the LHC. This infrastructure is called the Worldwide LHC Computing Grid (WLCG) [53–56]. The WLCG consists of four layers in total, namely the tiers 0, 1, 2 and 3. An overview of the WLCG structure is provided in Figure 2.3.

The Tier-0 corresponds to the CERN data center and is responsible for the storage of the original data received from the LHC. In addition, it provides a first processing of the data and sends this reconstruction output to the Tier-1. The Tier-1 consists of 13 large computer centers located around the world. In these computer centers, the events are fully reprocessed and stored. The next layer, the Tier-2, are universities and scientific institutes which provide additional storage space as well as CPU power for different analysis tasks. Tier-3 represents local computing resources that are accessed individually and do not officially belong to the WLCG.

3. Simulation and Reconstruction of Events

For analyses in particle physics it is crucial to compare the experimentally measured data with theory predictions in order to verify or to disprove them. For this purpose events are generated which simulate the signal and the background of a given process. In these simulations, every step of a proton-proton scattering process, from the initial hard collision of the two partons to the decay of the produced particles, and the detector response are modeled.

In the first part of this chapter, the structure of a proton-proton scattering process is explained. The programs used for a complete simulation of events are described in the second section, while the last section focuses on the reconstruction of physics objects from the measured and recorded data.

3.1. Proton-Proton Scattering Process

In Figure 3.1, an overview of the complexity of a proton-proton collision is shown. In order to simulate such a process properly, several generation steps are required, which are described in this section.

Hard Scattering

In proton-proton collisions at the LHC, the partons inside the two protons, carrying an unknown fraction of the initial proton momentum, take part in deep-inelastic scatterings. The probability to find a certain parton with a specific proton momentum fraction x at a momentum transfer μ^2 inside a proton is given by the parton distribution functions (PDFs), as shown in Figure 3.2. Since the evolution of the PDFs cannot be determined by a perturbative calculation, measurements of deep-inelastic experiments such as HERA are needed and the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [59–61] must be applied. The probability of a hard scattering process, which takes place

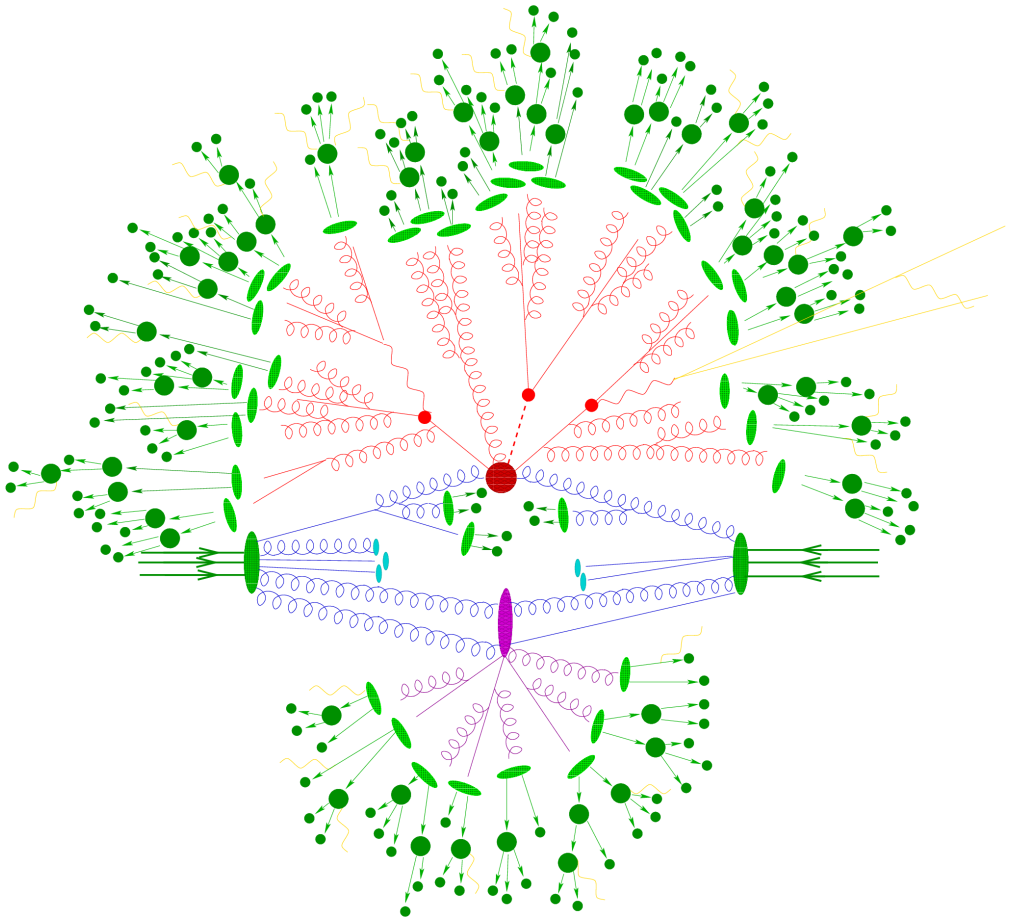


Figure 3.1.: Overview of the event generation process. The simulation of a proton-proton collision is shown schematically. It consists of a hard scattering process (indicated as a red circle in the center), which is surrounded by parton showers that represent gluon radiation, as well as the production and decay of hadrons. The hadronization and the decay of hadrons are illustrated by the light green and dark green structures, respectively, whereas soft photon radiations are marked by yellow lines. A secondary hard scattering process is represented by a purple oval. The figure is taken from [58].

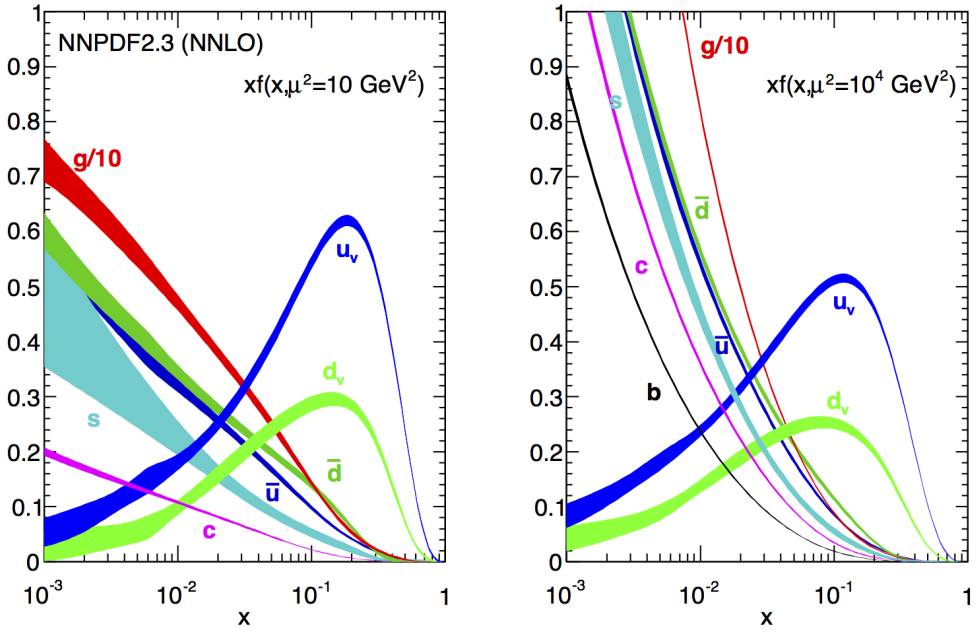


Figure 3.2.: The Neural Network Parton Distribution Function (NNPDF). The parton distribution functions for gluons, valence quarks and sea quarks for two different momentum transfers, namely $\mu^2 = 10\,\mathrm{GeV}^2$ (left) and $\mu^2 = 10^4\,\mathrm{GeV}^2$ (right), are shown. At a lower proton momentum fraction x, the distribution is dominated by gluons. The figure is taken from [63].

at a high momentum transfer, is calculated with the matrix element method (MEM) [62]. In this method, Feynman diagrams and transition amplitudes of a specific process as well as a set of PDFs are used to determine its occurrence probability.

Parton Shower

The partons of the proton are able to emit additional photons or gluons in a way that the conservation of the electric and the color charge is ensured. Since the electromagnetic coupling constant α_{em} is smaller than the strong coupling constant α_{s}, the radiation of a gluon is more likely. One distinguishes between initial state radiation (ISR) and final state radiation (FSR). Furthermore, the gluons can split into a quark-antiquark pair or into two gluons, which is described by the DGLAP equations. Hence, the parton shower, a cascade consisting of secondary partons, is produced.

Hadronization

The partons in the shower have to form particles which carry no color charge, as the confinement of the QCD forbids the observation of color charged particles in nature. This process is called hadronization, and cannot be described by perturbative theories due to the low energy scale. In order to simulate the hadronization process, phenomenological models, e. g. the Lund string model [64], are applied. In this model, all partons are treated as field lines, which are attracted to each other due to the gluon self-interaction and thus form a narrow tube. If the quarks are separated, this tube collapses, which causes the creation of additional quark antiquark pairs (i. e. mesons) and prevents the occurrence of color charged particles. As these mesons have a limited life span, they decay into stable particles afterwards.

Underlying Event and Pileup

In the hard scattering process, only one parton of each proton is considered. Since the other partons of the two protons are able to produce additional particles in the CMS detector, they needed to be simulated as well. This process is called underlying event (UE). Furthermore, it is possible that multiple proton-proton collisions overlap as the protons are accelerated in bunches containing 10^{11} protons each. The so-called in-time pileup (PU) corresponds to multiple collisions of protons in the same bunch, while out-of-time PU describes the overlap of collisions which do not originate from the same bunch crossing.

3.2. Event Generation

In order to generate an event, the physics of the scattering process described in Section 3.1 and the interaction of the produced particles with the sub-elements of the CMS detector need to be simulated. For this purpose Monte Carlo (MC) generators and a dedicated detector simulation software are used, which are shortly described in the following section.

3.2.1. Monte Carlo Generators

In order to simulate interactions observed in high-energy physics, event generators use the Monte Carlo method [65]. For a given process the event generators calculate Feynman diagrams in different orders of the strong coupling constant α_s. If an event is generated in leading order (LO), neither virtual particles in

loops nor additional radiation in the initial or final state exist. Hence, the diagrams that describe this event possess the smallest possible number of vertices. An event is generated in next-to-leading order (NLO) if further diagrams with an additional vertex are included in the calculation, i. e. one virtual loop or one ISR/FSR occurs. All MC generators used in this analysis are shortly introduced in the next paragraphs.

MadGraph5 and MadEvent

MadGraph5 [66] is a matrix element generator written in Python programming language, which calculates all Feynman diagrams in LO for a given process. Furthermore, it generates the mappings for the integration over the phase space. The output of MadGraph5 is then passed to MadEvent [67], which calculates the corresponding cross sections and simulates the events.

MadGraph5_aMC@NLO

MadGraph5_aMC@NLO [68] is an event generator that provides the automated computation of LO and NLO cross sections as well as their matching to parton shower simulations. It combines the features of the two event generators MadGraph5 (LO) and aMC@NLO (NLO). In NLO calculations, events can possess negative weights, which originate from subtraction terms. These subtraction terms prevent the double-counting of radiation processes and thus are dependent on the parton shower generator used afterwards.

POWHEG

Powheg (**PO**sitive **W**eight **H**ardest **E**mission **G**enerator) [69–71] is another MC generator that provides the generation of events at NLO. In contrast to MadGraph5_aMC@NLO, Powheg only calculates events with positive weights as it generates the hardest radiation first. The issue of double-counting is resolved since the later parton shower generators are not allowed to simulate a harder radiation. Hence, no subtraction terms are needed and Powheg is independent of the choice of the parton shower generator. As every single process was implemented by different authors, the user interface and the structure of the corresponding software package can often differ.

PYTHIA

PYTHIA [72, 73] is an event generator, which not only simulates the hard scattering process, but also the parton shower as well as the hadronization, the decays and the underlying events. Since the hard scattering process is only calculated at LO, it is mostly used as a parton shower generator. In this thesis, PYTHIA 8 is used for the simulation of the parton shower and the hadronization. The generated parton shower is p_{T}-ordered and the hadronization algorithm is based on the Lund string model.

3.2.2. Detector Simulation

Since the aforementioned MC generators do not simulate the interactions of the particles with the different sub-detectors of the CMS experiment, a comparison between the simulated events and the measured events is not yet possible. In order to simulate the behavior of the particles in the CMS detector, dedicated software is needed. GEANT4 [74, 75] provides a simulation of electromagnetic, hadronic and optical processes, hence it is possible to simulate the interactions of the particles with all parts of the CMS detector.

3.3. Event Reconstruction

For the reconstruction of events it is crucial to combine and to interpret the measured signals of each sub-detector of the CMS experiment correctly. In the following section, the technical requirements as well as different algorithms are explained, which are essential for a successful identification of various physics objects.

3.3.1. Technical Requirements

The trajectory of a charged particle, which is influenced by the magnetic field, possesses a different curvature and direction, which are directly related to its transverse momentum and its electric charge, respectively. This information can be used to identify charged particles. In order to reconstruct these trajectories, the CMS tracking software, called Combinatorial Track Finder (CTF) [76], is applied. This method uses Kalman filters [77–79] for the pattern recognition and the track fitting.

A primary vertex (PV) corresponds to the point where the hard collision occurs and is the origin of all particles created in this interaction. Due to pileup, usually more than one primary vertex exists in a single event. The primary vertices can be reconstructed via the Adaptive Vertex Fitting (AVF) method [80]. In this method, tracks that are not suitable for a specific vertex candidate are not ignored, but they receive a lower weight. Tracks which are compatible with the vertex candidate are thus assigned with higher weights. This compatibility between a track and a possible vertex is determined by a χ^2 minimization. The PV used for the event reconstruction corresponds to the vertex which possesses the highest weighted p_T^2 sum of its assigned tracks, while every other vertex is considered as pileup.

The total energy stored in the calorimeters is determined with a specific clustering algorithm. The purpose of such an algorithm is the detection and measurement of the energy and direction of stable neutral particles and the separation of these neutral particles from energy deposits caused by charged hadrons. Furthermore, it provides the identification of electrons as well as accompanying bremsstrahlung photons.

3.3.2. The Particle Flow Algorithm

The Particle Flow (PF) algorithm [81] links the information, delivered by the sub-elements of the CMS detector and reconstructed with the algorithms described above, in a way that individual physics objects are reconstructed. The PF algorithm defines the tracks in the silicon detector, the energy clusters in the ECAL and HCAL and the tracks in the muon system as fundamental "elements". These elements are then linked directly or indirectly to "blocks", which usually consist of two or three elements. In a final step, the PF algorithm interprets these blocks as particles, i. e. charged and neutral hadrons, muons, electrons and photons. Hence, an optimal use of the combination of subdetectors is guaranteed for each particle type. An overview of the main principle of the PF algorithm used in the CMS experiment is given in Figure 3.3.

3.3.3. Physics Objects

In this section, a more detailed description of the reconstruction of various physics objects, which are essential for this thesis, is given.

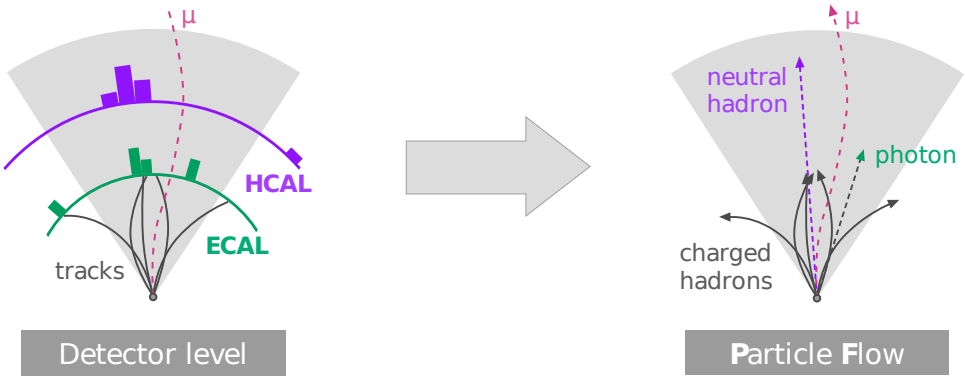


Figure 3.3.: Concept of the Particle Flow algorithm. The signals from the sub-detectors of the CMS experiment, e. g. the HCAL, ECAL and the silicon tracker, are combined and reconstructed via the Particle Flow algorithm to physics objects such as photons, neutral or charged hadrons. The figure is taken in revised form from [82].

Electrons

Due to bremsstrahlung, electrons already lose energy in the silicon tracker, which leads to characteristically short tracks. Hence, it is possible to pre-identify electrons with low transverse momenta using a Gaussian-sum filter (GSF) algorithm [83, 84]. In this algorithm, the energy loss in each layer is approximated by a mixture of Gaussian distributions, which possess different weights. The GSF algorithm fits the collected hits in the silicon detector and in the ECAL to estimate the track parameters and to reconstruct the trajectory of an electron. In case of an electron with a high transverse momentum of $p_{\mathrm{T}} > 5\,\mathrm{GeV}$, it is reconstructed by using ECAL-based information. In order to collect the bremsstrahlung energy, so-called "superclusters" [85] (i. e. clusters of clusters) are built, which consist of energy deposits that are located in a narrow area of the ECAL. The identified tracks and energy deposits eventually form the so-called "particle-flow electrons".

Muons

Since muons are significantly more massive than electrons, the energy loss due to bremsstrahlung is mostly negligible. Hence, muons are the only particles that cause hits in the muon system, as they are able to travel through the calorimeters. In order to reconstruct muons successfully, three possible reconstruction modes [86] exist. If only the track in the muon system is reconstructed, one receives a so-called "standalone muon". The other two approaches use addi-

tionally the reconstructed track of the inner silicon detector system. In case of a "global muon", the track of the muon system is extrapolated to the corresponding hits in the inner tracker, while for a "tracker muon" the extrapolation is operated the other way round.

Jets

A jet is defined as a narrow cone which consists of a cascade of hadrons. These hadrons are produced by gluons and quarks during the hadronization process (see also Section 3.1). As some hadrons are allowed to decay leptonically, the jets can additionally contain non-isolated photons and leptons. By combining and clustering hits and tracks in the sub-detectors, the origin of a jet, i. e. the initial parton, can be reconstructed.

In this thesis, a sequential jet clustering algorithm, namely the anti-k_T algorithm [87], is used for the reconstruction of jets. This algorithm is based on the distance d_{ij} between two particles and the distance $d_{i\mathrm{B}}$ between a particle and the beam, which are given by

$$d_{ij} = \min\left(k_{\mathrm{T},i}^{-2}, k_{\mathrm{T},j}^{-2}\right) \frac{\Delta R_{ij}^2}{R^2} \,, \tag{3.1}$$

$$d_{i\mathrm{B}} = k_{\mathrm{T},i}^{-2} \,. \tag{3.2}$$

Here, k_T denotes the transverse momentum of a particle and the distance of the two particles in the $\varphi - \eta$ plane is defined as $\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\varphi_i - \varphi_j)^2}$. The radius parameter R is referred to as cone radius and is usually set to 0.4 for analyses with the CMS experiment. The two distances d_{ij} and $d_{i\mathrm{B}}$ are calculated iteratively: the two particles with the smallest distance $d_{ij} < d_{i\mathrm{B}}$ are grouped into a new object and all distances are then recalculated. If the smallest distance found by the algorithm corresponds to $d_{i\mathrm{B}}$, the object is removed from the particle set and is referred to as a jet. The whole procedure is repeated until every particle is clustered into a jet. In Figure 3.4, an example of the jet reconstruction via the anti-k_T algorithm is illustrated.

In order to account for known detector effects and pileup interactions, the measured jet energy is corrected in data and in simulated events. The jet energy corrections (JEC) [88, 89] are divided into different levels. The level 1 (L1) correction is applied first and removes the energy which originates from pileup events.

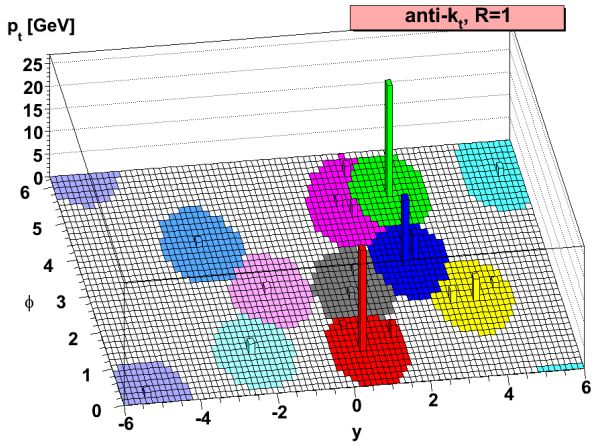


Figure 3.4.: The anti-k_T jet clustering algorithm. In the diagram, a flat projection of the HCAL in the $\varphi - y$ plane with energy deposits illustrated as towers is shown. The anti-k_T algorithm behaves like an idealized cone algorithm which successfully resolves overlapping jets. The figure is taken from [87].

These pileup offset corrections are determined from the simulation of a sample of QCD dijet events processed with and without pileup overlay. The L2L3 MC-truth corrections, i. e. the simulated jet response corrections, are determined by comparing the reconstructed p_T to the particle-level p_T^{gen} distribution. These corrections are given as a function of p_T and η. By applying the L2L3Residuals on data, last remaining differences with regard to the jet response in MC and data are corrected. The L2Residuals provide an η-dependent correction, while the L3Residuals correct the absolute jet energy scale. Another level, the L5 jet-flavor correction, is optional and not applied in this thesis.

Identification of b Jets

Since the decays of the Higgs boson into a bottom quark pair are analyzed in this thesis, it is essential to identify jets which originate from bottom quarks. The targeted identification of these b jets is referred to as "b tagging". The majority of the b tagging algorithms is based on the long life time of B hadrons which are present in b jets. This long life time is caused by the strongly suppressed decays of the bottom quark into an up quark or charm quark due to the values of the CKM matrix elements V_{ub} and V_{cb}. Furthermore, a decay of the bottom quark inside the same generation is kinematically not allowed, as the mass of the top quark is significantly higher. Hence, the B hadron travels a distance of a few millimeters before it decays into different particles. An ex-

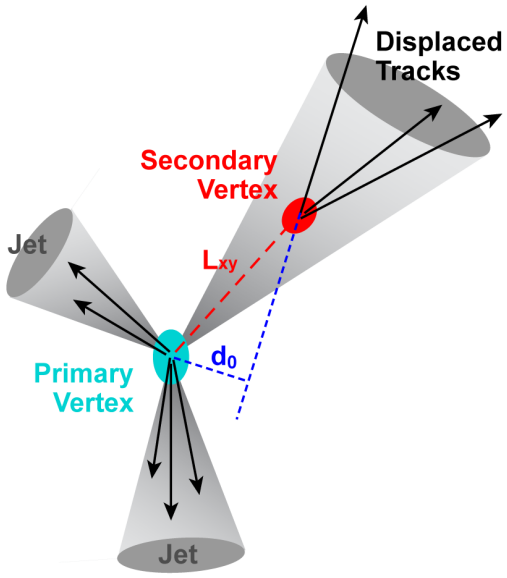


Figure 3.5.: Identification of b jets. A B hadron is produced in the primary vertex and travels, due to its life span, a distance L_{xy} before it decays into different particles. The tracks of these particles intersect at the secondary vertex, which is a prominent characteristic of b jets. The figure is taken from [90].

ample of such a process is shown in Figure 3.5. The B hadron decay creates a secondary vertex (SV), which is shifted from the primary vertex of the hard interaction.

In this thesis, an updated version of the Combined Secondary Vertex (CSV) algorithm [91] is used for the identification of b jets in Run II, which is referred to as CSVv2 [92]. The CSVv2 algorithm combines the information of displaced tracks with the information of secondary vertices that are associated to the jet by using a neural network. For the reconstruction of the secondary vertices the Inclusive Vertex Finder (IVF) algorithm [93] is applied.

Missing Transverse Energy

Since the two protons of the initial collision do not possess any transverse momenta, the vectorial sum of the transverse momenta of all particles in the final state should be equal to zero according to momentum conservation. However, it is possible that the measured sum does not vanish, as some particles interact only via the weak force and thus do not interact with the detector. This imbal-

ance in the transverse momentum measurement is called missing transverse energy (MET) and is given by

$$\not{E}_{\mathrm{T}} = \left| -\sum_{i=1}^{n} \vec{p}_{\mathrm{T},i} \right| , \tag{3.3}$$

where $\vec{p}_{\mathrm{T},i}$ corresponds to the transverse momentum vector of the reconstructed particle i and n denotes the total number of reconstructed particles. Since the neutrinos are the only particles of the SM that leave no hits in the detector, they can be observed indirectly via the measurement of MET.

4. Statistical Methods

For experiments in high energy physics, statistical methods are crucial for an efficient analysis of the measured data. Nowadays, the majority of these methods are implemented in appropriate software tools, which are presented in this chapter.

In the first part of this chapter, a method for parameter estimation, the maximum likelihood estimation, is introduced. A description of the methods used for the limit-setting procedure, the full CL_s and the asymptotic limit calculation, is given in the second and third section. In the last section, a multivariate method that is used for the classification of data sets, namely the boosted decision tree, is explained.

4.1. Maximum Likelihood Estimation

Due to the uncertainty of the measurement, it is essential to find the optimal method to successfully estimate the parameters and thus obtain the most precise results. For this purpose, the maximum likelihood estimation (MLE) method is applied. The following description is based on Reference [94].

For n given measured values $\vec{x} = \{x_1, x_2, \dots x_n\}$ a set of m parameters $\vec{a} = \{a_1, a_2, \dots a_m\}$ needs to be determined by calculating the likelihood function which is defined as

$$L(\vec{a}) = \prod_{i=1}^{n} f(x_i|\vec{a}) \ , \tag{4.1}$$

where $f(x_i|\vec{a})$ is the known probability density function for a measured value x_i. In order to achieve the best parameters $\hat{\vec{a}}$, the likelihood function needs to be maximized, i. e. $L(\hat{\vec{a}}) \geq L(\vec{a})$ for all possible parameter sets $\vec{a}$. Due to the large amount of multiplications, it is technically beneficial to transform the likelihood function with the negative natural logarithm, which leads to a

function containing sums instead of products:

$$-\ln L(\vec{a}) = -\sum_{i=1}^{n} \ln f(x_i|\vec{a}) \ . \tag{4.2}$$

The best parameters $\hat{\vec{a}}$ are thus obtained if Equation 4.2 fulfills two conditions:

$$\frac{\partial(-\ln L)}{\partial a_j} \stackrel{!}{=} 0 \ , \tag{4.3}$$

$$\left.\frac{\partial^2(-\ln L)}{\partial a_i \partial a_j}\right|_{\vec{a}=\hat{\vec{a}}} \text{ negative definite} \ . \tag{4.4}$$

Since binned distributions are analyzed in this thesis, the likelihood function is given by the product of Poisson probabilities for each bin i:

$$L(\vec{a}) = \prod_{i=1}^{m} P(n_i|\nu_i) = \prod_{i=1}^{m} \frac{\nu_i^{n_i}}{n_i!} \cdot \mathrm{e}^{-\nu_i} \ , \tag{4.5}$$

where m corresponds to the total number of bins and where the probability density function $f(x|\vec{a})$ is replaced by the mean value in the bin ν_i, which is defined as

$$\nu_i = n \cdot \int_{x_{i-1}}^{x_i} f(x|\vec{a}) \, \mathrm{d}x \ . \tag{4.6}$$

Here, n denotes the sample size.

In addition to the parameters of interest, $\vec{a}$, another set of parameters exists which describes rate and shape uncertainties. These parameters are called nuisance parameters $\vec{\lambda}$ and need to be considered in the likelihood function. Hence, a function of the form $L(\vec{a}, \vec{\lambda})$ is obtained.

In this thesis, the MLE method is performed with the COMBINE software [95], which is based on the ROOSTATS toolkit [96].

4.2. Full CL_s Exclusion Limits

In contrast to the measurement of established signal cross sections, one needs an appropriate limit-setting technique, the CL_s method [97–99], for the discovery

or exclusion of new physical processes. In this method, two different hypotheses are defined: the null hypothesis, namely the background hypothesis "b", and the signal+background hypothesis "s+b". This alternative hypothesis is favored if the null hypothesis is rejected to a sufficient degree.

In order to distinguish properly between the two hypotheses, a suitable test statistic q_μ needs to be defined which depends on the signal strength modifier $\mu = \sigma_{\text{obs}}/\sigma_{\text{exp}}$. According to the Neyman-Pearson lemma [100], the most powerful test statistic is given by the profile likelihood ratio. Therefore, the test statistic used in the CL_s method is defined as

$$q_\mu = -2 \ln \left(\frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})} \right) \ . \tag{4.7}$$

Here, $\hat{\hat{\theta}}$ corresponds to the set of nuisance parameters that maximizes the likelihood function L for a given μ, whereas $\hat{\mu}$ and $\hat{\theta}$ are the best values for the simultaneous maximization of L with regard to μ and θ.

By performing Monte Carlo toy experiments, the probability density functions of the two hypotheses, i. e. $f(q_\mu|\mu_{\text{b}})$ and $f(q_\mu|\mu_{\text{s}} + \mu_{\text{b}})$, are obtained. This information is then used to calculate the CL_s limit for an observed value q_μ^{obs} (see also Figure 4.1):

$$\text{CL}_{\text{s}} = \frac{p_{\text{s+b}}}{p_{\text{b}}} = \frac{\text{CL}_{\text{s+b}}}{1 - \text{CL}_{\text{b}}} \ , \tag{4.8}$$

with the so-called p-values

$$\text{CL}_{\text{s+b}} := p_{\text{s+b}} = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|\mu_{\text{s}} + \mu_{\text{b}}) \, \text{d}q_\mu \ , \tag{4.9}$$

$$\text{CL}_{\text{b}} := 1 - p_{\text{b}} = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|\mu_{\text{b}}) \, \text{d}q_\mu \ . \tag{4.10}$$

The first p-value, $p_{\text{s+b}}$, denotes the probability for the observation of $q_\mu \geq q_\mu^{\text{obs}}$ if a signal process exists, while $1 - p_{\text{b}}$ corresponds to the same probability in case of a background-only process.

For a pre-defined significance level α, the signal+background hypothesis is rejected, if $\text{CL}_{\text{s}} \leq \alpha$ is fulfilled. For measurements in high energy physics, the

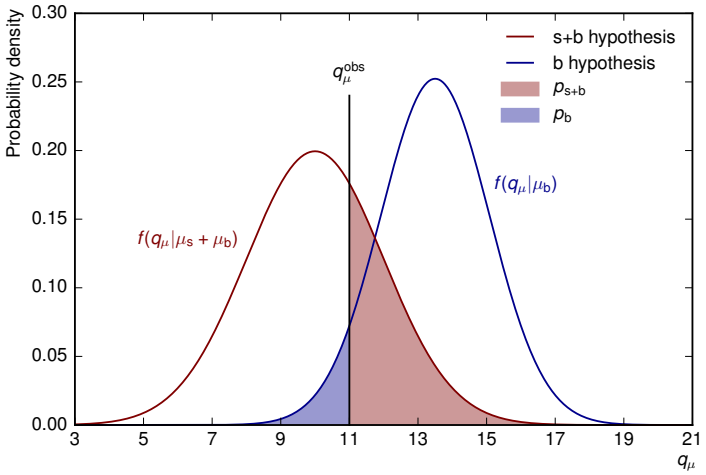


Figure 4.1.: The CL_s method for exclusion limit construction. The probability density functions of the two test statistics for the signal+background (s+b) hypothesis $f(q_\mu|\mu_s+\mu_b)$ and for the background-only (b) hypothesis $f(q_\mu|\mu_b)$ are shown. q_μ^{obs} corresponds to the observed value. The red area is defined as p_{s+b} and the blue area is referred to as p_b. The CL_s limit is thus determined by the ratio of these two areas.

significance level is usually chosen to be $\alpha = 0.05$, which corresponds to a confidence level (CL) of $(1-\alpha)\cdot 100\% = 95\%$.

4.3. Asymptotic Limit Calculation

Since the calculation of full CL_s limits may require a lot of computation power, especially in case of a large number of nuisance parameters, a faster and approximate method is employed instead in this thesis. In this method, explicit formulae for the asymptotic distributions of test statistics [101] are used which are based on the Wilks theorem [102]. Instead of the complete set of simulated data obtained with Monte Carlo toy experiments, only one representative data set, the so-called Asimov data set, is used for the asymptotic limit calculation. In case of large data samples, the results of the approximate method are in conformity with the full CL_s method. In this thesis, the calculation of asymptotic limits is performed with the COMBINE package.

4.4. Boosted Decision Trees

For the analysis of a specific process, an efficient binary classification of the measured data into a signal-like or background-like event is essential. Multivariate methods, e. g. boosted decision trees (BDTs) [103–105], are used to classify each event based on a given set of discriminating variables $\vec{x}$. In contrast to a simple cut-based classification method, the discriminating variables are combined into one final discriminator which allows a better separation than a consecutive application of various cuts on variables. Furthermore, correlations between the discriminating variables can become visible by using BDTs. In order to use such BDTs, the program "**T**oolkit for **M**ulti**V**ariate Data **A**nalysis with ROOT" (TMVA) [106] has been employed in this thesis.

Decision Trees

The concept of a single decision tree is illustrated in Figure 4.2. Consecutive "yes" or "no" decisions are taken on one single variable each, leading to multiple nodes. By fulfilling a stopping criterion, the events that end up in the final leaves are eventually classified as signal or background, depending on the majority of the training events which are located in the corresponding leaf.

The typically used stopping criterion D_{crit}, called the critical separation gain, is defined via the separation gain D which depends on the purity of the leaves as well as on the Gini index, a measure of statistical dispersion. The purity of

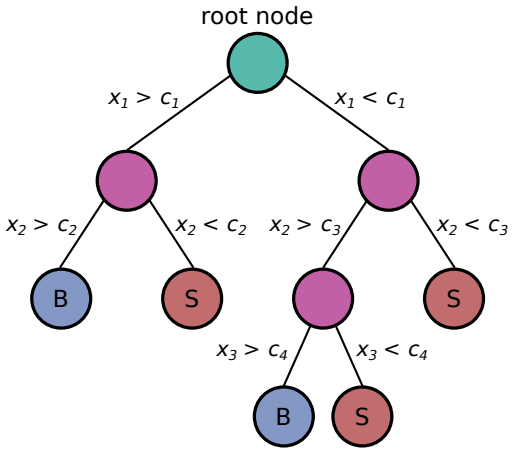


Figure 4.2.: Concept of a classification decision tree. Starting from the root node, cuts using the discriminating variables $\vec{x} = \{x_1, x_2, x_3, \dots\}$ are successively applied. This leads to a tree-like structure consisting of branches and nodes. The leaves at the end of the tree are labeled "S" for signal and "B" for background depending on the event class that represents the majority in the corresponding leaf.

each leaf is defined as

$$P = \frac{\sum_{i=1}^{n_\mathrm{s}} w_i}{\sum_{i=1}^{n_\mathrm{s}} w_i + \sum_{i=1}^{n_\mathrm{b}} w_i} \,, \tag{4.11}$$

where $\sum_{i=1}^{n_{\mathrm{s,b}}} w_i$ denotes the weight sum of the signal-like and background-like events, respectively. In the first run, all training inputs are assigned with $w_i = 1$, while in the following runs, each input gets a weight that is determined by a certain boosting algorithm, which will be explained later. The purity P is then used to calculate the Gini index which is given by

$$G = 4 \cdot P(1 - P) \,, \tag{4.12}$$

The maximal value of the Gini index, $G = 1$, is achieved for $P = 0.5$ which corresponds to a best separation of the events into signal and background. Using the Gini index of the parent node and its two daughter nodes, one receives the separation gain D, determined as

$$D = G_{\text{parent node}} - G_{\text{daughter node 1}} - G_{\text{daughter node 2}} \,. \tag{4.13}$$

If the separation gain is smaller than a pre-defined stopping criterion, i. e. $D < D_{\text{crit}}$, the training is terminated and the leaves are labeled as signal or background, depending on their purity. Alternative stopping criteria are for instance the maximum tree depth or the maximum number of leaves.

Boosting

In order to increase the separation power, not only one, but $M = 100 - 1000$ decision trees are used for the event classification. This "forest" of decision trees, called boosted decision tree, receives weights w_i for wrongly assigned events in a way that these events are taken more into account in the next iteration of the training. In this thesis, these weights are determined by the AdaBoost algorithm [107]. In this algorithm, the subsequent tree m receives a weight $\alpha_m > 1$, which is derived from the misclassification rate ε_m of the previous tree and is given by

$$\alpha_m = \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right) \ . \tag{4.14}$$

This weight is used to reweight the events of the corresponding tree:

$$w_i \to w_i \cdot \mathrm{e}^{\alpha_m S_m(\vec{x})} \ , \tag{4.15}$$

with $S_m(\vec{x}) = 0 \ (= 1)$ for wrong (correct) assignments. Hence, only wrongly assigned events are reweighted. For a set of discriminating variables $\vec{x}$, the final output of a BDT consisting of M decision trees is then defined as

$$y_{\text{boost}}(\vec{x}) = \frac{1}{M}\sum_{m=1}^{M} \alpha_m^{\beta} \cdot T_m(\vec{x}) \ , \tag{4.16}$$

where β is a parameter that determines the strength of the boosting. $T_m(\vec{x})$ denotes the binary output of tree m and is defined as $+1$ for signal leaves and -1 for background leaves, respectively.

Overtraining

Due to suboptimal parameter settings, the BDTs tend to learn statistical fluctuations of the training data set. This effect is called overtraining. If the output of an overtrained BDT is applied to an independent test data set, a worse separation power is observed since the learned fluctuations are no more present. Therefore, a similar performance in the training and in the test data set is desired. In order

to avoid a bias in the analysis, the trained and tested BDT output is applied to a third sample, the evaluation data set. This data set contains events that have never been used for the BDT training.

5. Search for the Associated Production of Single Top Quarks and Higgs Bosons in the H $\rightarrow$ b$\bar{\text{b}}$ Channel

In this chapter, the search for the associated production of single top quarks and Higgs bosons, which decay into bottom quark pairs, in dependency of the Higgs boson couplings is described. This thesis is based on the published analysis performed with the CMS experiment using the data set recorded in 2015 at a center-of-mass-energy of $\sqrt{s} = 13\,\text{TeV}$, corresponding to an integrated luminosity of $2.3\,\text{fb}^{-1}$ [3, 108].

In the first two sections, the analysis strategy and the topology of the signal and background processes are described. Section 5.3 focuses on the identification and reconstruction of physics objects. The fourth section describes the event selection and the fifth section gives an overview of the corrections applied for simulated events and for data. The event interpretation under the signal and under the background hypothesis and the event classification are covered in Section 5.6 and 5.7. The results, consisting of the systematic uncertainties and the exclusion limits, are presented in Section 5.8 and 5.9. The last part of this chapter focuses on an additional flavor classification of the dominating t$\bar{\text{t}}$ background, which has been performed exclusively for this thesis in order to further improve the published analysis. The last section presents the projection of the analysis to higher luminosities.

5.1. Search Strategy

Since the decay of a Higgs boson into a bottom quark pair is given by a branching ratio of 58.2 % [109], it is the most likely decay mode. Hence, this process can be used to analyze the Higgs boson couplings in the associated production

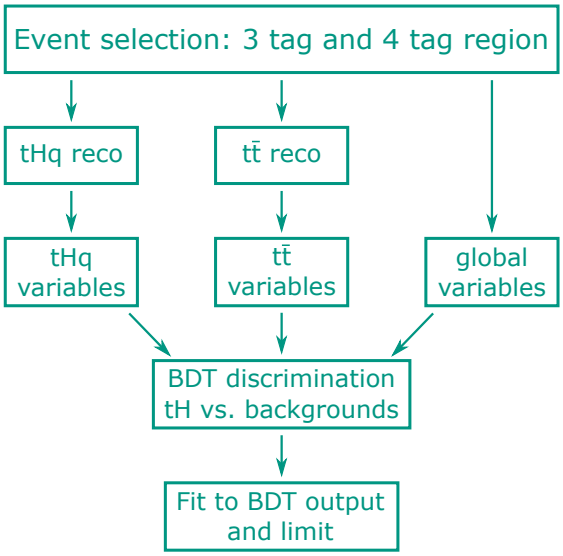


Figure 5.1.: Scheme of the analysis workflow. After the event selection into the two signal regions, the reconstruction under two hypotheses, namely the tHq and $t\bar{t}$ hypothesis, is performed by using BDTs. The reconstructed variables and the global variables, which are independent of any reconstruction, are then used for the training of an event classification BDT. In order to receive an exclusion limit, a fit to the event classification BDT output is applied.

of a single top quark with a Higgs boson (tH). However, the cross section for tH production, i. e. the tHq and tHW process, is very small compared to the cross sections of the background contributions. Therefore, a more complex analysis strategy, illustrated in Figure 5.1, is required in order to determine exclusion limits for this process. By using boosted decision trees (BDTs), each event is reconstructed under the tHq hypothesis and under the $t\bar{t}$ hypothesis as the production of top quark pairs is the dominating background. Variables that describe objects, such as the Higgs boson or the top quark, are obtained from these two reconstruction procedures and are used together with reconstruction-independent variables for the training of the final event classification BDT. The exclusion limit is then received by performing a fit to the event classification BDT output.

5.2. Event Topology

According to the signature of the events, one can discriminate between signal and background processes. For this distinction, it is essential to study the characteristics of the two signal processes and of the main background processes.

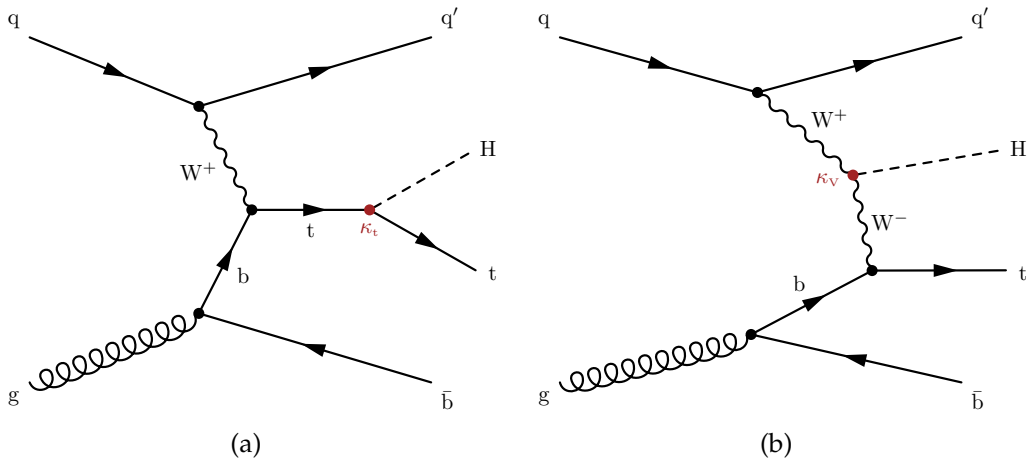


Figure 5.2.: The tHq signal process. For the production of a single top quark in association with a Higgs boson (tHq), two Feynman diagrams exist at leading order. The single top quark is produced via the *t* channel. The Higgs boson can be radiated either from the top quark (a) or from the virtual W boson (b).

In this section, an overview of the involved processes is given.

5.2.1. Signal Processes

Two signal processes exist for the associated production of a single top quark with an additional Higgs boson. They are distinguished by the production mode of the top quark. In case of the tHq process (Figure 5.2), the single top quark is produced via the *t* channel, while in the tHW process (Figure 5.3) it is produced in association with a W boson which is referred to as tW channel. In both signal processes the Higgs boson can be either radiated from the top quark or from the virtual W boson, which is described by the two coupling parameters κ_t and κ_V, respectively.

Higgs Boson Couplings to Top Quarks and to Vector Bosons

As already described in Section 1.3.1, the cross sections of the tHq and tHW process depend on the difference between κ_t and κ_V, which is illustrated in Figure 5.4. For specific $\kappa_t - \kappa_V$ configurations, a significantly higher cross section than for the SM case ($\kappa_t = \kappa_V = +1$), given by $\sigma^{SM}_{tHq} = 71\,\text{fb}$ and $\sigma^{SM}_{tHW} = 16\,\text{fb}$, is expected. Furthermore, the kinematics of the tH processes are dependent on the $\kappa_t - \kappa_V$ configuration, leading for instance to different p_T- and η-distributions of the Higgs boson (see also Appendix A.1). In order to analyze the Higgs boson

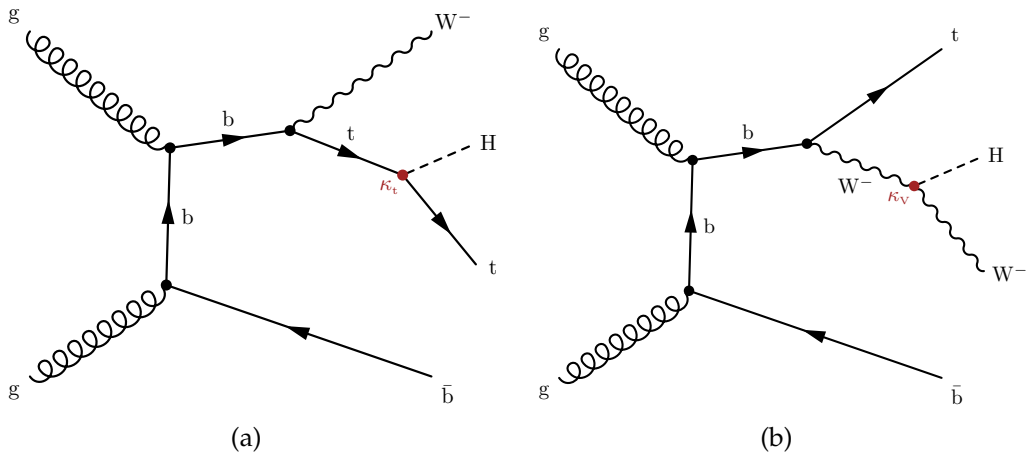


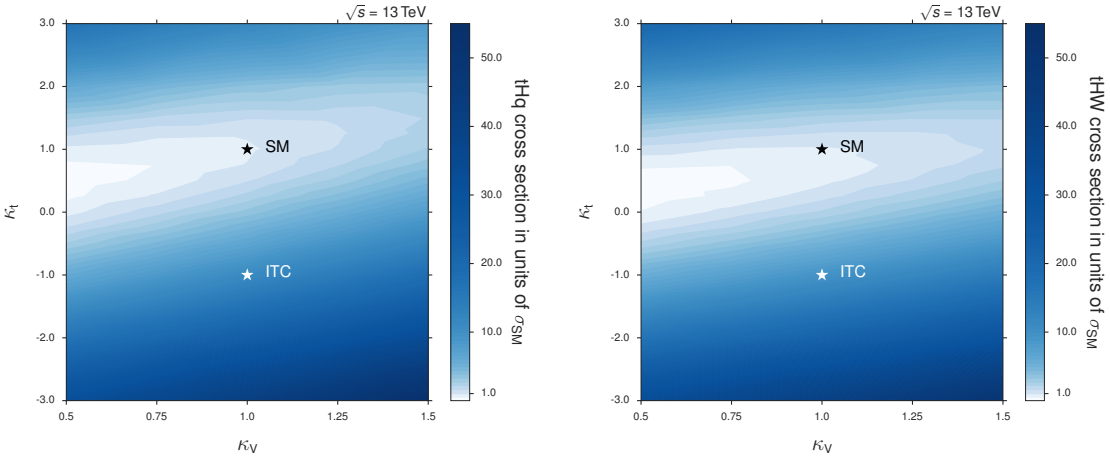
Figure 5.3.: The tHW signal process. For the production of a single top quark and a W boson in association with a Higgs boson (tHW), two Feynman diagrams exist at leading order. The single top quark is produced in association with a W boson, which is referred to as tW channel. Analogously to the tHq process, the Higgs boson can be radiated either from the top quark (a) or from the virtual W boson (b).

couplings to top quarks and to vector bosons, 51 different points in the $\kappa_t - \kappa_V$ plane, defined by $-3 \leq \kappa_t \leq +3$ and $\kappa_V = \{+0.5, +1, +1.5\}$, are chosen to be analyzed as described in Section 5.1.

In most analyses performed by the CMS and ATLAS collaborations in Run I (see also Figure 5.5), the measurements prefer the Higgs coupling parameters of the SM case. Nevertheless, an inverted Higgs-top coupling with regard to the SM, i. e. $\kappa_t = -1$, has not been excluded by the Higgs boson decay channels $H \to ZZ^*$, $H \to \tau^+\tau^-$, $H \to W^+W^-$ and $H \to b\bar{b}$. This scenario is called the inverted top coupling (ITC) case and provides an approximately ten times larger cross section than the SM case, $\sigma^{ITC}_{tHq} = 739\,\text{fb}$ and $\sigma^{ITC}_{tHW} = 147\,\text{fb}$. Due to their importance, this thesis is focused on the ITC and SM case.

Signature of the tHq Process

Since every event is reconstructed under the tHq hypothesis in order to discriminate between signal and background processes, it is crucial to study the final state of the tHq production. As illustrated in Figure 5.6, the signature of this process is given by four bottom quarks, one charged lepton and one neutrino and by one light-flavored quark, which is usually emitted in forward direction. One of the four bottom quarks originates from gluon splitting, which



(a) (b)

Figure 5.4.: Signal cross sections in the $\kappa_t - \kappa_V$ plane. The cross sections, normalized to the expected Standard Model cross section σ_{SM}, of the tHq (a) and the tHW signal process (b) are illustrated in dependency of the $\kappa_t - \kappa_V$ configuration. In both diagrams, two different $\kappa_t - \kappa_V$ configurations, namely the Standard Model (SM) case and the inverted top coupling (ITC) case, are highlighted. The cross section values of all $\kappa_t - \kappa_V$ configurations can be found in Appendix A.1 and A.2.

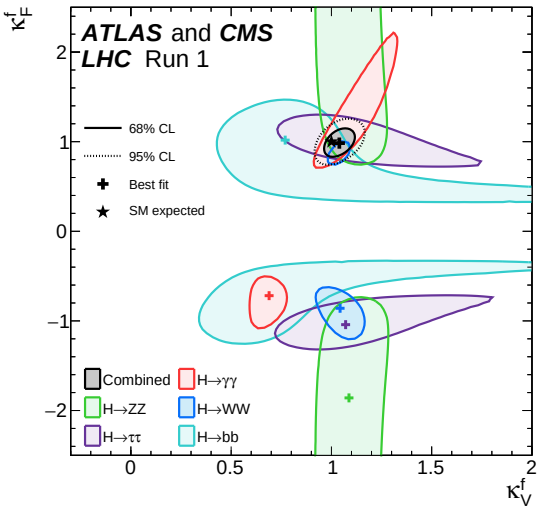


Figure 5.5.: Measurements of the Higgs boson couplings. The combined results of the CMS and ATLAS experiment at Run I for different Higgs boson decay modes are illustrated. The combination of all five decay channels prefers the SM case, i. e. $\kappa_t = \kappa_V = +1$. Nevertheless, four of five channels do not exclude the ITC case with an inverted Higgs top coupling parameter of $\kappa_t = -1$. The figure is taken from [110].

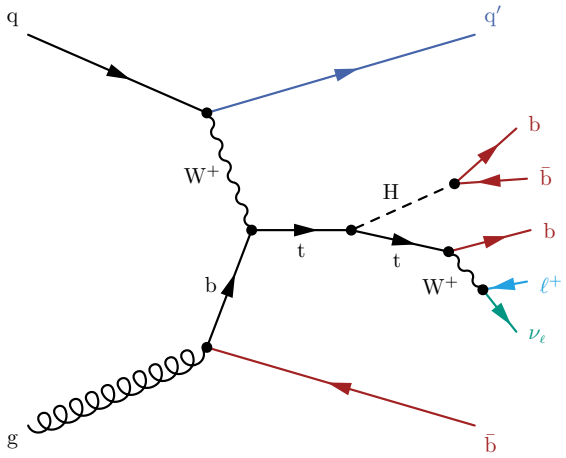


Figure 5.6.: Decay of the final state particles of the tHq process. A representative Feynman diagram at leading order of the tHq signal process which includes the decay of the final-state particles is illustrated. The signal signature is given by one light-flavored quark in forward direction (dark blue), four bottom quarks (red), one charged lepton (light blue) and one neutrino (green).

is required to provide a bottom quark in the initial state of the tHq process, and often remains undetected due to its rather low transverse momentum. In this case, only three bottom quarks can be reconstructed as a b jet. The charged lepton and the neutrino originate from the leptonically decaying W boson which is produced by the decay of the top quark.

The signature of the tHW process is similar to the tHq production as it also consists of four bottom quarks, one charged lepton and one neutrino. However, two W bosons instead of only one occur in the final state of the tHW production. One of the W bosons is required to decay hadronically, resulting in two light-flavored quarks that are not necessarily emitted in forward direction. Hence, the tHW process resembles semi-leptonic $t\bar{t}$ production.

5.2.2. Background Processes

The background consists of Standard Model processes which possess a signature similar to one of the two signal processes. In this section, all background processes that are considered in this thesis are described. The Feynman diagrams at leading order of the main background processes are illustrated in Figure 5.7.

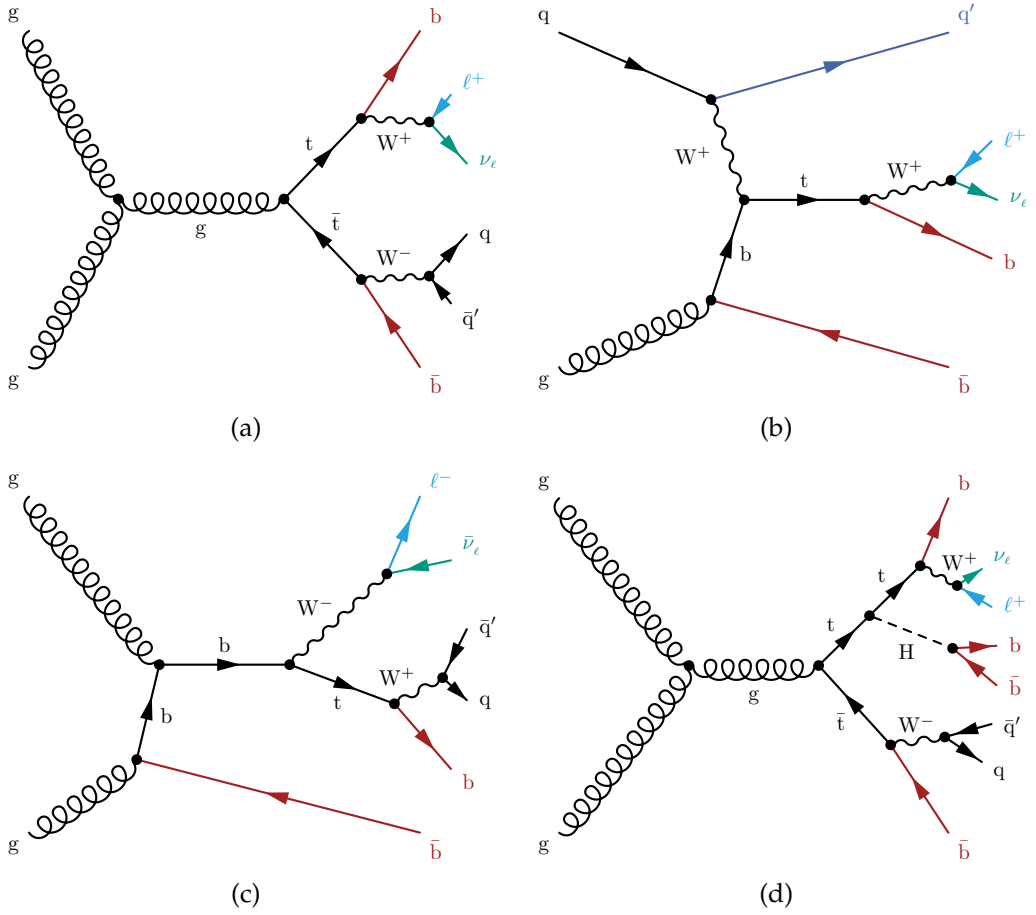


Figure 5.7.: The dominating background processes. Examples of Feynman diagrams at leading order of the main background processes: semi-leptonic $t\bar{t}$ production (a), production of single top quarks in the *t* channel (b) and in association with a W boson (c), respectively and $t\bar{t}H$ production with the decay $H \to b\bar{b}$ (d), are illustrated.

Top Quark Pair Production

Top quark pair production is the dominating background of this analysis. The highest contribution is given by semi-leptonic $t\bar{t}$ decays, illustrated in Figure 5.7(a). In comparison to the tHq signal process, a light quark in forward direction and two bottom quarks are missing in the signature of the semi-leptonic $t\bar{t}$ process. Nevertheless, two additional bottom quarks can occur due to initial or final state radiation or due to misidentification of the light quarks that result from the hadronic decay of the W boson. Furthermore, a light quark in forward direction can be radiated in the initial or final state of this process. Hence, one receives a signal-like signature. In this thesis, the full-leptonic $t\bar{t}$ process can also contribute to the background due to undetected leptons, while full-hadronic $t\bar{t}$ production is negligible.

Different cross-section ratios of the $t\bar{t}$ process with additional b jets to the $t\bar{t}$ process with additional light-flavored jets, measured by the CMS and ATLAS experiments, are observed when compared to the ratios determined with generated samples [111, 112]. To account for this mismodelling, the $t\bar{t}$ samples are split into five separately treated categories according to the original flavors of the additional jets that occur due to initial or final state radiation: $t\bar{t} + \text{lf}$, $t\bar{t} + c\bar{c}$, $t\bar{t} + 1\text{b}$, $t\bar{t} + b\bar{b}$ and $t\bar{t} + 2\text{b}$. The first category, $t\bar{t} + \text{lf}$, contains events with additional light-flavored (lf) jets. Additionally, events without any additional jets are taken into account in this category. The $t\bar{t} + c\bar{c}$ category consists of events with one or two additional jets from charm quarks. In case of additional jets that originate from bottom quarks, three categories are introduced. In the $t\bar{t} + 1\text{b}$ category, one of the two radiated b jets is undetected, while for $t\bar{t} + b\bar{b}$ both b jets can be reconstructed as spatially separated jets. In some cases, the two b jets are reconstructed as only one jet and the events fall into the $t\bar{t} + 2\text{b}$ category.

Single Top Quark Production

Another important contribution to the background is given by the production of single top quarks in the *t* channel 5.7(b) and in the tW channel 5.7(c). The *s*-channel production of single top quarks is negligible due to its low cross section. The *t*-channel production is dominant as one light quark in forward direction exists in its final state. Together with two additional bottom quarks that originate from gluon radiation, a signal-like signature is obtained.

$t\bar{t}H$ Production

The associated production of a top quark pair with a Higgs boson, which is emitted from one of the top quarks, needs also to be considered as a background process, in particular if the top quark pair decays semi-leptonically and if the Higgs boson decays into a bottom quark pair. In this case, the final state consists of four bottom quarks, one charged lepton and one neutrino as well as two light quarks that originate from the W boson (see also Figure 5.7(d)).

Miscellaneous Backgrounds

A minor contribution to the background is given by the top quark pair production in association with one vector boson, i. e. a W or Z boson. In case of the $t\bar{t} + W$ process, the additional W boson is preferably emitted in the initial state of the top quark pair production. In total, three W bosons exist in the final state, which can mimic the signal signature, if one of these W bosons decays leptonically and the two other W bosons hadronically. In the $t\bar{t} + Z$ production, the Z boson is emitted from one of the two top quarks. This process can create a signal-like signature if the top quark pair decays semi-leptonically and if the Z boson decays into a bottom quark pair. Since the cross sections of $t\bar{t} + W$ and $t\bar{t} + Z$ production are very small, their contribution to the background is insignificant.

Another minor contribution to the background is the production of a W or Z boson in association with jets. Despite the fact that it is unlikely to produce enough bottom quarks, these two processes are not negligible due to their large cross sections.

In the histograms shown in this chapter, the V + jets is grouped together with the $t\bar{t} + V$ production in one category called “Misc”.

Multijet QCD Production

Processes that contain multiple jets produced via the strong interaction are one of the most frequent processes at the LHC and are referred to as QCD production. They are able to mimic the signature of the two signal processes as leptons can be produced during the decay of D or B hadrons. In contrast to the lepton of the signal processes, these leptons are often non-isolated. By applying a cut on the missing transverse energy, the QCD background can be drastically

reduced to a negligible contribution and thus is not taken into account in this thesis.

5.3. Physics Objects Identification

This section describes the requirements that are applied in this analysis in order to define the physics objects. It is based on the general methods for the reconstruction of physics objects explained in Section 3.3.

Electrons

The electrons considered in this analysis need to pass the working point of the triggering MVA ID [113] that is defined at a signal efficiency of 80 % and must not originate from a photon conversion. Additionally, electron candidates that are located in the gap of the ECAL, i. e. $1.4442 < |\eta_{\text{sc}}| < 1.5660$, where η_{sc} is the pseudorapidity of the associated supercluster, are not reconstructed. The electron candidates need to be isolated, i. e. they need to fulfill a relative isolation of $I_\rho < 0.15$. This relative isolation is defined as:

$$I_\rho = \frac{1}{p_{\text{T}}} \left(\sum_i p_{\text{T},i}^{\text{CH}} + \max \left(0, \sum_i E_{\text{T},i}^{\text{NH}} + \sum_i E_{\text{T},i}^{\gamma} - \rho \cdot A_{\text{eff}} \right) \right) , \tag{5.1}$$

where A_{eff} is the effective area which compensates the neutral component of pileup and ρ denotes the average angular p_{T} density. The three sums $\sum_i p_{\text{T},i}^{\text{CH}}$, $\sum_i E_{\text{T},i}^{\text{NH}}$ and $\sum_i E_{\text{T},i}^{\gamma}$ correspond to the transverse momenta of charged hadrons (CH), and to the energy of neutral hadrons (NH) and photons (γ) deposited in a cone with $\Delta R = 0.3$ around the electron. Two types of electrons are defined, namely "tight" and "loose" electrons. A tight electron needs to satisfy $p_{\text{T}} > 30\,\text{GeV}$ and $|\eta| < 2.1$ and a loose electron only needs to fulfill $p_{\text{T}} > 15\,\text{GeV}$ and $|\eta| < 2.5$.

Muons

According to the guidelines of the Muon Physics Object Group (POG) [114], muons are reconstructed as global muons by using the tight Muon ID. The muons are required to be particle-flow muon candidates and their track must satisfy a goodness-of-fit of $\chi^2/\text{n.d.o.f} < 10$. Here, n.d.o.f. denotes the number of degrees of freedom. Furthermore, the global-muon track fit needs to include at least one hit in the muon chamber. Analogously to the electrons, "tight"

and "loose" muons are defined. A tight muon needs to satisfy $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, $|\eta| < 2.4$ and a relative $\Delta\beta$-corrected isolation of $I_{\Delta\beta} < 0.15$. $\Delta\beta$ corresponds to the average ratio of charged and neutral particles and corrects the deposit of the neutral particles in the cone. The relative isolation $I_{\Delta\beta}$ is given by

$$I_{\Delta\beta} = \frac{1}{p_{\mathrm{T}}} \left(\sum_i p_{\mathrm{T},i}^{\mathrm{CH}} + \max \left(0, \sum_i E_{\mathrm{T},i}^{\mathrm{NH}} + \sum_i E_{\mathrm{T},i}^{\gamma} - 0.5 \cdot \sum_i p_{\mathrm{T},i}^{\mathrm{CH(PU)}} \right) \right) , \quad (5.2)$$

with $\Delta\beta = 0.5$. Here, $\sum_i p_{\mathrm{T},i}^{\mathrm{CH(PU)}}$ denotes the sum of transverse momenta of charged hadrons that originate from pileup and that are located in a cone with $\Delta R = 0.4$ around the muon. A loose muon only needs to fulfill $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, $|\eta| < 2.4$ and $I_{\Delta\beta} < 0.2$.

Jets

In this analysis, jets are reconstructed with the anti-k_{T} algorithm with a cone radius of $R = 0.4$ (see also Section 3.3.3). If a jet is located closer than $\Delta R = 0.4$ to a tight lepton, the jet is removed from the jet collection. The reconstructed jets need to fulfill the loose particle-flow jet ID [115]. All jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are taken into account in this analysis. The energy of the jets is corrected as already described in Section 3.3.3. In order to be identified as b jets, the jets need to pass the requirements of the medium working point of the CSVv2 algorithm (CSVv2 > 0.8) [116]. Furthermore, b jets are required to fulfill $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$.

Missing Transverse Energy

The missing transverse energy (MET) $\vec{\not{E}}_{\mathrm{T}}$ is calculated as the negative vector sum of the transverse momenta of all reconstructed particles (see also Section 3.3.3). In order to account for the jet energy correction (JEC), the Type-1 MET correction [117] is applied. This correction replaces the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC has been applied.

W Bosons

As the neutrino, which originates together with a charged lepton from a leptonically decaying W boson, remains unobserved in the detector, it can be only reconstructed indirectly. While the transverse momentum of the neutrino can

be calculated from the missing transverse energy, the longitudinal component can be derived by assuming that the missing transverse energy is entirely due to the neutrino. The longitudinal momentum can then be determined by solving

$$m_{\mathrm{W}}^2 = \left(E_\ell + \sqrt{p_{\mathrm{T},\nu}^2 + p_{z,\nu}^2}\right)^2 - (\vec{p}_{\mathrm{T},\ell} + \vec{p}_{\mathrm{T},\nu})^2 - (p_{z,\ell} + p_{z,\nu})^2 \ . \tag{5.3}$$

The two solutions are given by

$$p_{z,\nu} = \frac{\mu \cdot p_{z,\ell}}{p_{\mathrm{T},\ell}^2} \pm \sqrt{\frac{\mu^2 \cdot p_{z,\ell}^2}{p_{\mathrm{T},\ell}^4} - \frac{E_\ell^2 \cdot p_{\mathrm{T},\nu}^2 - \mu^2}{p_{\mathrm{T},\ell}^2}} \ , \tag{5.4}$$

where the abbreviation μ is defined as

$$\mu = \frac{m_{\mathrm{W}}}{2} + \vec{p}_{\mathrm{T},\ell} \cdot \vec{p}_{\mathrm{T},\nu} \ . \tag{5.5}$$

The radicand in Equation 5.4 becomes negative if the reconstructed W boson mass exceeds the W boson pole mass of 80.4 GeV; this is caused by the finite $\not{E}_{\mathrm{T}}$ resolution. The imaginary component of the solutions is eliminated by modifying the x- and y-components of the neutrino transverse momentum while the measured $\not{E}_{\mathrm{T}}$ vector is kept fixed. In case of two real solutions for $p_{z,\nu}$, the solution with the lowest absolute value is chosen. More details on the reconstruction of W bosons can be found in Reference [118].

5.4. Event Selection

In order to distinguish signal events from the background, selection criteria are applied that guarantee the best signal-to-background ratio while keeping as many events as possible.

As no attempt is made to reconstruct tauons, only events containing an electron ("electron channel") or a muon ("muon channel") in the final state are considered and the analysis is performed in the combined "muon+electron" channel. Hence, the events are selected by applying the following triggers: `HLT_IsoTkMu20_vX` or `HLT_IsoMu20_vX` for the muon channel and `HLT_Ele27_eta2p1_WPLoose_Gsf_vX` for the electron channel. Every event in the muon (electron) channel needs to contain exactly one tight muon (electron), as defined in Section 5.3. If an event contains an additional loose muon or electron, it is rejected in order to suppress the contribution of Drell–Yan

events. Furthermore, a cut on the missing transverse energy in the muon channel ($\not{E}_T > 35\,\text{GeV}$) and in the electron channel ($\not{E}_T > 45\,\text{GeV}$) is applied to reject QCD multijet events.

As already described in Section 5.2.1, the final state of the two signal processes consists of four b-tagged jets. Since the bottom quark which originates from the gluon splitting often has a low transverse momentum, it cannot be reconstructed as a jet in many cases. Hence, one defines two signal regions, which require a different number of jets. In the 3 tag region, an event must consist of at least four jets with exactly three of them b-tagged, while the second signal region, referred to as 4 tag region, requires at least five jets with exactly four of them b-tagged in an event. Hence, an additional jet is required to account for the forward jet in the signal signature. In both signal regions, the untagged jets must satisfy $p_T > 30\,\text{GeV}$ in the central region ($|\eta| < 2.4$) and $p_T > 40\,\text{GeV}$ in the forward region ($2.4 \leq |\eta| \leq 4.7$), respectively.

In addition to the two signal regions, a control region (2 tag region) is defined which requires exactly two b-tagged jets, while the rest of the event selection is identical to the 3 tag region. Due to the requirement of two b jets, this region is enriched with $t\bar{t}$ events. The control region is used to search for discrepancies between simulation and data and to study the modeling of distributions in simulation.

The complete event selection for the two signal regions and the control region is summarized in Table 5.1. Table 5.2 gives an overview of the expected and observed event yields after applying the event selection.

Table 5.1.: Overview of the event selection criteria. Several requirements are applied in order to determine the two signal regions and the control region. Reconstructable jets are defined as jets that fulfill $p_T > 30\,\text{GeV}$ in the central region and $p_T > 40\,\text{GeV}$ in the forward region, respectively.

	signal regions		control region
	3 tag	4 tag	2 tag
# reconstructable jets	≥ 4	≥ 5	≥ 4
# b-tagged jets (CSVv2 > 0.8)	3	4	2
# tight leptons	1	1	1
# additional loose leptons	0	0	0
$\not{E}_T$ (muon channel)	$> 35\,\text{GeV}$	$> 35\,\text{GeV}$	$> 35\,\text{GeV}$
$\not{E}_T$ (electron channel)	$> 45\,\text{GeV}$	$> 45\,\text{GeV}$	$> 45\,\text{GeV}$

Table 5.2.: Event yields in the two signal regions. The events yields for the background processes and the tHq and tHW signal processes (SM and ITC scenarios) are denoted. The uncertainties include both statistical and systematic uncertainties. In addition, the observed event yields in data are shown.

	3 tag	4 tag
$t\bar{t} + lf$	2054 ± 721	20.6 ± 29.4
$t\bar{t} + c\bar{c}$	825 ± 607	37 ± 39
$t\bar{t} + 1b$	314 ± 184	18.1 ± 14.4
$t\bar{t} + b\bar{b}$	322 ± 265	69 ± 57
$t\bar{t} + 2b$	171 ± 135	12.6 ± 14.9
Single top	151 ± 51	5.8 ± 2.7
$t\bar{t}H$	19.6 ± 9.3	5.1 ± 2.8
$t\bar{t}Z$	9.3 ± 2.7	1.8 ± 0.7
$t\bar{t}W$	7.8 ± 2.5	0.4 ± 0.3
$W + jets$	42 ± 47	0.0 ± 0.0
$Z + jets$	9.3 ± 5.4	0.3 ± 0.3
Sum of backgrounds	3925 ± 1008	171 ± 80
tHq (SM)	0.75 ± 0.20	0.11 ± 0.04
tHW (SM)	0.59 ± 0.09	0.09 ± 0.02
tHq (ITC)	10.8 ± 3.5	1.6 ± 0.7
tHW (ITC)	6.5 ± 1.1	1.1 ± 0.3
Observed	3603	171

5.5. Corrections to Simulation and Data

It is crucial to simulate events accurately as the observed data is compared to a BDT output derived from Monte Carlo simulation events in order to obtain exclusion limits. Due to the complexity of the experiment, various discrepancies between data and simulation can arise, which need to be corrected by modifying the events. This section gives an overview of the corrections applied to simulation and data.

Pileup Reweighting

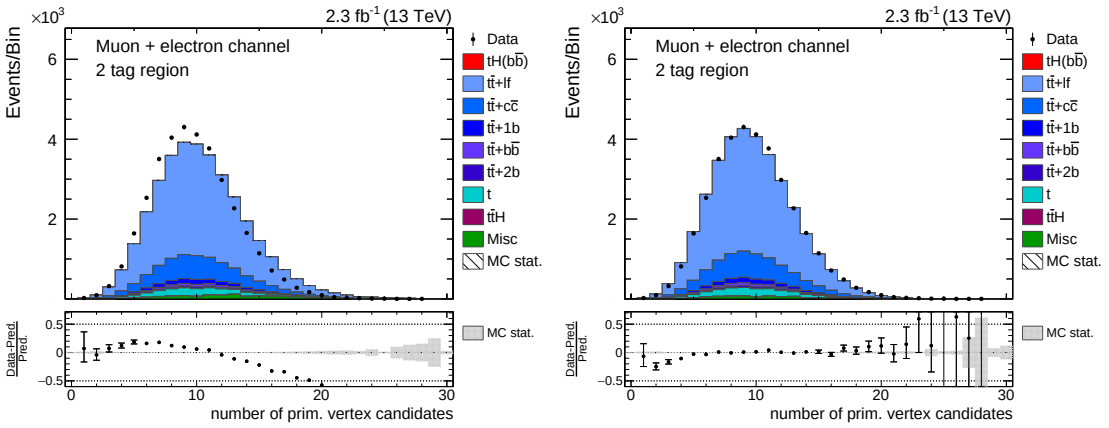
To account for pileup events, an expected modeling of the pileup shape, which is independent of the measured data, is applied on the simulated events. The number of primary vertices is a good measure of the probability density of pileup events and is highly sensitive to details of the reconstruction and to differences in the modeling of underlying events. As the number of primary vertices of the simulated events does not exactly concur with the actual observed number in data, the simulated events need to be reweighted based on the number of true pileup interactions, which is derived from the total cross section of inelastic proton-proton scattering and from a measurement of the instantaneous luminosity [119]. The effect of the pileup reweighting is illustrated in Figure 5.8.

Lepton Efficiency Scale Factors

Several discrepancies with respect to the muon and electron efficiencies between data and simulated events need to be corrected by applying appropriate scale factors.

The correction of the muon efficiencies comprises scale factors for the muon identification, isolation and trigger efficiencies which are treated separately. These correction factors are derived with the Tag-and-Probe method which uses narrow di-lepton resonances like Z or J/Ψ [120] and are provided by the Muon POG [121].

Separate scale factors for the electron reconstruction and for the triggering MVA ID, provided by the EGamma POG [122], are applied on the simulated electron efficiencies. These corrections are derived from $Z \to e^+e^-$ events using the Tag-and-Probe method [123]. Furthermore, the corrections of the electron



(a) (b)

Figure 5.8.: The number of primary vertices. In the distributions of the number of primary vertices before (a) and after (b) the application of the pileup reweighting, the simulated events are scaled to data. Slightly more primary vertices appear in the simulation than observed in data. After the reweighting, good agreement between simulation and data is observed.

isolation and trigger efficiency are also applied. For these efficiencies, privately calculated correction scale factors are used as they are not provided by the EGamma POG.

The lepton efficiency scale factors cause a reduction of the overall yield of $\approx 4\%$ and a slight change in the shape for some distributions.

b Tagging Efficiency

Since a discrepancy between data and simulation is observed for the distribution of the CSVv2 values, a reweighting procedure for the output of the CSVv2 algorithm is applied [124] to correct the shape of these distributions. Every event is assigned a weight such that the simulated events reproduce the complete measured CSVv2 distribution in data. The scale factors are a function of the jet CSVv2 value, jet p_{T} and jet η and are calculated for both heavy-flavored and light-flavored jets. The scale factors for heavy-flavored jets are determined with the Tag-and-Probe method applied in a control region that mostly contains di-leptonically decaying top quark pairs, while the scale factors for light-flavored jets are received in a region enriched with $\mathrm{Z}(\ell\ell)+\mathrm{jets}$ events. The scale factors for heavy-flavored jets take the light flavor contamination into

account and are determined as

$$\mathrm{SF_{HF}}(\mathrm{CSVv2}, p_\mathrm{T}, \eta) = \frac{\mathrm{Data} - \mathrm{MC_{LF}}}{\mathrm{MC_{HF}}} \,. \tag{5.6}$$

Here, $\mathrm{MC_{LF}}$ and $\mathrm{MC_{HF}}$ denote the simulated yields of light-flavored and heavy-flavored jets, respectively. By swapping the positions of $\mathrm{MC_{LF}}$ and $\mathrm{MC_{HF}}$ and by requiring one jet to be untagged instead of b-tagged, one receives analogously the scale factors $\mathrm{SF_{LF}}$ for light-flavored jets. More details on the scale factor estimation can be found in Reference [125].

Jet Pseudorapidity Mismodeling

The light-flavored jet in the forward region is one of the most prominent features of the tHq signal process. Though, a severe mismodeling of the jet pseudorapidity in the forward region is observed when comparing data with simulation. In particular, the pseudorapidity of the light-flavored jet is affected. In Figure 5.9, this mismodeling is shown in the 2 tag region .

The main contribution to the mismodeling is given by an overestimated number of jets in the simulated events for $2.6 \leq |\eta| \leq 3.4$. Good agreement between data and simulation is observed in the region with $|\eta| \approx 3.5$, but the discrepancy increases in the most forward region ($|\eta| > 3.5$). The mismodeling primarily occurs for jets with low transverse momenta, but remains visible at the higher end of the p_T spectrum.

The actual η values of data and simulation are redefined in order to be insensitive to the mismodeling:

$$\eta \mapsto \eta' = \begin{cases} \eta & \text{if} \quad |\eta| < 2.4 \,, \\ 2.8 \cdot \mathrm{sgn}(\eta) & \text{if} \quad 2.4 \leq |\eta| \leq 3.2 \,, \\ 3.5 \cdot \mathrm{sgn}(\eta) & \text{if} \quad |\eta| > 3.2 \,. \end{cases} \tag{5.7}$$

The bin boundaries are chosen in accordance with the derivation of the L2L3Residual corrections (see also Section 3.3) in these two bins due to a low number of jets in the forward region. Since the transformation is applied to both data and simulation, the modified pseudorapidity η' is an effective observable. Furthermore, the analysis is blind to the mismodeling in the forward region as the analysis only relies on good conformity of the simulated events with data in the transformed variables.

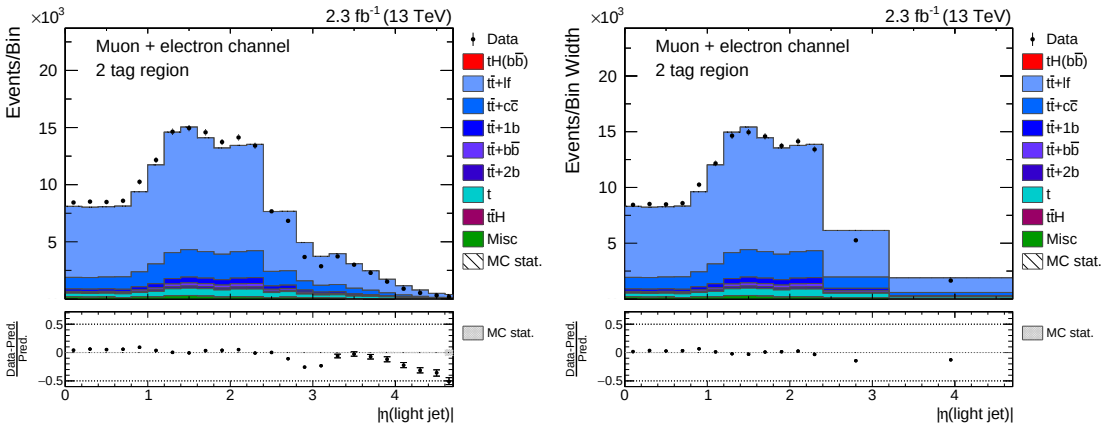


Figure 5.9.: The observed mismodeling of the jet pseudorapidity in the 2 tag region. In the distributions of the absolute pseudorapidity of the light-flavored jet before (a) and after (b) applying the transformation, the simulation is scaled to data. Improved agreement between the simulated and recorded events can be observed.

After applying the correction, a slightly different normalization in the forward region is still observed in the simulation. This is the reason that jets with transverse momenta of $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ for $|\eta| > 2.4$ are required for the event reconstruction (see also Table 5.1).

5.6. Event Reconstruction

Due to the high jet multiplicity after the event selection, a large number of possibilities exists to assign the jets to the final-state quarks of an event, which makes it difficult to find the correct assignment for the event reconstruction. For this purpose, boosted decision trees (BDTs) are applied in order to reconstruct the events.

As already mentioned in Section 5.1, every event is reconstructed under two different hypotheses, namely the tHq and the $\mathrm{t\bar{t}}$ hypothesis, where the jets are assigned to the quarks in the final states of the tHq signal process and the $\mathrm{t\bar{t}}$ background process, respectively. Since the tHW signal process kinematically lies between the tHq and $\mathrm{t\bar{t}}$ process and thus can be separated from the background with these two reconstruction hypotheses, no separate event reconstruction under a tHW hypothesis is performed. Based on the response

of the BDTs, the jet assignments are chosen. From both hypotheses, several variables are then derived for each event. These variables are used, together with reconstruction-independent variables, for the discrimination of the signal events from the background, which will be later described in Section 5.7.

5.6.1. Jet Assignment under the tHq Hypothesis

The aim of the tHq reconstruction is the assignment of the measured jets to the four quarks of the tHq signal process, i. e. the two bottom quarks originating from the decay of the Higgs boson, the bottom quark from the decay of the top quark and the light-flavored quark. As the cross section and the kinematics of the tHq process depend on the $\kappa_t - \kappa_V$ configuration, 51 different tHq reconstruction procedures need to be performed separately.

In order to find the best possible jet-quark assignment, 51 BDTs in total are trained in the 3 tag signal region. Due to the small number of events, no separate training is performed in the 4 tag region. The simulated tHq signal process is split into a training sample, used as input for the BDT training, a test sample, which is used for the search for possible overtraining, and a sample used for the evaluation of the training. Each BDT is trained with correct and wrong jet assignments from simulated events. A matchable assignment is defined as the configuration in which every reconstructed jet is within a cone of the size $\Delta R = 0.3$ around the assigned final-state quark. The best possible jet assignment, referred to as correct assignment, is subsequently chosen by selecting the matchable assignment that produces the highest BDT response in an event. The remaining assignments are then considered, together with assignments that failed to match all final-state quarks, as wrong. From all wrong assignments in an event, exactly one is randomly chosen. Hence, each event provides exactly one correct and exactly one wrong assignment. If an event does not possess a correct assignment, it is discarded for the training.

Since $n_{\text{jet}}!/(n_{\text{jet}} - 4)!$ possibilities exist to assign the jets for each event but only one of them can be the correct assignment, each considered assignment needs to satisfy a set of requirements to reduce the combinatorics. Jets matched to the bottom quarks must be located in the central region of the detector ($|\eta| < 2.4$) and must be b-tagged, while the jet assigned to the light-flavored quark needs to be untagged (CSVv2 < 0.8).

For each of the 51 tHq reconstruction procedures, the same set of variables,

listed in Table 5.3, is used for the training to ensure that the different reconstruction procedures are comparable. Out of the 15 variables, the invariant masses of the reconstructed Higgs boson and of the reconstructed top quark are the most important ones, followed by ΔR between the two jets from the Higgs boson decay. The distributions for the six most important variables for correct and wrong assignments for the ITC scenario ($\kappa_t = -1$, $\kappa_V = +1$) can be found in Figure 5.10. The remaining nine variables are shown in Appendix B.1 and B.2.

The set of variables has been optimized by calculating the area under the receiver-operating-characteristic (ROC) curve, which is used as a measure of performance, for each of the 51 BDT trainings. An example of a ROC curve is illustrated in Figure 5.11. The different ROC values in dependency of the $\kappa_t - \kappa_V$ configuration are shown in Figure 5.12. Overall, a good performance of the trainings can be observed. However, the separation between correct and wrong assignments is worse for points in the $\kappa_t - \kappa_V$ plane corresponding to lower cross sections of the process, as the training is effectively performed on a smaller number of events. The reduced number of events, caused by the application of negative Les Houches Event (LHE) [126, 127] weights, increases the chance of overtraining in these points, thus a smaller number of trees is used for the training of the corresponding BDTs. The parameter settings chosen for the 51 BDTs can be found in Table 5.4.

Exemplarily, the response of the tHq reconstruction BDT for the SM case ($\kappa_t = \kappa_V = +1$) and for the ITC case ($\kappa_t = -1$, $\kappa_V = +1$) are shown in Figure 5.13 for the training and for the test sample. In both cases, a good separation between correct and wrong jet assignments can be observed and no signs of overtraining are found.

After the successful training, the tHq reconstruction BDT response for all 51 $\kappa_t - \kappa_V$ configurations is applied to data and simulation events in all regions. Good agreement between simulation and data can be observed in Figure 5.14. Figure 5.15 shows the reconstruction efficiencies under the tHq hypothesis for the ITC scenario, obtained from the simulated tHq events. The matching efficiency for a complete tHq event is 47 % in the 3 tag region and 32 % in the 4 tag region, respectively. In both regions, the best efficiency is obtained for the reconstruction of the light-flavored jet as only one jet needs to be assigned to the final-state light-flavored quark, while three b-tagged jets must be matched correctly to the final-state bottom quarks.

Table 5.3.: Input variables for the tHq reconstruction BDTs. The variables are sorted by their importance in the training, averaged over all 51 BDTs.. In total, 15 variables are used for the training of the BDTs.

Variable	Description
$\log m(\mathrm{H})/\mathrm{GeV}$	Invariant mass of the reconstructed Higgs boson
$\log m(\mathrm{t})/\mathrm{GeV}$	Invariant mass of the reconstructed top quark
$\Delta R(\text{Higgs jets})$	ΔR between the two jets from the Higgs boson decay
$\Delta R(\mathrm{b_t}, \mathrm{W})$	ΔR between the jet assigned to the bottom quark from the top quark decay and the W boson
relative $\mathrm{H_T}$	Ratio of $p_\mathrm{T}(\mathrm{H}) + p_\mathrm{T}(\mathrm{t}) + p_\mathrm{T}(\text{light jet})$ to the scalar sum of p_T of all jets, charged lepton, and $\not{E}_\mathrm{T}$
$\cos\theta(\mathrm{b_t}, \ell)$	Cosine of the angle between the jet assigned to the bottom quark from the top quark decay and the charged lepton
$\mathrm{CSV}(\text{Higgs jet 2})$	Output of the CSVv2 b tagging algorithm for the second hardest jet assigned to the Higgs boson
$\mathrm{CSV}(\mathrm{b_t})$	Output of the CSVv2 b tagging algorithm for the jet assigned to the bottom quark from the top quark decay
$\lvert\eta(\text{light jet}) - \eta(\mathrm{b_t})\rvert$	Absolute difference of pseudorapidities of the light-flavored jet and of the b jet from the top quark decay
$\mathrm{CSV}(\text{Higgs jet 1})$	Output of the CSVv2 b tagging algorithm for the hardest jet assigned to the Higgs boson
$\lvert\eta(\mathrm{b_t})\rvert$	Absolute pseudorapidity of the jet assigned to the bottom quark of the top quark decay
$\lvert\eta(\mathrm{t}) - \eta(\mathrm{H})\rvert$	Absolute difference of pseudorapidities of reconstructed top quark and reconstructed Higgs boson
$\log\min(p_\mathrm{T}(\text{Higgs jets}))/\mathrm{GeV}$	Lower transverse momentum of the two jets assigned to the Higgs boson decay products
$\lvert\eta(\text{light jet})\rvert$	Absolute pseudorapidity of the light-flavored jet
$\Delta E(\text{light jet}, \mathrm{b_t})$	Energy difference between the light-flavored jet and the jet assigned to the bottom quark from the top quark decay

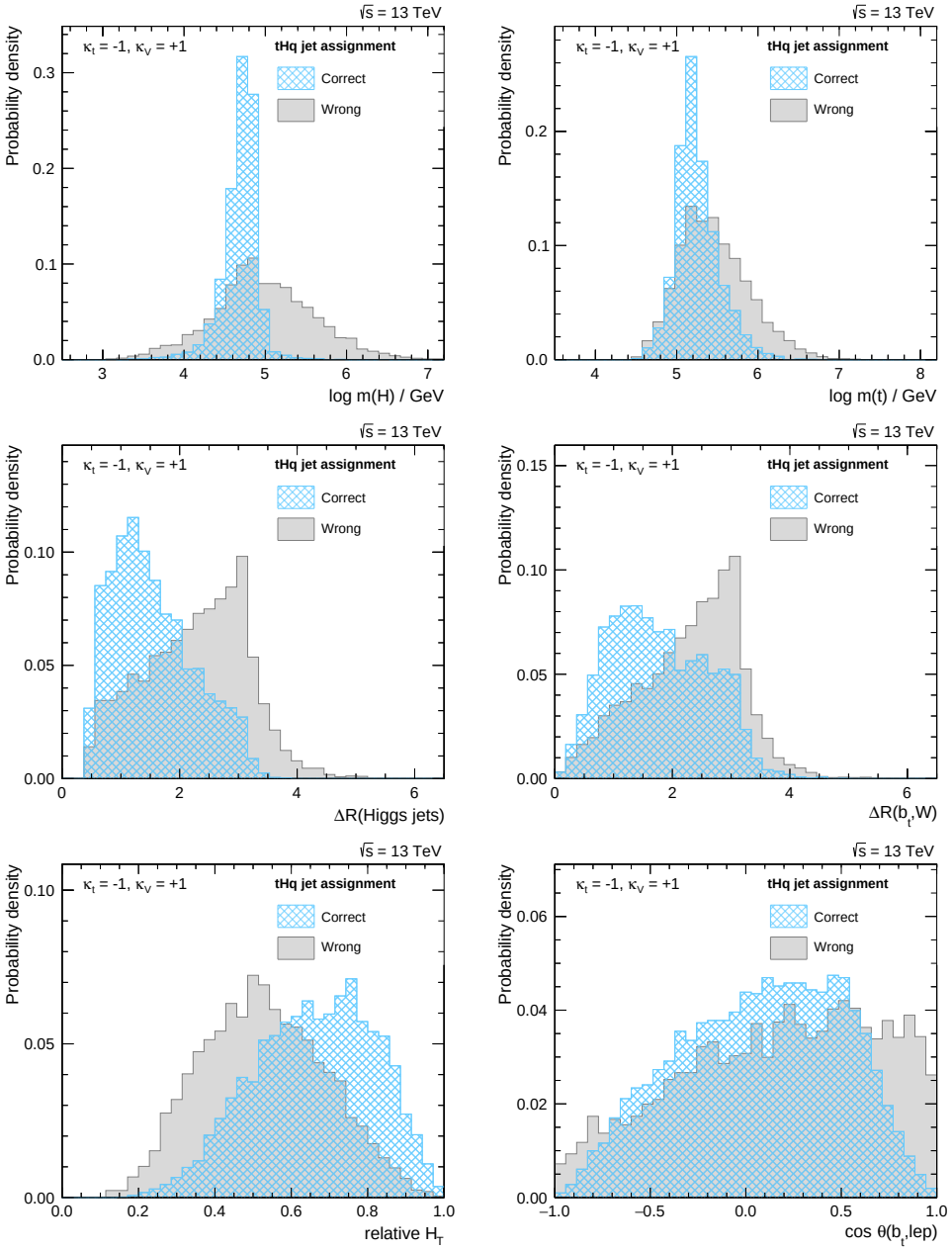


Figure 5.10.: The distributions of the six most discriminating variables of the tHq reconstruction. The distributions of the correct and wrong jet assignments for each variable are shown. The variables are sorted by their importance in the training. The description of the variables is listed in Table 5.3. The remaining variables can be found in Appendix B.1 and B.2.

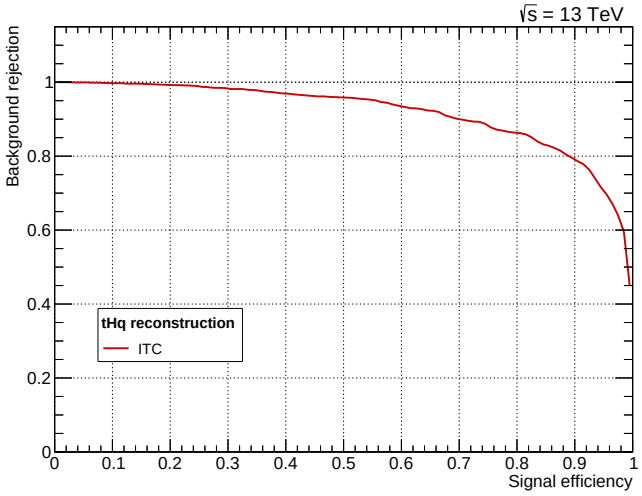


Figure 5.11.: The area under the ROC curve of the tHq reconstruction for the ITC case. The ROC curve has been obtained from the training of the tHq reconstruction BDT for $\kappa_t = -1$, $\kappa_V = +1$. The area under the ROC curve is a measure of the efficiency of the BDT training and corresponds to a value of 92.4 %.

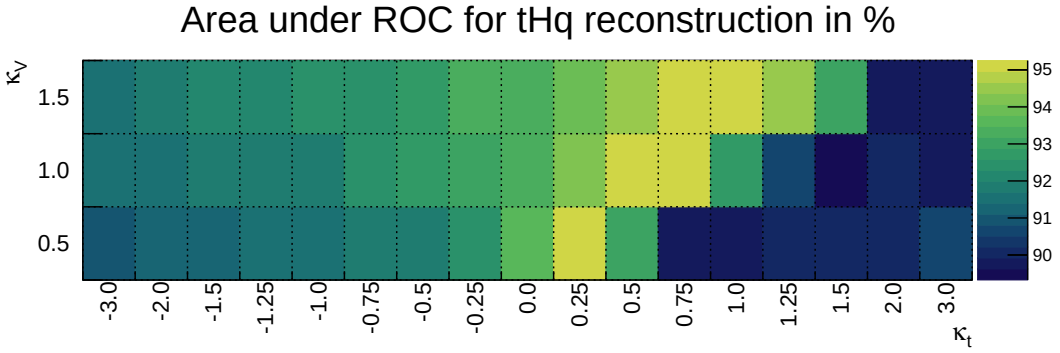
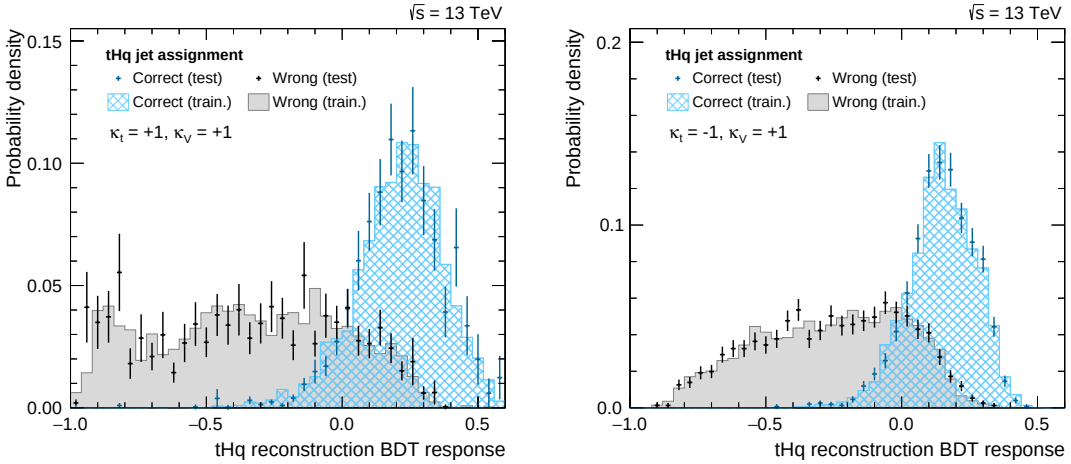


Figure 5.12.: The area under the ROC curve for all 51 tHq reconstruction procedures. An overall good performance of the training is observed as the ROC values are all higher than 89 %. For $\kappa_t - \kappa_V$ configurations with smaller cross sections, a slightly worse separation between correct and wrong assignments occurs.

Table 5.4.: Parameters used for the training of the tHq reconstruction BDTs. The lower number of trees is only used for points that have otherwise shown signs of overtraining. These points include the following $(\kappa_t|\kappa_V)$ value pairs: $(+1|+1)$, $(+1.25|+1)$, $(+1.5|+1)$, $(+1.5|+1.5)$, $(+2|+1.5)$ and $(+0.5|+0.5)$. After this reduction no sign of overtraining is found. Details on the parameters can be found in Reference [106].

Parameter	Value
`NTrees`	400/150
`MinNodeSize`	1
`MaxDepth`	3
`BoostType`	AdaBoost
`nCuts`	20
`AdaBoostBeta`	0.3
`SeparationType`	GiniIndex



(a) (b)

Figure 5.13.: Output values of the tHq reconstruction BDTs. A clear separation between correct and wrong jet assignments can be observed for the SM case (a) and for the ITC case (b). Furthermore, no signs of overtraining can be found.

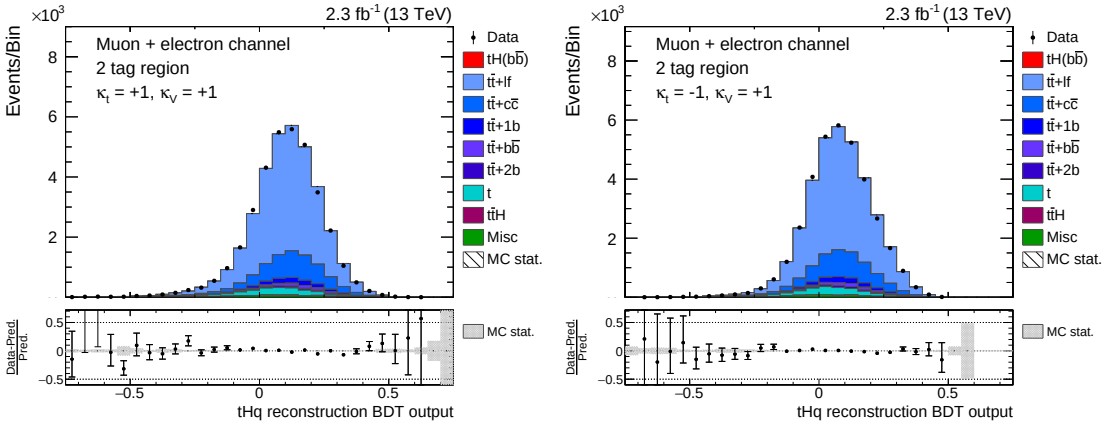


Figure 5.14.: Comparison of the tHq reconstruction BDT output between simulation and data. The highest BDT output value, i. e. the chosen correct jet assignment, per event is shown in the 2 tag control region for the SM (a) and for the ITC scenario (b). The simulated events are scaled to data. In both distributions, good agreement between data and simulated events is observed.

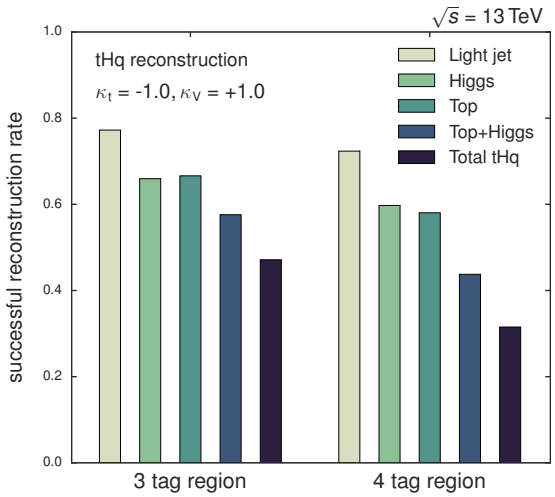


Figure 5.15.: Reconstruction efficiency of the tHq process for the ITC scenario. The efficiencies of the different objects of the tHq process and of the total tHq event are shown for the 3 tag and 4 tag region. Since the tHq reconstruction is exclusively trained in the 3 tag region, the reconstruction efficiency in the 3 tag region is higher than in the 4 tag region.

5.6.2. Jet Assignment under the $t\bar{t}$ Hypothesis

Semi-leptonic top quark pair decays are the most dominant background in the analysis. In order to provide an additional handle to suppress them, a dedicated jet assignment is developed. For this purpose, the semi-leptonic $t\bar{t}$ events are split into three samples: the training sample used as input for the BDT training, the test sample used to check for possible overtraining, and the evaluation sample.

Analogously to the tHq reconstruction, the measured jets are matched to the four final-state quarks of the semi-leptonic $t\bar{t}$ process, namely two bottom quarks from the top quark decays and two light-flavored quarks originating from the hadronically decaying W boson. The correct assignment is again found if all four final-state quarks can be matched to a jet within a cone size of $\Delta R = 0.3$. An event is not used for the training if no correct assignment can be found. The remaining possible jet assignments that do not correspond to the best correct assignment are considered as wrong assignments. In contrast to the tHq process, the kinematics of the $t\bar{t}$ process does not depend on the chosen $\kappa_t - \kappa_V$ configuration. Hence, a single $t\bar{t}$ reconstruction BDT is sufficient for all 51 studied $\kappa_t - \kappa_V$ configurations. The parameter settings for this reconstruction BDT are equal to the ones for the tHq reconstruction BDTs using the unreduced number of trees (see also Table 5.4).

A set of requirements is applied in order to reduce the possibilities for assigning the jets to the final-state quarks. Both jets that are assigned to the b quarks of the two top quark decays need to satisfy $|\eta| < 2.4$ and need to pass the requirements of the medium working point of the b tagging algorithm (CSVv2 > 0.8).

For the training of the $t\bar{t}$ reconstruction BDT, which is again performed in the 3 tag region, 11 input variables are used. A description of these variables can be found in Table 5.5. Out of these 11 variables, the invariant mass of the two jets assigned to the hadronically decaying W boson of the top quark and the difference between the invariant masses of the reconstructed hadronically decaying top quark and W boson are the most important ones. The distributions for the six most important variables for correct and wrong assignments can be found in Figure 5.16. The remaining five variables are shown in Appendix B.3.

Table 5.5.: Input variables for the $t\bar{t}$ reconstruction BDT. The variables are sorted by their importance in the training. The hadronically and leptonically decaying top quarks and W bosons are labeled with t_{had}, t_{lep}, W_{had} and W_{lep}. In total, 11 variables are used for the training of the BDT.

Variable	Description
$\log m(W_{had})/\text{GeV}$	Invariant mass of the two jets assigned to the W boson of t_{had}
$\log(m(t_{had}) - m(W_{had}))/\text{GeV}$	Difference between the invariant masses of reconstructed t_{had} and W_{had}
$\log m(t_{lep})/\text{GeV}$	Invariant mass of the reconstructed t_{lep}
CSV(W_{had} jet 1)	CSVv2 output of the hardest jet assigned to W_{had}
$\Delta R(b_{t_{lep}}, W_{lep})$	ΔR between the bottom quark of the reconstructed t_{lep} and W_{lep}
CSV(W_{had} jet 2)	CSVv2 output of the second hardest jet assigned to W_{had}
$\Delta R(W_{had}$ jets)	ΔR between the two jets assigned to the W boson of t_{had}
relative H_T	Ratio of $p_T(t_{had}) + p_T(t_{lep})$ to the scalar sum of p_T of all jets, charged lepton, and $\not{E}_T$
$\Delta R(b_{t_{had}}, W_{had})$	ΔR between the bottom quark of the reconstructed t_{had} and W_{had}
$\log p_T(t_{had})/\text{GeV}$	Transverse momentum of the reconstructed t_{had}
$\log p_T(t_{lep})/\text{GeV}$	Transverse momentum of the reconstructed t_{lep}

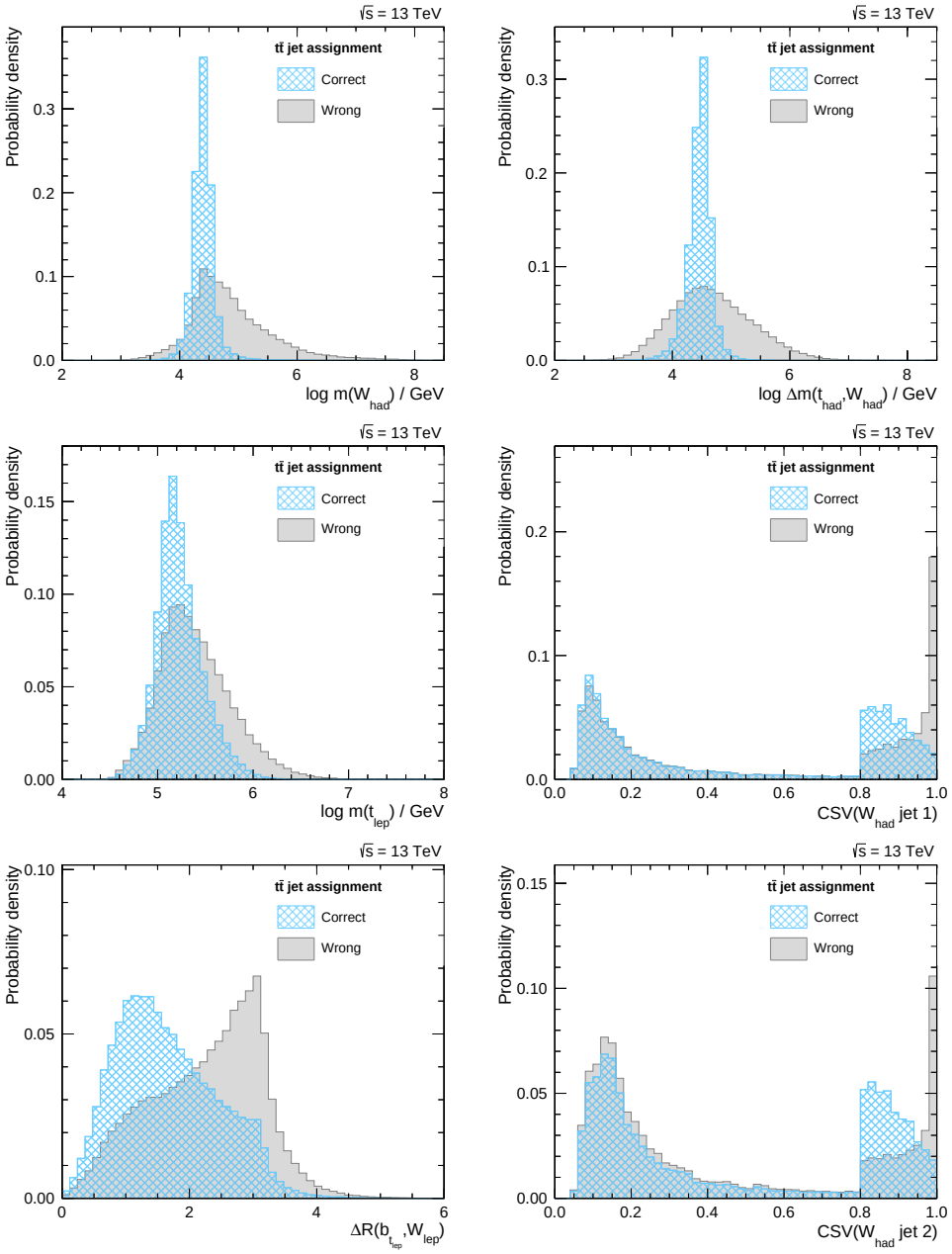


Figure 5.16.: The distributions of the six most discriminating variables of the $t\bar{t}$ reconstruction. The distributions of the correct and wrong jet assignments for each variable are shown. The variables are sorted by their importance in the training. The description of the variables is listed in Table 5.5. The remaining variables can be found in Appendix B.3.

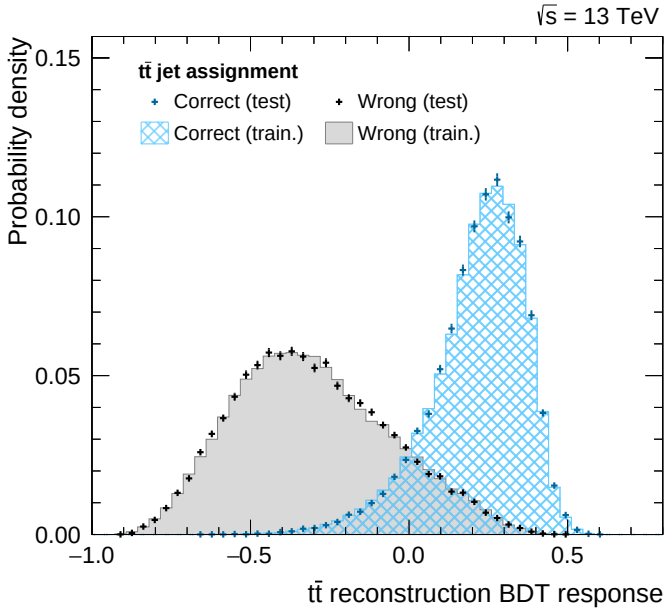


Figure 5.17.: Output value of the $t\bar{t}$ reconstruction BDT. A clear separation between correct and wrong jet assignments can be observed. Furthermore, no signs of overtraining can be found.

The response of the $t\bar{t}$ reconstruction BDT is shown in Figure 5.17 for the training and for the test sample. A good separation between correct and wrong jet assignments can be observed and no signs of overtraining are found.

After the successful training, the $t\bar{t}$ reconstruction BDT response is applied to data and simulation events in all regions. A reasonable agreement between simulation and data can be observed in Figure 5.18. Figure 5.19 shows the reconstruction efficiencies under the $t\bar{t}$ hypothesis, obtained from the simulated semi-leptonic $t\bar{t}$ events. The matching efficiency for a complete $t\bar{t}$ event is 68 % in the 3 tag region and 40 % in the 4 tag region, respectively.

Each event possesses 51 jet assignments under the tHq hypothesis and one under the $t\bar{t}$ hypothesis after performing the tHq and the $t\bar{t}$ reconstruction. The variables which are derived from these two reconstruction procedures are then used as input variables for the classification of events.

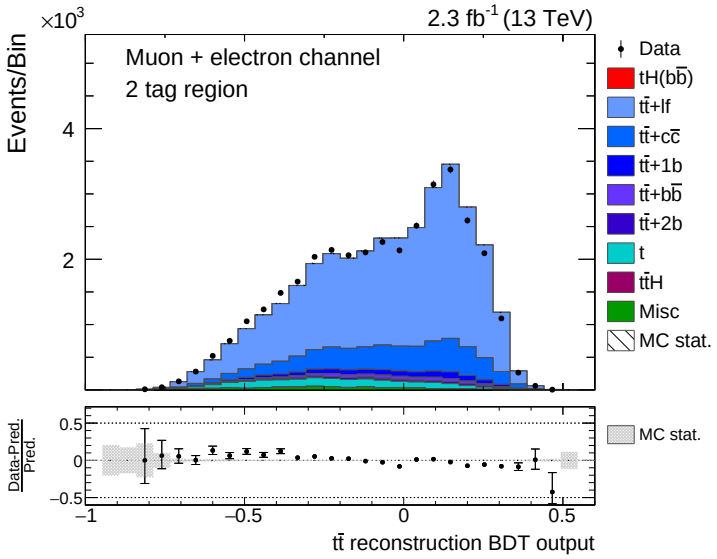


Figure 5.18.: Comparison of the tt̄ reconstruction BDT output between simulation and data. The highest BDT output value per event is shown in the 2 tag control region. The simulated events are scaled to data. Reasonable agreement between data and simulated events is observed.

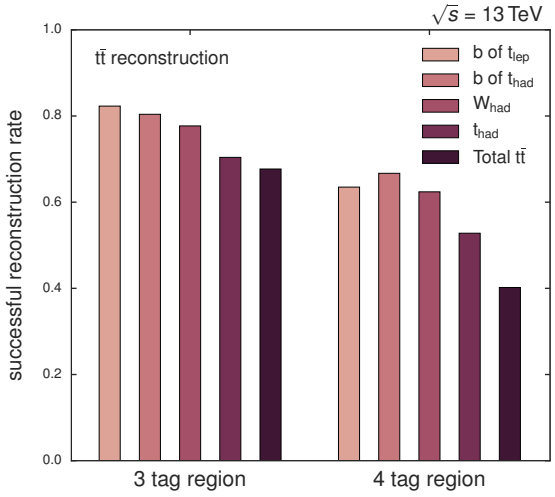


Figure 5.19.: Reconstruction efficiency of the semi-leptonic tt̄ process. The efficiencies of the different objects of the tt̄ process and of the total tt̄ event are shown for the 3 tag and 4 tag region. Since the tt̄ reconstruction is exclusively trained in the 3 tag region, the reconstruction efficiency in the 3 tag region is higher than in the 4 tag region.

5.7. Event Classification

Due to the small signal-to-background ratio, the discrimination between signal and background events is again performed by using BDTs. Three sets of variables are employed as input variables: one set of variables derived from the tHq reconstruction, one set of variables derived from the $t\bar{t}$ reconstruction and one set of variables independent of any reconstruction.

Since the variables reconstructed under the tHq hypothesis differ for each studied $\kappa_t - \kappa_V$ configuration, the event classification is performed separately for each of the 51 values in the $\kappa_t - \kappa_V$ plane. Hence, 51 BDTs in total need to be trained. The BDTs are only trained in the 3 tag region since the 4 tag region does not provide a sufficiently large number of events. As the second signal process, the tHW production, kinematically lies between the tHq signal process and the $t\bar{t}$ background process, it is not considered as a signal input. The dominant semi-leptonic $t\bar{t}$ production, the di-leptonic $t\bar{t}$ production and the $t\bar{t}H$ process are used as background events. The background events are scaled to their predicted cross sections and the signal is scaled such that the integral is consistent with the integral of the background events.

Similarly to the tHq reconstruction, the same set of variables, optimized by calculating the area under the ROC curve, is used for all classification BDTs. The ROC values for the different $\kappa_t - \kappa_V$ configurations are shown in Figure 5.20. Overall, a good performance of the training can be observed. However, the separation between signal and background events is worse for points in the $\kappa_t - \kappa_V$ plane corresponding to lower cross sections of the process, as the training is effectively performed on a smaller number of events. The reduced number of events, caused by the application of negative LHE weights, increases the chance of overtraining in these points, thus a smaller number of trees is used for the training of the corresponding BDTs. The parameter settings chosen for the 51 BDTs can be found in Table 5.6.

Exemplary, the response of the event classification BDT for the SM case and for the ITC case is shown in Figure 5.21 for the training and for the test sample. In both cases, a good separation between signal and background events can be observed and no signs of overtraining are found.

The optimized set of variables includes 15 variables out of three categories in total, which are listed in Table 5.7. The aplanarity, which provides information

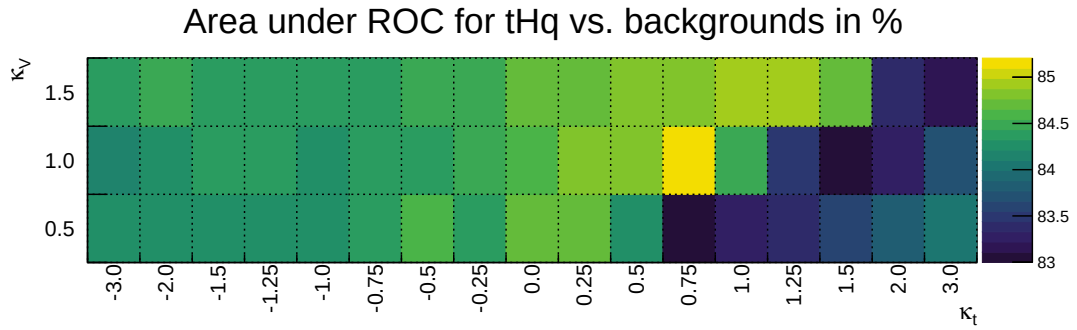


Figure 5.20.: The area under the ROC curve for all 51 event classifications. An overall good performance of the training is observed as the ROC values are at least higher than 83 %. For $\kappa_t - \kappa_V$ configurations with smaller cross sections, a slightly worse separation between signal and background events occurs.

Table 5.6.: Parameters used for the training of the event classification BDTs. For the following $\kappa_t - \kappa_V$ configurations, the lower number of trees is used to prevent overtraining: $\kappa_t \geq +0.5$ for $\kappa_V = +1.5$, $\kappa_t \geq +0.75$ for $\kappa_V = +1$, and $+0.25 \leq \kappa_t \leq +2$ in case of $\kappa_V = +0.5$. Details on the parameters can be found in Reference [106].

Parameter	Value
`NTrees`	400/100
`MinNodeSize`	1
`MaxDepth`	3
`BoostType`	AdaBoost
`nCuts`	20
`AdaBoostBeta`	0.3
`SeparationType`	GiniIndex

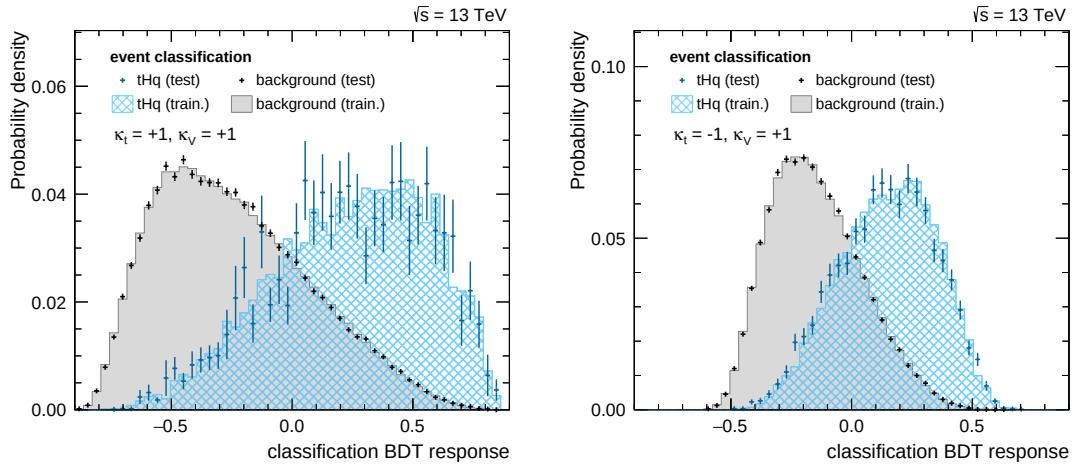


Figure 5.21.: Output values of the event classification BDTs. A clear separation between signal and background events can be observed for the SM case (a) and for the ITC case (b). Furthermore, no signs of overtraining can be found.

about the geometrical shape of the event in general and is further described in Reference [128], is the most important global variable. The invariant mass of the hadronically decaying top quark and the absolute pseudorapidity of the light-flavored jet in forward direction are the most important variables that depend on the tHq and $t\bar{t}$ reconstruction, respectively. The distributions for the six most important variables of the event classification for the ITC scenario ($\kappa_t = -1$, $\kappa_V = +1$) can be found in Figure 5.22. The remaining nine variables are shown in Appendix B.4 and B.5.

After the successful training, the event classification response for all 51 $\kappa_t - \kappa_V$ configurations is applied to data and simulation events in all regions. Good agreement between simulation and data can be observed in Figure 5.23. Furthermore, the ROC curves of the SM and ITC scenarios for the signal regions are illustrated in Figure 5.24. Although no separate training for the event classification has been performed in the 4 tag region, an efficiency comparable to the 3 tag region can be observed. The effectively smaller number of events causes a slightly worse efficiency for the SM case.

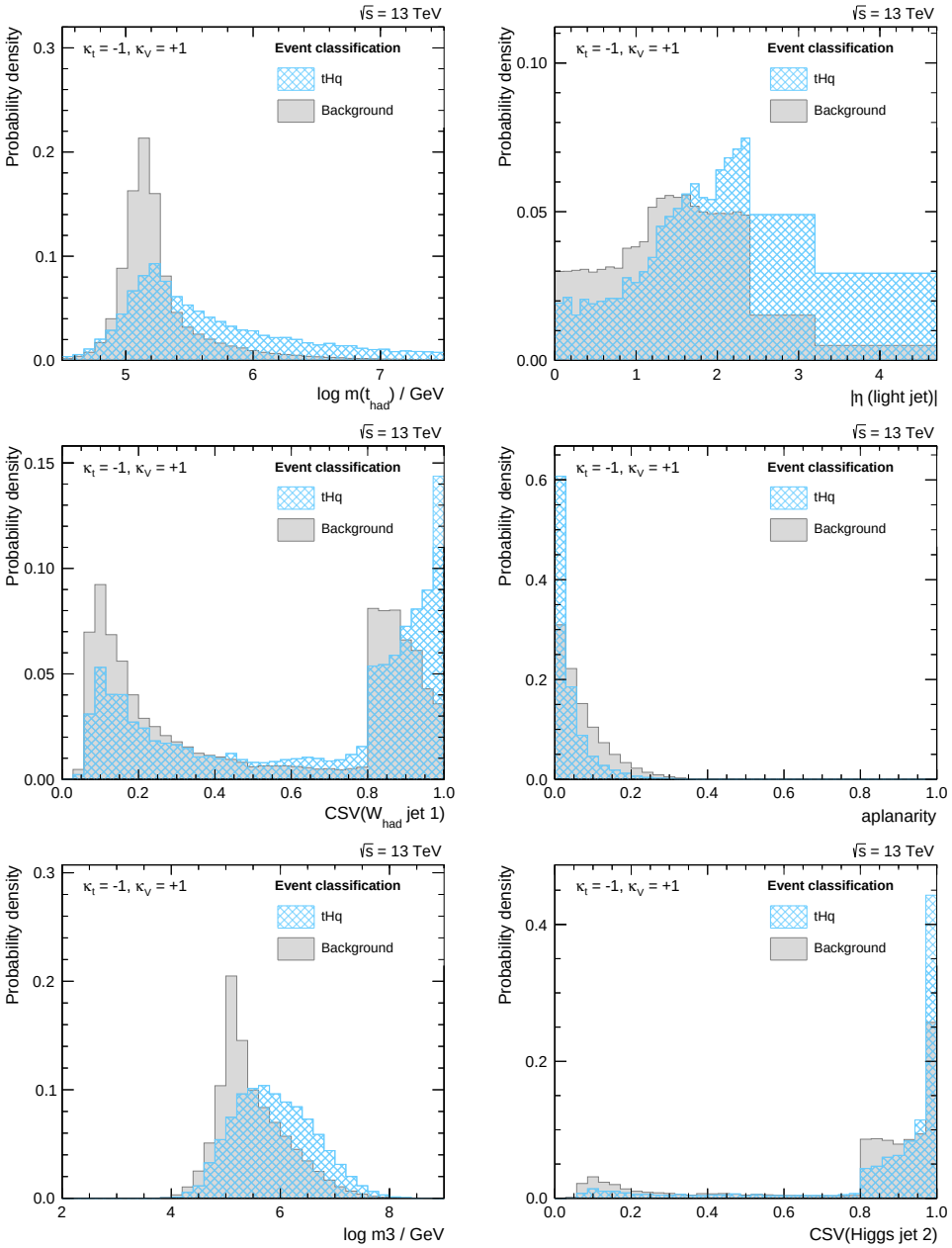
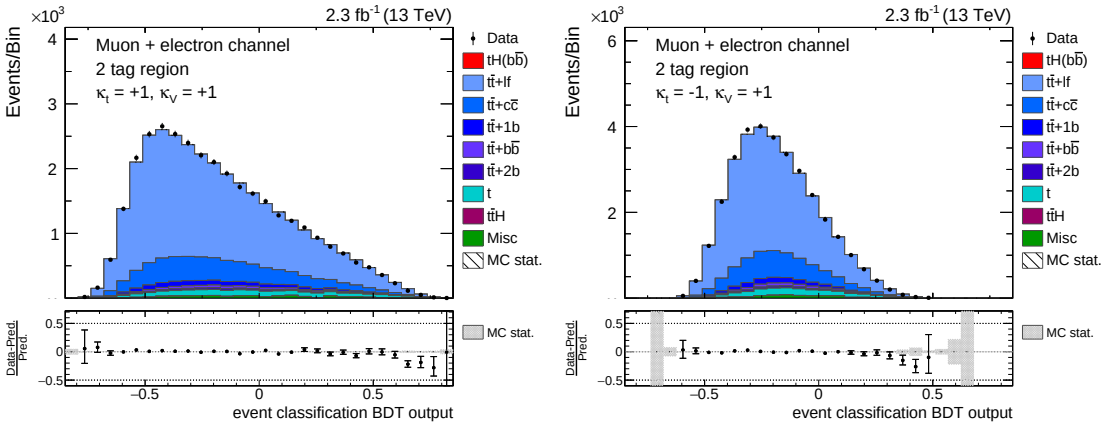


Figure 5.22.: The distributions of the six most discriminating variables of the event classification. The distributions of signal and background events for each variable are shown. The variables are sorted by their importance in the training. The description of the variables is listed in Table 5.7. The remaining variables can be found in Appendix B.4 and B.5.

Table 5.7.: Input variables for the event classification BDTs. The variables are sorted by their importance in the training within each category, averaged over all 51 BDTs. In total, 15 variables are used for the training of the BDTs.

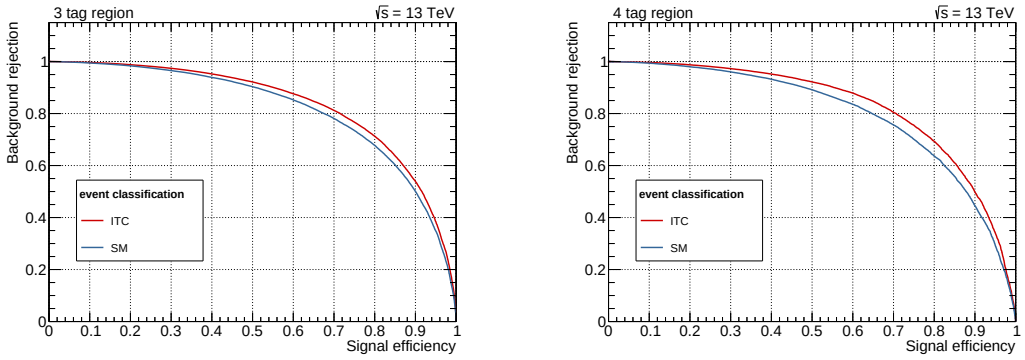
Variable	Description
Variables independent of any reconstruction	
aplanarity	Aplanarity of the event
$\log m3/\text{GeV}$	Invariant mass of the three hardest jets in the event
Fox-Wolfram #1	First Fox-Wolfram moment of the event[a]
$q(\ell)$	Electric charge of the lepton in units of e
Variables based on objects reconstructed under the $t\bar{t}$ hypothesis	
$\log m(t_{had})/\text{GeV}$	Invariant mass of t_{had}
CSV(W_{had} jet 1)	CSVv2 output of the hardest jet assigned to W_{had}
$\Delta R(W_{had}$ jets)	ΔR between the two jets from the decay of W_{had}
CSV(W_{had} jet 2)	CSVv2 output of the second hardest jet assigned to W_{had}
Variables based on objects reconstructed under the tHq hypothesis	
$\lvert\eta(\text{light jet})\rvert$	Absolute pseudorapidity of the light-flavored jet
CSV(Higgs jet 2)	CSVv2 output of the second hardest jet assigned to the Higgs boson
CSV(Higgs jet 1)	CSVv2 output of the hardest jet assigned to the Higgs boson
$\log p_T(\text{light jet})/\text{GeV}$	Transverse momentum of the light-flavored jet
$\log p_T(\text{Higgs})/\text{GeV}$	Transverse momentum of the Higgs boson
$\lvert\eta(t) - \eta(H)\rvert$	Absolute difference of pseudorapidities of reconstructed top quark and reconstructed Higgs boson
$\cos\theta(b_t, \ell)$	Cosine of the angle between the jet assigned to the bottom quark from the top quark decay and the charged lepton

[a] Further details on the Fox-Wolfram moments are described in Reference [129].



(a) (b)

Figure 5.23.: Comparison of the event classification BDT output between simulation and data. The output of the training is shown in the 2 tag control region for the SM (a) and for the ITC scenario (b). The simulated events are scaled to data. In both distributions, good agreement between data and simulated events is observed.



(a) (b)

Figure 5.24.: Comparison between the ROC curves of the SM and the ITC scenario. The ROC curves are shown for both scenarios in the 3 tag (a) and in the 4 tag region (b). A high efficiency can be observed in the two signal regions. Due to the effectively smaller number of events, a slightly worse efficiency is obtained for the SM case.

5.8. Systematic Uncertainties

Several sources of uncertainties, which can affect both simulated events and data, influence the sensitivity of the analysis. These uncertainties are divided into statistical (stat.) and systematic uncertainties. The systematic uncertainties are either of experimental (exp.) nature, for instance a limited energy resolution in the detector, or of theoretical (theo.) nature, e. g. the uncertainty on the production cross sections for specific processes. This section introduces all sources of uncertainties considered in the analysis, which have an effect on the event rate and/or on the shape of certain distributions.

Luminosity (exp., rate) An overall uncertainty on the luminosity measurement at $\sqrt{s} = 13\,\text{TeV}$ of 2.7 % [130] is assigned to all processes in the two signal regions.

Lepton Efficiencies (exp., rate) In order to account for uncertainties in the estimation of the lepton efficiencies, described in Section 5.5, a conservative overall uncertainty of 2 % is applied to all processes.

Pileup (exp., shape and rate) To estimate the uncertainty introduced by reweighting the distribution of the number of primary vertices, the cross section which is used to predict the number of pileup interactions in the simulated events is varied by ± 5 % from its nominal value of 69 mb [131, 132].

Jet Energy Resolution (exp., shape and rate) By increasing and decreasing the difference between the reconstructed jet energy and the true jet energy on particle level, the uncertainty that covers the jet energy resolution is evaluated for the generated jets. The scale factors and their uncertainties which are applied in this smearing process are given in Reference [133]. For the systematically changed simulated samples, the complete analysis chain (overview given in Figure 5.1) is repeated.

Jet Energy Scale (exp., shape and rate) The applied jet energy corrections (see Section 5.5) are varied within the provided uncertainties [134]. The complete analysis chain is repeated for these systematically shifted samples.

Unclustered Energy (exp., shape and rate) As described in Reference [135], the contributions of unclustered particles to $\not{E}_\text{T}$ are varied within their respective energy resolutions.

CSV Reweighting (exp., shape and rate) Various uncertainty sources are considered in the CSV reweighting procedure, described in Section 5.5, and are

treated as uncorrelated uncertainties. If the jet energy scale is changed according to its uncertainties, the change of the b tagging scale factors is re-evaluated and is considered as fully correlated to the shift of the energy scale. The impurity of the sample from which the scale factors were derived is taken into account as a second uncertainty source. Another source, namely the impact of statistical uncertainties during the determination of the scale factors, is propagated to an alternative set of scale factors. The statistical impact is described by two different nuisance parameters, which both have control over distortions in the CSV distribution. All of the uncertainties mentioned above are considered separately for heavy-flavored and light-flavored jets and are taken to be fully uncorrelated. Furthermore, two sets of weights are applied in order to change the contamination of charm jets in the samples which are used for the scale factor determination. A more detailed description of the uncertainties considered in the scale factor determination is given in Reference [125].

$\mathrm{t\bar{t}}$ + Heavy Flavor Rates (theo., rate) As described in Section 5.2.1, the $\mathrm{t\bar{t}}$ sample is split according to the flavor of the additionally produced jets. Thus, an independent scaling of these templates is enabled. Since no measurement of the normalization of the different heavy-flavor templates in control regions has achieved a precision better than 50 %, this conservative number is chosen as rate uncertainty for the $\mathrm{t\bar{t}} + \mathrm{1b}$, $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$, $\mathrm{t\bar{t}} + \mathrm{2b}$ and $\mathrm{t\bar{t}} + \mathrm{c\bar{c}}$ samples.

Q^2 Scale (theo., shape and rate) By using dedicated LHE weights, events in a simulated sample can be reweighted such that they emulate a sample produced with a varied Q^2 scale at matrix element level. These weights are available for all simulated samples but the single top sample. An uncorrelated Q^2 scale uncertainty is introduced for each process by reweighting the events in the final event classification output. The reweighted samples then correspond to a Q^2 scale of fourfold and quarter of the initial value. This uncertainty has a large impact on the shape of the event classification output and on the normalization of the individual processes. In case of the $\mathrm{t\bar{t}}$ process, the Q^2 scale uncertainty is determined by using dedicated samples which have been produced with different Q^2 scales in the parton shower (again fourfold and quarter of the initial Q^2 value). Hence, the Q^2 scale at matrix element and at parton level are varied simultaneously. As aforementioned, the LHE weights needed for the implementation of the Q^2 uncertainty are not available for the single top sample. Hence, a rate uncertainty of 4.0 % is assigned in analogy to the PDF rate uncertainties to cover the effect of the Q^2 scale variation.

Table 5.8.: Cross section uncertainties based on the choice of the PDF set. For three different production mechanisms, the uncertainty values obtained from References [138–141] are applied.

Process	gg in %	$q\bar{q}$ in %	qg in %
tHq			3.7
tHW			4.0
$t\bar{t}H$	3.6		
$t\bar{t}$	3.0		
$t\bar{t}V$		2.0	
Single top			4.0
W + jets		4.0	

PDF Scale (theo., rate) In Table 5.8, the applied uncertainties which affect the normalization of the different simulated processes in dependence of the chosen PDF set are listed. Uncertainties for processes with a common production mode are considered as fully correlated.

Bin-by-Bin Uncertainties (stat., shape) The finite size of the simulated samples requires an additional uncertainty. By applying the "Barlow-Beeston lite" method [136, 137], which introduces a nuisance parameter in the fit for each bin in each sample and each region, this uncertainty is evaluated. For every introduced nuisance parameter one bin is varied within its uncertainties whereas the other bins are scaled such that the normalization of the complete distribution is kept constant. Thus, hundreds of additional nuisance parameters are introduced, causing a significant increase of the needed computing power. In order to limit the computing time, only bins with a relative uncertainty above 5 % are considered for the evaluation of this uncertainty.

5.9. Exclusion Limits

Before the final upper limits are determined, a Maximum Likelihood Estimate fit of the event classification BDT output is performed for all 51 considered $\kappa_t - \kappa_V$ configurations simultaneously in the 3 tag and 4 tag region. In Figure 5.25, the postfit distributions for the SM scenario ($\kappa_t = \kappa_V = +1$) and for the ITC scenario ($\kappa_t = -1$, $\kappa_V = +1$) are shown. For both scenarios and both signal regions, good agreement between data and simulation is observed after the simulated events have been fitted to the observed events.

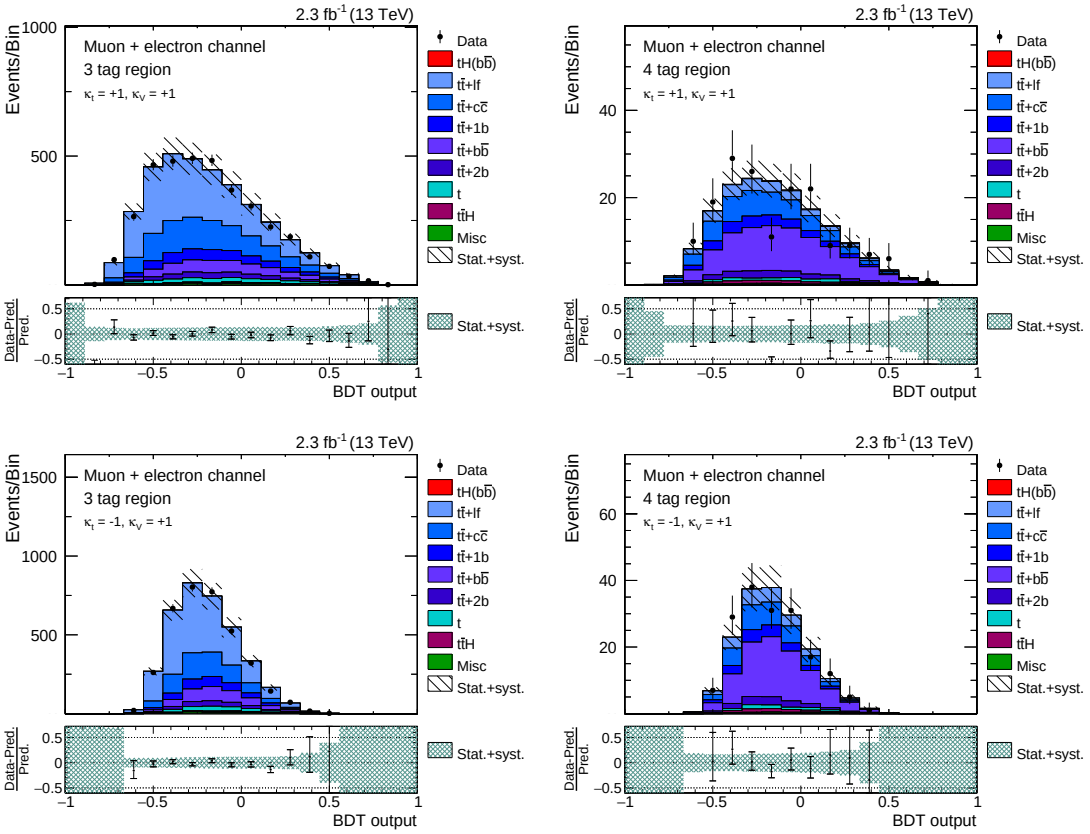


Figure 5.25.: Postfit distributions of the event classification BDT response. The postfit distributions are shown in the 3 tag (left) and 4 tag region (right) for the SM (top) and ITC scenario (bottom). Both statistical and systematic uncertainties are included in the uncertainty bands. Good agreement between data and simulation is observed after the fit.

Additionally, the impact of different uncertainties is analyzed by freezing specific uncertainties to their postfit values and hence effectively removing them from the limit calculation. The impact is determined in two ways: First, the impact of an uncertainty as sole uncertainty source is derived by freezing all other uncertainties to their postfit values. The effect is quantified by taking the relative limit increase on a limit where all uncertainties are frozen. Second, the impact from removing an uncertainty is determined by calculating the relative decrease of a limit with a certain frozen uncertainty with respect to the limit with all uncertainties included. In Figure 5.26, the impact of the uncertainties is illustrated for the ITC scenario. It can be seen that the analysis would benefit the most from a reduction of the jet energy and Q^2 scale uncertainty, followed by the uncertainties related to the CSV reweighting and to the cross section for the production of top quark pairs with additional heavy-flavored jets. The correlation matrix of all considered uncertainties is illustrated in Figure 5.27. No significantly high (anti-)correlations are apparent.

For 51 different $\kappa_t - \kappa_V$ configurations, asymptotic CL_s limits at 95 % CL are calculated with a simultaneous fit in the two signal regions. The determined expected and observed limits are illustrated in Figure 5.28. In order to smoothen the shape of the limit curves, a cubic spline fit has been applied to interpolate between the calculated values. For all points considered in the $\kappa_t - \kappa_V$ plane, the observed limits are well within the one standard deviation uncertainty band of the expected limits. In particular, the calculated values for the SM case and for the ITC case are listed in Table 5.9. The upper limit for the SM tH production is $116.1 \times \sigma_{SM}$ with an expected sensitivity of $99.5 \times \sigma_{SM}$. For the ITC scenario, the observed limit is determined to be $6.2 \times \sigma_{ITC}$ with an expected limit of $6.5 \times \sigma_{ITC}$.

The sensitivity of this analysis is already comparable to that of the Run I analysis [4] which was performed with measured data corresponding to an integrated luminosity of $19.8\,\text{fb}^{-1}$ and yielded an expected limit of $5.2 \times \sigma_{ITC}$ for the ITC scenario. This similar sensitivity achieved with only one-tenth of the measured luminosity can be explained by the increase of the cross section of the tH production, the inclusion of the tHW signal process, and a higher selection efficiency due to the use of the medium instead of the tight b tagging working point.

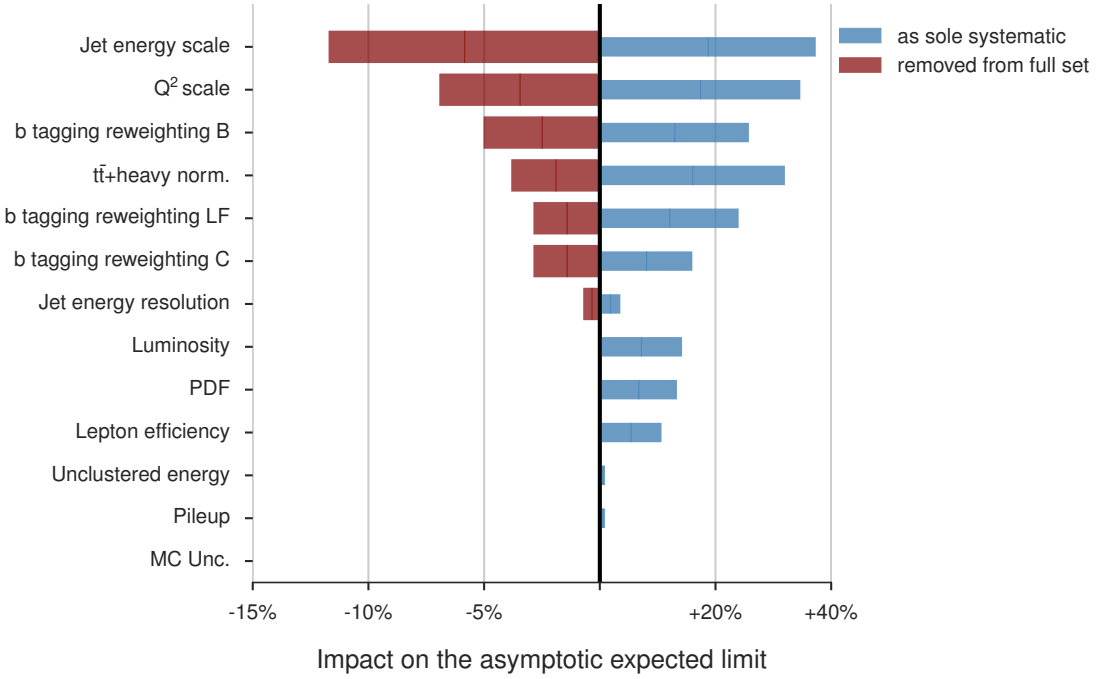


Figure 5.26.: Impact of systematic uncertainties on the expected asymptotic limit. The systematic uncertainties are either removed from the full uncertainty set by fixing them to their postfit values, or used as sole systematic uncertainty by fixing all other uncertainties to their postfit values. The changes shown in the diagram are determined relatively to the limit with all systematic uncertainties included (red bars) and to the limit where all uncertainties are fixed to their best postfit values (blue bars).

Table 5.9.: Expected and observed asymptotic CL_s limits. The limits at 95 % CL obtained from a simultaneous fit in the 3 tag and 4 tag region are denoted for the SM and for the ITC scenario. In addition, the 68 % and 95 % uncertainty values are shown. For both scenarios, good agreement between observed and expected limits is observed. The values of the observed limits are well within one standard deviation uncertainty of the expected limits. The calculated limits for all $\kappa_t - \kappa_V$ configurations are listed in Appendix C.1.

Scenario	Observed Limit	Expected Limit		
		Median	$\pm1\sigma$	$\pm2\sigma$
SM	116.1	99.5	[64.8, 162.2]	[45.7, 257.6]
ITC	6.2	6.5	[4.3, 10.3]	[3.1, 16.1]

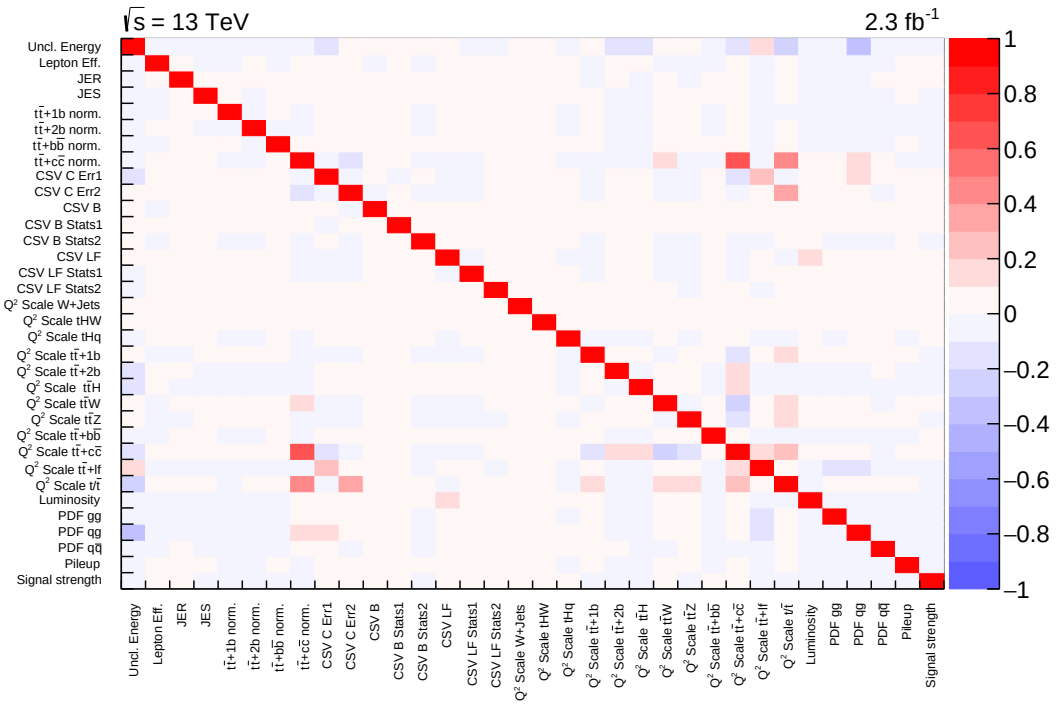
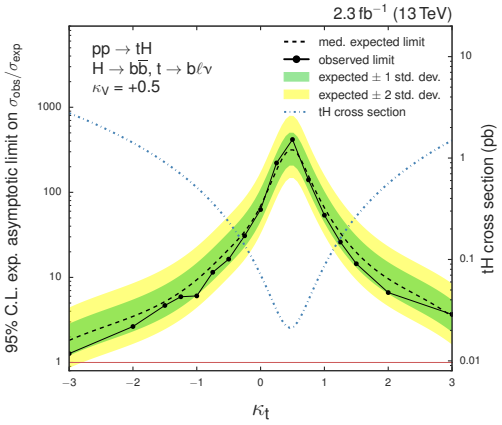
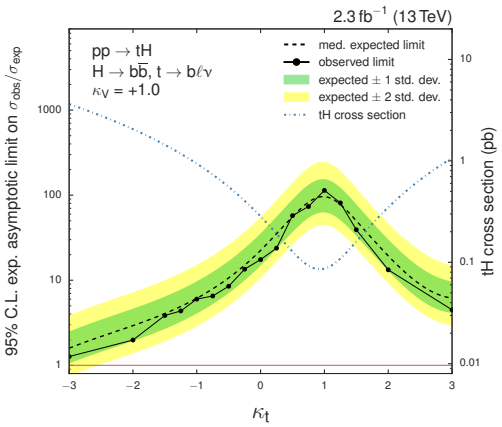


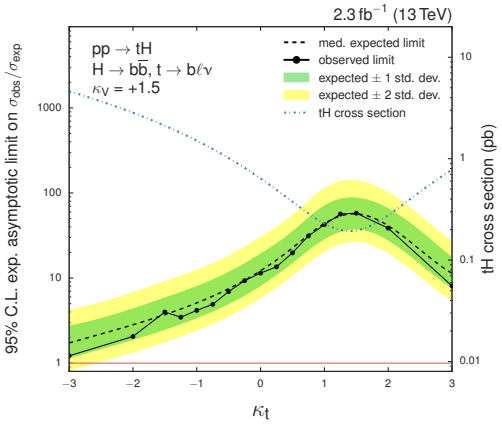
Figure 5.27.: Correlation matrix of all uncertainty sources. The correlation matrix is obtained from the fit. No significantly high (anti-)correlations between the uncertainty sources can be observed.



(a)



(b)



(c)

Figure 5.28.: Upper limits on tH scenarios with different $\kappa_t - \kappa_V$ configurations. The calculated limits are shown as a function of κ_t for $\kappa_V = +0.5$ (a), $\kappa_V = +1$ (b) and $\kappa_V = +1.5$ (c). Neither an excess nor a strong downward fluctuation can be observed. Additionally, the sum of the tHq and tHW cross sections are given by a blue dotted line.

5.10. Analysis with Flavor Classification of $t\bar{t}$ Background

As already stated in Section 5.9, one of the major impacts of the systematic uncertainties on the upper limits that need to be reduced is given by the b tagging reweighting and the normalization of the $t\bar{t}$ process in addition with heavy-flavored jets. For this purpose, an additional classification of the $t\bar{t}$ process based on the original flavor of the additional jets is performed in order to constrain these impacts and thus improve the upper limits. The requirements that are needed for this flavor classification are described in the following paragraphs.

Event Selection for Di-leptonic Control Region

In order to properly separate $t\bar{t} + lf$ events from the $t\bar{t} + 1b$, $t\bar{t} + 2b$ and $t\bar{t} + b\bar{b}$ events, a new control region is introduced which is dominated by di-leptonically decaying top quark pairs. As the final-state of the di-leptonically decaying $t\bar{t}$ process only consists of charged leptons, neutrinos and two b jets, a better separation power for the additional jets originating from initial or final state radiation is obtained. A representative Feynman diagram for the di-leptonic $t\bar{t}$ production with additional jets is illustrated in Figure 5.29.

Several selection criteria are applied in order to obtain a di-leptonic region that contains the same number of di-leptonic $t\bar{t}$ events with additional light-flavored, charm-flavored and bottom-flavored jets. Similarly to Section 5.4, events containing at least two electrons, two muons or one muon and one electron in the final state are considered and the analysis is performed in the combined "muon+electron" channel. By applying five different triggers, which are listed in Table 5.10, the corresponding events are selected. In the final state of di-leptonically decaying $t\bar{t}$ events, one of the two leptons is considered as a tight lepton, while the other lepton corresponds to a loose lepton. In order to gain more events, not exactly but at least one tight lepton and at least two loose leptons are required for the event selection. Furthermore, an overall cut on the missing transverse energy of 45 GeV is applied to reject QCD multijet events.

Since the final-state of the di-leptonic $t\bar{t}$ process consists of at least two bottom quarks, at least two b-tagged jets are required (CSVv2 > 0.8). Additionally, at least three jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ need to fulfill the loose working

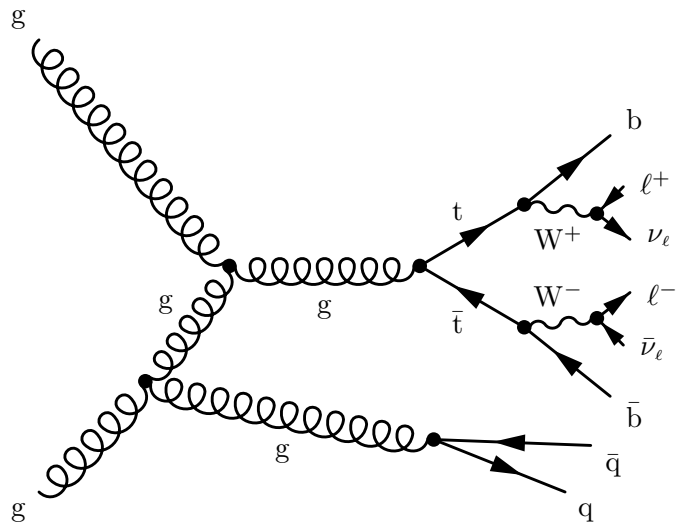


Figure 5.29.: The di-leptonic $t\bar{t}$ production with additional jets. The final state consists of up to four quarks, namely two bottom quarks each originating from the decay of the top quark and up to two quarks which are produced via initial state radiation. These quarks can be light-flavored (i. e. up, down or strange quarks), charm quarks or bottom quarks.

Table 5.10.: Triggers for the selection of di-leptonic events. For events with two electrons in the final state, one trigger is applied, while for events with two muons or one electron and one muon in the final state, two different triggers are required.

Type	Trigger
e e	`HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_vX`
e μ	`HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_vX`
e μ	`HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL_vX`
μ μ	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_vX`
μ μ	`HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_vX`

Table 5.11.: Overview of the event selection criteria for the di-leptonic control region. Several requirements are applied in order to determine the di-leptonic region. More details on the different b tagging working points can be found in Reference [92].

Criterion	Cut value
# medium b-tagged jets ($p_T > 30\,\text{GeV}$, $\lvert\eta\rvert < 2.4$, CSVv2 > 0.8)	≥ 2
# loose b-tagged jets ($p_T > 30\,\text{GeV}$, $\lvert\eta\rvert < 2.4$, CSVv2 > 0.46)	≥ 3
# tight leptons	≥ 1
# additional loose leptons	≥ 2
$\not{E}_T$ (muon+ electron channel)	$> 45\,\text{GeV}$

point of the CSVv2 algorithm (CSVv2 > 0.46) in order to obtain a comparable number of events for the $t\bar{t} + c\bar{c}$, $t\bar{t} + lf$ and for the $t\bar{t}$ processes that contain additional b jets.

The complete event selection for the di-leptonic region is summarized in Table 5.11. Table 5.12 gives an overview of the expected and observed event yields after applying the event selection.

Modified Lepton Efficiency Scale Factors

As already described in Section 5.5, lepton efficiency scale factors need to be applied in order to correct the observed discrepancy between data and simulation. In case of events that contain two instead of one lepton in their final state, dedicated trigger efficiency scale factors need to be calculated for events with two electrons, two muons and with one muon and one electron. Since these efficiency scale factors are not provided by the Muon and EGamma POG, privately produced correction factors are used. For the muon and electron identification and isolation efficiencies, the same correction factors are applied as for the single lepton events. A slight reduction of the overall yield and a small change in the shape for some distributions is caused by the lepton efficiency scale factors.

Table 5.12.: Event yields in the di-leptonic control region. The events yields for the background processes and the tHq and tHW signal processes (SM and ITC scenarios) are denoted. The uncertainties include both statistical and systematic uncertainties. In addition, the observed event yields in data are shown.

	3 tag
$t\bar{t} + lf$	138 ± 32
$t\bar{t} + c\bar{c}$	83 ± 48
$t\bar{t} + 1b$	27.7 ± 12.4
$t\bar{t} + b\bar{b}$	34 ± 21
$t\bar{t} + 2b$	14.8 ± 8.1
Single top	9.2 ± 2.1
$t\bar{t}H$	2.0 ± 0.8
$t\bar{t}Z$	4.4 ± 0.8
$t\bar{t}W$	1.8 ± 0.4
$W + jets$	2.0 ± 1.0
$Z + jets$	12.8 ± 3.7
Sum of backgrounds	330 ± 63
tHq (SM)	0.02 ± 0.01
tHW (SM)	0.07 ± 0.01
tHq (ITC)	0.2 ± 0.1
tHW (ITC)	0.8 ± 0.1
Observed	324

Flavor Classification

Similarly to the event classification, a dedicated BDT is trained in order to discriminate between $t\bar{t}$ events with additional light-flavored jets from $t\bar{t} + 1b$, $t\bar{t} + 2b$ and $t\bar{t} + b\bar{b}$ events. Since the kinematics for the top quark pair production do not depend on the chosen $\kappa_t - \kappa_V$ configuration, one BDT is sufficient for all 51 considered points in the $\kappa_t - \kappa_V$ plane.

For the training of the $t\bar{t}$ flavor classification BDT, which is performed in the di-leptonic region, 10 input variables are used. A description of these variables can be found in Table 5.13. In order to properly separate the $t\bar{t} + c\bar{c}$ process, which is neither considered as signal nor as background in the training, from the other processes, a dedicated charm-jet tagger [142–144] is used. The purpose of this c tagging algorithm, which is based on the CSV algorithm, is the discrimination of jets originating from a charm quark, from b jets and from light-flavored jets. As the properties of c jets are often distributed in between of those of b jets and light-flavored jets, a distinction between "charm-vs-light" (CvsL) and "charm-vs-bottom" (CvsB) is required. The most important variables are the CSVv2 output of the b jet which possesses the highest CSVv2 value, followed by the CvsL output of the third hardest jet. Neither the CSVv2 output of the two highest b jets nor the CvsL and CvsB outputs of the two hardest jets are considered in this training, as these jets mostly correspond to the two b jets of the top quark pair decay which are apparent in all considered di-leptonic $t\bar{t}$ samples. The distributions for the six most important variables of the flavor classification can be found in Figure 5.30. The remaining four variables are shown in Appendix B.6.

The chosen parameter settings of the $t\bar{t}$ flavor classification BDT are listed in Table 5.14. In Figure 5.31, the response of the flavor classification BDT for the training and the test sample can be seen. A good separation between events with additional light-flavored jets and events with b jets can be observed and no signs of overtraining are found.

Table 5.13.: Input variables for the flavor classification BDT. The variables are sorted by their importance in the training within each category. In total, 10 variables are used for the training of the BDT.

Variable	Description
CSV(b jet 3)	CSVv2 output of b jet with third highest CSVv2 value
CvsL(jet 3)	CvsL output of third hardest jet
CSV(b jet 2)	CSVv2 output of b jet with second highest CSVv2 value
CvsB(jet 3)	CvsB output of third hardest jet
CvsL(jet 4)	CvsL output of fourth hardest jet
CSV(b jet 4)	CSVv2 output of b jet with fourth highest CSVv2 value
n(jets)	Total number of jets in the final-state
n(jets, loose)	Number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ and CSVv2 > 0.46
n(jets, tight)	Number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ and CSVv2 > 0.935
n(jets, medium)	Number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ and CSVv2 > 0.8

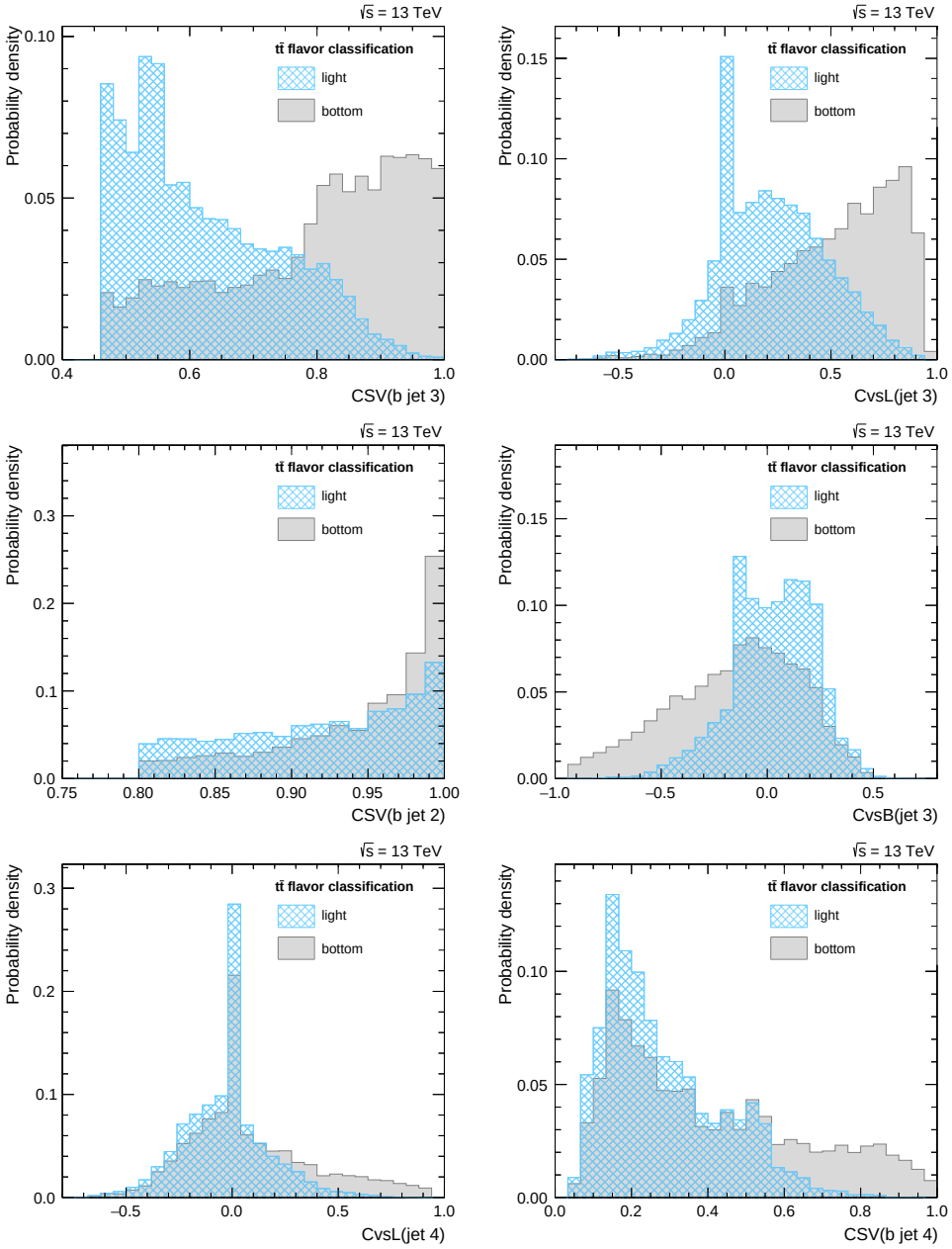


Figure 5.30.: The distributions of the six most discriminating variables of the flavor classification. The distributions of events with light-flavored and with heavy-flavored jets for each variable are shown. The variables are sorted by their importance in the training. The description of the variables is listed in Table 5.13. The remaining variables can be found in Appendix B.6.

Table 5.14.: Parameters used for the training of the flavor classification BDT. Due to the small number of events, a small number of trees has been chosen. Details on the parameters can be found in Reference [106].

Parameter	Value
`NTrees`	150
`MinNodeSize`	2
`MaxDepth`	2
`BoostType`	AdaBoost
`nCuts`	16
`AdaBoostBeta`	0.3
`SeparationType`	GiniIndex

After the successful training, the flavor classification response is applied to data and simulation events in the di-leptonic region. Reasonable agreement, considering the relatively small number of events, is observed in Figure 5.32(a). The shape of the flavor classification BDT output for the semi- and di-leptonically decaying top quark pairs is illustrated in Figure 5.32(b). A good separation of the $t\bar{t} + c\bar{c}$ sample from the other $t\bar{t}$ samples is observed. Additionally, the ROC curve is illustrated in Figure 5.33. A good separation is achieved for the flavor classification.

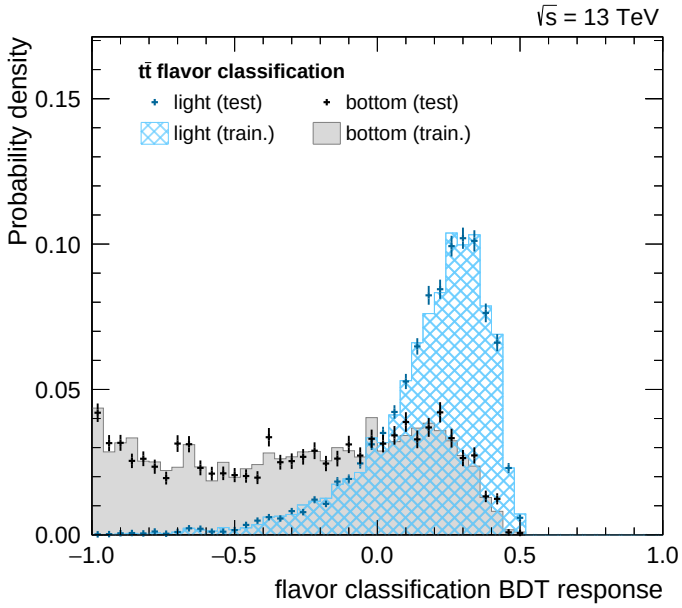
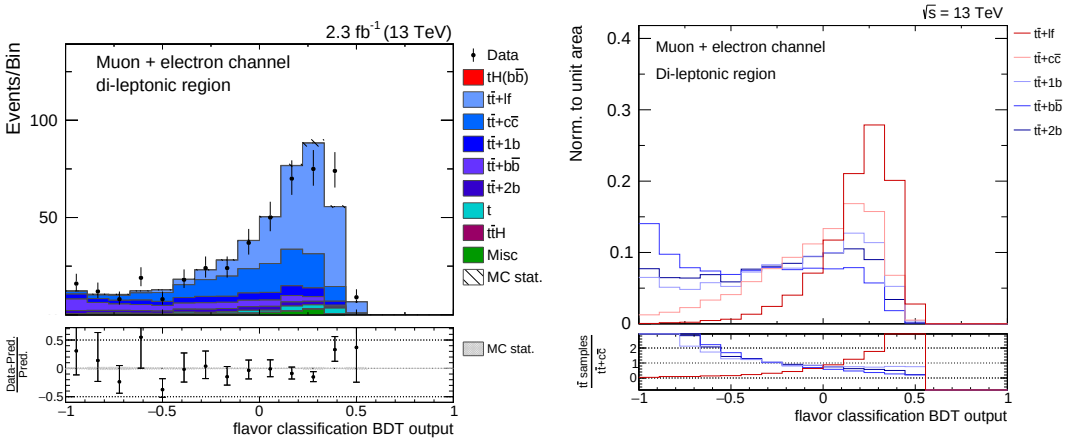


Figure 5.31.: Output value of the flavor classification BDT. A clear separation between additional light-flavored jets and jets that originate from bottom quarks can be observed. Furthermore, no signs of overtraining can be found.



(a) (b)

Figure 5.32.: Comparison of the flavor classification BDT output. The output of the training is shown in the di-leptonic control region (a). The simulated events are scaled to data. A reasonable agreement between data and simulated events is observed. The shape of the flavor classification BDT output for the $t\bar{t} + c\bar{c}$ process significantly differs from the $t\bar{t} + lf$ and the $t\bar{t}$ plus additional b jets distributions (b).

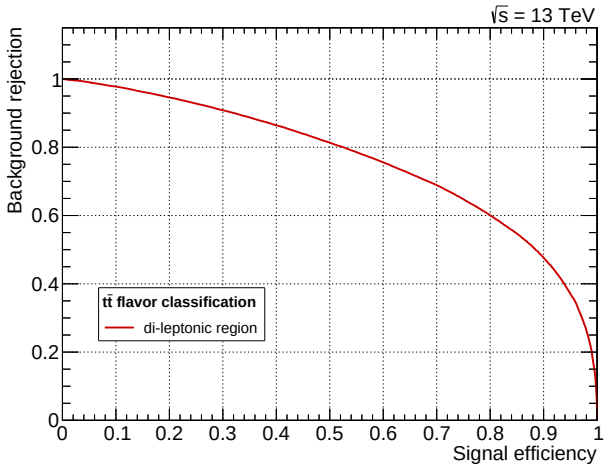


Figure 5.33.: ROC curve of the flavor classification BDT. The ROC curve is shown for the di-leptonic region. A high efficiency, corresponding to a ROC value of 83.4 %, can be observed.

Systematic Uncertainties and Exclusion Limits

The same systematic uncertainties used in the analysis without flavor classification and described in Section 5.8 are applied to the samples in the di-leptonic region. The corresponding correlation matrix is shown in Figure 5.34. No significantly high (anti-)correlations between the parameters are observed. Analogously to Section 5.9, a Maximum Likelihood Estimate fit of the flavor classification output is performed. The outcome can be seen in Figure 5.35. Reasonable agreement is observed after the fit. The upper limits for the SM and ITC scenario are then calculated by a simultaneous fit of the event classification BDT output in the 3 tag and 4 tag region and a fit of the flavor classification BDT output in the di-leptonic region.

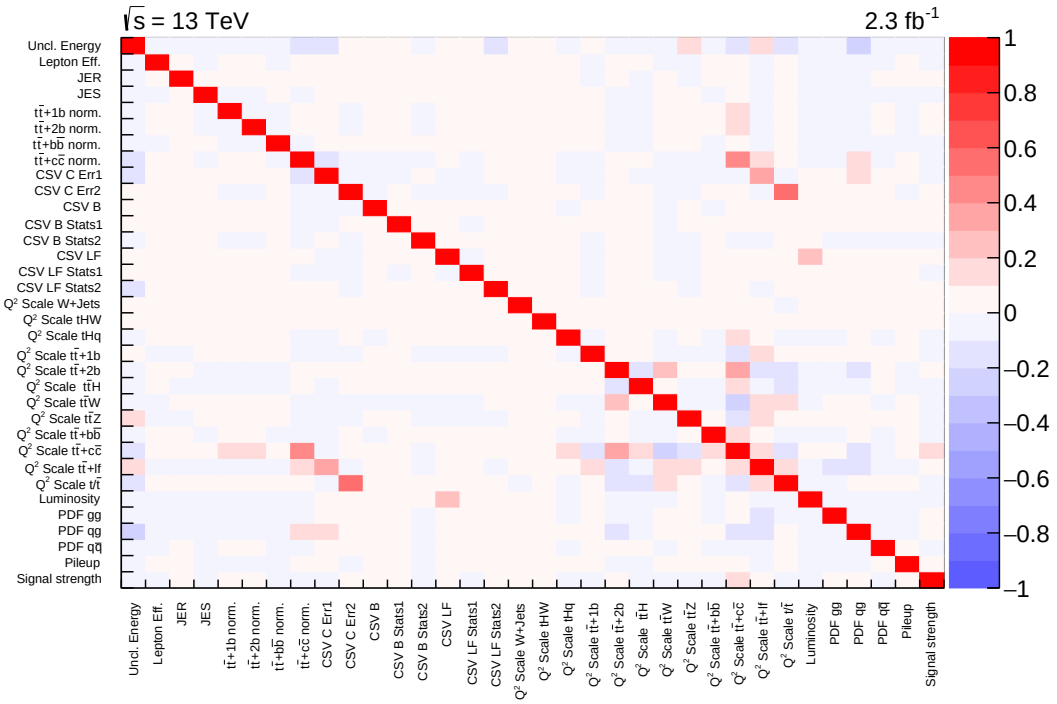


Figure 5.34.: Correlation matrix of all uncertainty sources. The correlation matrix is obtained from the fit. No significantly high (anti-)correlations between the uncertainty sources can be observed.

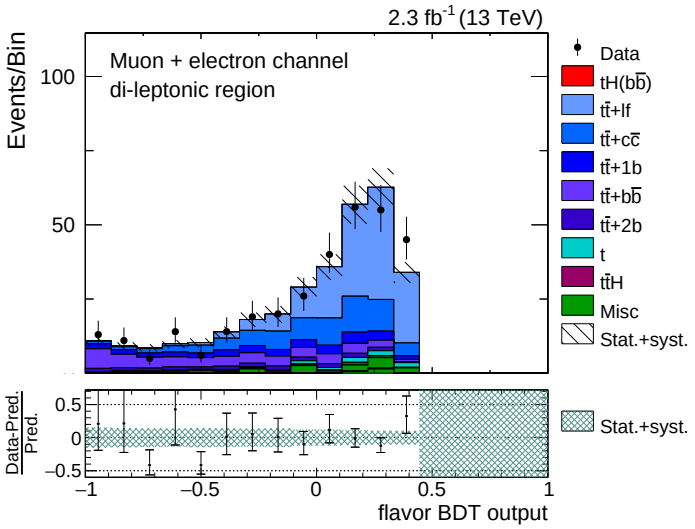
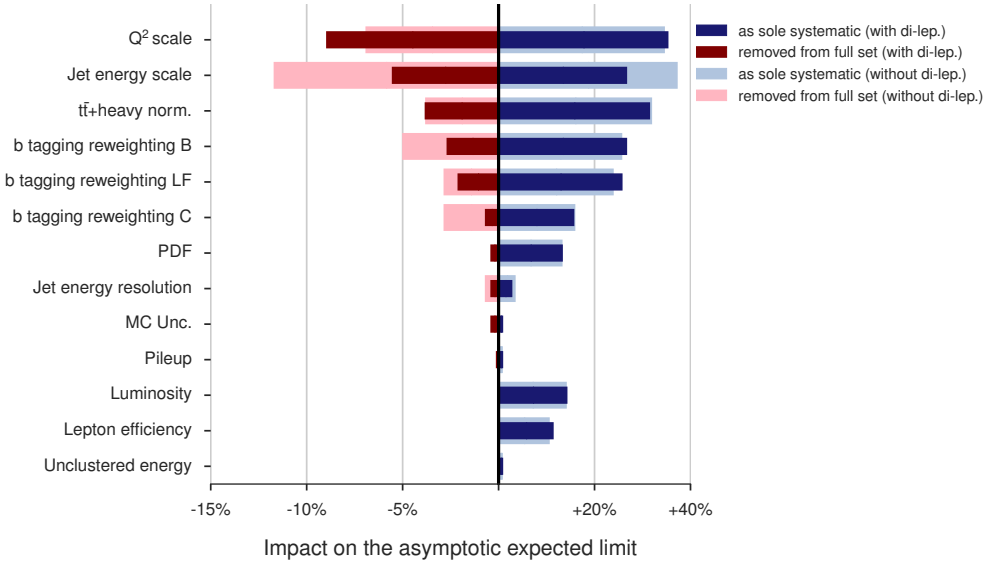


Figure 5.35.: Postfit distribution of the flavor classification BDT response. Both statistical and systematic uncertainties are included in the uncertainty bands. Reasonable agreement between data and simulation is observed after the fit.

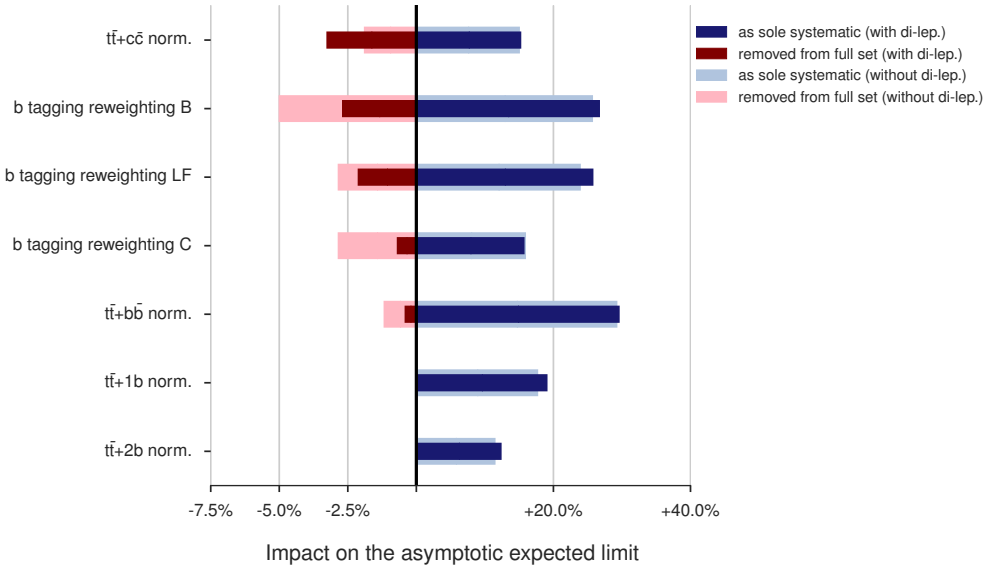
5.11. Result Comparison

After performing the limit calculation with the additional fit of the flavor classification BDT output in the di-leptonic region, the results need to be evaluated and compared to the formerly described analysis.

In Figure 5.36(a), the impact of all uncertainties considered in the analyses is compared. The dark bars correspond to the analysis including the additional flavor classification, the bright bars to the analysis without flavor classification. Especially if certain uncertainties are removed from the full set, significant differences between the two analyses arise. Due to a smaller number of jets and a better handle of the additionally produced jets in the di-leptonic region, a drastic reduction of the jet energy scale uncertainty is observed. The separation of $t\bar{t}$ events with additional light-flavored jets from $t\bar{t}$ events with additional b jets via the flavor classification BDT leads to a smaller impact of the b tagging reweighting uncertainties. The increased influence of the Q^2 scale uncertainty can be explained by the fact that this uncertainty consist of several parameters which seem to be slightly more correlated when the di-leptonic region is included in the limit calculation (see also Figure 5.27 and 5.34). These small correlations may sum up to a significant increase of the total Q^2 scale uncertainty.



(a)



(b)

Figure 5.36.: Comparison of the impact of the systematic uncertainties on the expected limit. The impact is directly compared for the limit calculation with (dark bars) and without (bright bars) including the di-leptonic region. In particular, significant differences can be seen, when certain uncertainties are removed from the full set: A drastic reduction of the jet energy scale and of the b tagging reweighting uncertainties is observed when including the di-leptonic region (a). The impact of the Q^2 scale uncertainty increases significantly. The separated $t\bar{t}$ and the b tagging uncertainties are shown (b). A decrease for the $t\bar{t} + b\bar{b}$ process is seen, whereas the $t\bar{t} + c\bar{c}$ uncertainty increases.

Table 5.15.: Comparison of expected and observed asymptotic CL_s limits. The limits at 95 % CL are compared for the limit calculation with and without including the di-leptonic region. The results are denoted for the SM and for the ITC scenario. In addition, the 68 % and 95 % uncertainty values are shown. By including the di-leptonic region, the expected limit is improved by 8.2 % (SM) and 15.4 % (ITC), respectively.

Scenario	Observed Limit	Expected Limit		
		Median	$\pm1\sigma$	$\pm2\sigma$
Fit in 3 tag and 4 tag region only				
SM	116.1	99.5	[64.8, 162.2]	[45.7, 257.6]
ITC	6.2	6.5	[4.3, 10.3]	[3.1, 16.1]
Fit including di-leptonic region				
SM	102.1	91.3	[60.7, 144.0]	[43.3, 224.3]
ITC	4.7	5.5	[3.7, 8.5]	[2.6, 13.0]

No change of the impact seems to be observed for the uncertainty of the $t\bar{t}+$ heavy flavor normalization although a dedicated separation of the di-leptonic $t\bar{t}$ events has been performed. This issue can be resolved by splitting this uncertainty group, as seen in Figure 5.36(b). It can be seen that the impact of the $t\bar{t}+b\bar{b}$ normalization decreases, while the one of the $t\bar{t}+c\bar{c}$ normalization increases. One possible reason for this increase is that the $t\bar{t}+c\bar{c}$ process has not been directly considered in the training of the flavor classification BDT. Thus, its uncertainty cannot be constrained in contrast to the uncertainty of the $t\bar{t}+b\bar{b}$ process.

In Table 5.15, an overview of the determined expected and observed limits for the SM and the ITC scenario is given for both analyses. In both analyses and for both scenarios, the observed limits are well within one standard deviation of the expected limit. The additional flavor classification leads to an expected limit of $91.3 \times \sigma_{SM}$ for the SM case and $5.5 \times \sigma_{ITC}$ for the ITC case. These results correspond to an improvement of the expected limits of 8.2 % (SM) and 15.4 % (ITC), respectively.

5.12. Projection to a Higher Luminosity

Since the CMS experiment collected data in 2016, an integrated luminosity significantly higher than $2.3\,\mathrm{fb}^{-1}$, measured in 2015, has been obtained. This leads to an increase of the event yields and thus to a reduction of the expected upper limits. In order to validate the improvements described in Section 5.10 and to demonstrate the increase of the analysis sensitivity, a projection to a higher luminosity of $25\,\mathrm{fb}^{-1}$ is performed for the SM and the ITC scenario.

Assuming the same systematic shape and rate uncertainties as given in Section 5.8 and using an Asimov data set (see also Section 4.3), the expected asymptotic $\mathrm{CL_s}$ limits are determined based on the simulation information only. The results are listed in Table 5.16. For the analysis which does not include the di-leptonic region, an upper limit of $42.3 \times \sigma_{\mathrm{SM}}$ in case of the SM scenario and of $2.8 \times \sigma_{\mathrm{ITC}}$ for the ITC case is expected. This corresponds to an improvement of 57 % with regard to the same analysis performed at $2.3\,\mathrm{fb}^{-1}$. The optimized analysis provides an improvement of 63 % when compared to the same analysis performed at $2.3\,\mathrm{fb}^{-1}$. With regard to the analysis without flavor classification, the calculated expected limits of $33.5 \times \sigma_{\mathrm{SM}}$ and $2.0 \times \sigma_{\mathrm{ITC}}$ of the optimized analysis correspond to an improvement of 20.8 % for the SM and 28.6 % for the ITC scenario, respectively. Furthermore, the sensitivity obtained with the improved analysis is almost sufficient to exclude the ITC scenario as the bottom edge of the 95 % uncertainty band is given by $1.0 \times \sigma_{\mathrm{ITC}}$.

It has to be noted that this projection uses a single JES nuisance parameter, which includes all JES uncertainty sources, to be comparable to the analysis performed with an integrated luminosity of $2.3\,\mathrm{fb}^{-1}$. However, for future analyses that reach a higher sensitivity it will be necessary to split up this nuisance parameter according to the different JES uncertainty sources available. These sources need to be treated separately in order to determine the correlations between the various sources and to estimate their impact on the exclusion limits. Thus, a more precise description of the JES uncertainty will be achieved.

Table 5.16.: Comparison of expected asymptotic CL_s limits at a luminosity of 25 fb^{-1}. The limits at 95 % CL are compared for the limit calculation with and without including the di-leptonic region. The results are denoted for the SM and for the ITC scenario. In addition, the 68 % and 95 % uncertainty values are shown. By including the di-leptonic region, the expected limit is improved by 20.8 % (SM) and 28.6 % (ITC), respectively.

Scenario	Expected Limit		
	Median	$\pm1\sigma$	$\pm2\sigma$
Fit in 3 tag and 4 tag region only			
SM	42.3	[27.7, 67.5]	[19.7, 108.0]
ITC	2.8	[1.8, 4.4]	[1.3, 7.0]
Fit including di-leptonic region			
SM	33.5	[22.9, 51.3]	[16.7, 77.9]
ITC	2.0	[1.4, 3.1]	[1.0, 4.6]

6. Conclusion

Since the discovery of a new boson in 2012, it is crucial to measure its properties precisely for answering the question whether the particle is really the Higgs boson predicted by the Standard Model (SM). Small deviations from SM predictions could point to physics beyond the Standard Model (BSM). For instance, such deviations can occur for the coupling strengths of the Higgs boson to top quarks (κ_t) and to massive vector bosons (κ_V). As the single top quark production in association with a Higgs boson (tH) is sensitive to the sign and the magnitude of these two couplings, the search for this rare process enables the study of different $\kappa_t - \kappa_V$ configurations. In this thesis, the decay of the Higgs boson into a bottom quark pair is considered for the search for the tH process. The 2015 data set of the CMS experiment, recorded at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and corresponding to an integrated luminosity of $2.3\,\text{fb}^{-1}$, has been analyzed.

In this analysis, based on the published analysis of the CMS collaboration [3], exclusion limits at 95 % CL were determined for 51 different $\kappa_t - \kappa_V$ configurations. In case of the SM scenario ($\kappa_t = \kappa_V = +1$), the observed upper limit for the tH production was found to be $116.1 \times \sigma_{SM}$ with an expected sensitivity of $99.5 \times \sigma_{SM}$. For the inverted top coupling (ITC) scenario ($\kappa_t = -1$, $\kappa_V = +1$), the observed limit was determined to be $6.2 \times \sigma_{ITC}$ with an expected limit of $6.5 \times \sigma_{ITC}$.

Due to the large contribution of the top quark pair production, it is essential to fully understand this dominant background process. For this purpose, an additional classification of the $t\bar{t}$ process based on the flavor of additionally radiated jets was performed for the SM and the ITC scenario. The expected limits were reduced to $91.3 \times \sigma_{SM}$ and $5.5 \times \sigma_{ITC}$. Compared to the analysis performed without an additional flavor classification, these results corresponds to an improvement of 8.2 % (SM) and 15.4 % (ITC), respectively. It can be concluded that the flavor classification leads to an improved description of the $t\bar{t}$ process and thus to a reduction of the systematic jet energy scale and b tagging reweighting uncertainties.

The amount of the data collected in 2015 is not yet sufficient to exclude any of the analyzed $\kappa_t - \kappa_V$ configurations. Nevertheless, the ongoing Run II of the LHC will provide additional data. Thus, a higher sensitivity and a further reduction of the expected limits will be achieved. In order to investigate the effect of increased event yields, a projection to an integrated luminosity of $25\,\text{fb}^{-1}$, which approximately corresponds to the amount of data measured in 2016, was performed. For the analysis without additional flavor classification, expected upper limits of $42.3 \times \sigma_{SM}$ and of $2.8 \times \sigma_{ITC}$ were determined. When including the flavor classification, the expected limits were found to be $33.5 \times \sigma_{SM}$ and $2.0 \times \sigma_{ITC}$, which corresponds to an improvement of 20.8 % (SM) and 28.6 % (ITC), respectively. Hence, the achieved sensitivity is almost sufficient to exclude the ITC scenario.

The increase of the center-of-mass energy up to $\sqrt{s} = 14\,\text{TeV}$ will lead to an even higher sensitivity of this analysis. Furthermore, the sensitivity can be improved by a dedicated reconstruction of the second signal process, the tHW production, and by a combination of the $H \to b\bar{b}$ search with analyses performed in other decay channels such as $H \to \gamma\gamma$, $H \to WW$ and $H \to \tau\tau$. As these improvements will bring the sensitivity of the search for the tH production to regions where several points of the $\kappa_t - \kappa_V$ plane can be excluded, this production mode will contribute to the knowledge on the Higgs boson couplings and therefore on the nature of this last discovered SM boson.

A. Signal Cross Sections and Shapes

Table A.1.: Production cross sections for tHq, depending on κ_t and κ_V. Obtained with MadGraph5_aMC@NLO at NLO in the four-flavor scheme. The quoted uncertainties on the cross section correspond to scale variations in %. Additional information can be found in Reference [145].

κ_t	κ_V	σ in pb	κ_t	κ_V	σ in pb	κ_t	κ_V	σ in pb
-3.0	$+0.5$	$2.260^{+1.9}_{-2.7}$	-3.0	$+1.0$	$2.991^{+2.1}_{-3.1}$	-3.0	$+1.5$	$3.845^{+2.6}_{-3.2}$
-2.0	$+0.5$	$1.160^{+2.0}_{-2.9}$	-2.0	$+1.0$	$1.706^{+2.6}_{-3.2}$	-2.0	$+1.5$	$2.371^{+2.5}_{-3.6}$
-1.5	$+0.5$	$0.748^{+2.1}_{-3.1}$	-1.5	$+1.0$	$1.205^{+2.5}_{-3.6}$	-1.5	$+1.5$	$1.784^{+2.7}_{-3.9}$
-1.25	$+0.5$	$0.573^{+2.1}_{-3.0}$	-1.25	$+1.0$	$0.987^{+2.6}_{-3.4}$	-1.25	$+1.5$	$1.518^{+2.8}_{-3.9}$
-1.0	$+0.5$	$0.472^{+2.3}_{-3.3}$	-1.0	$+1.0$	$0.793^{+2.7}_{-3.9}$	-1.0	$+1.5$	$1.287^{+3.0}_{-4.3}$
-0.75	$+0.5$	$0.300^{+2.5}_{-3.5}$	-0.75	$+1.0$	$0.621^{+2.9}_{-4.1}$	-0.75	$+1.5$	$1.067^{+3.1}_{-4.4}$
-0.5	$+0.5$	$0.198^{+2.8}_{-3.9}$	-0.5	$+1.0$	$0.472^{+3.2}_{-4.4}$	-0.5	$+1.5$	$0.874^{+3.4}_{-4.7}$
-0.25	$+0.5$	$0.119^{+3.1}_{-4.6}$	-0.25	$+1.0$	$0.351^{+3.5}_{-5.0}$	-0.25	$+1.5$	$0.703^{+3.6}_{-5.0}$
0.0	$+0.5$	$0.062^{+3.8}_{-5.6}$	0.0	$+1.0$	$0.248^{+3.9}_{-5.5}$	0.0	$+1.5$	$0.558^{+3.8}_{-5.4}$
$+0.25$	$+0.5$	$0.028^{+5.0}_{-7.1}$	$+0.25$	$+1.0$	$0.169^{+4.4}_{-6.2}$	$+0.25$	$+1.5$	$0.437^{+4.2}_{-6.1}$
$+0.5$	$+0.5$	$0.018^{+4.2}_{-6.7}$	$+0.5$	$+1.0$	$0.113^{+5.0}_{-7.1}$	$+0.5$	$+1.5$	$0.334^{+4.6}_{-6.5}$
$+0.75$	$+0.5$	$0.030^{+1.4}_{-2.9}$	$+0.75$	$+1.0$	$0.081^{+5.7}_{-7.6}$	$+0.75$	$+1.5$	$0.256^{+5.2}_{-7.2}$
$+1.0$	$+0.5$	$0.066^{+1.0}_{-3.6}$	$+1.0$	$+1.0$	$0.071^{+4.1}_{-6.7}$	$+1.0$	$+1.5$	$0.200^{+5.7}_{-7.6}$
$+1.25$	$+0.5$	$0.124^{+0.9}_{-3.7}$	$+1.25$	$+1.0$	$0.084^{+2.3}_{-4.6}$	$+1.25$	$+1.5$	$0.167^{+5.5}_{-7.5}$
$+1.5$	$+0.5$	$0.205^{+0.8}_{-3.7}$	$+1.5$	$+1.0$	$0.120^{+1.2}_{-2.9}$	$+1.5$	$+1.5$	$0.159^{+4.1}_{-6.7}$
$+2.0$	$+0.5$	$0.436^{+1.0}_{-3.6}$	$+2.0$	$+1.0$	$0.260^{+1.0}_{-3.6}$	$+2.0$	$+1.5$	$0.211^{+2.0}_{-3.9}$
$+3.0$	$+0.5$	$1.177^{+1.2}_{-3.2}$	$+3.0$	$+1.0$	$0.821^{+0.8}_{-3.7}$	$+3.0$	$+1.5$	$0.589^{+0.9}_{-3.7}$

Table A.2.: Production cross sections for tHW, depending on κ_t and κ_V. Obtained with MadGraph5_aMC@NLO at NLO in the five-flavor scheme. The quoted uncertainties on the cross section correspond to scale variations in %. Additional information can be found in Reference [145].

κ_t	κ_V	σ in pb	κ_t	κ_V	σ in pb	κ_t	κ_V	σ in pb
-3.0	$+0.5$	$0.514^{+2.3}_{-3.0}$	-3.0	$+1.0$	$0.641^{+2.3}_{-2.7}$	-3.0	$+1.5$	$0.783^{+2.1}_{-2.1}$
-2.0	$+0.5$	$0.255^{+2.3}_{-2.8}$	-2.0	$+1.0$	$0.346^{+2.2}_{-2.5}$	-2.0	$+1.5$	$0.457^{+2.1}_{-2.1}$
-1.5	$+0.5$	$0.159^{+2.3}_{-2.8}$	-1.5	$+1.0$	$0.253^{+2.1}_{-2.2}$	-1.5	$+1.5$	$0.329^{+1.9}_{-1.8}$
-1.25	$+0.5$	$0.120^{+2.2}_{-2.5}$	-1.25	$+1.0$	$0.188^{+2.0}_{-2.0}$	-1.25	$+1.5$	$0.275^{+1.9}_{-1.6}$
-1.0	$+0.5$	$0.087^{+2.1}_{-2.3}$	-1.0	$+1.0$	$0.147^{+2.0}_{-1.8}$	-1.0	$+1.5$	$0.224^{+1.9}_{-1.5}$
-0.75	$+0.5$	$0.059^{+2.0}_{-2.1}$	-0.75	$+1.0$	$0.110^{+2.0}_{-1.7}$	-0.75	$+1.5$	$0.180^{+1.8}_{-1.3}$
-0.5	$+0.5$	$0.037^{+1.9}_{-1.8}$	-0.5	$+1.0$	$0.080^{+1.7}_{-1.4}$	-0.5	$+1.5$	$0.141^{+1.6}_{-1.2}$
-0.25	$+0.5$	$0.020^{+1.8}_{-1.3}$	-0.25	$+1.0$	$0.055^{+1.6}_{-1.1}$	-0.25	$+1.5$	$0.108^{+1.6}_{-1.2}$
0.0	$+0.5$	$0.009^{+1.6}_{-1.3}$	0.0	$+1.0$	$0.036^{+1.5}_{-1.2}$	0.0	$+1.5$	$0.081^{+1.5}_{-1.2}$
$+0.25$	$+0.5$	$0.004^{+2.1}_{-2.0}$	$+0.25$	$+1.0$	$0.022^{+1.6}_{-1.5}$	$+0.25$	$+1.5$	$0.059^{+1.5}_{-1.4}$
$+0.5$	$+0.5$	$0.004^{+4.6}_{-6.1}$	$+0.5$	$+1.0$	$0.014^{+2.1}_{-2.0}$	$+0.5$	$+1.5$	$0.043^{+1.8}_{-1.7}$
$+0.75$	$+0.5$	$0.010^{+4.7}_{-6.3}$	$+0.75$	$+1.0$	$0.012^{+3.2}_{-3.9}$	$+0.75$	$+1.5$	$0.033^{+2.1}_{-2.0}$
$+1.0$	$+0.5$	$0.021^{+4.0}_{-5.5}$	$+1.0$	$+1.0$	$0.016^{+4.6}_{-6.1}$	$+1.0$	$+1.5$	$0.028^{+2.8}_{-3.0}$
$+1.25$	$+0.5$	$0.038^{+3.7}_{-5.2}$	$+1.25$	$+1.0$	$0.025^{+4.8}_{-5.4}$	$+1.25$	$+1.5$	$0.029^{+3.6}_{-4.7}$
$+1.5$	$+0.5$	$0.061^{+3.5}_{-4.9}$	$+1.5$	$+1.0$	$0.039^{+4.6}_{-6.3}$	$+1.5$	$+1.5$	$0.035^{+4.6}_{-6.0}$
$+2.0$	$+0.5$	$0.125^{+3.0}_{-4.3}$	$+2.0$	$+1.0$	$0.086^{+4.0}_{-5.5}$	$+2.0$	$+1.5$	$0.065^{+4.8}_{-6.5}$
$+3.0$	$+0.5$	$0.317^{+2.8}_{-4.0}$	$+3.0$	$+1.0$	$0.247^{+3.3}_{-4.6}$	$+3.0$	$+1.5$	$0.193^{+4.0}_{-5.6}$

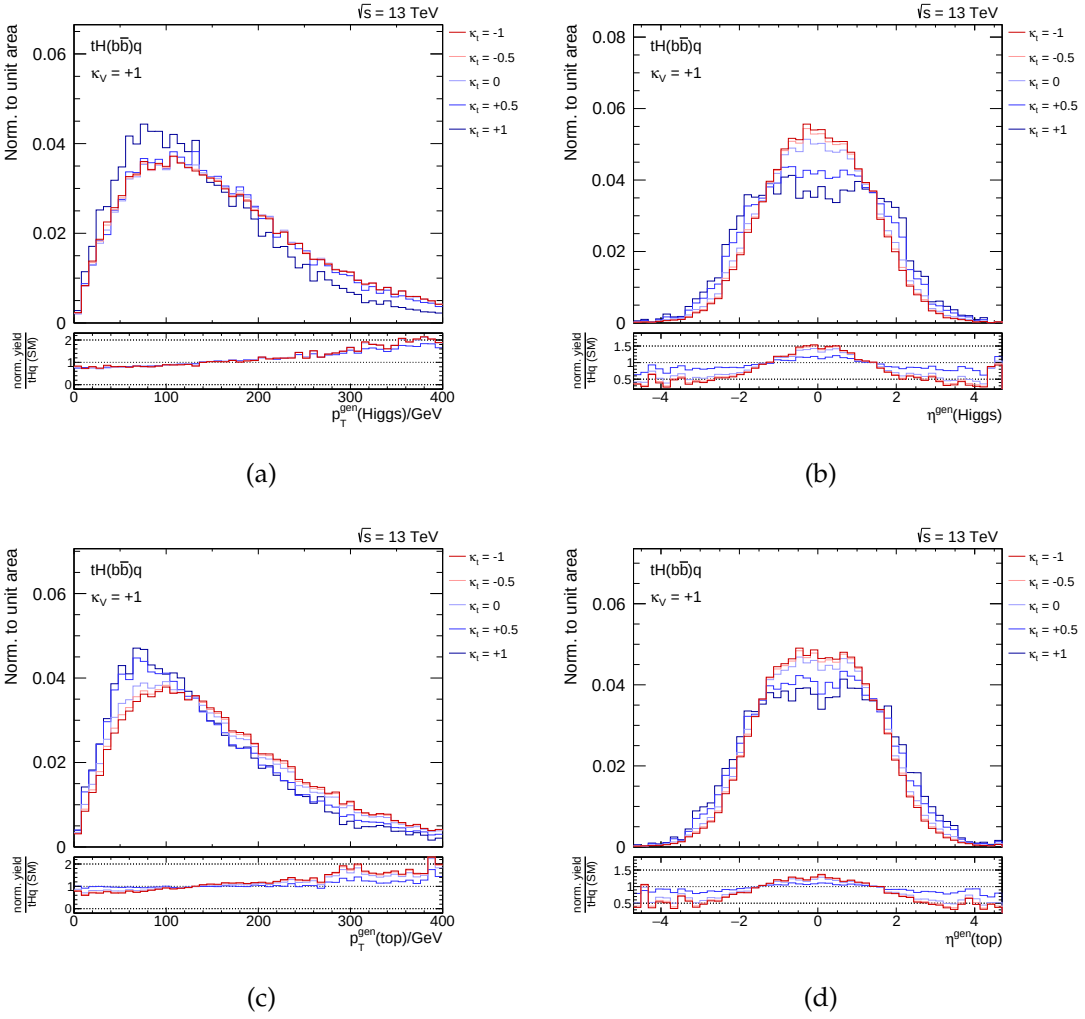


Figure A.1.: Change of the tHq kinematics in dependence of the $\kappa_t - \kappa_V$ configuration. The distributions of Higgs p_T (a), Higgs η (b), and p_T (c) and η (d) of the top quark, produced at generator level including the parton shower, are shown for $-1 \leq \kappa_t \leq +1$ and $\kappa_V = +1$. For higher κ_t values, the p_T distributions are shifted to higher values and more Higgs bosons and top quarks are observed in the central region (i. e. smaller absolute η values).

B. Variables of the Reconstruction and Classification

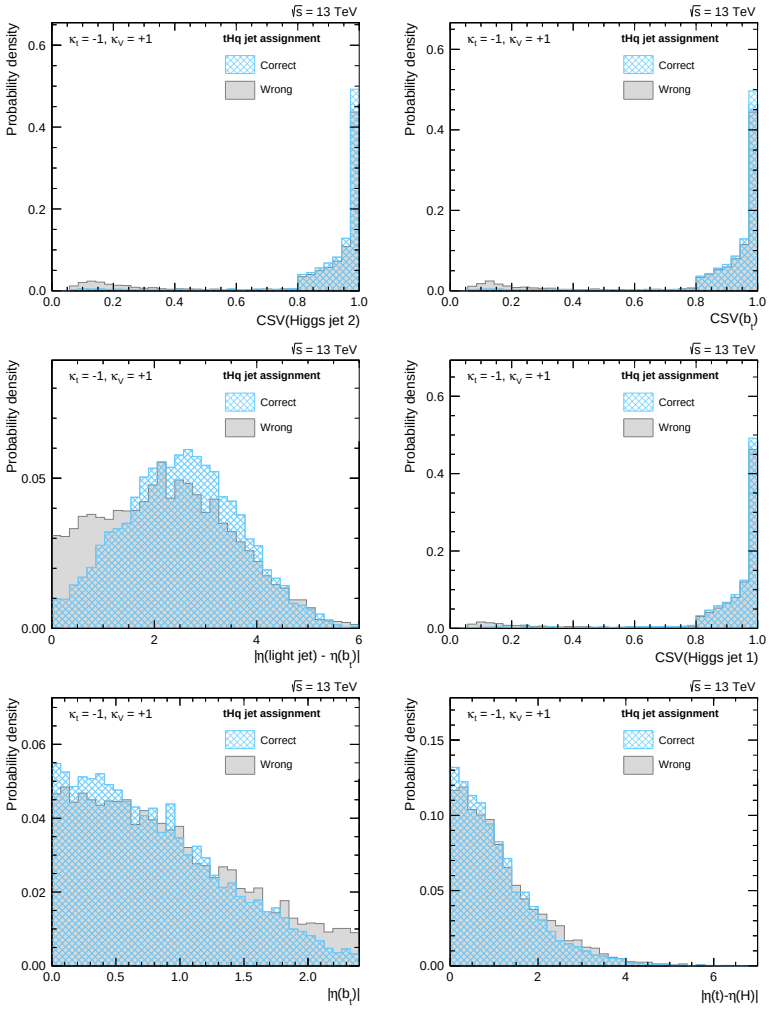


Figure B.1.: The distributions of variables ranked 7th to 12th place of the tHq reconstruction. The distributions of the correct and wrong jet assignments for each variable are shown. The description of the variables is listed in Table 5.3. The three remaining variables can be found in Appendix B.2.

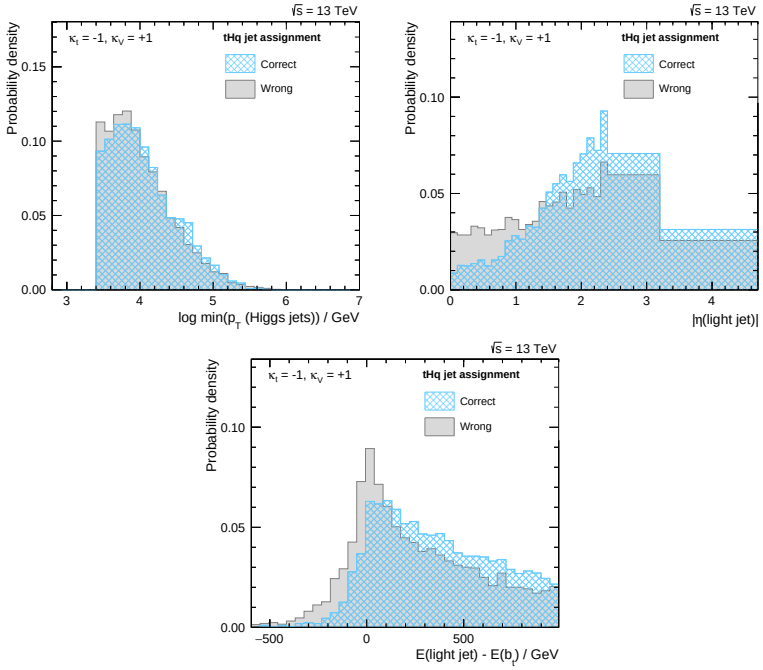


Figure B.2.: The distributions of variables ranked 13th to 15th place of the tHq reconstruction. The distributions of the correct and wrong jet assignments for each variable are shown. The description of the variables is listed in Table 5.3.

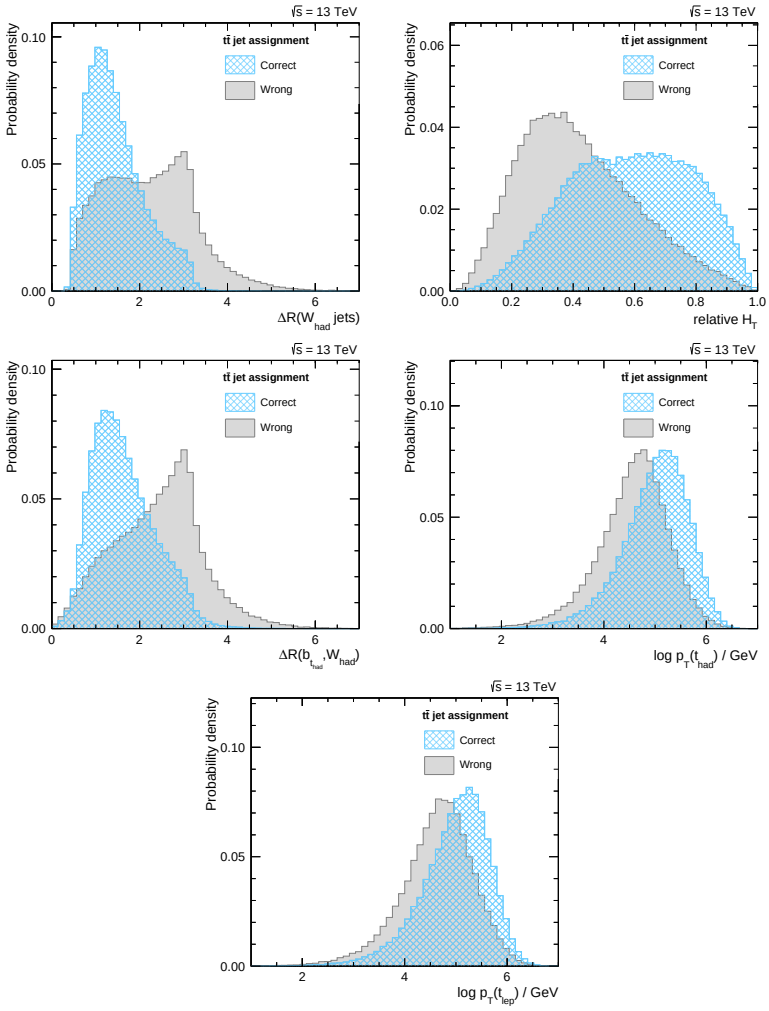


Figure B.3.: The distributions of variables ranked 7th to 11th place of the tt̄ reconstruction. The distributions of the correct and wrong jet assignments for each variable are shown. The description of the variables is listed in Table 5.5.

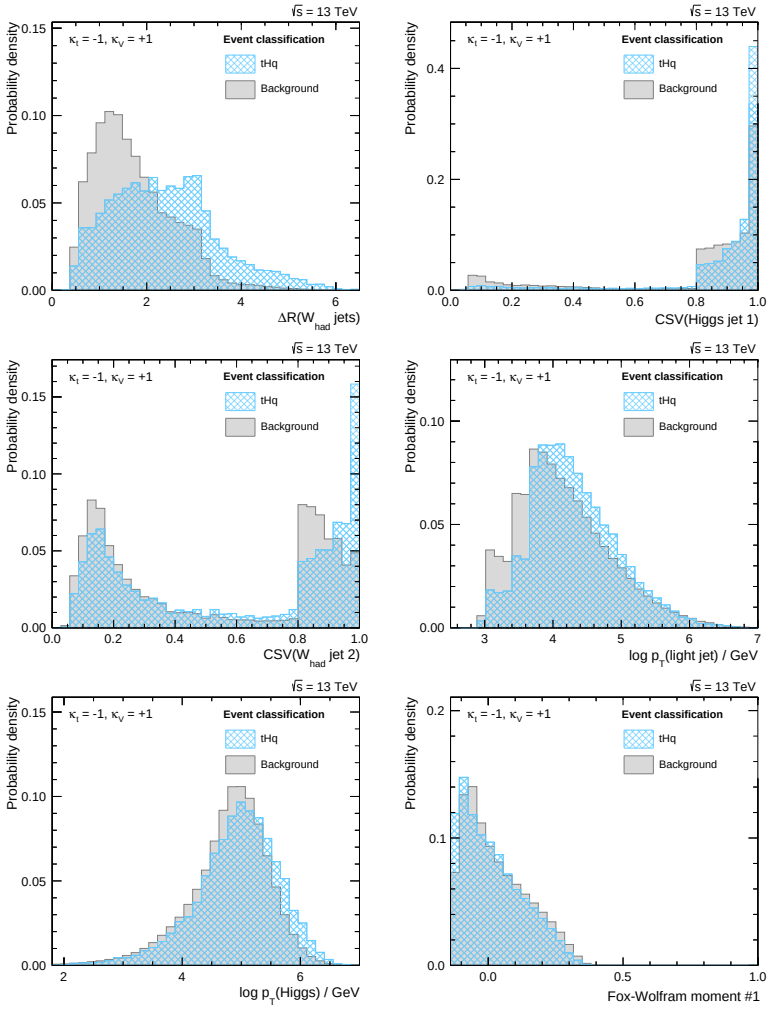


Figure B.4.: The distributions of variables ranked 7th to 12th place of the event classification. The distributions of signal and background events for each variable are shown. The description of the variables is listed in Table 5.7. The three remaining variables can be found in Appendix B.5.

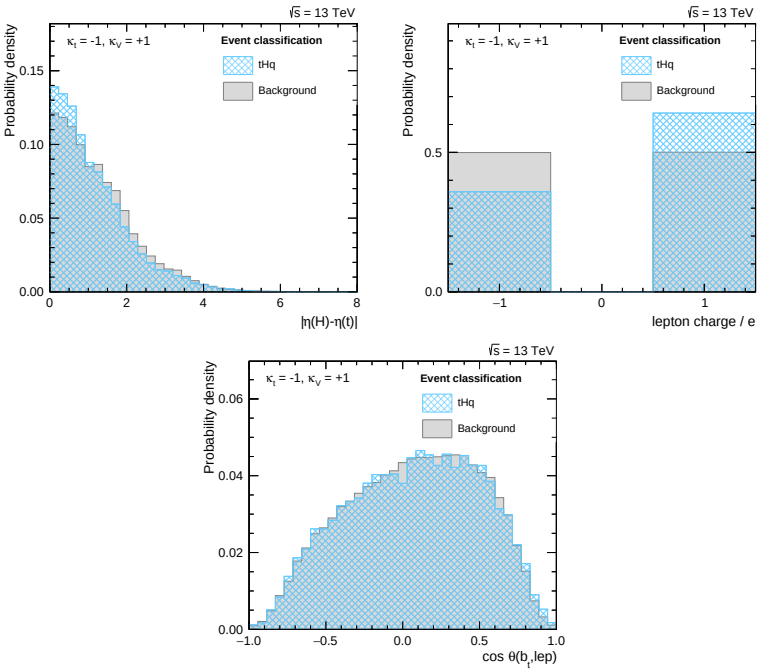


Figure B.5.: The distributions of variables ranked 13th to 15th place of the event classification. The distributions of signal and background events for each variable are shown. The description of the variables is listed in Table 5.7.

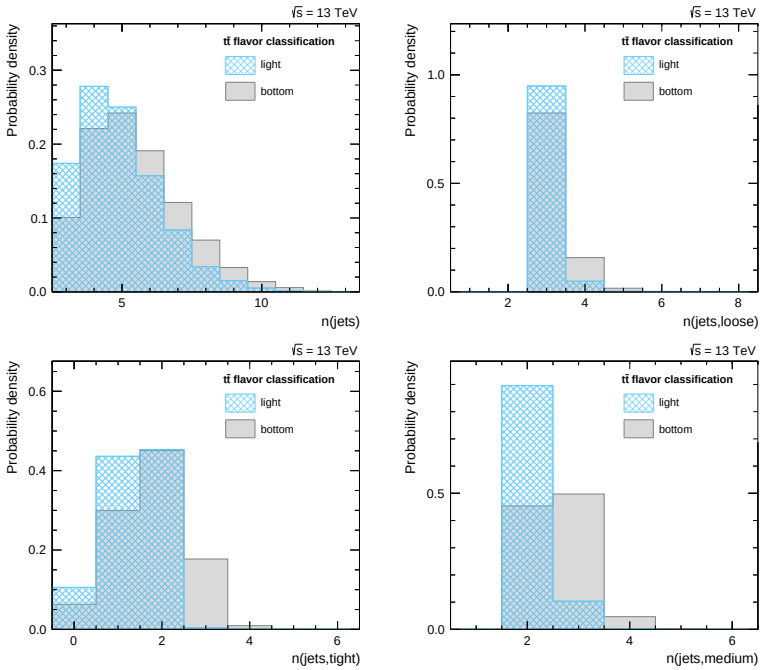


Figure B.6.: The distributions of variables ranked 7th to 10th place of the flavor classification. The distributions of events with light-flavored and with heavy-flavored jets for each variable are shown. The description of the variables is listed in Table 5.13.

C. Limit Values

Table C.1.: List of all expected and observed asymptotic limits at 95% CL for all studied points in the $\kappa_t - \kappa_V$ plane. The super- and subscribed values for the expected limit correspond to the $\pm 1\sigma$ uncertainty values for the studied points.

κ_t	$\kappa_V = +0.5$		$\kappa_V = +1.0$		$\kappa_V = +1.5$	
	obs.	exp.	obs.	exp.	obs.	exp.
-3.0	1.3	$1.9^{+1.1}_{-0.6}$	1.3	$1.6^{+0.9}_{-0.6}$	1.3	$1.8^{+1.1}_{-0.6}$
-2.0	2.7	$3.5^{+2.0}_{-1.2}$	2.1	$2.9^{+1.6}_{-1.0}$	2.1	$2.9^{+1.6}_{-1.0}$
-1.5	4.8	$5.3^{+3.1}_{-1.8}$	4.0	$4.1^{+2.5}_{-1.4}$	4.1	$3.7^{+2.1}_{-1.3}$
-1.25	6.1	$7.1^{+4.0}_{-2.4}$	4.5	$5.0^{+2.9}_{-1.7}$	3.6	$4.7^{+2.7}_{-1.6}$
-1.0	6.3	$9.2^{+5.3}_{-3.2}$	6.2	$6.5^{+3.8}_{-2.2}$	4.3	$5.2^{+3.0}_{-1.8}$
-0.75	11.9	$14.2^{+8.5}_{-4.9}$	6.8	$8.3^{+4.8}_{-2.8}$	5.1	$5.9^{+3.4}_{-2.0}$
-0.5	16.9	$21.7^{+12.5}_{-7.3}$	8.8	$10.7^{+6.2}_{-3.6}$	7.2	$7.8^{+4.6}_{-2.6}$
-0.25	32.0	$36.1^{+20.9}_{-12.4}$	14.0	$15.3^{+8.9}_{-5.2}$	9.7	$9.6^{+5.7}_{-3.3}$
0.0	65.3	$69.3^{+41.4}_{-23.6}$	18.1	$21.9^{+12.8}_{-7.5}$	11.9	$12.6^{+7.4}_{-4.3}$
$+0.25$	229.3	$184.0^{+107.8}_{-62.8}$	24.7	$34.9^{+20.2}_{-11.8}$	14.0	$16.9^{+9.9}_{-5.7}$
$+0.5$	431.9	$331.0^{+197.9}_{-113.8}$	59.0	$57.3^{+34.2}_{-19.5}$	20.5	$22.2^{+13.3}_{-7.5}$
$+0.75$	145.4	$157.3^{+88.4}_{-52.2}$	75.7	$86.8^{+51.9}_{-29.8}$	32.3	$32.6^{+19.1}_{-11.2}$
$+1.0$	55.9	$66.3^{+38.3}_{-22.0}$	116.1	$99.5^{+62.7}_{-34.7}$	44.0	$45.6^{+28.0}_{-15.8}$
$+1.25$	26.9	$34.8^{+19.8}_{-11.6}$	84.7	$78.3^{+45.2}_{-26.5}$	58.9	$54.3^{+33.3}_{-18.9}$
$+1.5$	15.0	$19.8^{+11.1}_{-6.6}$	40.9	$49.3^{+28.0}_{-16.6}$	59.6	$57.9^{+33.9}_{-19.6}$
$+2.0$	6.8	$9.8^{+5.5}_{-3.3}$	13.8	$19.9^{+11.2}_{-6.6}$	39.7	$42.1^{+24.3}_{-14.3}$
$+3.0$	3.8	$3.5^{+2.0}_{-1.2}$	4.6	$6.3^{+3.6}_{-2.1}$	8.4	$11.5^{+6.5}_{-3.8}$

List of Tables

List of Figures

Bibliography

[1] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, pp. 1–29, 2012.

[2] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, pp. 30–61, 2012.

[3] The CMS Collaboration, "Search for $H \to b\bar{b}$ in association with a single top quark as a test of Higgs boson couplings at $\sqrt{s} = 13$ TeV," CMS-PAS-HIG-16-019, CERN, Geneva, 2016.

[4] The CMS Collaboration, "Search for $H \to b\bar{b}$ in association with single top quarks as a test of Higgs couplings," CMS Physics Analysis Summary CMS-PAS-HIG-14-015, CERN, Geneva, 2014.

[5] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Physical Review Letters*, vol. 13, pp. 321–323, 1964.

[6] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Physical Review Letters*, vol. 13, pp. 508–509, 1964.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles," *Physical Review Letters*, vol. 13, pp. 585–587, 1964.

[8] A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik*, vol. 49, pp. 769–822, 1916.

[9] K. A. Olive, "Review of Particle Physics," *Chinese Physics C*, vol. 40, no. 10, p. 100001, 2016.

[10] M. Gell-Mann and Y. Neeman, "The eightfold way," WA Benjamin, Inc., New York, 1964.

[11] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the color octet gluon picture," *Physics Letters B*, vol. 47, no. 4, pp. 365 – 368, 1973.

[12] S. Tomonaga, "On a relativistically invariant formulation of the quantum theory of wave fields.," *Progress of Theoretical Physics*, vol. 1, no. 2, pp. 27–42, 1946.

[13] J. Schwinger, "Quantum electrodynamics. i. a covariant formulation," *Physical Review*, vol. 74, pp. 1439–1461, 1948.

[14] R. P. Feynman, "Space-time approach to quantum electrodynamics," *Physical Review*, vol. 76, pp. 769–789, 1949.

[15] S. Weinberg, "A model of leptons," *Physical Review Letters*, vol. 19, pp. 1264–1266, 1967.

[16] A. Salam, "Proceedings of the 8th Nobel Symposium, Stockholm, 1968," *Almquist and Wiksells*, 1968.

[17] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry," *Physical Review D*, vol. 2, pp. 1285–1292, 1970.

[18] Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Physical Review Letters*, vol. 81, pp. 1562–1567, 1998.

[19] J. Ellis, "Outstanding questions: physics beyond the standard model," *Philosophical Transactions of the Royal Society A*, vol. 370, no. 1961, pp. 818–830, 2012.

[20] J. D. Lykken, "Beyond the Standard Model," in *CERN Yellow Report CERN-2010-002, 101-109*, 2010.

[21] M. Peskin and D. Schroeder, "An introduction to quantum field theory," Westview Press, Boulder, CO, 1995.

[22] S. Dawson, "Introduction to electroweak symmetry breaking," in *Proceedings, Summer School in High-energy physics and cosmology: Trieste, Italy, June 29-July 17, 1998*, pp. 1–83, 1998.

[23] A. Zee, "Quantum field theory in a nutshell," Princeton university press, Princeton, NJ, 2010.

[24] L. Álvarez Gaumé and J. Ellis, "Eyes on a prize particle," *Nature Physics*, vol. 7, no. 1, pp. 2–3, 2011.

[25] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, "Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson," *Journal of High Energy Physics*, vol. 05, p. 022, 2013.

[26] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, "Higgs production in association with a single top quark at the LHC," *The European Physical Journal C*, vol. 75, no. 6, p. 267, 2015.

[27] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, "tWH associated production at the LHC," *arXiv:1607.05862*, 2016.

[28] S. Dittmaier *et al.*, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," *arXiv:1101.0593*, 2011.

[29] F. Abe *et al.*, "Observation of top quark production in $\bar{p}p$ collisions," *Physical Review Letters*, vol. 74, pp. 2626–2631, 1995.

[30] S. Abachi *et al.*, "Observation of the top quark," *Physical Review Letters*, vol. 74, pp. 2632–2637, 1995.

[31] The ATLAS, CDF, CMS and DØ Collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass," *arXiv:1403.4427*, 2014.

[32] C. Lefevre, "LHC: the guide (English version). Guide du LHC (version anglaise)," 2009.

[33] CERN, "The Large Hadron Collider," 2014. `http://home.cern/topics/large-hadron-collider`, last accessed on 2016-10-23.

[34] C. De Melis, "The CERN accelerator complex. Complexe des accélérateurs du CERN," 2016. `https://cds.cern.ch/record/2197559`, last accessed on 2016-10-18.

[35] CERN, "LHC experiments back in business at record energy," 2015. `http://home.cern/about/updates/2015/06/lhc-experiments-back-business-record-energy`, last accessed on 2016-10-23.

[36] CERN, "LHC performance reaches new highs," 2016. `https://home.cern/about/updates/2016/07/lhc-performance-reaches-new-highs`, last accessed on 2016-10-23.

[37] The CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008.

[38] CERN, "CMS: The Compact Muon Solenoid," 2012. `https://cds.cern.ch/record/1997263`, last accessed on 2016-11-22.

[39] The CMS Collaboration, "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays," *Journal of Instrumentation*, vol. 5, no. 03, p. T03021, 2010.

[40] S. R. Davis, "Interactive Slice of the CMS detector," 2016. `https://cds.cern.ch/record/2205172`, last accessed on 2016-10-18.

[41] V. Karimäki, M. Mannelli, P. Siegrist, H. Breuker, A. Caner, R. Castaldi, K. Freudenreich, G. Hall, R. Horisberger, M. Huhtinen, and A. Cattai, "The CMS tracker system project: Technical Design Report," Technical Design Report CMS, CERN, Geneva, 1997.

[42] The CMS Collaboration, "The CMS tracker: addendum to the Technical Design Report," Technical Design Report CMS, CERN, Geneva, 2000.

[43] The CMS Collaboration, "Tracker detector," 2011. `http://cms.web.cern.ch/news/tracker-detector`, last accessed on 2016-10-24.

[44] The CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report," Technical Design Report CMS, CERN, Geneva, 1997.

[45] P. Bloch, R. Brown, P. Lecoq, and H. Rykaczewski, "Changes to CMS ECAL electronics: addendum to the Technical Design Report," Technical Design Report CMS, CERN, Geneva, 2002.

[46] The CMS Collaboration, "Electromagnetic Calorimeter," 2011. `http://cms.web.cern.ch/news/electromagnetic-calorimeter`, last accessed on 2016-10-24.

[47] The CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report," Technical Design Report CMS, CERN, Geneva, 1997.

[48] The CMS Collaboration, "Hadron Calorimeter," 2011. `http://cms.web.cern.ch/news/hadron-calorimeter`, last accessed on 2016-10-24.

[49] The CMS Collaboration, "The CMS muon project: Technical Design Report," Technical Design Report CMS, CERN, Geneva, 1997.

[50] The CMS Collaboration, "Muon Detectors," 2011. `http://cms.web.cern.ch/news/muon-detectors`, last accessed on 2016-10-25.

[51] V. Khachatryan *et al.*, "The CMS trigger system," *Submitted to: Journal of Instrumentation, arXiv:1609.02366*, 2016.

[52] D. Acosta, "CMS Trigger Improvements Towards Run II," *Nuclear and Particle Physics Proceedings*, vol. 273–275, pp. 1008 – 1013, 2016. 37th International Conference on High Energy Physics (ICHEP).

[53] C. Eck *et al.*, "LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)," Technical Design Report LCG, CERN, Geneva, 2005.

[54] I. Bird *et al.*, "Update of the Computing Models of the WLCG and the LHC Experiments," CERN-LHCC-2014-014. LCG-TDR-002, 2014.

[55] WLCG Office, "Worldwide LHC Computing Grid (WLCG) – Tier centres," 2016. `http://wlcg-public.web.cern.ch/tier-centres`, last accessed on 2016-10-25.

[56] The CMS Collaboration, "Computing Grid," 2011. `http://cms.web.cern.ch/news/computing-grid`, last accessed on 2016-10-25.

[57] WLCG Project Office, "Documents & Reference - Tiers & Grid Images (2014)," 2014. `http://wlcg.web.cern.ch/documents-reference`, last accessed on 2016-10-18.

[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 (TASI 2014) Boulder, Colorado, June 2-27, 2014*, 2014.

[59] V. Gribov and L. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Soviet Journal of Nuclear Physics*, vol. 15, pp. 438–450, 1972.

[60] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nuclear Physics B*, vol. 126, p. 298, 1977.

[61] Y. Dokshitzer, "Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics," *Zh. Eksp. Teor. Fiz*, vol. 73, p. 1216, 1977.

[62] J. S. Gainer, J. Lykken, K. T. Matchev, S. Mrenna, and M. Park, "The Matrix Element Method: Past, Present, and Future," in *Proceedings, Community Summer Study 2013: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013.

[63] The NNPDF Collaboration, "Neural Network Parton Distribution Functions (NNPDF) – Images," 2016. `https://nnpdf.hepforge.org/images/`, last accessed on 2016-10-26.

[64] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics," *Physics Reports*, vol. 97, pp. 31–145, 1983.

[65] N. Metropolis and S. Ulam, "The Monte Carlo Method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[66] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond," *Journal of High Energy Physics*, vol. 06, p. 128, 2011.

[67] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *Journal of High Energy Physics*, vol. 02, p. 027, 2003.

[68] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, *et al.*, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *Journal of High Energy Physics*, vol. 07, p. 079, 2014.

[69] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms," *Journal of High Energy Physics*, vol. 11, p. 040, 2004.

[70] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *Journal of High Energy Physics*, vol. 11, p. 070, 2007.

[71] 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*, vol. 06, p. 043, 2010.

[72] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *Journal of High Energy Physics*, vol. 05, p. 026, 2006.

[73] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Computer Physics Communications*, vol. 178, pp. 852–867, 2008.

[74] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nuclear Instruments and Methods in Physics*, vol. A506, pp. 250–303, 2003.

[75] J. Allison *et al.*, "Geant4 developments and applications," *IEEE Transactions on Nuclear Science*, vol. 53, no. 1, pp. 270–278, 2006.

[76] S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *Journal of Instrumentation*, vol. 9, no. 10, p. P10009, 2014.

[77] P. Billoir, "Progressive track recognition with a Kalman-like fitting procedure," *Computer Physics Communications*, vol. 57, no. 1–3, pp. 390 – 394, 1989.

[78] 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*, vol. 294, no. 1–2, pp. 219 – 228, 1990.

[79] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2–3, pp. 444 – 450, 1987.

[80] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting," CMS-NOTE-2007-008, CERN, Geneva, 2007.

[81] The CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, Geneva, 2009.

[82] N. Bartosik, "PF concept in CMS," 2016. `http://bartosik.pp.ua/hep_sketches/cms_particle_flow`, last accessed on 2016-11-01.

[83] 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 Physics*, vol. 31, p. 9, Sept. 2005.

[84] V. Khachatryan *et al.*, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV," *Journal of Instrumentation*, vol. 10, no. 06, p. P06005, 2015.

[85] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, *et al.*, "Electron reconstruction in CMS," *Eur. Phys. J. C*, vol. 49, pp. 1099–1116, 2007.

[86] S. Chatrchyan *et al.*, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV," *Journal of Instrumentation*, vol. 7, p. P10002, 2012.

[87] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *Journal of High Energy Physics*, vol. 04, p. 063, 2008.

[88] V. Khachatryan *et al.*, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV," *Submitted to: Journal of Instrumentation, arXiv:1607.03663*, 2016.

[89] The CMS Collaboration: Jet Energy Resolution and Corrections (JERC) Subgroup, "Introduction to Jet Energy Corrections at CMS," 2016. `https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC`, last accessed on 2016-11-03.

[90] The DØ Collaboration, "Observation of Single Top Quark Production – B-Jet Identification," 2009. `https://www-d0.fnal.gov/Run2Physics/top/singletop_observation/`, last accessed on 2016-10-27.

[91] The CMS collaboration, "Identification of b-quark jets with the CMS experiment," *Journal of Instrumentation*, vol. 8, no. 04, p. P04013, 2013.

[92] The CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2," CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[93] V. Khachatryan *et al.*, "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV," *Journal of High Energy Physics*, vol. 03, p. 136, 2011.

[94] V. Blobel and E. Lohrmann, "Statistische und numerische Methoden der Datenanalyse," Teubner, Stuttgart Leipzig, 1998.

[95] The CMS Collaboration, "Documentation of the RooStats-based statistics tools for Higgs PAG," 2016. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit`, last accessed on 2016-11-10.

[96] The CMS Collaboration, "Welcome to the RooStats Wiki," 2016. `https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome`, last accessed on 2016-11-10.

[97] T. Junk, "Confidence level computation for combining searches with small statistics," *Nuclear Instruments and Methods in Physics A*, vol. 434, pp. 435–443, 1999.

[98] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)," 2000. CERN-OPEN-2000-205.

[99] A. L. Read, "Presentation of search results: the CL_s technique," *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, no. 10, p. 2693, 2002.

[100] 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. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, pp. 289–337, 1933.

[101] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *European Physical Journal C*, vol. 71, p. 1554, 2011.

[102] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.

[103] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees," CRC Press, New York, 1984.

[104] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, "The elements of statistical learning: data mining, inference, and prediction," Springer, Berlin, 2009.

[105] G. Bohm and G. Zech, "Introduction to statistics and data analysis for physicists," Verl. Dt. Elektronen-Synchrotron, Hamburg, 2010.

[106] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis," *Proceeding of Science*, vol. ACAT, p. 040, 2007.

[107] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.

[108] The CMS Collaboration, "Investigating the top-Yukawa coupling with the production of a Higgs boson in association with a single top quark in the $H \to b\bar{b}$ decay channel," CMS Physics Analysis Note, CMS AN-16-065, CERN, 2016.

[109] CERN, "SM Higgs Branching Ratios and Total Decay Widths (update in CERN Report4 2016)," 2016. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR#Higgs_2_fermions`, last accessed on 2016-11-29.

[110] The ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV," *Journal of High Energy Physics*, vol. 08, p. 045, 2016.

[111] The CMS Collaboration, "Measurement of the cross section ratio $\sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$ in pp collisions at $\sqrt{s} = 8\,\text{TeV}$," *Physics Letters B*, vol. 746, pp. 132 – 153, 2015.

[112] The ATLAS Collaboration, "Study of heavy-flavor quarks produced in association with top-quark pairs at $\sqrt{s} = 7\,\text{TeV}$ using the ATLAS detector," *Physical Review D*, vol. 89, no. 7, p. 072012, 2014.

[113] The CMS Collaboration, "Multivariate Electron Identification for Run2 – Triggering electron MVA details and working points," 2016. `https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2#Triggering_electron_MVA_details`, last accessed on 2016-11-30.

[114] The CMS Collaboration, "Baseline muon selections for Run-II – Muon Identification," 2016. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2#Tight_Muon`, last accessed on 2016-12-01.

[115] The CMS Collaboration, "Jet Identification – Recommendations for 13 TeV data analysis (74X, 76X, 80X)," 2016. `https://twiki.cern.ch/twiki/bin/view/CMS/JetID#Recommendations_for_13_TeV_data`, last accessed on 2016-12-01.

[116] The CMS Collaboration, "Usage of b Tag Objects for 13 TeV Data with 25ns bunch spacing and 76X ReReco – Supported Algorithms and Operating Points," 2016. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation76X, last accessed on 2016-12-01.

[117] The CMS Collaboration, "MET Corrections and Uncertainties for Run-II," 2016. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETRun2Corrections, last accessed on 2016-12-02.

[118] T. Chwalek, "Measurement of the W-Boson Helicity-Fractions in Top-Quark Decays with the CDF II Experiment and Prospects for an Early $t\bar{t}$ Cross-Section Measurement with the CMS Experiment," Karlsruhe Institute of Technology, CERN-THESIS-2010-255, 2010.

[119] The CMS Collaboration, "Pileup Reweighting Utilities," 2016. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities, last accessed on 2016-12-03.

[120] The CMS Collaboration, "Muon T&P Instructions for Run-II," 2016. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonTagAndProbeTreesRun2, last accessed on 2016-12-04.

[121] The CMS Collaboration, "Reference muon id, isolation and trigger efficiencies for Run-II," 2016. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2, last accessed on 2016-12-04.

[122] The CMS Collaboration, "Electron and photon ID – Efficiencies and scale factors," 2016. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Efficiencies_and_scale_factors, last accessed on 2016-12-04.

[123] The CMS Collaboration, "Details of the Tag and Probe procedure for Egamma," 2016. https://twiki.cern.ch/twiki/bin/view/CMS/ElectronScaleFactorsRun2, last accessed on 2016-12-04.

[124] The CMS Collaboration, "Btag Shape Calibration – Event reweighting using scale factors calculated with a tag and probe method," 2016. https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration, last accessed on 2016-12-04.

[125] The CMS Collaboration, "Calibration of the Combined Secondary Vertex b-Tagging discriminant using dileptonic ttbar and Drell-Yan events," CMS Physics Analysis Note, CMS AN-13-130, CERN, 2013.

[126] E. Boos *et al.*, "Generic user process interface for event generators," in *Physics at TeV colliders. Proceedings, Euro Summer School, Les Houches, France, May 21-June 1, 2001*, 2001.

[127] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, *et al.*, "A Standard format for Les Houches event files," *Computer Physics Communications*, vol. 176, pp. 300–304, 2007.

[128] V. D. Barger, J. Ohnemus, and R. J. N. Phillips, "Event shape criteria for single lepton top signals," *Physical Review D*, vol. 48, pp. 3953–3956, 1993.

[129] G. C. Fox and S. Wolfram, "Observables for the analysis of event shapes in e^+ e^- annihilation and other processes," *Physical Review Letters*, vol. 41, no. 23, p. 1581, 1978.

[130] "CMS Luminosity Measurement for the 2015 Data Taking Period," CMS-PAS-LUM-15-001, CERN, Geneva, 2016.

[131] The CMS Collaboration, "Estimating Systematic Errors Due to Pileup Modeling," 2013. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors`, last accessed on 2016-12-13.

[132] The CMS Collaboration, "Utilities for Accessing Pileup Information for Data – Pileup JSON Files For Run II," 2016. `https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#Pileup_JSON_Files_For_Run_II`, last accessed on 2016-12-13.

[133] The CMS Collaboration, "Jet Energy Resolution – JER Scaling factors and Uncertainty for 13 TeV (2015 and 2016)," 2016. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution#JER_Scaling_factors_and_Uncertai`, last accessed on 2016-12-13.

[134] The CMS Collaboration, "Recommended Jet Energy Corrections and Uncertainties For Data and MC – Jet Energy Corrections in Run2," 2016. `https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC`, last accessed on 2016-12-13.

[135] The CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV," *Journal of Instrumentation*, vol. 10, no. 02, p. P02006, 2015.

[136] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples," *Computer Physics Communications*, vol. 77, no. 2, pp. 219 – 228, 1993.

[137] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra," in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, pp. 115–120, 2011.

[138] The CMS Collaboration, "SM Higgs production cross sections at $\sqrt{s} = 13$ TeV (update in CERN Report4 2016)," 2016. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV`, last accessed on 2016-12-14.

[139] The CMS Collaboration, "NNLO+NNLL top-quark-pair cross sections," 2016. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO`, last accessed on 2016-12-14.

[140] 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," *Journal of High Energy Physics*, vol. 02, p. 113, 2016.

[141] The CMS Collaboration, "NLO single-top channel cross sections," 2016. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec`, last accessed on 2016-12-14.

[142] The CMS Collaboration, "Documentation on c-tag MVA trainings," 2016. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagCTagger`, last accessed on 2016-12-16.

[143] S. Moortgat, "Development of new charm-tagging methods for the search for Flavour Changing top-quark dark matter interactions at the LHC," Vrije Universiteit Brussel, Master thesis, 2015.

[144] The ATLAS Collaboration, "Performance and Calibration of the JetFitterCharm Algorithm for c-Jet Identification," ATL-PHYS-PUB-2015-001, CERN, Geneva, 2015.

[145] The CMS Collaboration, "Modelling of the single top-quark production in association with the Higgs boson at 13 TeV," 2016. https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV, last accessed on 2016-12-18.

Danksagung

An erster Stelle möchte ich mich bei Herrn Prof. Dr. Thomas Müller für die Aufnahme in seine Arbeitsgruppe und für die Ermöglichung dieser Masterarbeit bedanken.

Ebenfalls möchte ich Herrn Prof. Dr. Ulrich Husemann für die Übernahme des Korreferats danken.

Ein großer Dank geht an Dr. Thorsten Chwalek für die hervorragende Betreuung sowie für das ausführliche Korrekturlesen dieser Arbeit.

Besonders bedanken möchte ich mich bei meinen ehemaligen Analysekollegen Dr. Simon Fink und Dr. Benedikt Maier, auf deren Arbeit diese Masterarbeit aufbaut und die sich stets Zeit genommen haben, mich in die Analyse einzuarbeiten und geduldig meine vielen Fragen zu beantworten. Weiterhin möchte ich Dr. Frank Roscher für seine Anregungen, seine Unterstützung bezüglich allem, was mit Programmieren und git zu tun hat, sowie für seine Geduld mit mir danken. Mein Dank gilt ebenso Nils Faltermann für das Korrekturlesen der „Rohfassung“ dieser Masterarbeit.

Bedanken möchte ich mich bei Kevin Flöh, mit dem ich – zusammen mit Dr. Simon Fink, Dr. Benedikt Maier und Nils Faltermann – die Analyse, auf welche diese Arbeit beruht, durchgeführt habe. Mein Dank gilt auch Johann Rauser und David Seith für die Vorbereitung der Analyse auf die 2016er Daten, sodass sie Kevin und mir während dem Schreiben unserer Arbeiten den Rücken frei hielten. Darüber hinaus möchte ich allen anderen Mitgliedern der Arbeitsgruppe für ein ausgezeichnetes Arbeitsklima danken.

Generell möchte ich mich beim gesamten Institut für eine angenehme und auch unterhaltsame Arbeitsatmosphäre danken. Besonderer Dank geht hier an Frau Bräunling für die organisatorischen Angelegenheiten sowie an das EKP-Admin-Team, die stets für einen reibungslosen Ablauf im technischen Bereich sorgen.

Ein großer Dank geht an alle meine Freunde sowie an meine WG, die stets ein offenes Ohr für meine Sorgen haben und mich auch auf andere Gedanken gebracht haben, wenn es mal notwendig war.

Abschließend möchte ich mich allerliebst bei meinen Eltern danken, für die moralische und finanzielle Unterstützung während meines Studiums sowie für ihr Vertrauen in mir. Ohne sie wäre mein Studium niemals möglich gewesen.