Search for vector-like T' quarks using tools for the analysis of jet substructure with the CMS experiment

Dissertation zur Erlangung des Doktorgrades
an der Fakultät für
Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Rebekka Sophie Höing

Hamburg

2014

Tag der Disputation:	11. Dezember 2014
Folgende Gutachter empfehlen die Annahme der Dissertation:	Dr. Alexander Schmidt Prof. Dr. Johannes Haller

Abstract

A search for pairs of vector-like T' quark produced in proton-proton collisions recorded with the CMS experiment at $\sqrt{s} = 8$ TeV is presented. The search is optimized for decays of T' quarks to top quarks and Higgs bosons, where the top quarks and Higgs bosons decay hadronically. The T'-quark mass range between 500 and 1000 GeV is investigated. The top quarks and Higgs bosons produced in decays of the heavy T' quarks acquire large Lorentz boosts. The signatures of these particles in the detector can overlap and are therefore difficult to resolve using classical jet reconstruction methods.

Large-radius jets are reconstructed and subjets formed from their constituents. The decay products of particles with large Lorentz boosts are highly collimated and can all be found within a single one of these large-radius jets. Top jets containing hadronic top-quark decays are identified with a top-tagging algorithm that analyzes the jet substructure. A b-tagging algorithm is applied to the reconstructed subjets in order to find bottom quarks within the jet substructure. In order to identify Higgs bosons with large Lorentz boosts decaying to pairs of bottom quarks, the Higgs-tagging algorithm searches for two b-tagged subjets within a single jet. This is the first application of a top-tagging algorithm in conjunction with subjet b-tagging in an analysis of CMS data. Also, a Higgs-tagging algorithm is used for the first time in a search for new physics.

The main background contributions to this analysis consist of pair-produced top quarks and QCD-multijet events. More than 99% of these events are rejected by the event selection based on the new jet-substructure methods, while 6-8% of the signal events are retained. A description for the QCD-multijet background is obtained from data in a method also using jet-substructure information. Bayesian exclusion limits are derived from a likelihood ratio in which two discriminating variables are combined. T' quarks with masses below 745 GeV are excluded at 95% confidence level for exclusive decays of T' $\rightarrow$ tH. Furthermore, results for all combinations of the decay modes T' $\rightarrow$ tH, T' $\rightarrow$ tZ, and T'$\rightarrow$ bW are obtained. A statistical combination with other searches for T' quarks is performed. For different decay modes of the T' quark, the resulting mass limits range from 697 to 782 GeV.

Kurzfassung

Eine Suche nach in Proton-Proton-Kollisionen produzierten Paaren von vektorartigen T'-Quarks in den mit dem CMS-Experiment bei einer Schwerpunktsenergie von 8 TeV aufgezeichneten Daten wird vorgestellt. Diese Suche ist für Zerfälle der T'-Quarks in Top-Quarks und Higgs-Bosonen optimiert, in denen die Top-Quarks und Higgs-Bosonen hadronisch zerfallen. Ein Massenbereich für das T'-Quark von 500 bis 1000 GeV wird untersucht. Die in Zerfällen der schweren T'-Quarks produzierten Top-Quarks und Higgs-Bosonen weisen großen Lorentz-Boost auf. Das kann dazu führen, dass die Signaturen der Teilchen im Detektor überlappen. Dies erschwert die klassische Rekonstruktion der verschiedenen Teilchen in einzelnen Jets.

Jets mit großen Radien werden rekonstruiert und Subjets aus ihren Bestandteilen geformt. Die Zerfallsprodukte von Teilchen mit großem Lorentz-Boost liegen sehr nah beieinander und können daher allesamt innerhalb eines einzelnen Jets gefunden werden. Sogenannte Top-Jets enthalten hadronische Zerfälle von Top-Quarks. Top-Tagging-Algorithmen dienen ihrer Identifizierung mittels Analyse der Jetsubstruktur. Ein b-Tagging-Algorithmus wird auf die rekonstruierten Subjets angewandt, um Bottom-Quarks in der Jetsubstruktur zu finden. Um Higgs-Bosonen mit großen Lorentz-Boosts zu erkennen, die in Paare von Bottom-Quarks zerfallen, sucht der Higgs-Tagging-Algorithmus innerhalb der Jets nach zwei Subjets, die vom Subjet-b-Tagging-Algorithmus markiert wurden. Dies ist die erste Verwendung eines Top-Tagging-Algorithmus in Kombination mit einem Subjet-b-Tagging-Algorithmus in einer Analyse von CMS Daten. Außerdem wird zum ersten Mal ein Higgs-Tagging-Algorithmus in einer Suche nach neuer Physik angewendet.

Der Untergrund zu dieser Analyse besteht hauptsächlich aus in Paaren produzierten Top-Quarks und QCD-Multijet-Ereignissen. Mehr als 99% dieser Ereignisse werden in der Ereignissselektion aussortiert, während 6-8% der Signalereignisse ausgewählt werden. Der Untergrundbeitrag von QCD-Multijet-Ereignissen wird mit Hilfe von gemessenen Daten beschrieben. Die verwendete Methode basiert ebenfalls auf Informationen über die Substruktur von Jets. Bayessche Ausschlussgrenzen werden mit Hilfe einer Likelihood-Variable bestimmt, in der zwei zwischen Untergrund und Signal diskriminierende Variablen zusammengefasst werden. Unter der Annahme, dass nur Zerfälle von T' $\rightarrow$ tH möglich sind, werden T'-Quarks mit geringeren Massen als 745 GeV mit 95% C.L. ausgeschlossen. Außerdem werden Ergebnisse für alle erlaubten Kombinationen der drei Zerfallsmoden T'$\rightarrow$ tH, T' $\rightarrow$ tZ und T' $\rightarrow$ bW produziert. Eine statistische Kombination der Analyse mit anderen Suchen nach T'-Quarks wird durchgeführt. Hier werden Massenausschlussgrenzen zwischen 697 und 782 GeV für verschiedene Zerfallsmoden gesetzt.

List of publications

My research during the three years of my PhD studies resulted in the following publication:

CMS Collaboration, "Search for top-Higgs resonances in all-hadronic final states using jet substructure methods", CMS Physics Analysis Summary CMS-PAS-B2G-14-002, 2014. `https://cds.cern.ch/record/1706121?ln=en`

Furthermore, the following publication is in preparation:

CMS Collaboration, "Search for top Higgs resonances in all hadronic final state using jet substructure methods", Journal of High Energy Physics (2014)

Contents

1 Introduction

In 2014, physicists worldwide celebrate the 60th anniversary of CERN. Since the ratification of the Conseil Européen pour la Recherche Nucléaire (CERN) by its twelve original member states on September 29th 1954, physicists at CERN have contributed greatly to the understanding of modern particle physics. Outstanding research results include the discovery of the W and Z bosons in 1983 [1–4], and the first time production of anti-hydrogen in 1995 [5]. Recently, much publicity was given to the long awaited discovery of the Higgs boson [6, 7], that was achieved in the summer of 2012 by the CMS and ATLAS collaborations in analyses of collisions of particles provided by the Large Hadron Collider (LHC). Besides the many scientific accomplishments, the well-functioning international collaboration at CERN is noteworthy. Today, there are 21 member states and scientists from more than 60 countries shape the different research programs at CERN, setting an example for peaceful collaboration regardless of world politics or diplomatic conflicts.

While the Higgs boson is often described as the last building block completing the standard model of particle physics, it certainly does not mark the end of the era of research in particle physics. Many fundamental questions remain unsolved to date. For instance, the observed matter-antimatter asymmetry in our universe, as well as the origin of the gravitationally interacting dark matter are yet to be understood. Another issue is the so-called hierarchy problem: loop corrections appear in the calculation of the Higgs boson mass [8]. If the standard model is to remain valid up to large energy scales such as the Planck scale, these loop corrections give rise to large divergencies. The corrections needed to cancel these divergencies exceed the actual mass of the Higgs boson itself by several orders of magnitude, which can be perceived as unnatural. The standard model in its current form does not incorporate solutions to any of these problems. Consequently, physics beyond the standard model of some form must exist. The theoretical physics community is providing numerous hypothetical solutions to these issues. Many of them are being tested in the experiments at the LHC.

The discovery of the Higgs boson at the LHC had a great impact on the landscape of physics beyond the standard model. Many theories have been rigorously constrained or even completely excluded by the discovery. One example is the extension of the standard model quark sector to a sequential fourth generation exhibiting similar properties as the already known quarks. These kind of models seemed very appealing because of their simplicity. There is no intrinsic feature of the standard model limiting its quark sector to three generations. However, the existence of additional heavy particles of that kind would drastically enhance certain Higgs boson production modes with respect to the standard model expectation. The values measured by the CMS and ATLAS collaborations show no significant divergencies from the standard model prediction, thus excluding a sequential fourth generation of quarks at very high confidence level [9].

The discovery of the Higgs boson has also renewed interest in other areas of research though. In light of the exclusion of a sequential fourth generation of quarks, models with additional vector-like quarks have gained attractiveness. They are now the simplest possible extension of the quark sector still compatible with current measurements of standard

model parameters. Vector-like quarks are part of many models for physics beyond the standard model, e.g., little-Higgs models, composite-Higgs models, or models of extra dimensions [10–12]. All of these models propose solutions to the hierarchy problem and predict the existence of new particles at the TeV scale probed in LHC physics.

Vector-like quarks differ from the standard model quarks in one important aspect: their behavior under the weak interaction. Both, the right-handed and left-handed components of vector-like quarks have couplings to weak currents. They are very heavy and couple mainly to third generation quarks, leading to unique decay signatures including W, Z, and Higgs bosons, as well as top and bottom quarks. The particles are therefore also referred to as "heavy top partners".

So far, no evidence for physics beyond the standard model has been found in any collider experiment. As many scenarios predicting light new particles have been excluded in previous measurements, the focus is now shifting towards searches for heavier resonances. In the decay of these potentially very massive particles, the daughter particles are expected to obtain large Lorentz boosts. The subsequent decay of these daughter particles results in particular signatures in the particle detector: the decay products of the daughter particles are extremely collimated. If the Lorentz boost of the decaying particle is large enough, the decay products can even be collimated to an extent where the entire decay is contained in a single particle jet. New analysis techniques have been developed in recent years to identify such decays within the substructure of the jets. The use of these substructure tools is opening up a new window for the examination of hadronic final states in many sectors of particle physics.

In this work, a search for pair produced vector-like T' quarks is presented. Large sensitivity for the specific case in which both T' quarks decay into a top quark and a Higgs boson was the central design goal for this analysis. However, all possible decay channels of T' quarks are examined in this search. Only events without isolated leptons in the final state are considered. A T'-quark mass range of 500 GeV to 1 TeV is analyzed, meaning that the decay products of the T' quarks are likely to be produced with large Lorentz boosts[1]. Novel tools for the analysis of jet substructure, including the HEPTopTagger [13] and subjet b-tagging algorithms [14], are used for the first time. In the HEPTopTagger algorithm, subjets of large particle jets are reconstructed. Their properties are then used to identify hadronic top-quark decays within the original jet. While the identification of jets with bottom-quark content using b tagging algorithms is a well established method in particle physics analyses, the application of these algorithms to subjets of larger jets is a new approach. Subjet b tagging improves the performance of top-tagging algorithms and is also used to identify Higgs bosons with large Lorentz boosts decaying to $b\bar{b}$ pairs.

At the beginning of this thesis in chapter 2, an overview of the main concepts of the standard model of particle physics is given. This chapter also includes an introduction to models for physics beyond the standard model that predict the existence of vector-like quarks. In chapter 3, a description of the Large Hadron Collider and the main components of the CMS experiment is provided. Following this, an outlook to the planned upgrades of

[1] In this work so-called natural units are used. In this convention, electron volts (eV) are used as unit for energy. At the same time, the speed of light c and Planck's constant $\hbar$ are set to unity, $c = \hbar = 1$. Thus, masses are measured in units of electron volts as well. The corresponding unit for spatial distances and time is eV^{-1}. Natural units are commonly used in particle physics in order to simplify calculations.

the detector for future operation of the LHC is presented. Information on event simulation with Monte Carlo generators can be found in chapter 4. The particle reconstruction with the particle-flow algorithm is described in chapter 5 with a focus on the clustering of jets, which are of great importance in this analysis. Algorithms for the identification of bottom quarks within jets are introduced in section 5.2.4. Novel techniques for the analysis of jet substructure are employed, they are detailed in section 5.2.5. In chapter 6 the main concepts of Bayesian statistics, and their application in this analysis using the theta framework are outlined.

The introduction of these general concepts is followed by a detailed description of the search for pair-produced vector-like T' quarks in all-hadronic final states in chapter 7. An overview of the analysis strategy is provided in section 7.1. The specifics of the used datasets and simulated samples can be found in section 7.2, followed by a description of the event selection in section 7.3. A data-driven approach is used to model the background contribution from QCD-multijet events. Specifics on the method used in the modelling are provided in 7.4. The sources of systematic uncertainties in this analysis and their effect on the results are listed in section 7.5. In section 7.6, the results of the search are discussed.

In chapter 8, the previously presented search is combined with other searches for vector-like quarks. The potential of future searches for vector-like quarks in the single production channel or at higher center-of-mass energies is evaluated in chapter 9. The work presented in this thesis is concluded in chapter 10.

2 The standard model and vector-like quarks

An overview of the theoretical concepts relevant for the work presented in this thesis is given in this chapter. In the first section, the main properties of the standard model of particle physics and its particle content are described, mostly following the description in [15, 16].

The second half of this chapter concerns ideas for physics beyond the standard model (BSM theories). Many BSM theories have been developed in the last decades. Their aim is to provide solutions for certain issues that are not addressed in the standard model of particle physics in its current form. Here, the focus is set on BSM theories involving vector-like quarks.

2.1 The standard model of particle physics

The standard model of particle physics describes the nature of interactions between structureless, point-like elementary particles. In the last decades, the standard model has been thoroughly investigated in numerous experiments. It was found to be extremely successful: all particles predicted by the standard model have been discovered in experiments. Impressively high precision has also been achieved in the experimental determination of other model parameters [16]. The results of these measurements show very good agreement with the values predicted by the theory.

Three fundamental interactions, or forces, are described by the standard model of particle physics: the electromagnetic interaction, the weak interaction, and the strong interaction. The interactions of the standard model are described by fields and mediated by spin-1 particles, the gauge bosons. The properties of these field quanta are summarized in table 2.1. Gravity as a fourth fundamental interaction cannot be included in the mathematical framework of the standard model. Its effects on the particles under study are negligible though. A different kind of charge is associated with each of the fundamental interactions. Only particles carrying these charges are affected by the corresponding interactions.

The spin-1/2 matter particles in the standard model are called fermions and come in two categories: quarks and leptons. Matter in the universe as we know it today is composed of these fermions. While quarks are affected by all three forces, the leptons do not take part in the strong interaction. The fermions are organized in pairs. There are three pairs of quarks and leptons each, as illustrated in tables 2.2 and 2.3, making up the three fermion generations. Ordinary matter usually consists of first generation fermions only. Each lepton generation is made up of one electrical charged and one neutral lepton. Because of their missing electric charge, the latter so-called neutrinos take part in the weak interaction only, which makes their detection very difficult even in experimental setups dedicated explicitly to neutrino physics. Each quark generation consists of one up-type quark with a non-integer electric charge of $\frac{2}{3}$ and one down-type quark of electric charge $-\frac{1}{3}$.

Interaction	Boson	Symbol	Mass	Electric Charge
Electromagnetic	Photon	γ	0	0
Weak	W	W^+/W^-	80.385 ± 0.015 GeV	1/-1
	Z	Z	91.1876 ± 0.0021 GeV	0
Strong	Gluon	g	0	0

Table 2.1: Force mediating bosons in the standard model [16].

Generation	Lepton	Symbol	Mass [MeV]	Electric Charge
1	Electron	e	$0.51 \pm (1.1 \times 10^{-8})$	-1
	Electron neutrino	ν_e	$< 2 \times 10^{-6}$	0
2	Muon	μ	$105.7 \pm (3.5 \times 10^{-6})$	-1
	Muon neutrino	ν_μ	$< 2 \times 10^{-6}$	0
3	Tau	τ	1776.82 ± 0.16	-1
	Tau neutrino	ν_τ	$< 2 \times 10^{-6}$	0

Table 2.2: Leptons in the standard model [16].

Generation	Quark Flavor	Symbol	Mass [GeV]	Electric Charge
1	Up	u	$(2.3^{+0.7}_{-0.5}) \times 10^{-3}$	$+\frac{2}{3}$
	Down	d	$(4.8^{+0.5}_{-0.3}) \times 10^{-3}$	$-\frac{1}{3}$
2	Charm	c	1.275 ± 0.025	$+\frac{2}{3}$
	Strange	s	$(95 \pm 5) \times 10^{-3}$	$-\frac{1}{3}$
3	Top	t	$173.34 \pm 0.27(stat.) \pm 0.71(syst.)$	$+\frac{2}{3}$
	Bottom	b	4.18 ± 0.03	$-\frac{1}{3}$

Table 2.3: Quarks in the standard model [16]. The quoted top-quark mass is taken from the recent combination of Tevatron and LHC measurements [17]. The other quark masses are quoted in the mass-independent subtraction scheme $\overline{MS}$ [16].

For every electrically charged particle, the standard model contains also a corresponding anti-particle, that has the exact same properties, except for the fact that it has an electric charge of the opposite sign. Whether anti-particles exist also for the standard model neutrinos, is not yet clear. Some models assume, that the neutrinos are their own antiparticles. Particles with this property are called Majorana particles, other particles are referred to as Dirac particles.

The physics of the standard model is described in the framework of Lagrangian field theory. The concepts of quantum mechanics and special relativity are merged into a quantum field theory. In classical mechanics, the Lagrangian of a physical system is given by $L = T - V$, where T and V are the kinetic and potential energy. This Lagrangian is used to describe discrete systems with coordinates $q_i(t)$. In the framework of the standard model, the so-called Lagrange density is used instead. It is a function of fields $\phi(x_\mu)$ with continuous parameters x_μ: $\mathcal{L}(\phi, \frac{\partial\phi}{\partial x_\mu}, x_\mu)$. The integral over the Lagrange density gives the action of the physical system. For simplicity, the Lagrange density is also commonly referred to as the "Lagrangian".

The standard model is built on the $SU(3) \times SU(2) \times U(1)_Y$ symmetry group and was developed based on the concept of gauge-symmetry transformations of fields such as $G^a_\mu \rightarrow G^a_\mu - \frac{1}{g}\partial_\mu\alpha_a$. Gauge invariance implies, that the corresponding Lagrangian is not affected by gauge transformations. The theories describing the interactions of the standard model are based on these symmetry groups. According to the Noether Theorem, any symmetry of the action of a physical system, i.e., the integral over its Lagrange density, corresponds to a conservation law. The conserved quantities corresponding to the interactions are the charges of the physical system.

2.1.1 Quantum chromodynamics

The strong interaction is described by the theory of quantum chromodynamics (QCD), which is based on the SU(3) symmetry group. It is mediated by massless gauge bosons named gluons and affects all particles that carry color charge. Three types of strong charges, or colors, and their corresponding anti-colors exist according to the standard model. While quarks only have a single color charge, and anti-quarks one anti-color charge per particle correspondingly, each gluon carries color and anti-color at the same time. There are eight gluons, one for each linear combination of colors and anti-colors that is not color neutral.

The phase transformation of the three quark-color fields q_1, q_2, and q_3 are described by the SU(3) group. The free Lagrangian

$$\mathcal{L} = \bar{q}_j(i\gamma^\mu\partial_\mu - m)q_j \tag{2.1}$$

with the colors $j = 1, 2, 3$ needs to be invariant under local color-phase transformations written as

$$q(x) \to Uq(x) \equiv e^{i\alpha_a(x)T_a}q(x). \tag{2.2}$$

The generators of the SU(3) group are eight linearly independent, traceless 3×3 matrices T_a. All of the elements of the SU(3) group can be expressed in terms of these generators. The corresponding group parameters are denoted by α_a. A covariant derivative

$$D_\mu = \partial_\mu + igT_aG_\mu^a \tag{2.3}$$

is introduced, as well as eight gauge fields G_μ^a representing the eight gluons. To ensure gauge invariance of the Lagrange density, the gauge fields need to transform as

$$G_\mu^a \to G_\mu^a - \frac{1}{g}\partial_\mu\alpha_a - f_{abc}\alpha_b G_\mu^c, \tag{2.4}$$

where the f_{abc} are the so-called structure constants of the group. Finally, the gauge-invariant Lagrange density of QCD is obtained via addition of a kinetic energy term for each of the gauge fields:

$$\mathcal{L} = \bar{q}(i\gamma^\mu\partial_\mu - m)q - g(\bar{q}\gamma^\mu T_a q)G_\mu^a - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}. \tag{2.5}$$

The requirement of local gauge invariance implies, that the gauge bosons of QCD, the gluons, are massless. Because of the non-Abelian structure of SU(3), the gluons themselves carry color charge. This allows for self-interaction between gluons, so that three or four gluon vertices can be realized. Vertices denote the interaction points between a number of particles. The self-interaction between gluons is a distinct feature of the strong interaction, these kind of vertices are not realized for photons or the Gauge bosons of the weak interaction. This leads to special property of the strong interaction: It grows stronger with increasing distance. For this reason, color-charged particles cannot be found in unbound states. Quarks only exist in bound states, the so-called hadrons. There are two types of hadrons: mesons consisting of quark-anti-quark pairs and baryons consisting of three quarks each. All hadrons are color neutral and can therefore exist as free particles.

In the attempt to separate color charged particles, new colored particles are generated. These newly generated particles then form additional color-neutral states. At very small distances, where the strong force is weaker, the interaction of the quarks is similar to that of free particles. This concept is known as asymptotic freedom.

2.1.2 The weak interaction

The theory of weak interactions describes the mixing between different quark generations. In the mass-eigenstate basis, quarks are represented as doublets consisting of a single up-type quark, and a linear combination of weak eigenstates of the down-type quarks:

$$\begin{pmatrix} \mathrm{u} \\ \mathrm{d'} \end{pmatrix} \begin{pmatrix} \mathrm{c} \\ \mathrm{s'} \end{pmatrix} \begin{pmatrix} \mathrm{t} \\ \mathrm{b'} \end{pmatrix}. \tag{2.6}$$

The linear combinations d', s', and b' are given by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix [18, 19]:

$$\begin{pmatrix} \mathrm{d'} \\ \mathrm{s'} \\ \mathrm{b'} \end{pmatrix} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix}. \tag{2.7}$$

In common formulations of the standard model, the neutrinos are assumed to be massless. Measurements of solar neutrinos, as well as results of other neutrino experiments, give compelling evidence of neutrino-flavor oscillations though, see for example [20, 21]. These oscillations are only possible if the neutrinos are massive. The mixing mechanism for leptons is then described by the PMNS matrix which has a similar structure as the CKM matrix.

2.1.3 The electroweak interaction and the BEH mechanism

As proposed by Glashow, Weinberg and Salam [22, 23], the weak and electromagnetic interactions are interconnected and can be unified into a single theory of electroweak interaction. The electroweak theory is based on the $SU(2) \times U(1)_Y$ gauge group. The Pauli Matrices τ_i are the generators of the SU(2) gauge group. The components of the weak isospin can be expressed in terms of the Pauli Matrices: $T_i = \frac{\tau_i}{2}$ (with i = 1,2,3). The three generators correspond to the three massless gauge fields W_μ^i. An additional single, massless gauge field B_μ is introduced to correspond to the hypercharge $Y = 2(Q - T_3)$, where T_3 denotes the third component of the weak isospin and Q the electrical charge. The hypercharge acts as generator of the abelian group U(1).

In 1957, Robert Marshak and George Sudarshan proposed a vector-axial vector (V-A) structure for the currents in the Lagrangian of the weak interaction [24]. In this framework, fermion fields are decomposed into their left-handed and right-handed components. The left-handed fermions are then placed in SU(2) doublets χ, while the right handed fermions appear as SU(2) singlets ψ. In this representation the leptons can be written as

$$\begin{pmatrix} \nu_i \\ \ell_i \end{pmatrix}_L, \quad \ell_{iR}, \quad \nu_{iR} \tag{2.8}$$

and, respectively, the quarks carrying color $\alpha = 1, 2, 3$ as

$$\begin{pmatrix} u_i^\alpha \\ d_i^\alpha \end{pmatrix}_L, \quad u_{iR}^\alpha, \quad d_{iR}^\alpha, \tag{2.9}$$

where $i = 1, 2, 3$ stands for the fermion generation.

In case of the left-handed doublets, the third component of the isospin is $T_3 \neq 0$. For the right-handed singlets on the other hand, T_3 is equal to 0. This implies, that only left-handed particles and right-handed anti-particles transform under SU(2) transformations: parity is not conserved in the weak interaction. This parity violation in weak interactions had been observed previous to the theoretical explanation in decays of $^{60}_{27}$Co in the Wu experiment in 1956 [25]. Finding a correct theoretical description for the observed behavior of weak interactions was rather challenging. The solution is to describe them as a combination of vector currents $\bar{\psi}\gamma^\mu\psi$ and axial-vector currents $\bar{\psi}\gamma^\mu\gamma^5\psi$. These currents behave differently under parity transformation. The V-A current for a left-handed fermion can be written as

$$\frac{1}{2}(\bar{\psi}\gamma^\mu\psi - \bar{\psi}\gamma^\mu\gamma^5\psi). \tag{2.10}$$

Parity is violated because of the interference of the vector and axial-vector terms in the interaction. Also so-called charge conjugation transformations which transform particles into their anti-particles are not allowed for charged particles in the weak interaction. These would involve particles and anti-particles of the same chirality which conflicts with the observations. However, if parity transformations are combined with charge-conjugation transformations, the resulting CP transformations are conserved.

The physical, neutral gauge fields A_μ and Z_μ, corresponding to the photon γ and the Z^0 boson, respectively, are orthogonal combinations of the gauge fields W_μ^3 and B_μ. They can therefore be written as

$$A_\mu = W_\mu^3 sin(\Theta_W) + B_\mu cos(\Theta_W) \tag{2.11}$$

$$Z_\mu = W_\mu^3 cos(\Theta_W) - B_\mu sin(\Theta_W) \tag{2.12}$$

with the Weinberg angle Θ_W, which quantifies the mixing between SU(2) and U(1). The charged W^+ and W^- bosons can be expressed in terms of the gauge fields W_μ^1 and W_μ^2:

$$W_\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \tag{2.13}$$

The physical $W^\pm$ and Z^0 bosons being combinations of massless gauge fields, would be expected to be massless as well. Experimental results conflict with this assumption though: the gauge bosons of the weak interaction have been shown to be massive. The most current measured values for the masses of the W and Z bosons are given in table 2.1. Also, the requirement of gauge invariance prohibits the addition of mass terms for the gauge bosons to the Lagrange density.

The W and Z bosons acquire their masses in a different way, in a mechanism first proposed by Brout, Englert and Higgs in 1964: the BEH mechanism [26–28], in which the W and Z bosons acquire their masses via spontaneous symmetry breaking.

In the formulation of the mechanism, four scalar fields are introduced that are arranged

in an isospin doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.14}$$

where

$$\phi^+ = (\phi_1 + i\phi_2)/\sqrt{2} \quad \text{and} \quad \phi^0 = (\phi_3 + i\phi_4)/\sqrt{2}. \tag{2.15}$$

The weak hypercharge of this doublet is Y = 1. The gauge-invariant Lagrange density of the scalar fields contains three massless gauge bosons $W^a_\mu(x)$, with a = 1,2,3; and the Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2. \tag{2.16}$$

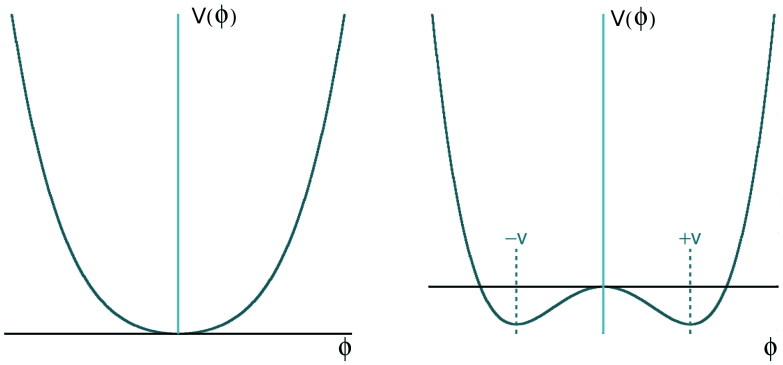

Figure 2.1: Higgs potential for different values of the parameter μ^2. Left: $\mu^2 > 0$, $\lambda > 0$. Right: $\mu^2 < 0$, $\lambda > 0$.

The choice of the parameters $\mu^2 > 0$ and $\lambda > 0$ determines the form of the potential. Two examples are shown in figure 2.1. Values $\mu^2 > 0$ and $\lambda > 0$ result in a potential that is symmetric with respect to the $V(\phi)$ axis as shown in the left plot of figure 2.1. This potential has a single absolute minimum at $\phi = 0$. If the values for μ and λ are set to $\mu^2 < 0$ and $\lambda > 0$, there is no longer a single minimum but a manifold of minimal values of the potential at $|\phi| > 0$. This manifold is invariant under SU(2) transformations. The choice of a single vacuum expectation value ϕ_0 for the fields $\phi(x)$ of, e.g.,

$$\phi_0 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.17}$$

breaks the SU(2) and $U(1)_Y$ gauge symmetries.
One can expand about the vacuum given in equation 2.17, and replace the fields in the Lagrange density with

$$\phi(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.18}$$

After this expansion, the theory has four degrees of freedom corresponding to the four scalar fields. In the local-gauge-symmetry breaking, three of these originally four degrees of freedom disappear in the mass acquisition of the three gauge bosons $W^{\pm}$ and Z^0. As the electromagnetic U(1) symmetry is not broken in this case, the photon is left massless. The remaining scalar field $h(x)$ can be identified as the Higgs field. The quanta of this Higgs field are the electrically neutral Higgs bosons H^0. Properties of the Higgs boson are described in section 2.1.5. The charged standard model fermions acquire their masses through Yukawa couplings to the Higgs field. For the charged leptons these couplings assume the form

$$\mathcal{L}_{Yukawa} = -G_\ell[\bar{\psi_R}(\phi^\dagger\psi_L) + (\bar{\psi_L}\phi)\psi_R] = -\frac{vG_\ell}{\sqrt{2}}\bar{\ell}\ell - \frac{G_\ell}{\sqrt{2}}\bar{\ell}\ell H, \tag{2.19}$$

where G_ℓ is chosen in such a way that the mass of the charged lepton $M_\ell = \frac{vG_\ell}{\sqrt{2}}$. Analogous terms appear for the Yukawa couplings of standard model quarks. The strength of the Yukawa couplings of fermions to the Higgs boson is proportional to the fermion mass in the framework of the standard model. The couplings of the W and Z bosons to the Higgs field are given by a kinetic term in the Lagrangian of the scalar fields of the form

$$\mathcal{L} = |D_\mu\phi|^2 = |\left(i\partial_\mu - g\frac{1}{2}\tau\cdot W_\mu - g'\frac{Y}{2}B_\mu\right)\phi|^2. \tag{2.20}$$

2.1.4 Properties of the top quark

The top quark plays a special role in the standard model of particle physics and in many models of physics beyond the standard model. It is distinguished from the other standard model fermions by its large mass. The current world-average measured value for the top quark mass is $173.34 \pm 0.27(stat.) \pm 0.71(syst.)$ GeV [17]. Large center-of-mass-energies are needed to produce such massive particles in experiments. Because of this, the top quark was discovered only in 1995 with the CDF and D0 experiments at the Tevatron proton-anti-proton collider at $\sqrt{s} = 1.9$ TeV [29, 30]. The large top-quark mass corresponds to an extremely short lifetime $\tau_t \propto \frac{1}{\Gamma_t}$ of $5 \cdot 10^{-25}$ s, which is smaller than the hadronization time scale. This means, that the top quark decays, before it can be bound in a hadron. In this way, kinematic information of the top quark is passed on to its decay products, without being distorted by hadronization effects.

In the standard model framework, top quarks are mainly produced in processes mediated by the strong interaction as particle-antiparticle pairs. At the LHC with its large center of mass energies, gluon fusion is the dominant production mode for $t\bar{t}$ pairs. This process is illustrated on the left-hand side of figure 2.2. Single production of top quarks via the weak interaction is also possible. The dominant process for weak production of top quarks is the t-channel production. The Feynman diagram for this production mode is shown on the right-hand side of figure 2.2. The most accurate measurement of the cross section for t-channel single top production at a center of mass energy of 8 TeV is $\sigma_{t-channel} = 83.6 \pm 2.3 \pm 7.1 \pm 2.2$ pb and was obtained from CMS data [31]. The most accurate measurement to date for the $t\bar{t}$ cross section was performed by the ATLAS collaboration and gives a value of $\sigma_{t\bar{t}} = 242.4 \pm 1.7 \pm 5.5 \pm 7.5 \pm 4.2$ pb [32]. The four individually quoted uncertainties are the statistical uncertainty, the systematic uncertainty arising from the general experimental and analysis setup, and the uncertainties in the

measurements of the integrated luminosity and the LHC beam energy.

The top quark decays almost exclusively to a W boson and a bottom quark via weak interaction, because the CKM-matrix element V_{tb} is ≈ 1. Top-quark decays are usually classified by the products of the consequent W-boson decay as either leptonic, in case of decays of $W \to \ell + \nu$, or hadronic, for decays of $W \to q\bar{q}'$. In the hadronic case, only decays to (u,d) or (c,s) are kinematically allowed. Since three different color charges can be carried by quarks, there are a total of six hadronic and three leptonic decay modes.

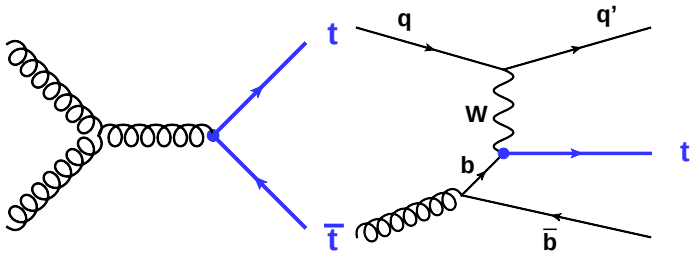

Figure 2.2: Mechanisms of top quark production. Left: top quark pair production via gluon fusion. Right: electroweak t-channel single top quark production.

2.1.5 Properties of the Higgs boson

From the first proposal of its existence in 1964, it took almost forty years for the Higgs boson to be discovered by the ATLAS and CMS experiments in 2012 [6, 7].

The CMS collaboration measured a value of $125.03^{+0.26}_{-0.27}(\text{stat.})^{+0.13}_{-0.15}(\text{syst.})$ GeV for the mass of the Higgs boson, using the full datasets recorded at 7 TeV and 8 TeV [33]. Measurements of $H \to \gamma\gamma$ and $H \to ZZ$ decays were used for the mass determination, as these yield the best resolution.

The main Higgs-production modes are shown in figure 2.3. Most Higgs bosons are produced via gluon or vector-boson fusion, but also the production in association with vector bosons or top quarks has a sizeable cross section.

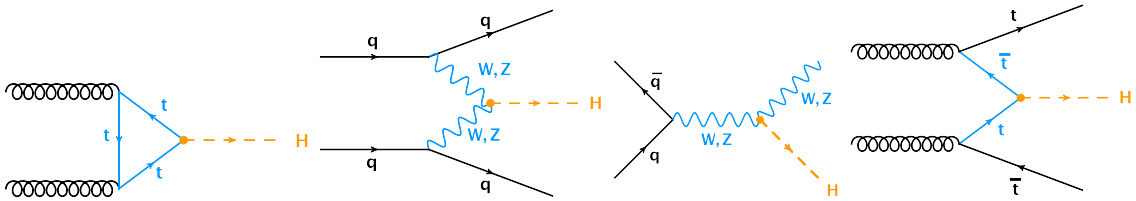

Figure 2.3: The main Higgs-boson production mechanisms. The Feynman diagrams are arranged according to the cross sections of the shown processes. In decreasing order of cross section from left to right: gluon fusion, vector-boson fusion, associated production of vector bosons, and associated production of top quarks.

The theoretical prediction for the branching fractions of the different decay modes of the Higgs boson strongly depend on the mass of the decaying Higgs boson. This is illustrated

in figure 2.4. At the measured mass of about 125 GeV, decays to bottom-quark pairs are most likely. Despite their considerably smaller branching fractions, the decay channels to Z-boson and photon pairs yield much higher sensitivity than the $H \to b\bar{b}$ channel though, as leptons and photons can be detected with much higher efficiency and better resolution in the CMS and ATLAS experiments. The suppression of background processes arising from QCD-multijet production is also much easier when the signal events contain isolated leptons.

To date, all measurements of properties of the newly discovered particle, including the particle mass, its spin, and the couplings to other standard model particles, are in agreement with the hypothesis, that this particle is indeed the standard model Higgs boson [33–36].

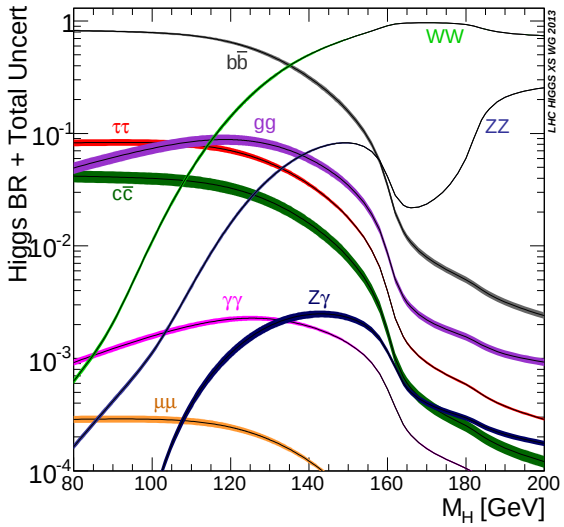

Figure 2.4: Predicted branching fractions for the different decay modes of the standard model Higgs boson with respect to the Higgs-boson mass [37].

2.1.6 Implications of the Higgs-boson discovery for models predicting a fourth generation of quarks

There is no intrinsic feature of the standard model that limits the number of quark generations to exactly three. An extension of the standard model quark sector to another generation is attractive because of its simplicity. With the measurements of the Higgs boson couplings at the LHC [33], strong limits have been set on these kind of models though. One important production mechanism for Higgs bosons is the production via fermion loops. The Yukawa couplings of standard model fermions to the Higgs boson are proportional to the fermion masses. Therefore, production in fermion loops involving new heavy, chiral quarks would give a sizeable contribution to the production cross section via gluon fusion. An enhancement of about a factor 9 would be expected due to the large masses of these hypothetical new quarks.

This expectation conflicts strongly with results of the measurement at the LHC, as

illustrated in figure 2.5. It shows the result of a combined fit of electroweak precision observables and the signal strengths in the Higgs-decay channels to $\gamma\gamma$, WW, ZZ, $b\bar{b}$, and $\tau\tau$ measured in LHC data, as well as the $p\bar{p} \to H \to b\bar{b}$ signal strength obtained from Tevatron measurements [9]. A model including a fourth generation of chiral quarks, that possess the same properties as standard model quarks, is very incompatible with the measured parameters, especially with the $H \to \gamma\gamma$ signal strength. Such a model is excluded at 5.3 standard deviations in this fit.

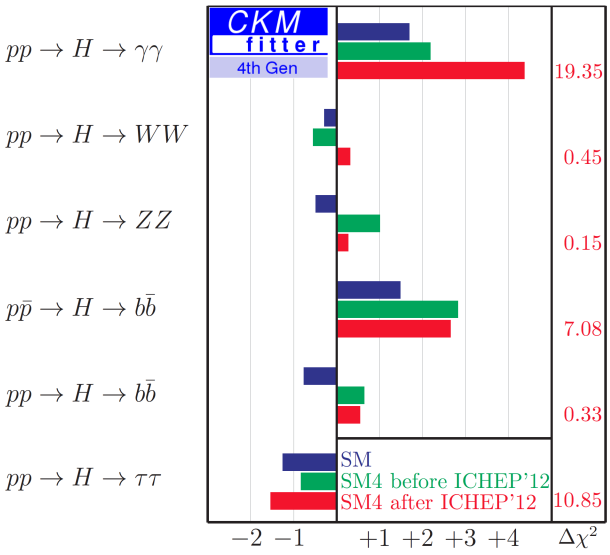

Figure 2.5: Result of a combined fit of electroweak precision observables and Higgs signal strength [9]. The statistical compatibility of the measured values of different Higgs-boson decay modes with the standard model (blue) and an extended standard model including a sequential fourth quark generation (red) is shown. The compabitibilty of a fourth-generation model when excluding the Higgs-signal-strength measurements at LHC is shown for comparison in green.

2.2 Vector-like quarks in physics beyond the standard model

The standard model of particle physics is found to be extremely successful at predicting particles and numerical values for other parameters. However, some issues are not addressed by the standard model in the current framework. One of the problems of the standard model is the instability of the Higgs-boson mass due to radiative corrections. These corrections are proportional to the square of the scale, up to which the theory is expected to be valid. They can therefore be much larger than the mass of the Higgs boson itself. Large contributions to the radiative corrections come from one-loop diagrams of particles with sizeable couplings to the Higgs boson. Such particles are top quarks, the gauge bosons of the weak interaction $W^{\pm}$, and Z, and the Higgs boson itself [10]. The most important one-loop corrections to the Higgs-boson mass are illustrated in figure 2.6. This issue is usually referred to as the "hierarchy problem" [8].

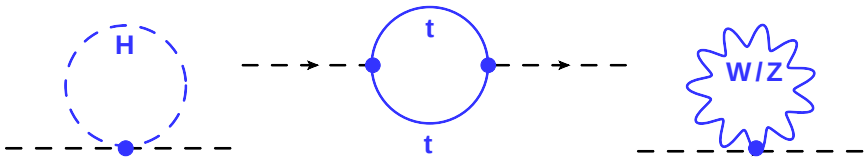

Figure 2.6: One loop corrections to the Higgs mass in the standard model.

A very popular model that can provide a solution of the hierarchy problem is the theory of supersymmetry (SUSY) [8,38,39]. In this model, the space-time symmetry is extended and supersymmetric partners for all standard model particles are introduced. Standard model fermions obtain bosonic partners, fermionic partners are predicted for the bosons of the standard model. In the original concept of SUSY, the supersymmetric particles are assumed to have masses identical to those of their standard model partners. This way, the loop corrections to the Higgs-boson mass would be cancelled in a very elegant way. To date, no evidence for the existence of SUSY has been observed [40, 41]. The SUSY particles are more massive than originally predicted and super-symmetry must be broken, in case SUSY does exist.

As no superpartners have been found at the energy scales examined so far, the loop corrections cannot cancelled entirely. In the Minimal Supersymmetric Standard Model (MSSM), a partial cancellation of the loop corrections is still expected though. Thus, the question of Higgs,boson mass stability is reduced to the so-called "little hierarchy problem". For energy scales Λ of about 10 TeV, i.e., scales in the order of magnitude of center-of-mass energies reached at the LHC, the top-quark loop contributes to the total Higgs-boson mass as

$$-\frac{3}{8\pi^2}\lambda_t^2\Lambda^2 \sim -(2\text{TeV})^2. \tag{2.21}$$

In order to avoid fine tuning of the corrections to the Higgs-boson mass above a 10% level, new physics beyond the standard model is needed to cut off the top-quark loop at a scale of $\Lambda_{top} < 2$ TeV. Otherwise, large divergencies cannot be prevented. There are several theories providing solutions to the little hierarchy problem without the introduction of supersymmetry. These models predict new particles with masses of a few TeV. They can affect the couplings of the standard model quarks to the Higgs boson and thus reduce the

impact of the top-quark loop contribution. These hypothetical new particles are similar to the top quark with respect to their quantum numbers and couplings to the Higgs boson and are therefore often referred to as "top partners". Candidate particles are vector-like quarks which are predicted by several models for physics beyond the standard model, such as little-Higgs models, models of extra-dimensions, and composite-Higgs models. The main ideas of these models are outlined in the following sections. More details on vector-like quarks can be found in section 2.2.4.

2.2.1 Little-Higgs models

One ansatz to the solution of the little hierarchy problem are little-Higgs models. In these models, there is no need for fine tuning up to a cut-off scale $\Lambda >> 1$ TeV and the theory stays perturbative up to this scale [42]. Little-Higgs models involve three scales Λ, f, and M_{weak}. They relate to each other as

$$\Lambda \sim 4\pi f \sim (4\pi)^2 M_{weak}. \tag{2.22}$$

Only physics at energies below the UV cut-off scale λ is described in little-Higgs models. Furthermore, the scalar mass is not affected by physics beyond this scale. The established standard model particles are found at the scale M_{weak}, while the supplementary particle content introduced in little-Higgs models is found at the scale f.

In little-Higgs models, the Higgs boson is a pseudo-Nambu-Goldstone boson of an approximate global SU(3) symmetry. This symmetry is spontaneously broken to a SU(2) symmetry [10]. The Nambu-Goldstone boson therefore corresponds to a SU(2) doublet field h. Nambu-Goldstone bosons do not participate in gauge interactions and do not have Yukawa couplings, though. In order to correctly describe the Higgs-boson properties expected in the standard model, these interactions have to be added to the model manually. Introduction of new one-loop quadratic divergences has to be avoided in this step, though. When gauging the SU(3) symmetry via introduction of new SU(3)-invariant, covariant derivatives containing the eight gauge bosons of SU(3), a quadratically divergent diagram does appear. However, this diagram does not contribute to the Higgs mass.

Nevertheless, this gauging gives rise to a new problem: the Nambu-Goldstone bosons, and with them the Higgs boson, disappear in the mass generation of the gauge bosons that correspond to the broken generators of SU(3). One therefore needs to introduce two versions of the Nambu-Goldstone bosons ϕ_1 and ϕ_2, and, consequently, two covariant derivatives to break the SU(3)×SU(3) symmetry. The resulting diagrams again involve only one of the two fields exclusively, therefore only one of them is "eaten" in the mass generation. The outcome of this procedure is a single Nambu-Goldstone boson, which can be identified as the Higgs boson.

No new quadratically divergent contributions to the Higgs-boson mass are introduced in this approach. The symmetry is broken collectively, which means, that neither of the couplings to ϕ_1 and ϕ_2 completely vanishes in the symmetry breaking. Because of this, no quadratically divergent terms can appear in the Higgs potential at one-loop level, as none of the quadratically divergent one-loop diagrams involve both of these couplings. However, the divergencies giving rise to the little hierarchy problem in the standard model have not been dissolved in this process yet.

The largest quadratic divergence to the Higgs-boson mass in the standard model is

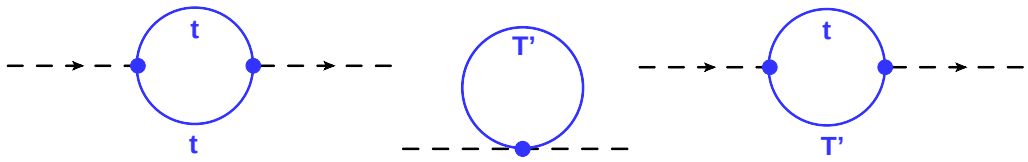

Figure 2.7: One loop corrections to the Higgs mass in the top sector of little-Higgs models.

caused by top-quark loops, because of the large Yukawa-couplings of the top quark to the Higgs boson. In little-Higgs models, new particles at the TeV scale f are predicted. The standard model quark doublets are extended to triplets $\Psi \equiv$ (t, b, T') containing a vector-like, heavy top partner T'. These triplets transform under the SU(3) gauge symmetry. The couplings of the Higgs boson to the top quark and the new vector-like T' quark are described by a relatively independent sector which is not influenced by the usual Higgs dynamics and vice versa [43]. The one-loop contributions to the Higgs-boson mass in this sector are shown in figure 2.7. The couplings of the top quark to the Higgs boson are modified with respect to the standard model values when adding the dynamics of the T' quark to the model.

2.2.2 Models of extra-dimensions

Another approach to the stabilization of the Higgs-boson mass is to supplement the four dimensions of space time by additional spacial dimensions. In models of universal extra-dimensions, all of the dimensions are pervaded by the physical fields. The extra-dimensions are compactified in orbifolds with radii R. This causes Kaluza-Klein excitations to appear which can be described as standing waves in these compactified extra-dimensions.

In the Randall-Sundrum Model, a fifth dimension is compactified on a $S^1/\mathbb{Z}_2$ orbifold. There is then a manifold with the ultraviolet boundary $y = 0$ and the infrared boundary $y = \pi R$. These names stem from the fact that the 4D fields at the infrared boundary are red shifted with respect to the ultraviolet boundary. The effective mass scales at the infrared boundary are of the order of TeV. In case the Higgs-field is localized at this boundary, the hierarchy problem is solved naturally. The extra-dimensions are hidden at low energy and the known standard model physics can be identified as the low-energy spectrum of the theory [12, 44].

For all fields of the standard model corresponding Kaluza-Klein excitations are introduced [45]. The so-called Kaluza-Klein mass towers consist of all possible invariant mass values for these resonances. In order to obtain the chiral zero-mode fermions of the standard model, one needs $S^1/\mathbb{Z}_2$ orbifolding. Vector-like towers of states exist above the chiral zero mode of each $SU(2)_L$ doublet or singlet field of the standard model. For the third generation of standard model quarks, three degenerate Kaluza-Klein excitations with electric charges of 5/3, 2/3, and -1/3 are predicted. These Kaluza-Klein excitations are vector-like quarks with large couplings to the third-generation quarks of the standard model [46].

2.2.3 Composite-Higgs models

In composite-Higgs models, the Higgs boson is assumed not to be an elementary particle. Instead, it is a bound state of new so-called ultra-fermions formed by a newly introduced ultra-color interaction [47, 48]. Here, the Higgs boson is the pseudo-Goldstone boson that corresponds to the spontaneous breaking of the global approximate symmetries associated with the ultra-color interaction. One possibility is the introduction of a $SO(5) \times U(1)_X$ symmetry [49].

Here, the hierarchy problem is not solved via loop cancellation as it is the case in previously presented models for physics beyond the standard model. Instead, the problem is simply avoided as the Higgs boson is treated as a pseudo-Goldstone boson. Therefore, its mass is protected by the global approximate symmetry of the theory. Thus, also the relatively small observed mass of the Higgs boson can be explained.

When constructing composite-Higgs models, new particles, the ultra-fermions, are introduced [11]. In contrast to the standard model fermions, these ultra-fermions are vector-like. The pseudo-Goldstone boson identified as the Higgs boson is a bound state of two of these ultra-fermions. They are subject to the weak SU(2)×U(1) interaction of the standard model and to a new force, the ultra-color force with a gauge group of higher order, e.g., SO(5).

In a first phase transition, the ultra-color force becomes strong enough to cause the formation of an ultra-fermion condensate. In this first phase transition the $SU(5) \times U(1)_X$ symmetry is broken down to a $SO(5) \times U(1)_X$ symmetry. The potential of the ultra-fermion condensate has several almost degenerate minima. Only some of them break the $SU(2) \times U(1)$ symmetry. In order for the latter symmetry not to be broken automatically in the formation of the ultra-fermion condensate, it must be possible to embed it in the new symmetry to which the original flavor symmetry of the ultra-color force is broken down. This is the case for $SO(5) \times U(1)_X$ symmetry.

Next, another weak gauge force described by a subgroup of the $SO(5) \times U(1)_X$ is defined. It destabilizes the vacuum, selecting a specific physical vacuum-expectation value. This leads to the spontaneous breaking of the $SU(2) \times U(1)$ symmetry, giving masses to the gauge bosons of the weak interaction.

2.2.4 Properties of vector-like quarks

The vector-like quarks appearing in all the previously described models are spin-1/2 particles that carry color charge. Their left-handed and right-handed representations have identical electroweak and color quantum numbers. Therefore, left- and right-handed particles exhibit indistinguishable behavior also under electroweak gauge-group transformations [50]. This implies, that the weak interaction has purely vectorial structure for these particles, as opposed to the V-A structure of the weak interaction of standard model fermions.

Vector-like quarks are not a simple extension of the standard model quark sector to a fourth generation of chiral quarks. Assuming perturbativity and a single Higgs doublet, these kind of theories have been excluded by the recent discovery of the Higgs boson as described in section 2.1.6. With the exclusion of a sequential fourth generation of quarks, vector-like quarks are the most simple possible extension of the fermion sector compatible with experimental data. In models predicting vector-like quarks, the Higgs boson can also

be produced in vector-like fermion loops. For vector-like fermions, the Yukawa couplings to the Higgs boson do not have to be proportional to the fermion mass though, because these particles can acquire their masses via a simple mass term in the Lagrangian of the form $\mathcal{L} = M\psi\bar{\psi}$. Therefore, their contribution does not have to affect the Higgs-production cross section in such a dramatic way. The expected changes are so small that they are well contained in the uncertainties in the measurements of cross section for Higgs-boson production to date. Vector-like quarks are not excluded by the discovery of the Higgs boson.

There are seven representations of vector-like quarks that lead to renormalizable couplings to the standard model particles and have definite $SU(3)_C \times SU(2)_L \times U(1)_Y$ quantum numbers at the same time: two singlets

$$\mathrm{T'}_{L,R} \quad \mathrm{B'}_{L,R}, \tag{2.23}$$

three doublets

$$(\mathrm{X, T'})_{L,R} \quad (\mathrm{T', B'})_{L,R} \quad (\mathrm{B', Y})_{L,R}, \tag{2.24}$$

and two triplets

$$(\mathrm{X, T', B'})_{L,R} \quad (\mathrm{T', B', Y})_{L,R}. \tag{2.25}$$

The charge of the T' quark is 2/3, the B' quark carries a charge of -1/3. The X and Y quarks have charges of 5/3 and -4/3 respectively and are therefore also referred to as "exotic vector-like quarks".

An addition of the $\mathrm{T'}_{L,R}$ field to the standard model framework can have an effect on the couplings of the up-type quarks, i.e., the up, charm, and top quark. Non-zero $\mathrm{T'}_{L,R}$ components lead to changes in the couplings of these quarks to the massive gauge bosons and the Higgs boson with respect to their standard model values [51,52]. Mixed couplings of standard model quarks and the new vector-like quarks to the W, Z, and Higgs bosons are introduced. These couplings can be described in terms of the same parameters defining the couplings of two standard model quarks to these bosons. Therefore, the latter are also affected when new vector-like particles are added to the quark sector. Measurements of standard model parameters place strong experimental constraints on any deviations of the couplings to the W, Z, and Higgs bosons for the up quark and the charm quark. Not much information about the couplings to the top quark can be derived from these fits though, leaving these couplings less constrained [53]. Therefore, it is a valid assumption that the mixing of third generation standard model quarks with the heavy vector-like T' quark is dominant.

In case of doublet or triplet representations of the vector-like quarks, the heavy new particles share a single mass term in the Lagrange density. The mixing with the third generation of standard model quarks leads to a mass splitting between the heavy new quarks though: $m_{\mathrm{T'}} \geq m_{\mathrm{X}}$, $m_{\mathrm{B'}} \geq m_{\mathrm{Y}}$, the T' quark can be either heavier or lighter than the B' quark in different models. The mass difference between the different particles does not exceed a few GeV. Therefore, decays between the different vector-like quarks are suppressed. The possible decay modes of the vector-like T' quark are shown in figure 2.8. The decay modes of the T' quark are T'→tH, T'→tZ, and T'→bW; those of the B' quark are B'→bH, B'→bZ, and B'→tW.

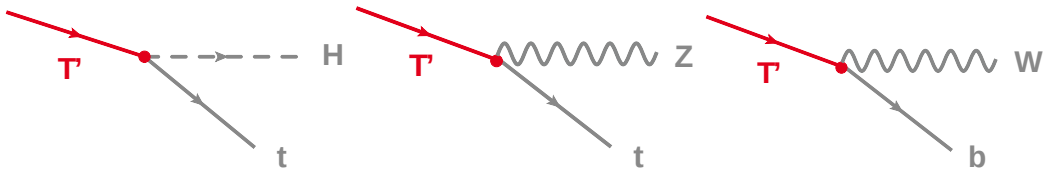

Figure 2.8: Decay modes of the vector-like T' quark.

As vector-like quarks carry color charge, pairs of these particles can be produced in simple QCD interactions. Feynman diagrams for the pairwise production mode are displayed in figure 2.9.

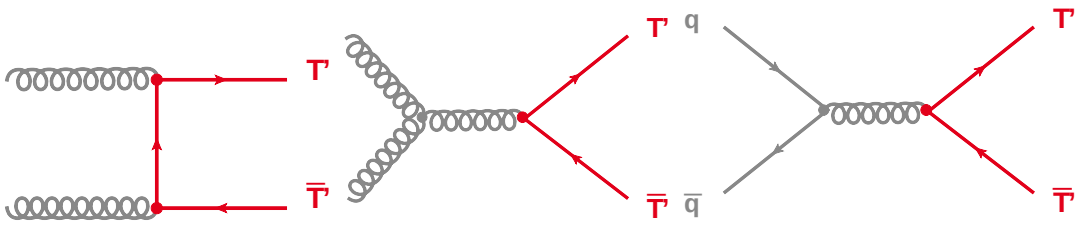

Figure 2.9: Feynman diagrams of T'-quark pair production.

As shown in figure 2.11, the cross section for pair production of vector-like quarks is the same for all types of vector-like quarks and depends only on the mass of the produced particles [54]. In addition, vector-like quarks can be produced singly in association with top or bottom quarks. Exemplary Feynman diagrams for single T'-quark production can be found in figure 2.10.

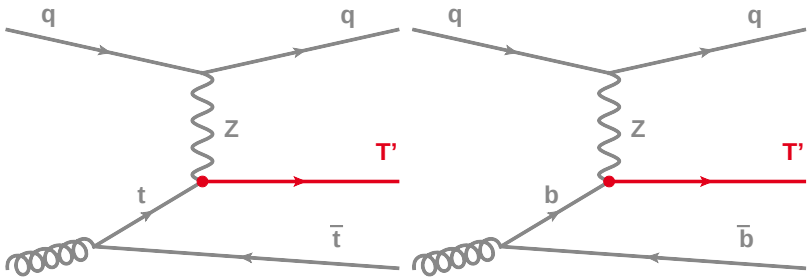

Figure 2.10: Feynman diagram of single T'-quark production in association with a top quark (left) and a bottom quark (right).

The calculation of the production cross section for singly produced T' quarks is more complicated than in the case of pair production. In single production, the cross section depends on a number of model parameters, e.g., the couplings to the standard model gauge bosons or whether the produced particle is a singlet or part of a doublet or triplet. It also differs for the different quark types.

The cross section for single production of vector-like T' quarks is constrained by mea-

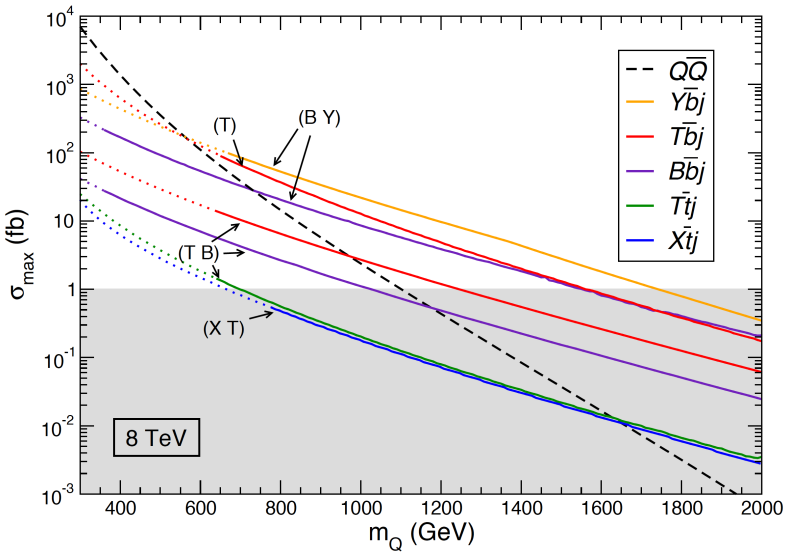

Figure 2.11: Maximum production cross sections of vector-like quarks at the LHC with 8 TeV. The black dotted line corresponds to the pair-production cross section, which is identical for all types of vector-like quarks. The cross sections for single production of the different quarks are drawn as colored lines. In mass ranges excluded by direct searches, the cross section is drawn as a dotted line. The area shaded in gray marks cross sections below 1 fb, which are out of reach for analyses of the collected approximately 20 fb$^-$1 of data at $\sqrt{s}$ =8 TeV [50].

surements of electroweak precision observables and findings in flavor physics [50, 55, 56]. It is proportional to the couplings of vector-like quarks to the W and Z bosons. Mixing of bottom quarks with vector-like B' quarks affects the Zb$\bar{\text{b}}$ coupling at tree level. Furthermore, modifications at one-loop level are introduced by a mixing of top quarks and vector-like T' quarks. Changes in the Zb$\bar{\text{b}}$ vertex lead to different values for four of the so-called Z-pole parameters: the forward-backward asymmetry A^b_{FB}, the asymmetry parameter A_b, the hadronic branching fraction $R_b = \Gamma(\text{Z} \to \text{b}\bar{\text{b}})/\Gamma(\text{Z} \to \text{hadrons})$, and the analogously defined hadronic branching fraction R_c. Fits of the Z-pole variables are frequently used to test models for physics beyond the standard model [57].

Another set of parameters used to constrain the single vector-like quark production rate are the oblique parameters S and T. These parameters are so-called Peskin-Takeuchi parameters [58] used in certain parameterizations of radiative corrections to the electroweak sector. They are obtained in fits to electroweak precision data. The contributions ΔS and ΔT of vector-like quarks to the two oblique parameters can be computed for any model predicting vector-like quarks.

Measurements of these two sets of parameters are used to constrain the mixing of the vector-like quarks with the standard model quarks of the third generation, and, consequently, the production cross section for single vector-like quarks. In case the vector-like quark multiplets do not contain B' quarks, the constraints from the oblique parameters

are most powerful. In the other case, there are tree level corrections to the $Zb\bar{b}$ vertex introduced by the mixing of the B' quark with the bottom quark. This results in stronger limits from the Z pole variables.

Figure 2.11 shows the maximum production cross sections for vector-like quarks at a center of mass energy of 8 TeV that have been calculated using PROTOS at tree level [50, 59]. Cross sections are shown for the multiplets with the largest allowed mixing to standard model quarks taking into account the constraints described above. The model and particle-type independent cross section for pair production is drawn as a black, dashed line. The different single production cross sections are indicated by colored lines. At low quark masses, the pair production is expected to be dominant. The production cross section in general decreases for higher masses of the vector-like quarks, but the slope of the single production cross section curves is much smaller than for the pair-production cross section. Therefore, the contribution of singly produced vector-like quarks becomes more important with increasing quark masses.

Three different models for single T'-quark production are included in this plot. The largest single-production cross section for the T' quark can be expected for T' singlets produced in association with a bottom quark and a jet. This production channel is expected to become dominant over the pair production already for T' quark masses slightly above 600 GeV. The production cross section for T' quarks in (T', B') doublets is about one order of magnitude smaller, surpassing the pair-production cross section at T' quark masses of about 1 TeV. In addition, production in association with a top quark and a light jet is possible for T' quarks from (T', B') doublets, and also for those in (X, T') doublets. For this production mode, the cross section is very small though. Already for masses of about 700 GeV the cross section falls below 1 fb. With the approximately 20 fb^{-1} of data collected at LHC at $\sqrt{s} = 8$ TeV, no sensitivity for processes with such small expected cross sections can be achieved.

In the future, the LHC will be operated at larger center-of-mass energies. Calculations are also available for $\sqrt{s} = 13$ TeV [50]. Overall, the cross sections for vector-like quark production increase with respect to those at $\sqrt{s} = 8$ TeV. A more detailed discussion of the impact of the planned increase of center-of-mass energy on searches for vector-like quarks and further aspects important to future analyses can be found in chapter 9.

As the simulation of pair production of vector-like quarks is less complicated because of its model independence, most experimental searches to date focus on this production mode. The interest in single production of these particles is rising though.

3 Experimental setup

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a proton-proton ring accelerator and collider located in the proximity of Geneva, Switzerland, at the European Centre for Nuclear Research (CERN). A very detailed description of the LHC can be found in [60].

The circular tunnel containing the LHC has a circumfence of about 27 km and was originally built for LEP, an electron-positron collider which was operated by CERN from 1989 until 2000. The design center-of-mass-energy of the LHC is $\sqrt{s} = 14$ TeV. During the first three years of runtime, it has been operated at lower center-of-mass-energies of $\sqrt{s} = 7$ TeV in the 2010/2011 data taking and $\sqrt{s} = 8$ TeV in 2012. After the technical shutdown in 2013/2014, operation will resume at a center-of-mass-energy of $\sqrt{s} = 13$ TeV in the spring of 2015.

As the LHC is a particle-particle accelerator, the two beams counter-rotate in separate pipes before they are brought to collision in one of the four interaction points. Before being injected to the actual LHC rings, the particles pass through a chain of pre-accelerators which is sketched in figure 3.1 and described in great detail in [61].

The LHC design was restricted by the constraints given by the pre-existing single LEP tunnel. Because of the rather small diameter of the tunnel, a "two-in-one" design was adapted for the super-conducting magnets of the LHC: the magnetic flux circulates through both beam channels, while a single cold mass and cryostat contains the windings for both channels. Also, the geometry of the tunnel made an accelerator design with eight rather long straight and eight curved segments necessary. While the long straight segments were needed to reduce the energy loss due to synchrotron radiation for LEP, a proton-proton collider would ideally consist of longer curved segments.

Six large experiments have been built for the LHC. The experimental caverns for the two high-luminosity, multi-purpose detectors "Compact Muon Solenoid" (CMS) [63] and ATLAS [64] are at straight sections of the accelerator on opposite sides of the ring. The "Total Elastic and Diffractive Cross Section Measurement" (TOTEM) [65] experiment is also located in the experimental cavern of CMS, while the "Large Hadron Collider forward" (LHCf) [66] experiment is contained in the ATLAS cavern. The purpose of TOTEM is to measure particles at very small angles with respect to the beampipe. LHCf is built to measure properties of neutral pions. At the two other collision points there are the LHCb experiment [67], which was designed specifically for the measurement of physics involving bottom quarks, and "A Large Ion Collider Experiment" (ALICE) [68] which examines collisions of heavy ions.

To deliver a peak instantaneous luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} for the high-luminosity experiments CMS and ATLAS is the design goal of the LHC. The instantaneous luminosity can be expressed as:

$$\mathcal{L} = \frac{N_b^2 \; n_b \; f_{rev} \; \gamma_r}{f \; \pi \; \epsilon_n \; \beta^*} \cdot F \tag{3.1}$$

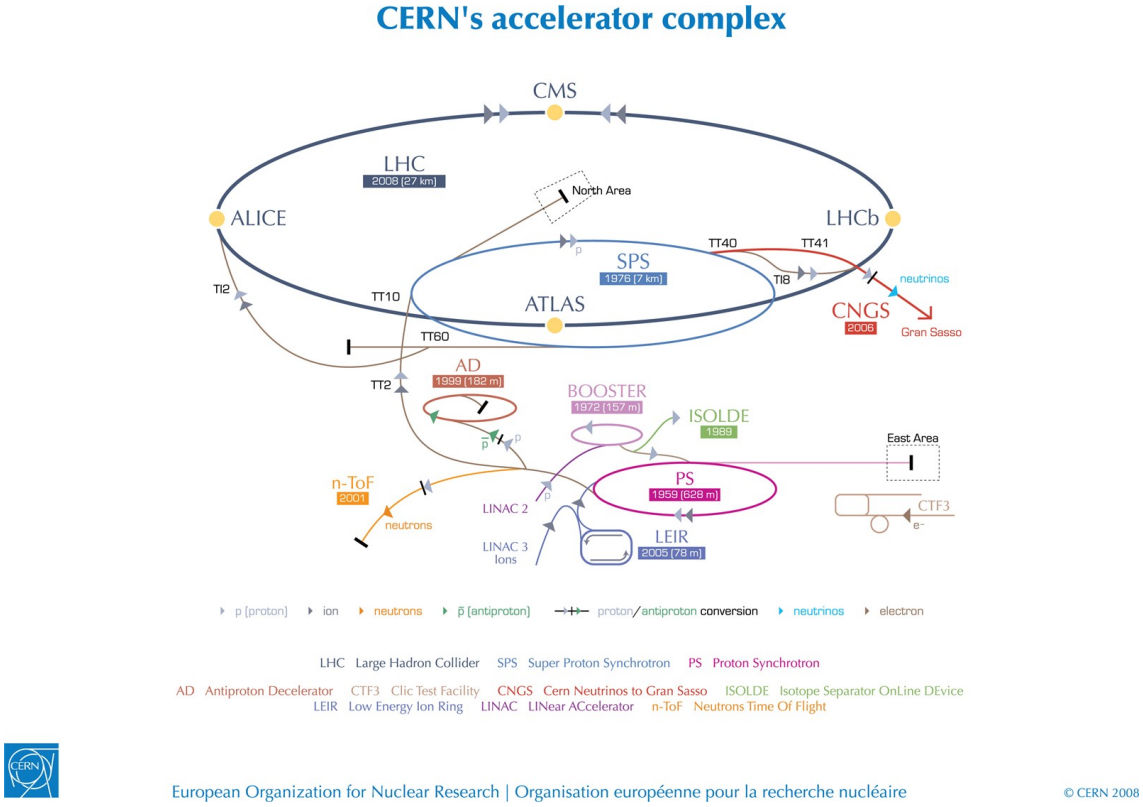

Figure 3.1: Layout of the CERN accelerator complex [62].

with the number of particles per bunch N_b, the number of bunches per beam n_b, the frequency of the beam revolution f_{rev}, the relativistic gamma factor γ_r, the normalized transverse beam emittance ϵ_n, the beta function at the collision point β^*, and a geometric factor F to take into account the luminosity reduction due to the crossing angle at the interaction point [60]. For a given process with the cross section $\sigma_{process}$, the number of events generated per second, n_{events}, relates to the machine luminosity as

$$n_{events} = \mathcal{L} \cdot \sigma_{process}. \tag{3.2}$$

While high instantenous luminosities are beneficial for generating large event rates, they introduce complications at analysis level. At high instantenous luminosities, there is a large probability for so-called pileup to occur. If more than one proton-proton collision takes place in a single beam crossing, the contributions to the event content coming from these additional processes are denoted as pileup. Coping with the pileup contamination of the events recorded in the high-luminosity environment of the LHC is one of the special challenges in analyzing LHC data.

The integrated luminosity $L = \int \mathcal{L} dt$ yields the total number of events in a certain timespan. The integrated luminosity recorded by the CMS experiment in the course of operation at the center-of-mass-energy of 8 TeV in 2012 amounted to 19.7 fb^{-1}.

3.2 The CMS experiment

The main motivation in designing the CMS detector was to facilitate the search for the Higgs boson and thus confirm the theory of electroweak-symmetry breaking. The versatility of the CMS layout allows for the measurement of a wide variety of physics processes though, ranging from precision measurements of standard model parameters to searches for physics beyond the standard model. The name of the detector mirrors the main concept of the detector design: it is based on the powerful solenoid magnet, which allows for extremely precise momentum measurements of charged particles, and employs a high-resolution muon detection system. The cylindrical layout of the CMS detector is sketched in figure 3.2. It is 21.6 m long, has a diameter of 14.6 m and a total weight of 12500 t. The different detector systems are built in layers in the barrel region of the detector as well as in the two endcaps.

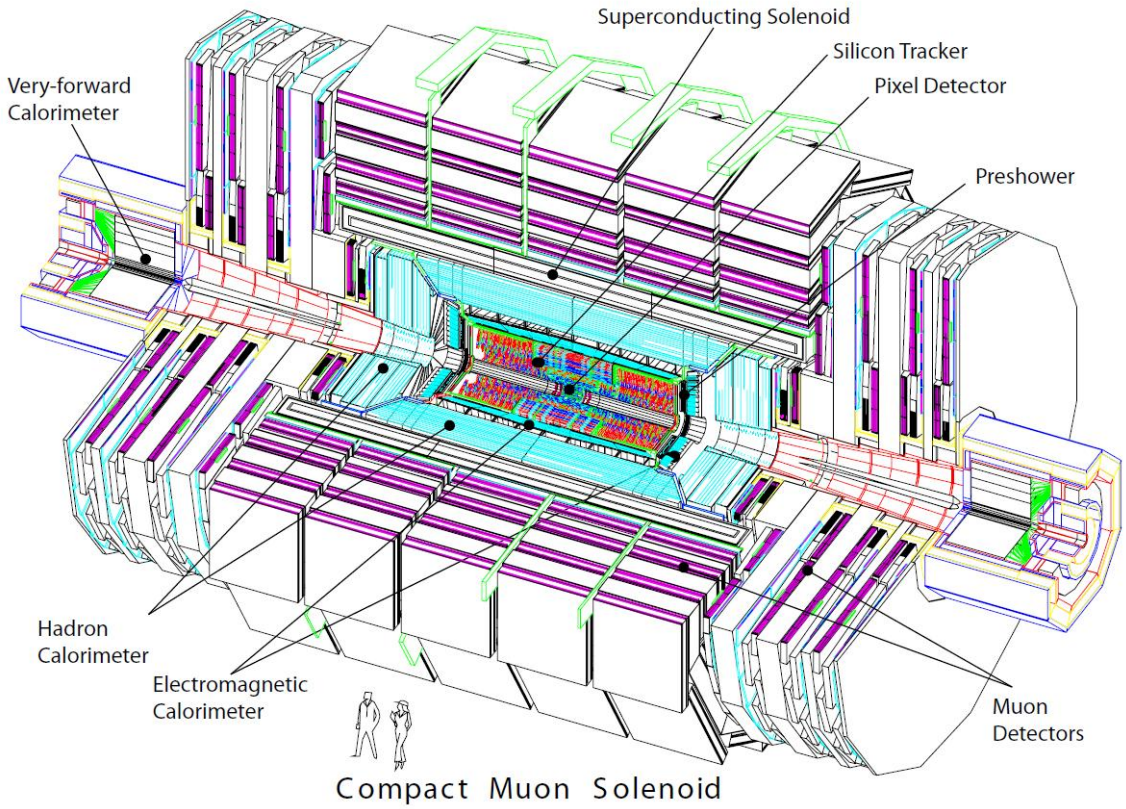

Figure 3.2: Illustration of the CMS detector [63].

The beam-collision point at the center of the CMS experiment is enclosed by the innermost component of the detector: the silicon tracking system. It is followed by the calorimetry system, which measures the energy of particles. The inner electromagnetic calorimeter (ECAL) is optimized for measurements of electrons and photons, the hadronic calorimeter (HCAL) for that of hadrons. Both, the tracking and the calorimetry systems

are contained within the super-conducting solenoid magnet. The cylindrical magnet of 12.5 m length and 6 m diameter stores energies up to 2.6 GJ. With four layers of winding, a field of 3.8 T is generated. A 10000 t iron yoke returns the magnetic flux. The large amount of iron is the main contribution to the large weight of the detector. The yoke is built in five rings of 2.536 m width, which can be moved individually to facilitate access to internal detector systems for maintenance. The four-layer muon system is placed within the iron flux-return yoke structure. Each detector system is described in more detail below.

The coordinate system used for description of detector properties and in analysis of physics events is tailored to the cylindrical design of the CMS detector. The nominal interaction point at the center of the detector defines the origin of the coordinate system. The x-axis is directed radially inwards, pointing from the origin towards the center of the LHC collider ring. The y-axis points upwards, while the z-axis is aligned with the beam line in anti-clockwise direction. The azimuthal angle ϕ is measured in the (x,y)-plane with respect to the x-axis, the polar angle θ is given in the (y,z)-plane with respect to the z-axis. Often, θ is replaced by the pseudo rapidity η, which is defined as:

$$\eta = -ln\left(tan\left(\frac{\theta}{2}\right)\right). \tag{3.3}$$

Differences in pseudorapidity $\Delta\eta$ are Lorentz invariant. Therefore, using the pseudorapidity instead of θ is advantageous for applications in high-energy physics.

3.2.1 Tracking system

The purpose of the tracking system is to precisely measure particle trajectories. The high luminosity delivered by the LHC calls for very radiation-hard materials at this innermost part of the detector. High resolution and fast response are also needed to reconstruct the particle trajectories and match them to the correct bunch crossing. The challenge in designing the tracking system is the trade-off between the desired fine granularity as well as fast response on one side and, on the other side, the requirement to keep the material budget for read-out and cooling technology as low as possible. To accommodate these needs, the CMS tracking system is based entirely on silicon-detector technology.
Figure 3.3 depicts the layout of the tracking system. Directly surrounding the collision point, the pixel detector is made up of three layers of silicon-pixel modules in the barrel region and two layers each in the two endcaps. With its 66 million pixels it can provide three very precise measurements of a single particle track.

The silicon-strip detector is designed in two parts. The inner barrel tracker (TIB) covers the radius from 20 to 55 cm and is made up of four layers of silicon-strip detectors, while the corresponding inner endcaps of the tracker (TID) consist of three layers each. The point resolution of the first two layers of silicon-strip modules in the barrel is 23 μm, that of the two outer barrel layers and the endcaps 35 μm. Another six layers of silicon-strip modules extend to a radius of 115 cm from the interaction point and make up the tracker outer barrel (TOB). Here, the single-point resolution varies between 35 and 53 μm. The tracking system is completed by the outer endcaps (TEC), each of them consisting of nine layers that carry up to seven rings of silicon-micro-strip detectors. To enable measurements of the z-coordinate in the barrel region and the r-coordinate in the endcaps, additional strip modules are mounted at a stereo angle of 100 mrad and back-to-back to the modules in the first two layers of the TIB, TOB, and their endcaps. This leads to a single-point

resolution of 230 μm in the TIB and 530 μm in the TOB. In total, the silicon-strip detector of CMS has an active silicon area of 198 m^2 made up of 9.3 million strips.

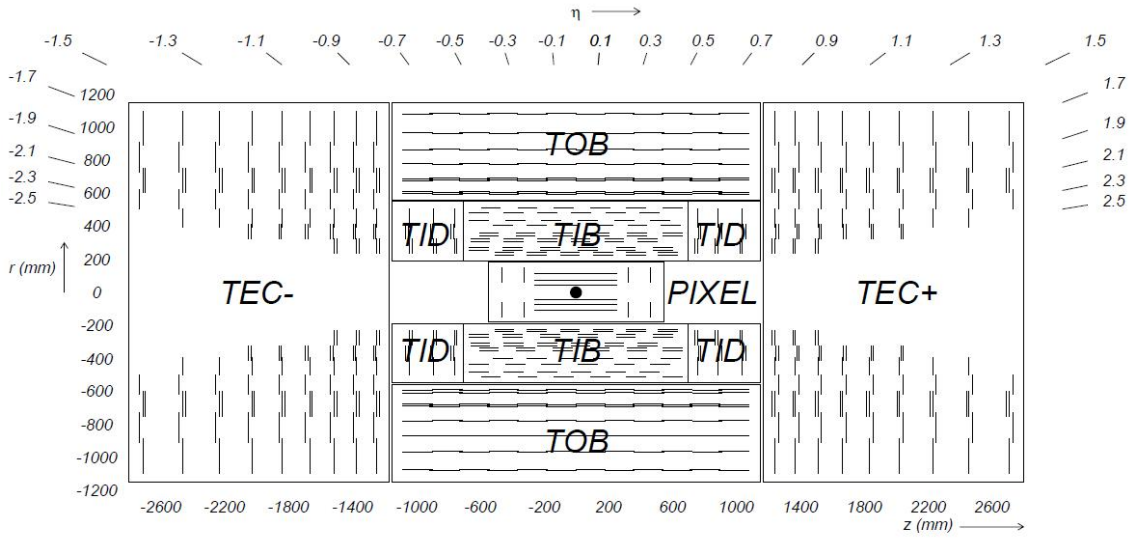

Figure 3.3: Layout of the CMS tracker system [63].

3.2.2 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) of the CMS detector is a lead-tungstate ($PbWO_4$) crystal calorimeter with a hermetic, homogeneous layout. The design goal for this part of the detector was to facilitate the detection of H$\rightarrow \gamma\gamma$ decays. The short radiation length (0.89 cm) and small Molière radius (2.2!cm) of lead tungstate allow for a very fine granularity. The material also has a very high density of 8.28 g/cm^3, thus accommodating the space limitations within the solenoid of the CMS experiment. The short scintillation-decay time of the used crystals is of the same order of magnitude as the LHC bunch-crossing time. The radiation hardness of the material helps to keep damages to the calorimeter at a minimum in spite of the high intensity of radiation relatively close to the interaction point. It is not possible to fully avoid radiation damage leading to wavelength-dependent loss of transparency in the crystals though. Therefore, the transparency of the crystals is monitored by laser-light injection. Correction factors for the ECAL measurements are applied to account for this effect.

The barrel region of the ECAL is made up of 61200 crystals, which corresponds to 360 divisions in ϕ and 2×85 in η. Each crystal has a length of 23 cm corresponding to 25.8 X_0, where X_0 denotes the electromagnetic interaction length of the material. The total barrel volume amounts to 8.14 m^3. The barrel ECAL is complemented by two endcaps which cover the rapidity range of $1.479 < |\eta| < 3.0$. Each endcap is made up of 7324 crystals with a length of 22 cm (24.7 X_0), amounting to a total volume of 2.9 m^3. The pyramidical shape of the individual crystals in the barrel region results in non-uniform light collection. To reduce this effect, one lateral face of the crystals is depolished. The endcap is not affected by this because the crystals in this region are placed almost parallel

to each other.

The light emission of lead-tungstate crystals is highly temperature dependent. Therefore, the ECAL is thermally screened from the adjacent tracking systems and readout electronics. A dedicated cooling system is installed to keep the crystals at a temperature of exactly 18 °C. The light output of the calorimeter material is relatively low and needs to be amplified by photo detectors. Not all types of photo detectors are suited for the task of operating in the 3.8 T magnetic field of the CMS solenoid. Two kinds of photo detectors are used in the ECAL system: two avalanche photo diodes (APD) are mounted to each crystal in the barrel region, while a single vacuum photo triode (VPT) is applied at the back of each crystal in the endcap regions.

In addition to the crystal calorimeters, there are pre-shower detectors mounted in front of the endcaps that cover the range of $1.653 < |\eta| < 2.6$. Here, two-layer sampling calorimeters are used for detection. A first layer of lead in which photons and electrons initiate electromagnetic showers is followed by a second layer consisting of silicon-strip sensors. Here, the shower characteristics are measured. The pre-shower calorimeter has a thickness of 20 cm.

3.2.3 Hadron calorimeter

The hadron calorimeter (HCAL) is particularly important for the measurement of hadron jets and so-called missing transverse energy, an imbalance in the total measured transverse energy caused by neutrinos or weakly interacting particles in models for new physics, which leave the detector without interacting with its material. The HCAL is placed in the remaining space between the electromagnetic calorimeter and the magnet coil. This spatial restriction limits the amount of material that can be used to absorb the hadronic shower. As this may not be sufficient for highly-energetic hadron showers, an additional outer hadron calorimeter (HO) system is installed outside the solenoid to measure the tails of these showers.

The inner part of the HCAL consists of sampling calorimeters with brass absorber plates. 18 wedges are assembled to form detector rings, making the detector hermetic in ϕ. This is of special importance for measuring the missing transverse energy of an event. The thickness of the absorber plates ranges from 50.5 mm to 56.5 mm. The plastic scintillators between these absorber plates are segmented in η and ϕ , where $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. This corresponds to a total of 70,000 tiles. The 16 absorber layers and the scintillator plates in each of these (η, ϕ) segments make up the so-called HCAL towers. Figure 3.4 shows the segmentation in the (r,z)-plane. As the towers are not built perpendicular to the beam pipe, their total thickness increases with η. The innermost and outermost layers of each segment are made of stainless steel to ensure the stability of the detector system.

The detector region covered by the inner HCAL endcaps incorporates about 34% of all particles in the final state. This leads to very high event rates and calls for radiation-hard material. A precise measurement of single particles in the endcaps is challenging due to large pileup and magnetic-field effects. Therefore, the emphasis was laid upon the hermetical design, closing all gaps between the inner barrel HCAL and the endcaps. Thicker brass-absorber plates of 79 mm are installed in this part of the detector. The scintillator material is contained in 9 mm wide spaces between the absorber plates.

In the central pseudorapidity region, the width of the inner calorimeter towers does not suffice to absorb all hadronic showers. In order to also measure the tails of highly-energetic

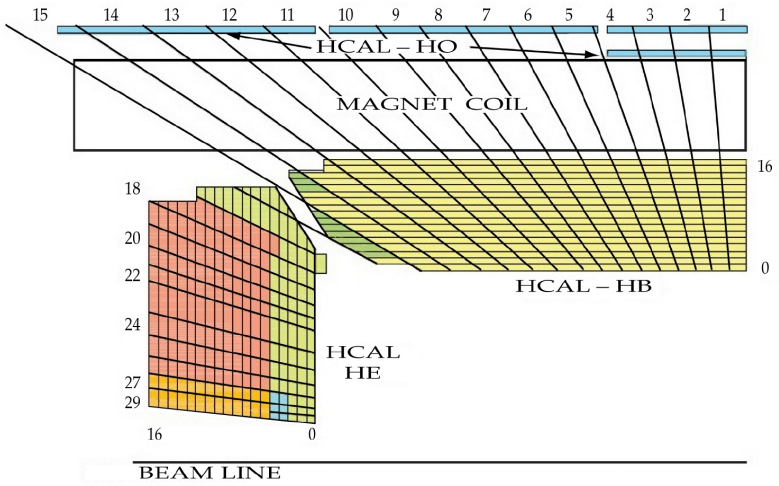

Figure 3.4: Illustration of the HCAL tower segmentation in the (r-z)-plane [63].

showers that extend beyond the inner HCAL, another calorimeter system is installed as the first detector component outside of the solenoid. It functions as an additional absorber for the hadron showers. The layout of this outer hadron calorimeter (HO) is heavily constrained by the spatial requirements of the muon system, which is described in the section below. Taking into account the need for support structures, only a 16 mm space remains for the scintillator tiles of the HO. The absorber depth of the inner HCAL and solenoid is significantly smaller in the central η region. To compensate this, an additional 19.5 cm thick iron absorber plate with a scintillator layer on each side is installed here.

Including the outer hadron calorimeter, the depth of the HCAL is at least 11.8 interaction lengths λ_I over the full covered η range, except for the barrel-endcap boundary region. By also accounting for the tails of hadronic showers, the accuracy of energy measurements is considerably improved and thus also the measurement of the missing transverse energy. The HCAL system is completed by a forward hadron calorimeter at a distance of 11.2 m from the interaction point. It covers a range of $3 < |\eta| < 5.2$ very close to the beampipe. Here, Cherenkov-based, very radiation-hard technology is put to use.

3.2.4 Muon system

The muon system of the CMS experiment can reconstruct momentum and charge of muons over the full kinematic range of the LHC. Good muon identification is crucial for a large number of physics analyses. The muon system is the outermost component of the CMS detector. Therefore, it is screened from most other particles by the hadron calorimeter and the solenoid magnet. This clean environment and the fact that muons are not much affected by radiative losses when crossing the inner detector systems make their detection relatively easy. The placement of the muon chambers within the iron flux-return yoke yields a very good muon-momentum resolution.

The layout of the muon system is illustrated in figure 3.5. Gaseous particle detectors are used in the muon system. The detection planes in the barrel section and the two endcaps add up to 25,000 m^2. In the barrel region of η <1.2 the magnetic field is uniform and

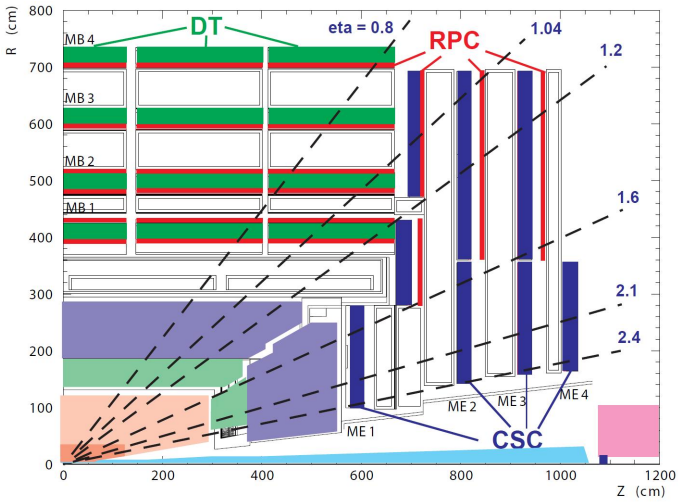

Figure 3.5: Layout of the CMS muon system [63].

mainly contained in the return yoke. Here, drift tubes are used for the muon detection. They are organized in four stations that are separated from each other by layers of the flux-return yoke. In each of the four stations the coordinate in the (r-ϕ)-plane is measured, while a measurement of the z-coordinate is only performed in the first three stations.

In the endcap region between $0.9 < |\eta| < 2.4$, four stations of cathode strip chambers (CSC) are installed. CSCs have fast response time and a fine segmentation and are therefore better suited to operate in this region with higher background rates and rather large, non-uniform magnetic field than the drift chambers used in the barrel region. The offline-muon-reconstruction efficiency measured in simulation is typically 95-99% except in transition regions between the drift chamber and CSC systems, and the transition regions between the wheels of the iron flux-return yoke.

In addition to this muon-detection system, a complementary system of resistive plate chambers (RPC) is installed in the barrel and endcap regions for trigger purposes. The RPCs are operated in avalanche mode and provide a fast trigger with a sharp transverse-momentum threshold in the pseudorapidity range of $|\eta| <1.6$. The transverse momentum is defined as the fraction of the total momentum of an object that lies in the plane perpendicular to the beam axis. There are six layers of RPCs in the barrel region and three in each endcap.

3.2.5 Trigger system

The LHC is designed to provide high instantaneous luminosity to its experiments. In the 2012 data taking instantaneous luminosities up to 7.7 Hz/nb were recorded. It is impossible to store and process the corresponding vast amount of data. Therefore, the number of events is reduced by the trigger system which only marks events to be further processed if they contain certain types of physics objects. It operates in two steps: the hardware-based level-1 (L1) trigger and the software-based high-level trigger (HLT).

The L1 trigger consists of custom-designed electronics that analyze coarsely-segmented data from the calorimeter and muon systems. The complete high-resolution data is held

in memories for the time needed for evaluation by the L1 trigger. The functioning of the L1 trigger can be described in three tiers: first, the events are rated by the trigger primitive generators (TPGs), based on entries of a single detector system, for instance calorimeter trigger towers or track segments. Then, the regional triggers use the information provided by the TPGs to determine object candidates through pattern logic. These object candidates are ranked according to energy or momentum and quality. Finally, the global calorimeter trigger and global muon trigger identify the highest-ranking calorimeter or muon objects and pass them to the global trigger. Here the actual decision whether to keep the event is taken and it is stored for evaluation by the HLT. Large parts of the L1 trigger electronics are installed in the service cavern neighboring the actual experimental cavern of the CMS experiment.

The HLT is a software system implemented in a filtering farm. It can use information from the full dataset for evaluation of the event. The calculations made by the HLT can be of similar complexity as those made in the offline analysis software. In a first step, only information from the calorimeter and muon detectors is used for classification. Full track reconstruction from stored tracker information requires large amounts of CPU time. In order to reduce the needed CPU time, electron and muon candidates can be confirmed in an intermediate step using partial tracker information, such as individual pixel hits. Afterwards, full particle tracks are reconstructed. In order to further reduce the required CPU time, only tracks in the proximity of an object identified by the L1 trigger are fully reconstructed. The HLT is described in more detail in [69].

3.2.6 Luminosity system

An accurately measured luminosity is an important ingredient for many precision measurements in the CMS analysis program. Two components of the CMS detector are used to measure the instantaneous luminosity: the forward hadron calorimeter and the pixel detector [70]. The tower-occupancy method is based on information from the forward hadronic calorimeters (HF). This method exploits the linear dependence of the luminosity on the average transverse energy per calorimeter tower. Because of its pileup dependency, the HCAL response was not linear in the run conditions of the 2012 LHC operation. Because of this, making accurate measurements is difficult using this method. It is therefore only used as a cross check of the actual measurement. The latter is performed with the pixel detector.

In the pixel-cluster-counting method, the luminosity is derived from the average number of pixel clusters per event $\langle n \rangle$. The luminosity can be written as

$$L = \frac{f_{rev} \cdot \langle n \rangle}{\sigma_T \cdot n_1} \tag{3.4}$$

where f_{rev} is the orbital beam-revolution frequency of the LHC, σ_T the total inelastic cross section and n_1 the number of pixel clusters per inelastic collision. The value of the so-called visible cross section $\sigma_{vis} = \sigma_T n_1$ then needs to be calibrated in a Van der Meer scan procedure [71]. The principle of this method is to determine size and shape of the interaction region by measuring how the relative interaction rate depends on the transverse beam separation. With the proton-density profile in x and y direction $F(x, y) =$

$f_x(x)f_y(y)$, the instantaneous luminosity can be expressed as

$$\mathcal{L}_0 = \frac{N_1\ N_2\ f_{rev}\ F(0,0)}{\int f_x(\Delta x)d\Delta x \cdot \int f_y(\Delta y)d\Delta y}. \tag{3.5}$$

The measurement of these profiles was performed in November 2012 and consisted of five scans in the horizontal and vertical direction each [70]. The cumulative luminosity delivered to the CMS experiment by the LHC over time is shown in figure 3.6. A precision of 2.6% was achieved for the luminosity measurement of the 2012 dataset [70].

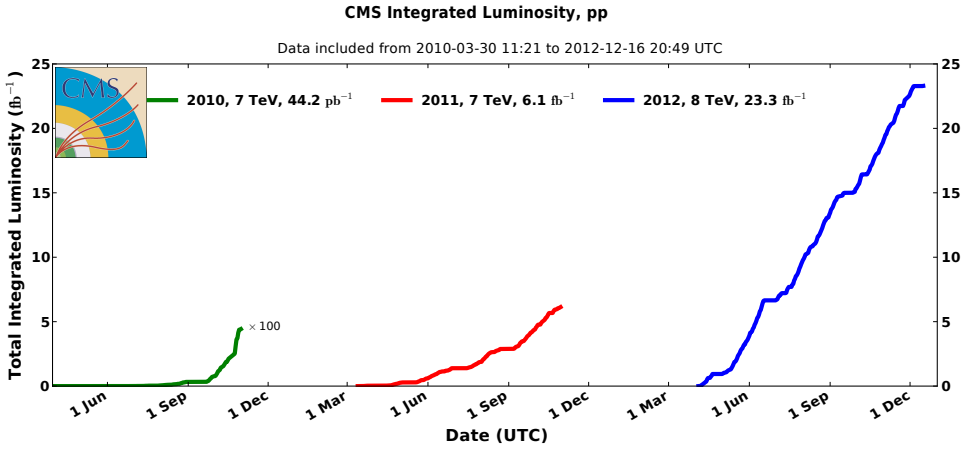

Figure 3.6: Cumulative luminosity in proton-proton collisions delivered to the CMS experiment over time starting with the first data taking in 2010 up until the beginning of the long shutdown 1 in December 2012. The blue curve depicts the cumulative luminosity for the 2012 data taking at $\sqrt{s} = 8$ TeV [72].

3.3 Future LHC operations and planned detector upgrades

The detector in the previously described layout will resume operation after the current long shutdown (LS1) in March 2015 with the start of the LHC Run 2 [73]. The center-of-mass-energy of the LHC will be increased to 13 TeV in a first step and, after approximately one year of data taking, to 14 TeV. In the summer of 2018, the second long shutdown (LS2) will take place. During this shut down, the injector chain of the accelerator will be improved to increase the intensity and lower the emittance of the proton bunches injected into the LHC ring. While LS1 was only used for maintenance of the detector, the layout of several detector components will be updated during LS2 to improve the overall detector performance and meet the demands of much brighter delivered particle bunches [74].

Some of the most important changes to the detector layout are outlined in the following: a fourth layer of CSCs in the barrel region and RPCs in the endcaps of the muon system will improve the muon-trigger performance. Furthermore, a depth segmentation will be introduced in the hadron calorimeters to better compensate for radiation damage of the scintillator material. The pixel detector will be replaced completely. With respect to the current design, the new pixel detector will have an additional layer of detectors in the barrel and endcap regions, which amounts to four barrel and three endcap pixel layers. This improvement will enable a tracking performance similar to the presently observed performance even in a more difficult environment with high tracker occupancies.

In Run 3, starting after LS2 in 2020, the LHC will again start operation at a center-of-mass-energy of 14 TeV. The goal is to record ~ 500 fb^{-1} of data under these conditions before the third long shutdown (LS3) begins in 2022. By the end of Run 3 the luminosity delivered by the LHC will exceed the accelerators design goals, reaching luminosities of about $2 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$. During LS3 not only the detectors but also the LHC machine itself will be upgraded following plans of the high luminosity LHC program (HL-LHC). After these upgrades the goal for the HL-LHC is to provide approximately 2500 fb^{-1} of data. The average number of pile-up vertices per event in this environment is expected to be about 140, which leads to a number of difficulties including high tracking fake rates and resolution losses in the calorimetry system. Therefore, extensive changes in the detector layout are being proposed to cope with these new challenges.

The most significant change to the CMS detector in the phase 2 upgrade will be the complete replacement of the tracking system. The material of the new tracking system will be radiation harder and more compact to better accommodate the space limitations within the CMS detector. Also, the granularity of the pixel detector and the outer tracking system will be improved and the acceptance of the tracking system extended to high pseudorapidities. The main goals of the phase two upgrade are to ensure the longevity of all detector systems and to improve coverage and performance in the forward region which will become increasingly important in future measurements.

4 Event simulation with Monte Carlo generators

Simulations of physics processes are frequently used in modern high-energy physics, e.g., in the development and validation of measurement techniques. Sometimes, simulated events are even used directly to model background processes in analyses, comparing distributions of physical observables in measured data to those predicted by the simulation. For these purposes, the events taking place in the detector must be modelled as accurately as possible. Monte Carlo techniques are employed in the simulation. First, probability distributions for the particles contained in the events are obtained by calculating the Feynman diagrams of the physics process that is evaluated. The Monte Carlo generators then produce hypothetical events using pseudo-random numbers following these probability distributions. This way, the quantum-mechanic nature of particle interactions is taken into account.

The process of event generation can be roughly divided into five steps in most MC generator programs [75]: it begins with the hard scattering process, which is the process with the highest momentum transfer in an event. Parton distribution functions (PDF) describe the momentum distributions of the partons of the particles brought to collision, providing probability distributions for incoming partons from lowest-order perturbation theory. The PDFs cannot be calculated in perturbation theory but are determined in global fits to experimental data from deep inelastic scattering, Drell-Yan processes, or jet production. Such fits are for example performed by the CTEQ group, who then provide the results of their fits for use in the Monte Carlo event generation [76]. With this input, the expected cross section of a process is computed. The order up to which matrix elements contribute to this computation depends on the Monte Carlo generator program used. For example, the frequently used leading-order generator MadGraph [77,78] can, in principle, take into account any number of final-state particles. In practice, the accuracy is limited by the availability of computing resources.

At the parton shower stage, bremsstrahlung is emitted and gluons emerge from accelerated particles in scattering processes, causing cascades of partons. These cascades need to be simulated in a probabilistic manner. The simulation of parton showers is an iterative process starting from the high-p_T particles in the hard process and proceeding in each step to lower momentum scales until perturbation theory finally breaks down. The phase space is thus filled up by large numbers of mostly soft gluons. At this point, hadronization models are applied. Due to quark confinement, hadrons are formed that are actually visible in detectors. As most of these hadrons are in fact unstable, their decays must be modelled in a final step.

Besides the actual process of interest, also the so-called underlying event has to be modelled. Interactions of other constituents of the two colliding hadrons lead to a contamination of the event with soft hadrons. There is more than one model available to describe the involved multi-parton interactions. This makes a correct simulation of the contribution of the underlying event rather difficult. Often, the modelling of the underlying event is included in the simulation of the parton showering. For this, certain parameters in the programs used for simulation of the showering are tuned to match previous experimental

observations. The values of these parameters can only be determined in this empirical way.

The approximations used in modeling of the parton showers are only valid in the collinear and soft limits [79]. This sort of simulation does not describe hard jets well. It therefor does not provide a very reliable modeling of events with well-separated, hard jets. The solution to this is so-called matrix-element matching. In this method, tree-level matrix elements for multijet events are combined with parton showers from other event generators, that describe the internal structure of the jets and the soft radiation more accurately.

Matrix element generators, such as MadGraph or Powheg [80–82], are used to obtain the tree-level matrix elements of an event. For a process with a given set of initial- and final-state particles the program identifies the contributing Feynman diagrams and determines the matrix elements. The generated events are then passed to other Monte Carlo programs for simulation of the showering such as Pythia [83] or Herwig [84]. The produced parton showers are then matched to the matrix elements [85]. A caveat in this sort of combination is double counting of parts of the phase space which needs to be avoided. Different methods to correct for double counting effects are implemented in both Pythia and Herwig.

The ultimate goal is to compare these generated events to measured data. In a last step, the hypothetical detector response to the generated particles is obtained. The Geant4 program [86] simulates the readout signals using a detailed model of the CMS detector. After the application of the detector simulation, the same methods are utilized for reconstruction of physics objects as described in chapter 5 and the events generated in the Monte Carlo simulation can be directly compared to measured data.

5 Reconstruction of physics objects and jet substructure

Dedicated reconstruction algorithms use the detector information provided by the CMS experiment to reconstruct the actual particle content of the recorded events. Simple detector hits are converted into physics objects such as charged particles or jets that are used in the physics analyses. Details on this reconstruction procedure are given in the first part of this chapter.

Jets are physics objects composed of many constituents which means that, in contrast to, e.g., reconstructed charged leptons, they do have a substructure. Several algorithms for the examination of jet substructure are introduced towards the end of this chapter.

5.1 Particle reconstruction using the particle-flow algorithm

The particle-flow algorithm exploits the fact, that most stable particles leave traces in more than one detector system [87]. Information from all CMS sub-detectors is taken into account to reconstruct the stable particles in the event, i.e., electrons, muons, photons, charged hadrons, and neutral hadrons. These particles are then used to determine further event properties such as the missing transverse energy or to build jets. This use of the full detector information enables very precise measurements of particle direction and energy and more accurate identification of the particle type. A good illustration of this improvement is, for example, the calorimetry system: the granularity of the HCAL is about 25 times coarser than that of the ECAL. When examining jets with relatively high transverse momenta of about 100 GeV or more, the HCAL system alone would not be able to spatially resolve charged and neutral hadrons. Combining HCAL and ECAL information improves the hadron energy resolution to about 10% at transverse momenta of 100 GeV [87]. A neutral hadron can be identified if an energy excess is observed in an HCAL cell which has no charged-hadron track pointing to it.

In a first step, so-called particle-flow elements are reconstructed using single detector systems. These particle-flow elements include charged-particle tracks from the silicon-tracker system, calorimeter clusters, as well as track segments from the muon system. A dedicated linking algorithm then matches the elements from different systems to blocks, to gather all available information for the particle identification with the particle-flow algorithm.

5.1.1 Charged-particle tracks

For the reconstruction of charged-particle tracks in the silicon tracker, an iterative tracking algorithm is used, that ensures high efficiency and a low fake rate [88, 89]. In the seed generation, track candidates are formed from two or three tracker hits. Using a Kalman filtering technique [90], further hits along the expected flight direction of the particle are assigned to the track seed. Only tracks passing certain quaility criteria are retained.

In the first iteration of track identification, very strong criteria are applied in seeding and reconstruction of tracks, resulting in a very small fake rate. The identified tracks need to consist of three pixel hits, have a $p_T > 0.8$ GeV and originate very close to the interaction point. Tracker hits that can be unambiguously assigned to the thus identified tracks are removed from the event. As these strict selection criteria compromise the reconstruction efficiency, the track seeding criteria are loosened in the second iteration in which tracks with only two tracker hits are reconstructed. In a third iteration also tracks mit low momenta are identified. Finally, tracks that do not originate close to the beam spot are searched for in three further iterations. These kind of tracks are left by secondary charged particles which can be produced in photon conversions or nuclear interactions. Already identified tracks are removed from the event, reducing the number of available tracker hits. Thus, the probability to assign tracker hits wrongly decreases with each iteration in the track reconstruction, lowering the fake rate.

5.1.2 Vertices

An accurate position measurement of the primary vertex (PV) is a crucial ingredient for a number of techniques used in the event reconstruction, including track seeding or identification of b-hadron decays [91]. The compatibility with the beam spot is used as the main selection criterion for tracks that are used as input to the reconstruction of the PV. The deterministic annealing method [92] is applied in order to identify clusters of tracks that are close to each other in z-direction at their point of closest approach to the beam-line. The primary interaction vertex is defined by the cluster with the highest value of $\sum(p_T^{track})^2$. Finally, an adaptive vertex fit [93] is performed to find the exact position of this primary vertex. A difficulty in vertex fitting is to ensure correct treatment of mis-associated tracks. In this adaptive method, weights are applied to the tracks in order to reduce the impact of outlying tracks on the final vertex coordinates. The vertex position is found in an iterative Kalman filtering procedure [90].

5.1.3 Calorimeter clusters

When reconstructing the event content, the demands for precise information from the calorimeter system are manifold: energy and direction of photons, electrons, stable charged and neutral hadrons need to be measured as accurately as possible. Neutral particles are to be distinguished from the energy deposits left by charged hadrons. A three-step calorimeter-clustering algorithm has been developed to accommodate all of these tasks. It is applied separately to entries of each sub-system of the calorimeter, i.e., ECAL barrel and endcap, HCAL barrel and endcap, as well as first and second layer of the pre-shower detector. No clustering is performed in the hadron forward sub-detector, where each single cell is counted as an individual cluster. In a first step, local maxima in the calorimeter-cell energy are identified. If their energy exceeds the given threshold, they are used as cluster seeds. Next, topological clusters are formed. Neighboring cells sharing at least one side with the cluster seed or other cells already associated to this seed are examined. If the energy deposited in these cells is large enough they are merged into the topological cluster. If a topological cluster contains more than one seed, a “particle-flow cluster” is constructed for each seed. The energy of cells in a topological cluster is shared between the multiple particle-flow clusters according to the distance between cell and each of the

particle-flow clusters. The concrete distribution of the cell energy is determined in an iterative procedure.

5.1.4 Muon tracks

The reconstruction of the muon tracks used as input for the particle-flow algorithm starts with a detector-level reconstruction of track segments. These segments are found in fits to a number of aligned muon-detector hits [94, 95]. Track segments are identified in the chambers of each muon sub-system individually. In order to form so-called stand-alone muon tracks from these segments, a Kalman filtering technique is used [90]. In this procedure, track segments from the innermost muon chambers are used as seeds. Matching segments from the next layers are linked to the track in a stepwise procedure, taking into account the stored information about position, momentum, and direction of the track segments, until the outermost chamber is reached. The energy loss in the material and the non-uniform magnetic field need to be considered in these calculations. Once all segments from the innermost to the outermost layer of the muon system have been examined, a Kalman filter is applied a second time. This second iteration of Kalman filtering begins on the outside and finishes with the innermost layer. The track parameters are defined in this step of the reconstruction. Finally, the track is extrapolated towards the interaction point to check compatibility with the interaction vertex in order to identify so-called global muons. The muon reconstruction is finalized by the particle-flow algorithm described below.

5.1.5 Particle-flow algorithm

The different particle-flow elements need to be matched if they are caused by the same particle passing through the detector. Double counting from different detector systems needs to be avoided in this process. Each pair of elements can be linked, for example by extrapolating tracks from one detector system to another. The quality of these links is rated by a distance measurement in the (η, ϕ)-plane between the extrapolated track and the element it is being linked to. The thus constructed blocks contain several elements which have been directly or indirectly linked.

Finally, the particle-flow algorithm reconstructs particles from each block of elements provided by the linking algorithm, starting with the muon identification. Global muons consist of a charged-particle track in the tracker that is linked to a muon track in the muon system. They are identified as particle-flow muons if the combined momentum measurement of both systems is compatible with that of the tracking system alone [87]. For so-called tracker muons the reconstruction starts with a charged-particle track that is extrapolated to the muon system, where it is matched to muon-track segments. If at least one matching muon-track segment can be found, the charged-particle tracker track is marked as a tracker-muon track. Less than 1% of the muons from collisions do not leave any trace in the tracking system and can be detected in the muon chambers only [96]. This third type of reconstructed muons are the previously mentioned stand-alone muons. Once the muon identification has finished, all tracks associated to the muons are removed from the block.

The next step is the electron identification: in a pre-identification stage, electron-track candidates are identified by means of their shortness and their energy loss due

to Bremsstrahlung. Then, a Gaussian-sum filter attempts to reconstruct a trajectory through the full tracking system into the ECAL. If the linked calorimeter clusters and tracks fulfill the requirements, they are assigned to the particle-flow electron and removed from the block. At each intersection point of the track with a tracker layer, tangents from possible electron tracks are extrapolated to the ECAL to account for the energy loss due to Bremsstrahlung. If one of the tangents matches an ECAL cluster, the latter is assumed to arise from a Bremsstrahlung photon and is linked to the particle track.

The remaining tracks are linked to the ECAL and HCAL clusters closest to them in the (η, ϕ)-plane. For 0.2% of these tracks, the uncertainty of the transverse-momentum measurement is smaller than the relative expected energy resolution for charged hadrons. In this case, the track is discarded. 90% of these rejected tracks are fake tracks, the energy of the non-fake 10% of these tracks can instead be obtained from the independent measurements in the calorimetry system. The momentum of tracks passing this selection is compared to the energy of the linked calorimeter clusters to identify neutral particles in the block. If more than one track is linked to a calorimeter cluster, the sum of the track momenta is used in the comparison. If the cluster linked to the track has an energy smaller than the charged-particle momentum, neighboring ECAL clusters are linked to the track in addition. Each of the tracks is associated to a particle-flow charged hadron whose momentum is taken from the tracker measurement. If the track momentum is compatible with the energy of the calorimeter clusters the track is linked to, the hadron momentum is re-calculated also taking into account the energy measurement in the calorimetry system. In case the energy of the closest calorimeter clusters exceeds the momentum of the linked track by more than the expected calorimeter resolution, a particle-flow photon is reconstructed from the ECAL energy deposits. If the energy excess is even larger than the total energy deposited in the ECAL clusters in question, a particle-flow neutral hadron is reconstructed from the HCAL clusters in the block. All calorimeter clusters that are not linked to particle tracks give rise to particle-flow photons or particle-flow neutral hadrons.

Finally, further physics objects can be constructed using the particle-flow particles as building blocks. One example are jets, which are formed from these building blocks using jet clustering algorithms. Jets are of great importance to this analysis and will be described in detail in section 5.2. Another value determined using the particle-flow particles as input is the so-called missing transverse energy $E_{T,miss}$. Neutrinos and some hypothetical weakly interacting particles predicted in models for physics beyond the standard model leave the CMS detector without interacting with its material. Therefore, the only obtainable trace of such a particle having been produced is an imbalance in the transverse momentum of all particles reconstructed in an event. The sum of the transverse momenta of all particle-flow particles in an event gives the negative of the missing transverse energy of this event. The measurement of the missing transverse energy is very delicate as it heavily depends on correct measurements of all particles in the event. Details on the commissioning of the missing transverse energy can be found in [97].

5.2 Jets

While charged muons and electrons leave clear individual tracks in the detector, gluons and quarks, which are subject to the strong interaction, hadronize directly upon production in the hard interaction. Therefore, entire particle showers leave their traces in the detector instead of individual particles. For CMS analyses, a particle shower is reconstructed as a single object though: jets are formed in the event reconstruction. Special clustering algorithms associate the different shower components to the jets. Typically the energy of a jet originates to 65% from charged particles, 25% from photons and only to 10% from neutral hadrons. Instead of relying fully on stand alone measurements of the particle showers in the electromagnetic calorimeter and the hadronic calorimeter, the particle-flow algorithm also uses information from other detector systems for the jet reconstruction. The electromagnetic calorimeter and especially the tracking system clearly outperform the hadronic calorimeter with respect to energy resolution. Only the 10% fraction of neutral hadrons cannot be matched to charged particle tracks. For the remaining 90% of the jet constituents, supplementary information is obtained from the tracking system. For these particles, the hadronic and electromagnetic calorimeters can both be used for the energy measurements. Therefore, use of the particle-flow algorithm improves the reconstruction of the jet energy significantly [87].

5.2.1 Jet-clustering algorithms

The most basic approach to the definition of a jet is implemented in fixed cone algorithms. In these algorithms, one simply uses a cone of fixed size around a seed particle to define a jet. In iterative cone algorithms, a jet is seeded by a hard cluster and close particles are clustered into the jet until a stable cone is found [94]. These simple algorithms are not collinear safe though which can cause divergences in higher order perturbative calculations. In collinear safe algorithms, splitting of a single jet constituent into two particles with identical direction does not affect the jet properties. Another desirable quality of jet algorithms is so-called infrared safety, i.e., robustness against emission of soft radiation. An infrared- and collinear-safe alternative to the simple cone algorithms are the k_T-like algorithms, which are most widely used in CMS [98]. They cluster the candidate jet constituents in a sequential, iterative procedure. The jets are constructed hierarchically. For each pair of four-vector inputs, two parameters are calculated: the pairwise distance parameter

$$d_{ij} = min(k_{T,i}^n, k_{T,j}^n)\frac{\Delta R_{ij}^2}{R^2} \tag{5.1}$$

and the beam-distance parameter

$$d_{iB} = k_{T,i}^n. \tag{5.2}$$

In these equations, the transverse momentum of the i-th particle with respect to the beam axis is denoted as $k_{T,i}$, the distance between the two particles i and j in rapidity y and azimuthal angle ϕ with ΔR_{ij}. R is is the radius of the jet. Different jet-clustering algorithms are defined by the choice of the parameter n in equations 5.1 and 5.2: in the k_T algorithm [99] n is set to $n = 2$, for the anti-k_T algorithm [100], which is most widely used in CMS, $n = -2$. In the Cambridge/Aachen algorithm [101] $n = 0$ is used, resulting also in a constant value for the beam distance of $d_{iB} = 1$.

In the jet clustering process, the minimum of all pairwise distance parameters d_{ij} and the beam distances d_{iB} is determined. If the minimum is one of the d_{ij}, the four-vectors of the two particles i, j are added and considered as a single new particle in the following. If the minimum is one of the d_{iB}, the particle i is classified as a final jet and removed from the list prior to the next iteration. The jet clustering is continued until all input particles have been associated to jets.

The choice of the parameter n and the associated jet algorithm influences the topology of the clustered jets. While the anti-k_T algorithm preferably merges constituents with high transverse momenta, the Cambridge/Aachen algorithm does not employ a momentum-based weighting at all. The jet constituents are grouped according to their spatial relations only. This makes Cambridge/Aachen jets more suitable for examinations of the jet substructure as described in section 5.2.5. The k_T algorithm is mainly used to reconstruct low momentum jets and is very sensitive to low-p_T pileup contributions. The treatment of pileup in CMS analyses is described below.

While the k_T and Cambridge/Aachen algorithms produce jets with irregular shapes, the anti-k_T algorithm gives results that are more similar to those of idealized cone algorithms [102].

5.2.2 Charged-hadron subtraction

Pileup is the contamination of an event caused by interactions taking place in the detector in addition to the event of interest. In the high luminosity environment at the LHC, the correct treatment of pileup events is a challenge in the event reconstruction in general, but especially when clustering jets from the particle-flow particles [103]. With the increasing interest in analysis of the jet substructure, correct description of this internal structure of jets becomes even more important. Several tools are available to mitigate the effects of pileup on the jet properties in CMS, *e. g.* , so-called charged-hadron subtraction.

In the charged-hadron subtraction (CHS) method, any charged hadron reconstructed by the particle-flow algorithm is discarded prior to the jet clustering, if it can be unambiguously matched to one of the pileup vertices in the event. The vertices in the event are ordered by the magnitude of the sum of squared transverse momenta of tracks $\sum |p_T^{track}|^2$. The vertex leading in this quantity is identified as the primary vertex of the event. All other vertices are treated as pileup vertices. The charged-hadron tracks are matched to the vertices using a χ^2-per-degree-of-freedom criterion. Only tracks that cannot be matched to one of the pileup vertices with a $\chi^2/\mathrm{D.o.F.} < 20$ are considered in the jet clustering.

5.2.3 Jet energy corrections and resolution

Jet energy corrections are applied to data and simulated events. This is necessary to translate the jet energy measured in the detector to the energy of the true partons initiating the jet. The response of the calorimeter system to the energy of the measured particles is not linear. Therefore, a number of corrections have to be applied to obtain a response that is independent of η and p_T. In CMS, a factorized approach is used for the correction of the jet energy in which a number of individual corrections are applied sequentially [104]. The order in which these corrections are applied cannot be changed.

The steps of the jet energy correction sequence are:

1. Offset correction: energy coming from interactions of other partons within the colliding protons, i.e., the underlying event, and pileup is removed. Also, extra energy contributions due to electronic noise are discarded in this step. For this subtraction, the jet-area method is employed [105]. The corresponding correction factor is based on the median of the distribution of the jet transverse momentum per jet area $p_T, i/A_i$, where i runs over all jets in an event. This quantity is obtained from jets clustered with the k_T algorithm, which reconstructs a large number of very soft jets per event. The corrections applied to the jets after this step are then independent of the instantaneous luminosity and, therefore, the pileup.

2. Relative jet energy scale correction: this correction ensures uniformity in the jet response vs. η by correcting the response of jets in the full η range to that of a jet in the central $|\eta| < 1.3$ region.

3. Absolute jet energy scale correction: in this step, the transverse momenta of the reconstructed jets are corrected in such a way, that they are, on average, equal to those of the jets clustered directly from the generated particles in the Monte Carlo simulation.

The correction factors derived for data recorded by the CMS detector at $\sqrt{s} = 8$ TeV for jet clustered with the anti-k_T algorithm using a distance parameter $R = 0.5$ (AK5 jets) can be found in figure 5.1. On the left-hand side, the correction factors for the offset correction are shown for data and simulated events. Four different pileup scenarios are considered, each corresponding to a different number of reconstructed primary vertices N_{PV}. The magnitude of this correction factor increases with the number of reconstructed vertices. The relative jet energy correction factors are displayed on the right-hand side of figure 5.1. These corrections are derived from Monte Carlo simulation. Larger corrections are needed for jets with lower transverse momenta.

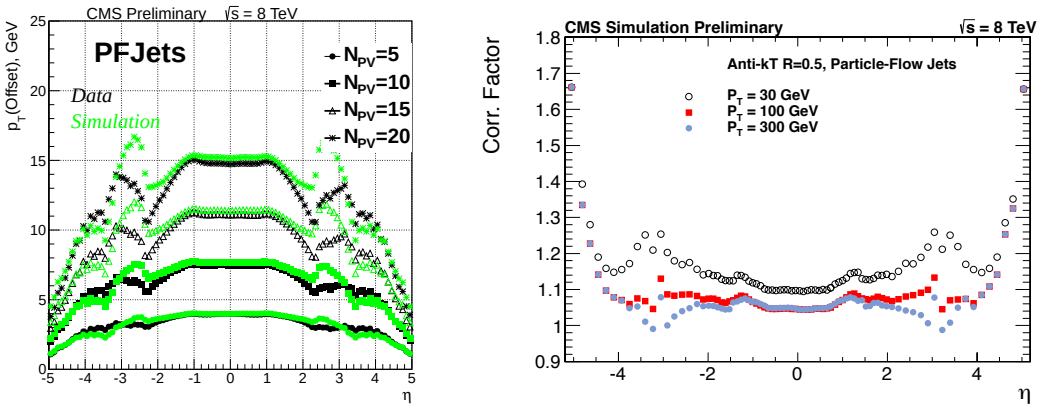

Figure 5.1: Left: magnitude of the p_T-offset correction in bins of η for different pileup scenarios. The number of reconstructed primary vertices is denoted as N_PV. The correction factor is derived separately for data and simulated events. Right: relative jet energy corrections derived from Monte Carlo simulation [106].

After these corrections have been used, additional residual corrections need to be applied to data only despite the overall successful modeling of the jet energy response. These corrections differ, depending on the jet η. Overall, they do not exceed 10% for any value of η. The absolute uncertainty on the jet energy scale is p_T dependent but overall smaller than 4% for any jet with $p_T > 40$ GeV. While the other jet energy correction factors are obtained from Monte Carlo simulation, the residual corrections are measured in p_T-balanced dijet events in data. This method was developed for the UA2 experiment at the CERN SP$\bar{\text{P}}$S and the Tevatron experiments CDF and D0 [107–109].

Since jets are part of the input to the determination of several event properties, e.g., the missing transverse energy, the effect of the jet energy corrections is also propagated to these quantities.

Not only the jet energy response, but also the resolution can be different in data and Monte Carlo simulation. The jet energy resolution of AK5 jets is also measured in dijet events, and in a second method using events with photons [110]. Scale factors are then computed for a number of pseudorapidity ranges. They are applied to the jets in simulated events. In data, the measured jet p_T resolution for particle-flow jets with $p_T > 100$ GeV and $|\eta| < 0.5$ is typically better than 10%. In simulated events the core of the distribution of the jet p_T resolution is broader than observed in data by 10-20% [104].

5.2.4 Identification of jets with bottom-quark content using b-tagging algorithms

Identification of jets containing bottom quarks is an important ingredient to many physics analyses performed with the CMS detector. Many particles under study at the LHC are very likely to decay into bottom quarks. For example, in decays of the top quark, the branching fraction Br(t$\to$bW) amounts to almost 100% and the recently discovered Higgs boson has a branching fraction Br(H$\to$b$\bar{\text{b}}$) of approximately 58% [37].

The bottom quark has a rather large mass compared to the light up, down, and strange quarks. Its fragmentation function also differs from those of the lighter quarks. The momentum spectrum of bottom-quark-decay products tends to be rather hard [111]. Also, b hadrons are relatively long lived, leading to secondary vertices displaced from the primary vertex of the event. In consequence, when physics processes involving bottom quarks occur in the detector, the hadronization of the bottom quark gives rise to jets with special properties. These characteristics can be exploited to distinguish these so-called b jets from regular light jets originating from u,d, or s quarks or gluons. Charm-quark hadronization is another background for b-jet identification. The background from charm quarks is more difficult to reduce because c jets can exhibit similar properties as b jets.

Several b-tagging algorithms are available in CMS software to identify b jets. The more simple and robust algorithms use a single observable for discriminating between b and light jets, but also more complicated multivariate algorithms have been developed to increase the discrimination power. The b-tagging algorithms in CMS are usually applied to AK5 jets clustered from particle-flow particles [100], but they can also be applied to jets clustered from other objects or subjets of large radius jets as described in section 5.2.5.2.

Only well-reconstructed tracks are used as input for the b-tagging algorithms. Each track has to fulfill certain requirements [14]: there have to be at least eight tracker hits associated with the track, two of these from the pixel detector. The transverse momentum of the track must be larger than 1 GeV. Furthermore, the point of closest approach between

jet axis and track must be at maximum 5 cm displaced from the primary vertex. At the same time, the minimum distance between the track and primary vertex in the transverse plane (along the beam axis) must not be larger than 0.2 cm (17 cm). The angular distance between tracks and the jet axis is limited to $\Delta R < 0.5$. Finally, the fit of the tracker hits has to have a value of χ^2 per degrees of freedom smaller than 5.

A clean set of tracks is needed as input for the b-tagging algorithms. The distance between selected tracks and the jet axis at their point of closest approach $P_{T/A}$ is required to be smaller than 700 μm. At the same time, the point $P_{T/A}$ must be located within 5 cm of the primary vertex. These two criteria efficiently reject tracks from pileup [111].

The information used as input for the b-tagging algorithms can be split into two categories: properties of the charged-particle tracks contained in the jets on one hand and, on the other hand, attributes of the decay vertices of the b hadrons. These secondary vertices are usually displaced from the primary vertex of the hard interaction due to the larger lifetime of the b hadrons that allows them to travel some distance within the detector before decaying.

b tagging using particle track information

The track-impact parameter (IP) is a powerful handle to distinguish tracks originating from b-hadron decays from tracks originating from the primary vertex, i.e., the prompt tracks. It is defined as the minimal distance between the track and the primary vertex as illustrated in figure 5.2.

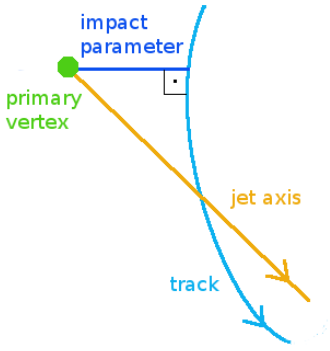

Figure 5.2: Illustration of the track IP which is defined as the minimal distance between a track and the primary vertex.

A three dimensional IP is calculated. The sign of the IP is defined as the scalar product of the jet direction and a vector pointing from the primary vertex to the point of closest approach between primary vertex and track. By calculating the IP significance S_{IP}, which is the ratio of the IP and its estimated uncertainty, the resolution of the IP is taken into account. This reduces a possible bias in the discrimination introduced by tracks with very large IP but only very poor resolution.

The track-counting algorithms in CMS are based entirely on the IP significance. All tracks contained in a jet are ranked by values of the IP significance. The track-counting-high-efficiency (TCHE) algorithm then uses the IP significance of the second ranking track

as discriminator, while the track-counting-high-purity (TCHP) algorithm uses the one of the third ranking track. The jet-probability algorithm combines the IP information of all tracks into a single discriminator using a likelihood method.

Exemplary distributions of the track IP and the discriminator obtained from the TCHP algorithm are shown in figure 5.3. An inclusive multijet sample containing at least one jet with a transverse momentum of 60-500 GeV was used to obtain these distributions. The flavor composition of the samples is obtained from the Monte Carlo simulation. The shape of the distribution in data is described well by the simulation. The jets initiated by bottom quarks tend to have larger values of IP and, in consequence, also of the TCHP discrimator, while the majority of light jets are found to have rather low values of the same variables.

The TCHP and jet-probability algorithms are the only b-tagging algorithms purely based on track displacement that are used in analyses of the 8 TeV dataset.

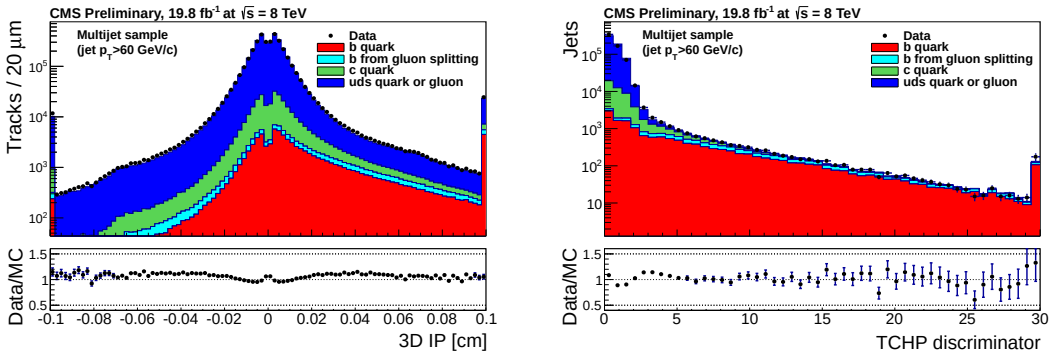

Figure 5.3: Distributions of the track IP(left) and the discriminator of the TCHP algorithm (right) [14].

b tagging using secondary-vertex information

A reconstructed secondary vertex within a jet is a strong indication for the jet to originate from bottom-quark hadronization. The secondary-vertex candidates used in the b-tagging algorithms in CMS must be clearly distinguishable from the primary vertex: their radial distance must have a significance exceeding 3 σ. Also, no more than 65% of the tracks associated with the candidate may be shared with the primary vertex. The flight direction of the secondary vertex has to be contained in a cone of $\Delta R < 0.5$ around the jet direction. In order to reject background arising from interactions of particles with detector material or decays of long-lived mesons, the secondary-vertex candidates must not be displaced further than 2.5 cm from the primary vertex and their masses should not be compatible with the K^0 mass. The multiplicity of vertices fulfilling these requirements is shown on the left-hand side of figure 5.4. The majority of background events from light-quark and gluon jets does not present a secondary vertex, while most of the jets originating from bottom quarks contain at least one secondary vertex.

In the simple-secondary-vertex algorithm, the flight distance significance is used to discriminate between b jets and light jets.

5.2.4.1 The combined secondary vertex algorithm

The efficiency of secondary-vertex reconstruction consequently limits the efficiencies of purely secondary-vertex based b-tagging algorithms. Their efficiencies do not exceed approximately 65%. To avoid this limitation, the combined secondary vertex (CSV) algorithm is based on a multivariate approach using secondary-vertex information as well as track-based variables as input. This means, that b-jet identification is also possible when no secondary vertex can be reconstructed.

In the CSV algorithm, likelihood ratios are built from a number of track-based and secondary-vertex variables, that are then used as discriminating variables. In case the secondary-vertex reconstruction fails, the algorithm attempts to identify a "pseudo vertex" from tracks with a large IP significance $S_{IP} > 2$. For jets containing real or pseudo secondary vertices, the algorithm considers several properties of the vertex in the likelihood ratio, including the flight-distance significance, vertex mass and number of tracks at the vertex. If neither a real nor a pseudo vertex can be reconstructed, the algorithm relies fully on the track-based variables. These include, e.g., the IP significances of each track and the track multiplicity.

The resulting discriminating variable is shown on the right-hand side of figure 5.4. For jets originating from bottom quarks, the discriminating variable assumes large values, while light-quark and gluon jets are found to result in low values. Varying the threshold of the discriminating variable above which a jet is marked as originating from a bottom quark or b hadron, three operating or working points are defined for the CSV algorithm: the loose working point corresponds to a misidentification rate of 10%, the medium working point to 1%, and the tight one to only 0.1% misidentification rate in events with typical $t\bar{t}$ kinematics.

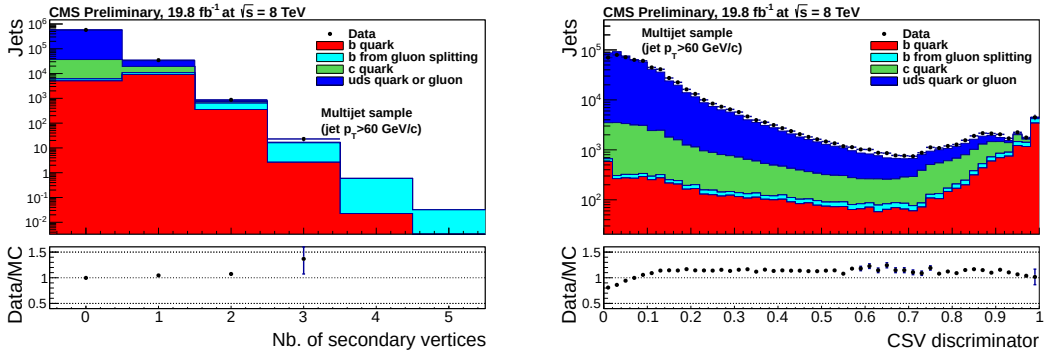

Figure 5.4: Distributions of the secondary-vertex multiplicity (left) and the discriminator of the CSV algorithm (right) [14].

5.2.4.2 Performance of the CSV algorithm in data and simulation

In general, the data are described well by the simulation in all distributions used in the b-tagging algorithms. Even in the tail regions of the input distributions, the discrepancies do not exceed 20% [14]. Scale factors are derived to compensate the observed small differences

in b-tagging performance between data and simulation. The b-tagging efficiency ϵ_b and the probability to misidentify a light jet, ϵ_{misID}, need to be measured for this purpose.

In order to determine ϵ_{misID}, negative taggers are used. The negative taggers do not differ from the actual b-tagging algorithms, except for the fact that they use only tracks with negative IPs and secondary vertices with negative decay lengths as input. The sign of the decay lengths is determined by the position of the secondary vertex along the jet direction with respect to the primary vertex. For light jets the discriminator distributions of the actual b-tagging algorithms and negative taggers are expected to be the same except for a difference in sign, while real b jets only rarely pass the selection of the negative taggers. In figure 5.5 the negative and the original discriminator distributions are shown for the TCHP, the jet-probability and the CSV algorithms in events triggered by a single high-p_T jet. The misidentification probability in data, ϵ^{data}_{misID}, can be derived from the rate of events tagged by the negative tagger ϵ_-. After correcting the measured ϵ^{data}_{misID} for the slight heavy flavor contamination in the negative tag rate and some higher order asymmetries of the negative and positive tag rates, the scale factor $SF_{light} = \epsilon^{data}_{misID}/\epsilon^{sim.}_{misID}$ can be applied in simulated events to correct for the observed differences with respect to data in misidentification rate of light jets.

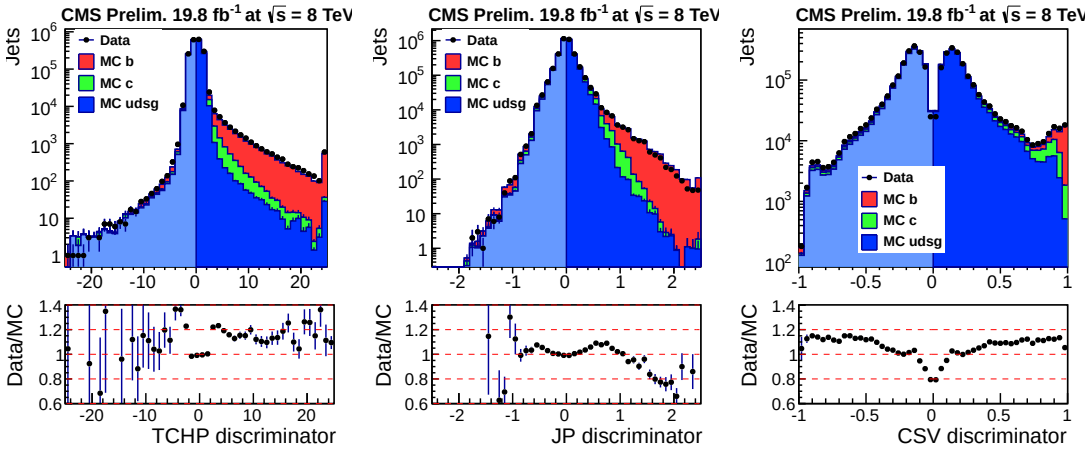

Figure 5.5: Comparison of discriminator variables in data and simulation for events passing a trigger with a jet-p_T threshold of 40 GeV. The discriminator of the actual b-tagging algorithm is shown on the positive x-axis range, that of the negative tagger on the negative range [14]. The light jet contribution is shown in two shades of blue, where the lighter shade corresponds to the negatively tagged jets. The simulation is normalized to the number of events measured in data.

In the measurement of the b-tagging efficiency, a number of methods are employed, each yielding especially high precision in a different kinematic region [14, 111]. Some of the methods use multijet events as input, in which a muon is found within a $\Delta R < 0.4$ cone around the axis of a jet. This jet is referred to as a "muon jet". When also including cascade decays of b $\to$ c $\to \ell$, the semileptonic branching fraction of b hadrons is approx-

imately 20%. Therefore, there is a high probability that a muon jet originates from a b quark rather than from a quark of another flavor. Since the CMS detector is very well suited for identification of muons, a muon-jet-enriched sample can easily be obtained.

There are several methods using variables that discriminate between b jets and light jets for determination of the b-tagging efficiency from muon-jet-enriched samples. These variables include the muon transverse momentum relative to the jet axis p_T^{rel}, the 3D IP of the muon track and the likelihood discriminator LT used in the jet-probability b-tagging algorithm. The shapes of a b-jet and a light-jet template are obtained from simulation and fitted to the distribution of these discriminating variables in data. First, the fit is performed in the full data sample and, in a second step, in a subset of events in which the muon jet passes the b-tagging requirement.

Complimentary, the b-tagging efficiency is also measured in $\mathrm{t\bar{t}}$ events which are naturally enriched in b jets. In events with two leptonic top decays, the flavor tag matching (FTM) method is used for the measurement, while in semileptonic $\mathrm{t\bar{t}}$ events the flavor tag consistency (FTC) is employed. The FTC and FTM methods are likelihood methods based on the distributions of the multiplicity of b tags in data and simulation. The actual b-tagging efficiency is obtained in log-likelihood fits in which the efficiency and $\mathrm{t\bar{t}}$ cross sections are free parameters.

Further methods for the determination of the b-tagging efficiency are explained in [14]. A weighted mean of all the measurements of the b-tagging efficiency scale factors is finally calculated in the different jet-p_T and jet-η bins, taking into account systematic uncertainties and overlap in the events used in the different measurements. The results of this combination for the CSV algorithm at medium working point are shown in figure 5.6. The results obtained in the different measurements are all consistent with each other within the statistical and systematic uncertainties. These scale factors can then be applied to simulated events in the analyses.

5.2.5 Jet substructure

The phase space for low-mass particles has been thoroughly explored in many analyses using the first LHC data but also in previous experiments. No evidence for any particles predicted in extensions of the standard model has been found in the low-mass range. Therefore, interest in searches for more massive particles is increasing. When progressing to higher mass ranges, new challenges arise in the analyses. The decay products of high-mass resonances can be highly energetic corresponding to large Lorentz boosts of their decay products. Many of the new massive particles are expected to have large branching fractions for decays to top quarks, due to the large top-quark mass of approximately 173 GeV. Top quarks are not detected directly but via their decay products. The top quark decays almost exclusively into a bottom quark and a W boson. The latter consequently decays either leptonically ($W \to \ell + \nu_\ell$) or hadronically ($W \to q + \bar{q}'$). The decay products of a top quark with a Lorentz boost of $\gamma = \frac{E}{m}$ are produced with a distance of approximately $\Delta R = \frac{2}{\gamma}$ [102]. In case of leptonic top decays, this means that the lepton, which is usually well-isolated and a good handle to identify these sort of decays, may overlap with the b jet. For hadronic decays, in which the top quark needs to be reconstructed from the two jets produced in the W-boson decay and the additional b jet, the identification is complicated to an even higher degree by the large Lorentz boost: in many cases it is not possible to reconstruct the three jets initiated by the decay products of the top quark

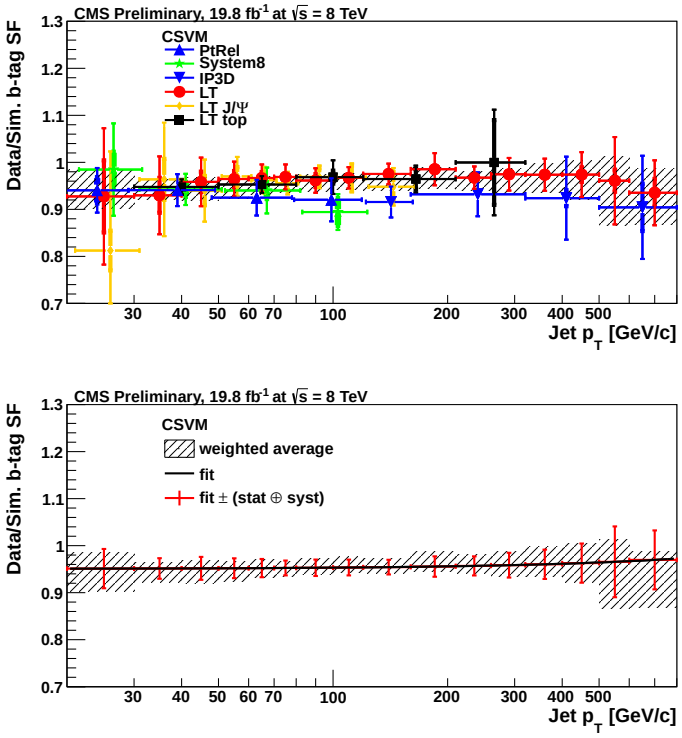

Figure 5.6: Measurements of the data/MC scale factors for the b-tagging efficiency with the CSV algorithm at medium working point [14]. The individual measurements from different methods applied to muon events, as well as the measurement from the so-called LTtop method in dileptonic $t\bar{t}$ events are shown on top. Statistical and sum of statistical and systematic uncertainties are given by the inner and outer uncertainty bars for each measurement. The hatched area shows the combined measurement. The bottom plot shows a parameterization of the combined measurement following $SF_b(p_T) = \alpha(1+\beta p_T)/(1+\gamma p_T)$. The uncertainty bars correspond to the uncertainties of the combination in each bin.

individually. This effect is illustrated on the left-hand side of figure 5.7. In these cases a different approach for the top-quark identification is used: analysis of jet substructure.

When using jet-substructure techniques, so-called "fat jets" with rather large cone radii of $R = 0.8 - 1.5$ are used as input to the algorithms for substructure analysis. In the boosted regime, all decay products of a top quark can be clustered into a single fat jet as illustrated on the right-hand side of figure 5.7. The substructure of these large jets is then examined. For this purpose, subjets are reconstructed. Fat jets clustered by the Cambridge/Aachen algorithm (CA jets) are most suitable for reconstruction of subjets. This algorithm clusters purely according to spatial properties of the jet constituents, not giving any preference, e.g., to high p_T constituents. Therefore, spatially separated subjets can be found more easily in the substructure of CA jets. Top-tagging algorithms then make kinematic selections based on properties of the subjets to identify jets that contain

decaying top quarks, the top jets. Details on such algorithms are given in sections 5.2.5.1 and 5.2.5.3.

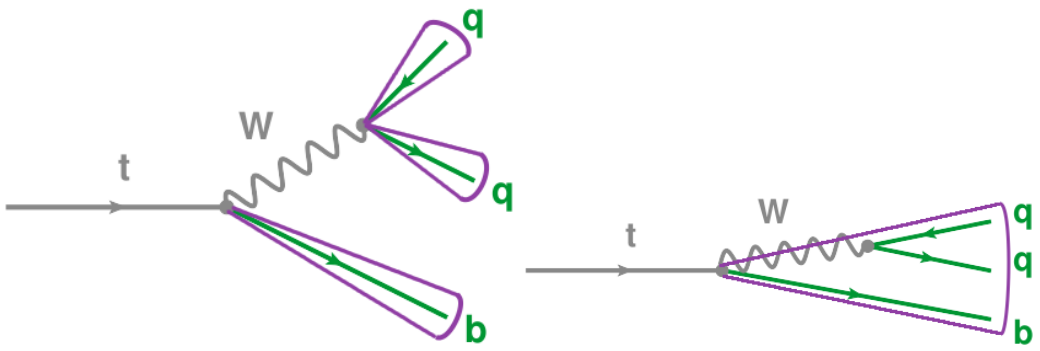

Figure 5.7: Reconstruction of hadronic top-quark decays in the resolvable (left) and boosted (right) regime.

Not only top-quark decays may be merged into fat jets in the boosted regime. The daughter particles of highly boosted W and Higgs bosons exhibit the same behavior. Here, two subjets are merged into a fat jet as illustrated in figure 5.8 for the example of a decaying Higgs boson. More details on identification of Higgs-boson decays in jet substructure is given in section 5.2.5.2, methods for W tagging are described in 5.2.5.3.

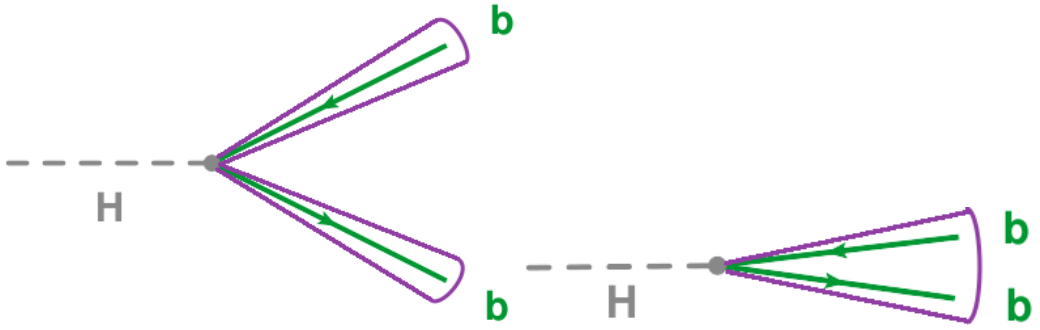

Figure 5.8: Reconstruction of hadronic Higgs-boson decays in the resolvable (left) and boosted (right) regime.

5.2.5.1 The HEPTopTagger algorithm

The HEPTopTagger algorithm is applied to Cambridge/Aachen jets with a cone radius of $R = 1.5$ [13, 102]. These jets will be referred to as CA15 jets in the following. The substructure of the CA15 jets is found by reversing the clustering history in a stepwise manner as illustrated in figure 5.9.

In the declustering procedure, the CA15 jet itself is first split into the two clusters that were merged to form the final jet. If one of the thus obtained parent clusters has a mass smaller than 30 GeV, it is saved as a subjet of the CA15 jet. Else, the splitting procedure is repeated using this cluster as input. A mass drop criterion is applied to the results of

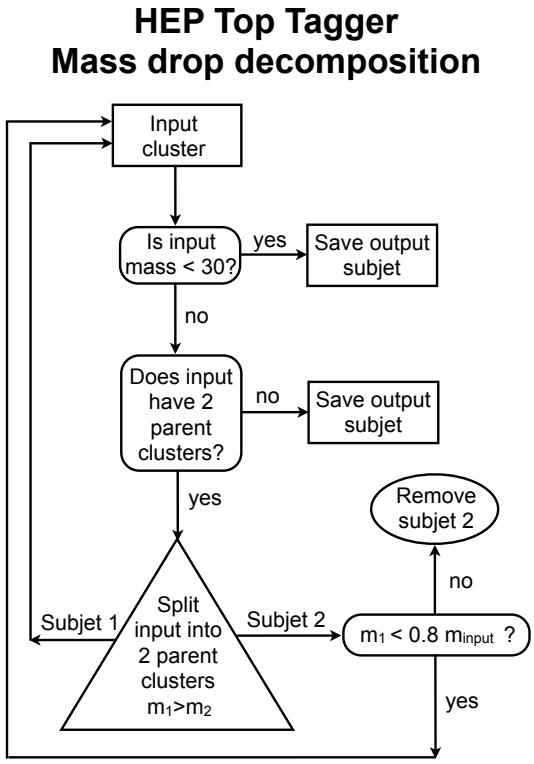

Figure 5.9: Flow chart for the mass-drop decomposition in the HEPTopTagger algorithm [112].

each splitting step: if one of the two parent clusters obtained in the splitting accounts for less than 20% of the mass of the daughter cluster prior to splitting, it is discarded. This iterative declustering is stopped only when there are no subclusters left to be split further. Subclusters are saved as subjets in two cases: either their masses are lower than 30 GeV, or they cannot be split into two parent clusters anymore because the starting point of the clustering history has been reached. This declustering procedure can result in any number of subjets, but only CA15 jets with three or more subjets are considered by the HEPTopTagger algorithm. The algorithm always fails for jets with two or less subjets.

In the next step, the filtering algorithm [113], a jet-grooming procedure, is applied to every possible combination of three subjets. In this procedure, the constituents of the three subjets are reclustered using a variable distance parameter

$$R_{filt} = min(0.3, \Delta R_{ij}/2), \tag{5.3}$$

where ΔR_{ij} is the radial distance between the two closest subjets i and j. Out of the resulting subjets of this reclustering procedure, only the five with the highest transverse momentum, the filtered subjets, are considered further. This filtering procedure is pictured in figure 5.10.

The filtered mass is then given by the invariant mass of all five filtered subjets. Out of

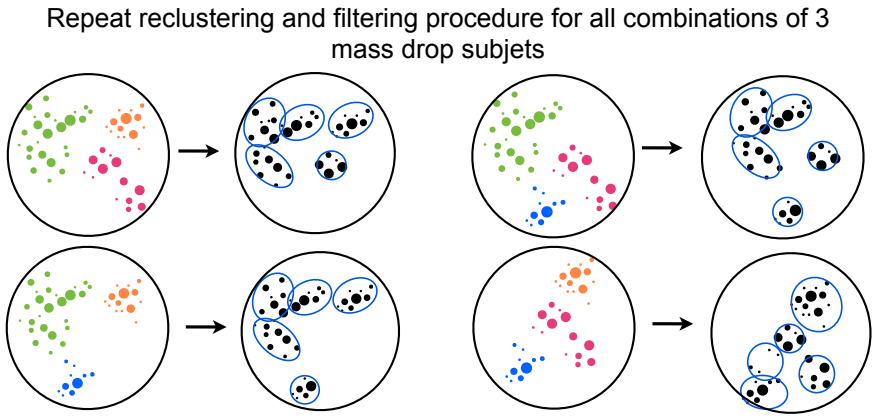

Figure 5.10: Illustration of the filtering of all combinations of three subjets after the mass-drop decomposition [112].

all combinations of three pre-filtering subjets, the one leading to the filtered mass closest to the actual top-quark mass is considered. Only the five filtered subjets are retained and all other constituents of the original CA15 jet are discarded.

Finally, the constituents of the five filtered subjets are clustered once more by a modified Cambridge/Aachen algorithm which forces the number of subjets to be exactly three as illustrated in figure 5.11.

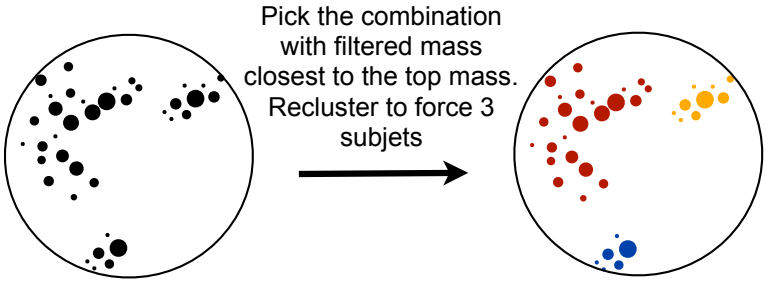

Figure 5.11: Selection of the combination of three pre-filtering subjets with the filtered mass closest to the actual top-quark mass. The five retained filtered subjets are reclustered to force exactly three subjets [112].

The kinematic selection of the HEPTopTagger algorithm is then applied to these filtered CA15 jets. It classifies jets as top tagged or non-top tagged on the basis of a number of criteria applied to the CA15 jet itself and its three subjets. The subjets are ordered according to their p_T. Only CA15 jets with a transverse momentum larger than 200 GeV are considered. Then, four mass quantities are calculated: the top-jet mass m_{123} is the invariant mass of all three filtered subjets, m_{12}, m_{13}, and m_{23} are the pairwise invariant masses of each combination of two subjets. A top-mass window is applied by requiring 140 GeV$< m_{123} <$ 250 GeV. Finally, one of these three selection conditions related to the

W-boson mass has to be fulfilled:

1. $0.2 < \arctan\left(\frac{m_{13}}{m_{12}}\right) < 1.3$ and $R_{min} < \frac{m_{23}}{m_{123}} < R_{max}$
2. $R^2_{min} \cdot \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R^2_{max} \cdot \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right)$ and $\frac{m_{23}}{m_{123}} > 0.35$
3. $R^2_{min} \cdot \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R^2_{max} \cdot \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right)$ and $\frac{m_{23}}{m_{123}} > 0.35$,

with $R_{min} = (1 - f_W) \times \frac{m_W}{m_t}$ and $R_{max} = (1 + f_W) \times \frac{m_W}{m_t}$. f_W is the width of a window around the W-boson mass and set to $f_W = 0.15$.

The effect of these three conditions can be more easily understood with the help of the two Dalitz plots in figure 5.12. In case an event passes one of the three conditions, it can be found in the A-shaped region bordered by the black lines in the two plots. The left-hand part of this figure shows the distribution for a sample of jets actually containing top-quark decays obtained from simulated $t\bar{t}$ events. Here, a large part of the jets populate the A-shaped region. The accumulation of entries in the bottom left corner of the plot is caused by events in which the decay products of the top quark are only partly contained within the CA15 jet. These jets do not pass the top-tagging criteria. The plot on the right-hand side of figure 5.12 shows the same distribution for a sample of light-flavor jets obtained from background events. Here, only a small minority of events pass the selection criteria of the HEPTopTagger algorithm.

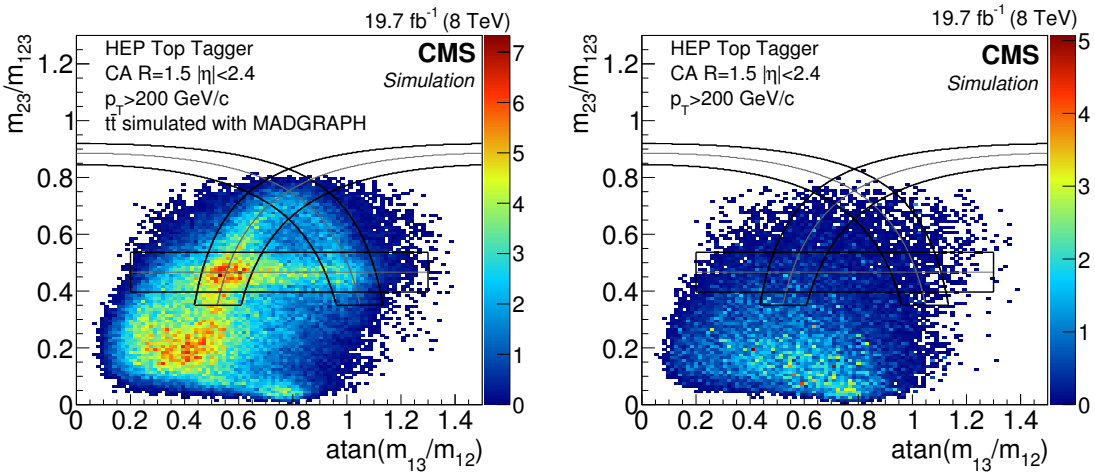

Figure 5.12: "Dalitz plots" illustrating the selection criteria of the HEPTopTagger algorithm in $t\bar{t}$ events (left) and background events (right) [102].

5.2.5.2 Identification of bottom quarks in jet substructure

The b-tagging algorithms described in section 5.2.4 are originally constructed for application to AK5 jets. In the boosted regime, also jets containing bottom quarks can present as subjets of larger radius jets. Thus, identification of subjets initiated by bottom quarks can provide valuable information about the origin of a fat jet.

Identification of bottom quarks in substructure of top-tagged jets

Identification of the bottom quark produced in the decay of a top quark with large Lorentz boost within the jet substructure can improve the performance of top-tagging algorithms [14].

The following studies of b-tagging in top-tagged jets are performed using samples of simulated T'T' $\rightarrow$ tHtH events, where the mass of the T' quarks is 1 TeV. The HEP-TopTagger algorithm is used for top tagging of CA15 jets, their subjets are obtained as described in section 5.2.5.1. There are two possibilities for enhancing the top-tagging performance using b tags:

1. B tagging the CA15 jet itself
2. B tagging one or more of its subjets.

Studies of both approaches are summarized below. The CSV-b-tagging algorithm described in section 5.2.4.1 is used for b-jet identification in all studies presented in this section.

Per default, the input to the CSV algorithm consists of all tracks within a distance of $\Delta R < 0.3$, because it was developed for the application to AK5 jets. When applied directly to the much larger CA15 jets, all tracks within a distance of $\Delta R < 1.5$ are considered for b tagging though. For tagging of the subjets of fat jets, the original parameter of $\Delta R < 0.3$ is maintained.

In figure 5.13, the misidentification rate measured in a simulated QCD-multijet sample is plotted against the b-tagging efficiency derived from the T'-quark sample at different working points of the CSV-b-tagging algorithm. CA15 jets in a transverse momentum range of 200 GeV $< p_T <$ 400 GeV are used for these studies. In the application to subjets of the fat jets the b-tagging algorithm performs much better than in the application to the CA15 jets themselves. Both a lower mistag rate and a higher b-tagging efficiency are measured for b-tagging of the subjets.

The top-tagging performance of the HEPTopTagger algorithm can be improved by requesting also one subjet of the top-tagged jet to be b-tagged by the CSV algorithm. The effect of this additional b-tag requirement on the top-tagging efficiency is shown on the left-hand side of figure 5.14. Requiring a subjet b-tag on the top-tagged CA15 jet reduces the top-tagging efficiency slightly. However, the effect on the misidentification rate of light jets, displayed in the right plot in figure 5.14, is much more pronounced: when requiring an additional CSV subjet b tag at the medium working point, the misidentification rate is decreased by approximately a factor of 10 to less than percent level. This significant gain in purity can compensate the slight loss in efficiency.

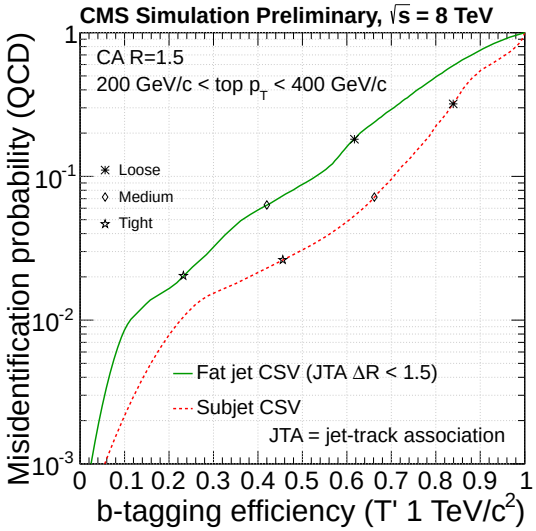

Figure 5.13: Light-jet-misidentification rate vs. b-jet-identification efficiency for the CSV b-tagging algorithm applied to CA15 jets (green) and their subjets (red) [14].

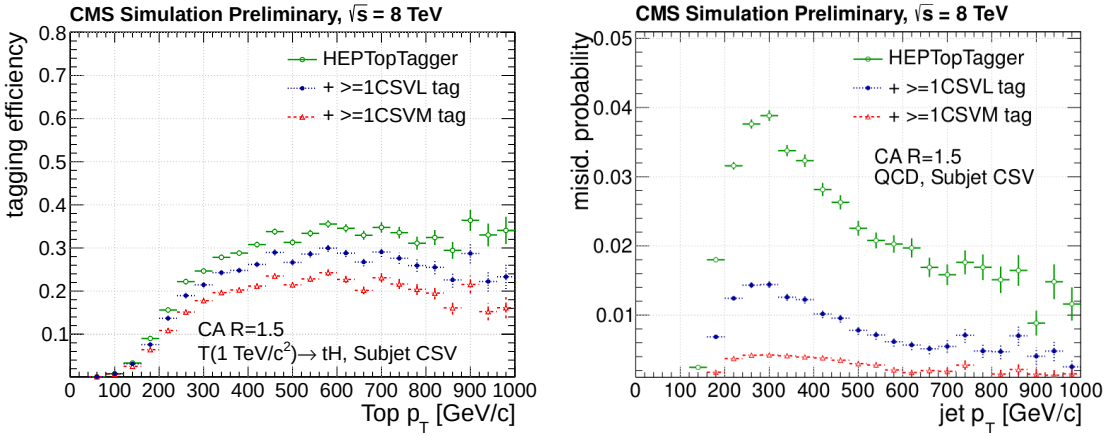

Figure 5.14: Top-tag efficiency in top jets (left) and mistag rate measured in light jets from a QCD-multijet sample (right) for the default HEPTopTagger algorithm (green), the HEPTopTagger algorithm with an additional CSV b tag at loose working point (blue,) and the HEPTopTagger algorithm with an additional CSV b tag at medium working point (red) [14].

Identification of bottom-quark pairs in jet substructure for identification of Higgs-boson decays

Decays of the standard model Higgs boson to $b\bar{b}$ pairs have a branching fraction of approximately 58% [37]. Therefore, tagging of $b\bar{b}$ pairs in jet substructure can be used to identify decays of Higgs bosons with large Lorentz boost. Studies of Higgs-tagging algorithms are performed using pruned Cambridge/Aachen jets with a cone radius of $R = 0.8$ (CA8 jets) in B'B' $\rightarrow$ bHbH events with B'-quark masses of 1 TeV.

The jet-pruning algorithm [114, 115] reclusters a Cambridge/Aachen jet starting with all of its jet constituents while applying certain pruning criteria. These criteria involve the hardness z of a combination of two subclusters i and j, defined as

$$z = \frac{min(p_T^i, p_T^j)}{p_T^{combination}}. \tag{5.4}$$

If the combination of two subclusters is softer than $z_{cut} = 0.1$, the softer contributing subcluster is discarded if, at the same time, the angle ΔR between the two subclusters is wider than

$$D_{cut} = 0.5 \times \frac{m^{CA8}}{p_T^{CA8}}, \tag{5.5}$$

where m^{CA8} and p_T^{CA8} are mass and transverse momentum of the original CA8 jet.

In Higgs tagging algorithms, the subjets of CA8 jets are obtained by undoing the last step in this jet pruning algorithm. The momenta of the resulting subjets are strongly correlated with the direction of the decay products of the Higgs boson. A mass requirement of $75 < m_{jet} < 135$ GeV is placed on the mass of the pruned jet in order to reduce the QCD-multijet background contribution. Using the remaining events, the efficiency and mistag probability of two b-tagging approaches are compared:

1. b tagging applied directly to CA8 jets
2. Higgs tagging, i.e., identification of two b-tagged subjets in combination with a Higgs-boson- mass window requirement. For both of these b tags the same operating point of the CSV b-tagging algorithm is chosen.

In figure 5.15, the results of the comparison of tagging performance are shown for jets in a range of $300 < p_T < 500$ GeV (left) and high p_T jets with $p_T > 700$ GeV (right). The tagging efficiencies are determined in events containing H $\rightarrow$ $b\bar{b}$ decays, while the misidentification rates are measured in QCD-multijet events. For the moderately boosted jets in the lower p_T range, the Higgs tagging algorithm clearly yields better results than regular b tagging of the CA8 jet. At low efficiencies, it results in a far higher purity, while at high efficiencies the light-jet rejection of the two taggers is at about the same level. For the decays of the highly-boosted Higgs bosons in the sample of high-p_T jets, the performance of the Higgs-tagging algorithm and the direct application of the b-tagging algorithm to the CA8 jets are very similar, as shown in the right plot in figure 5.15. Compared to regular b tagging of CA8 jets, a gain in purity can be achieved in the medium-boosted regime when using the Higgs-tagging algorithm in this form, while no loss in performance is observed even at very high jet transverse momenta.

In figure 5.16, the Higgs-tagging efficiency is compared for CA8 jets containing decays of different heavy particles, including Higgs bosons, W and Z bosons, and top quarks.

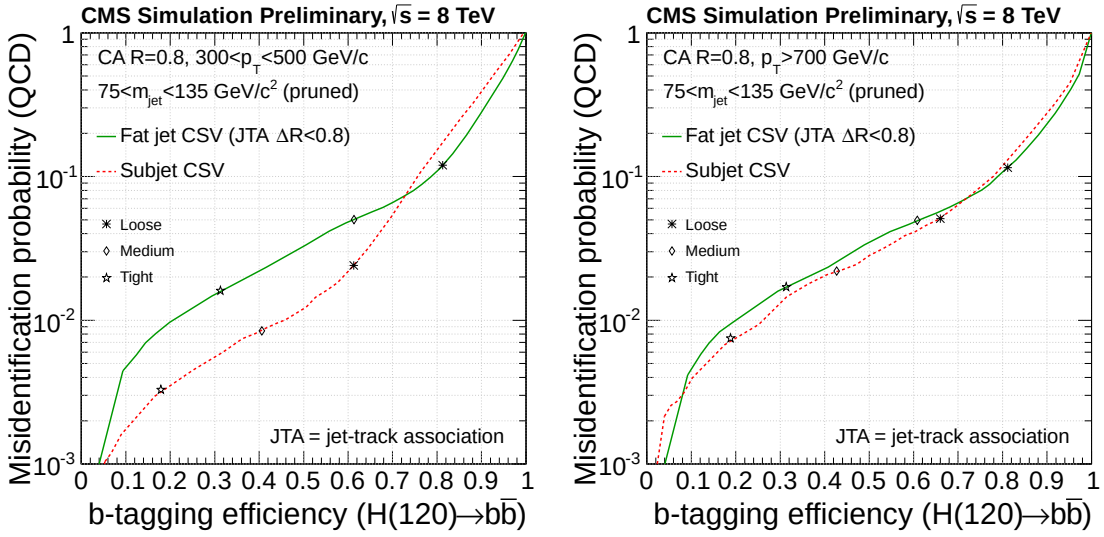

Figure 5.15: Performance of the Higgs-tagging algorithm (red) compared to the performance of direct application of the b-tagging algorithm to CA8 jets. In this comparison the CSV b-tagging algorithm at medium working point is used. Left: $300 < p_T(jet) < 500$ GeV. Right: $p_T(jet) > 700$ GeV [14].

Also, jets from QCD-multijet production are examined which exclusively consist of lighter particles. Here, a CA8 jet is classified as Higgs tagged, if its pruned mass falls into the mass window of 75-135 GeV and two of its subjets are tagged by the CSV-b-tagging algorithm at loose operating point. The efficiencies are calculated with respect to all CA8 jets with a $p_T > 300$ GeV prior to application of any pruned-jet-mass selection criterion.

A high Higgs-tagging efficiency of 40-50% is achieved over the full p_T range for jets containing $H \to b\bar{b}$ decays. The QCD-multijet background is reduced most effieciently to only 0.4%. But also the W/Z- and top-jet backgrounds in which a real bottom-quark contribution is expected are reduced. Only a single bottom quark is produced in a top-quark decay. Requiring two b-tagged subjets per CA8 jet is therefore a good handle to supress this background contribution. Also, the mass of a CA8 jet containing a top-quark decay is not expected to fall into the Higgs-boson-mass window of 75-135 GeV, wich leads to a further reduction of the misidentification rate in top jets.

5.2.5.3 Further jet substructure tools

Besides the previously described HEPTopTagger and subjet-b-tagging algorithms, there are several other jet-substructure tools that are used in CMS analyses. Brief descriptions of these methods are given in this section, more details can be found in [102, 115].

CMS Top Tagger

Like the HEPTopTagger algorithm, the CMS Top Tagger algorithm is used to identify hadronic decays of top quarks with high Lorentz boost within large radius Cambridge/Aachen jets. The CMS Top Tagger is applied to CA8 jets [98]. It is based on

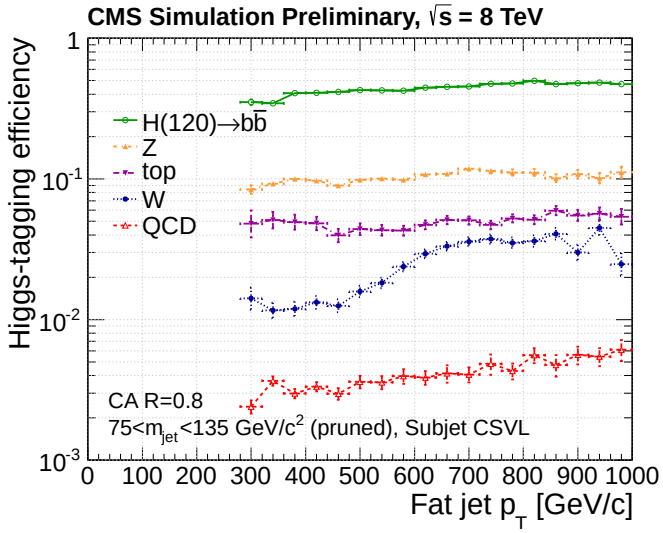

Figure 5.16: Higgs-tagging efficiency for jets containing decays of different heavy particles and light QCD jets. The efficiencies are calculated with respect to all CA8 jets before the pruned-jet-mass selection [14].

the JHU Top Tagger [116] and employs a different jet-declustering method than the HEP-TopTagger algorithm making it more suitable for highly boosted jets. As illustrated in figure 5.17, the CA8 input jets are decomposed in two tiers: the primary and the secondary jet decomposition.

In the primary decomposition, the jet clustering of the Cambridge/Aachen algorithm is reverted. When moving back one step in the clustering history of the CA8 jet, the adjacency criterion

$$\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.4 - 0.0004 \times p_T^{input} \tag{5.6}$$

must be fulfilled for the two parent subjets. Otherwise, the CA8 jet has only one subjet. Here, p_T^{input} is the transverse momentum of the input cluster, in this first declustering step the original CA8 jet. If one of the parent clusters does not satisfy the momentum-fraction criterion

$$p_T^{subcluster} > 0.05 \times p_T^{input}, \tag{5.7}$$

it is discarded. In this case, another step in the jet-clustering history is reverted and the adjacency and momentum-fraction criteria are evaluated once more. This procedure is repeated until two subclusters satisfying both criteria have been identified. If this is not possible, the jet has only a single subjet. Once these two subjets are obtained, the same criteria are applied another time in the splitting of the two subclusters in the secondary decomposition. This final decomposition step results in either two, three, or four final subjets. CA8 jets with $p_T > 350$ GeV containing at least three subjets are tagged by the CMS Top Tagger, if their jet masses and on of the minimum pairwise masses of their three subjets are compatible with the top-quark mass and the W-boson mass, respectively. The CMS Top Tagger is used in a number of CMS analyses in the boosted regime, for example [117, 118].

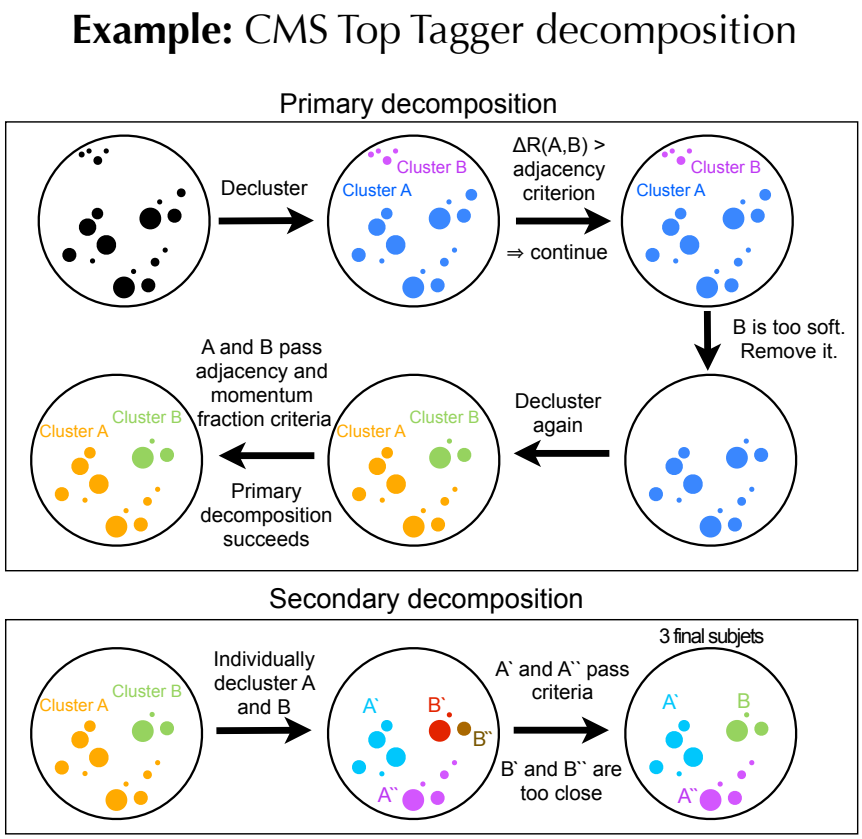

Figure 5.17: Illustration of the jet decomposition in the CMS Top Tagger algorithm [112].

Shower deconstruction tagger

The shower deconstruction tagger is another algorithm to identify hadronic top-quark decays merged into CA jets. Here, the k_T jet clustering algorithm is applied to all jet constituents of CA8 or CA15 jets to find micro jets [119]. These micro jets must have a transverse momentum of at least 10 GeV to be considered by the top-tagging algorithm. The cone size of the micro jets decreases in three steps with increasing jet p_T from $R = 0.3$ to $R = 0.1$. After application of a W-boson-mass window and top-quark-mass window selection, the discriminating variable χ is calculated and used to distinguish top jets from background. χ is approximately the ratio of two likelihoods: the likelihood for a top jet to have the exact structure of the CA15 jet under study and the likelihood for a light jet from a QCD-multijet background event to produce this structure. Having been developed only very recently, the shower deconstruction tagger has not yet been used in any physics analyses in CMS to date. First performance studies are documented in [120].

N-subjettiness

The N-subjettiness τ_N is a measure of how compatible a jet is with having a certain number of subjets N [121]. For application of this algorithm, candidate subjets of fat jets are defined with the exclusive-k_T algorithm [99, 122]. A certain number N of subjets is

forced in this clustering procedure. The shape variable τ_N is defined as

$$\tau_N = \frac{\sum_i^n p_{T,i} min\{\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i}\}}{\sum_i^n p_{T,i} R} \tag{5.8}$$

summing over all of the n particle-flow jet constituents. R is the jet radius of the input jet. With this definition, τ_N is a p_T-weighted sum of the angular distance of the jet constituents to the closest candidate-subjets axis R_{ij}. This means, that small values of τ_N indicate that the jet is likely to have N or less subjets. Often, ratios of two N-subjettiness variables are used as discriminators, e.g., τ_3/τ_2 is very suitable to identify top jets, which are expected to have three subjets rather than two or less. The N-subjettiness is used for example in [118].

W tagging

A number of variables is suited to distinguish jets containing W-boson decays from light QCD jets as described in detail in [115]. One of them is the pruned-jet mass which is the mass of a Cambridge/Aachen jet after application of the jet-pruning algorithm. The jet-pruning algorithm is applied in the same way as in the Higgs-tagging algorithm described in section 5.2.5.2. The N-subjettiness ratio τ_2/τ_1 can be used to identify jets with exactly two subjets. If a so-called mass-drop criterion is to be applied, the last step of the pruned-jet clustering is undone. The mass ratio between the most massive subcluster and the original pruned jet is the mass drop $\mu = \frac{m_1}{m_{jet}}$. For applications of W-boson tagging in CMS analyses see for example [123]. Other variables with high potential to improve the performance of W-boson tagging include the jet charge [124], a measure for the electric charge of the particle from which the jet originates, the generalized energy correlation functions C_2^β [125], and the Q jet volatility Γ_{Qjet} [126]. None of these variables have been applied in searches for BSM physics to date.

5.2.6 Performance of the HEPTopTagger and subjet-b-tagging algorithms in data and simulation

As some variables used in the algorithms for analysis of the jet substructure are not perfectly described by the simulation, their efficiency measured in data and simulation can differ. Therefore, scale factors are derived for efficiency and misidentification rate to compensate for the observed differences.

Performance of the subjet-b-tagging algorithm in data and simulation

In order to derive scale factors for the subjet-b-tagging performance, semileptonic $\mathrm{t\bar{t}}$ events are used, in which the hadronic top-quark decay is merged into a single jet that is then tagged by the HEPTopTagger algorithm [14]. The flavor-tag consistency method also used in the efficiency measurement for the default CSV-b-tagging algorithm (see section 5.2.4.2) is slightly modified to measure the b-tagging efficiency on subjets. An inclusive measurement of the b-tagging scale factor, as well as three separate measurements in different transverse momentum categories are performed. The results found in jet substructure for the CSV-b-tagging algorithm are shown in table 5.1. They are consistent with the scale

factors measured for b tagging in AK5 jets. Therefore, the regular b-tagging scale factors can also be used as correction factors in the application of the subjet-b-tagging algorithm.

p_T of CA15 jet	Scale factor CSVM
Inclusive ($p_T > 150$ GeV)	0.979 ± 0.023
$150 \leq p_T < 350$ GeV	$0.978^{+0.023}_{-0.023}$
$p_T \geq 350$ GeV	$0.993^{+0.034}_{-0.034}$
$p_T \geq 450$ GeV	$0.997^{+0.067}_{-0.067}$

Table 5.1: Scale factors for subjet b-tagging with the CSV algorithm in CA15 jets top-tagged using the HEPTopTagger algorithm. All values correspond to the medium working point of the CSV-b-tagging algorithm. [14]

Performance of the HEPTopTagger algorithm in data and simulation

One example for discrepancies between data and simulated events in jet substructure variables is the distribution of the invariant mass of the three filtered subjets shown in figure 5.18. Therefore, data-Monte Carlo scale factors are also derived for the HEPTopTagger.

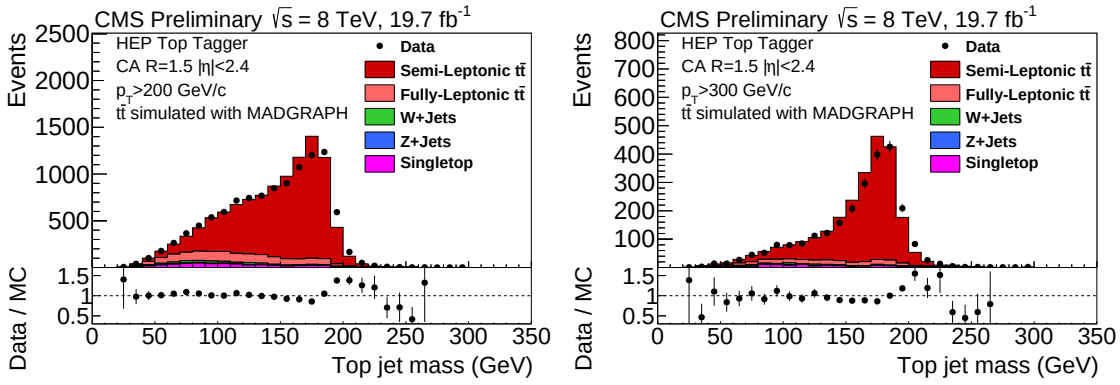

Figure 5.18: Distribution of the invariant mass of the three filtered subjets of CA15 jets with a transverse momentum $200 < p_T < 300$ GeV (left) and $300 < p_T < 400$ GeV (right) [102]. The observed mismodelling of the top-quark-mass peak might be due to a discrepancy in jet energy resolution observed between data and simulation [104].

The scale factors for the HEPTopTagger algorithm are measured in a tag-and-probe approach in semileptonic $t\bar{t}$ events [102]. In order to select the leptonic top-quark decay, i.e., the tag, exactly one high-p_T muon and an AK5 jet tagged by the CSV-b-tagging algorithm at the medium working point are required. The selected events are split into two hemispheres: the leptonic one with a radial distance to the muon of $\Delta R < \pi/2$ and

the hadronic hemisphere with $\Delta R \geq \pi/2$. The performance of the algorithm is then probed on the p_T-leading CA15 jet in the hadronic hemisphere with a $p_T > 200$ GeV that contains a b-tagged subjet. The denominator of the efficiency is the full number of such CA15 jets, the enumerator is given by the subset of jets which are actually tagged by the HEPTopTagger algorithm. The scale factor is obtained by dividing the efficiency measured in data by that measured in simulation. The measurement is performed in two regions of pseudorapidity and three jet-p_T categories, the results can be found in table 5.2.

Jet p_T	$\lvert\eta\rvert < 1.0$	$1.0 < \lvert\eta\rvert < 2.4$
Inclusive ($200 < p_T < 250$ GeV)	0.92 ± 0.04	0.93 ± 0.06
$250 \leq p_T < 400$ GeV	0.95 ±0.03	0.95 ± 0.05
$p_T \geq 400$ GeV	1.36 ± 0.07	0.95± 0.15

Table 5.2: Scale factors for the HEPTopTagger efficiency measured in CA15 jets in semileptonic $t\bar{t}$ events that contain a b-tagged subjet [102].

6 Statistical methods

The objective of searches in high-energy physics is to observe new physics processes. In case of a discovery of new physics, a discovery significance is reported. Else, exclusion limits are determined, e.g., for the cross section of the process of interest. Methods of Bayesian statistics can be employed to obtain such exclusion limits. The main concepts of Bayesian statistics are provided in the first section of this chapter. Afterwards, an introduction to limit setting using the theta framework [127] is given in section 6.2. Theta is a framework for statistical modeling and was designed specifically for applications in high-energy physics. Afterwards, an alternative method for limit determination based on frequentist statistics is introduced in section 6.3. In the course of physics analyses, it is often necessary to asses the agreement between different distributions. The χ^2 method is often used for these kind of tests, as explained in the last section of this chapter.

6.1 Bayesian statistics

The key concept in Bayesian statistics is to summarize all knowledge about a certain model parameter θ in a so-called posterior distribution $p(\theta|x)$ [16, 128]. Given a specific dataset x, the posterior distribution expresses a degree of belief for the parameter θ to assume certain values. An important ingredient in the determination of the posterior distribution are the so-called prior distributions. A prior distribution expresses subjective assumptions about the probability density functions of a certain model parameter before the evaluation of measured data. There is always some arbitrariness in the choice of the prior distribution. Therefore, it is valuable to provide the prior distributions used in the statistical evaluation along with the results, when using methods based on Bayesian statistics in a physics analysis.

In general, the posterior distribution of a certain parameter θ can be determined using Bayes' theorem

$$p(\theta|x) = \frac{L(x|\theta) \cdot \pi(\theta)}{f(x)}, \tag{6.1}$$

where $\pi(\theta)$ is the assumed prior distribution for θ, and $L(x|\theta)$ the likelihood function describing the probability density function for a certain dataset x as a function of the parameter θ. The function $f(x)$ is needed only for normalization of the posterior distribution.

The Bayesian interval $[\theta_{low}, \theta_{up}]$ at a confidence level β is defined as

$$\beta = \int_{\theta_{low}}^{\theta_{up}} p(\theta|x) d\theta. \tag{6.2}$$

In elementary particle physics, the parameter of interest θ usually is the mean value of the Poisson distribution of the number of counted signal events n. The parameter θ can be related to the cross section of the process under study and cannot assume negative values. Therefore, one often uses a prior of the form

$$\pi(s) = \begin{cases} 0 & \text{for } \theta \leq 0 \\ 1 & \text{for } \theta > 0 \end{cases} \tag{6.3}$$

to describe the assumptions about the parameter of interest. These kind of priors are called improper, as their integral is larger than unity which is not expected for a probability density function. In most applications of these priors, this is not problematic because only the posterior distribution for θ is of interest. If the likelihood function in 6.1 falls quickly enough for large values of x, the posterior function is still well defined. A caveat in Bayesian statistics is the impossibility to define a prior that does not already contain any information about the parameter of interest. On first glance, setting the prior function to a constant value seems to introduce little information. A flat distribution in a certain parameterization of the model may not be flat when moving to a different parameterization though. This kind of arbitrariness cannot be avoided in Bayesian statistics. Priors of the form given in equation 6.3 in general lead to good coverage when setting upper exclusion limits.

In case exclusion limits are to be set for the parameter θ, the upper limit s_{up} on θ at confidence level β can be derived from

$$\beta = \int_0^{s_{up}} p(s|n)ds. \tag{6.4}$$

Use of the likelihood function for Poisson-distributed n and the prior in equation 6.3 results in a relation for s_{up} that can be solved numerically.

Usually, a model depends on more parameters than just the parameter of interest θ. These additional parameters are often referred to as "nuisance parameters". In practice, the systematic uncertainties of an analysis are identified with these nuisance parameters. In this case, the posterior distribution of the parameter of interest also exhibits a dependence on the nuisance parameters. To obtain the marginal posterior for the parameter θ, which has no dependence on the nuisance parameter, an integration over all nuisance parameters in the model is performed:

$$p(\theta|x) = \int p(\theta, \nu|x)d\nu. \tag{6.5}$$

Often, a Gaussian function is chosen to describe the prior of nuisance parameters. For parameters that cannot assume negative values, e.g., event rates, the choice of a Gaussian function around the expected mean value leads to an unphysical truncation of the prior distribution at zero. This can introduce a bias towards higher values. Therefore, log-normal functions are more suited to describe the prior distributions of these kind of parameters.

6.2 Deriving Bayesian exclusion limits with the theta framework

A template-based approach to statistical evaluation of the measured data is employed in the theta framework [127]. The expected data distributions are described by the sum of different histograms, the templates. Individual templates are obtained from Monte Carlo simulation or data driven background descriptions for each of the different expected physics processes contributing to the distribution of the data. These templates represent

the probability density functions of the different contributions. Also the measured data is provided to the theta framework in form of histograms. The data in each bin of the histogram are expected to be Poisson distributed.

The nuisance parameters can affect the statistical model in two ways: if a systematic uncertainty only affects the overall normalization of the templates, i.e., the expected event yield of a certain contributing physics process, a Gaussian is chosen for the prior distribution of the corresponding nuisance parameter. In this case, only an estimate of the uncertainty in the event rate needs to be provided and the entire histogram is scaled according to this estimate. For parameters that cannot assume negative values, e.g., event rates, not the event yield μ itself, but a parameterization $\mu = e^{\lambda\nu}$ with the nuisance parameter ν and a constant λ is chosen to avoid the effects of truncation explained above. This way, the event yield does not assume negative values regardless of the Gaussian prior distribution for the corresponding nuisance parameter ν. If also the shape of the template is affected by the uncertainty, i.e., the impact on the content of histogram bins is observed to vary between different bins, additional histograms quantifying this effect need to be provided. These histograms are obtained in a variation of the nuisance parameter by one standard deviation in upward and downward directions in the production of the templates. The prior of the associated nuisance parameter ν again follows a Gaussian distribution. A variation of the nuisance parameter by one standard deviation then results in the additionally provided systematically varied template, for other variations a bin-by-bin interpolation between the provided nominal and systematically varied templates is performed. This procedure is referred to as "template morphing".

The statistical model is evaluated individually in each bin of the provided templates. The expected mean value m_i per bin i of the histogram is

$$m_i(x, \nu) = \theta \cdot T_{signal} + \sum_j y_j \cdot T_j(x, \nu), \tag{6.6}$$

where the index j runs over all expected background processes, θ is the parameter corresponding to the signal cross section, the y_j are other real-valued coefficients, and the T_j and T_{signal} are the templates provided for each of the background processes and the expected signal, respectively.

The templates for the contributing physics processes are derived using Monte Carlo simulation techniques. The statistical accuracy of the estimated contribution in each bin of the histograms depends on the number of simulated events falling into this bin. In a method proposed by Barlow and Beeston [129], separate nuisance parameters are assigned to each bin of the histograms for every contributing process, to take into account the limited size of the simulated samples. As this procedure complicates the computation of limits to a great extent, a simplified approach proposed in [130] is implemented in the theta framework. Here, one nuisance parameter per histogram bin is assigned. The impact of these specific nuisance parameters on the results is determined analytically before starting the actual limit-setting procedure.

When Bayesian exclusion limits are set using the theta framework, the nuisance parameters in the posterior distribution of the parameter of interest are integrated out according to equation 6.5 using Markov-chain Monte Carlo methods [16, 131].

Expected exclusion limits for the cross section of the hypothetical signal are obtained in fits to so-called toy datasets. These toy datasets contain artificial data events created in

a random manner. The probability density functions for different model parameters, provided in the form of the templates and the prior distributions of the nuisance parameters, are taken into account in the generation of these toy datasets. For each of the toy datasets, a hypothetical limit is calculated in a Bayesian procedure, resulting in a probability distribution for the expected limit. Usually the expected limits, i.e., the median value of this distribution, as well as symmetric intervals around the median value containing 68% and 95% of the distribution are quoted. The expected limits with their uncertainty intervals quantify the sensitivity of an analysis for setting exclusion limits on a certain parameter. The observed limit is obtained in a fit to the actually measured data.

6.3 Deriving exclusion limits using frequentist statistics

In frequentist statistics, the frequency of a certain outcome of a repeatable experiment determines the probability. No probability for a single hypothesis or parameter is defined in the frequentist approach. In contrast to the Bayesian methods described above, no prior beliefs are incorporated in methods based on frequentist statistics. A likelihood function can be defined using the probability distribution function $f(\mathbf{x}, \theta)$ where $\mathbf{x} = (x_1, ..., x_N)$ is a set of N measured quantities and $\theta = (\theta_1, ..., \theta_n)$ a set of n unknown parameters. For statistically independent measurements x_i following a probability distribution function $f(x, \theta)$, the joint probability distribution for all x_i factorizes, resulting in a likelihood function

$$L(\theta) = \prod_{i=1}^{N} f(x_i, \theta). \tag{6.7}$$

In the maximum likelihood method, the set of values θ that result in the maximum likelihood give the point estimators of the different parameters.

In order to derive exclusion limits in a frequentist approach, a test statistic of the observable under study is defined. Often likelihood variables are used as test statistics. Two types of hypotheses are formulated: the background only hypothesis and the signal + background hypotheses including contributions arising from the hypothetical new physics with different signal strengths μ. A confidence level CL_x is assigned to each hypothesis, where

$$CL_x = P_x(Q \leq Q_{obs}), \tag{6.8}$$

i.e., the probability for the test statistic to assume a value Q not greater than the value Q_{obs} observed in data. In the actual limit computation, the assumed signal strength μ in the signal + background hypothesis is varied. The limit at confidence level α is set to that value of the signal strength μ, for which $1 - \mathrm{CL}_x \leq \alpha$.

A caveat in this limit setting procedure is the possibility to overestimate the sensitivity to a certain model, due to underfluctuations in the measured data. In the CL_S method, the frequentist limit setting procedure is slightly modified. Here, the background only hypothesis is also taken into account and the ratio

$$CL_S = \frac{CL_{S+B}}{CL_B} \tag{6.9}$$

is evaluated for limit setting instead of CL_{S+B}. This way, less weight is given to signal + background

hypotheses with small signal strengths μ that are very similar to the background only hypothesis.

Combinations of frequentist and Bayesian methods are also used in high energy physics, e.g., to treat the systematic uncertainties in a Bayesian approach and then conduct a frequentist statistical test of the method.

6.4 χ^2 tests and the p-value

The objective of so-called χ^2 tests is to quantify the level of agreement between two distributions. A common application in high energy physics is the comparison of two histograms containing Poisson distributed numbers. In this case the χ^2 value

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \nu_i)^2}{\nu_i + n_i} \tag{6.10}$$

is defined, where the n_i are the contents of the N bins of one histogram, those of the second histogram are the ν_i. The χ^2 value can be used as a goodness-of-fit statistic between the two histograms. The actual level of compatibility between the two histograms is then quantified by the p-value. It is the probability to obtain values for the goodness-of-fit statistic larger or equal to the one measured for the two histograms in question. A large p-value therefore indicates a good agreement between the two histograms.

7 Search for pair-produced T' quarks in all-hadronic final states

In this chapter, a search for pair-produced vector-like T' quarks decaying into all-hadronic final states is presented. After an overview of the analysis strategy in section 7.1, the datasets and simulated samples used in this analysis are listed in section 7.2. Details on the event selection can be found in section 7.3. In section 7.4, the method used to derive the QCD-multijet background contribution from data is explained. The different sources of systematic uncertainties are discussed and their impact on the analysis are quantified in section 7.5. The results of the analysis are presented in section 7.6.

7.1 Analysis Strategy

There are three decay modes for vector-like T' quarks: T' $\rightarrow$ tH, T'$\rightarrow$ tZ, and T' $\rightarrow$ bW. Depending on the decay channel, the event signatures of the vector-like quarks in the detector can be very different. The search for vector-like quarks presented in this chapter is optimized for decays of the T' quark to top quarks and Higgs bosons. The T' quarks are produced in particle-antiparticle pairs. Therefore, two top quarks and two Higgs bosons are expected to be contained in each event, as illustrated in the Feynman diagram in figure 7.1.

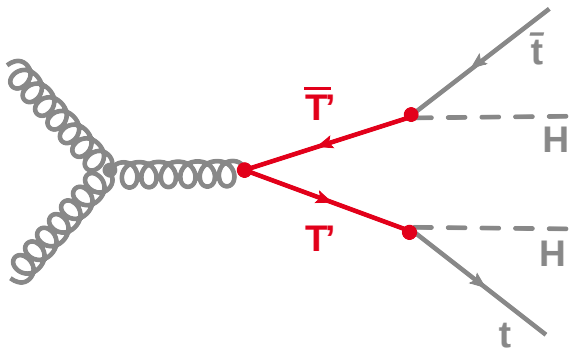

Figure 7.1: Feynman diagram of T'-quark pair production with two decays of T' $\rightarrow$ tH.

This analysis specializes on hadronic decays of the top quark and the Higgs boson. A large number of hadronic decays occur in the T' $\rightarrow$ tH channel. Approximately two thirds of all top quarks decay hadronically and the hadronic decay mode of the Higgs boson, H$\rightarrow$ b$\bar{\text{b}}$, has a branching fraction of about than 58%. This makes analyses in the all-hadronic channel highly interesting.

The absence of isolated charged leptons is a challenge for analyses with hadronic decays though. For the event filters applied in the data taking it is much easier to identify events in which a lepton is present. Also in the offline event selection, the requirement of a

well-reconstructed isolated lepton is a good handle to reduce the number of selected background events, especially the contribution from QCD-multijet events. Another advantage of analyses in the leptonic channel is the better resolution for measurements of leptons compared to measurements of jets. Mass spectra of resonances can be reconstructed much more precisely in leptonic decay channels, as the uncertainty in the 4-vectors of jets are much larger than for charged leptons.

This search in the hadronic channel is further complicated by the large multiplicity of jets in the final state: the pair of T' quarks decays into two top quarks and two Higgs bosons. Each hadronically decaying top quark contributes three jets to the event topology, where at least one of them is a b jet. Two further b jets are added by each decaying Higgs boson. This amounts to a total of ten jets in the final state of the hypothetical signal events.

Previous searches for T' quarks have excluded the possibility of low-mass vector-like quarks. With increasing sensitivity of the analyses, higher limits are set for the mass of the particles. In decays of high-mass resonances, the daughter particles are expected to have a large Lorentz boost. When these boosted top quarks and Higgs bosons decay, their decay products are likely to be found within small spatial distances. Therefore, the efficiency to reconstruct each of the decay products as an individual jet decreases.

To counteract the obstacles in this demanding decay channel, a novel approach is adopted in this analysis: instead of attempting to reconstruct and correctly identify each of these ten jets individually, large-radius Cambridge/Aachen jets are clustered in the event reconstruction. As described in section 5.2.5, it is then likely that all decay products of a decaying boosted top quark or Higgs boson are merged into a single one of these Cambridge/Aachen jets. In the event selection, innovative jet-substructure tools are then employed to identify signal-like events with jets containing hadronic decays of top quarks or Higgs bosons. In this search, the HEPTopTagger algorithm as well as the novel technique of subjet b tagging are applied for the first time in an analysis of data recorded with the CMS experiment.

The QCD-multijet background for this analysis is estimated from data via an ABCD method. In this method, the information provided by the jet-substructure tools is used as well, as described in section 7.4.

Two variables with promising sensitivity to the T'-quark signal are identified: the H_T, defined as the scalar sum of the p_T of all subjets of filtered CA15 jets with a $p_T > 150$ GeV, and the so-called Higgs-candidate mass, reconstructed from the subjets of a Higgs-tagged jet. To exploit the sensitivity of both of these variables, they are combined into a single, more powerful discriminating variable using a likelihood-ratio method.

In addition to this measurement in the T'T'$\rightarrow$ tHtH decay channel, all other possible decay modes of the vector-like T' quark are considered as well. For this purpose, a scan of the branching fractions into each decay mode is performed. Results are then obtained for each combination of the three decay modes T' $\rightarrow$ tH, T' $\rightarrow$ tZ, and T' $\rightarrow$ bW.

7.2 Datasets and simulated samples

7.2.1 Simulated samples and datasets used in this analysis

The full 2012 dataset recorded with the CMS detector at a center-of-mass-energy of 8 TeV is analyzed in this search. A jet-based trigger requiring H_T^{calo} to be larger than 750 GeV is used in the event filter, where H_T^{calo} is the scalar sum of the transverse momenta of all calorimeter jets in the event[1]. The names of the specific data samples with their corresponding integrated luminosity are given in table 7.1.

Dataset	Luminosity (pb^{-1})
`/Jet/Run2012A-22Jan2013-v1/AOD`	888
`/JetHT/Run2012B-22Jan2013-v1/AOD`	4403
`/JetHT/Run2012C-22Jan2013-v1/AOD`	7052
`/JetHT/Run2012D-22Jan2013-v1/AOD`	7414

Table 7.1: Used datasets and the corresponding integrated luminosities.

The main backgrounds to this analysis are top-quark pair and QCD-multijet production. In the optimization of the event selection, events simulated using Monte Carlo techniques are used to describe distributions of the expected background and signal events.

More information on the functioning of the programs used in the simulation of these samples can be found in chapter 4. The background from pair-produced top quarks is modelled using the POWHEG matrix-element generator [132] interfaced with the PYTHIA6 [83] program for the simulation of the hadronic showering. The phase space is vastly constrained in the event selection, reducing the background contribution from top-quark pair production significantly.

Studies are also performed using the MADGRAPH program as matrix-element generator in the simulation of $t\bar{t}$ background. In general, no significant differences are observed in the description of event variables between the samples generated using POWHEG and MADGRAPH. Detailed results of the comparison between the two samples can be found in appendix B.

The templates and event rates for the modeling of the QCD-multijet background distribution are derived from data in this analysis using an ABCD method. The specifics of this ABCD method can be found in section 7.4. Simulated samples are used in the validation of the ABCD method and in the optimization of the event selection though. They are generated with the MADGRAPH event generator interfaced with PYTHIA6 for showering.

Each signal event contains two T'-quark decays. For each pairwise combination of the three decay modes T'$\rightarrow$ tH, T' $\rightarrow$ tZ, and T'$\rightarrow$ bW, individual Monte Carlo samples are produced, resulting in six signal samples per T'-quark-mass point. The mass of the hypothetical T' quark is varied between 500 GeV and 1 TeV in steps of 100 GeV. In the production of all signal samples, the MADGRAPH matrix-element generator is interfaced with PYTHIA6 for shower generation. The mass of the Higgs boson is set to a value of 120 GeV in the generation of these samples. As explained in section 2.1.5, the branching

[1] This definition of the H_T variable is different from the definition used later on in the analysis. In this work, H_T refers to the quantity obtained from the subjets of the CA15 jets in the event, if not explicitely stated otherwise.

fractions of the Higgs-boson decays strongly depend on the mass of the Higgs boson. Therefore, the branching fractions for the Higgs-boson-decay modes are corrected to the value corresponding to the measured Higgs-boson mass of approximately 125 GeV in the simulated signal samples [33].

The CTEQ6l1 [76] parton-distribution functions are used in the generation of the signal samples and the $t\bar{t}$ background.

Additional background from $t\bar{t}H$ events is considered, but the contribution in the signal region is found to be negligible. The same is true for W/Z + $b\bar{b}$ events with hadronically decaying W or Z bosons. For examination of the latter two processes, MC samples are produced using the MadGraph event generator. The $t\bar{t}H$ events are simulated with Pythia6.

The total inclusive production cross sections used in the normalization of all simulated background samples are listed in table 7.2, those for the T'-quark signal assuming different T'-quark masses in table 7.3. They are computed with the Top++2.0 program [133] with exact NNLO and full NNLL soft gluon re-summation. MSTW2008nnlo68cl PDF [134] and version 5.9.0 of LHAPDF [135] are used in these computations.

Sample	Cross section	
$t\bar{t}$	245.8 pb	NNLO [134]
QCD H_T 500-1000 GeV	8426.0 pb	LO
QCD $H_T > 1000$ GeV	204.0 pb	LO
Z + $b\bar{b}$ with Z→$b\bar{b}$	61.43 pb	LO
W + $b\bar{b}$ with W→qq'	27.16 pb	LO
$t\bar{t}$ + H	0.075 pb	NLO [37]

Table 7.2: Cross sections for the background samples obtained from Monte Carlo simulation. The quoted leading-order (LO) cross sections are determined using MadGraph or Pythia6.

T'-quark mass	Cross section
500 GeV	0.59 pb
600 GeV	0.174 pb
700 GeV	0.0585 pb
800 GeV	0.0213 pb
900 GeV	0.0083 pb
1000 GeV	0.00336 pb

Table 7.3: Cross sections for the simulated signal samples at different T'-quark-mass points.

7.2.2 Application of jet energy corrections

As explained in section 5.2.3, corrections have to be applied to the reconstructed jets in order to obtain a jet energy response that is independent of η and p_T. No dedicated jet energy corrections for CA15 jets are available for CMS analyses. Therefore, the correction factors derived for anti-k_T jets with an R parameter of 0.7 (AK7 jets) [136] are used to correct the jet energy scale of the larger radius CA15 jets. This choice was made, because no correction factors were derived for jets with larger radii than 0.7. In order to validate this approach, studies using samples simulated with Monte Carlo generators are performed: jets clustered from the particle-flow particles are compared to so-called generator jets. Generator jets are clustered directly from the stable particles generated in the Monte Carlo simulation. Figure 7.2 shows the jet response as function of the transverse momentum of the generator jet p_T^{GEN} (upper left), the pseudorapidity η^{GEN} of the generator jet (upper right), and the number of primary vertices in the event (bottom row). The curves drawn in magenta are obtained in a comparison of reconstructed CA15 jets to generator CA15 jets. Here, the reconstructed CA15 jets are corrected using the jet energy scale corrections derived for ak7 particle-flow jets after charged hadron subtraction. The black curve shows a similar comparison of corrected AK5 jets to generator AK5 jets. Dedicated jet energy corrections are available for AK5 particle-flow jets after charged hadron subtraction though. These jets are widely used in CMS analyses of final states with smaller Lorentz boost. They can therefore be used as a reference case to assess how suitable the ak7 jet energy corrections are for application to CA15 jets. The differences in response between the such corrected AK5 and CA15 jets do not exceed the 4% level, the largest deviations are found for large values of the number of primary vertices and at large η. The ak7 jet energy correction factors are therefore suitable for application to CA15 jets.

The H_T variable is calculated from the p_T of subjets and is very important to this analysis. Therefore, the jet energy scale of the subjets of the CA15 jets also needs to be considered. As for the CA15 jets themselves, no dedicated corrections for the jet energy scale are available for the subjets of CA15 jets. Also, no scale factors to correct for potential differences between the jet energy resolution in data and simulation are available for the subjets. Corrections of the jet energy resolution are only provided for AK5 jets in CMS.

The effects of the corrections of the jet energy scale and those of the jet energy resolution are therefore studied for the subjets of CA15 jets using $\mathrm{t\bar{t}}$ events with one leptonically and one hadronically decaying top quark. These kind of events are selected requiring a single well-isolated muon. No other leptons can be contained in these events. Using the location of the selected muon in the detector, each event is split into a leptonic and a hadronic hemisphere, where $|\phi - \phi_\mu| > \frac{2}{3}\pi$ in the hadronic hemisphere and $|\phi - \phi_\mu| < \frac{2}{3}\pi$ in the leptonic one. In addition, there has to be at least one AK5 jet in the leptonic hemisphere. This jet has to be b tagged using the CSV algorithm. In the hadronic hemisphere, a single CA15 jet has to be found. For the application of the HEPTopTagger algorithm, jets from the jet collection used as input for the HEPTopTagger algorithm are matched to the filtered CA15 jets. In the following, a CA15 jet is referred to as tagged by the HEPTopTagger algorithm, if the matching jet from the jet collection used as input to the HEPTopTagger is tagged by this algorithm. In the event selection used for this study, the selected CA15 jet in the hadronic hemisphere needs to be tagged by the HEPTopTagger

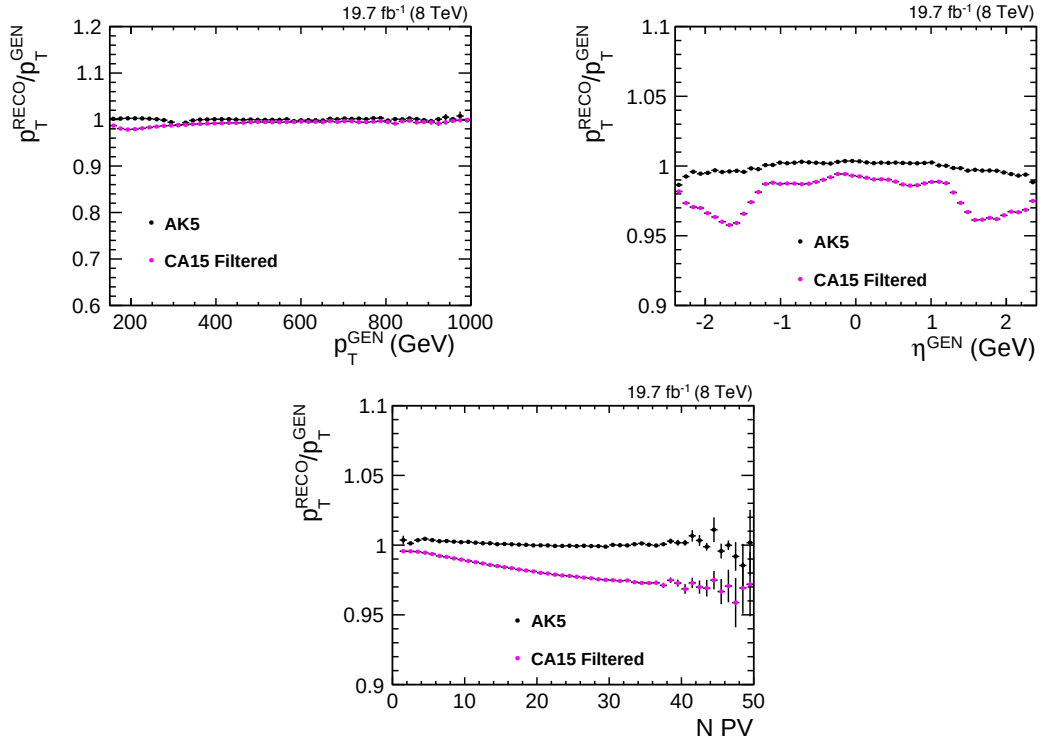

Figure 7.2: Jet response as a function of the generator-jet p_T (upper left), generator-jet η (upper right), and number of primary vertices (bottom row). Comparison of jet response of AK5 jets (black) corrected using the energy-scale corrections for AK5 jets to the response of CA15 jets (magenta) corrected using the energy-scale corrections derived for ak7 jets [137].

algorithm. Exactly one of the subjets of the CA15 jet also has to be b tagged. With this event selection, a sample containing mainly $t\bar{t}$ events and only very few events from background processes such as W-boson-plus-jets production is obtained.

The two subjets of the CA15 jet that are not b tagged are used to reconstruct the W boson from the hadronic top decay. Figure 7.3 shows the invariant mass of these two subjets in data and simulation. In the top left plot, no jet energy corrections are applied, while in the top right plot only the jet energy scale is corrected but not the jet energy resolution. The plot on the bottom left shows the distribution after application of the jet energy corrections and those for the jet energy resolution. Both, the values for the jet energy corrections and the corrections for the jet energy resolution that are applied here were originally derived for AK5 jets.

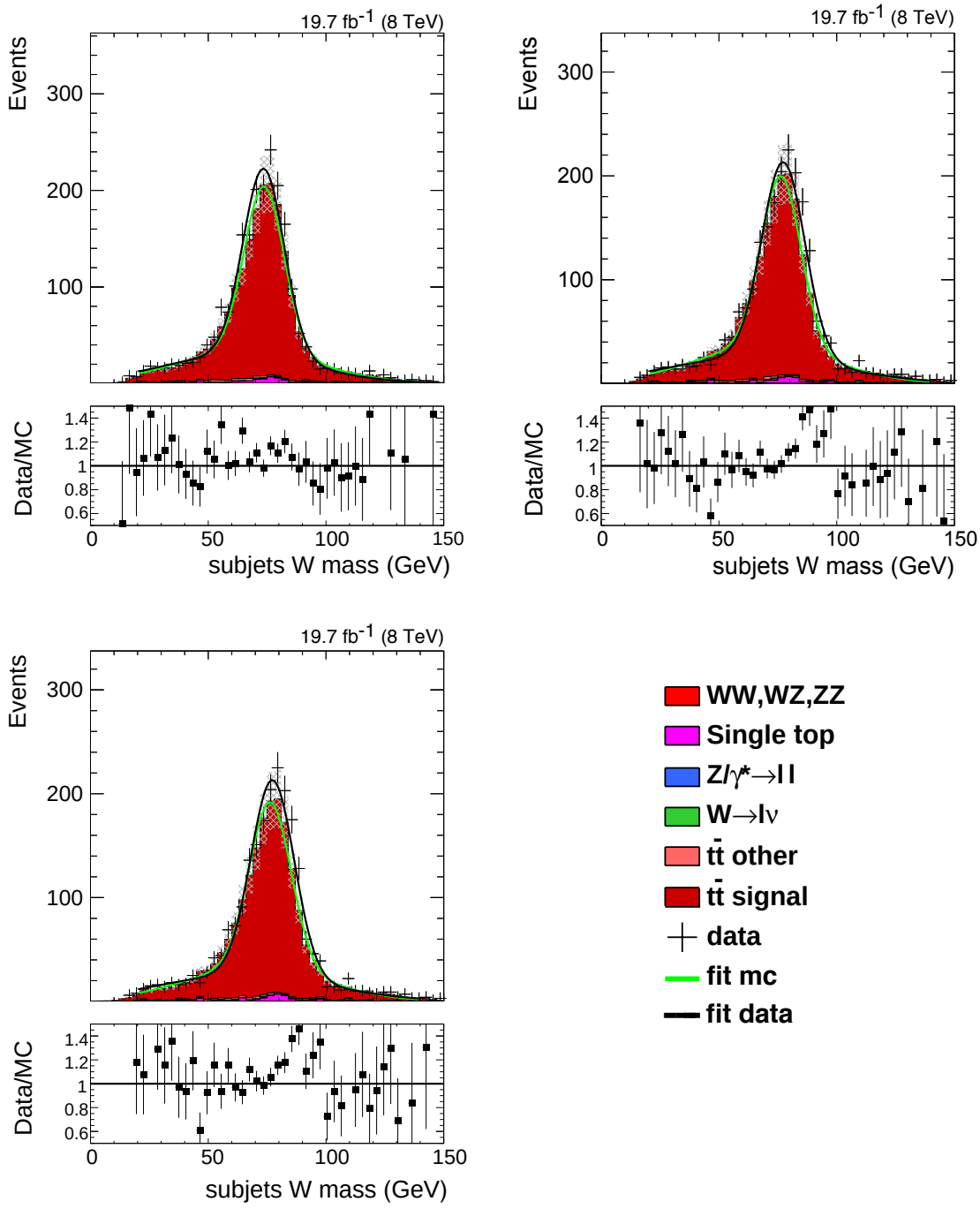

Figure 7.3: Distribution of the W-boson mass obtained from a semi-leptonic $t\bar{t}$ sample, using uncorrected subjets (top left), jet energy scale corrected subjets (top right), and subjets with corrected jet energy scale and resolution (bottom left) [137].

In all three plots, the data are well described by the simulation taking into account the statistical uncertainties. Double Gaussian functions are fitted to the distributions in data and simulation. The result of the fit to data is drawn as a black line, the result obtained from simulation as a green line. The corresponding fit parameters are summarized in table 7.4. In all three setups, the parameters from the fits to data and simulated events are well compatible within their uncertainties. When applying the jet energy corrections, the peak of the distribution is shifted to higher values closer to the actual W-boson mass. Additional application of the correction of the jet energy resolution does not affect the results much though. The compatibility of the results obtained from data and simulation is even slightly impaired by the application of these scale factors. Therefore, no corrections of the jet energy resolution are applied to the subjets of CA15 jets. As the absolute value of the jet energy scale of the subjets is not relevant for this analysis, the jet energy scale corrections are also no applied.

	σ		mean	
	Data	Simulation	Data	Simulation
No corrections	8.72±0.36	8.45±0.12	73.51±0.34	73.69±0.10
With JEC	9.18±0.38	8.48±0.13	77.33±0.35	76.30±0.11
With JEC and JER	9.18±0.38	8.85±0.14	77.33±0.35	76.33±0.11

Table 7.4: Results of a fit to the W-boson mass obtained from a semi-leptonic $t\bar{t}$ sample in data and simulation, for different treatments of the jet energy corrections for subjets.

7.2.3 Subjet b-tagging scale factors

To compensate the differences between data and simulated events observed in the measurement of the subjet b-tagging performance (see section 5.2.6), scale factors are applied to the simulation. One set of scale factors is available for the efficiency, with which jets with real bottom-quark content are identified. For the misidentification rates, two sets of scale factors are provided: one for jets with charm-quark content and another one for light jets.

Instead of applying weights to entire events, the scale factors are used to update the b-tagging status on jet-by-jet basis. If the scale factor SF is smaller than one, the status of a fraction $f = 1 - SF$ of subjets is changed from b-tagged to non-b-tagged in a random manner. In case the scale factor is larger than one, the individual b-tagging efficiencies ϵ_{MC} for all simulated samples are needed for its application. Now, the b-tagging status is changed from "non b tagged" to "b tagged" for a fraction of jets $f = \frac{1-SF}{1-1/\epsilon_{MC}}$. The subjets, for which the b-tagging status is updated, are chosen with help of a random-number generator. To ensure the reproducibility of results, it is important to always use

the same seed to initialize the random-number generation for a certain event. In case of the misidentification rates, the status is updated from "non b tagged" to "b tagged" in an equivalent way.

7.2.4 Top-tagging scale factors

The scale factors for the top-tagging efficiency of the HEPTopTagger algorithm are applied as weights to entire events, in contrast to the jet-by-jet ansatz employed in the application of the subjet b-tagging scale factors. The specifics of the determination of the scale factor derivation for the HEPTopTagger algorithm are given in section 5.2.6. The magnitude of the applied weight depends on the number N_{top} of CA15 jets tagged by the HEPTopTagger algorithm found in a certain event. The scale factor is applied N_{top} times per event.

In both, T'- and top-pair production events, two real top quarks can be found per event. Therefore, a non-negligible misidentification rate is only expected for QCD-multijet events in this analysis. As this background is obtained in a data-driven method, no scale factor for the top-misidentification rate is applied to simulated samples.

7.2.5 Pileup reweighting

In the generation of the simulated events using Monte Carlo event generators, a certain distribution of the number of pileup interactions is assumed. In case, the actual number of pileup interactions in data has not yet been measured, the expected number of pileup interactions can be calculated using the inelastic proton-proton interaction cross section and the luminosity. Here, one assumes that this distribution is a Poisson distribution around the expected number of pileup vertices. Usually, the distribution measured in data differs from that used in the Monte Carlo simulation. The distribution of the number of two-particle interactions of the simulated samples is therefore reweighted to the corresponding distribution obtained from data.

7.3 Event selection

In this search for pair-produced T' quarks, the event selection is optimized for the identification of T'-quark decays to top quarks and Higgs bosons. The top quarks and Higgs bosons are assumed to decay hadronically. Ten particles are produced in this final state: six bottom quarks and four light quarks. Each of these quarks initiates a shower of collimated particles in the detector. These showers are reconstructed as jets by the particle-flow algorithm, as described in section 5.2.6. The radial distance R between the daughter particles depends on the transverse momentum of the decaying resonance. The p_T distributions for generated top quarks and generated Higgs bosons in simulated signal events are shown in figure 7.4 for three samples generated with different hypotheses for the T'-quark mass. Assumption of larger T'-quark mass in the simulation results in harder p_T spectra of the decay products. Overall, the majority of particles in the signal samples considered here have transverse momenta between 200 and 400 GeV. Because of the corresponding rather high Lorentz boost of the decaying top quarks and Higgs bosons, dedicated tools are used to search for their decay products in the substructure of CA15 jets. A description of these methods can be found in section 5.2.5.

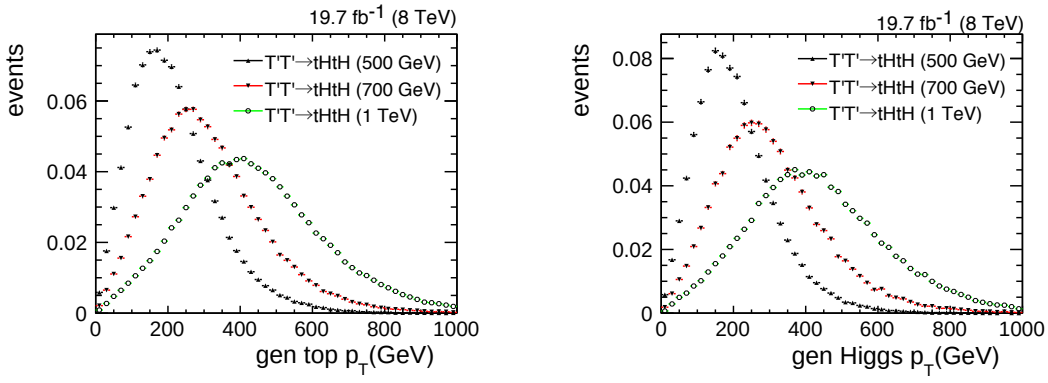

Figure 7.4: Left: transverse momentum of hadronically decaying top quarks in signal events at the generator level. Right: transverse momentum of Higgs bosons decaying to $b\bar{b}$ pairs in signal events at the generator level [137].

7.3.1 Preselection

In this analysis, the event property H_T is defined as the scalar sum of the transverse momenta of the subjets of all CA15 jets with a $p_T > 150$ GeV. Because of the large number of quarks produced in the decays of the T'-quark pairs, H_T is expected to be large in signal events. Consequently, a trigger with a rather high H_T threshold was used in the data acquisition to ensure selection of a large number of signal-like events. The used trigger requires the H_T^{calo} variable to be larger than 750 GeV. H_T^{calo} is defined as the scalar sum of the transverse momenta of all so-called calorimeter jets in the event. Only calorimeter information is used as input to the jet clustering algorithms when clustering calorimeter jets. The efficiency of this trigger in dependence of the H_T calculated from subjets is shown in figure 7.5. It is calculated in dependence of the H_T calculated from

subjets. The distribution is shown for data, and also for simulated $t\bar{t}$ background and signal events. The signal sample used in this plot is produced assuming a T'-quark mass of 700 GeV. The events used to calculate the trigger efficiency in data are recorded using another H_T-based trigger which functions in a similar way as the one used for the actual analysis except that it uses a lower H_T^{calo} threshold of only 650 GeV. This trigger is prescaled, meaning that only a certain fraction of the events actually passing the trigger requirements are stored. Therefore, it is not suited to select a dataset used for complete physics analyses. The denominator for the trigger efficiency curves consists of all events passing the full event selection described in this chapter, except for the trigger and H_T requirements. The events in the numerator satisfy the same selection criteria but also pass the trigger requiring $H_T^{calo} > 750$ GeV.

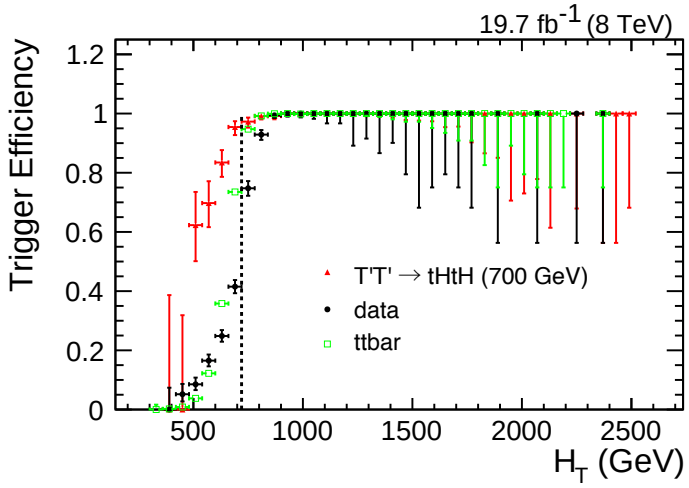

Figure 7.5: Efficiency of the trigger measured as a function of the offline H_T variable. Performance curves are shown for data and simulation, for a $t\bar{t}$ background and a signal sample. The signal events were simulated assuming a T'-quark mass of 700 GeV. The dashed line indicates the event selection criterion $H_T >$ 720 GeV. At this point, the trigger is nearly fully efficient in all samples. The simulated events are reweighted in order to correct for the remaining discrepancies between data and simulation [137].

In order to be able to correctly compare data to simulation in the further stages of the analysis, the trigger of choice should yield the same efficiency in data and simulated events in the region of interest. Therefore, a requirement on the H_T calculated from subjets is introduced to avoid the region of phase space, in which large discrepancies between data and simulation are observed. In this analysis, events are discarded, if their H_T calculated from subjets is smaller than 720 GeV. For selected events with $H_T < 900$ GeV, slight differences between data and simulated events are observed nevertheless. Introduction of a stricter H_T requirement of, e.g., $H_T > 900$ GeV compromises the signal selection efficiency to great extend though. Instead, correction factors are applied, rescaling the simulated signal and $t\bar{t}$ background samples to match the trigger efficiency measured in data. Only the first two bins of the H_T distribution are affected by this reweighting

procedure though. The majority of the selected signal events are found at large values of H_T. Less than 10% of the signal events simulated with a T'-quark mass of 700 GeV are found in these first two bins of the H_T distribution.

7.3.2 Jet-multiplicity selection

After the preselection of events described in the previous section, substructure methods are used to further reduce the background contributions. The scope of the event selection described here is to identify at least two CA15 jets, one of them containing a merged hadronic top-quark decay, the other a $H \to b\bar{b}$ decay. The multiplicity of CA15 jets with a transverse momentum of at least 150 GeV after the preselection is shown in figure 7.6. The contribution of $t\bar{t}$-background events and the QCD-multijet events are drawn as colored histograms. The distribution of these background events can thus be directly compared to those of signal events, that are shown for three T'-quark mass hypotheses of $m(T') = 500$ GeV, $m(T') = 700$ GeV, and $m(T') = 1000$ GeV. Events contained in the first two bins of this distribution are discarded in the event selection.

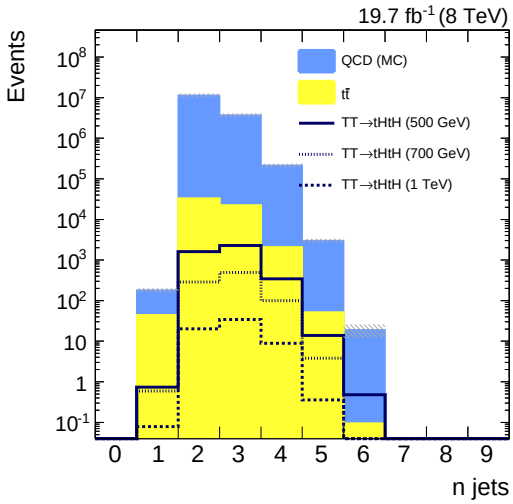

Figure 7.6: Multiplicity of CA15 jets with $p_T > 200$ GeV after the preselection. Events with at least two of these jets are selected. The statistical uncertainty in the two background contributions is depicted by the grey hashed error bands.

Two types of CA15 jets are used in the event selection. The subjets of CA15 jets in the so-called "HEPTopTagger-jet collection" are identified in the procedure described in section 5.2.5.1. In this method, the subjets are chosen in a way that ensures good compatibility of invariant mass of the three reconstructed subjets with the top quark mass. Therefore, jets from this collection are only suited for identification of top quark decays within jets. In this analysis, also Higgs decays are searched for within CA15 jets. For this purpose, jets from another jet collection, the "filtered-jet collection", are used. The two jet collections are superimposable, meaning that each jet in the HEPTopTagger-jet collection can be directly matched to a jet in the filtered-jet collection. In fact, the jets only differ in the composition of their subjets.

7.3.3 Identification of the top-candidate jet

After the jet-multiplicity selection, the HEPTopTagger algorithm is used to analyze the substructure of all CA15 jets in the remaining events. For this purpose, CA15 jets from the HEPTopTagger-jet collection are used. The algorithm is optimized to identify hadronic decays of moderately boosted top quarks within CA15 jets. The multiplicity of jets that have a $p_T > 200$ GeV and are tagged by the HEPTopTagger algorithm is shown in figure 7.7. Events contained in the first bin of the histogram are discarded when asking for at least one jet that is tagged by the HEPTopTagger algorithm to be contained in the event. The number of QCD-multijet background events is very effectively reduced by almost 95% in this selection step.

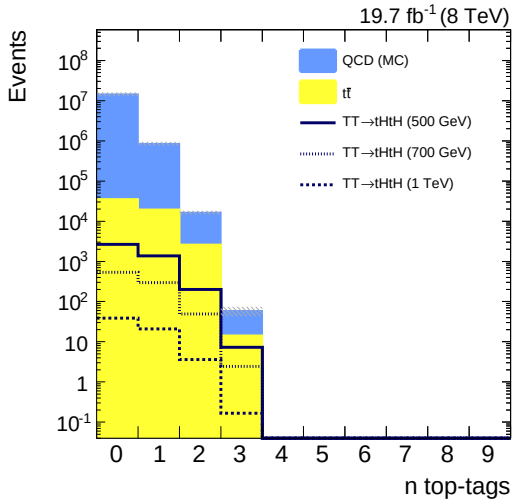

Figure 7.7: Multiplicity of CA15 jets with $p_T > 200$ GeV that are tagged by the HEPTopTagger algorithm in events passing the preselection and containing at least two CA15 jets. Events with at least one of these jets are selected. The statistical uncertainty in the two background contributions is depicted by the grey hashed error bands.

As discussed in section 5.2.5.2, the misidentification rate of the HEPTopTagger algorithm can be dramatically reduced by requiring one of the subjets of top-tagged jets to be tagged by the CSV-b-tagging algorithm at medium working point (CSVM algorithm). In all events passing the previous selection steps, this combination of the HEPTopTagger and CSVM b-tagging algorithm is applied to the CA15 jets. The multiplicity of CA15 jets tagged by this combination of algorithms is displayed in figure 7.8. When rejecting all events found in the first bin of the histogram, the QCD-multijet contribution is once more reduced by a large fraction. Out of the events that are passing the default HEPTopTagger selection despite the fact that they do not contain real top quarks, only 12% are mistakenly selected with this improved top-identification method. All CA15 jets passing the combined requirements of the HEPTopTagger and CSV algorithms are ordered according to their transverse momentum. The jet with the highest p_T is referred to as the "top-candidate jet".

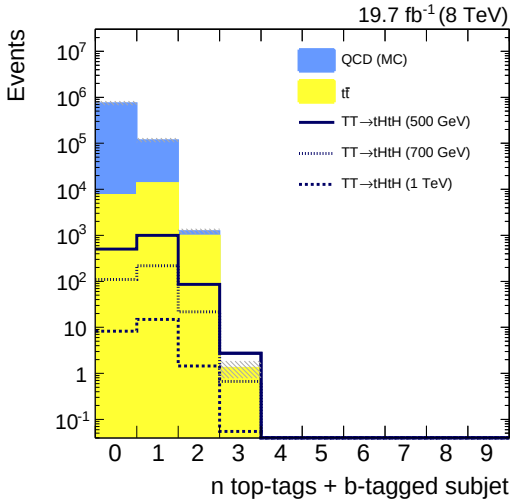

Figure 7.8: Multiplicity of CA15 jets with $p_T > 200$ GeV that are tagged by the HEPTopTagger and contain a subjet that is b tagged by the CSV algorithm at medium working point. Events with at least one of these jets are selected. In this distributions only events passing the preselection and containing at least two CA15 jets are regarded. Furthermore, one of the CA15 jets in all of these events is tagged by the HEPTopTagger algorithm. Events in the first bin contain a top-tagged CA15 jet, but no b-tagged subjet is found within this jet. The statistical uncertainty in the two background contributions is depicted by the grey hashed error bands.

7.3.4 Identification of the Higgs-candidate jet

Finally, a Higgs-tagging technique is used to identify one of the Higgs-boson decays in the signal events. More details on Higgs-tagging algorithms are provided in section 5.2.5. First, the CSVM b-tagging algorithm is applied to all subjets of the CA15 jets in the filtered-jet collection of the remaining events. In case a jet contains two subjets that are tagged by the CSVM algorithm, the invariant mass of the two b-tagged subjets is calculated. If the resulting invariant mass is larger than 60 GeV, the jet is Higgs tagged. Figure 7.9 shows the multiplicity of Higgs-tagged CA15 jets for all events containing a top-candidate jet. The Higgs-tagged CA15 jet with the highest transverse momentum is identified as the "Higgs-candidate jet", unless the corresponding jet from the HEPTopTagger jet collection has already been selected as top-candidate jet in the previous selection step. If there are no Higgs-tagged jets in the event besides the one matched to the top-candidate jet, the event is not selected. Otherwise, the Higgs-tagged jet with the second leading p_T becomes the Higgs-candidate jet.

By rejecting all events containing no Higgs-candidate jet, the QCD-multijet background is further reduced by more than a factor of 200. However, in this step of the event selection the rate of $t\bar{t}$-background events is affected heavily as well. Only 4% of the previously selected $t\bar{t}$-background events pass this last selection step.

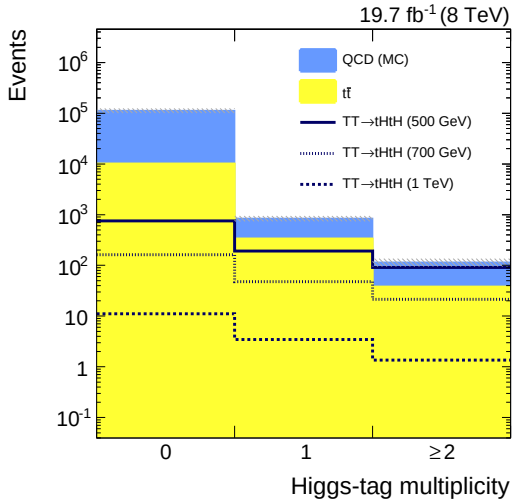

Figure 7.9: Multiplicity of Higgs-tagged CA15 jets with $p_T > 150$ GeV in events that contain a top-candidate jet, pass the preselection, and contain two or more CA15 jets. Events with three or more Higgs tags are rare. They are included in the ≥ 2 Higgs-tag category. The statistical uncertainty in the two background contributions is depicted by the grey hashed error bands.

7.3.5 Definition of event categories

In the following steps of the analysis, the selected events are split into two categories. Events with exactly one Higgs-tagged jet form the "single Higgs-tag category", which corresponds to the second bin in the histogram shown in figure 7.9. The third bin of this histogram contains all events with two or more Higgs-tagged jets. These are the events of the "multi Higgs-tag category". This splitting is introduced in order to exploit the excellent signal to background ratio in the multi Higgs-tag category. Considering, for example, the signal sample generated under the assumption of a T'-quark mass of 700 GeV where each of the T'-quarks decays to a top quark and a Higgs boson, an overall signal to background ratio of 6% is found after the full selection. Splitting the events into the two Higgs-multiplicity categories reduces the signal-to-background ratio slightly to 5% in the single Higgs-tag category, but results in a much improved ratio of 16% in the multi Higgs-tag category. The splitting is therefore expected to have a positive effect on the overall sensitivity of the analysis.

7.3.6 Results of the event selection

The presented event selection reduces the background processes very efficiently. The dominant QCD-multijet background is reduced to about the same level as the background from top-pair production in the jet-substructure selection. The combination of top and Higgs tagging is capable of reducing both of these background processes.

The selection efficiency for all simulated signal samples used in this analysis is shown in figure 7.10 for the inclusive selection (top), the single Higgs-tag category (middle), and also the multi Higgs-tag category (bottom). Here, not only signal samples simulated with decays of T'T'$\rightarrow$ tHtH, but also those assuming other decay modes of the T' quarks are considered. Each bin of the histogram corresponds to a certain T'-quark mass assumption. The selection efficiency for each sample can be read off the y-axis.

The signal selection efficiency is better for higher T'-quark masses. This is due to the larger boost of the top quarks and Higgs bosons in these samples. In the lower-mass region, the decay products of the top quarks and Higgs bosons may not always be merged fully into a single CA15 jet. Thus, the selection efficiencies of the HEPTopTagger and Higgs-tagging algorithms are compromised to a certain extent. In case both T' quarks decay to a top quark and Higgs boson, the best signal selection efficiency is obtained, since this is the final state the event selection is optimized for. The selection does not explicitly require two top-candidate jets and two Higgs-candidate jets though, but only asks for one of each kind. Therefore, good signal efficiency is also achieved for the samples containing only one T' quark decaying into a top quark and a Higgs boson. Decays to top quarks and Z bosons can have signatures very similar to T' $\rightarrow$ tH decays, while T' $\rightarrow$ bW decays result in rather different topologies. Therefore, signal samples with Z bosons in the final state are favored by the event selection with respect to those containing W bosons.

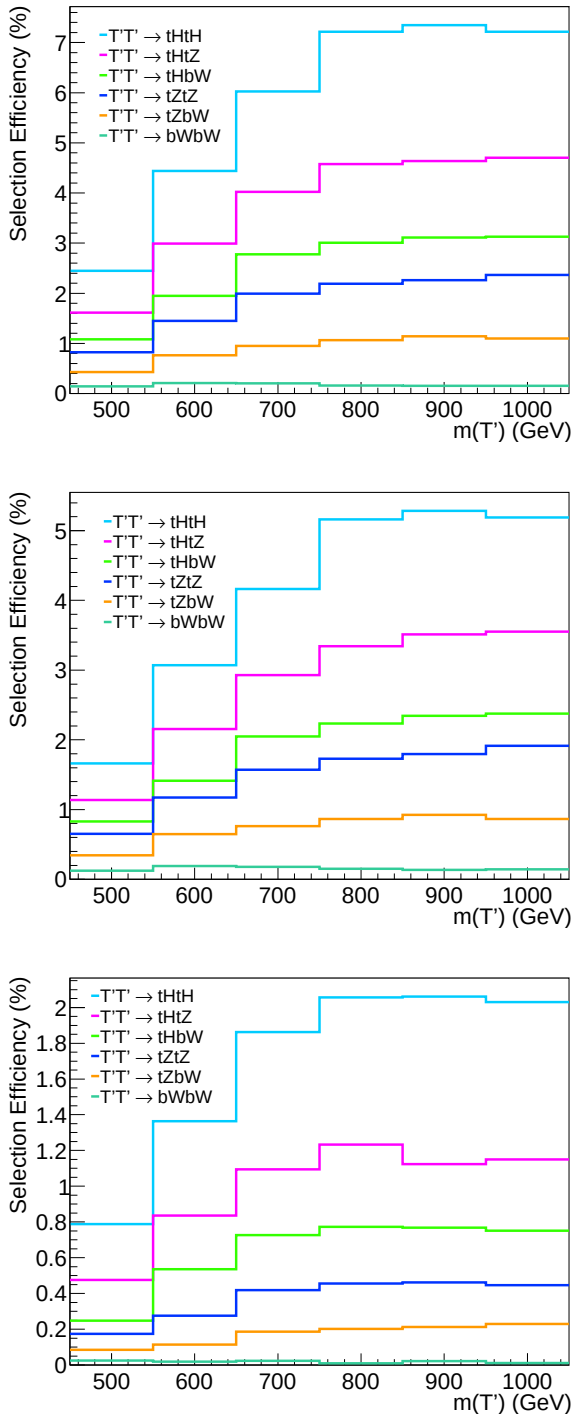

Figure 7.10: Percentual selection efficiencies for six signal samples simulated assuming different decay modes of the T' quark, for the inclusive selection (top), single Higgs-tag category (middle), and multi Higgs-tag category (bottom). The efficiencies are calculated with respect to an inclusive sample without restriction to any decay mode and before application of any selection criteria.

Besides the optimization of the signal-to-background ratio with dedicated event selection criteria, another handle to improve the sensitivity of the analysis is to find variables whose shapes can be used to discriminate signal from background. Two variables well suited for this purpose are identified in this analysis: the H_T variable calculated from subjets as previously explained, and the Higgs-candidate mass, defined as the invariant mass of the two b-tagged subjets in the Higgs-candidate jet. Figure 7.11 shows the distributions of both variables, H_T on the left and the Higgs-candidate mass on the right. The distributions in the top row are obtained from all events passing the event selection, those in the second row from events in the single Higgs-tag category, and those in the bottom row from all events containing two or more Higgs-tagged jets.

The apparent structures and fluctuations in the QCD-multijet contribution are due to the very small number of simulated events that fulfill all event-selection criteria. Very large event weights are applied to some of these events. This leads to artificial structures in the distributions of some variables. A better description for the QCD-multijet background contribution in the very constrained phase space defined by the rather strict event selection is derived from data. The specifics of this method are provided in section 7.4.

A difference in shape between the distributions obtained from background processes and the ones from signal processes is clearly visible. The difference becomes more pronounced as the T'-quark mass used in the simulation of the signal samples increases. In general, signal events tend to have higher H_T than the background events. The Higgs-candidate mass distributions from signal samples feature a peak-like structure slightly below the expected Higgs mass of approximately 125 GeV, which is not visible in the distribution of background. The Higgs-candidate jets in $\mathrm{t\bar{t}}$ and QCD-multijet events are in fact misidentified jets without real Higgs-boson content, as no real Higgs bosons are present in these samples. Therefore, the mass of these jets is not expected to match the Higgs-boson mass of 125 GeV. The H_T and Higgs-candidate mass variables are used in the statistical evaluation of the results described in section 7.6.

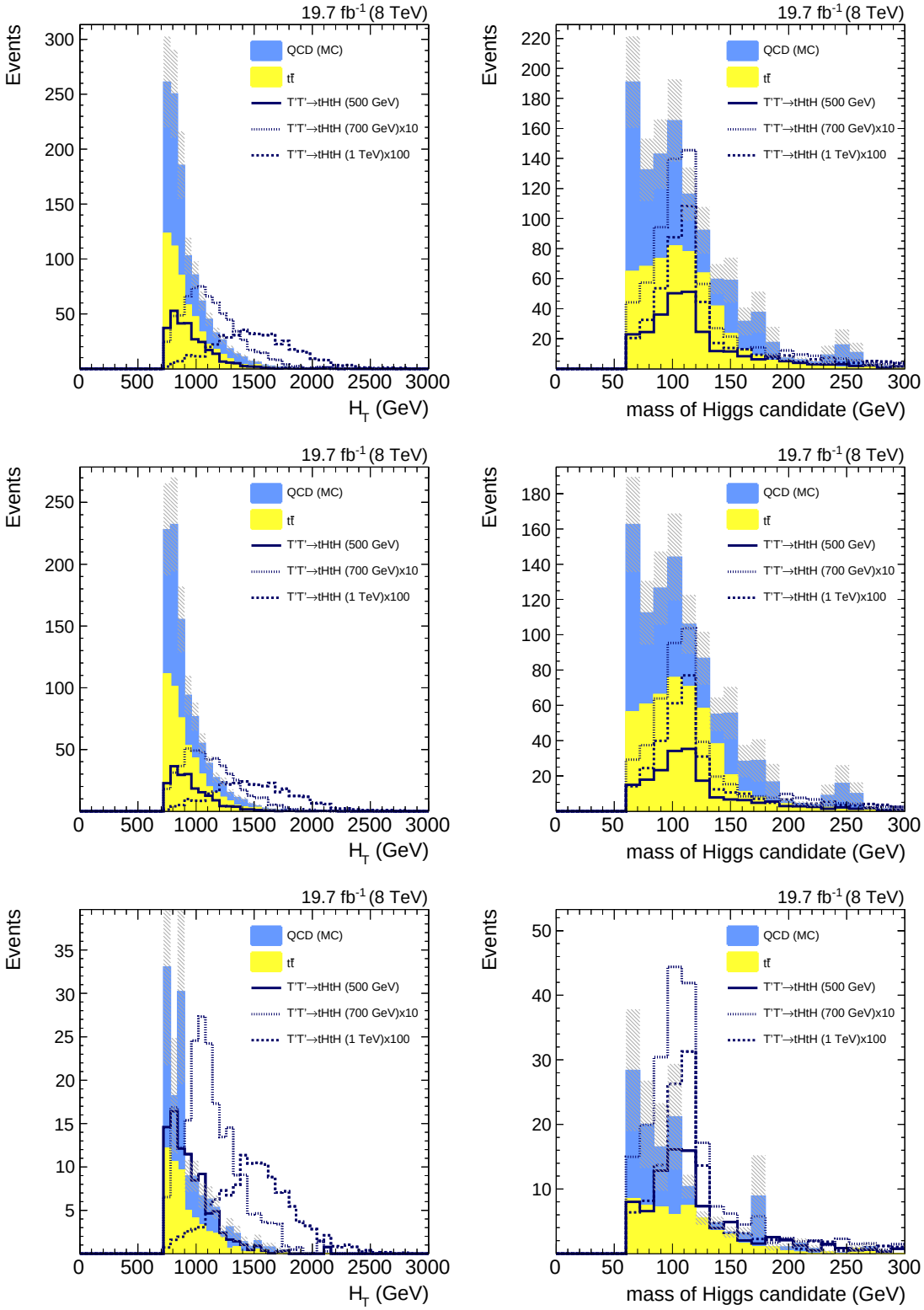

Figure 7.11: Distributions of the H_T (left) and mass of the Higgs-candidate jet (right) variable after the inclusive event selection (top), in the single Higgs-tag category (middle), and the multi Higgs-tag category (bottom). The statistical uncertainty in the two background contributions is depicted by the hashed error bands.

7.4 The QCD-multijet background

Only a very small fraction of the QCD-multijet background events is retained in the event selection described in the previous section. The special properties of events in this small corner of the kinematic phase space can lead to inconsistencies in the modelling of the remaining events. Therefore, the event yield and also the shape of distributions of QCD-multijet background events is not estimated using the samples simulated using Monte Carlo techniques in this analysis. Instead, the number of selected events and the templates describing the shape of the distributions that are used as input for the statistical evaluation of the search results are derived from data.

7.4.1 The ABCD method

The so-called ABCD method is employed to derive a suitable estimate of the QCD-multijet event yield from data. In this method, three exclusive sideband regions A, B, and C are obtained by inverting two uncorrelated requirements in the default event selection. The signal region defined by the regular event selection is referred to as region D in this naming scheme. The definition of the four regions is illustrated in table 7.5: both criteria are applied in the default event selection that is used to obtain the events in signal region D. Inversion of one of the two criteria at a time results in regions B and C. Both criteria are inverted at the same time to select the events in region A.

Region A: Inversion of both criteria	**Region B:** Inversion of criterion 1
Region C: Inversion of criterion 2	**Region D:** Application of both criteria

Table 7.5: Illustration of the four regions in the ABCD method.

If the two criteria chosen for inversion are uncorrelated, the four rates N_i of QCD-multijet events selected in each of the four regions exhibit the following dependence:

$$\frac{N(Region\,A)}{N(Region\,B)} = \frac{N(Region\,C)}{N(Region\,D)}. \tag{7.1}$$

The number of QCD multijet events in the signal region D can then be easily calculated to be

$$N(Region\,D) = \frac{N(Region\,C) \cdot N(Region\,B)}{N(Region\,A)}. \tag{7.2}$$

In order to derive the QCD-multijet event yield in the signal region from data, the event selection needs to be inverted in such a way, that the sideband regions predominantly contain QCD-multijet events and only few other events. The event content of the different regions is checked using simulated samples. If the sideband regions contain mainly simulated QCD-multijet events with only little contamination of other simulated processes, also

the data in each of the sideband regions can be assumed to consist almost exclusively of QCD-multijet events. In order to obtain a data-driven estimate of the number of selected QCD-multijet events in the signal region, a projection into region D can then be made with the event yields in regions A, B, and C measured in data following equation 7.2.

In the application of the ABCD method presented here, two of the jet-substructure selection requirements are inverted: the Higgs-taging and HEPTopTagger conditions.

When inverting the Higgs-tagging selection criterion, all events containing Higgs-tagged CA15 jets are rejected. Here, the definition of a Higgs tag is altered though. Instead of requiring two CSV b tags at medium working point, the algorithm is operated at the loose working point. Thus, an even larger number of events is excluded from the regions obtained by inversion of the Higgs tag. Events in which the top-candidate jet is simultaneously selected by the CSVL Higgs-tagging algorithm are not discarded. As the events are not part of the signal region, where top-candidate jet and Higgs-candidate jet cannot be identical, they may be contained in the sideband region without violating the required exclusiveness of the four regions.

The inversion of the HEPTopTagger algorithm is implemented in a slightly more complicated way. Instead of simply asking for events with no jets tagged by the HEPTopTagger algorithm, certain parts of the algorithm itself are inverted, defining a new anti-HEPTopTagger algorithm. In the anti-HEPTopTagger algorithm, the ratio m_{23}/m_{123} must be smaller than 0.35 and the jet mass must not fall into the top mass window, i.e., it is required that $m_{CA15} < 140$ GeV or $m_{CA15} > 250$ GeV. All other configurations remain the same as described in section 5.2.5.1. When asking for at least one jet that is tagged by the anti-HEPTopTagger algorithm, the selected events are kinematically similar to those in the signal region. The first jet passing the criteria of the anti-HEPTopTagger algorithm is identified as the anti-top-candidate jet. In case the Higgs-tagging criterion is not inverted simultaneously, the jet may in addition pass the requirements of the Higgs-tagging algorithm. The anti-top-candidate jet must not be identical with the Higgs-candidate jet though. In addition to these criteria, no CA15 jet tagged by the regular HEPTopTagger algorithm may be found in the events.

The number of data and simulated events selected in each of the three sideband regions are provided in table 7.6 for the inclusive event selection, table 7.7 for events in the single Higgs-tag category, and table 7.8 for the multi Higgs-tag category.

In all event categories, the expected number of QCD-multijet events in the sideband regions clearly exceeds the expected contributions from other processes for all three sideband regions A, B, and C. The contributions from different signal processes and the $t\bar{t}H$, $Wb\bar{b}$ and $Zb\bar{b}$ backgrounds to the overall expected event yield are at percent level or lower and can be neglected. Dedicated studies show that the slight signal contamination has no significant impact on the results of this ABCD method. These studies are documented in section 7.4.4. There are sizeable contributions from $t\bar{t}$-background events in all three sideband regions though. As the $t\bar{t}$ background is modelled very accurately in Monte Carlo simulations, the expected contribution according to the simulation is subtracted from the data in order to obtain a correct estimate for the number of QCD-multijet events.

	Inverted Higgs tag		**Regular Higgs tag**	
	Region A		**Region B**	
Inverted HEPTopTag	Data	1152640	Data	9541
	QCD-multijet	1078720 ± 2258	QCD-multijet	6592 ± 162
	$t\bar{t}$	6176 ± 37	$t\bar{t}$	328.4 ± 9
	$t\bar{t}H$	12± 0.2	$t\bar{t}H$	5 ± 0.1
	$Wb\bar{b}$	28± 9	$Wb\bar{b}$	2 ± 2
	$Zb\bar{b}$	9± 7	$Zb\bar{b}$	7 ± 5
	Signal T'T' → tHtH		Signal T'T' → tHtH	
	m(T') = 500 GeV	331 ± 8	m(T') = 500 GeV	172 ± 6
	m(T') = 700 GeV	85 ± 2	m(T') = 700 GeV	47 ± 1
	m(T') = 1000 GeV	7 ± 0.1	m(T') = 1000 GeV	3 ± 0.1
	Signal T'T' → tZtZ		Signal T'T' → tZtZ	
	m(T') = 500 GeV	367 ±7	m(T') = 500 GeV	54±3
	m(T') = 700 GeV	88 ± 1	m(T') = 700 GeV	14± 0.4
	m(T') = 1000 GeV	7±0.1	m(T') = 1000 GeV	1 ±0.03
	Signal T'T' → bWbW		Signal T'T' → bWbW	
	m(T') = 500 GeV	1442 ±14	m(T') = 500 GeV	72 ±3
	m(T') = 700 GeV	225± 2	m(T') = 700 GeV	13 ±0.4
	m(T') = 1000 GeV	13± 0.1	m(T') = 1000 GeV	1±0.02
	Region C		**Region D**	
Regular HEPTopTag	Data	140911	Data	1560
	QCD-multijet	92605 ± 646	QCD-multijet	577 ± 48
	$t\bar{t}$	10939 ± 45	$t\bar{t}$	541 ± 9
	$t\bar{t}H$	19 ± 0.2	$t\bar{t}H$	10.4 ± 0.2
	$Wb\bar{b}$	-	$Wb\bar{b}$	-
	$Zb\bar{b}$	-	$Zb\bar{b}$	-
	Signal T'T' → tHtH		Signal T'T' → tHtH	
	m(T') = 500 GeV	468 ± 10	m(T') = 500 GeV	283 ± 8
	m(T') = 700 GeV	97 ± 2	m(T') = 700 GeV	69 ± 2
	m(T') = 1000 GeV	7 ± 0.1	m(T') = 1000 GeV	5 ± 0.1
	Signal T'T' → tZtZ		Signal T'T' → tZtZ	
	m(T') = 500 GeV	531 ±8	m(T') = 500 GeV	95± 4
	m(T') = 700 GeV	112±1	m(T') = 700 GeV	23±1
	m(T') = 1000 GeV	8± 0.1	m(T') = 1000 GeV	2 ± 0.04
	Signal T'T' → bWbW		Signal T'T' → bWbW	
	m(T') = 500 GeV	274±6	m(T') = 500 GeV	17 ±1
	m(T') = 700 GeV	35±1	m(T') = 700 GeV	2± 0.2
	m(T') = 1000 GeV	2± 0.04	m(T') = 1000 GeV	0.1± 0.01

Table 7.6: Results of the event selection in the four regions ABCD, where region D is the signal region of the analysis. All events before splitting into Higgs-tag multiplicity categories are shown here. The numbers are obtained from the recorded data and the simulated samples described in chapter 7.2.

	Inverted Higgs tag		**Regular Higgs tag**	
	Region A		**Region B**	
Inverted HEPTopTag	Data	1152640	Data	8384
	QCD-multijet	1078720 ± 2258	QCD-multijet	5829 ± 154
	$t\bar{t}$	6176 ± 37	$t\bar{t}$	295 ± 8
	$t\bar{t}H$	12 ±0.2	$t\bar{t}H$	4 ± 0.1
	$Wb\bar{b}$	28± 9	$Wb\bar{b}$	-
	$Zb\bar{b}$	9± 7	$Zb\bar{b}$	4 ± 4
	Signal T'T' → tHtH		Signal T'T' → tHtH	
	m(T') = 500 GeV	331 ± 8	m(T') = 500 GeV	119 ± 5
	m(T') = 700 GeV	85 ± 2	m(T') = 700 GeV	32 ± 1
	m(T') = 1000 GeV	7 ± 0.1	m(T') = 1000 GeV	2 ± 0.1
	Signal T'T' → tZtZ		Signal T'T' → tZtZ	
	m(T') = 500 GeV	368±7	m(T') = 500 GeV	41 ±2
	m(T') = 700 GeV	88±1	m(T') = 700 GeV	11 ±0.4
	m(T') = 1000 GeV	7±0.1	m(T') = 1000 GeV	1±0.03
	Signal T'T' → bWbW		Signal T'T' → bWbW	
	m(T') = 500 GeV	1442 ±14	m(T') = 500 GeV	65 ±3
	m(T') = 700 GeV	225±2	m(T') = 700 GeV	12 ± 0.4
	m(T') = 1000 GeV	13±0.1	m(T') = 1000 GeV	1±0.02
	Region C		**Region D**	
Regular HEPTopTag	Data	140911	Data	1355
	QCD-multijet	92605 ± 646	QCD-multijet	500 ± 45
	$t\bar{t}$	10939 ± 45	$t\bar{t}$	486 ± 8
	$t\bar{t}H$	19 ± 0.2	$t\bar{t}H$	9 ± 0.1
	$Wb\bar{b}$	-	$Wb\bar{b}$	-
	$Zb\bar{b}$	-	$Zb\bar{b}$	-
	Signal T'T' → tHtH		Signal T'T' → tHtH	
	m(T') = 500 GeV	468 ± 10	m(T') = 500 GeV	192 ± 7
	m(T') = 700 GeV	97 ± 2	m(T') = 700 GeV	48 ± 1
	m(T') = 1000 GeV	7 ± 0.1	m(T') = 1000 GeV	3 ± 0.1
	Signal T'T' → tZtZ		Signal T'T' → tZtZ	
	m(T') = 500 GeV	531 ±8	m(T') = 500 GeV	75±3
	m(T') = 700 GeV	112±1	m(T') = 700 GeV	18± 0.5
	m(T') = 1000 GeV	8±0.1	m(T') = 1000 GeV	1 ±0.03
	Signal T'T' → bWbW		Signal T'T' → bWbW	
	m(T') = 500 GeV	274±6	m(T') = 500 GeV	14 ±1
	m(T') = 700 GeV	35 ± 1	m(T') = 700 GeV	2± 0.2
	m(T') = 1000 GeV	2 ±0.04	m(T') = 1000 GeV	0.1 ±0.01

Table 7.7: Results of the event selection in the four regions ABCD, where region D is the signal region of the analysis. All events falling into the single Higgs-tag category are shown here. The numbers are obtained from the recorded data and the simulated samples described in chapter 7.2.

	Inverted Higgs tag		**Regular Higgs tag**	
	Region A		**Region B**	
Inverted HEPTopTag	Data	1152640	Data	1157
	QCD-multijet	1078720 ± 2258	QCD-multijet	761 ± 52
	$t\bar{t}$	6176 ± 37	$t\bar{t}$	34 ± 3
	$t\bar{t}H$	12 ±0.2	$t\bar{t}H$	1±0.1
	$Wb\bar{b}$	28± 9	$Wb\bar{b}$	2 ± 2
	$Zb\bar{b}$	9± 7	$Zb\bar{b}$	4 ± 4
	Signal T'T' → tHtH		Signal T'T' → tHtH	
	m(T') = 500 GeV	331 ± 8	m(T') = 500 GeV	54 ± 4
	m(T') = 700 GeV	85 ± 2	m(T') = 700 GeV	15 ± 1
	m(T') = 1000 GeV	7 ± 0.1	m(T') = 1000 GeV	1 ± 0.04
	Signal T'T' → tZtZ		Signal T'T' → tZtZ	
	m(T') = 500 GeV	368 ± 7	m(T') = 500 GeV	13 ±1
	m(T') = 700 GeV	88±1	m(T') = 700 GeV	3 ±0.2
	m(T') = 1000 GeV	7± 0.1	m(T') = 1000 GeV	0.2±0.01
	Signal T'T' → bWbW		Signal T'T' → bWbW	
	m(T') = 500 GeV	1442 ±14	m(T') = 500 GeV	6±1
	m(T') = 700 GeV	225 ± 2	m(T') = 700 GeV	1±0.2
	m(T') = 1000 GeV	13± 0.1	m(T') = 1000 GeV	0.1±0.01
	Region C		**Region D**	
Regular HEPTopTag	Data	140911	Data	205
	QCD-multijet	92605 ± 646	QCD-multijet	77 ± 16
	$t\bar{t}$	10939 ± 45	$t\bar{t}$	55 ± 3
	$t\bar{t}H$	19 ± 0.2	$t\bar{t}H$	2 ± 0.1
	$Wb\bar{b}$	-	$Wb\bar{b}$	-
	$Zb\bar{b}$	-	$Zb\bar{b}$	-
	Signal T'T' → tHtH		Signal T'T' → tHtH	
	m(T') = 500 GeV	468 ± 10	m(T') = 500 GeV	91 ± 5
	m(T') = 700 GeV	97 ± 2	m(T') = 700 GeV	21 ± 1
	m(T') = 1000 GeV	7 ± 0.1	m(T') = 1000 GeV	1 ± 0.05
	Signal T'T' → tZtZ		Signal T'T' → tZtZ	
	m(T') = 500 GeV	531±8	m(T') = 500 GeV	20 ±2
	m(T') = 700 GeV	112±1	m(T') = 700 GeV	5 ±0.3
	m(T') = 1000 GeV	8±0.1	m(T') = 1000 GeV	0.3 ±0.02
	Signal T'T' → bWbW		Signal T'T' → bWbW	
	m(T') = 500 GeV	274±6	m(T') = 500 GeV	3±1
	m(T') = 700 GeV	35±1	m(T') = 700 GeV	0.3 ±0.1
	m(T') = 1000 GeV	2± 0.04	m(T') = 1000 GeV	0.01 ± 0.002

Table 7.8: Results of the event selection in the four regions ABCD, where region D is the signal region of the analysis. All events falling into the multi Higgs-tag category are shown here. The numbers are obtained from the recorded data and the simulated samples described in chapter 7.2.

7.4.2 Validation of the ABCD method

In the first step of the validation of the ABCD method employed here, the ratios A/B and C/D are calculated for the expected number of QCD-multijet events in each region. These numbers are obtained from Monte Carlo simulation. The ratios can be found in table 7.9. In all event categories the values for the two ratios A/B and C/D are well compatible within uncertainties. This verifies, that the selection criteria that were chosen for inversion are indeed uncorrelated.

	Inclusive selection	Single Higgs-tag category	Multi Higgs-tag category
A/B	164 ± 4	185 ± 5	1417 ± 97
C/D	160 ± 13	185 ± 17	1203 ± 250

Table 7.9: Ratios of expected QCD-multijet events from simulation. The quoted uncertainties are purely statistical.

If inversion of the chosen selection criteria does not much alter the kinematic behavior of the selected events, also the shapes of the distributions in the sideband regions are expected to be similar to those in the signal region. Figure 7.12 shows shape consistency checks between regions B and D for the H_T (left) and Higgs-candidate mass (right) distributions in simulated QCD-multijet events after the inclusive selection (top), and also in the single (middle) and multi (bottom) Higgs-tag categories.

Good agreement is found in all variables in the shape comparison between sideband region B and the signal region D. A χ^2 compatibility test of the two distributions, as described in section 6.4, is conducted in each event category and the resulting χ^2/DoF values and p values are summarized in table 7.10. They also indicate good shape agreement.

	H_T		Higgs-candidate mass	
	χ^2/DoF	p value	χ^2/DoF	p value
Inclusive selection	0.6	0.86	1.3	0.18
Single Higgs-tag category	0.8	0.71	0.9	0.56
Multi Higgs-tag category	0.9	0.49	0.4	0.82

Table 7.10: χ^2 per degree of freedom and corresponding p value for the shape comparison between the normalized distributions in signal region D and sideband region B.

In addition, the shape agreement between sideband regions A and C is tested, to check how inversion of the top-tagging criterion affects the shape of the distributions used in the further analysis. This check is only possible for the inclusive selection, as none of the events selected in sideband regions A and C contain Higgs-candidate jets by definition. For the same reason, only the distribution of the H_T variable can be checked. This comparison is shown in figure 7.13. No significant deviations are observed.

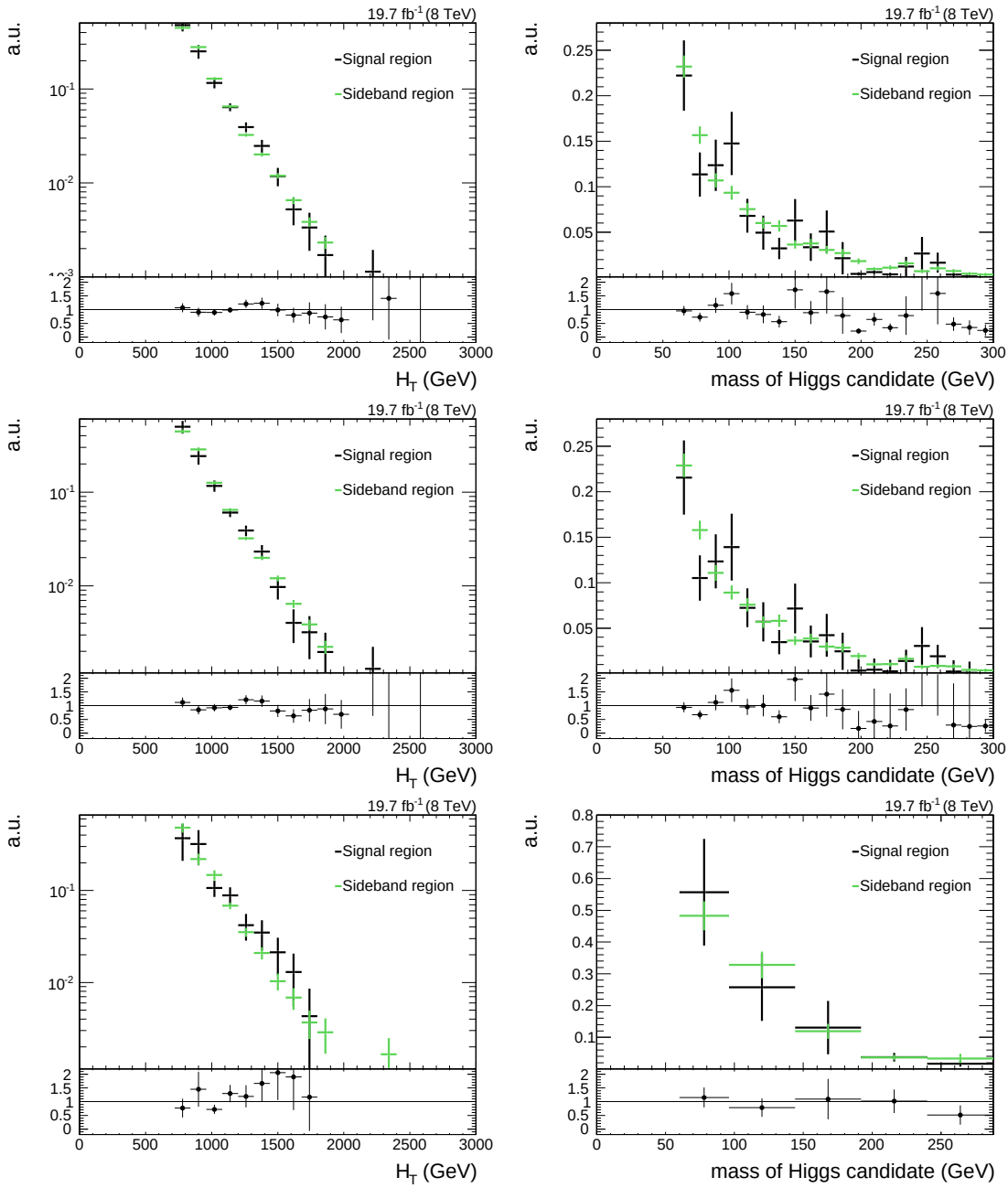

Figure 7.12: Shape compatibility tests between simulated QCD-multijet events in sideband region B (green) and the signal region D (black) for the H_T (left) and the Higgs-candidate mass (right) distributions after the inclusive selection (top), in the single Higgs-tag category (middle), and the multi Higgs-tag category (bottom). The distributions have been normalized to unity for this shape comparison. The ratio of the two histograms is displayed in the bottom part of each plot.

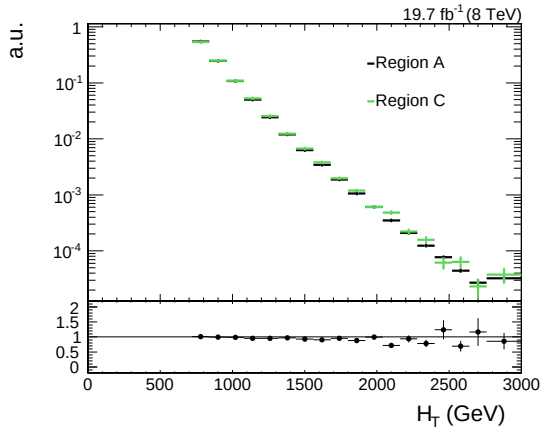

Figure 7.13: Shape-compatibility tests between simulated QCD-multijet events in sideband regions A and C for the H_T distribution after the inclusive selections. Both distributions have been normalized to unity for this shape comparison. The ratio of the two histograms is displayed in the bottom part of the plot.

Similar shape comparisons between the distributions in regions A and B, or regions C and D, reveal considerable shape discrepancies. However, this behavior is well understood. The QCD-multijet background contributions in the regions B and D are selected with a Higgs-tagging criterion. They consist mainly of events in which gluons are split into $b\bar{b}$ pairs. These kind of events are rejected by the inversion of the Higgs-tagging criterion, leading to changes in the event kinematics. This is not problematic with respect to the validity of the ABCD method though, as the transition from region D to region C introduces similar changes to the distributions as a transition from region B to region A. The good shape agreement between the H_T distributions obtained from events in region A and region C as seen in figure 7.13 is a confirmation of this behavior.

7.4.3 Deriving the model for QCD-multijet background from data

After validation of the ABCD method in the chosen setup, the actual model for QCD-multijet events in the signal region is derived from data. Before calculating the event rate for the QCD-multijet model from data, the expected contamination of $t\bar{t}$ events in each of the sideband regions is subtracted from the distributions measured in data. With the resulting estimates of the QCD-multijet contributions in the three sideband regions, the projection into the signal region D is made following equation 7.2. The number of selected data and expected $t\bar{t}$ events in the three sideband regions A, B, and C, their respective difference, as well as the projection into the signal region D are summarized in table 7.11 for the inclusive selection, and also in the single and multi Higgs-tag categories. The quoted uncertainties in the predicted event rate in the signal region are purely statistical. In this method, the difference of selected data and simulated $t\bar{t}$ events gives the number of expected QCD-multijet events in each of the sideband regions A, B, and C. The statistical uncertainty in these numbers is propagated in the calculation of the expected number of QCD-multijet events in the signal region D according to equation 7.2.

The shapes of the QCD-multijet distributions for the variables that are used as input

Inclusive selection			
Region A		Region B	
Data	1152640	Data	9541
$t\bar{t}$	6176	$t\bar{t}$	328
Data $-t\bar{t}$	1146464	Data $-t\bar{t}$	9213
Region C		Region D	
Data	140911		
$t\bar{t}$	10939	Prediction	1044 ± 11
Data $-t\bar{t}$	129972		
Single Higgs-tag category			
Region A		Region B	
Data	1152640	Data	8384
$t\bar{t}$	6176	$t\bar{t}$	294.7
Data $-t\bar{t}$	1146464	Data $-t\bar{t}$	8089.3
Region C		Region D	
Data	140911		
$t\bar{t}$	10938.7	Prediction	917 ± 11
Data $-t\bar{t}$	129972.3		
Multi Higgs-tag category			
Region A		Region B	
Data	1152640	Data	1157
$t\bar{t}$	6176.0	$t\bar{t}$	33.7
Data $-t\bar{t}$	1146464.0	Data $-t\bar{t}$	1123.3
Region C		Region D	
Data	140911		
$t\bar{t}$	10938.7	Prediction	127 ± 4
Data $-t\bar{t}$	129972.3		

Table 7.11: Measured and predicted event rates in the sideband and signal regions, respectively, as obtained in the ABCD method. The expected $t\bar{t}$ contamination is subtracted from the nominal yield in the sidebands. The prediction based on equation 7.2 is given in the lower right quadrant for each event category. The quoted uncertainty in the prediction is purely statistical due to the sample sizes in the sideband regions.

to the statistical analysis, in particular H_T and the Higgs-candidate mass, need to be modelled as well. Since the shape compatibility tests in figure 7.12 show good shape agreement between the distributions in sideband region B and the signal region D for the simulated samples, the distributions of data in sideband region B, after subtraction of the $t\bar{t}$ contamination, are chosen to model the shapes of the QCD-multijet distributions in the signal region. Figure 7.14 compares the shapes of the thus obtained templates for the QCD-multijet distributions to the distributions of QCD-multijet events from Monte Carlo simulation in the signal region.

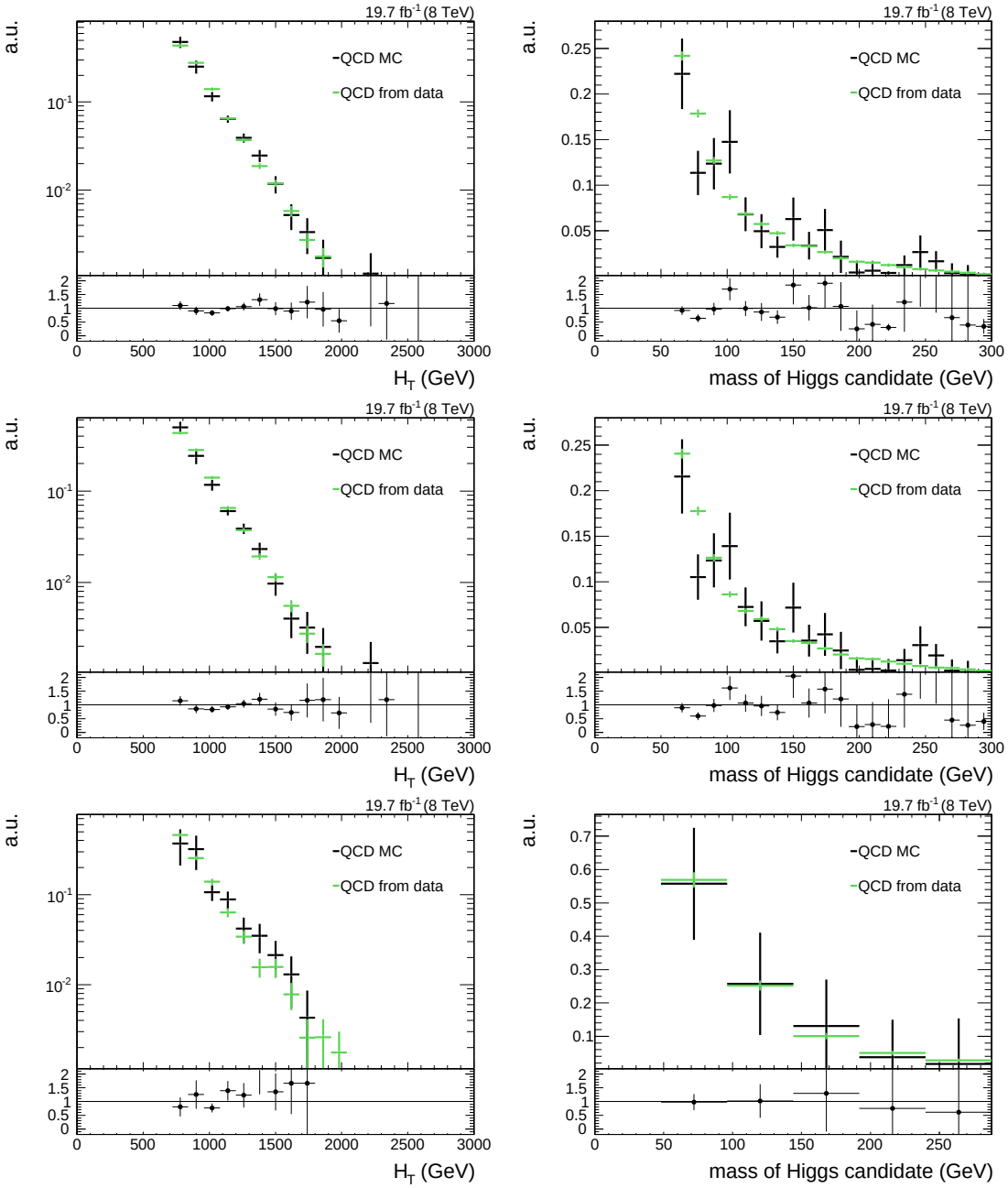

Figure 7.14: Comparison between simulated QCD-multijet events (black) in the signal region D and the QCD model derived from data (green) for the H_T (left) and the Higgs-candidate mass (right) distributions after the inclusive selection (top), in the single Higgs-tag category (middle), and the multi Higgs-tag category (bottom). The distributions have been normalized to unity for this shape comparison. The ratio of the two histograms is displayed in the bottom part of each plot.

Overall, no significant shape differences are observed between the model derived from data and Monte Carlo prediction. However, the statistical uncertainties are much improved when using the model from data rather than the prediction from Monte Carlo simulation. Furthermore, the prediction for the event rate is changed significantly: the number of QCD-multijet events in the signal region after the inclusive selection is predicted to be 579 in Monte Carlo simulation. This rate is approximately doubled to 1044 when estimated via the ABCD method. In the single Higgs-tag category, the number of events is increased from 502 to 917, in the multi Higgs-tag category from 77 to 127. This demonstrates once more the need for a data-driven estimation of the QCD-multijet background contribution.

7.4.4 Signal contamination study

In the QCD-multijet enriched sideband regions A, B, and C, a slight contamination from signal events is predicted by simulation. In order to estimate the effect this signal contamination has on the QCD-multijet model from data, not only the expected $t\bar{t}$ contribution in the sidebands, but also that of simulated signal events are subtracted from the data in this study. Here, signal samples simulated assuming a branching fraction Br(T' $\rightarrow$ tH) = 100% were used, as the largest signal impact is expected for contamination of this type of events. The effect on the event rate in the signal region is documented in table 7.12.

Inclusive selection	
Mass point of injected signal	Impact
$m_{T'} = 500$ GeV	2.2%
$m_{T'} = 700$ GeV	0.5%
$m_{T'} = 1000$ GeV	0.03%
Single Higgs-tag category	
Mass point of injected signal	Impact
$m_{T'} = 500$ GeV	1.8%
$m_{T'} = 700$ GeV	0.5%
$m_{T'} = 1000$ GeV	0.02%
Multi Higgs-tag category	
Mass point of injected signal	Impact
$m_{T'} = 500$ GeV	5.2%
$m_{T'} = 700$ GeV	1.4%
$m_{T'} = 1000$ GeV	0.1%

Table 7.12: Impact of signal contamination in the sideband regions on the prediction of the QCD-multijet event rate after the inclusive selection, and also in the single and multi Higgs-tag categories. Signal samples modelled with a branching fraction of Br(T'$\rightarrow$ tH) = 100% are used for this study.

The prediction for the number of QCD-multijet events in the signal region is affected only slightly. A hypothetical contamination by signal events containing T' quarks with a mass of 500 GeV changes the event rate by about 2% in the inclusive and single Higgs-tag categories, and by about 5% in the multi Higgs-tag category. Because of the smaller production cross sections for signal samples simulated with higher T'-quark masses, the

effect is even smaller for these samples. The impact is below percent level in these cases.

Overall, the impact of the signal contamination is negligible. The impact of a hypothetical contamination by signal events with a T' quark mass of 500 GeV on the shape template for the QCD-multijet model in the multi Higgs-tag category is illustrated in figure 7.15.

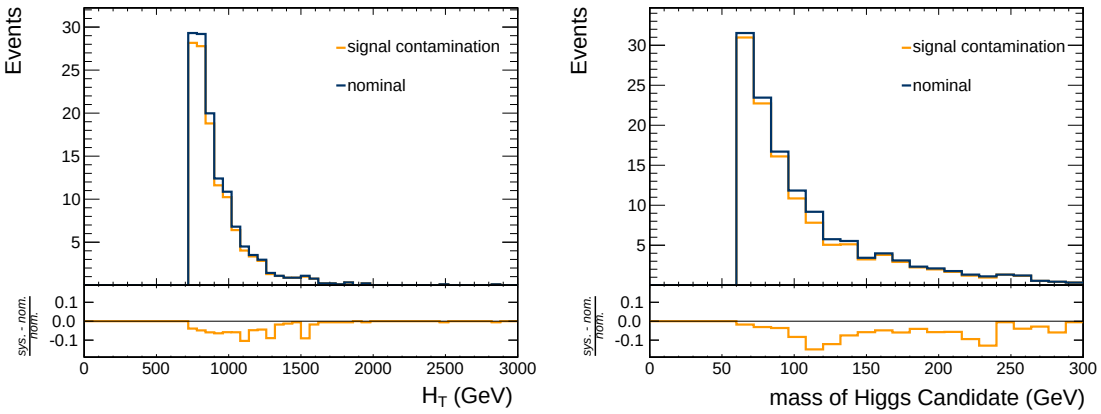

Figure 7.15: Impact of a possible signal contamination in the sideband regions on the shape of the H_T (left) and Higgs-candidate mass (right) distributions in the multi Higgs-tag category for the QCD-multijet contribution derived from data. A signal sample simulated with a T' quark mass of 500 GeV is used in this study.

7.5 Systematic uncertainties

Besides the statistical uncertainties in the data and the simulated samples, also systematic uncertainties have an impact on the sensitivity of the measurement. In this section, the different sources of systematic uncertainties and their impact on the analysis are described. A summary of the effect of systematic uncertainties on the number of selected events can be found towards the end of this section in table 7.15. In general, uncertainties can be divided into two categories: those affecting only the number of selected events in the analysis, and those changing also the shape of certain distributions.

7.5.1 Luminosity

The cluster counting method is employed for luminosity measurements in CMS. A description of this method can be found in section 3.2.6. The estimated uncertainty in the luminosity measurement is 2.5% (*syst.*) + 0.5% (*stat.*) = 2.6%(*total*) [70]. Since all samples that were generated using Monte Carlo techniques are scaled to match the measured luminosity for data, this uncertainty is taken into account for all simulated samples.

7.5.2 Cross sections

The $t\bar{t}$ background events and signal events are simulated using cross sections calculated at leading order, but afterwards scaled to next-to-next-to-leading order cross sections. An uncertainty of 13% is assigned to the used $t\bar{t}$ cross section [134]. The uncertainties in the signal cross sections do not need to be provided as the cross section of these samples are not considered in the fitting procedure. The QCD-multijet background is derived from data as described in section 7.4. The uncertainty in the theoretical prediction of the QCD-multijet production cross section is therefore not relevant for this analysis.

7.5.3 Parton distribution function

The parton distribution functions (PDFs) used in the Monte Carlo simulation of $t\bar{t}$ and signal events for this analysis are provided by the CTEQ group [76]. In the determination of these PDFs in an n-dimensional fit an experimental uncertainty arises. It needs to be taken into account as a systematic uncertainty in the number of selected events and shape of distributions obtained from the Monte Carlo simulation. In order to do so, the assumption is made that an expansion of the χ^2 goodness-of-fit distribution around the global minima a_i^0 of the n fit parameters a_i of the form

$$\Delta\chi^2 = \chi^2 - \chi^2_{min} = \sum_{i,j}^{n} H_{ij}(a_i - a_i^0) \cdot (a_j - a_j^0), \tag{7.3}$$

with the Hessian matrix $H_{ij} = \frac{\partial^2\chi}{2\partial a_i \partial a_j}$ is valid [138]. This Hessian matrix has a set of n independent eigenvectors and eigenvalues. $2n$ so-called error PDFs result from upward and downward variation of each of the n fit parameters along the direction of the eigenvectors. These error PDFs are then used to reweight the simulated samples. In each bin of the distributions obtained from simulation, a shift is introduced by each of the individual eigenvectors. All variations per bin are added in quadrature resulting in the actual uncertainty induced by experimental uncertainty in the PDF measurement.

Here, the CTEQ6 and CTEQ10 [76] set of eigenvectors are used in the reweighting procedure for signal and $t\bar{t}$ events respectively. The resulting uncertainties in the expected number of selected events in the $t\bar{t}$ and signal samples range between 3% and 8%. Figure 7.16 illustrates how the shape of the H_T and Higgs-candidate mass distributions are affected by this source of systematic uncertainty. Overall, the shape of the distributions is rather robust against changes in the PDFs. For signal, only the effect on the signal acceptance but not on the calculated cross section is taken into account.

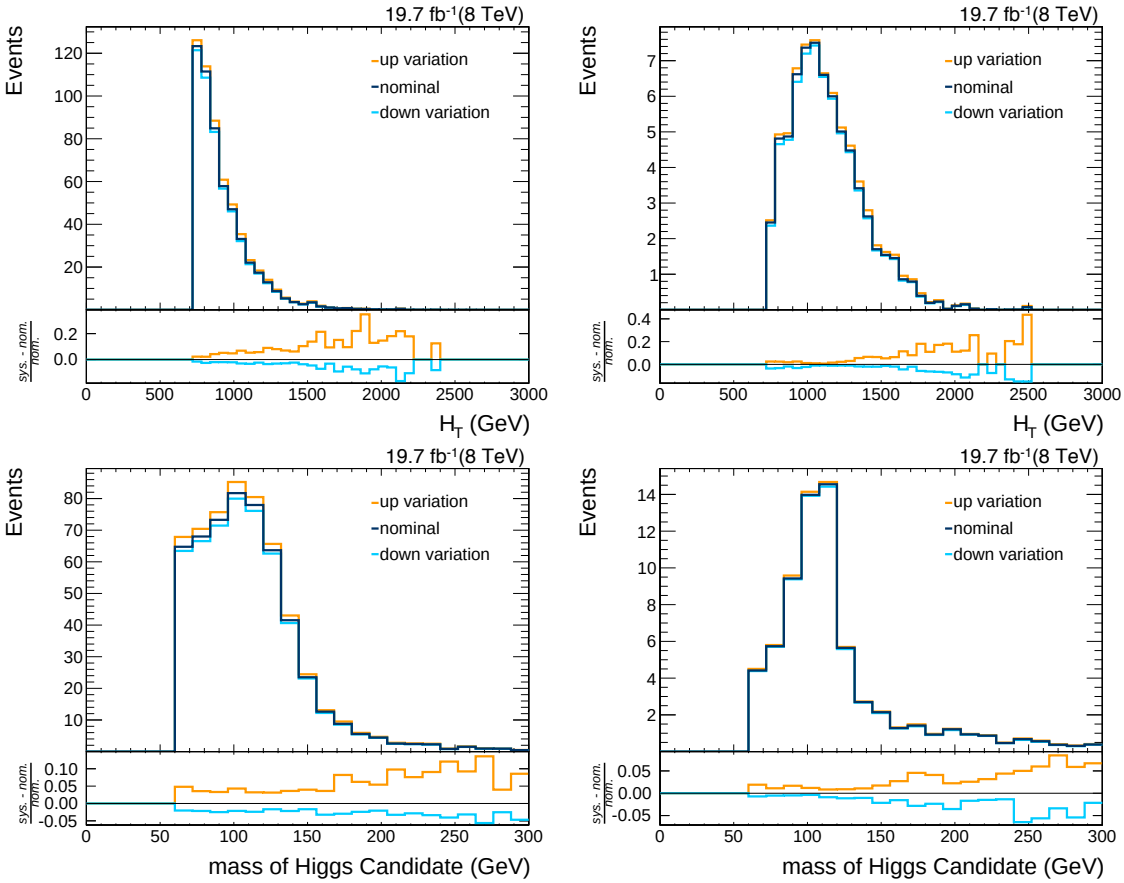

Figure 7.16: Expected impact of the uncertainties in the parton distributions functions on rate and shape of the H_T (top) and Higgs-candidate mass (bottom) distributions for $t\bar{t}$ events (left) and signal events simulated with a T'-quark mass of 700 GeV (right). The relative impact of the upward and downward variation of the uncertainty with respect to the nominal distribution is shown in the bottom part of each plot.

7.5.4 Renormalization and factorization scale

The renormalization and factorization scale, commonly referred to as the Q^2 scale, is another important source of systematic uncertainty introduced in the Monte Carlo simulation of $t\bar{t}$ events. Protons are not elementary particles but composite objects. Thus, the parton distributions functions f_i of the different partons i within the proton contribute to the cross section of a process $pp \to X$. This cross section can be written as

$$\sigma_{pp\to X}(\mu_F, \mu_R) = \sum_{a,b} \int dx_a dx_b f_a(x, \mu_F) f_b(x, \mu_F) \hat{\sigma}_{ab\to X}(x_a, x_b, \mu_R). \tag{7.4}$$

The partonic cross section $\hat{\sigma}$ can be calculated perturbatively [139]. It depends on the renormalization scale μ_R, which corresponds to the scale at that the running coupling α_s is evaluated. The dependency of the cross section on the factorization scale μ_F is introduced via the parton distribution functions f_i. This scale is needed to cancel large logarithms in the calculations and thus ensure perturbative convergence. Such logarithms can appear in the calculation of higher-order corrections to the cross section. The value of the scales μ_F and μ_R in the Monte Carlo generation is usually chosen to match a typical momentum transfer Q^2 of the simulated process, so that $\mu_F = \mu_R = Q^2$. The uncertainty in this scale needs to be taken into account as a systematic uncertainty in physics analyses using simulated samples.

Dedicated simulated samples were produced to evaluate the effect of the Q^2-scale uncertainty in the $t\bar{t}$ contribution. In the production of these samples, the value assumed by μ_F and μ_R is varied up and down to $2 \cdot Q^2$ and $\frac{1}{2} \cdot Q^2$ respectively. After application of the full event selection, these samples describe the impact of this Q^2 uncertainty in the selected $t\bar{t}$ events in this analysis.

The size of the samples produced with Q^2-scale variations is much smaller than that of the nominal POWHEG $t\bar{t}$ sample used in this analysis. Therefore, large statistical fluctuations are observed, especially in the tail regions of the distributions. In order not to be limited in sensitivity by the small sample size, the uncertainty is assumed to affect solely the event rate and not the shape of the $t\bar{t}$ distributions. Figure 7.17 shows shape comparisons between the nominal $t\bar{t}$ sample and the samples produced with systematic variations of the Q^2 scale. The H_T distributions are shown in the top row of the figure, those of the Higgs-candidate mass in the bottom row. In the plots on the left-hand side, the green curve corresponds to events produced with increased Q^2 scale. The impact of a reduced Q^2 scale is shown on the right-hand side. All of the distributions are scaled to unity for easier shape comparison.

The nominal H_T and Higgs-candidate mass distributions and those obtained from the samples produced with systematically varied Q^2 scale agree well within statistical uncertainties. In table 7.13, the results of χ^2 tests of the shape agreement are documented. The χ^2/DoF values also confirm the good shape agreement. Therefore, the assumption that only the event rate of the $t\bar{t}$ sample is affected by systematic Q^2 variations is valid. With a magnitude of 34% the Q^2-scale uncertainty is the largest uncertainty in this analysis.

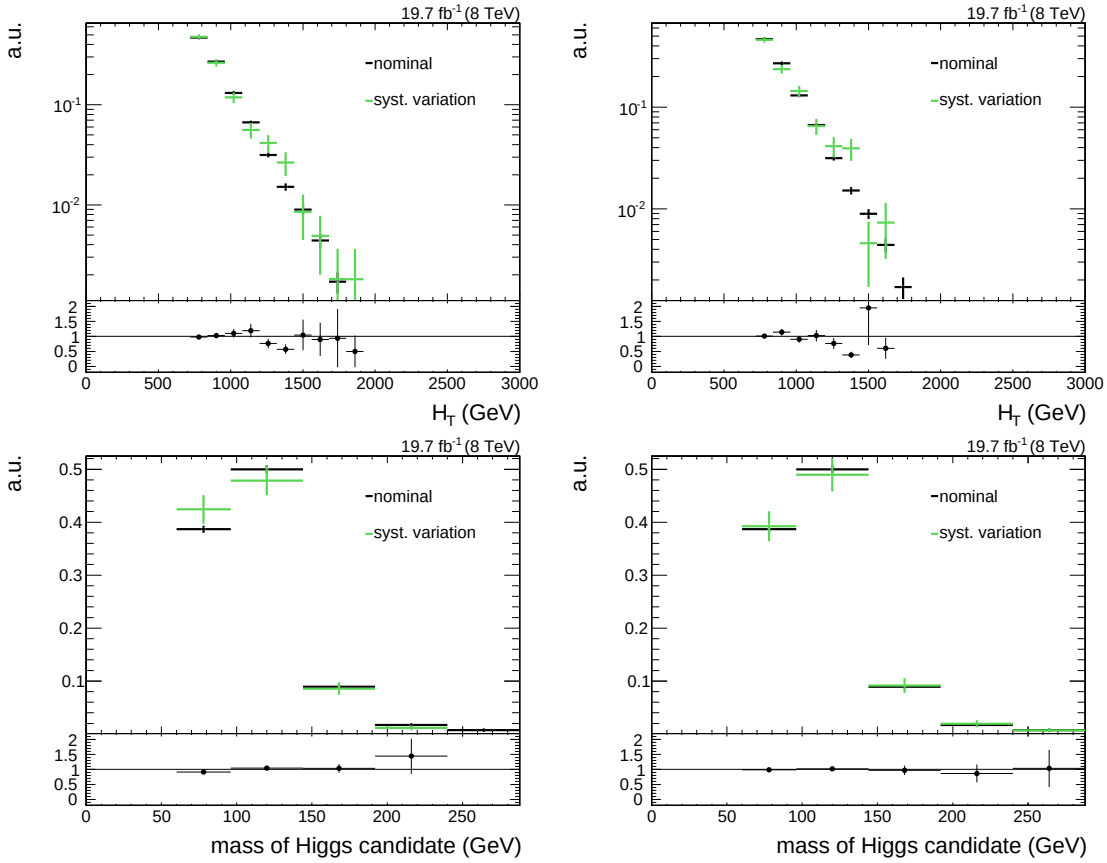

Figure 7.17: Shape comparison for the H_T (top) and Higgs-candidate mass (bottom) distributions between the nominal $t\bar{t}$ sample and the $t\bar{t}$ samples produced with varied Q^2 scale. Left: upward variation. Right: downward variation. All distributions are normalized to unity. The ratio of the two histograms is displayed in the bottom part of each plot.

m(Higgs candidate)		
	p value	χ^2/DoF
Scale up	0.31	1.19
Scale down	0.99	0.08
H_T		
	p value	χ^2/DoF
Scale up	0.72	0.68
Scale down	0.09	1.77

Table 7.13: χ^2 per degree of freedom and corresponding p value for the shape comparison between the normalized distributions obtained from the nominal $t\bar{t}$ sample and the $t\bar{t}$ samples produced with varied Q^2 scale.

7.5.5 Jet energy corrections

The jet energy corrections applied to CA15 jets in this analysis are derived for AK7 particle-flow jets clustered after charged-hadron subtraction as described in section 7.2.2. The uncertainties in the used jet energy corrections can have an impact on the acceptance of the event selection and also on the shape of distributions. To evaluate this effect, the four-momenta of CA15 jets are varied up and down according to the provided uncertainties in the applied corrections. This uncertainty is found to have only minor effects on the acceptance in this analysis. The resulting changes in the number of selected signal and $\mathrm{t\bar{t}}$ events do not exceed 1.3% for any of the simulated samples. Also, the impact on the shapes of distributions is negligible.

No jet energy corrections were applied to the subjets of CA15 jets in this analysis (see section 7.2.2). A systematic uncertainty in the actual subjet energy scale is applied though. The accuracy of the subjet energy scale is assumed to be similar to that of AK5 jets. Therefore, the uncertainties provided for the corrections derived for AK5 particle-flow jets after charged-hadron subtraction are used to evaluate the systematic uncertainty due to the non-exact knowledge of the subjet energy scale. The impact of this uncertainty in the shape of the H_T and Higgs-candidate mass distributions is illustrated in figure 7.18, for $\mathrm{t\bar{t}}$ events and for signal events simulated with a T'-quark mass of 700 GeV.

While small shape changing effects can be seen over the whole range of the distributions, they are more pronounced in the high-H_T regions, with respect to the nominal bin content. The overall uncertainty in the $\mathrm{t\bar{t}}$ event rate is about 5%. For the signal samples it ranges from about 4%, when assuming $m(T') = 500$ GeV, to less than 1% for the samples simulated with T'-quark masses larger than 700 GeV. Variations of the subjet energy scale have a direct impact on the distribution of H_T. The H_T spectrum is much harder for signal samples produced with a high T'-quark mass. Therefore, the H_T selection criterion rejects a much smaller fraction of signal events in the case of high T'-quark masses, making the final event rate more robust against variations of the jet energy scale.

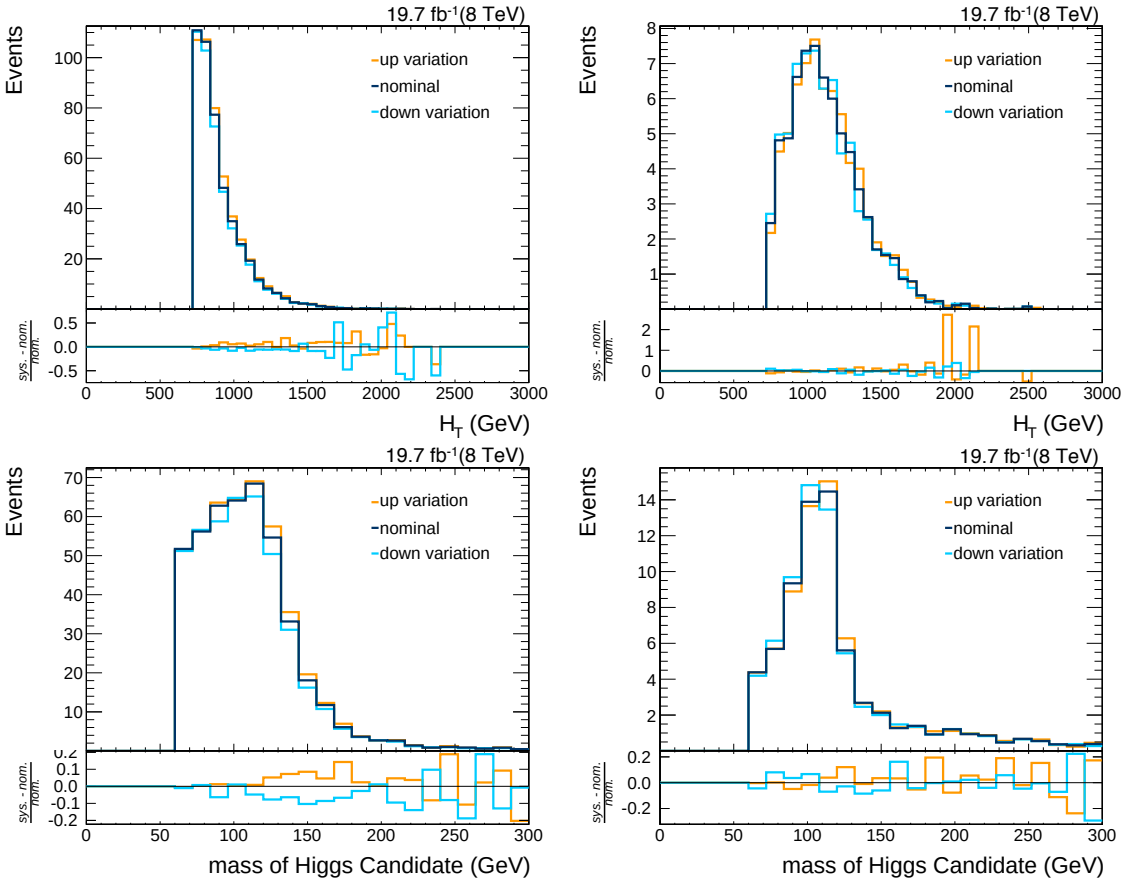

Figure 7.18: Impact of the uncertainties in the subjet energy scale on rate and shape of the H_T (top) and Higgs-candidate mass (bottom) distributions for $t\bar{t}$ events (left) and signal events simulated with a T'-quark mass of 700 GeV (right). The relative impact of the upward and downward variation of the uncertainty with respect to the nominal distribution is shown in the bottom part of each plot.

7.5.6 Trigger reweighting

As described in section 7.3, a trigger with a threshold of $H_T^{calo} > 750$ GeV is used. Slight differences in trigger efficiency are observed between data and Monte Carlo simulation in the low-H_T region. To correct for these differences, a scale factor SF_{trig} is applied to the simulated events. The effect of the systematic uncertainty in this scale factor is estimated through variation of the scale factor by $\pm 0.5 \cdot (1 - SF_{trig})$. Figure 7.19 shows the impact of this uncertainty in the H_T and Higgs-candidate mass distributions of the selected $t\bar{t}$ events on the left-hand side and signal events simulated with a T'-quark mass of 700 GeV on the right-hand side.

In the H_T distribution, displayed on the left-hand side of figure 7.19, only the first two bins, in which the discrepancy in trigger efficiency is found, are affected. The impact on the first bin of the distribution is rather large. The tail region of this distribution, which is expected to be populated by the hypothetical signal events, is not affected by this

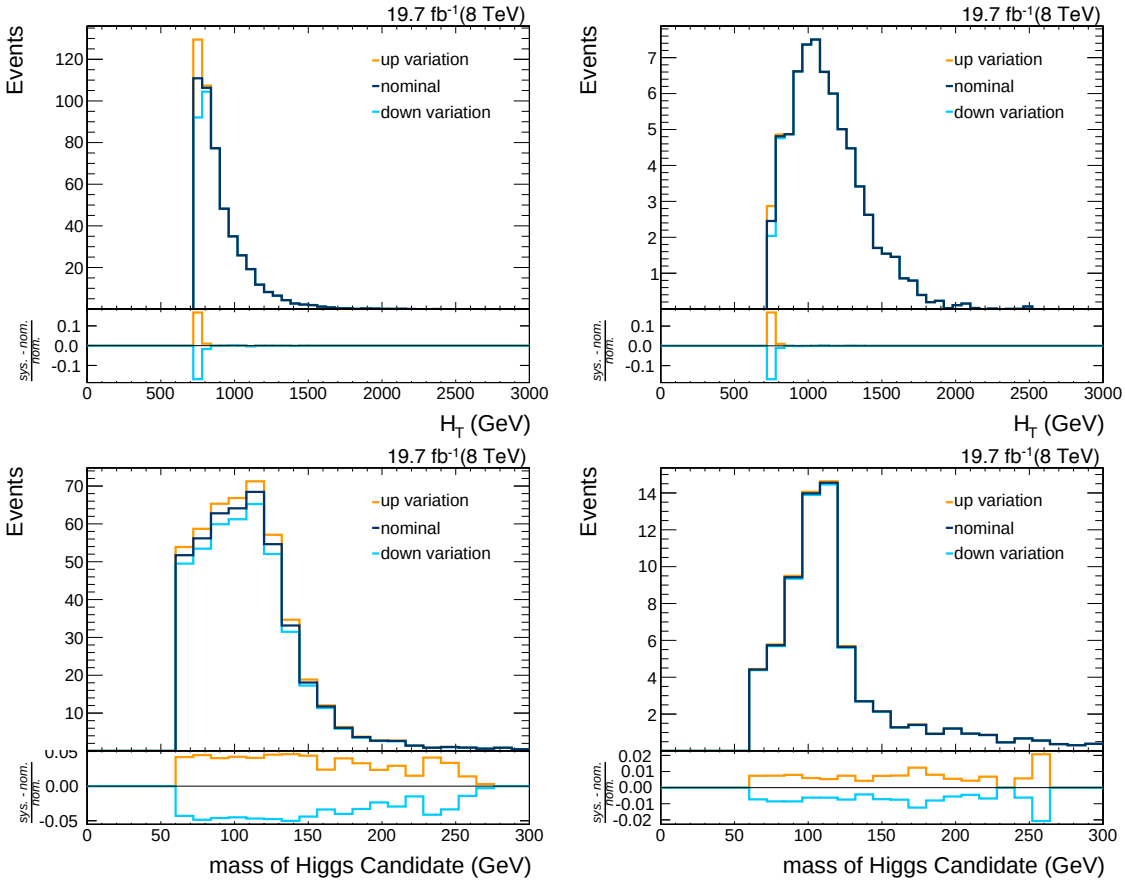

Figure 7.19: Impact of the trigger efficiency reweighting uncertainty on rate and shape of the H_T (top) and Higgs-candidate mass (bottom) distributions for $t\bar{t}$ events (left) and signal events simulated with a T'-quark mass of 700 GeV (right). The relative impact of the upward and downward variation of the uncertainty with respect to the nominal distribution is shown in the bottom part of each plot.

uncertainty though. In the Higgs-candidate mass distribution the systematic variation is observed to have an effect over the full range of the variable, albeit much smaller than in the first bin of the H_T distribution. For the Higgs-candidate mass variable, no significant changes in shape are caused by the systematic variation of SF_{trig}.

For $t\bar{t}$ events, the upward variation of the trigger scale factor leads to a 6.9% change in the rate of selected events. The effect is less pronounced for signal events. For these, it ranges from 2.5% for the sample generated with a T'-quark mass of 500 GeV to less than 0.1% under assumption of a T'-quark mass of 1000 GeV.

7.5.7 b-tagging scale factor

Scale factors are applied to account for differences in the b-tagging performance observed between data and simulated events. Individual scale factors are available for the actual b-tagging efficiency [14], the misidentification rate for charm quarks, and that for light

quarks. The specifics on the application of these scale factors can be found in section 7.2.3.

The uncertainties in the scale factors for the b-tag efficiency and charm-quark misidentification rate are assumed to be fully correlated. In order to quantify the effect of these uncertainties in the analysis, the scale factors for the b-tagging efficiency and charm-quark misidentification rate are varied simultaneously in the upward or downward direction, according to the provided uncertainty in the scale factor measurement. These two scale factors are also referred to as "heavy-flavor scale factors" in the following. The impact of the uncertainty in the scale factor determined for the misidentification rate of light quarks is evaluated independently, again by variation of the scale factor value within the respective uncertainties.

A fixed cone size of $\Delta R < 0.3$ is used in the CSV-b-tagging algorithm for association of tracks to the jets, as described in section 5.2.4. In the case of subjet b tagging, this can lead to tracks being ambiguously assigned to two very close subjets. In figure 7.20, the ΔR distance between the two closest subjets found the Higgs-candidate jet for three different signal samples and QCD-multijet events with large H_T values is shown.

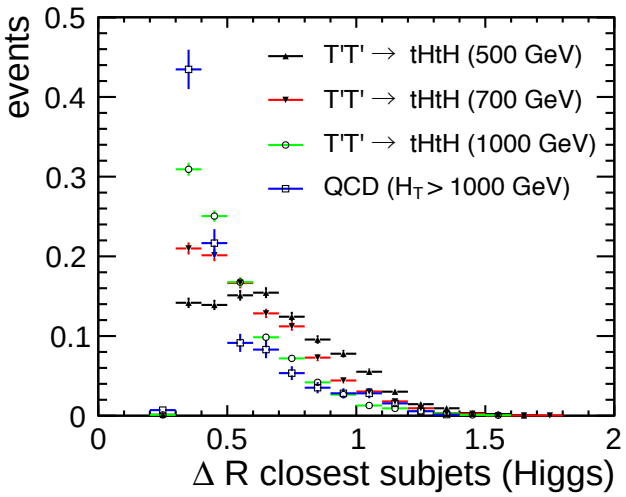

Figure 7.20: ΔR between the two closest subjets in Higgs-candidate jets for three signal samples simulated under different hypotheses for the T'-quark mass and for QCD-multijet events from Monte Carlo simulation [137].

By construction, subjets cannot be closer than $\Delta R = 0.3$, as this is the cone radius of the subjets considered in this analysis. In order to consider possible correlation effects of the track sharing of slightly further separated subjets, the uncertainty in the b-tagging and b-misidentification scale factors is doubled for pairs of subjets found at $0.3 < \Delta R < 0.4$. In table 7.14, the fraction of events affected by this procedure is listed. Because only few events contain subjets with such small pairwise separation, the overall effect of this increase in uncertainty is small. The total b-tagging scale factor uncertainty is changed by less than 2%, the effect on the b-misidentification scale factor uncertainty is below percent level.

T'-quark mass	Higgs-tagged jets	Top-tagged jets
500 GeV	15%	13%
700 GeV	22%	17%
1000 GeV	31%	31%

Table 7.14: Fraction of signal events containing Higgs-tagged or top-tagged CA15 jets with subjets separated by less than $\Delta R = 0.4$.

The outcome of this systematic variation of the scale factors is illustrated in figure 7.21 for the heavy-flavor scale factors and 7.22 for the scale factor for the light-flavor-misidentification rate.

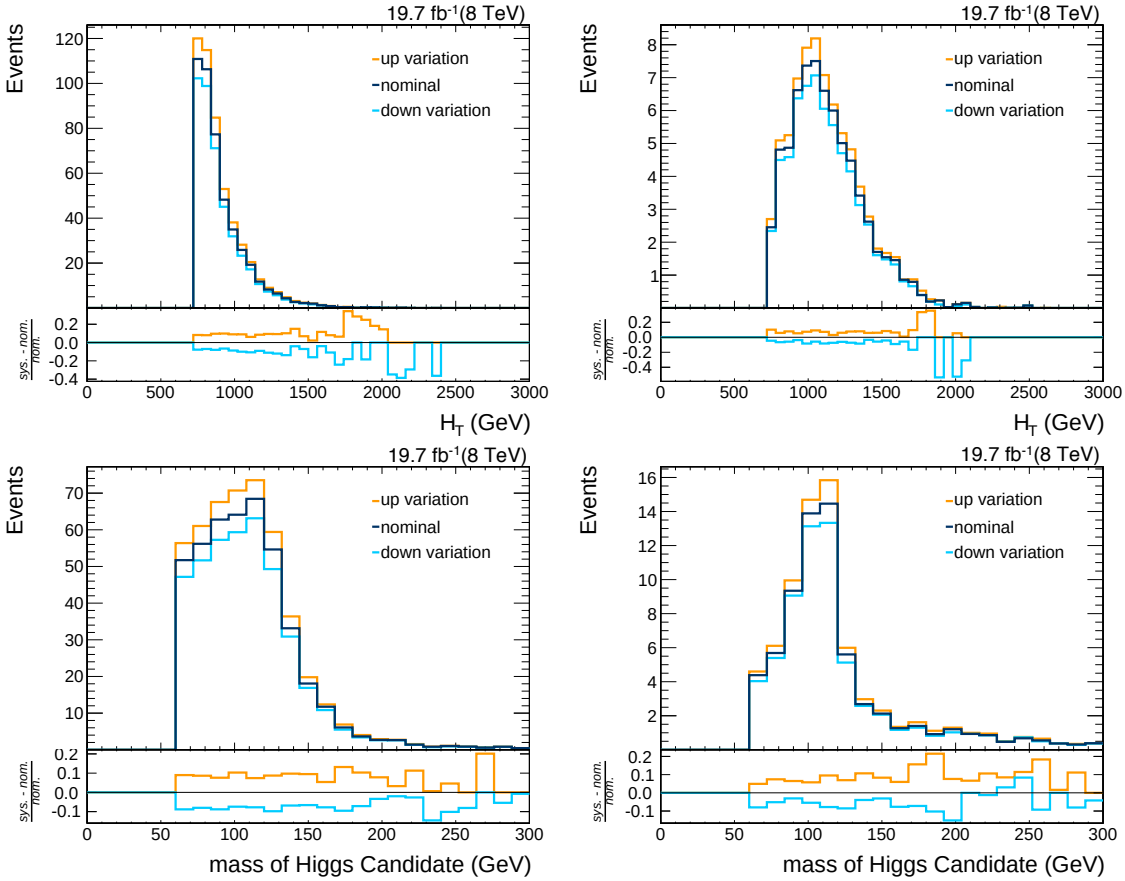

Figure 7.21: Impact of the uncertainties in the subjet-b-tagging and charm-quark-misidentification scale factors on rate and shape of the H_T (top) and Higgs-candidate mass (bottom) distributions. The distributions are obtained from the $t\bar{t}$ sample (left) and the signal sample simulated with a T'-quark mass of 700 GeV (right). The b-tagging uncertainties for close subjets with $0.3 < \Delta R < 0.4$ are increased by a factor 2. The relative impact of the upward and downward variation of the uncertainty with respect to the nominal distribution is shown in the bottom part of each plot.

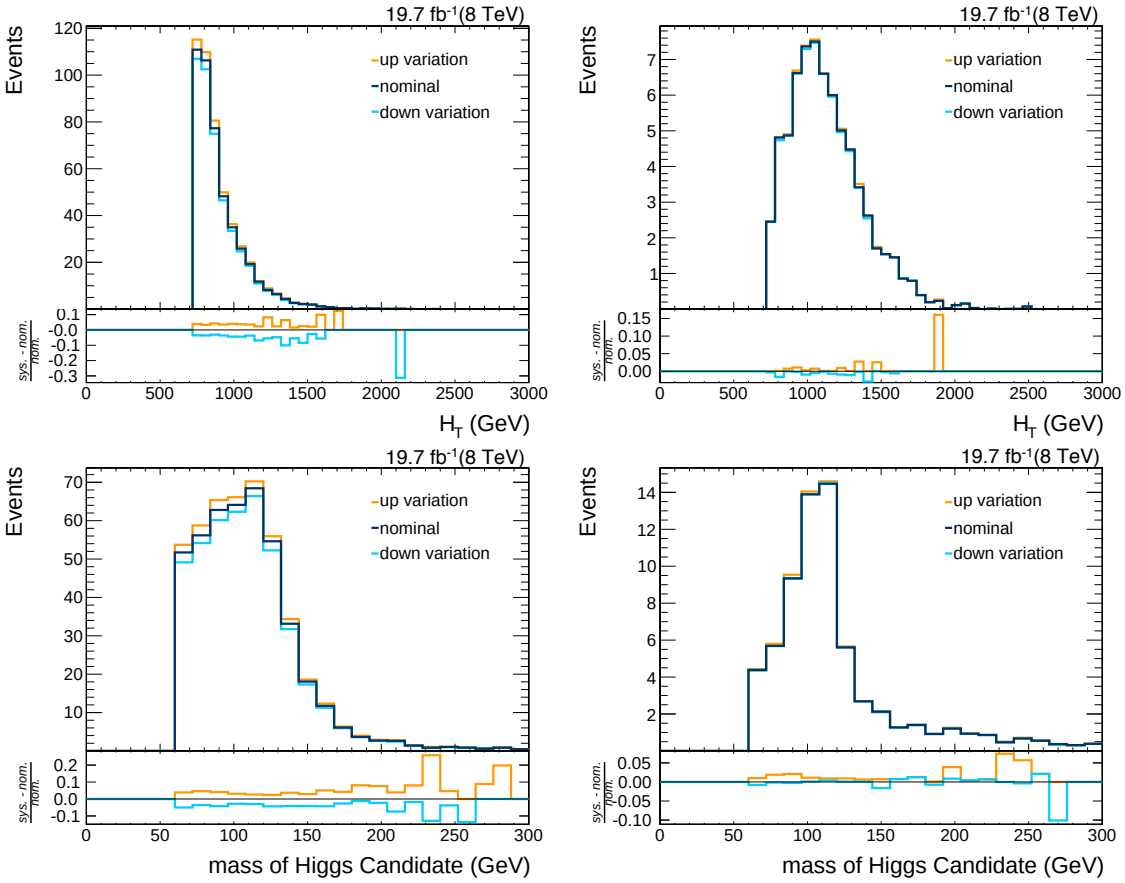

Figure 7.22: Impact of the uncertainty in the light-quark misidentification scale factor on rate and shape of the H_T (top) and Higgs-candidate mass (bottom) distributions. The distributions are obtained from the $t\bar{t}$ sample (left) and the signal sample simulated with a T'-quark mass of 700 GeV (right). The b-tagging uncertainties for close subjets with $0.3 < \Delta R < 0.4$ are increased by a factor 2. The relative impact of the upward and downward variation of the uncertainty with respect to the nominal distribution is shown in the bottom part of each plot.

In the top row of both figures, the H_T distributions are shown for $t\bar{t}$ background events on the left-hand side and signal events simulated with a T'-quark mass of 700 GeV on the right-hand side. The corresponding distributions of the Higgs-candidate mass can be found in the bottom rows of the two figures. The changes in the shape of the distributions induced by these systematic uncertainties are taken into account in this analysis. The overall effect of the uncertainty in the heavy-flavor scale factor on the number of selected events ranges between 6% and 9% for the $t\bar{t}$ background and the different signal samples. This rather large impact is due to the requirement of at least three b-tagged subjets in the event selection. The magnitude of the systematic uncertainty induced by variation of the light-flavor misidentification scale factor is smaller, ranging between 0.5% and 4.2%.

7.5.8 HEPTopTagger scale factor

In addition to the nominal values of the top-tagging efficiency scale factors for the HEP-TopTagger, also uncertainties in their measurement need to be taken into account [102]. Their effect on the analysis is evaluated by varying the applied scale factors up and down according to the uncertainties in their measured value. As illustrated in figure 7.23, the shapes of the H_T and Higgs-candidate mass distributions are changed by this uncertainty to some extent in both $t\bar{t}$ and signal events generated with a T'-quark mass of 700 GeV. The overall magnitude of this uncertainty is rather small and does not exceed 3% for any of the simulated samples used in this analysis.

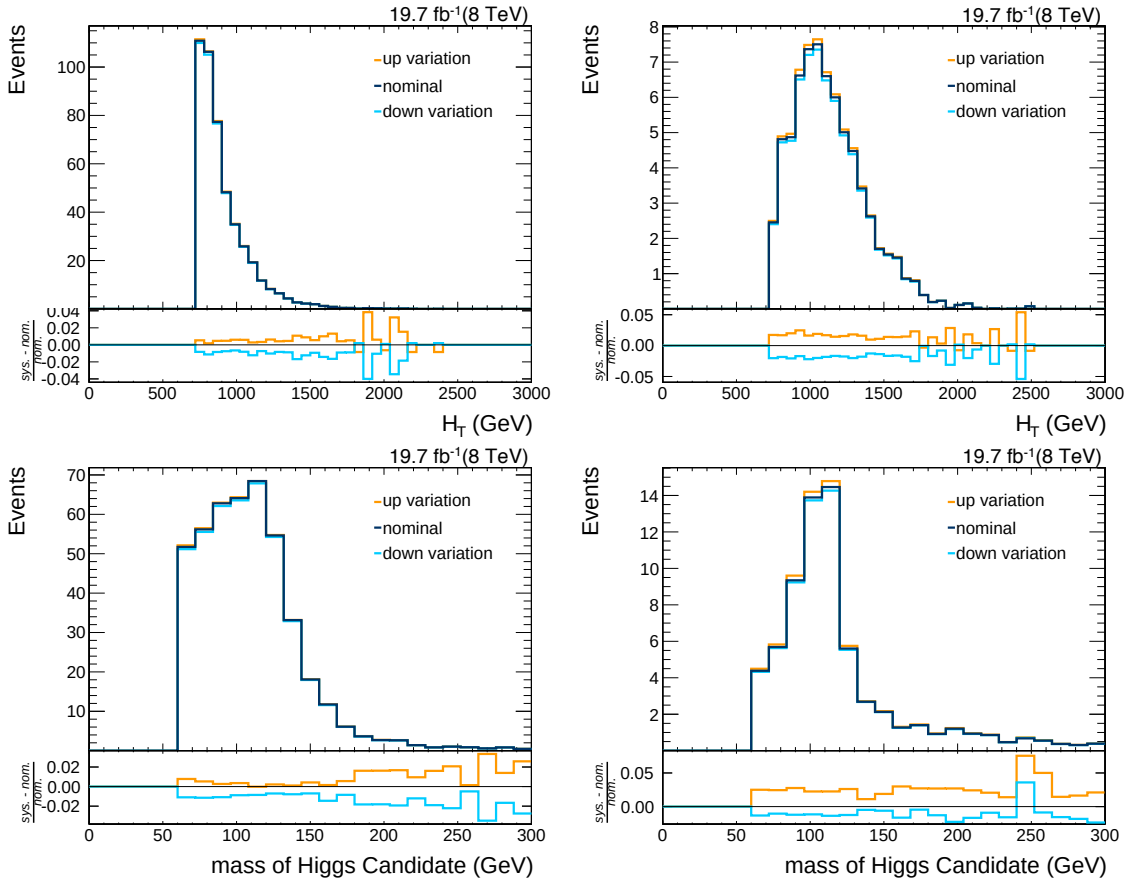

Figure 7.23: Impact of the uncertainty in the HEPTopTagger scale factors on rate and shape of the H_T (top) and Higgs-candidate mass (bottom) distributions. The distributions are obtained from the $t\bar{t}$ sample (left) and the signal sample simulated with a T'-quark mass of 700 GeV (right). The relative impact of the upward and downward variation of the uncertainty with respect to the nominal distribution is shown in the bottom part of each plot.

7.5.9 QCD-multijet background model derived from data

The QCD-multijet background is estimated in an ABCD method, as described in section 7.4. It is not directly affected by the previously described systematic uncertainties in simulated samples. Part of the method is the subtraction of $t\bar{t}$ contributions in the sideband regions A,B, and C though. This contamination of $t\bar{t}$ events in these regions is estimated using simulated samples. The rate of selected simulated $t\bar{t}$ events is strongly affected by the systematic uncertainty due to the Q^2 scale variation. Therefore, this uncertainty is propagated to the estimate of the QCD-multijet background derived from data.

For this propagation, the impact of the Q^2-scale uncertainty in the $t\bar{t}$ rate in all three sideband regions is determined. The resulting higher or lower number of $t\bar{t}$ events is then subtracted from the data in each sideband region. As the results of these subtractions are used in the calculation of the QCD-multijet event rate in the signal region, the QCD-multijet event rate is in second order also affected by the Q^2 scale uncertainty in the $t\bar{t}$ event simulation. The effect on the event rate is smaller than 1% for both the upward and downward variation of the Q^2 scale though.

When deriving the shape of the QCD-multijet contribution from sideband region B, the $t\bar{t}$ distribution is also subtracted from the data. Subtraction of $t\bar{t}$ distributions that are modified according to the Q^2 scale uncertainty can lead to shape changes in the distributions of the data-driven QCD-multijet model for the signal region. The effects on the shape are found to be negligible though. In the statistical analysis, the propagated Q^2-scale uncertainty in the QCD-multijet background is treated as fully correlated with the Q^2-scale uncertainty in the $t\bar{t}$ background contribution.

In addition to the uncertainty from the propagation of the Q^2-scale uncertainty, an overall rate uncertainty of 10% in the single Higgs-tag category and 20% in the multi Higgs-tag category is applied to the QCD-multijet model from data. This way the statistical uncertainties in the calculations of the event-rate for the QCD-multijet model are taken into account. The slight observed changes due to signal contamination discussed in section 7.4.4 are also well contained in these overall event rate uncertainties.

In table 7.15, the uncertainties in the rate of selected $t\bar{t}$-background and T'-quark-signal events are summarized. The largest impact on the rate of selected $t\bar{t}$-background events is caused by the Q^2 scale uncertainty, followed by that of the subjet-b-tagging scale, and the trigger-reweighting uncertainty. For signal, the subjet-b-tagging scale is the leading uncertainty. The impact of jet energy correction uncertainies on the rate of selected signal events decreases with increasing T'-quark mass. The H_T spectrum is directly affected by the jet energy corrections. Because of the harder H_T spectrum in signal samples simulated with a large T'-quark mass, a smaller fraction of these events is rejected by the H_T-selection criterion resulting in a smaller percentual impact of this uncertainty.

All of these uncertainties are taken into account as nuisance parameters in the statistical analysis described in section 7.6.

			T'T' → tHtH signal					
		$t\bar{t}$	m(T') = 500 GeV	m(T') = 600 GeV	m(T') = 700 GeV	m(T') = 800 GeV	m(T') = 900 GeV	m(T') = 1 TeV
Cross section	+	13	2.11	1.94	1.82	1.75	1.71	1.72
	−	-13	-2.18	-2.13	-2.06	-2.03	-2.03	-1.92
Q^2 scale	+	11	-	-	-	-	-	-
	−	34	-	-	-	-	-	-
JEC	+	0.5	1.3	0.5	0.3	0.5	0.2	0.4
CA15	−	0.5	0.3	0.5	0.3	0.5	0.2	0.4
JEC	+	-4.1	-2.8	-2.0	-0.7	-0.5	-0.3	-0.4
Subjets	−	5.0	3.7	2.0	0.7	0.3	0.2	0.1
PDF	+	-4.4	-2.4	-2.6	-3.2	-4.8	-4.1	-5.8
	−	8.0	3.9	3.9	4.8	5.9	5.4	7.3
b-tag	+	-7.5	-6.8	-6.4	-6.5	-6.6	-7.4	-8.0
scale	−	9.2	6.0	6.7	7.1	5.7	7.5	7.8
b-mistag	+	-3.2	-0.7	-0.5	-0.6	-1.3	-0.7	-1.0
scale	−	4.2	1.2	1.2	0.9	0.6	0.8	0.8
Top-tag	+	-0.4	-1.7	-2.0	-0.8	-2.3	-2.2	-2.3
scale	−	0.9	1.6	1.5	1.7	1.8	.8	1.8
Trigger	+	6.9	2.5	1.4	0.7	0.5	0.2	< 0.1
weight	−	-2.0	-2.5	-1.3	-0.6	-0.2	< 0.1	< 0.1

Table 7.15: Impact of systematic uncertainties on the rate of simulated $t\bar{t}$ background and T'-quark-signal samples in percent. For each uncertainty the impact of the upward (+) and downward (−) variation is quoted. The abbreviation "JEC" stands for "jet energy corrections", "PDF" for "parton distribution function".

7.6 Results

The results of the search for pair-produced T' quarks are presented in this section. Exclusion limits on the cross section for T'-quark pair production are set assuming a branching fraction Br(T' → tH) = 100% and also for mixed decay modes. Results are provided for all allowed combinations of branching fractions for the three decay modes T'→ tH, T'→ tZ, and T'→ bW. In comparisons with the theory prediction for the cross section, also exclusion limits on the T'-quark mass are derived.

7.6.1 Results for Br(T' → tH) = 100%

The distributions of the H_T and Higgs-candidate mass variables are shown in figure 7.24 for the inclusive selection, in figure 7.25 for the single Higgs-tag category, and in figure 7.26 for the multi Higgs-tag category. The QCD-multijet background shown in these figures is derived from data, following the method described in section 7.4, the $t\bar{t}$ contribution is taken from Monte Carlo simulation. The mass of the hypothetical T' quark was set to 500 GeV, 700 GeV, and 1000 GeV in the simulation of the three signal samples shown in these plots.

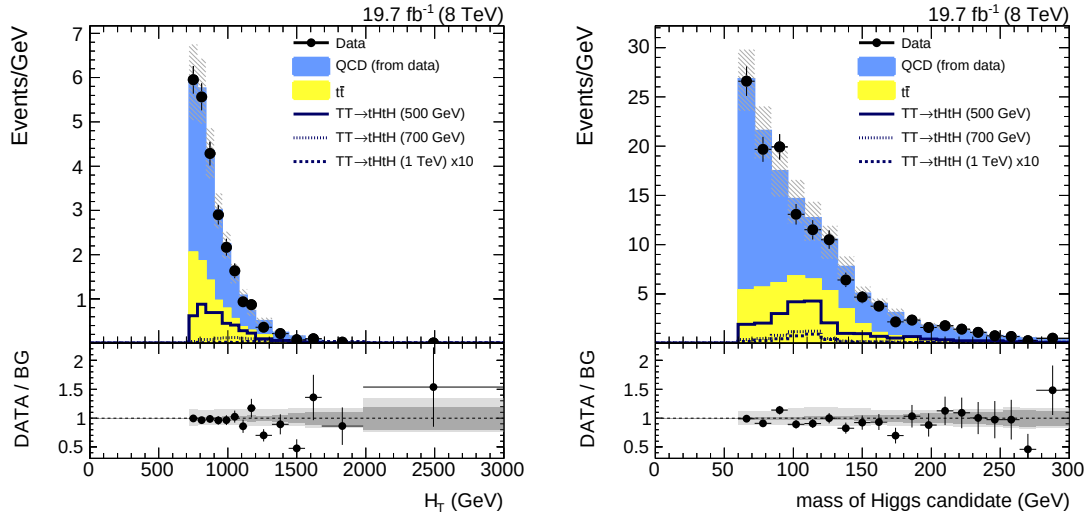

Figure 7.24: Distributions of the H_T (left) and Higgs-candidate mass (right) variables after the full selection. The quadratic sum of all systematic and statistical uncertainties in the two background contributions is depicted by the hashed error bands in the stack plot. In the ratio plot, the central, darker grey band corresponds to the statistical uncertainty, the outer lighter grey band to the quadratic sum of statistical and systematic uncertainties.

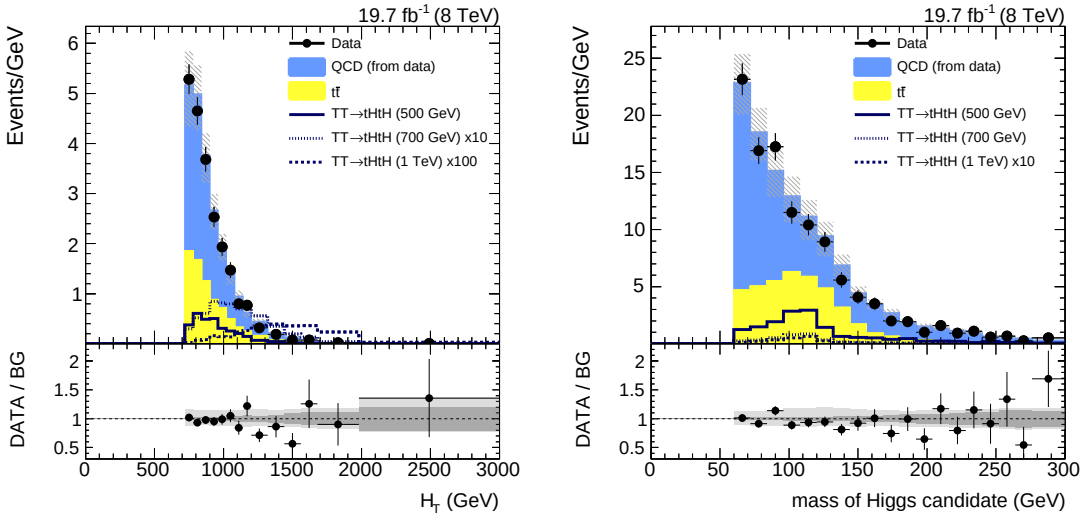

Figure 7.25: Distributions of the H_T (left) and Higgs-candidate mass (right) variables in the single Higgs-tag category. The quadratic sum of all systematic and statistical uncertainties in the two background contributions is depicted by the hashed error bands in the stack plot. In the ratio plot, the central, darker grey band corresponds to the statistical uncertainty, the outer lighter grey band to the quadratic sum of statistical and systematic uncertainties.

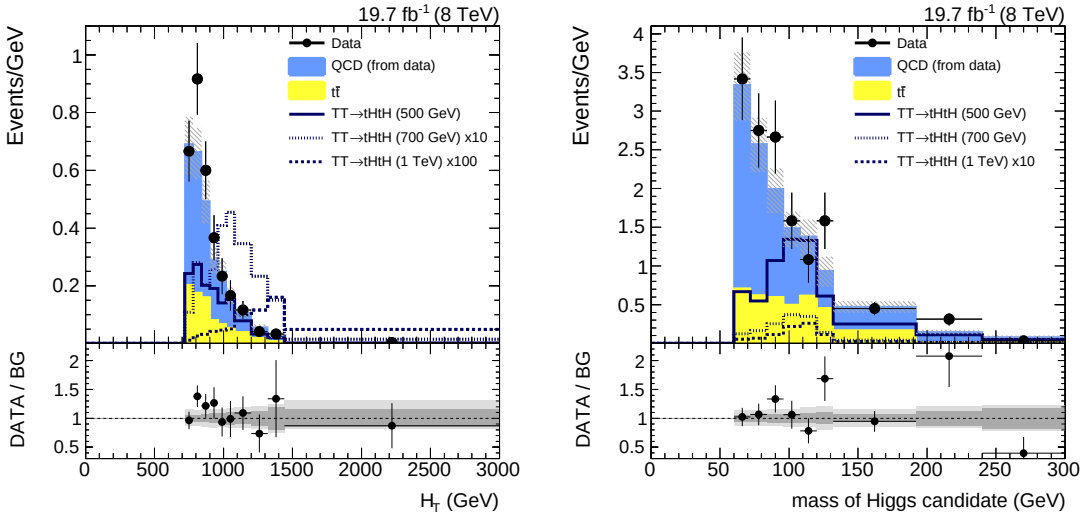

Figure 7.26: Distributions of the H_T (left) and Higgs candidate mass (right) variables in the multi Higgs-tag category. The quadratic sum of all systematic and statistical uncertainties in the two background contributions is depicted by the hashed error bands in the stack plot. In the ratio plot, the central, darker grey band corresponds to the statistical uncertainty, the outer lighter grey band to the quadratic sum of statistical and systematic uncertainties.

The distribution of signal events in all three simulated samples is clearly different from the distribution of the expected background events for both, the H_T and the Higgs-candidate mass variables. Signal events tend to have a harder H_T spectrum than background events, especially under the assumption of large T'-quark masses. In the distributions of the Higgs-candidate mass in signal events, shown in the right plots of figures 7.24-7.26, a peak-like structure can be seen at a value corresponding approximately to the Higgs boson mass of 125 GeV. The distribution of expected background events does not feature this structure though. Because of these shape differences between expected signal and background distributions, these two variables are well suited for discrimination between signal and background events in the statistical interpretation of the search results.

No excess of data over the background expectation is observed in any of the distributions. Bayesian exclusion limits on the cross section for T'-quark pair production are set and used to also derive exclusion limits on the T'-quark mass in comparisons to the theory prediction for the production cross section. The theta framework [127] is used for the statistical evaluation of the results. More information on the theta framework and Bayesian statistics are provided in chapter 6.

Figure 7.27 shows the expected upper exclusion limits at 95% confidence level on the cross section for pairwise T'-quark production as a function of the T'-quark mass.

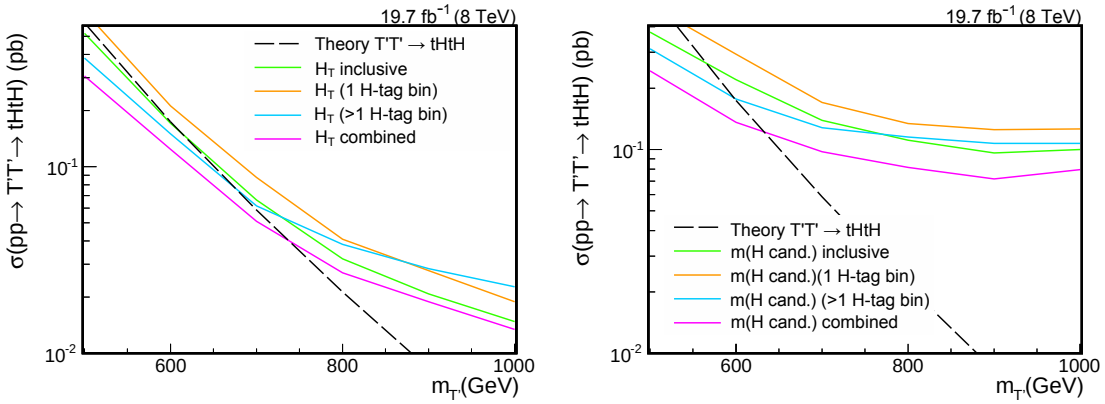

Figure 7.27: Expected upper exclusion limits at 95% confidence level on the cross section for T'-quark pair production extracted from the H_T variable (left) and the Higgs-candidate mass distribution (right). The limits are obtained using four different analysis setups: analysis of all events passing the event selection before categorization (green), analysis of events in the single (orange) or multi (blue) Higgs-tag category only, and simultaneous analysis of the events in the two categories (magenta).

These limits are obtained from the distributions of the H_T and Higgs-candidate mass distributions in figures 7.24-7.26. The exclusion limits shown in the left plot of figure 7.27 are derived using information of the H_T distributions only. While an overall loss in sensitivity is observed when analyzing only events in the single Higgs-tag category, a slightly better cross section limit can be set when using only events of the multi Higgs-tag category as input for the statistical analysis. The best result is obtained when performing the fit simultaneously in the single and multi Higgs-tag categories. The same pattern is seen

when fitting the distribution of the Higgs-candidate mass in the different setups.

The theory expectation for the cross section of T'-quark pair production as a function of the T'-quark mass is depicted as a dashed black line in these plots. The crossing point of this theory curve with the cross section limits indicates the expected limit on the T'-quark mass. While the expected mass limit from the H_T distributions is clearly better than that from the Higgs-candidate mass distribution, more stringent limits on the cross sections for low masses of the hypothetical T' quark can be achieved using the latter variable. Both variables, H_T and the Higgs-candidate mass, contribute to the sensitivity of this analysis.

To exploit the sensitivity of both, the H_T and Higgs-candidate mass variables, they are combined using a multivariate technique. To determine the optimal technique for this combination, the correlation between the two variables is examined. Table 7.16 shows the percentual correlation observed in signal events simulated with T'-quark masses of 500 GeV, 700 GeV, and 1000 GeV in the single and multi Higgs-tag categories. A visualization of the correlation can be found in the scatter plots of the two variables displayed in figure 7.28 for signal events generated with a T'-quark mass of 700 GeV.

$m_{\mathrm{T'}}$	500 GeV	700 GeV	1000 GeV
Single Higgs-tag	21.5	8.0	8.6
Multi Higgs-tag	28.1	7.2	11.5

Table 7.16: Correlation between the H_T and the Higgs-candidate mass variables for the single and multi Higgs-tag categories. These numbers are obtained from simulated events assuming three different mass hypotheses. The correlation is expressed in percent. The signal samples used for this studies are simulated assuming Br(T' $\rightarrow$ tH)= 100%

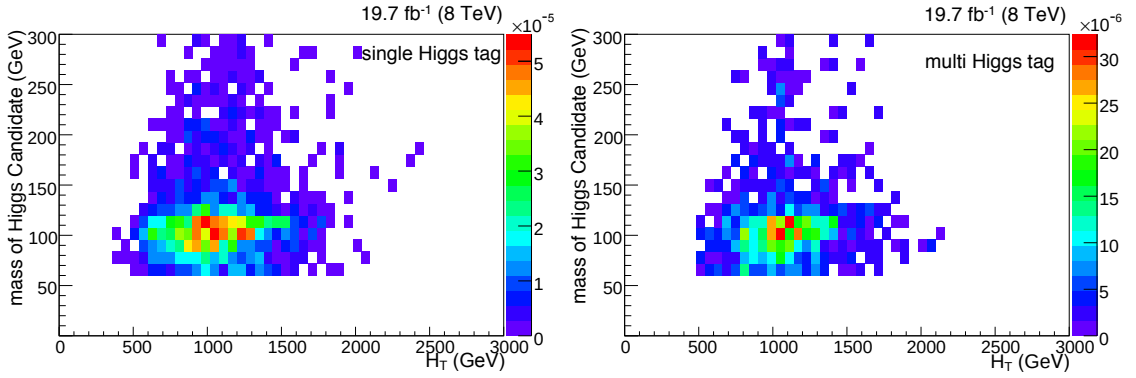

Figure 7.28: Correlation between the H_T and the Higgs-candidate mass variables for the single (left) and multi (right) Higgs-tag categories. These numbers are obtained from simulated events assuming a T'-quark mass of 700 GeV.

Correlations of less than 30% are observed for signal events simulated with a T'-quark mass of 500 GeV. In samples produced assuming higher T'-quark masses, the correlations are found to be even smaller, of the order of 10%. Likelihood ratios are a suitable method to combine variables with such small correlations.

Using the probability density functions P_{signal} and $P_{background}$ for signal and background events in each variable, a likelihood discriminating variable L can be defined as

$$L = ln\left(1 + \frac{P_{signal}(H_T)}{P_{background}(H_T)} \cdot \frac{P_{signal}(m(Higgs\ candidate))}{P_{background}(m(Higgs\ candidate))}\right). \tag{7.5}$$

The discriminating variable L combines the information provided by the two input variables. In order to obtain the probability density function for background events $P_{background}$, the distribution of simulated $t\bar{t}$ events and that of QCD-multijet events derived from data using the ABCD method are added. The probability density function for signal events P_{signal} is taken directly from Monte Carlo simulation. An individual likelihood distribution L is determined for each T'-quark mass hypothesis, in both the single and the multi Higgs-tag categories. The resulting twelve likelihood distributions of data compared to the expected background and signal distributions are shown in figures 7.29-7.31.

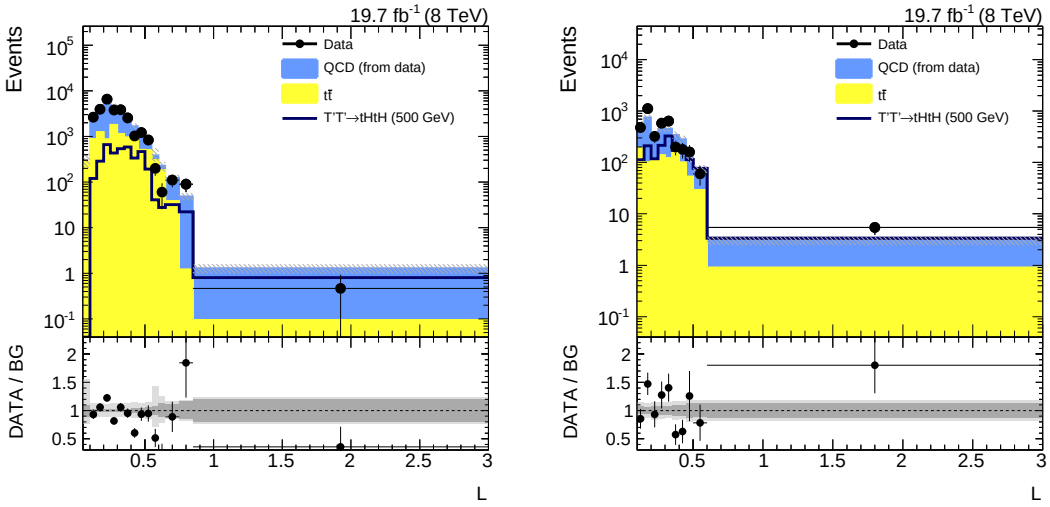

Figure 7.29: Likelihood discriminating variable L constructed using signal simulated with a T'-quark mass of 500 GeV in the single (left) and multi (right) Higgs-tag categories. The quadratic sum of all systematic and statistical uncertainties in the two background contributions is depicted by the hashed error bands in the stack plot. In the ratio plot, the central, darker grey band corresponds to the statistical uncertainty, the outer lighter grey band to the quadratic sum of statistical and systematic uncertainties.

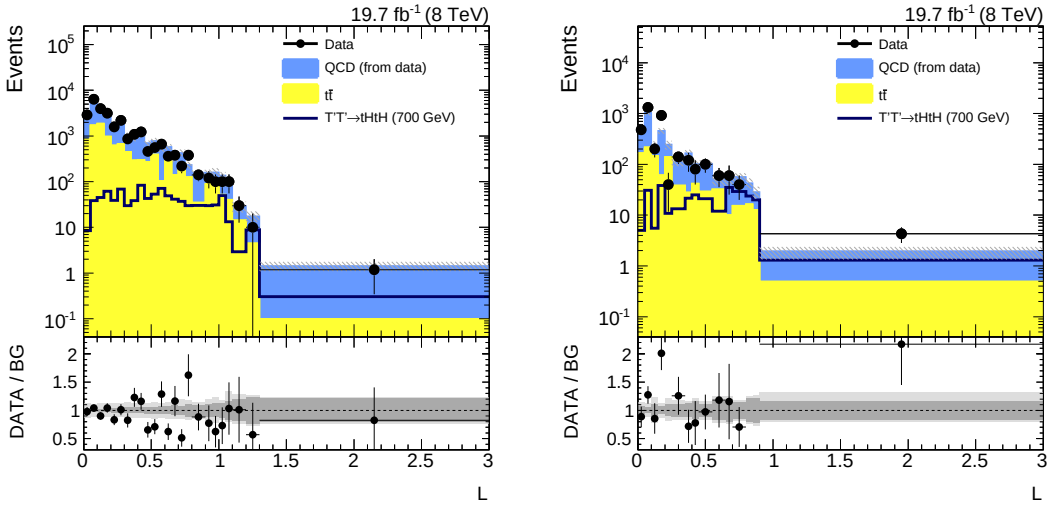

Figure 7.30: Likelihood discriminating variable L constructed using signal simulated with a T' quark mass of 700 GeV in the single (left) and multi (right) Higgs-tag categories. The quadratic sum of all systematic and statistical uncertainties in the two background contributions is depicted by the hashed error bands in the stack plot. In the ratio plot, the central, darker grey band corresponds to the statistical uncertainty, the outer lighter grey band to the quadratic sum of statistical and systematic uncertainties.

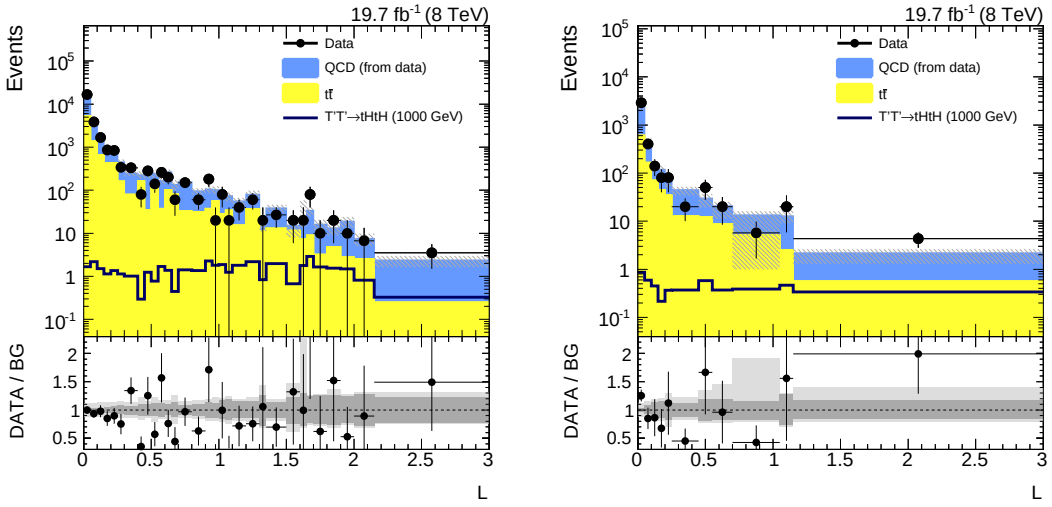

Figure 7.31: Likelihood discriminating variable L constructed using signal simulated with a T'-quark mass of 1000 GeV in the single (left) and multi (right) Higgs-tag categories. The quadratic sum of all systematic and statistical uncertainties in the two background contributions is depicted by the hashed error bands in the stack plot. In the ratio plot, the central, darker grey band corresponds to the statistical uncertainty, the outer lighter grey band to the quadratic sum of statistical and systematic uncertainties.

The great discrimination power of the likelihood variables for all T'-quark mass hypotheses is visible in these plots: while the signal events tend to present large values of L, the background events populate the region with lower L values. This behavior can be seen in both, the single and the multi Higgs-tag categories. The good agreement of data with the background expectation seen in figures 7.29-7.31 was forecast by the consistency between data and background expectation found in the distributions of the individual input variables H_T and the Higgs-candidate mass.

Figure 7.32 illustrates the improvement in sensitivity when using the likelihood distributions L as input for the Bayesian limit setting procedure. The left plot of the figure illustrates once more, how the analysis benefits from the categorization of the selected events. The exclusion limits obtained using events from the single and multi Higgs-tag categories only are clearly improved when fitting the events in both categories simultaneously. The result of the simultaneous fit produces the most stringent expected exclusion limits.

In the right plot, the improvement of the limits due to the combination of the two sensitive variables into a single likelihood variable L is shown. Here, the strongest expected limits for each variable are shown that are obtained by fitting each variable in the two event categories simultaneously. Though the expected limit from the Higgs-candidate mass distribution is less stringent than the expected limit from the H_T distributions for all but the lowest T'-quark mass hypothesis, the sensitivity improves over the whole mass range when extracting the expected limits from the likelihood ratio of the two variables instead of the H_T variable alone. The distribution of the likelihood variable L in the two event categories is therefore used for the determination of exclusion limits from data.

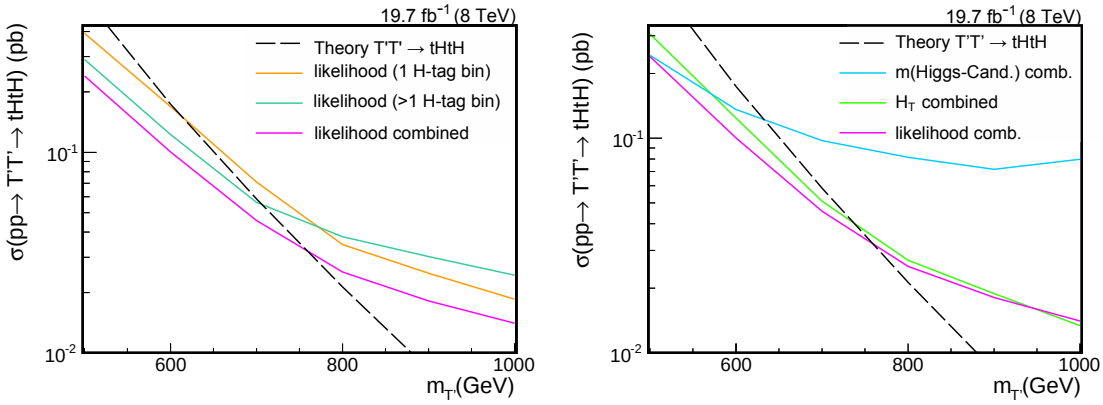

Figure 7.32: Comparison of the expected upper exclusion limits at 95% confidence level on the cross section for T'-quark pair production. Left: limits obtained from the likelihood variable L in the single (orange) and multi (blue) Higgs-tag categories and in a simultaneous analysis of both categories (magenta). Right: limits obtained from the H_T (green), Higgs-candidate mass (blue), and likelihood (magenta) variables in simultaneous analysis of the single and multi Higgs-tag categories.

In the statistical analysis, posterior distributions for all considered nuisance parameters are determined. Figure 7.33 shows the distribution of the likelihood variable L calculated under the 700 GeV T'-quark mass hypothesis. In these plots, the background contributions are scaled to the central values of the posterior distributions of the nuisance parameters corresponding to the background event rates. Also, template morphing effects on the shapes of the contributions are taken into account. In both, the single Higgs-tag category in the left plot and the multi Higgs-tag category in the right plot, the data is described well by the backgrounds over the whole range of the variables.

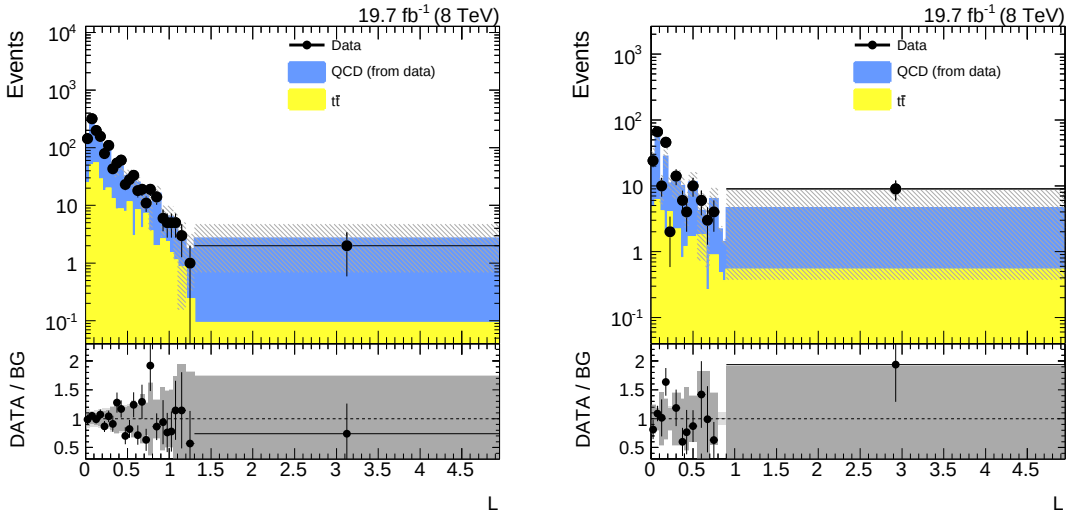

Figure 7.33: Distributions of the likelihood variable L in the single (left) and multi (right) Higgs-tag categories resulting from the statistical analysis.

The results of the statistical analysis are listed in table 7.17 for the individual nuisance parameters. In the theta framework, the prior distributions for all nuisance are normalized to have a central value of 0 and a width of ± 1 before the fit. The central values determined in the limit setting procedure are listed in the second column of table 7.17, the corresponding uncertainties in the last column. All of the obtained values are compatible with the prior distributions within uncertainties.

The expected and observed Bayesian upper exclusion limits at 95% confidence level on the cross section for T'-quark pair production under assumption of a branching fraction Br(T'$\rightarrow$ tHtH) = 100% can be found in figure 7.34. Better sensitivity is achieved for signal samples simulated with high T'-quark masses, since the analysis was optimized for events containing boosted top quarks and Higgs bosons. These are more likely to be produced in decays of very heavy T' quarks. The observed lower limit on the T'-quark mass of 745 GeV is slightly lower than the expected limit of 773 GeV because of a minor upward fluctuation of the data in the likelihood variable distribution for high T'-quark mass points. It is well covered by the uncertainties though. Over the entire examined mass range, the observed limit on the cross section does not exceed the 95% confidence level band around the median expected limit.

Nuisance parameter	Post-fit value	Post-fit σ
$t\bar{t}$ cross section	-0.44	0.96
Q^2 scale	1.30	0.67
JEC subjets	-0.08	0.67
PDF	-0.16	0.77
b-tag SF	-0.38	0.88
b-mistag SF	-0.17	0.75
Top-tag SF	-0.030	0.83
Trigger weight	-0.023	0.75
Luminosity	-0.08	0.96
QCD rate single Higgs tag	1.36	0.57
QCD rate multi Higgs tag	1.39	0.43

Table 7.17: Post-fit values for the nuisance parameters.

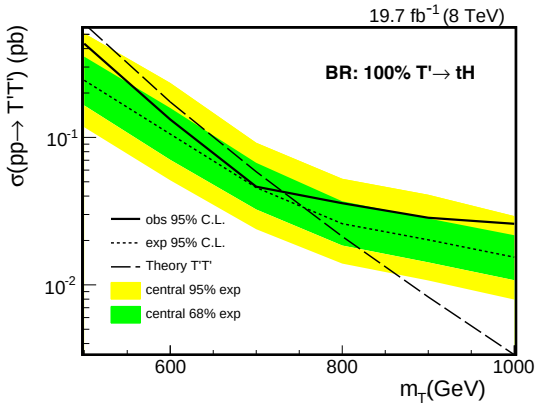

Figure 7.34: Expected and observed upper exclusion limits at 95% confidence level on the cross section for T'-quark pair production. The limits are derived from the likelihood variable L in simultaneous analysis of events in the single and multi Higgs-tag categories. The median expected limit is depicted by the dotted black curve. The 68% and 95% confidence level bands for the expected limit are drawn in green and yellow, respectively. The solid black curve corresponds to the observed limit. The theory prediction for the production cross section is represented by the dashed black line.

7.6.2 Results for all possible branching fractions

The vector-like T' quark does not decay exclusively to top quarks and Higgs bosons. The decays T'→ tZ and T'→ bW are allowed as well. Therefore, also results for other signal compositions are of interest. A scan of all possible branching fractions is performed. The results are presented in this section. For this scan, the branching fractions Br(T'→ tH), Br(T'→ tZ), and Br(T'→ bW) are varied simultaneously. As only these three decay modes are allowed for vector-like T' particles, the three branching fractions must always add up to one:

$$\mathrm{Br(T' \to tH)} + \mathrm{Br(T' \to\ tZ)} + \mathrm{Br(T' \to\ bW)} = 1. \tag{7.6}$$

As documented in section 7.3, also non-negligible parts of the signal samples simulated for other decay modes than T' → tH pass the event selection of this analysis. Therefore, the analysis can be expected to be sensitive not only to decays to top quarks and Higgs bosons, but also to other final state compositions. Figure 7.35 shows the distributions of selected signal events for the H_T and Higgs-candidate mass variables.

The selected events are split into the single and multi Higgs-tag categories. Each colored line corresponds to a signal sample with a different composition of final states. The distributions are scaled to the number of selected events predicted for each sample by the simulation. The selection efficiency clearly decreases for samples with final-state compositions that are less similar to those of the T'T' → tHtH sample. Events with final states containing top quarks and Higgs or Z bosons are selected more frequently than those containing bottom quarks and W bosons. However, the shapes of the distributions are rather similar for all final states examined in this study. Especially in the H_T distributions, only small shape variations are observed. In the Higgs-candidate mass distribution, a peak is found between 100 and 120 GeV for samples containing T' → tH decays, which is shifted to lower masses for samples containing only decays of T' → tZ or T' → bW. These events can pass the event selection, if a Z- or W-boson decay is misidentified as a Higgs-boson decay. Only one likelihood variable per T'-quark-mass hypothesis is computed using the T'T' → tHtH samples as input. In order to save computation time, this variable is used to classify the events of all other signal samples as well. Because of the overall small changes in shape between the different signal samples, no large gain in sensitivity is expected if a dedicated likelihood ratio was derived for each signal composition.

To perform the actual scan of all possible combinations of branching fractions, the shapes of the likelihood distributions found for the different signal samples are combined accordingly. The samples named T'T' → tHtH, T'T' → tZtZ, and T'T' → bWbW are simulated assuming 100% branching fraction to tH, tZ, or bW respectively. For the samples with mixed final states, e.g., the T'T'→ tHtZ sample, 50% branching fraction are assumed in the simulation for each of the contributing decay modes. This means that only 50% of these signal samples are made up of events where the T' quarks decay into different particles, while the other half is also made up of events in which both T' quarks decay exclusively into either tH or tZ. This caveat needs to be considered when computing the weights with which the different signal samples contribute to a given combination of branching fractions.

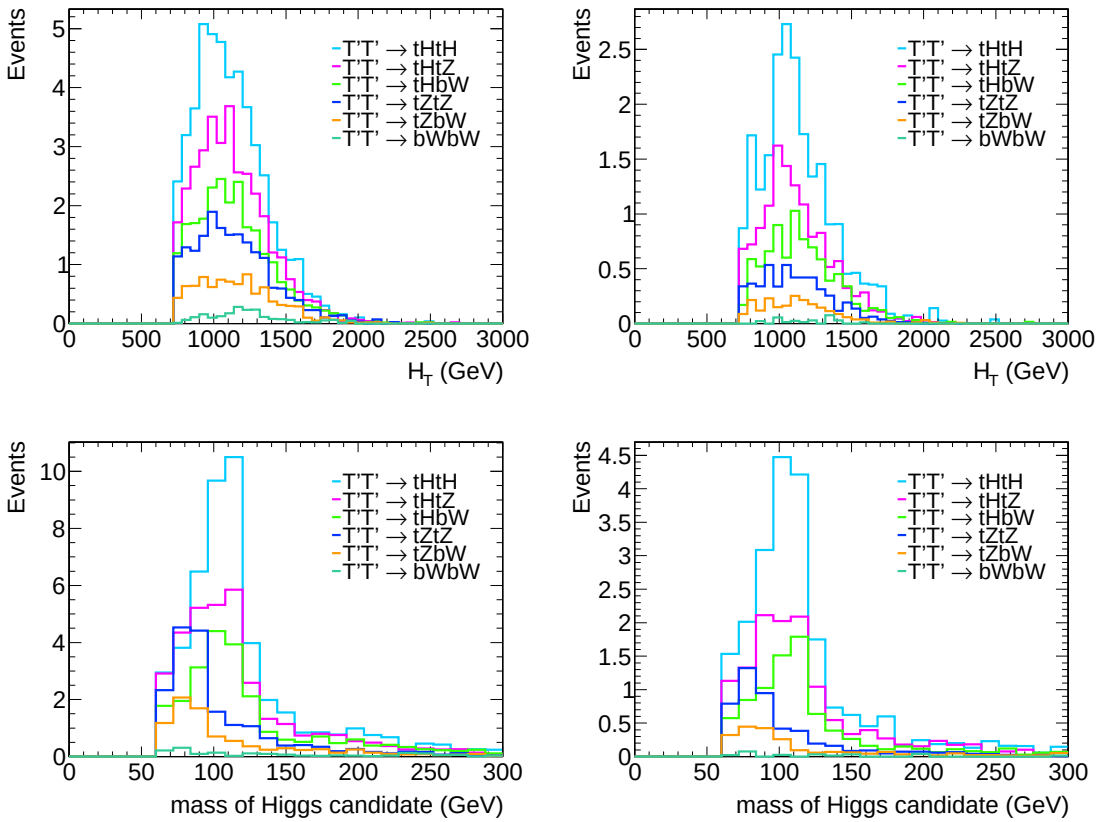

Figure 7.35: Comparison of shape and rate of the distributions of selected signal events for samples that were simulated assuming different branching fractions in the decay of the T' quarks. The results for the single Higgs-tag category are shown in the left plots, those for the multi Higgs-tag category in the right ones. The H_T distributions are displayed in the top row, the distributions of the Higgs-candidate mass in the bottom row.

The three branching fractions are varied from 0 to 1 in steps of 0.01. For each point in this scan of branching fractions that fulfills equation 7.6, expected and observed exclusion limits for the cross section are computed for each of the six T'-quark mass hypotheses. The resulting expected and observed exclusion limits for the cross section for T'-quark pair production are shown in figures 7.36-7.38, for signal samples simulated with a T'-quark mass of 500 GeV, 700 GeV, and 1000 GeV. Each point in the triangles corresponds to a certain combination of branching fractions. The branching fraction Br(T' $\rightarrow$ tH) is plotted on the x-axis, while the y-axis gives the value of the branching fraction Br(T' $\rightarrow$ tZ). The corresponding value of the third branching fraction Br(T' $\rightarrow$ bW) can be calculated using equation 7.6. Dark blue colors correspond to strong cross section limits, light red colors to weaker limits. The used color scale is not the same in all of the plots. As expected, the most stringent limits can be set for signal compositions resulting in many decays to top quarks and Higgs bosons, because the analysis was optimized for these type of decays. The results for these samples can be found in the bottom right corner of the triangular plots. Also, for sample compositions with non-negligible top-quark and Z-boson contributions in the final state, good sensitivity is achieved. Overall, more stringent cross section limits are set when assuming large T'-quark masses. This is due to the higher Lorentz boost of the daughter particles of the T' quark, that improves the performance of the jet-substructure methods used in the event selection.

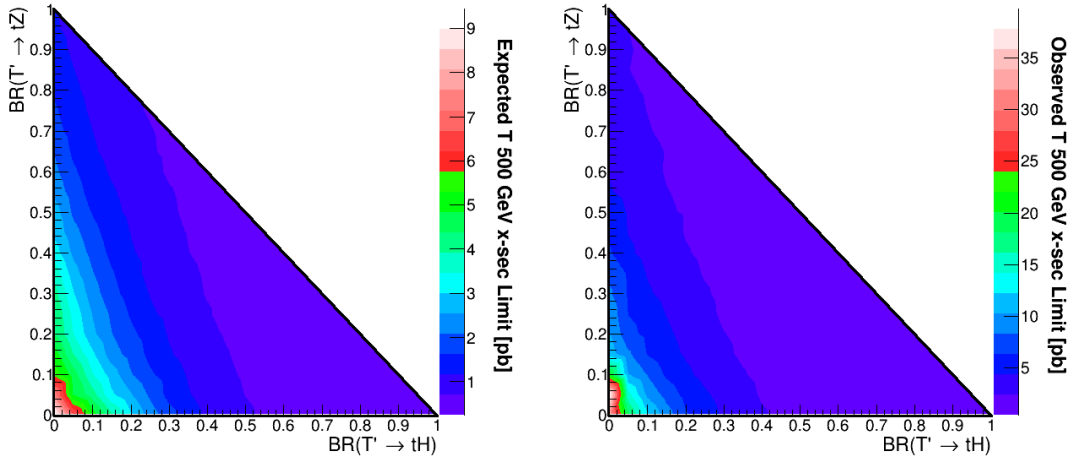

Figure 7.36: Expected (left) and observed (right) upper exclusion limits for the cross section for T'-quark pair production at 95% confidence level for any possible branching fraction assuming a T'-quark mass of 500 GeV. The branching fraction to bottom quarks and W bosons is given by Br(T' $\rightarrow$ bW) = 1 - Br(T' $\rightarrow$ tH) - Br(T' $\rightarrow$ tZ).

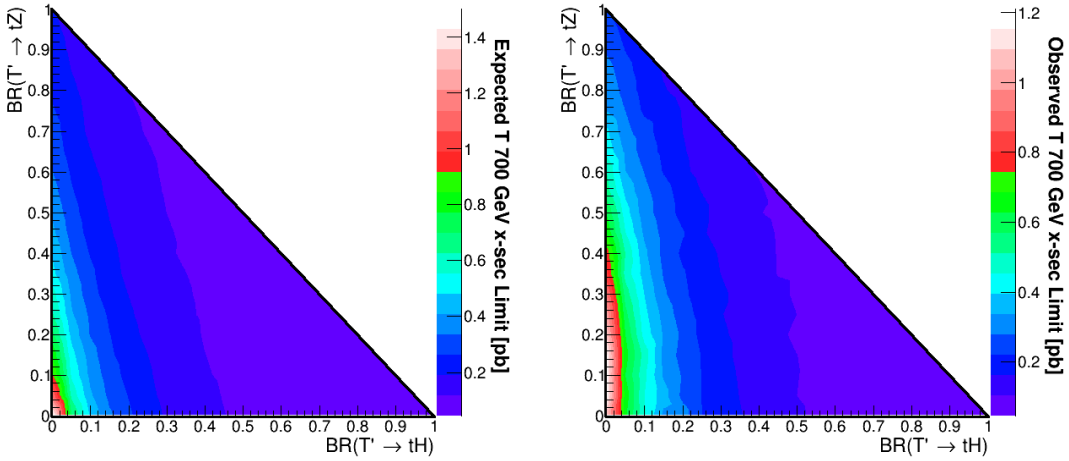

Figure 7.37: Expected (left) and observed (right) upper exclusion limits for the cross section for T'-quark pair production at 95% confidence level for any possible branching fraction assuming a T'-quark mass of 700 GeV. The branching fraction to bottom quarks and W bosons is given by Br(T' → bW) = 1 - Br(T' → tH) - Br(T' → tZ).

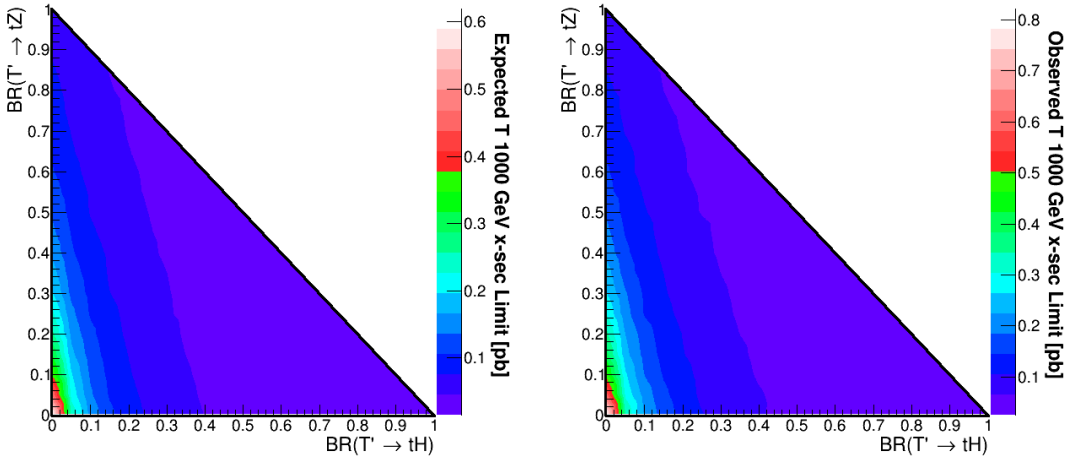

Figure 7.38: Expected (left) and observed (right) upper exclusion limits for the cross section for T'-quark pair production at 95% confidence level for any possible branching fraction assuming a T'-quark mass of 1000 GeV. The branching fraction to bottom quarks and W bosons is given by Br(T' → bW) = 1 - Br(T' → tH) - Br(T' → tZ).

Furthermore, lower exclusion limits for the T'-quark mass are set at 95% confidence level for each point in the branching fraction scan. They are obtained in the same way as in the case of Br(T' → tH) = 100% described in the previous section: the mass limit is marked by the crossing point of the measured cross section limit curve and the curve describing the theoretically predicted cross sections. The thus obtained mass limits are also reported using triangular plots in which each point corresponds to a certain mix of branching fractions. The triangular plot showing the expected mass limits can be found on the left-hand side of figure 7.39, the observed mass limits are shown in the plot on the right-hand side of the same figure. The most stringent mass limits are, again, obtained for signal samples containing a large fraction of T' → tH decays.

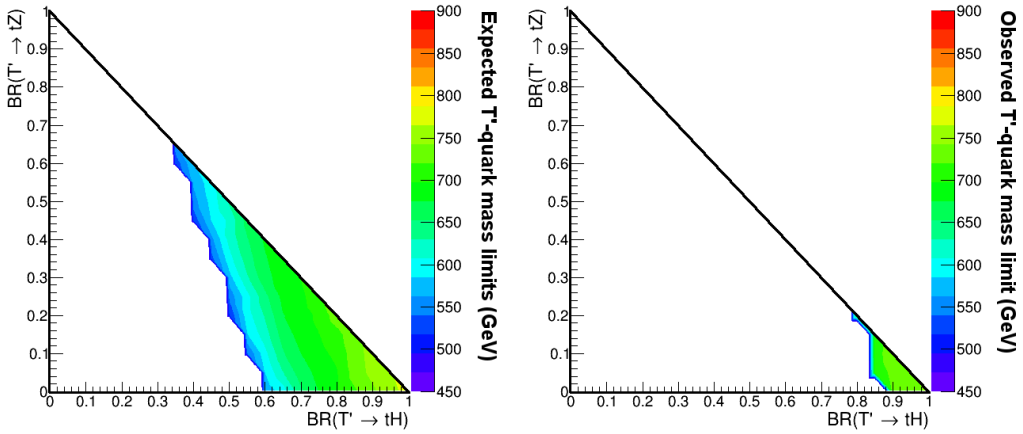

Figure 7.39: Expected (left) and observed (right) lower exclusion limits for the T'-quark mass for any possible combination of branching fractions. The branching fraction to bottom quarks and W bosons is given by Br(T' → bW) = 1 - Br(T' → tH) - Br(T' → tZ).

8 Combination with other searches for vector-like T' quarks

In addition to the analysis presented in the previous chapter, two other searches for pair produced T' quarks have been performed in the data recorded with the CMS experiment at $\sqrt{s} = 8$ TeV and published to date: an inclusive search in the single-lepton and multi-lepton channel [123], and a search for decays of T'T'$\rightarrow$ tHtH where the Higgs boson decays into two photons [140]. Neither of these analyses provide any evidence for the existence of T' quarks in the examined T'-quark mass range. The results of these two analyses are combined with the previously described search in the all-hadronic channel that is optimized for decays of T'T'$\rightarrow$ tHtH. The combination of the three analyses yields the possibility to further improve the mass and cross-section exlcusion limits for pair-produced T' quarks.

In section 8.1, a brief introduction is given into the analysis strategies of the searches for pair-produced T' quarks in the CMS data recorded at $\sqrt{s} = 8$ TeV that are combined with the search described in chapter 7. More detailed descriptions of these analyses can be found in [123, 140]. In section 8.2, the details of the actual statistical combination are explained and the results of the combination are presented.

8.1 Overview of other CMS searches for vector-like T' quarks at $\sqrt{s} = 8$ TeV

8.1.1 Inclusive single-lepton analysis

The single lepton analysis uses a multivariate approach to distinguish the T'-quark signal from background processes [123]. Events with a single electron or muon with $p_T > 32$ GeV are analyzed in two categories: events containing a CA8 jet that can be W tagged, and events that contain no such jet. A boosted decision tree (BDT) is trained combining the discriminating power of a number of variables. Amongst these are several variables obtained in the analysis of jet substructure, such as the multiplicity of W and top tags. More information on W tagging and the used CMS top tagger algorithm are provided in section 5.2.5.3. A separate BDT is trained for each considered T'-quark mass hypothesis. BDT discriminator distributions are obtained separately for the W-tag and no-W-tag event categories in both, the electron and muon channel. These four distributions are used in the limit setting procedure.

8.1.2 Inclusive multi-lepton analysis

The multi-lepton analysis introduces three categories for events containing more than one electron or muon: the same-sign dilepton category, the opposite-sign dilepton category, and the trilepton category [123]. Each category is sensitive to a different decay channel of the pair of T' quarks. Two different event selections are developed for the opposite-sign dilepton category. The first is optimized for decays of T'T' $\rightarrow$ bWbW, the other for decays

of T'T' → tZtZ. The same-sign dilepton and trilepton categories mainly contain events in which at least one of the T' quarks decays to a top quark and a Higgs or Z boson. No decays of T'T' → bWbW are expected in these categories. Counting experiments are performed in all of the event categories to derive limits on the T'-quark mass.

8.1.3 T'T' → tHtH (H → $\gamma\gamma$) analysis

This analysis is also optimized for decays of both T' quarks to top quarks and Higgs bosons. Here, the Higgs bosons do not decay hadronically into bottom-quark pairs, but into two photons [140]. This allows for a very precise reconstruction of the Higgs-boson mass, making this variable a good handle to discriminate signal from background. A disadvantage in this analysis approach is the relatively low branching fraction Br(H → $\gamma\gamma$) which results in a rather low signal-selection efficiency.

In this analysis, events containing two photons and two jets are classified as leptonic, if one or more leptons are found in the event. Hadronic events have to contain at least one b-tagged jet in order to be taken into account in the analysis. Only events with a large hadronic activity are selected by requiring $H_T \geq 1000$ GeV in hadronic events and $H_T \geq 770$ GeV in leptonic events. Counting experiments are then performed in the leptonic and hadronic event categories simultaneously.

8.2 Combination

Different analyses can only be combined easily into a common statistical analysis if there is no overlap in the selected data events. No such overlap is found for the three T'-quark searches that are combined in this study. This means that no additional event selection criteria, such as, e.g., lepton vetoes, need to be applied in any of the analyses. The single-lepton and multi-lepton analyses are combined into a single leptonic analysis in the following.

All sources of systematic uncertainties that are taken into account in the statistical combination are listed in table 8.1. Some of these are shared between the different analyses, e.g., the Q^2 scale uncertainty for simulated $t\bar{t}$ background events or the uncertainties in the jet energy resolution. They are therefore treated as fully correlated between the analyses. Some analyses expect contributions of the same background processes. For those backgrounds that are described by Monte Carlo simulation, the same samples are used by all analyses. Details on the sources of systematic uncertainty that are relevant for the all-hadronic T'T' → tHtH anlaysis are provided in section 7.5. The sources of systematic uncertainty that are specific to the H → $\gamma\gamma$ and the leptonic analyses can be found in the corresponding documentation [123, 140].

Bayesian upper exclusion limits on the cross section for pair production of T' quarks are derived using the theta framework as described in section 6 taking into account all of the nuisance parameters listed in table 8.1.

	T'T'→tHtH (all-hadronic)	T'T'→tHtH (H→ $\gamma\gamma$)	Inclusive leptonic
$t\bar{t}$ matching	-	-	✓
$t\bar{t}$ Q^2 scale	**C**	-	**C**
b-tag SF	**C**	-	**C**
b-tag mistag SF	✓	-	-
top-tag SF	✓	-	-
Photon ID	-	✓	-
Photon energy scale	-	✓	-
Photon energy resolution	-	✓	-
Jet energy resolution	**C**	**C**	**C**
Jet energy scale	**C**	**C**	**C**
Lepton ID	-	**C**	**C**
Luminosity	**C**	**C**	**C**
PDF $t\bar{t}$	**C**	**C**	-
Pileup jet ID	-	✓	-
Data driven background estimate	✓	✓	✓
Trigger	✓	✓	✓
$t\bar{t}$ cross section	**C**	-	**C**
PDF $t\bar{t}H$	-	✓	-
Vertex efficiency	-	✓	-

Table 8.1: Nuisance parameters considered in the combination of the three analyses. Nuisance parameters that are taken into account by an analysis are marked with a ✓ symbol. A bold capital **C** denotes uncertainties that are treated as correlated between different analyses.

In figure 8.1, the expected upper exlcusion limits at 95% confidence level on the cross section for T'-quark pair production are shown. A comparison is made between the limits that are obtained from the individual analyses and those from the combination of all analyses. All of the limits shown in this plot are computed assuming a branching fraction Br(T' → tH) of 100%.

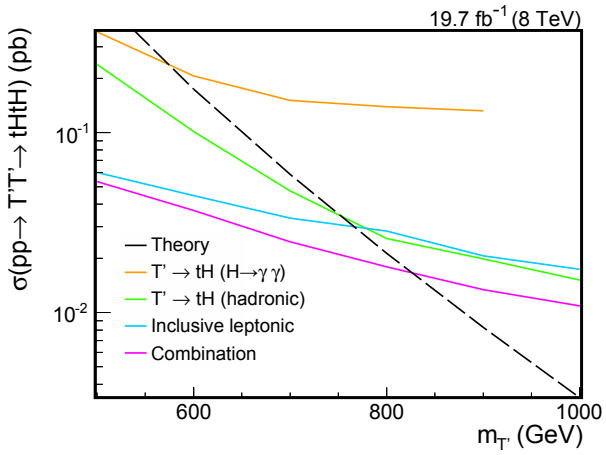

Figure 8.1: Expected upper exclusion limits at 95% confidence level on the cross section for T'-quark pair production assuming Br(T' → tH) = 100%. Different colors show the limits resulting from the three individual analyses. The result of the combination of all analyses is shown in magenta, that of the all-hadronic T' → tH search in green. The blue line shows the result of the search with leptons, the orange curve that of the search in events with H→ $\gamma\gamma$ decays.

The analysis in the all-hadronic channel optimized for decays of T'T' → tHtH described in chapter 7 is the most sensitive in the high T'-quark mass range above approximately 750 GeV. For lower masses it is surpassed in sensitivity only by the inclusive combination of the two leptonic analyses. The analysis with two photons in the final state is also designed for final states with top quarks and Higgs bosons specifically. It does not reach the same sensitivity as the all-hadronic T'T' → tHtH analysis though. This is due to the much smaller branching fraction for H → $\gamma\gamma$ decays. This analysis is expected to give a more important contribution in the analysis of the larger data sets that will be recorded at $\sqrt{s}$ = 13 TeV in future LHC operation. The very clean final state of this analysis channel will be of advantage handling the expected large pileup contamination in these events. The combination of all three analyses improves the sensitivity over the full mass range compared to the individual analyses. A lower exclusion limit on the T'-quark mass of 843 GeV is expected when taking into account the results of all searches.

The expected and observed exclusion limits on the cross section for T'-quark pair production at 95% confidence level obtained from the combination of all analyses for Br(T' → tH) = 100% are displayed in figure 8.2. The observed limit is slightly worse than the expected limit over the full mass range. However, it does not exceed the 95% confidence level band around the expected limit, that is drawn in yellow in this plot. Due to a upward fluctuation of the observed cross section limit at a T'-quark mass of 800 GeV, the observed mass limit of 767 GeV is lower than the expected mass limit of 839 GeV.

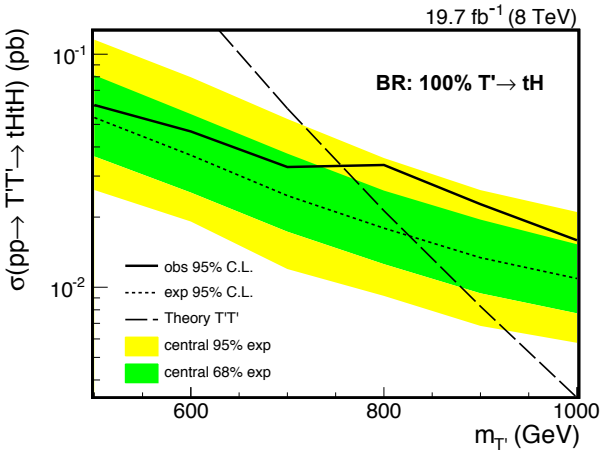

Figure 8.2: Expected and observed limits at 95% confidence level on the cross section for T'-quark pair production obtained in the combination of all analyses. The median expected limit is depicted by the dotted black curve. The 68% and 95% confidence level bands for the expected limit are drawn in green and yellow, respectively. The solid black curve corresponds to the observed limit. The theory prediction for the production cross section is represented by the dashed black line.

Figure 8.3 shows the central values and 1-σ intervals of the posterior distributions obtained in the statistical evaluation for all considered nuisance parameters. All of these post-fit values are compatible with the prior distributions assigned to the nuisance parameters within uncertainties.

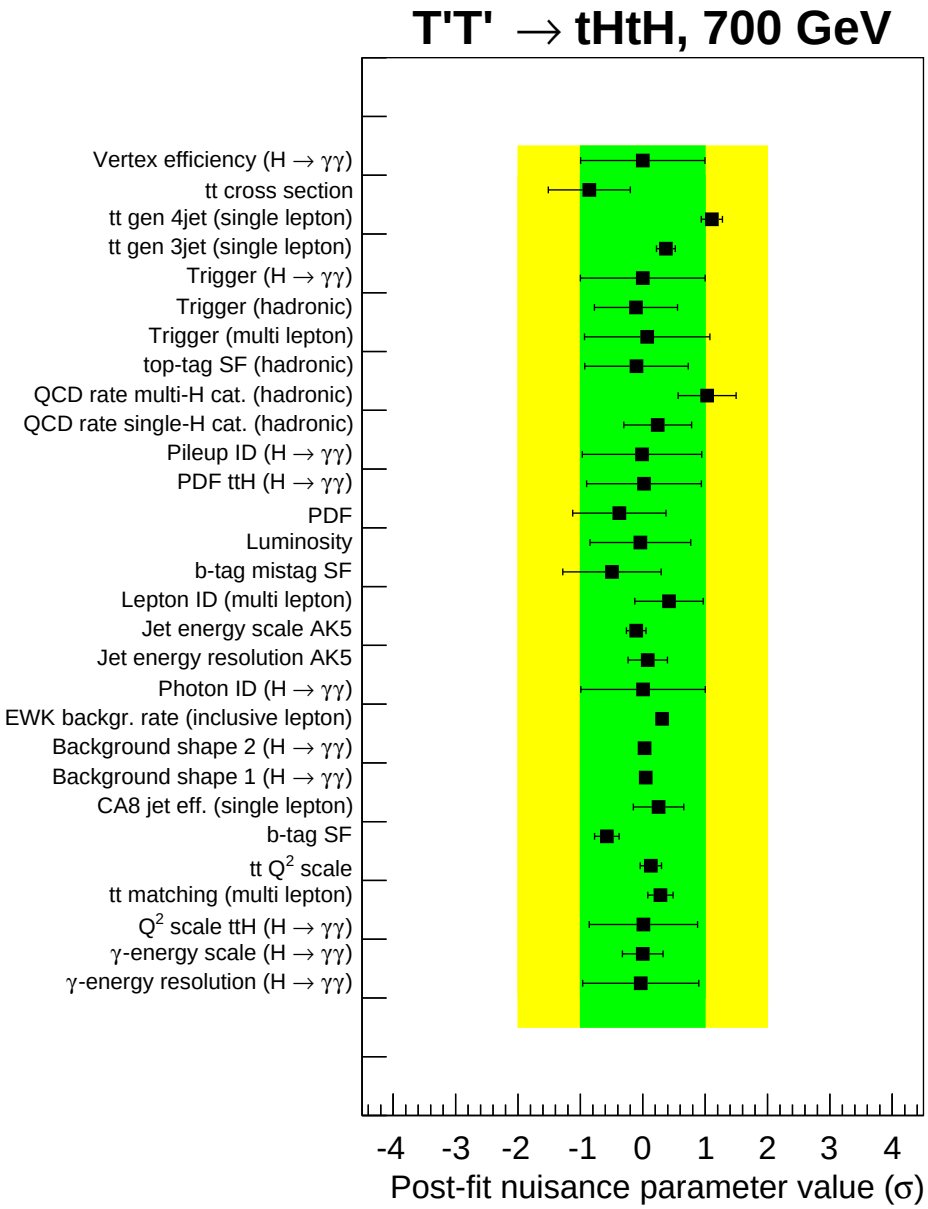

Figure 8.3: Central values and 1-σ intervals of the posterior distributions determined for the nuisance parameters in the statistical analysis performed simultaneously for all three analyses. The 1- and 2-σ intervals of the prior distributions are drawn as green and yellow bands.

A scan of all possible branching fractions for the T'-quark decays as described in section 7.6.2 is performed also in this combination of the three analyses. All possible branching fractions for T'-quark decays are scanned. The results obtained in this scan can be found in figure 8.4.

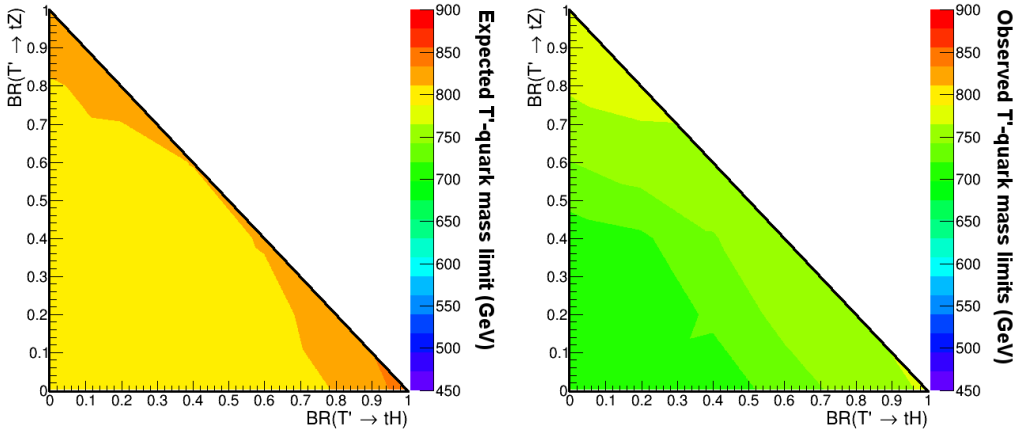

Figure 8.4: Expected (left) and observed (right) lower exclusion limits at 95% confidence level on the T'-quark mass for any possible combination of branching fractions. The branching fraction to bottom quarks and W bosons is given by Br(T' → bW) = 1 - Br(T' → tH)- Br(T' → tZ).

The expected lower exclusion limits at 95% level on the T'-quark mass are shown in the left plot, the corresponding observed limits in the right plot of the figure. When combining the different analyses, good sensitivity is achieved in the whole range of the triangle. The observed mass limits range from 697 GeV, for Br(T'→ tZ) = 20% and Br(T'→ bW) = 80%, to 782 GeV for Br(T'→ tZ) = 100%. These are the most stringent limits on the T'-quark mass set in CMS analyses to date. The best expected T'-quark mass limit of 839 GeV is achieved for Br(T'→ tH) = 100%.

9 Outlook to future analyses with vector-like quarks

In this chapter, analysis possibilities promising new insight about vector-like quarks in the future are described. All of the previously described analyses are searching for pair-produced T' quarks at a center-of-mass energy of 8 TeV. New, complimentary approaches can be the search for singly produced vector-like T' quarks and searches in data recorded at higher center-of-mass energies of 13/14 TeV in future LHC runs. First feasibility studies for a search for singly produced T' quarks at $\sqrt{s} = 8$ TeV and pair-produced T' quarks at 13 TeV are presented below. In these studies a branching fraction Br(T' $\rightarrow$ tH) = 100% is assumed.

9.1 Searching for single production of vector-like T' quarks

At the moment, analyses searching for singly produced vector-like quarks are still in the development stage, no results have been published to date. The previous searches for pair-produced T' quarks provide mass limits of at least 697 GeV for any decay mode of the T' quark (see chapter 8). As shown in figure 9.1, assuming a T'-quark mass larger than approximately 700 GeV, the cross section for single T'-quark production is of the same order of magnitude or even larger than the cross section for pair production of the particle.

In the study described below, the focus is set on the shape differences between distributions of single T'-quark signal and background events. The analysis setup is identical to that used for the search for pair-produced T' quarks described in chapter 7. Only one alteration is made to the event selection described in section 7.3: the selected events are not split into categories according to Higgs-tag multiplicity, as only a single Higgs-jet per event can be expected in events with singly produced T' quarks.

The signal samples are simulated using the MadGraph Monte Carlo event generator interfaced with Pythia for simulation of particle showering. Three signal samples are produced, setting the mass of the T' quark to three different values: 700 GeV, 1000 GeV, and 1200 GeV. In the simulation, the branching fraction of the T' quark is set to Br(T' $\rightarrow$ tH) = 100%. No restriction is made with respect to the decay modes of the produced top quarks and Higgs bosons. The $t\bar{t}$ and QCD-multijet background contributions are also modelled with Monte Carlo generators. Here, the samples listed in section 7.2 are used.

Table 9.1 summarizes the impact of the different selection criteria on the three simulated signal samples. The percentages quoted in the table are calculated with respect to the total number of generated events before any selection criteria. The most striking effect comes from the high threshold of $H_T^{calo} > 750$ GeV in the used trigger. While large fractions of samples generated with high T'-quark masses pass the trigger threshold, the sample generated with a T'-quark mass of 700 GeV is reduced by more than a factor of 2.

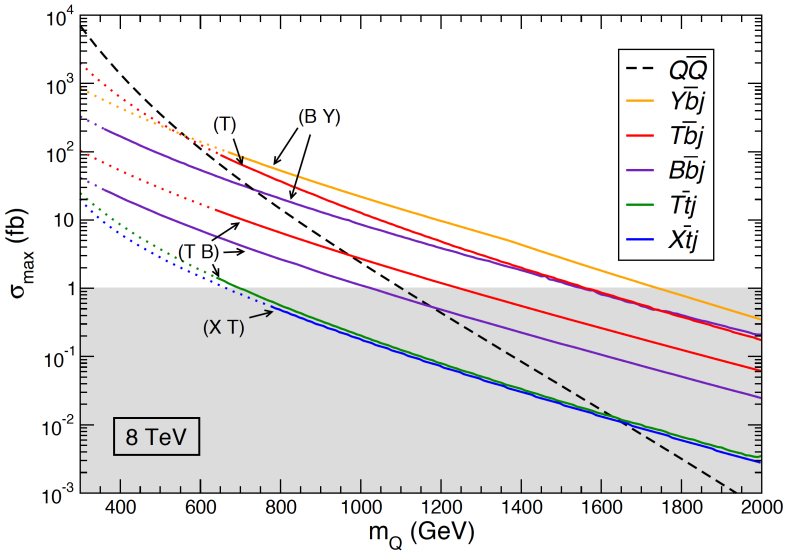

Figure 9.1: Maximum production cross sections for vector-like quarks at the LHC at a center-of-mass energy of 8 TeV [50]. The black dotted line corresponds to the cross section for pair production, which is identical for all types of vector-like quarks. The cross sections for single production of the different quarks are drawn as colored lines. The shaded area marks cross sections below 1 fb, which are out of reach for analyses of the collected approximately 20 fb^{-1} of data at $\sqrt{s}$ =8 TeV.

$m_{\text{T'}}$ (GeV)	Trigger	$\geq$ 2 CA15 jets	$\geq$ 1 HTT	$\geq$ 1 HTT + subjet b-tag	Higgs tag	$H_T >$ 720 GeV
700	47.5	46.8	13.6	9.0	1.6	1.2
1000	80.2	78.6	23.7	15.3	2.9	2.7
1200	88.3	86.4	25.2	15.5	2.9	2.9

Table 9.1: Fraction of events passing the event selection after each selection step in percent assuming a branching fraction of Br(T' $\rightarrow$ tH) = 100%. These percentages are calculated with respect to the full number of simulated events before application of any selection criteria. Values are predicted by the Monte Carlo simulation for three different signal samples containing singly produced T' quarks. The abbreviation HTT refers to HepTopTagged jets.

Figure 9.2 shows the H_T and Higgs-candidate mass distributions, that are used for discrimination between background and signal events in the search for pair-produced T' quarks. The distributions for both backgrounds are obtained from events simulated using Monte Carlo techniques. As explained in section 2.2.4, the cross section for single production of vector-like quarks is strongly model dependent. The signal distributions in figure 9.2 are scaled to an arbitrary number of events for this shape comparison.

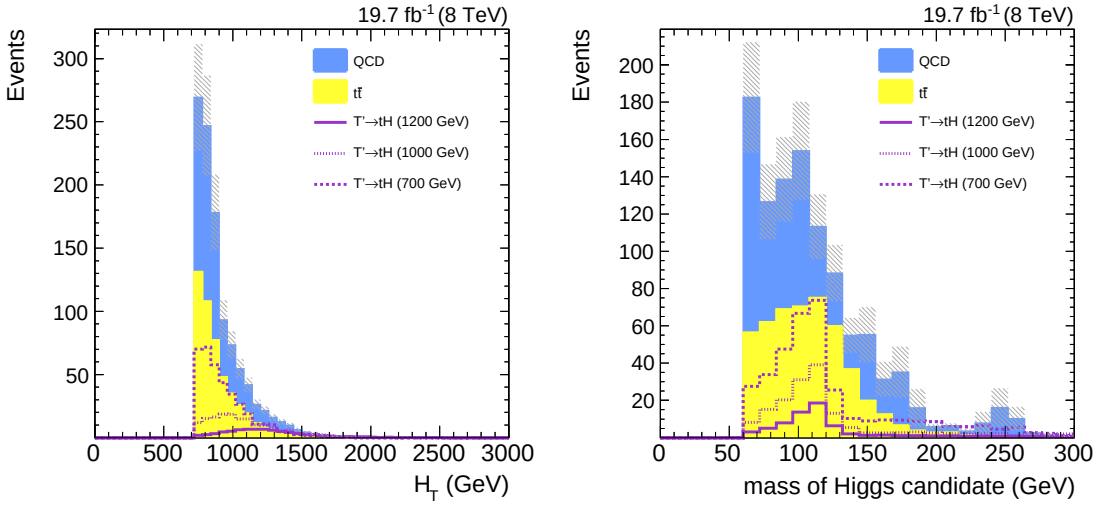

Figure 9.2: The H_T (left) and Higgs-candidate mass (right) variables as described in chapter 7. The colored histograms show the expected background estimated using Monte Carlo simulation. The hashed gray areas represent the statistical uncertainty of the expected background contribution. The expected contributions from single T'-quark prodcution are taken from Monte Carlo simulation. The samples are produced assuming three different T'-quark masses. The signal distributions are drawn as the violet lines. Their normalization is arbitrary.

For high T' quark masses, the H_T distribution shown in the left plot of figure 9.2 is also suitable to discriminate signal events due to single T'-quark production from the background. For a T'-quark mass of 1000 or 1200 GeV, the shape of the distribution of signal events is clearly different from that of the two background contributions. Here, the effect of the high H_T^{calo}-trigger threshold of 750 GeV becomes apparent once more: for the sample produced with a T'-quark mass of 700 GeV, the resulting H_T distribution is clearly truncated by the selection requirement of $H_T > 720$ GeV. This offline H_T selection is introduced to select a phase space region with good trigger efficiency. As there is only a single T'-quark decay contributing to the H_T in the samples used in this study, the smaller value of H_T with respect to events containing pair-produced T' quarks of the same mass is expected. The truncation of the H_T distribution compromises the event selection efficiency heavily for this sample. As discussed in the beginning of section 7.3, a relaxation of this requirement is not possible when using the same trigger as in the analysis described in chapter 7, though. Therefore, use of the H_T variable in the statistical evaluation can only lead to good sensitivity in searches for singly produced T' quarks with a mass larger than 700 GeV. To exploit the sensitivity of the H_T variable in an optimal way, a different trigger

should be used for searches of singly produced T' quarks.

The right plot in figure 9.2 shows the distribution of the Higgs-candidate mass, which is the invariant mass of the two b-tagged subjets in the Higgs-candidate jet. This variable yields very good discrimination power for all of the three examined signal samples. The peak around the expected Higgs boson mass of about 125 GeV is reconstructed with a better resolution than in the simulation of pair-produced T' quarks shown on the righthand-side of figure 7.11. In case of single production of T' quarks, the probability to correctly identify the Higgs jet from the T'-quark decay is larger than in the analysis of the pair-produced quarks, as there are not as many particles in the final state. This improved resolution is expected to have a positive effect on the sensitivity of the analysis.

The final state resulting from singly produced T' quarks is more simple compared to that found in events containing pair-produced T' quarks. This can be used to improve the sensitivity of the analysis. Figure 9.3 shows the invariant mass of the Higgs- and top-candidate CA15 jets. For pair-produced T' quarks, ambiguities can arise in the assignment of the top- and Higgs-candidate jets to one of the T'-quark decays. This is not the case for signal from singly produced T' quarks. The distribution of the reconstructed T'-quark mass in signal events has a rather narrow peak very close to the T'-quark mass that was assumed in the Monte Carlo simulation of the samples.

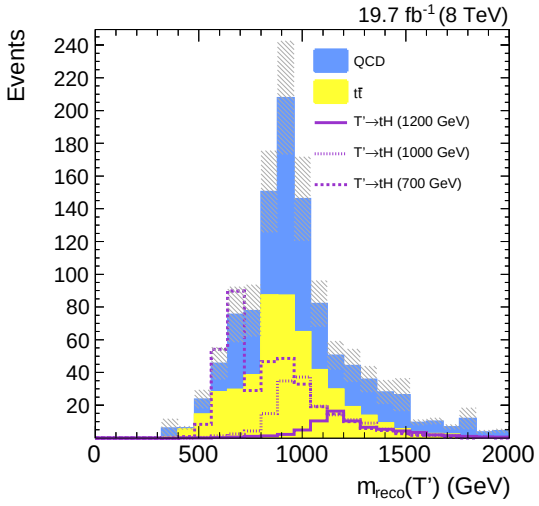

Figure 9.3: The invariant mass of the Higgs- and top-candidate jets. The colored histograms show the expected background estimated using Monte Carlo simulated samples. The hashed gray areas stand for the statistical uncertainty of the expected background contribution. The expected contributions from single T'-quark prodcution are taken from Monte Carlo simulation. The samples are produced assuming three different T'-quark masses. The signal distributions are drawn as the violet lines. Their normalization is arbitrary.

In the search for pair-produced T' quarks, the sensitivity improves greatly when splitting the selected events into two categories according to their Higgs-tag multiplicity. This is not an option in case of a search for single T'-quarks. Only one Higgs boson is expected in events with singly produced T'-quarks decaying to top quarks and Higgs bosons. In events containing T'T'→ tHtH decays, two or more Higgs-tagged jets are found in 28-32% of the selected events. In simulation of singly produced T' quarks, only 1-2% of the events passing the full event selection can be found in the multi Higgs-tag category. Therefore, this categorization will not improve the sensitivity of the search for single T'-quark production.

In the method for the estimation of the QCD-multijet background described in section 7.4, the shapes of distributions are obtained from a sideband region that does not contain a top-candidate jet by definition. Therefore, this method is not suitable for description of the distribution of the reconstructed T'-quark mass in QCD-multijet events. If this variable was to be used in an analysis, a new sideband region, in which all distributions are described properly, would have to be identified to model the QCD-multijet background in a data driven way. For these first studies, the distributions of QCD-multijet background events are taken from simulated events.

Overall, these studies of the single T'-quark search channel promise good sensitivity. In order to adapt the analysis described in chapter 7 in the search for singly-produced T' quarks decaying to top quarks and Higgs bosons, a number of modifications must be made to ensure good sensitivity to the signal. For example, a different trigger strategy has to be developed, and the offline H_T selection criterion needs to be relaxed. In order to use the well-discriminating reconstructed T'-quark-mass variable in the statistical analysis, a different data-driven model for the QCD-multijet background has to be obtained. The findings of the already conducted search for pair-produced T' quarks can be valuable input for the design of a search strategy dedicated to single T'-quarks.

9.2 Prospects of searches for vector-like T' quarks at 13 TeV

In the next data taking period at the LHC starting in the first half of 2015, the center-of-mass energy will be increased from 8 to 13 TeV. Numerous studies are performed by the experimental collaborations to estimate the performance and analysis prospects at this higher center-of-mass energy. A number of aspects have to be considered to estimate the potential of future searches for vector-like quarks. They are outlined in the following.

Figure 9.4 shows the production cross sections for vector-like quarks at a center-of-mass energy of 13 TeV in different signal models. They are enhanced for both, pair and single production of the particles, with respect to the values provided for $\sqrt{s} = 8$ TeV in figure 9.1 by about a factor of 10. For models in which the vector-like T' quarks appear as singlets the single production becomes dominant over the pair production already at about 700 GeV. The same is true for vector-like quarks that are part of a (B, Y) doublet. As T' quarks with lower masses have already been excluded in the searches conducted at $\sqrt{s} = 8$ TeV, combined searches for single and pair production of the particles in higher mass ranges will be of particular interest in future analyses.

Not only the cross section of the different signal processes, but also the cross section for the production of background events increases with the larger center-of-mass energy. For $t\bar{t}$ production, one of the main backgrounds to T'-quark production, the cross section

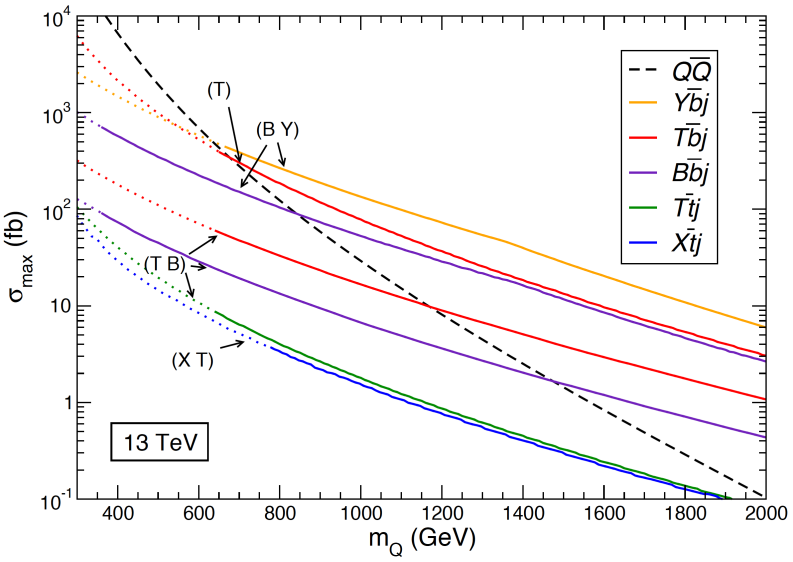

Figure 9.4: Maximum production cross sections of vector-like quarks at the LHC at a center-of-mass energy of 13 TeV [50]. The black dotted line corresponds to the cross section for pair production, which is identical for all types of vector-like quarks. The cross sections for single production of the different quarks are drawn as colored lines.

increases by approximately a factor 3 from 245.8 pb to 806.1 pb according to simulation [134]. For pair production of T' quarks, the increase in production cross section is predicted to be more pronounced though, as seen in the plots in figure 9.1. This can be expected to lead to an improved signal to background rate in future analyses.

The search for pair-produced T' quarks decaying to top quarks and Higgs bosons in the fully hadronic channel described in chapter 7 is found to yield better signal selection efficiencies for large T'-quark mass, because the used jet-substructure tools perform better in topologies with large Lorentz boosts. The Lorentz boost of the top quarks and Higgs bosons in the signal events depends on the mass of the decaying T' quark. As the analyses are moving to higher mass ranges, the Lorentz boost is expected to become more pronounced.

In the plots in figure 9.5, the distributions of a number of variables are compared for simulated signal events generated with different T'-quark masses and center-of-mass energies. In these samples, the T' quarks are produced in pairs. The distributions obtained from the sample generated assuming a T'-quark mass of 1 TeV and a center-of-mass energy of 8 TeV are compared to those from a sample assuming the same T'-quark mass but a higher center-of-mass energy of 13 TeV. This comparison illustrates the effect of the increased center-of-mass energy. The two complimentary orange and green curves show the distributions expected at a center-of-mass energy of 13 TeV for T' quark masses of 2 and 3 TeV, respectively. These allow for an estimation of the search potential in extremely boosted final states. All of the distributions displayed in figure 9.5 are normalized to unity for this shape comparison.

The plot on the top left of figure 9.5 shows the transverse-momentum distribution of all CA15 jets with a $p_T > 150$ GeV. Comparison of the two curves obtained from samples generated with a T'-quark mass of 1 TeV show, that an increase in center-of-mass energy hardly affects the p_T spectrum. Moving to larger T'-quark masses, the shape of the p_T distribution changes more significantly. While the maximum of the distributions is found in a similar p_T range, a much larger fraction of the jets with very large transverse momenta is found in the samples with high T'-quark mass.

The multiplicity of CA15 jets is displayed on the top right of figure 9.5. Overall, more CA15 jets are expected in events generated at $\sqrt{s} = 13$ TeV. This corresponds directly to the larger values of H_T observed when increasing the center-of-mass energy to 13 TeV, as illustrated in the bottom left plot. Also, larger values for the T'-quark mass result in harder H_T spectra. This is an indication that the H_T variable will be very suitable to identify signal events also in analyses of the data recorded in the future LHC runs.

In the bottom right plot, the multiplicity of CA15 jets tagged by the HEPTopTagger algorithm is shown. No large difference is observed between the two samples generated with a T'-quark mass of 1 TeV at the two different center-of-mass energies. This corresponds to the similarity of the p_T spectra obtained from these two samples. For the samples with higher T'-quark mass, less jets are tagged by the HEPTopTagger algorithm though. The HEPTopTagger algorithm is optimized for the identification of hadronic top-quark decays in the moderately boosted regime, corresponding to values of the fat-jet p_T between 200 and 400 GeV. A much smaller fraction of the CA15 jets in the samples generated with T'-quark masses of 2 and 3 TeV fall into this category. This suggests that other top-tagging techniques might be more suited for future analyses, for example the CMS Top Tagger algorithm which is described in section 5.2.5.3. This algorithm is found to be more efficient in the very boosted regime where jets with $p_T > 400$ GeV are observed [102].

While a number of new challenges arise for analyses of data recorded at higher center-of-mass energies, including larger numbers of pile-up events and increased background event rates, overall the increased energy is expected to positively affect the search potential for vector-like quarks. Especially the increased signal-production cross section is an advantage. Further interesting insight into the vector-like quark sector can surely be expected from analyses of the 13 TeV data.

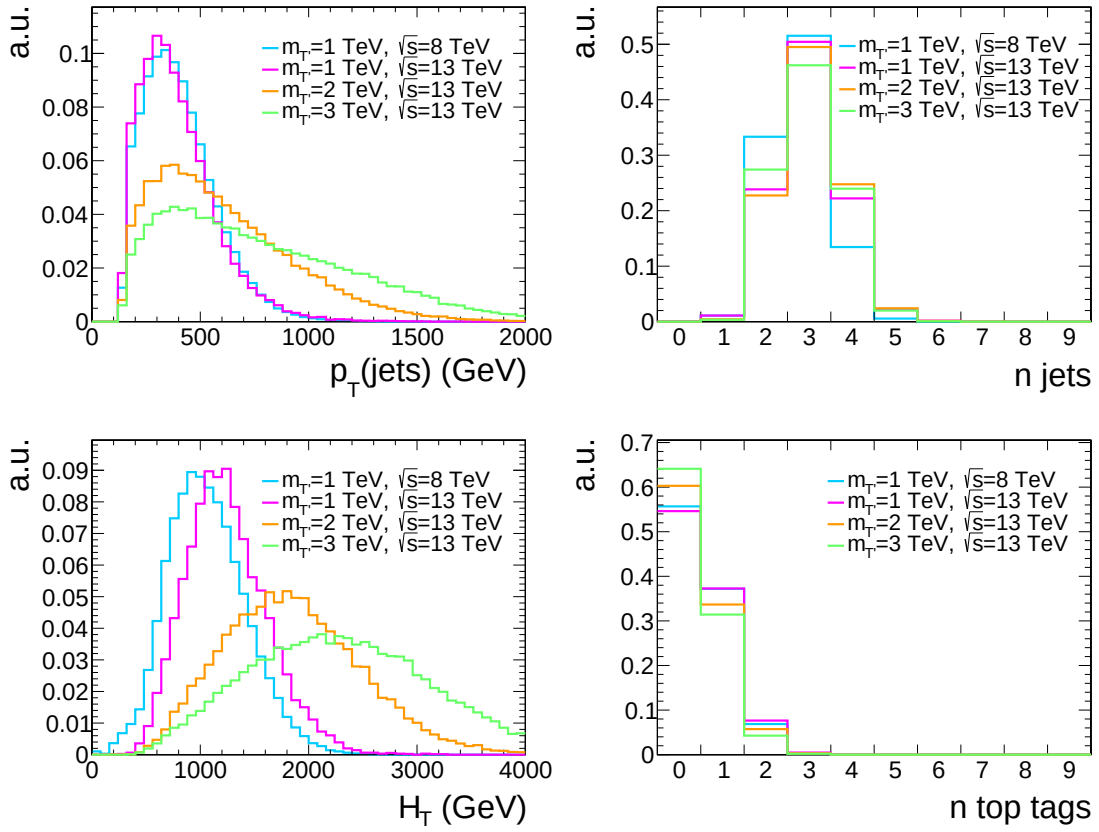

Figure 9.5: Shape comparison of the p_T of the CA15 jets (top left), the CA15 jet multiplicity (top right), the H_T (bottom left), and the HEPTopTag-multiplicity (bottom right) distributions for samples generated assuming values of 1, 2, and 3 TeV for the T'-quark mass and values of 8 and 13 TeV for the center-of-mass energy. The distributions are normalized to unity for this shape comparison.

10 Summary

The complete dataset recorded with the CMS experiment at the CERN LHC at a center-of-mass energy of 8 TeV was analyzed in this thesis. It amounts to an integrated luminosity of 19.7 fb^{-1}. A search for pair-produced vector-like T' quarks was conducted. The analysis strategy was optimized for decays of the T' quarks to top quarks and Higgs bosons, where both, the top quark and the Higgs boson, decay hadronically. In the hadronic decays of the two top quarks and two Higgs bosons, a total of ten quarks are produced. Each of them can initiate an individual jet in the detector. As the decaying T' quarks are very heavy, their daughter particles are produced with large Lorentz boosts. Their decay products may therefore be so collimated, that overlapping signatures occur in the detector. This constitutes a challenge for analyses of these final states using classical event reconstruction techniques.

Novel analysis tools have been developed which allowed for circumvention of these complications in this search. Instead of reconstructing ten individual jets corresponding to the ten expected final state particles, larger radius jets were reconstructed that contained more than one of the decay products of the top quarks or Higgs bosons. Dedicated algorithms were applied to reconstruct and analyze the substructure of these large radius Cambridge/Aachen jets (CA jets). If the Lorentz boost of the decaying top quarks or Higgs bosons is large, all of their decay products can be found within the substructure of a single CA jet.

Specifically, the HEPTopTagger algorithm was used for the first time in an analysis of CMS data in the presented search. The HEPTopTagger algorithm was designed to identify hadronic decays of top quarks with large Lorentz boosts in the substructure of CA jets. It was shown to differentiate efficiently between jets containing top quark decays and other jets that originate from QCD processes. Another innovative technique was developed specifically for this analysis: the application of b-tagging algorithms to the subjets of CA jets. The combined secondary vertex (CSV) b-tagging algorithm is the most advanced tool for identification of jets containing bottom quarks that is available in CMS software. It is widely used in CMS analyses for the examination of jets that are reconstructed with the anti-k_T algorithm. The analysis presented in this thesis was the first to incorporate application of the CSV algorithm to subjets of CA jets. This subjet b-tagging technique entered the analysis in two significant instances: it improved the performance of the HEPTopTagger algorithm and was used for Higgs tagging, i.e., the identification of decays of boosted Higgs bosons to bottom-quark pairs within the substructure of CA jets.

The dominant background processes in this analysis are $\text{t}\bar{\text{t}}$ and QCD-multijet production. Application of the HEPTopTagger algorithm mainly reduces the QCD-multijet background, while the Higgs-tagging algorithm is a handle to also suppress the contribution of $\text{t}\bar{\text{t}}$ background events. A selection efficiency of 5-8% with respect to the number of events passing the trigger threshold was achieved for signal events containing T'T' $\rightarrow$ tHtH decays. Less than 0.8% of the simulated $\text{t}\bar{\text{t}}$ events passing the threshold are retained after the full event selection. An even more significant reduction to only 0.003% was achieved for the QCD-multijet background contribution. After the event selection with jet sub-

structure tools, the magnitude of the QCD-multijet contribution was found to be of the same order as that of the $t\bar{t}$ contribution. The use of novel techniques for jet-substructure analysis made the all-hadronic decay channel accessible.

The production cross section of QCD-multijet events is much larger than that of other processes taken into account in this analysis. Only a very small fraction of the events produced in the Monte Carlo simulation of these processes was retained by the event selection. The selected events fall into a very small region of the kinematic phase space whose special properties may not be accurately modeled in the simulation of the physics processes. The description of the QCD-multijet background contribution to the signal region was therefore derived using a data-driven technique instead. In the application of this method, two of the main jet-substructure related event-selection criteria were inverted to obtain QCD-multijet dominated and signal depleted sideband regions. The model for the QCD-multijet contribution in the signal region was derived using the events from these three sideband regions. In this data-driven estimation, the expected event rate was approximately doubled with respect to the number of selected events predicted by the Monte Carlo simulation. The statistical uncertainty in the shape of the background contribution was reduced drastically.

Numerous sources of systematic uncertainties were considered. The largest uncertainties were introduced by the choice for the renormalization and factorization scale used in the modeling of the $t\bar{t}$ background events. Furthermore, scale factors were applied to simulated $t\bar{t}$-background and signal events, in order to correct for observed differences in performance of the subjet b tagging in data and simulation. The uncertainty in these scale factors was the second leading uncertainty in the analysis.

In absence of any evidence for the existence of the searched for T' quark, Bayesian exclusion limits on the cross section for T'-quark pair production were derived. Two variables with good discrimination power between signal and background events were identified: the H_T variable, defined as the scalar sum of the transverse momenta of all subjets in the event, and the mass of the Higgs-candidate jet found in the event selection. This mass was reconstructed from the two b-tagged subjets in the Higgs-tagged CA jet. To exploit the discrimination power of both variables in an optimal way, they were combined in a single likelihood-ratio variable. One likelihood variable per examined T'-quark mass point was calculated. Studies of the expected limits showed that the overall sensitivity of the analysis was further improved by splitting the selected events into two categories according to the multiplicity of Higgs-tagged CA jets in the event.

In the comparison of the obtained limits on the cross section for pair production of T' quarks with the theory prediction, a lower exclusion limit on the T'-quark mass of 745 GeV at 95% confidence level was set under the assumption of Br(T'$\rightarrow$tH) = 100%, which is slightly lower than the expected limit of 773 GeV. Besides decays of T' quarks to top quarks and Higgs bosons, also decays of T' $\rightarrow$ tZ and T' $\rightarrow$ bW are predicted in the corresponding models. Therefore, supplementary results were produced in a scan of all possible combinations of branching ratios. The limits on the T'-quark mass extend from 745 GeV for Br(T' $\rightarrow$ tH) = 100% to 698 GeV in case of Br(T' $\rightarrow$ tH) = 80% and Br(T' $\rightarrow$ tZ) = 20%.

The analysis presented in this thesis results in the most stringent limits derived from CMS data for decays of T'T' $\rightarrow$ tHtH. It is the only analysis published to date that is optimized to these kind of decays in the absence of leptons. The ATLAS collaboration published another analysis optimized specifically for T'-quark decays to top quarks and

Higgs bosons. This search was performed in events with leptons in the final state. Lower exclusion limits on the T'-quark mass of approximately 800 GeV were obtained assuming a branching ratio Br(T' $\rightarrow$ tH) = 100% [141]. In the statistical combination of all published searches for T' quarks in the 8 TeV CMS data, a limit of 782 GeV was obtained.

At the center-of-mass energy of 13 TeV at which the LHC will be operated in the next data taking period, the cross section for pair production of vector-like quarks is expected to increase by almost a factor of 10 with respect to the cross sections predicted for the production at $\sqrt{s} = 8$ TeV. Taking into account the high mass limits set in searches in the 8 TeV data, future analyses will need to focus on higher mass ranges for the new particles and consider masses larger than 1 TeV for the hypothetical particles . Therefore, jet substructure methods designed for the analysis of final states with very large Lorentz boosts will continue to be of high importance to analyses in this field.

First studies also showed a good potential for searches for the single production of vector-like T' quarks. To date, most analyses focused on the pairwise production of T' quarks, which is the dominant production mode for vector-like quarks of lower masses. For masses exceeding the current exclusion limits, the cross section for single production of T' quarks is predicted to be larger than that for pair production. Furthermore, it is important to examine all possible production modes in order to assess the different facets of the models under study and improve the potential for future discovery of the vector-like quarks. Searches for vector-like quarks will remain highly interesting in the future.

A Impact of systematic uncertainties on shapes of observables

The impact of different systematic uncertainties on the shapes of the Higgs-candidate mass and H_T distributions of all events passing the event selection is discussed in section 7.5. In this section, supplementary information is given: the impact of the systematic uncertainties on the distributions for observables after splitting of the events into the single and multi Higgs-tag categories is illustrated in figures A.1-A.10.

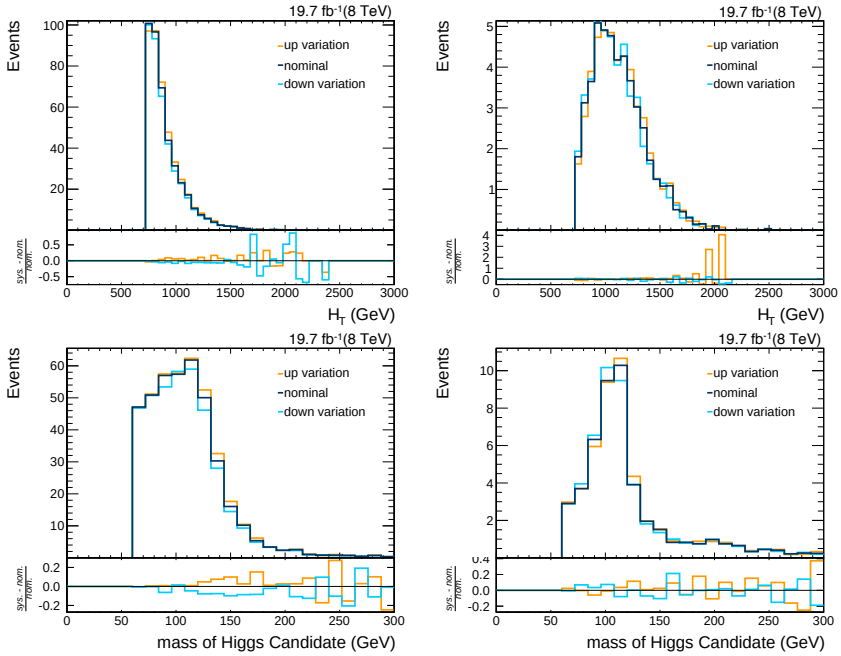

Figure A.1: Impact of uncertainties in the subjet energy corrections in the single Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

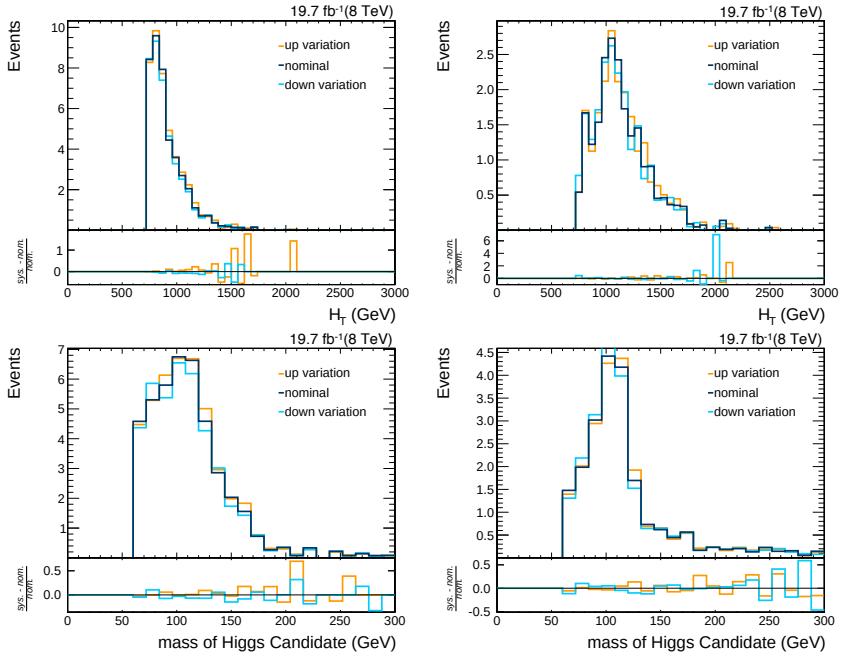

Figure A.2: Impact of uncertainties in the subjet energy corrections in the multi Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

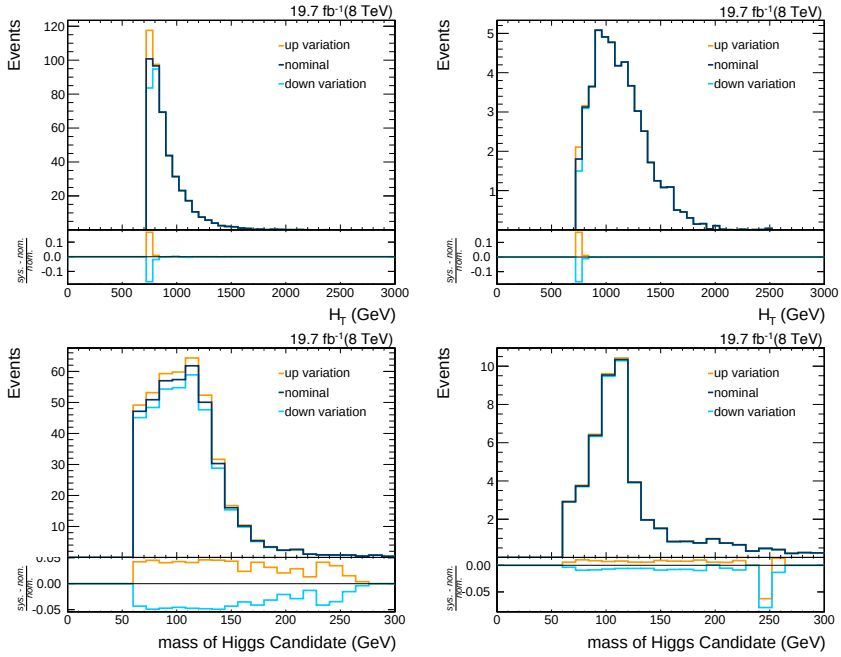

Figure A.3: Impact of uncertainties in the trigger scale factors in the single Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

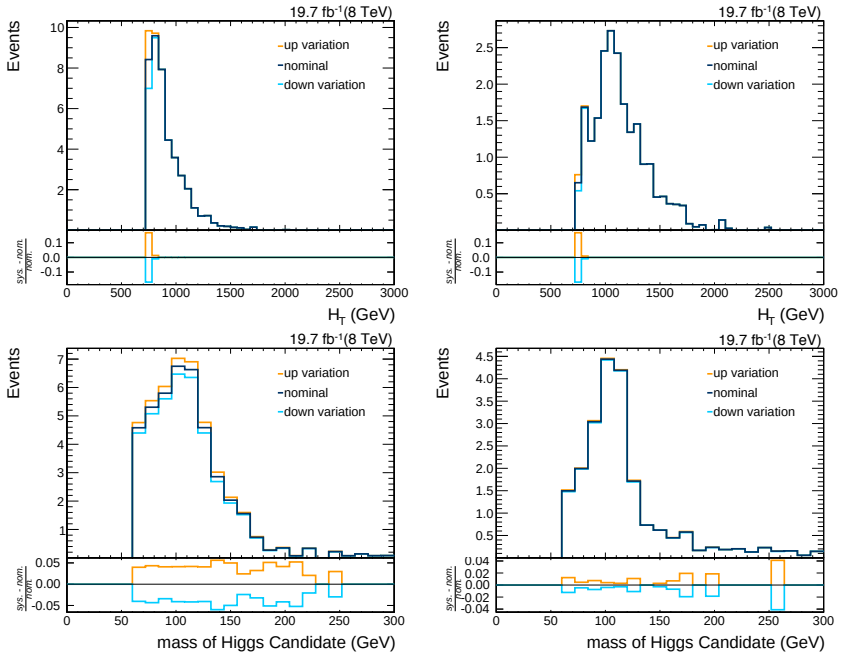

Figure A.4: Impact of uncertainties in the trigger scale factors in the multi Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

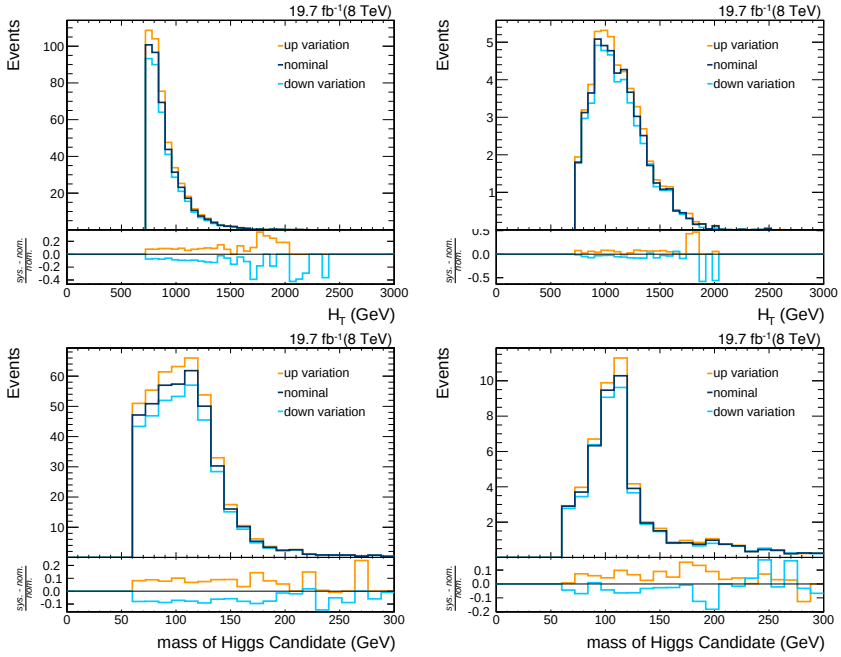

Figure A.5: Impact of uncertainties in the b-tagging scale factors in the single Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

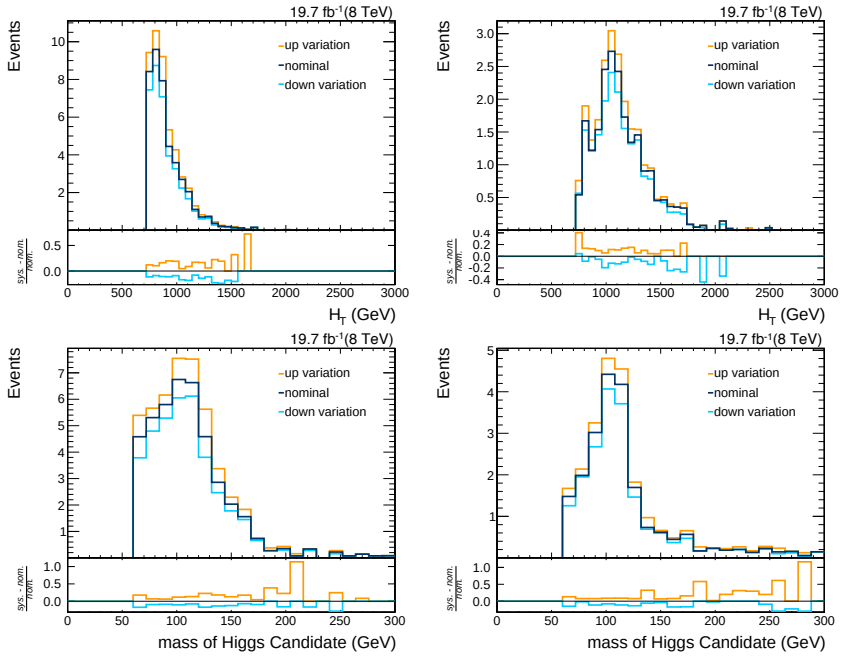

Figure A.6: Impact of uncertainties in the b-tagging scale factors in the multi Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

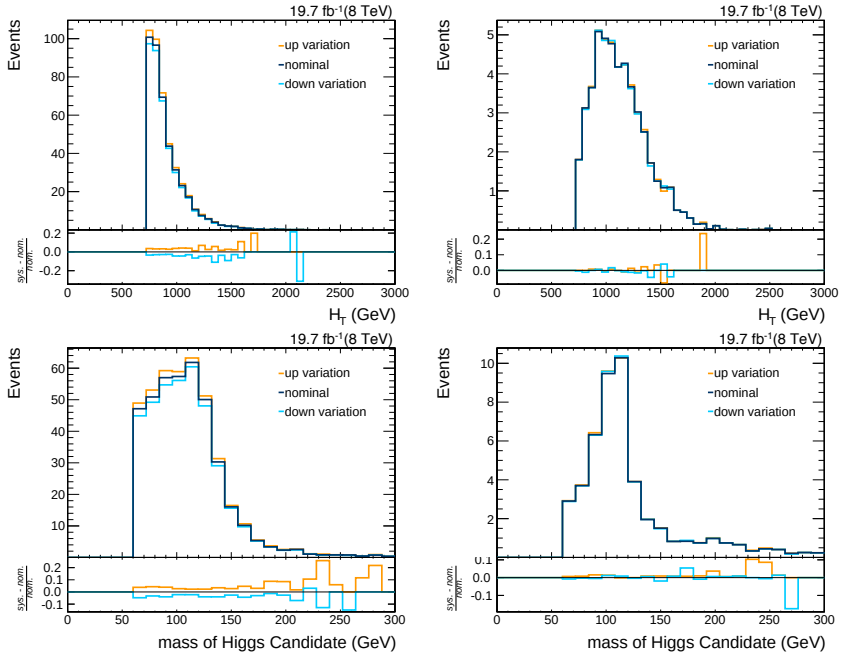

Figure A.7: Impact of uncertainties on the b-misidentification scale factors in the single Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

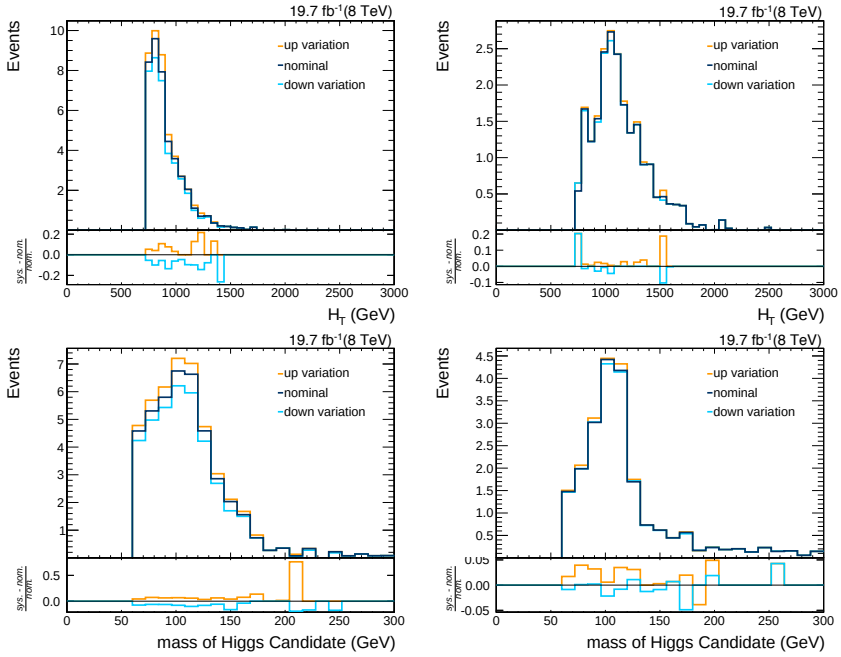

Figure A.8: Impact of uncertainties in the b-misidentification scale factors in the multi Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

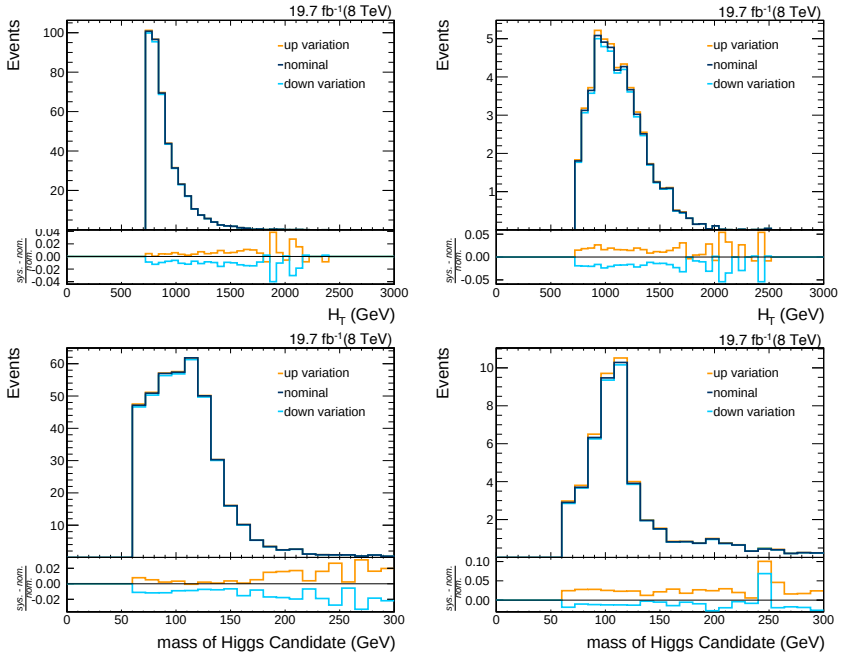

Figure A.9: Impact of uncertainties in the top-tagging scale factors in the single Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

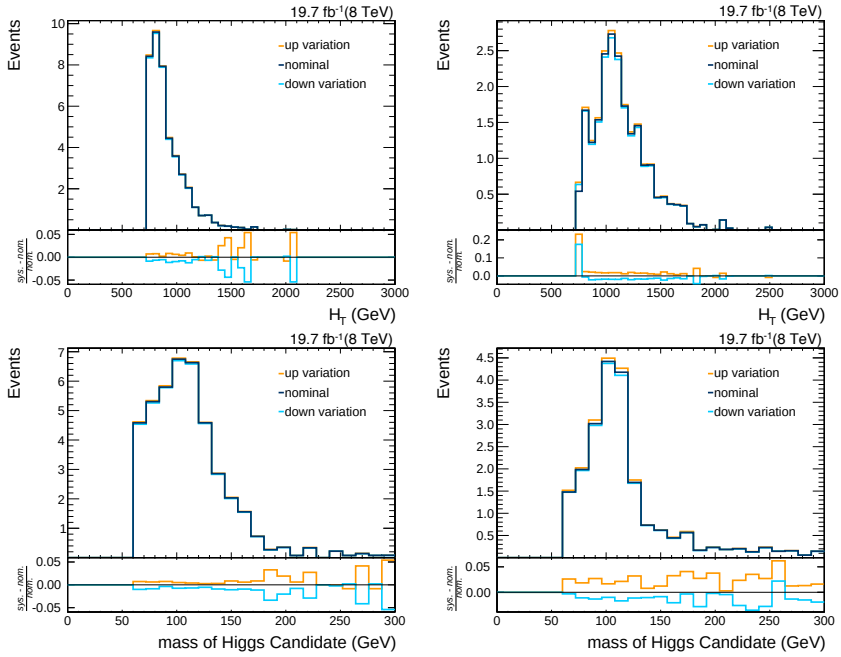

Figure A.10: Impact of uncertainties in the top-tagging scale factors in the multi Higgs-tag category. Top: H_T; Bottom: m(Higgs candidate); Left: $t\bar{t}$; Right: T'T' $\rightarrow$ tHtH signal (700 GeV).

B Monte Carlo generator studies in modeling of $t\bar{t}$ background events

In this section, distributions of $t\bar{t}$ background events generated with the MadGraph matrix element generator are compared to those for events generated with the Powheg Monte Carlo generator. Pythia is used for the shower simulation in the production of both samples. The names of the samples used in this study are listed in table B.1.

$t\bar{t}$ background
MadGraph samples
/TTJets_SemiLeptMGDecays_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A_ext-v1/AODSIM
/TTJets_HadronicMGDecays_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A_ext-v1/AODSIM
Powheg samples
/TT_CT10_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM

Table B.1: Simulated samples for description of the $t\bar{t}$ background.

For the $t\bar{t}$ sample simulated with MadGraph, differences in the transverse-momentum distribution have been observed between data and simulated events [142]. Scale factors are applied to each event to correct for these differences. The scale factors SF are calculated using the p_T of the generated top quark and anti-top quark in each event:

$$SF = \sqrt{e^{0.156-0.00137p_T(\mathrm{t})} \cdot e^{0.156-0.00137p_T(\bar{\mathrm{t}})}}. \tag{B.1}$$

In figure B.1 the H_T (left) and Higgs-candidate mass (right) distributions are compared between the $t\bar{t}$ samples generated with MadGraph and Powheg. The shapes agree well between the two samples, only a difference in normalization can be observed. Overall, more MadGraph simulated events than events simulated using Powheg are retained in the event selection. This difference is well covered by the rather large systematic uncertainties on the $t\bar{t}$ background contribution.

A comparison of the distributions of events simulated with Powheg before application of the p_T-reweighting scale factors to those of p_T-reweighted events simulated with MadGraph is provided in figure B.2. Also in this case, no significant shape differences are observed. The difference in normalization is reverted in this case, leading to a slightly

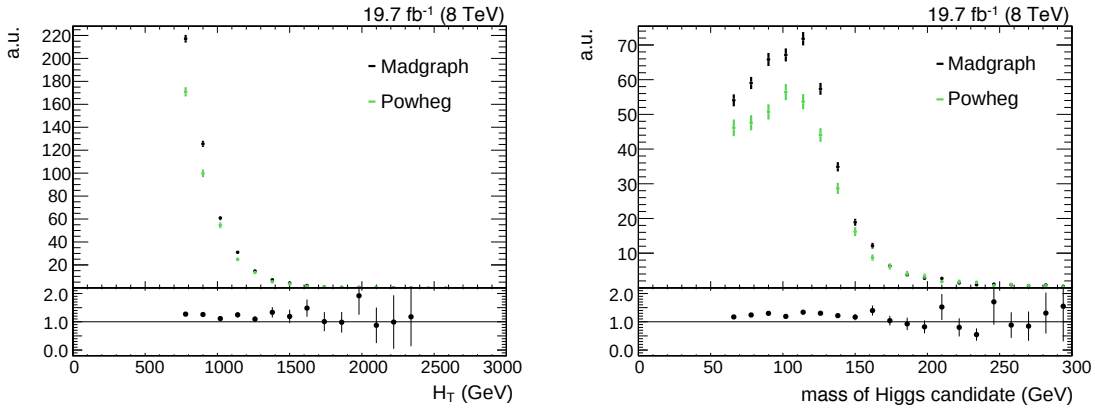

Figure B.1: Comparison of the H_T (left) and Higgs-candidate mass (right) distributions for $t\bar{t}$ events simulated with MadGraph and Powheg. The p_T-reweighting scale factors are applied to the events in both samples.

larger number of selected events from Powheg simulation. Also this discrepancy is accounted for by the uncertainties on the $t\bar{t}$ background contribution.

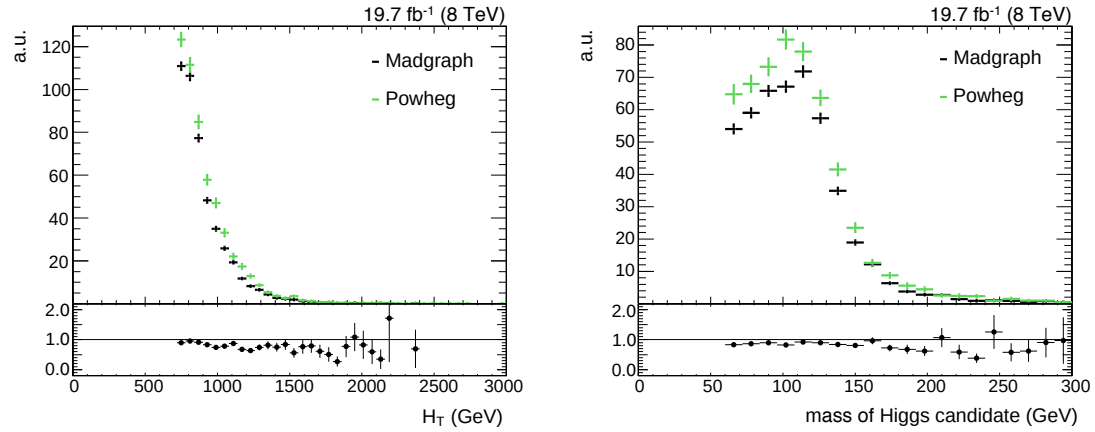

Figure B.2: Comparison of the H_T (left) and Higgs-candidate mass (right) distributions for $t\bar{t}$ events simulated with MadGraph and Powheg. The p_T-reweighting scale factors are applied events in the MadGraph sample only.

No significant differences are observed between the samples generated with Powheg and MadGraph. The sample generated with Powheg is used in this analysis.

C Additional information on the event selection efficiency

	Trigger	$H_T >$ 720 GeV	≥ 2 CA15 jets	≥ 1 HTT	≥ 1 HTT + subjet b-tag	Higgs-tag	
						1	≥ 2
Data	19673546	15084796	15084199	1039546	164265	1355	205
QCD	22733196	15335568	15335431	846159	103987	502.5	77
$t\bar{t}$	71838	57227	57183	22206	14640	486	55
Signal: Br(T' $\rightarrow$ tH) = 100%							
500 GeV	4939.9	4234.3	4233.6	1586.9	1084.6	192.0	91.0
600 GeV	2273.1	2090.8	2089.6	811.4	557.1	105.3	46.7
700 GeV	928.1	887.4	886.8	350.2	240.7	47.9	21.4
800 GeV	375.3	366.8	366.6	145.4	99.5	21.7	8.6
900 GeV	154.2	152.2	152.0	59.3	39.8	8.7	3.4
1000 GeV	63.8	63.3	63.3	24.6	16.3	3.5	1.4
Signal: Br(T' $\rightarrow$ tH) = 50%, Br(T' $\rightarrow$ tZ) = 50%							
500 GeV	4698.4	4013.7	4008.5	1526.8	975.3	131.7	55.1
600 GeV	2111.7	1937.1	1934.1	758.8	487.1	73.6	28.5
700 GeV	879.6	838.0	836.1	329.7	212.1	33.6	12.6
800 GeV	359.2	349.3	348.3	135.5	86.5	14.1	5.2
900 GeV	148.2	145.6	145.0	55.8	35.5	5.8	1.8
1000 GeV	62.4	61.7	61.4	23.7	14.8	2.4	0.8
Signal: Br(T' $\rightarrow$ tH) = 50%, Br(T' $\rightarrow$ bW) = 50%							
500 GeV	5798.4	5167.1	5162.3	1324.1	783.0	95.8	28.8
600 GeV	2425.1	2281.7	2279.3	606.3	361.7	48.5	18.4
700 GeV	957.7	925.8	924.3	248.4	150.7	23.7	8.4
800 GeV	378.8	371.5	370.5	97.2	58.3	9.4	3.3
900 GeV	153.7	151.9	151.3	39.1	23.4	3.9	1.3
1000 GeV	63.9	63.4	63.1	16.0	9.4	1.6	0.5

Table C.1: Resulting event yields after each selection step as predicted by the Monte Carlo simulation. The abbreviation HTT refers to HepTopTagged jets. No restriction to leptonic or hadronic final states is made in the production of the simulated events. (Continued on following page.)

	Trigger	H_T > 720 GeV	≥ 2 CA15 jets	≥ 1 HTT	≥ 1 HTT + subjet b-tag	Higgs-tag	
						1	≥ 2
Signal: Br(T' → tZ) = 100%							
500 GeV	4464.9	3837.3	3829.6	1467.1	847.5	75.3	20.1
600 GeV	1990.3	1821.6	1816.7	709.6	417.8	40.1	9.4
700 GeV	834.2	790.9	788.1	308.1	183.7	18.1	4.8
800 GeV	342.6	331.9	330.2	127.8	76.8	7.3	1.9
900 GeV	142.7	139.6	138.6	53.2	31.9	3.0	0.8
1000 GeV	60.4	59.5	58.9	22.2	13.2	1.3	0.3
Signal: Br(T' → tZ) = 50%, Br(T' → bW) = 50%							
500 GeV	5587.3	4995.9	4986.8	1275.0	670.3	40.0	9.8
600 GeV	2287.7	2144.8	2139.6	554.9	295.2	22.2	3.9
700 GeV	912.1	878.5	876.0	224.1	120.9	8.8	2.2
800 GeV	364.0	356.0	354.0	89.0	48.5	3.6	0.8
900 GeV	149.2	147.0	146.0	37.0	20.0	1.5	0.4
1000 GeV	62.3	61.6	61.0	15.1	8.1	0.6	0.2
Signal: Br(T' → bW) = 100%							
500 GeV	6666.9	6153.4	6141.5	953.5	399.7	14.2	2.9
600 GeV	2557.5	2448.3	2443.7	348.1	145.3	6.6	0.6
700 GeV	974.7	949.2	946.3	124.8	49.9	2.1	0.3
800 GeV	379.2	373.1	371.4	45.0	17.3	0.6	0.04
900 GeV	152.7	151.1	150.1	16.9	6.6	0.2	0.03
1000 GeV	63.2	62.7	62.1	6.6	2.4	0.1	0.01

Table C.2: Resulting event yields after each selection step as predicted by the Monte Carlo simulation. The abbreviation HTT refers to HepTopTagged jets. No restriction to leptonic or hadronic final states is made in the production of the simulated events. (Continued from previous page.)

	≥ 1 Higgs-tags	1 Higgs-tag	≥ 2 Higgs-tags
Signal: Br(T' $\rightarrow$ tH) = 100%			
$m_{T'} = 500\ GeV$	2.5	1.7	0.8
$m_{T'} = 600\ GeV$	4.4	3.1	1.4
$m_{T'} = 700\ GeV$	6.0	4.2	1.9
$m_{T'} = 800\ GeV$	7.2	5.3	2.0
$m_{T'} = 900\ GeV$	7.3	5.3	2.1
$m_{T'} = 1000\ GeV$	7.2	5.2	2.0
Signal: Br(T' $\rightarrow$ tH) = 50%, Br(T' $\rightarrow$ tZ) = 50%			
$m_{T'} = 500\ GeV$	1.6	1.1	0.5
$m_{T'} = 600\ GeV$	3.0	2.1	0.8
$m_{T'} = 700\ GeV$	4.0	2.9	1.1
$m_{T'} = 800\ GeV$	4.6	3.3	1.2
$m_{T'} = 900\ GeV$	4.6	3.5	1.1
$m_{T'} = 1000\ GeV$	4.7	3.6	1.2
Signal: Br(T' $\rightarrow$ tH) = 50%, Br(T' $\rightarrow$ bW) = 50%			
$m_{T'} = 500\ GeV$	1.1	0.8	0.2
$m_{T'} = 600\ GeV$	1.9	1.4	0.5
$m_{T'} = 700\ GeV$	2.8	2.0	0.7
$m_{T'} = 800\ GeV$	3.0	2.2	0.8
$m_{T'} = 900\ GeV$	3.1	2.3	0.8
$m_{T'} = 1000\ GeV$	3.1	2.4	0.8

Table C.3: Selection efficiencies in percent for different signal decay modes and mass points for the three event categories. The efficiencies are calculated with respect to an inclusive sample without restriction to any decay modes of the top/bottom quarks and W/Z/Higgs bosons and before application of any selection criteria. (Continued on following page.)

	≥ 1 Higgs-tags	1 Higgs-tag	≥ 2 Higgs-tags
Signal: Br(T' $\rightarrow$ tZ) = 100%			
$m_{T'} = 500\ GeV$	0.8	0.7	0.2
$m_{T'} = 600\ GeV$	1.5	1.2	0.3
$m_{T'} = 700\ GeV$	2.0	1.6	0.4
$m_{T'} = 800\ GeV$	2.2	1.7	0.5
$m_{T'} = 900\ GeV$	2.3	1.8	0.5
$m_{T'} = 1000\ GeV$	2.4	1.9	0.4
Signal: Br(T' $\rightarrow$ tZ) = 50%, Br(T' $\rightarrow$ bW) = 50%			
$m_{T'} = 500\ GeV$	0.4	0.3	0.1
$m_{T'} = 600\ GeV$	0.8	0.6	0.1
$m_{T'} = 700\ GeV$	1.0	0.8	0.2
$m_{T'} = 800\ GeV$	1.1	0.9	0.2
$m_{T'} = 900\ GeV$	1.1	0.9	0.2
$m_{T'} = 1000\ GeV$	1.1	0.9	0.2
Signal: Br(T' $\rightarrow$ bW) = 100%			
$m_{T'} = 500\ GeV$	0.1	0.1	0.02
$m_{T'} = 600\ GeV$	0.1	0.1	0.02
$m_{T'} = 700\ GeV$	0.2	0.2	0.02
$m_{T'} = 800\ GeV$	0.2	0.2	0.01
$m_{T'} = 900\ GeV$	0.2	0.1	0.02
$m_{T'} = 1000\ GeV$	0.2	0.1	0.01

Table C.4: Selection efficiencies in percent for different signal decay modes and mass points in the three event categories. The efficiencies are calculated with respect to an inclusive sample without restriction to any decay modes of the top/bottom quarks and W/Z/Higgs bosons and before application of any selection criteria. (Continued from previous page.)

D Exclusion limits obtained in the scan of branching fractions

Complementary to the graphical presentation of the results obtained in the limit setting procedure in the scan of all branching fractions, which are given in section 7.6, the results are listed in tables in this section. Table D.1 contains the lower exclusion limits on the T'-quark mass at 95% confidence level obtained in statistical analyses of the results using differently composed signal samples. The branching fractions assumed when obtaining the limits are quoted in the first three columns of the table, followed by the observed and median expected limit on the T'-quark mass. In the last two columns, the one and two sigma intervals around the median expected limit can be found.

Branching fraction			Observed	Expected	Expected$\pm1\sigma$	Expected$\pm2\sigma$
T' $\to$ bW	T' $\to$ tZ	T' $\to$ tH	limit	limit		
0.0	0.2	0.8	698	732	[596,795]	[<500,851]
0.0	0.15	0.85	715	734	[633,798]	[<500,857]
0.0	0.1	0.9	725	751	[639,806]	[<500,862]
0.0	0.05	0.95	739	763	[655,827]	[538,873]
0.0	0.0	1.0	745	773	[664,832]	[557,875]
0.05	0.1	0.85	716	732	[619,798]	[<500,856]
0.05	0.05	0.9	724	749	[633,812]	[503,858]
0.05	0.0	0.95	731	757	[650,817]	[534,865]
0.1	0.05	0.85	708	730	[595,795]	[<500,849]
0.1	0.0	0.9	720	737	[599,799]	[<500,859]

Table D.1: Branching fractions considered in the scan and the resulting expected and observed limits on the mass of the T' quark in GeV. Only combinations for which a non-null observed limit is found are reported. $<$ 500 indicates that the considered limit lies outside the scanned mass region between 500 and 1000 GeV.

Tables D.2-D.4 contain the upper exclusion limits at 95% confidence level on the cross section of T'-quark pair production obtained in the scan of the branching fractions. An individual limit is set for each T'-quark mass hypothesis, ranging from $m(\text{T'}) = 500$ GeV to $m(\text{T'}) = 1000$ GeV in steps of 100 GeV. Again, the final-state composition of the signal samples is given by the three branching fractions quoted in the first two columns. For each signal composition and T'-quark mass point, the expected limit with the corresponding uncertainties is stated in brackets in the upper part of the rows, the observed limit can be found in the lower part of each row.

bW	tZ	tH	$m_{T'}$=500 GeV	$m_{T'}$=600 GeV	$m_{T'}$=700 GeV	$m_{T'}$=800 GeV	$m_{T'}$=900 GeV	$m_{T'}$=1000 GeV
0	0	100	$(0.24^{+0.11}_{-0.08})$ 0.432	$(0.105^{+0.053}_{-0.035})$ 0.132	$(0.046^{+0.021}_{-0.014})$ 0.046	$(0.026^{+0.011}_{-0.008})$ 0.036	$(0.020^{+0.008}_{-0.006})$ 0.029	$(0.015^{+0.007}_{-0.004})$ 0.026
0	20	80	$(0.30^{+0.12}_{-0.10})$ 0.576	$(0.118^{+0.063}_{-0.036})$ 0.157	$(0.054^{+0.025}_{-0.014})$ 0.059	$(0.032^{+0.014}_{-0.010})$ 0.046	$(0.023^{+0.011}_{-0.006})$ 0.036	$(0.018^{+0.008}_{-0.005})$ 0.029
0	40	60	$(0.39^{+0.21}_{-0.12})$ 0.866	$(0.143^{+0.074}_{-0.043})$ 0.191	$(0.067^{+0.027}_{-0.019})$ 0.076	$(0.041^{+0.019}_{-0.012})$ 0.057	$(0.030^{+0.014}_{-0.009})$ 0.043	$(0.023^{+0.010}_{-0.007})$ 0.036
0	60	40	$(0.53^{+0.30}_{-0.19})$ 1.121	$(0.186^{+0.097}_{-0.061})$ 0.263	$(0.085^{+0.041}_{-0.025})$ 0.102	$(0.056^{+0.026}_{-0.016})$ 0.082	$(0.039^{+0.017}_{-0.011})$ 0.052	$(0.030^{+0.012}_{-0.009})$ 0.043
0	80	20	$(0.75^{+0.43}_{-0.27})$ 1.743	$(0.27^{+0.13}_{-0.08})$ 0.353	$(0.118^{+0.057}_{-0.033})$ 0.148	$(0.081^{+0.038}_{-0.024})$ 0.110	$(0.055^{+0.024}_{-0.016})$ 0.069	$(0.043^{+0.018}_{-0.012})$ 0.058
0	100	0	$(1.19^{+0.69}_{-0.43})$ 3.010	$(0.50^{+0.23}_{-0.17})$ 0.690	$(0.196^{+0.084}_{-0.057})$ 0.229	$(0.127^{+0.055}_{-0.037})$ 0.181	$(0.087^{+0.037}_{-0.026})$ 0.102	$(0.066^{+0.029}_{-0.020})$ 0.094
20	0	80	$(0.34^{+0.17}_{-0.11})$ 0.656	$(0.137^{+0.062}_{-0.043})$ 0.155	$(0.061^{+0.027}_{-0.018})$ 0.061	$(0.035^{+0.015}_{-0.010})$ 0.049	$(0.026^{+0.011}_{-0.008})$ 0.038	$(0.020^{+0.009}_{-0.006})$ 0.033

Table D.2: Branching fractions considered in the scan and the resulting (median expected) observed upper exlusion limits at 95% confidence level on the cross section of T'-quark pair production for different masses of the T' quark. The central 68% interval around the median expected limit are given by the + and − interval.

bW	tZ	tH	$m_{T'}$=500 GeV	$m_{T'}$=600 GeV	$m_{T'}$=700 GeV	$m_{T'}$=800 GeV	$m_{T'}$=900 GeV	$m_{T'}$=1000 GeV
20	20	60	$(0.46^{+0.24}_{-0.15})$ 0.934	$(0.165^{+0.080}_{-0.052})$ 0.206	$(0.076^{+0.033}_{-0.022})$ 0.081	$(0.045^{+0.019}_{-0.014})$ 0.060	$(0.033^{+0.014}_{-0.010})$ 0.049	$(0.025^{+0.011}_{-0.007})$ 0.039
20	40	40	$(0.63^{+0.33}_{-0.20})$ 1.338	$(0.22^{+0.11}_{-0.07})$ 0.278	$(0.099^{+0.046}_{-0.031})$ 0.116	$(0.062^{+0.029}_{-0.018})$ 0.079	$(0.043^{+0.018}_{-0.011})$ 0.061	$(0.035^{+0.015}_{-0.010})$ 0.051
20	60	20	$(0.93^{+0.50}_{-0.31})$ 1.962	$(0.33^{+0.17}_{-0.10})$ 0.452	$(0.146^{+0.061}_{-0.043})$ 0.171	$(0.089^{+0.040}_{-0.025})$ 0.114	$(0.064^{+0.027}_{-0.020})$ 0.081	$(0.048^{+0.022}_{-0.014})$ 0.067
20	80	0	$(1.54^{+0.92}_{-0.53})$ 3.545	$(0.63^{+0.31}_{-0.20})$ 0.806	$(0.25^{+0.11}_{-0.07})$ 0.323	$(0.150^{+0.063}_{-0.042})$ 0.161	$(0.108^{+0.043}_{-0.033})$ 0.122	$(0.084^{+0.037}_{-0.027})$ 0.113
40	0	60	$(0.56^{+0.29}_{-0.19})$ 1.136	$(0.191^{+0.092}_{-0.058})$ 0.206	$(0.087^{+0.037}_{-0.024})$ 0.090	$(0.050^{+0.022}_{-0.014})$ 0.070	$(0.036^{+0.017}_{-0.010})$ 0.053	$(0.028^{+0.013}_{-0.007})$ 0.046
40	20	40	$(0.79^{+0.46}_{-0.26})$ 1.648	$(0.26^{+0.14}_{-0.08})$ 0.314	$(0.114^{+0.057}_{-0.032})$ 0.129	$(0.070^{+0.030}_{-0.020})$ 0.085	$(0.049^{+0.021}_{-0.013})$ 0.072	$(0.037^{+0.016}_{-0.010})$ 0.059
40	40	20	$(1.20^{+0.81}_{-0.42})$ 2.522	$(0.41^{+0.18}_{-0.14})$ 0.533	$(0.170^{+0.081}_{-0.052})$ 0.210	$(0.103^{+0.049}_{-0.027})$ 0.120	$(0.072^{+0.035}_{-0.020})$ 0.099	$(0.059^{+0.025}_{-0.017})$ 0.081

Table D.3: Branching fractions considered in the scan and the resulting (median expected) observed upper exlusion limits at 95% confidence level on the cross section of T'-quark pair production for different masses of the T' quark. The central 68% interval around the median expected limit are given by the + and − interval.

bW	tZ	tH	$m_{T'}$=500 GeV	$m_{T'}$=600 GeV	$m_{T'}$=700 GeV	$m_{T'}$=800 GeV	$m_{T'}$=900 GeV	$m_{T'}$=1000 GeV
40	60	0	$(2.2^{+1.2}_{-0.8})$ 4.248	$(0.79^{+0.45}_{-0.26})$ 1.189	$(0.34^{+0.16}_{-0.10})$ 0.516	$(0.178^{+0.087}_{-0.050})$ 0.175	$(0.131^{+0.065}_{-0.037})$ 0.151	$(0.108^{+0.049}_{-0.032})$ 0.152
60	0	40	$(1.09^{+0.76}_{-0.39})$ 2.606	$(0.31^{+0.16}_{-0.10})$ 0.381	$(0.137^{+0.067}_{-0.040})$ 0.137	$(0.082^{+0.038}_{-0.023})$ 0.105	$(0.056^{+0.025}_{-0.016})$ 0.079	$(0.045^{+0.018}_{-0.013})$ 0.067
60	20	20	$(1.72^{+1.38}_{-0.63})$ 3.751	$(0.50^{+0.26}_{-0.16})$ 0.622	$(0.212^{+0.098}_{-0.064})$ 0.266	$(0.124^{+0.055}_{-0.035})$ 0.152	$(0.087^{+0.040}_{-0.024})$ 0.123	$(0.070^{+0.033}_{-0.019})$ 0.104
60	40	0	$(3.3^{+2.2}_{-1.2})$ 6.594	$(1.16^{+0.68}_{-0.41})$ 2.028	$(0.48^{+0.23}_{-0.14})$ 0.781	$(0.25^{+0.11}_{-0.07})$ 0.215	$(0.183^{+0.081}_{-0.056})$ 0.208	$(0.159^{+0.079}_{-0.050})$ 0.213
80	0	20	$(3.1^{+3.1}_{-1.4})$ 6.577	$(0.65^{+0.33}_{-0.22})$ 0.725	$(0.28^{+0.14}_{-0.09})$ 0.307	$(0.174^{+0.076}_{-0.052})$ 0.202	$(0.112^{+0.047}_{-0.033})$ 0.160	$(0.087^{+0.040}_{-0.024})$ 0.130
80	20	0	$(4.4^{+4.4}_{-1.7})$ 10.448	$(1.67^{+0.96}_{-0.60})$ 3.033	$(0.73^{+0.36}_{-0.23})$ 1.093	$(0.40^{+0.20}_{-0.12})$ 0.301	$(0.28^{+0.13}_{-0.09})$ 0.317	$(0.27^{+0.14}_{-0.08})$ 0.357
100	0	0	$(9.^{+11.}_{-5.})$ 33.062	$(3.1^{+2.6}_{-1.2})$ 4.811	$(1.5^{+1.6}_{-0.6})$ 1.090	$(1.8^{+4.2}_{-0.8})$ 1.524	$(0.57^{+0.30}_{-0.18})$ 0.781	$(0.61^{+0.33}_{-0.19})$ 0.813

Table D.4: Branching fractions considered in the scan and the resulting (median expected) observed upper exlusion limits at 95% confidence level on the cross section of T'-quark pair production for different masses of the T' quark. The central 68% interval around the median expected limit are given by the + and − interval.

E Limits on the T'-quark mass obtained in the combination of three searches for T' quarks

The statistical combination of the results of different searches for vector-like T' quarks in CMS data at $\sqrt{s} = 8$ TeV is described in chapter 8. Complementary to the graphical presentation of the results obtained in the limit setting procedure in the scan of all branching fractions, the results are listed in tabular form in this section. Table E.1 contains the upper limits on the T'-quark mass obtained from statistical analyses using differently composed signal samples. The branching fractions assumed when obtaining the limits are quoted in the first three columns of the table, followed by the observed and expected limit on the T'-quark mass. In the last two columns, the one and two sigma intervals around the median expected limit can be found.

Branching fraction			Observed	Expected	Expected	Expected
T' $\to$ bW	T' $\to$ tZ	T' $\to$ tH	limit	limit	$\pm1\sigma$	$\pm2\sigma$
0.0	1.0	0.0	782	832	[783,876]	[750,> 1000]
0.0	0.8	0.2	772	820	[778,871]	[731,896]
0.0	0.6	0.4	757	810	[772,868]	[725,897]
0.0	0.4	0.6	758	812	[773,871]	[725,> 1000]
0.0	0.2	0.8	762	824	[775,880]	[727,> 1000]
0.0	0.0	1.0	767	839	[781,888]	[728,> 1000]
0.2	0.8	0.0	768	806	[773,866]	[730,895]
0.2	0.6	0.2	756	798	[765,863]	[698,895]
0.2	0.4	0.4	741	796	[756,861]	[695,897]
0.2	0.2	0.6	749	799	[759,865]	[695,898]
0.2	0.0	0.8	751	811	[769,877]	[714,> 1000]
0.4	0.6	0.0	742	795	[755,858]	[699,890]
0.4	0.4	0.2	715	792	[744,853]	[685,893]
0.4	0.2	0.4	725	790	[741,857]	[686,895]
0.4	0.0	0.6	730	793	[748,864]	[688,897]
0.6	0.4	0.0	708	788	[743,851]	[682,888]
0.6	0.2	0.2	698	788	[733,852]	[680,891]
0.6	0.0	0.4	703	793	[737,861]	[682,894]
0.8	0.2	0.0	697	790	[735,852]	[674,891]
0.8	0.0	0.2	708	792	[738,863]	[680,> 1000]
1.0	0.0	0.0	720	799	[752,871]	[684,> 1000]

Table E.1: Results of the combination of different searches for T' quarks. Sample of branching fractions considered in the scan and the resulting expected and observed limits on the mass of the T' quark in GeV. > 1000 indicates that the considered limit lies outside the scanned mass region between 500 and 1000 GeV.

Bibliography

[1] UA1 Collaboration, "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV", *Physics Letters B* **122** (1983), no. 1, 103 – 116. doi:10.1016/0370-2693(83)91177-2.

[2] UA2 Collaboration, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider", *Physics Letters B* **122** (1983), no. 56, 476 – 485. doi:10.1016/0370-2693(83)91605-2.

[3] UA1 Collaboration, "Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider", *Physics Letters B* **126** (1983), no. 5, 398 – 410. doi:10.1016/0370-2693(83)90188-0.

[4] UA2 Collaboration, "Evidence for Z0 → e+ e- at the CERN anti-p p Collider", *Physics Letters* **B129** (1983) 130–140. doi:10.1016/0370-2693(83)90744-X.

[5] G. Baur, G. Boero, A. Brauksiepe et al., "Production of antihydrogen", *Physics Letters B* **368** (1996), no. 3, 251 – 258. doi:10.1016/0370-2693(96)00005-6.

[6] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* **716** (2012), no. 1, 1 – 29. doi:10.1016/j.physletb.2012.08.020.

[7] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** (2012), no. 1, 30 – 61. doi:10.1016/j.physletb.2012.08.021.

[8] S. Martin, "A Supersymmetry primer", *Advanced Series on Directions in High Energy Physics* **21** (2010) 1–153. doi:10.1142/9789814307505_0001.

[9] O. Eberhardt, G. Herbert, H. Lacker et al., "Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations", *Phys.Rev.Lett.* **109** (2012) 241802. doi:10.1103/PhysRevLett.109.241802.

[10] M. Schmaltz and D. Tucker-Smith, "Little Higgs Theories", *Annual Review of Nucleal Particle Science* **55** (2005) 229–270. doi:10.1146/annurev.nucl.55.090704.151502.

[11] M. Dugan, H. Georgi, and D. Kaplan, "Anatomy of a composite Higgs model", *Nuclear Physics B* **254** (1985), no. 0, 299 – 326. doi:10.1016/0550-3213(85)90221-4.

[12] L. Randall and R. Sundrum, "Large Mass Hierarchy from a Small Extra Dimension", *Physical Review Letters* **83** (1999) 3370–3373. doi:10.1103/PhysRevLett.83.3370.

[13] T. Plehn, M. Spannowsky, M. Takeuchi et al., "Stop reconstruction with tagged tops", *Journal of High Energy Physics* **2010** (2010), no. 10,. doi:10.1007/JHEP10(2010)078.

[14] CMS Collaboration, "Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events", CMS Physics Analysis Summary CMS-PAS-BTV-13-001, 2013. https://cds.cern.ch/record/1581306.

[15] F. Halzen and A. Martin, "Quarks and Leptons: An Introductory Course in Modern Particle Physics". John Wiley and Sons, Inc, Chichester, 1984. ISBN 0-471-88741-2.

[16] K. Olive et al. (Particle Data Group), "Review of Particle Physics", *Chinese Physic C* **38** (2014) 090001. doi:10.1088/1674-1137/38/9/090001.

[17] ATLAS, CDF, CMS, D0 Collaboration, "First combination of Tevatron and LHC measurements of the top-quark mass", CMS Physics Analysis Summary CMS-PAS-TOP-13-014, 2014. https://cds.cern.ch/record/1669819.

[18] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Physical Review Letters* **10** (1963) 531–533. doi:10.1103/PhysRevLett.10.531.

[19] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Progress of Theoretical Physics* **49** (1973), no. 2, 652–657. doi:10.1143/PTP.49.652.

[20] R. Davis, D. Harmer, and K. Hoffman, "Search for Neutrinos from the Sun", *Physical Review Letters* **20** (1968) 1205–1209. doi:10.1103/PhysRevLett.20.1205.

[21] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Physical Review Letters* **81** (1998) 1562–1567. doi:10.1103/PhysRevLett.81.1562.

[22] S. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** (1961), no. 4, 579 – 588. doi:10.1016/0029-5582(61)90469-2.

[23] S. Weinberg, "A Model of Leptons", *Physical Review Letters* **19** (1967) 1264–1266. doi:10.1103/PhysRevLett.19.1264.

[24] E. Sudarshan and R. Marshak, "Chirality Invariance and the Universal Fermi Interaction", *Physical Review* **109** (1958) 1860–1862. doi:10.1103/PhysRev.109.1860.2.

[25] C. Wu, E. Ambler, R. Hayward et al., "Experimental Test of Parity Conservation in Beta Decay", *Physical Review* **105** (1957) 1413–1415. doi:10.1103/PhysRev.105.1413.

[26] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Physical Review Letters* **13** (1964) 321–323. doi:10.1103/PhysRevLett.13.321.

[27] P. Higgs, "Broken symmetries, massless particles and gauge fields", *Physics Letters* **12** (1964), no. 2, 132 – 133. doi:10.1016/0031-9163(64)91136-9.

[28] P. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Physical Review Letters* **13** (1964) 508–509. doi:10.1103/PhysRevLett.13.508.

[29] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Physical Review Letters* **74** (1995) 2626–2631. doi:10.1103/PhysRevLett.74.2626.

[30] D0 Collaboration, "Observation of the top quark", *Physical Review Letters* **74** (1995) 2632–2637. doi:10.1103/PhysRevLett.74.2632.

[31] CMS Collaboration, "Measurement of the t-channel single-top-quark production cross section and of the $| V_{tb} |$ CKM matrix element in pp collisions at $\sqrt{s}$= 8 TeV", *Journal of High Energy Physics* **1406** (2014) 090. doi:10.1007/JHEP06(2014)090.

[32] ATLAS Collaboration, "Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector", arXiv:1406.5375.

[33] CMS Collaboration, "Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model", CMS Physics Analysis Summary CMS-PAS-HIG-14-009, 2014. https://cds.cern.ch/record/1728249.

[34] CMS Collaboration, "Evidence for the direct decay of the 125 GeV Higgs boson to fermions", *Nature Physics* **10** (2014). doi:10.1038/nphys3005.

[35] "Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC", *Physics Letters B* **726** (2013), no. 13, 88 – 119. doi:10.1016/j.physletb.2013.08.010.

[36] "Evidence for the spin-0 nature of the Higgs boson using ATLAS data", *Physics Letters B* **726** (2013), no. 13, 120 – 144. doi:10.1016/j.physletb.2013.08.026.

[37] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group", Report CERN-2013-004, 2013. https://cds.cern.ch/record/1559921.

[38] J. Wess and B. Zumino, "A lagrangian model invariant under supergauge transformations", *Physics Letters B* **49** (1974), no. 1, 52 – 54. doi:10.1016/0370-2693(74)90578-4.

[39] J. Wess and B. Zumino, "Supergauge transformations in four dimensions", *Nuclear Physics B* **70** (1974), no. 1, 39 – 50. doi:10.1016/0550-3213(74)90355-1.

[40] CMS Collaboration, "CMS Supersymmetry Public Results". Topic revision: r279 - 16 Sep 2014, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS.

[41] ATLAS Collaboration, "ATLAS Experiment - Public Results: ATLAS Supersymmetry searches". Topic revision: r435 - 22 Oct 2014, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults.

[42] N. Arkani-Hamed, A. Cohen, and H. Georgi, "Electroweak symmetry breaking from dimensional deconstruction", *Physics Letters* **B513** (2001) 232–240. doi:10.1016/S0370-2693(01)00741-9.

[43] M. Perelstein, M. Peskin, and A. Pierce, "Top quarks and electroweak symmetry breaking in little Higgs models", *Physical Review* **D69** (2004) 075002. doi:10.1103/PhysRevD.69.075002.

[44] R. Sundrum, "Tasi 2004 lectures: To the fifth dimension and back", arXiv:hep-th/0508134.

[45] T. Rizzo, "Probes of universal extra dimensions at colliders", *Physical Review* **D64** (2001) 095010. doi:10.1103/PhysRevD.64.095010.

[46] M. Carena, E. Ponton, J. Santiago et al., "Electroweak constraints on warped models with custodial symmetry", *Physical Review* **D76** (2007) 035006. doi:10.1103/PhysRevD.76.035006.

[47] H. Georgi and D. Kaplan, "Composite Higgs and custodial SU(2)", *Physics Letters B* **145** (1984), no. 34, 216 – 220. doi:10.1016/0370-2693(84)90341-1.

[48] R. Contino, "The Higgs as a Composite Nambu-Goldstone Boson", arXiv:1005.4269.

[49] K. Agashe, R. Contino, and A. Pomarol, "The Minimal composite Higgs model", *Nuclear Physics B* **719** (2005) 165–187. doi:10.1016/j.nuclphysb.2005.04.035.

[50] J. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer et al., "Handbook of vectorlike quarks: Mixing and single production", *Physical Review D* **88** (2013) 094010. doi:10.1103/PhysRevD.88.094010.

[51] F. del Aguila, M. Perez-Victoria, and J. Santiago, "Effective description of quark mixing", *Physics Letters* **B492** (2000) 98–106. doi:10.1016/S0370-2693(00)01071-6.

[52] F. del Aguila, M. Perez-Victoria, and J. Santiago, "Observable contributions of new exotic quarks to quark mixing", *Journal of High Energy Physics* **0009** (2000) 011. doi:10.1088/1126-6708/2000/09/011.

[53] J. Aguilar-Saavedra, "Effects of mixing with quark singlets", *Physical Review D* **67** (2003) 035003. doi:10.1103/PhysRevD.67.035003.

[54] A. De Simone, O. Matsedonskyi, R. Rattazzi et al., "A First Top Partner Hunter's Guide", *Journal of High Energy Physics* **1304** (2013) 004. doi:10.1007/JHEP04(2013)004.

[55] G. Cacciapaglia, A. Deandrea, D. Harada et al., "Bounds and Decays of New Heavy Vector-like Top Partners", *Journal of High Energy Physics* **1011** (2010) 159. doi:10.1007/JHEP11(2010)159.

[56] G. Isidori, Y. Nir, and G. Perez, "Flavor Physics Constraints for Physics Beyond the Standard Model", *Annual Review of Nuclear and Particle Science* **60** (2010) 355. doi:10.1146/annurev.nucl.012809.104534.

[57] S. Heinemeyer, W. Hollik, A. Weber et al., "Z Pole Observables in the MSSM", *Journal of High Energy Physics* **0804** (2008) 039. doi:10.1088/1126-6708/2008/04/039.

[58] M. Peskin and T. Takeuchi, "New constraint on a strongly interacting Higgs sector", *Physical Review Letters* **65** (1990) 964–967. doi:10.1103/PhysRevLett.65.964.

[59] J. AguilarSaavedra, "Identifying top partners at LHC", *Journal of High Energy Physics* **2009** (2009), no. 11, 030. doi:10.1088/1126-6708/2009/11/030.

[60] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (2008), no. 08, S08001. doi:10.1088/1748-0221/3/08/S08001.

[61] M. Benedikt, P. Collier, V. Mertens et al., "LHC Design Report", volume 3: The LHC injector chain. CERN, Geneva, 2004. CERN-2004-003-V-3, http://cdsweb.cern.ch/record/823808.

[62] C. Lefèvre, "The CERN accelerator complex". CERN-DI-0812015, http://cdsweb.cern.ch/record/1260465, 2008.

[63] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08004. doi:10.1088/1748-0221/3/08/S08004.

[64] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08003. doi:10.1088/1748-0221/3/08/S08003.

[65] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08007. doi:10.1088/1748-0221/3/08/S08007.

[66] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08006. doi:10.1088/1748-0221/3/08/S08006.

[67] LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08005. doi:10.1088/1748-0221/3/08/S08005.

[68] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08002. doi:10.1088/1748-0221/3/08/S08002.

[69] CMS Trigger and Data Acquisition Group, "The CMS high level trigger", *The European Physical Journal* **C46** (2006) 605–667. doi:10.1140/epjc/s2006-02495-8.

[70] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013. https://cds.cern.ch/record/1598864.

[71] CMS Collaboration, "Absolute Calibration of Luminosity Measurement at CMS: Summer 2011 Update", CMS Physics Analysis Summary CMS-PAS-EWK-11-001, 2011. https://cds.cern.ch/record/1376102.

[72] CMS Collaboration, "CMS Luminosity Public Results". Topic revision: r101 - 24 Jan 2014, https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[73] CMS Collaboration, "Technical proposal for the upgrade of the CMS detector through 2020", Technical Report CERN-LHCC-2011-006. LHCC-P-004, 2011. https://cds.cern.ch/record/1355706.

[74] J. Mans, CMS Collaboration, "CMS Phase 2 Upgrade: Scope and R & D goals". https://indico.cern.ch/event/315626/session/11/contribution/22/material/slides/1.pdf, 2014.

[75] M. Seymour and M. Marx, "Monte Carlo Event Generators", arXiv:1304.6677.

[76] J. Pumplin, D. Stump, J. Huston et al., "New generation of parton distributions with uncertainties from global QCD analysis", *Journal of High Energy Physics* **0207** (2002) 012. doi:10.1088/1126-6708/2002/07/012.

[77] T. Stelzer and W. Long, "Automatic generation of tree level helicity amplitudes", *Computer Physics Communications* **81** (1994) 357–371. doi:10.1016/0010-4655(94)90084-1.

[78] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5 : Going Beyond", *Journal of High Energy Physics* **1106** (2011) 128. doi:10.1007/JHEP06(2011)128.

[79] M. Dobbs, S. Frixione, E. Laenen et al., "Les Houches guidebook to Monte Carlo generators for hadron collider physics", arXiv:hep-ph/0403045.

[80] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *Journal of High Energy Physics* **0411** (2004) 040. doi:10.1088/1126-6708/2004/11/040.

[81] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *Journal of High Energy Physics* **2007** (2007), no. 11, 070. doi:10.1088/1126-6708/2007/11/070.

[82] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *Journal of High Energy Physics* **2010** (2010), no. 6,. doi:10.1007/JHEP06(2010)043.

[83] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *Journal of High Energy Physics* **0605** (2006) 026. doi:10.1088/1126-6708/2006/05/026.

[84] G. Corcella, I. Knowles, G. Marchesini et al., "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", *Journal of High Energy Physics* **0101** (2001) 010. doi:10.1088/1126-6708/2001/01/010.

[85] M. Mangano, M. Moretti, F. Piccinini et al., "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *Journal of High Energy Physics* **0701** (2007) 013. doi:10.1088/1126-6708/2007/01/013.

[86] S. Agostinelli, J. Allison, K. Amako et al., "Geant4a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303. doi:10.1016/S0168-9002(03)01368-8.

[87] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009. https://cds.cern.ch/record/1194487.

[88] W. Adam, B. Mangano, T. Speer et al., "Track Reconstruction in the CMS tracker", CMS Note CMS-NOTE-2006-041, 2006. https://cds.cern.ch/record/934067.

[89] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* **9** (2014), no. 10, P10009. doi:10.1088/1748-0221/9/10/P10009.

[90] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Fluids Engineering* **82** (1960) 35–45. doi:10.1115/1.3662552.

[91] S. Cucciarelli, M. Konecki, D. Kotlinski et al., "Track reconstruction, primary vertex finding and seed generation with the Pixel Detector", CMS Note CMS-NOTE-2006-026, 2006. https://cds.cern.ch/record/927384.

[92] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239. doi:10.1109/5.726788.

[93] W. Waltenberger, R. Frhwirth, and P. Vanlaer, "Adaptive vertex fitting", *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 12, N343. doi:10.1088/0954-3899/34/12/N01.

[94] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006. CMS-TDR-008-1, CERN-LHCC-2006-001, http://cdsweb.cern.ch/record/922757.

[95] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-003, 2010. https://cds.cern.ch/record/1279347.

[96] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-MUO-10-002, 2010. https://cds.cern.ch/record/1279140.

[97] CMS Collaboration, "Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector", CMS Physics Analysis Summary CMS-PAS-JME-12-002, 2013. https://cds.cern.ch/record/1543527.

[98] CMS Collaboration, "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging", CMS Physics Analysis Summary CMS-PAS-JME-09-001, 2009. https://cds.cern.ch/record/1194489.

[99] S. Ellis and D. Soper, "Successive combination jet algorithm for hadron collisions", *Physical Review* **D48** (1993) 3160–3166. doi:10.1103/PhysRevD.48.3160.

[100] M. Cacciari, G. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *Journal of High Energy Physics* **0804** (2008) 063. doi:10.1088/1126-6708/2008/04/063.

[101] Y. Dokshitzer, G. Leder, S. Moretti et al., "Better jet clustering algorithms", *Journal of High Energy Physics* **9708** (1997) 001. doi:10.1088/1126-6708/1997/08/001.

[102] CMS Collaboration, "Boosted Top Jet Tagging at CMS", CMS Physics Analysis Summary CMS-PAS-JME-13-007, 2013. https://cds.cern.ch/record/1647419.

[103] CMS Collaboration, "Study of Pileup Removal Algorithms for Jets", CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014. https://cds.cern.ch/record/1751454.

[104] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation* **6** (2011), no. 11, P11002. doi:10.1088/1748-0221/6/11/P11002.

[105] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Physics Letters* **B659** (2008) 119–126. doi:10.1016/j.physletb.2007.09.077.

[106] CMS Collaboration, "8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS", CMS Detector Performance Summary CMS-DP-2013-033, 2013. https://cds.cern.ch/record/1627305.

[107] UA2 Collaboration, "Measurement of production and properties of jets at the CERN $\bar{\text{p}}\text{p}$ collider", *Zeitschrift für Physik C Particles and Fields* **20** (1983), no. 2, 117–134. doi:10.1007/BF01573214.

[108] D0 Collaboration, "Determination of the absolute jet energy scale in the D calorimeters", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **424** (1999), no. 23, 352 – 394. doi:10.1016/S0168-9002(98)01368-0.

[109] CDF Collaboration, "Determination of the jet energy scale at the Collider Detector at Fermilab", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **566** (2006), no. 2, 375 – 412. doi:10.1016/j.nima.2006.05.269.

[110] CMS Collaboration, "Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-014, 2011. https://cds.cern.ch/record/1339945.

[111] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *Journal of Instrumentation* **8** (2013) P04013. doi:10.1088/1748-0221/8/04/P04013.

[112] J. Conway, J. Dolen, R. Erbacher et al., "Boosted Top Jet Tagging at CMS", CMS Note 2013/029, 2013.

[113] J. M. Butterworth, A. R. Davison, M. Rubin et al., "Jet substructure as a new Higgs search channel at the LHC", *Phys.Rev.Lett.* **100** (2008) 242001. doi:10.1103/PhysRevLett.100.242001.

[114] S. Ellis, C. Vermilion, and J. Walsh, "Techniques for improved heavy particle searches with jet substructure", *Physical Review D* **80** (2009) 051501. doi:10.1103/PhysRevD.80.051501.

[115] CMS Collaboration, "Identification techniques for highly boosted W bosons that decay into hadrons", arXiv:1410.4227.

[116] D. Kaplan, K. Rehermann, M. Schwartz et al., "Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks", *Physical Review Letters* **101** (2008) 142001. doi:10.1103/PhysRevLett.101.142001.

[117] CMS Collaboration, "Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s}$=8 TeV", *Physical Review Letters* **111** (2013) 211804. doi:10.1103/PhysRevLett.111.211804.

[118] CMS Collaboration, "Search for $W' \to t\bar{b}$ in the all-hadronic final state", CMS Physics Analysis Summary CMS-PAS-B2G-12-009, 2014. https://cds.cern.ch/record/1751504.

[119] D. Soper and M. Spannowsky, "Finding top quarks with shower deconstruction", *Physical Review* **D87** (2013), no. 5, 054012, arXiv:1211.3140. doi:10.1103/PhysRevD.87.054012.

[120] CMS Collaboration, "Boosted Top Jet Tagging at CMS", Technical Report CMS DP-2014/036, 2014. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMSDP-2014/036.

[121] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *Journal of High Energy Physics* **1103** (2011) 015. doi:10.1007/JHEP03(2011)015.

[122] S. Catani, Y. Dokshitzer, M. Seymour et al., "Longitudinally-invariant k-clustering algorithms for hadron-hadron collisions", *Nuclear Physics B* **406** (1993), no. 12, 187 – 224. doi:10.1016/0550-3213(93)90166-M.

[123] CMS Collaboration, "Inclusive search for a vector-like T quark with charge $\frac{2}{3}$ in pp collisions at $\sqrt{s} = 8$ TeV", *Physics Letters B* **729** (2014), no. 0, 149 – 171. doi:10.1016/j.physletb.2014.01.006.

[124] D. Krohn, M. Schwartz, T. Lin et al., "Jet Charge at the LHC", *Physical Review Letters* **110** (2013), no. 21, 212001. doi:10.1103/PhysRevLett.110.212001.

[125] A. Larkoski, G. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure", *Journal of High Energy Physics* **1306** (2013) 108. doi:10.1007/JHEP06(2013)108.

[126] S. Ellis, A. Hornig, T. Roy et al., "Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure", *Physical Review Letters* **108** (2012) 182003. doi:10.1103/PhysRevLett.108.182003.

[127] "Theta framework web page". http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto/.

[128] F. James, "Statistical Methods in Experimental Physics: 2nd Edition". World Scientific Publishing Company, 2006. ISBN-13: 978-9812705273.

[129] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Computer Physics Communications* **77** (1993), no. 2, 219 – 228. doi:10.1016/0010-4655(93)90005-W.

[130] J. Conway, "Nuisance parameters in likelihoods for multisource spectra", *Proceedings of PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding* (2011). https://cds.cern.ch/record/1306523.

[131] R. Neal, "Probabilistic Inference Using Markov Chain Monte Carlo Methods", Technical Report CRG-TR-93-1, 1993. http://www.cs.toronto.edu/~radford/ftp/review.pdf.

[132] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *Journal of High Energy Physics* **0709** (2007) 126. doi:10.1088/1126-6708/2007/09/126.

[133] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", arXiv:1112.5675.

[134] M. Czakon, P. Fiedler, and A. Mitov, "The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$", *Physical Review Letters* **110** (2013) 252004. doi:10.1103/PhysRevLett.110.252004.

[135] M. Cacciari, M. Czakon, M. Mangano et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Physics Letters B* **710** (2012) 612–622. doi:10.1016/j.physletb.2012.03.013.

[136] CMS Collaboration, "Jet Energy Corrections and Uncertainties. Detector Performance Plots for 2012.", CMS Detector Performance Summary CMS-DP-2012-012, 2012. https://cds.cern.ch/record/1460989.

[137] R. Höing, I. Marchesini, A. Schmidt et al., "Search for top-Higgs resonances in all-hadronic final states using substructure methods", *CMS Analysis Note* **CMS-AN-13-129** (2013). Internal documentation.

[138] J. Pumplin, D. Stump, R. Brock et al., "Uncertainties of predictions from parton distribution functions. 2. The Hessian method", *Physical Review* **D65** (2001) 014013. doi:10.1103/PhysRevD.65.014013.

[139] J. Campbell, J. Huston, and W. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics", *Reports on Progress in Physics* **70** (2007) 89. doi:10.1088/0034-4885/70/1/R02.

[140] CMS Collaboration, "Search for pair production of vector-like partners of the top quark (T), with T $\to$ tH, H $\to \gamma\gamma$", CMS Physics Analysis Summary CMS-PAS-B2G-14-003, 2014. https://cds.cern.ch/record/1709129.

[141] ATLAS Collaboration, "Search for heavy top-like quarks decaying to a Higgs boson and a top quark in the lepton plus jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", Technical Report ATLAS-CONF-2013-018, 2013. https://cds.cern.ch/record/1525525.

[142] CMS Collaboration, "Measurement of the differential $t\bar{t}$ cross section in the dilepton channel at 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-028, 2013. https://cds.cern.ch/record/1523664?ln=en.

Viele Personen haben mich auf unterschiedliche Weise bei der Fertigstellung dieser Arbeit unterstützt. Mein besonderer Dank gilt meinem Doktorvater Alexander Schmidt für die engagierte Betreuung in den letzten drei Jahren. Deine unerschütterliche Begeisterung für die Analyse war immer sehr motivierend für mich. Auch für die Gelegenheit einige Monate am CERN zu arbeiten und viele Konferenzen und Sommerschulen besuchen zu dürfen, bin ich dir sehr dankbar. Das waren wirklich sehr spannende und wertvolle Erfahrungen für mich.

Auch Ivan Marchesini stand mir immer mit Ratschlägen und Hilfestellungen bei der Konzeption und Durchführung meiner Analyse zur Seite. Ich habe sehr viel von dir gelernt, ganz herzlichen Dank für die immer freundliche und geduldige Betreuung.

Vielen Dank auch an die weiteren Gutachter meiner Dissertation und Disputation Johannes Haller, Günter Sigl, Peter Schleper und Christian Sander, die alle im vollen Terminkalender der Vorweihnachtszeit noch Zeit für meine Prüfung gefunden haben.

Die drei Jahre meiner Promotion in Hamburg werde ich sicherlich in sehr guter Erinnerung behalten. Die offene, freundliche und hilfsbereite Atmosphäre in unserer Arbeitsgruppe hat dazu beigetragen, dass ich immer gerne in das Institut am DESY gekommen bin. Vielen Dank dafür an alle, die in den letzten drei Jahren Mitglied unserer Arbeitsgruppe waren, insbesondere natürlich an die Damen aus Büro Nr. 103.

Und ganz sicher nicht zuletzt: Ganz herzlichen Dank an meine Familie. Begonnen mit dem Umzug aus Aachen bis zur stressigen Abgabephase in den letzten Monaten wart ihr immer für mich da, wenn ich Hilfe brauchte. Dass ihr zu meiner Disputation nach Hamburg gekommen seid, hat den Tag für mich noch schöner gemacht.

Dankeschön!