Search for vector-like quarks using jet substructure techniques with the CMS experiment

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik, Informatik und

Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

Dominik Nowatschin

aus Tettnang

Hamburg

2017

Gutachter der Dissertation:	Prof. Dr. Johannes Haller Dr. Andreas Meyer
Zusammensetzung der Prüfungskommission:	Prof. Dr. Johannes Haller Dr. Andreas Meyer Prof. Dr. Caren Hagner Prof. Dr. Gudrid Moortgat-Pick Dr. Alexander Schmidt
Vorsitzender der Prüfungskommission:	Prof. Dr. Caren Hagner
Datum der Disputation:	03.07.2017
Vorsitzender Fach-Promotionsausschusses Physik:	Prof. Dr. Wolfgang Hansen
Leiterin des Fachbereichs Physik:	Prof. Dr. Michael Potthoff
Dekan der Fakultät für Mathematik, Informatik und Naturwissenschaften:	Prof. Dr. Heinrich Graener

List of Publications

The research presented in this thesis has resulted in the following public document and publication:

CMS Collaboration, "Search for vector-like quark pair production in final states with leptons and boosted Higgs bosons at $\sqrt{s} = 13$ TeV", CMS-PAS-B2G-16-011, 2016

CMS Collaboration, "Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure techniques at $\sqrt{s} = 13$ TeV", 2017, submitted to JHEP, `arXiv:1706.03408`

Abstract

In this thesis, a search for pair produced vector-like T quarks in pp collision data from the LHC at $\sqrt{s} = 13\,\text{TeV}$ is presented. The data were collected with the CMS detector and correspond to an integrated luminosity of up to $2.6\,\text{fb}^{-1}$. Vector-like quarks are hypothetical new particles predicted by many extensions of the Standard Model in which the Higgs boson is a composite state of an unknown strong interaction. Vector-like T quarks are assumed to decay via three different decay modes to either bW, tZ or tH, with branching fractions that are not fixed and can vary depending on the particular model featuring vector-like quarks.

This search focuses on decays of the $T\overline{T}$ system in which at least one muon or electron is present in the final state, and in which at least one of the T quarks decays to a top quark and a Higgs boson. As the T quarks are expected to be quite heavy, their decay products are significantly Lorentz-boosted in the reference frame of the $T\overline{T}$ system. The subsequent decay products of the Higgs boson are then emitted with a very small angle between them. This search is optimised for the main decay channel of the Higgs boson to two bottom quarks and attempts to reconstruct the two b quarks within a single jet with a large cone size. Dedicated jet substructure techniques, in combination with algorithms to identify jets originating from the fragmentation of a b quark, are then used to reconstruct the entire $H \to b\bar{b}$ decay.

The event categories of this search are also combined with the categories of a search for pair-produced T quarks that focuses on the $T \to bW$ decay. This approach provides a high sensitivity to $T\overline{T}$ production for many different combinations of branching fractions to the three possible decay modes.

No excess of the data above the expected background contribution from the Standard Model is observed in any of the final event categories. Upper limits on the $T\overline{T}$ production cross section are calculated at 95% confidence level and lower mass limits for vector-like T quarks are set for various assumptions of their decays: for branching fractions corresponding to an electroweak isospin singlet T quark, masses below 860 GeV are excluded, while for an electroweak isospin doublet, T quarks are excluded with masses up to 830 GeV. For other branching fraction combinations, lower mass limits are derived ranging from 710 GeV to 940 GeV. These exclusion limits are comparable to and for some branching fraction assumptions even higher than exclusion limits set by the CMS experiment using a data set corresponding to $19.7\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$.

Kurzfassung

In der vorliegenden Arbeit wird eine Suche nach paarproduzierten vektorartigen T-Quarks präsentiert, die pp-Kollisionsdaten des LHC Beschleunigers bei einer Schwerpunktsenergie von 13 TeV nutzt. Diese Daten wurden mit dem CMS Detektor aufgezeichnet und entsprechen einer integrierten Luminosität von $2,6\,\mathrm{fb}^{-1}$. Vektorartige Quarks sind hypothetische neue Teilchen die in vielen Erweiterungen des Standardmodells vorhergesagt werden, in welchen das Higgs-Boson ein zusammengesetztes Teilchen einer unbekannten starken Wechselwirkung ist.

Diese hier vorgestellte Suche konzentriert sich hauptsächlich auf Zerfälle des $T\overline{T}$-Systems, in welchen wenigstens ein Myon oder Elektron erzeugt wird und mindestens ein T-Quark in ein Top-Quark und ein Higgs-Boson zerfällt. Da man erwartet, dass T-Quarks sehr hohe Massen besitzen, folgt, dass deren Zerfallsprodukte einen hohen Lorentz-Boost im Schwerpunktssystem des $T\overline{T}$-Systems besitzen. Die resultierenden Zerfallsprodukte des Higgs-Bosons haben daher häufig einen sehr kleinen Austrittswinkel. Diese Suche ist optimiert für den Hauptzerfallskanal des Higgs-Bosons in zwei Bottom Quarks und versucht, diese beiden b-Quarks in einem einzigen Jet mit einem großen Radius zu rekonstruieren. Desweiteren werden speziell dafür vorgesehene Substrukturmethoden verwendet um, in Kombination mit weiteren Algorithmen zur Identifizierung von Jets aus der Fragmentierung eines b-Quarks, den kompletten $H \to b\bar{b}$ Zerfallsprozess zu rekonstruieren.

Die Ereigniskategorien, die in dieser Suche definiert sind, werden außerdem mit den Kategorien einer weiteren Suche nach paarproduzierten T-Quarks kombiniert, die sich auf $T \to bW$ Zerfälle konzentriert. Diese Vorgehensweise ermöglicht eine hohe Sensitivität für $T\overline{T}$-Produktion für zahlreiche Kombinationen des T-Quark Verzweigungsverhältnisses zu den verschiedenen Zerfallskanälen.

In keiner der finalen Ereigniskategorien wird ein Exzess der Daten über dem erwarteten Standardmodelluntergrund beobachtet. Obere Ausschlussgrenzen für den $T\overline{T}$-Produktionswirkungsquerschnitt werden mit einem Konfidenzniveau von 95% berechnet. Untere Grenzen auf die Masse von vektorartigen T-Quarks werden unter Annahme verschiedener Verzweigungsverhältnisse gesetzt: Für den Fall, dass T-Quarks in einem elektroschwachen Isospin-Singlet erscheinen werden Massen unterhalb von 860 GeV ausgeschlossen. Unter der Annahme, dass T-Quarks in einem elektroschwachen Isospin-Doublet realisiert sind, werden Massen bis 830 GeV ausgeschlossen. Untere Ausschlussgrenzen für weitere Kombinationen der Verzweigungsverhältnisse des T-Quarks werden berechnet, diese reichen je nach Annahme des Verzweigungsverhältnisses von 710 GeV bis 940 GeV.

Diese Ausschlussgrenzen sind vergleichbar mit und teilweise sogar stärker als Grenzen, die in mehreren Suchen am CMS Experiment bei einer Schwerpunktsenergie von $\sqrt{s} = 8\,\mathrm{TeV}$ und mit einem Datensatz, der einer integrierten Luminosität von $19,7\,\mathrm{fb}^{-1}$ entspricht, durchgeführt wurden .

Contents

1 Introduction

The search for a fundamental mechanism to describe the laws of nature has passed through many different stages over the last century. From the development of quantum mechanics at the beginning of the 20th century until the formulation of the Standard Model of particle physics in the 1970s, there have been both times in which theorists could barely keep up with new observations and phases in which the prediction of new particles preceded their actual discovery by many years. The culmination of this process was the discovery of the Higgs boson by the ATLAS and CMS collaborations almost 50 years after its postulation in 2012 which marked the end of an era dedicated to the validation of the Standard Model.

The discovery of the Higgs boson has now initiated a new stage in the high energy physics and the current situation is in some ways quite different from previous times. In the past, physicists were often confronted with direct measurements contesting their theories e. g. when the development of particle accelerators and novel measuring resulted in an unforeseen discovery of a large number of new particles. The Standard Model, however, agrees with the outcomes of a large number of high energy experiments to an enormous precision, yet, it faces problems that the majority of physicists believe can only be solved if new, unknown mechanisms exist.

One of its open questions is, for instance, why there is such a huge difference between the electroweak scale and the gravitational scale. In any physical theory obeying the principle of naturalness, such large discrepancies are excluded by symmetries within the theory. While there are symmetries in the Standard Model that protect the masses of the fermions and massive gauge bosons, there is none that keeps the mass of the Higgs boson at its observed low value. The only way to explain this in the current formulation of the Standard Model is by fine-tuning the bare Higgs boson mass with an immense accuracy which contradicts the idea of naturalness.

Many extensions of the Standard Model introduce new particles to avoid this fine tuning. While some of the most prominent theoretical models, which are supersymmetric extensions of the Standard Model, postulate bosonic partners of the SM quarks that serve this purpose, a wide range of alternative models predict fermionic partners. These new resonances are expected to have masses around the TeV scale if they were to restore the naturalness of the electroweak scale without introducing new sources of fine tuning. It would therefore be possible to observe them at the LHC already with a relatively small data set after its upgrade to a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Consequently, the first data set collected after the restart of LHC in 2015 is of high significance in the search

for extensions of the Standard Model motivated by the idea of naturalness.

This thesis presents a search for such fermionic resonances, so-called vector-like T quarks, in proton-proton collision data from the LHC. The first data set produced in 2015 at $\sqrt{s} = 13\,\text{TeV}$ is studied which was recorded by the CMS experiment and corresponds to an integrated luminosity of up to $2.6\,\text{fb}^{-1}$. Jet substructure techniques are used in this search to identify hadronically decaying Higgs bosons with very large transverse momenta, which are possible decay products of a T quark. The main background processes arise from $t\bar{t}$ + jets and W + jets production and are estimated from simulation, with corrections applied to account for differences observed between data and simulation in dedicated control regions.

This thesis is structured as follows: the phenomenological properties of vector-like quarks are introduced in Chapter 2 along with an introduction of the Standard Model and an overview over theoretical models that incorporate these new particles. The LHC and the CMS experiment with technical details about its subdetectors are then described in Chapter 3. Chapter 4 provides a brief summary of Monte-Carlo simulation techniques to compute expected yields and the shape of various observables for many physical processes. The most important algorithms to reconstruct hard scattering processes occurring in proton-proton collisions are outlined in Chapter 5. In Chapter 6, the search for vector-like quarks is described in detail, including an outline of the search strategy, the estimation of background processes and systematic uncertainties, and a discussion of the final results. This search is then combined with a complementary analysis searching for vector-like T quarks decaying to a bottom quark and a W boson to increase the sensitivity for other decay modes of the T quark. The combination strategy and final results are presented in Chapter 7. Chapter 8 finally summarises the contents of this thesis and provides an outlook into future searches for such new resonances.

The author of this thesis has contributed to its content as outlined in the following: he has been the main author of the search for vector-like T quarks decaying to boosted Higgs bosons which is described in Chapter 6. This search resulted in a public preliminary result released by the CMS collaboration [1]. He has worked as co-author in the combination of that search with a complementary search for vector-like quarks which is outlined in Chapter 7, contributing both to the estimation of correlated systematic uncertainties and to the technical implementation of the combination. This combination, along with an update of the preliminary result, has been published by the CMS Collaboration and submitted to the Journal of High Energy Physics [2]. Additionally, he performed a study about the impact of nuclear interactions on b tagging algorithms which is described in Section 5.7.3. This study was performed in the scope of a campaign of the b tagging working group within the CMS collaboration to improve the performance of b tagging algorithms for very high-p_T jets.

2 The Standard Model and possible extensions

2.1 The Standard Model of particle physics

All fundamental particles known to date and the interactions between them are described in the *Standard Model* (SM) of particle physics. The Standard Model has been developed over many years, starting in the 60s with a series of papers by Glashow, Weinberg and Salam about the theory of electroweak unification [3–5]. It has had a remarkable success since then, providing exact predictions about particle dynamics that have been experimentally confirmed to an enormous precision. The latest milestone in its verification was the proof of the Brout-Englert-Higgs mechanism [6–11] through the discovery of the Higgs boson by the ATLAS and CMS Collaborations [12–14]. This section provides a brief introduction into the particle content of the Standard Model and its mathematical formulation, more detailed descriptions can be found in a broad range of review articles and textbooks, e. g. in Refs. [15, 16]. Throughout this chapter and the rest of this thesis, natural units are used by setting $c = \hbar = 1$.

The particles in the Standard Model can be grouped into *fermions* and *bosons* with half-integer and integer spin, respectively. The group of bosons comprise the *gauge bosons* with spin 1 which mediate the three fundamental interactions between all particles in the Standard Model: the *electromagnetic interaction*, which is mediated by the photon (γ) and is responsible e. g. for the formation of atoms and molecules, the *weak interaction*, which is mediated by the Z and the $W^{\pm}$ bosons and is responsible e. g. for radioactive β-decays, and the *strong interaction* which is mediated by the gluon (g) and binds together quarks to form hadrons. A fourth hypothetical gauge boson is the *graviton* which mediates the gravitational force and would have a spin of 2. It is, however, not yet possible to incorporate gravity into the Standard Model. The final particle in the group of bosons is the aforementioned *Higgs boson* with spin 0 which is a consequence of the Brout–Englert–Higgs mechanism by which the Z and $W^{\pm}$ bosons as well as the fermions acquire their masses (this is described in more detail in Section 2.1.3).

The fermions in the Standard Model can again be divided into two groups, the *quarks* and *leptons*, which provide the fundamental constituents of matter and differ by the interactions they are subject to. For both, quarks and leptons, six different particles exist distinguished by their *flavour* which are grouped into three *generations*. There are three up-type quarks with an electromagnetic charge of $+2e/3$, the *up* (u), *charm* (c) and *top* (t) quarks, and three down-type quarks with an electromagnetic charge of $-1e/3$, the *down*

(d), *strange* (s) and *bottom* (b) quarks. In the leptonic sector there exist three charged particles, the *electron* (e), the *muon* (μ) and the *tau lepton* (τ) with electromagnetic charges of $+1e$ and three neutrinos, the electron-, muon- and tau-neutrino ($\nu_{e,\mu,\tau}$) which are uncharged and only interact via the weak force. In the Standard Model, they are regarded as massless, though recent observations of neutrino oscillations indicate that they have a non-zero mass [17–21]. For each fermion in the Standard Model, there exists in addition an *anti-particle* with the exact same mass, but opposite-signed quantum numbers. An overview over the particle content of the Standard Model can be found in Fig. 2.1.

The Standard Model is described mathematically as a *quantum field theory* so that elementary particles correspond to excitations of fundamental fields. The dynamics of these fields are described by the *Lagrangian density* (or simply *Lagrangian*) $\mathcal{L}$. From this, one can derive relativistic wave equations for each type of particle e. g. the Dirac equation for spin-1/2 particles

$$i\gamma^\mu \partial_\mu \psi - m\psi = 0 \tag{2.1}$$

where ψ is a Dirac spinor and γ^μ are the gamma matrices.

In the limit of weak interactions between the fields, decay rates and scattering cross sections can be calculated perturbatively. These quantities are proportional to the *matrix element* (ME) describing the transition probability between the in- and out-states ψ_i and ψ_o

$$|\mathcal{M}|^2_{i,o} = \langle \psi_i | S | \psi_o \rangle^2. \tag{2.2}$$

Here, S is the scattering matrix which can be directly related to the corresponding interaction terms in the Lagrangian using the *Feynman rules*. These processes can also be graphically represented through *Feynman diagrams*. The matrix element can be calculated perturbatively by expanding it in powers of the coupling constant of the corresponding interaction.

Including higher-order terms in the perturbation series of such a calculation increases the accuracy of the calculation and corresponds to adding external legs and internal loops in the corresponding Feynman diagram. These loops can lead to divergent terms (so-called *ultra-violet* divergences) which can be eliminated by using *renormalised* quantities instead of the "bare" ones. The renormalised quantities depend on a chosen *renormalisation scale* μ_R which is introduced in the renormalisation procedure. If an infinite number of orders would be included in the calculation of a given process, the outcome would be independent of the chosen renormalisation scale. This is, however, unfeasible and usually only the first few orders of the perturbation series are included in such a calculation. As a result, the renormalised parameters exhibit a dependency on the chosen renormalisation scale, an example for which is the "running" of the electromagnetic coupling constant α_{em}. Usually, the renormalisation scale is set to the energy scale of the physical process under consideration.

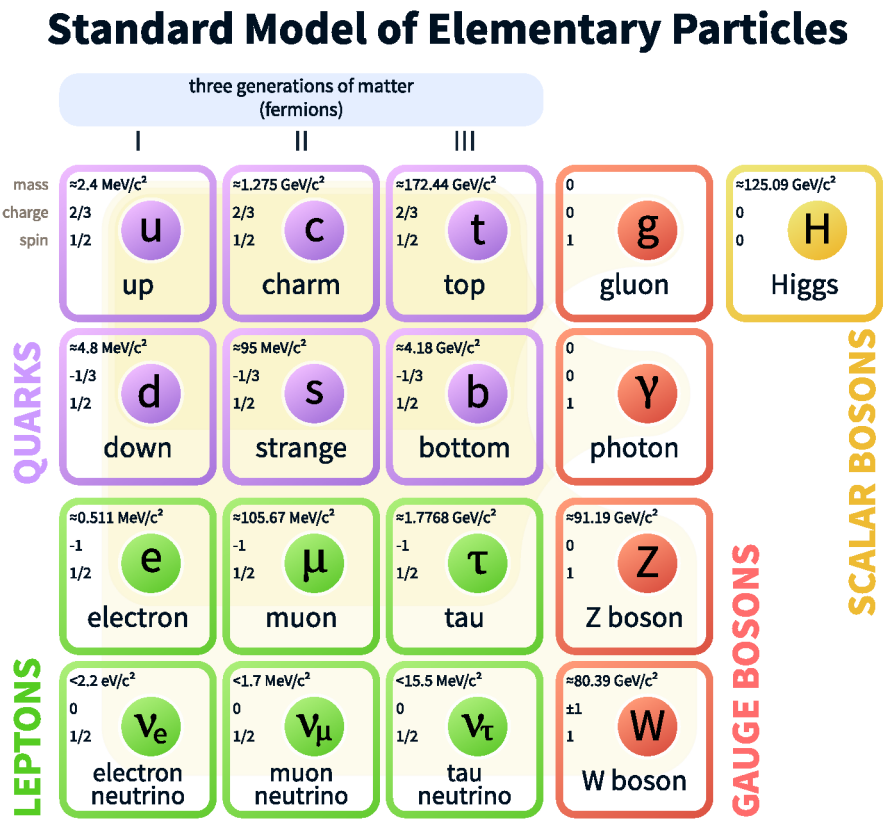

Figure 2.1: Particles in the Standard Model, grouped into quarks, leptons, gauge boson and scalar bosons, along with their masses, electromagnetic charges and values for spin. Taken from Ref. [22].

The SM Lagrangian is invariant under *local gauge transformations* which form the Lie group

$$SU(3)_C \times SU(2)_L \times U(1)_Y$$

where the subgroups $SU(3)_C$ and $SU(2)_L \times U(1)_Y$ describe the strong force and the *electroweak interaction*, respectively. The strong interaction and the particles that it couples to, quarks and gluons, are described in Section 2.1.1. The electroweak interaction, which unifies the electromagnetic and weak forces, is described in Section 2.1.2. The Brout-Englert-Higgs mechanism that leads to the breaking of the electroweak symmetry and creates mass terms for the gauge bosons of the weak interaction is described in more detail in Section 2.1.3.

2.1.1 Quarks and the strong interaction

Quarks are the only particles in the Standard Model that are subject to all fundamental interactions, including the strong interaction. They appear in nature as bound states of the strong interaction which is described by a field theory referred to as *Quantum Chromodynamics* (QCD). Each quark carries either a red, blue or green *colour charge* (or anti-red, anti-blue or anti-green) and transforms as a triplet under the QCD gauge group $SU(3)_C$. This gauge group is formed by eight generators represented by the Gell-Mann-matrices T^a that correspond to eight coloured gluon fields G_a^μ where $a = 1 \dots 8$. The kinetic and interaction terms of the QCD Lagrangian then read as

$$\mathcal{L}_{\text{QCD}} = \bar{q}_i i\gamma_\mu(\partial^\mu - ig_S T^a G_a^\mu)q_i - \frac{1}{4}G_{\mu\nu}G^{\mu\nu} \tag{2.3}$$

with

$$G^{\mu\nu} \equiv (\partial^\mu G_a^\nu - \partial^\nu G_a^\mu + g_S f_{abc} G_b^\mu G_c^\nu)T^a \tag{2.4}$$

where q_i are the quark fields with i running over all quark flavours, m_i are their masses, g_S is the strong coupling constant and f_{abc} are the structure constants of the $SU(3)$ group.

QCD is a *Non-Abelian* theory which and consequently, no gluon in the colour singlet state and all gluons are self-interacting. This is consistent with the absence of experimental evidence for uncoloured gluons in nature. This is important as this has various phenomenological implications and amongst others opens a variety of particle production processes at the LHC via *gluon fusion*.

An important feature of the strong interaction is that its renormalised coupling constant g_S is inversely proportional to the energy scale of the physical process. As a result, at low energies, g_S becomes very large and results in a strong binding force between quarks which is known as *confinement* and cannot be calculated perturbatively anymore. At high energies, however, g_S gets very low so that quarks are in principle free particles, an effect referred to as *asymptotic freedom*. In nature, quarks never appear as free states but are always contained in bound states like mesons or baryons. At colliders it is possible to produce separate quarks which, as they travel onwards, create a shower of light bound states in a process which is referred to as *hadronisation*. Such a shower can be measured in suitable detectors and is commonly known as *jet*.

The coupling of quarks to the weak force after electroweak symmetry-breaking is proportional to the elements of the *Cabibbo–Kobayashi–Maskawa* (CKM) matrix [23, 24] which determines the mixing of the mass eigenstates of the quarks q and the electroweak eigenstates q':

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.5}$$

The best-fit values for its parameters, assuming unitarity and the presence of only three quark generations, are [25]

$$V_{CKM} = \begin{pmatrix} 0.97427 & 0.22534 & 0.00351 \\ 0.22520 & 0.97344 & 0.0412 \\ 0.00867 & 0.0404 & 0.999146 \end{pmatrix} \tag{2.6}$$

Quarks can therefore transform from one flavour into another one through a charged weak current. It can be seen that each quark preferentially decays to the next lightest quark with a branching ratio that is determined from the elements of the CKM matrix.

2.1.2 The electroweak interaction

The electromagnetic and the charged weak interactions are described by the gauge groups $U(1)_{em}$ and $SU(2)_L$, respectively. The generator of the $U(1)_{em}$ group corresponds to the photon while two of the generators of the $SU(2)_L$ group correspond to the $W^{\pm}$ bosons. The subscript L denotes the fact that the charged weak interaction only couples to fermions with left-handed chiralities and naturally, the neutral weak interaction would be expected to follow the same behaviour with its mediator, the Z boson, corresponding to the third generator of the $SU(2)_L$ group. However, the Z boson has been observed to couple to both left- and right-handed fermions with differing coupling strengths. To resolve this contradiction, the concept of *electroweak unification* has been developed according to which the Z and γ bosons are mixtures of the gauge bosons of a unified interaction.

This unified force, referred to as *electroweak* interaction, is described by the symmetry group $SU(2)_L \times U(1)_Y$. There are four generators of this group and correspondingly, four vector fields, the B^μ and the W_i^μ (with $i = 1 \ldots 3$). The generators corresponding to the W_i^μ fields can be represented by the three Pauli-matrices T_i. Left-handed quark and lepton fields transform as doublets under the $SU(2)_L$ group (which is referred to as *weak isospin*). A quark doublet consists of an up-type and down-type quark of the same generation and a lepton doublet consists of a charged lepton and its associated neutrino. Right-handed fields transform as singlets under $SU(2)_L$ and there are no right-handed neutrino fields as they have never been observed in nature. Furthermore, all fields have a *weak hypercharge* $Y_W = 2(Q - T_3)$ that determines their coupling to the gauge field of the $U(1)_Y$ B^μ, where Q is the electromagnetic charge and T_3 is the third component of the weak isospin.

The physical states of the gauge bosons observed in nature emerge after a breaking of the electroweak symmetry group $SU(2)_L \times U(1)_Y$. This breaking, referred to *electroweak symmetry breaking*, is mediated by the Brout-Englert-Higgs mechanism which is described in more detail in Section 2.1.3.

The $W^{\pm}$ bosons correspond to the $W_{1,2}$ bosons before electroweak symmetry breaking

while the γ and Z bosons are mixtures of the B and W_3 bosons given by

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \tag{2.7}$$

where θ_W is the so-called *Weinberg angle*.

The Lagrangian of the electroweak sector before spontaneous symmetry breaking is then given by

$$\mathcal{L}_{\mathrm{EW}} = \bar{\psi}_L i \not{D}_L \psi_L + \bar{\psi}_R i \not{D}_R \psi_R - \frac{1}{4}(B_{\mu\nu}B^{\mu\nu} + W_{\mu\nu}W^{\mu\nu}), \tag{2.8}$$

with

$$\begin{aligned} \not{D}_L &\equiv \gamma_\mu(\partial^\mu + g' Y_W B^\mu - i g_W T^i W_i^\mu), \\ \not{D}_R &\equiv \gamma_\mu(\partial^\mu + g' Y_W B^\mu), \\ B^{\mu\nu} &\equiv \partial^\mu B^\nu - \partial^\nu B^\mu, \\ W^{\mu\nu} &\equiv (\partial^\mu W_i^\nu - \partial^\nu W_i^\mu + g_W \epsilon_{ijk} W_j^\mu W_k^\nu) T^i, \end{aligned}$$

where ψ_L and ψ_R are left- and right-handed SM fermionic fields, g' and g_W are the coupling constants of the gauge fields B^μ and W_i^μ, respectively, and ϵ_{ijk} is the structure constant of the $SU(2)$ Lie group.

2.1.3 The Brout-Englert-Higgs mechanism

Gauge invariance of the Lagrangian would require all gauge bosons to be massless. While this is the case for the photon (and also the gluon), the $W^\pm$ and the Z bosons are observed to be massive particles with masses of 80.385 ± 0.015 GeV and 91.1876 ± 0.0021 GeV [25], respectively. Furthermore, the differing mass terms between quarks (and leptons) belonging to the same $SU(2)_L$ doublet violate the gauge symmetry of the $U(2)_L \times U(1)_Y$ group.

This contradiction is resolved by introducing a formalism referred to as *Brout-Englert-Higgs mechanism* to induce the spontaneous breaking of the electroweak symmetry group. In this mechanism, a new complex doublet field called the *Higgs field* is introduced

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

which is associated with the *Higgs potential*

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2.$$

If the parameters μ^2 and λ are chosen such that $\mu^2 < 0$ and $\lambda > 0$, the Higgs potential is symmetric with respect to the origin but has a degenerate minimum at a non-zero value.

Consequently, the Higgs field obtains a *vacuum expectation value* (VEV) in its ground state

$$v = \sqrt{\frac{-\mu^2}{2\lambda}} \approx 246\,\text{GeV}.$$

This non-zero VEV corresponds to a *spontaneous breaking* of the symmetry of the Higgs potential. According to *Goldstone's theorem* [26–28], this results in three massless so-called *Goldstone bosons* that correspond to three of the four degrees of freedom of the Higgs field. These three Goldstone bosons become the longitudinal degrees of freedom of the $W^{\pm}$ and Z bosons. The remaining fourth degree of freedom of the Higgs field corresponds to a new particle with mass $m_H = \sqrt{2}\mu$ which has spin 0 and is referred to as *Higgs boson*. This field is invariant under gauge transformations and by choosing a unitary gauge, the gauge bosons of the weak interaction acquire mass terms

$$m_W = \frac{g_W v}{2}$$
$$m_Z = \frac{g_W v}{2\cos\theta_W}.$$

Besides the weak gauge bosons, the Higgs boson can couple also to itself and to the fermions in the Standard Model. The couplings to fermions are referred to as *Yukawa couplings* and generate mass terms for the fermions that do not violate the $SU(2)_L$ symmetry of the SM Lagrangian. The masses are given by

$$m_f = \frac{y_f v}{\sqrt{2}}$$

and are thus proportional to the individual Yukawa couplings y_f for each fermion.

The Higgs boson has a large number of decay modes, with branching fractions shown in Fig. 2.2 as a function of the Higgs boson mass. At the observed value of 125 GeV, the dominant decay mode is $H \to b\bar{b}$ with a branching fraction of 58%, followed by the decays $H \to W^+W^-$, $H \to gg$ and $H \to \tau\bar{\tau}$ with the values 22%, 8.6% and 6.3%, respectively. These branching fractions result from the couplings and thus partial decay widths being proportional to the masses of the corresponding particles. For the decay mode to two W bosons, the available phase space is rather limited as the two W bosons cannot be produced on-shell, thus reducing the corresponding branching fraction. The $H \to gg$ decay is mediated by a loop diagram as there is no direct coupling of the Higgs bosons to gluons. The same applies to the decay mode to two photons which has a very low branching fraction but, due to the very clear signature with low background contributions, provides a discovery channel with large sensitivity.

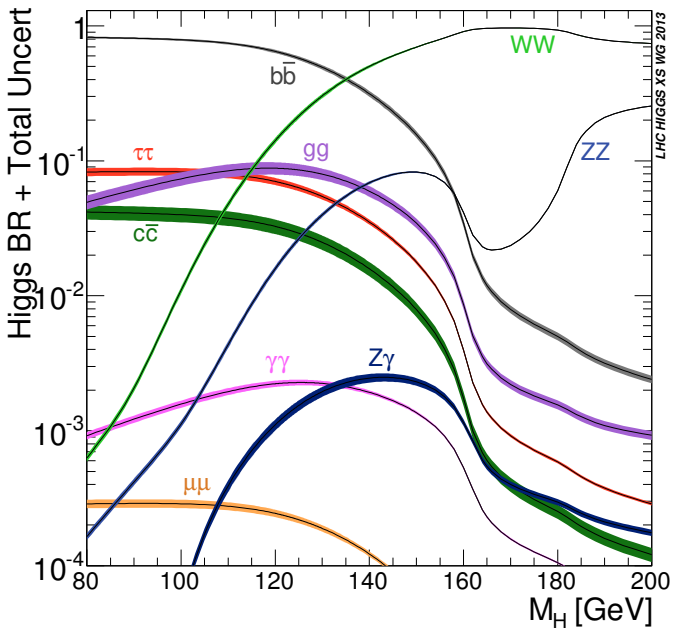

Figure 2.2: Branching fractions for various decay modes of the Higgs boson as a function of the Higgs boson mass. Taken from Ref. [29].

2.2 Shortcomings of the Standard Model

Despite the enormous success of the Standard Model, which has been confirmed in many experiments to a high precision, it has some imperfections and fails to explain all phenomena observed in nature. Some of the most important of these shortcomings are summarised in the following, followed by a discussion of several approaches to resolve them.

Dark matter According to Kepler's law of planetary motion, the velocity of an astronomical object on an orbit around a massive centre (like a star in a galaxy or a galaxy in a galaxy cluster) is proportional to $\sqrt{M(r)r^{-1}}$ where r is the radial distance and $M(r)$ is the mass contained within the orbit. Astronomical observations of many galaxies have shown that the velocity is nearly constant over a large range of the radial distance r, implying a mass density $\rho(r) \propto r^{-2}$. This, however, cannot be explained by the accumulated mass of the luminous matter observed in these galaxies alone which indicates the existence of additional non-luminous matter which is commonly referred to as *dark matter* (for a good overview over dark matter, see e. g. Refs. [30, 31]).

Dark matter particles can therefore only couple weakly to SM particles and further cosmological observations imply that they also have to be weakly self-interacting [32]. Potential candidates for dark matter in the Standard Model would be the neutrinos

as they have these properties and are found to have non-vanishing masses. However, they cannot alone explain the large abundance of non-luminous matter which indicates the existence of an unknown type of matter [33].

Baryogenesis The generation of nucleons in the early universe, which are the basic constituents of the bulk of existing matter, occurred through a process referred to as *Baryogenesis*. The CPT-symmetry of the SM Lagrangian then implies that matter and anti-matter would be produced in equal shares in this process. This is, however, in contrast to the large abundance of matter in the universe, which is the fundamental basis of our own existence. To explain this large asymmetry and incorporate Baryogenesis in a model describing the evolution of the early universe, three necessary conditions were described by Andrei Sakharov [34]. The first of these conditions is that a Lagrangian describing particle dynamics in the early universe must include terms that violate Baryon number conservation. Secondly, the C- and CP-symmetry of this Lagrangian must be broken as well. As a third condition, the rate of a baryon-generating process has to be slower than the rate of the expansion of the universe. The Standard Model partially satisfies these requirements by having CP-violating terms in the CKM matrix but not to an extent that could alone explain the observed matter/anti-matter asymmetry [35].

Naturalness One of the most widely debated subjects in the context of elementary particle physics over the last few decades is the idea of *naturalness*. This states that small parameters in an effective field theory – which the Standard Model is expected to be – correspond to an enhancement of the symmetry of the theory if the parameters are set to zero (see e. g. Ref. [36] for one of the first accounts of this topic). While this is true for the masses of gauge bosons and fermions in the SM Lagrangian, which are protected by gauge and chiral symmetry, respectively, there is no corresponding symmetry to keep the Higgs boson mass as low as it has been observed. Writing down the higher-order expansion of its renormalised mass results in quantum corrections of the form

$$\Delta m_H \propto y_f^2 \Lambda^2$$

where y_f is the Yukawa coupling for a fermion f and Λ is the cutoff scale up to which the Standard Model is considered to be valid. If one sets Λ to the Planck scale M_P, thus assuming that there are no new physical dynamics up to this scale[1], all quantum corrections have to cancel each other out with an enormous precision to keep the Higgs boson mass at its observed low value. This requires an immense

[1] Above the Planck scale, there must exist new dynamics since at this energy scale, the strength of gravity becomes comparable to the other fundamental forces and can no longer be neglected in a fully comprehensive field theory.

fine tuning of the bare mass term of the Higgs boson which is usually considered as unnatural and also referred to as *Hierarchy problem*.

There are many solutions proposed to resolve some or all of the problems of the Standard Model. One suggestion is the idea of the *multiverse* (see e. g. Ref. [37]) which assumes that the universe we are living in is just one of many simultaneously existing universes that happens to satisfy all necessary conditions to enable our existence (this does not have to be the case in other universes). This is in principle an application of the *anthropic principle* which justifies unnatural parameters in a theoretical framework by the fact that, if they would not have the values that we observe, we would not exist and therefore also not be able to make such observations [38].

The most widespread belief is, however, that there is a more fundamental theory than the Standard Model in which the aforementioned problems are resolved e. g. by introducing new symmetries. In such a theory, the Standard Model is an effective field theory describing interactions and particle dynamics at low energies. The consequence is then that new physical dynamics, commonly labelled as *physics beyond the Standard Model*, exist at higher energy scales that might, however, be observable in current collider experiments.

A variety of theoretical frameworks exists which describe such new physical dynamics and which include new particles, interactions and symmetries. Some of the most popular among these models are *supersymmetric* extensions of the Standard Model in which each SM particle has one or more supersymmetric partners with a spin differing by $1/2$[2]. Bosonic partners of the SM fermions can then e. g. cancel loop corrections to the Higgs boson mass, thus resolving the previously described hierarchy problem. Furthermore, many supersymmetric models feature at least one stable, weakly interacting particle which is a natural candidate to explain dark matter.

Due to the very large number of free parameters in supersymmetric models, there is an immense number of possible realisations of this theory, not all of which, however, can resolve all of the previously listed problems of the Standard Model. A considerable portion of this parameter space has also already been excluded in direct searches at the LHC (see e. g. Ref. [40] for a summary of constraints on the supersymmetric parameter space after the first data-taking period of the LHC at $\sqrt{s} = 13\,\text{TeV}$).

It is therefore worth exploring alternatives to supersymmetric models and one of the most prominent frameworks, referred to as *Composite Higgs* (CH) model or *Compositeness*, is presented in Section 2.3. A common feature in many of these alternative models are *vector-like quarks* (VLQ) which are hypothetical new particles that are likely to have masses around the TeV scale if they exist. These particles are the main target of the searches presented in Chapters 6 and 7 and are described in detail in Section 2.4.

[2] A detailed review of supersymmetric models can e. g. be found in Ref. [39].

2.3 Composite Higgs models

The basic idea of Composite Higgs models was laid out in a series of papers in the 80s [41–47] which proposed that the Higgs boson is not an elementary particle but a composite state of an underlying strong sector. This section gives an overview over these models and is based on Refs. [48–50] which provide detailed reviews about a broad range of CH models.

An equivalent mechanism, referred to as *chiral symmetry breaking*, is already realised in the Standard Model. This is briefly described in Section 2.3.1 in order to explain the fundamental idea of CH models. These are then outlined in Section 2.3.2 along with some of their general properties. Section 2.3.3 describes a popular feature of many CH models to generate mass terms for the SM fermions, referred to as *partial compositeness*. Finally, in Section 2.3.4 a particular set of models incorporating compositeness is described, so-called *warped extra dimension models*.

2.3.1 Chiral symmetry breaking in QCD

Chiral symmetry breaking refers to the symmetry of the strong sector of the SM Lagrangian under chiral transformations that belong to the group $SU(2)_L \times SU(2)_R$, denoted as *chiral symmetry*. A quark condensate with a VEV proportional to the *confinement scale* Λ_{QCD} of QCD[3] leads to a spontaneous breaking of this symmetry and consequently creates three Goldstone bosons, the pions. As the quarks have, however, small but non-vanishing masses and therefore explicitly break the chiral symmetry in the low-energy regime, the pions are in fact *pseudo-Goldstone bosons* and have small masses themselves. As for large energy scales above Λ_{QCD}, chiral symmetry is approximately conserved – at least in the part of the strong SM Lagrangian containing the light quarks u and d– it prevents the masses of the light quarks and of the pions from obtaining loop corrections linearly proportional to these scales.

2.3.2 General properties of composite Higgs models

The idea of a composite Higgs follows a similar principle as chiral symmetry breaking: an additional strong sector is introduced corresponding to the symmetry group $\mathcal{G}$ which is broken at the *compositeness scale* f to a subgroup $\mathcal{H}$. In most CH models, the Higgs boson is a Goldstone boson resulting from this symmetry breaking[4]. This requires that

[3] Below this scale, the strong interaction can in general not be explained perturbatively any longer due to the increase of the coupling constant of QCD g_S. This also leads to the confinement of hadrons and to the formation of the mentioned quark condensate.

[4] There exist models (see e. g. Ref. [51, 52]) in which the Higgs boson is a generic bound state of the new strong interaction instead of being a Goldstone boson of the spontaneous symmetry breaking of the strong sector. These models are, however, not considered in the following as they tend to yield predictions that are not compatible with observed properties of the Standard Model.

the electroweak symmetry group $SU(2)_L \times U(1)_Y$ is embedded in the unbroken subgroup $\mathcal{H}$ and that there is at least one $SU(2)_L$ doublet corresponding to the Higgs doublet in the coset $\mathcal{G}/\mathcal{H}$.

The SM fields for the fermions and gauge bosons are considered as external to the new strong sector and are denoted as "elementary". They break the symmetry of the strong sector explicitly, generating a mass for the Higgs boson (which therefore becomes a pseudo-Goldstone boson) and creating the Higgs potential through quantum loops. This eventually gives rise to electroweak symmetry breaking and the generation of mass terms for the $W^{\pm}$ and Z bosons through the non-vanishing VEV of the Higgs field.

There exists a variety of possible realisations of compositeness such as "Little Higgs" [53–60] and "Holographic composite Higgs" models [61–64]. All such models can be characterised by a set of parameters Λ, a and b that determine the leading terms in the radiatively created Higgs potential [48]

$$\Delta V(h) = \frac{y_{SM}^2 \Lambda^2}{16\pi^2}\left(-a|h|^2 + b\frac{|h|^4}{2f^2}\right) \tag{2.9}$$

where $h = \sqrt{2}H$ is the Higgs field, y_{SM} is a generic SM coupling e. g. the top Yukawa coupling $y_{SM}^2 \sim y_t^2$ and Λ, a and b have to satisfy the relations

$$(246\,\mathrm{GeV})^2 = v^2 = \frac{a}{b}f^2, \qquad (125\,\mathrm{GeV})^2 = m_h^2 = 4bv^2\frac{y_{SM}^2\Lambda^2}{16\pi^2 f^2}. \tag{2.10}$$

The Λ parameter approximately corresponds to the mass of additional bound states that appear in CH models. If there would be no fine tuning at all in the final model, the compositeness scale and the VEV of the Higgs field would satisfy

$$\xi = \frac{v}{f} \approx 1, \tag{2.11}$$

setting the masses of the new bound states close to the Higgs boson mass. However, electroweak precision tests and measurements of the Higgs coupling, which are discussed in more detail in Section 2.4.2, imply $\xi \lesssim 0.1$. This introduces a moderate amount of fine tuning of the parameters in the theory regardless of the particular realisation and places new bound states at the TeV scale.

2.3.3 Partial compositeness and the generation of fermion masses

As the Higgs boson is not an elementary scalar in CH models, these models have to provide a mechanism to generate Yukawa couplings between the SM fermion fields and the Higgs field which would then become mass terms for the SM fermions. A mechanism realised in many CH models is referred to as *partial compositeness* [62, 65–68]. This introduces

mixing terms between the SM fermionic fields and fermionic fields of the strong sector, resulting in effective couplings between the two sectors of the form

$$\mathcal{L}_{int} \propto \lambda_L \bar{\psi}_L \Psi_R + \lambda_R \bar{\psi}_R \Psi_L \tag{2.12}$$

where $\Psi_{L,R}$ are the fermionic fields of the new strong sector. These new particles must have the same quantum numbers as the SM fermions they mix with and are often referred to as their "partner" particles. The partners of the SM quarks are vector-like quarks which are discussed in more detail in Section 2.4.

At low energies, the observed mass eigenstates of the SM fermions are then mixings between $\psi_{L,R}$ and $\Psi_{L,R}$ where the mixing angles depend on the coupling strengths $\lambda_{L,R}$ and the masses of the new fermionic partners. In many models, the highest mixing is expected for the right-handed top quark singlet t_R so that this is considered a *partially composite* particle while all other SM fermion mass eigenstates can be viewed as elementary in good approximation [69,70]. This mixing gives rise to the mass terms of the SM fermions and also creates effective Yukawa couplings of the SM fermions to the Higgs field.

2.3.4 Warped extra dimension models

A particular group of models in which compositeness is often realised are *warped extra dimension* models (also referred to as *Randall-Sundrum* (RS) models). These models are based on Ref. [71], detailed reviews abut this class of SM extensions can be found e. g. in Refs. [62,72,73]. They assume that there is an additional compactified fifth dimension that expands the four-dimensional Minkowski spacetime and follows an *anti-de Sitter* (AdS) metric. In this metric, the infinitesimal line element is

$$\mathrm{d}s^2 = e^{-2ky} \eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu + \mathrm{d}y^2 \tag{2.13}$$

where k is the curvature of the AdS metric, $\eta_{\mu\nu}$ is the Minkowski metric, x^μ and x^ν are the coordinates of the four-dimensional Minkowski space and y is the added fifth dimension. This means that the four spacetime dimensions are *warped* along the y-direction and their curvature depends on the y-position in the fifth dimension. There are two so-called *branes* at fixed points of the fifth dimension: an *infra-red* (IR) brane and an *ultra-violet* (UV) brane, also referred to as *Planckbrane*. A solution to the hierarchy problem is then provided by requiring that gravity has a large coupling strength on the UV brane which decreases exponentially by a factor of e^{-ky} on the bulk between the two branes. Practically, this means that the Planck scale M_P is in fact of the same order as the electroweak scale on the UV brane and the large discrepancy between the two scales appears only close to the IR brane. As only the Higgs field has to be localised near or on the IR brane to solve the hierarchy problem, other SM fields can have trajectories anywhere in the bulk. As this

could lead to new sources of fine-tuning to maintain a light Higgs boson, many RS models adopt the framework of compositeness described in Section 2.3.2. This is realised in most models by assuming a global $SO(5) \times U(1)$ symmetry in the bulk which is spontaneously broken to $SO(4) \times U(1)$ near the IR brane due to the boundary conditions of the 5D space [65].

Furthermore, according to a conjecture known as *Ads/CFT correspondence* [74], there is an operator of a *conformal field theory* (CFT) associated to each field in the bulk. These operators are identified with the fermionic operators Ψ introduced in Section 2.3.3 so that partial compositeness is naturally incorporated in RS models to generate mass terms for the SM fermions [62]. These fermionic resonances are often vector-like quarks which are discussed in the following section.

2.4 Vector-like quarks

Composite Higgs models in which the SM fermions acquire their masses through the mechanism of partial compositeness (see Section 2.3.3) predict fermionic resonances which must have the same quantum numbers as their SM counterparts. The main difference to the SM fermions is that they exhibit a *vector* symmetry under the electroweak symmetry group $SU(2)_L \times U(1)_Y$ i. e. both left- and right-handed fields transform in the same way under this symmetry group. For this reason, these new fermions are commonly referred to as *vector-like* fermions. This property is necessary as otherwise, their mass terms would have to be generated through effective Yukawa couplings to the Higgs field, equivalently to the SM fermions. This is, however strongly disfavoured both by electroweak precision measurements [75] and direct observations of the loop-mediated Hgg and $H\gamma\gamma$ couplings of the Higgs boson at the LHC [76].

The partners of the SM quarks therefore transform as triplets under the QCD gauge group $SU(3)_C$ as well and are consequently assigned to the group of quarks. These vector-like quarks are the target of the searches presented in Chapters 6 and 7 of this thesis[5]. As already mentioned in Section 2.3.2, the constraint on the ratio of the Higgs VEV and compositeness scale $\xi \lesssim 0.1$ motivates masses for these particles around the TeV scale which would make them observable at the LHC at $\sqrt{s} = 13\,\text{TeV}$. As will be explained in the following, the lightest of these new quarks are most likely partners of third-generation quarks, the T and B quarks.

Phenomenological aspects of vector-like quarks in general, and of T and B quarks in particular, are discussed in Section 2.4.1, including production mechanisms and potential signatures. Current constraints on phenomenological parameters of vector-like quarks, most notably their masses, from both direct and indirect measurements are described in

[5] Vector-like leptons can appear as well in many models (see e. g. Refs. [77, 78]) but are not the focus of the search presented in this thesis and therefore not considered in the following.

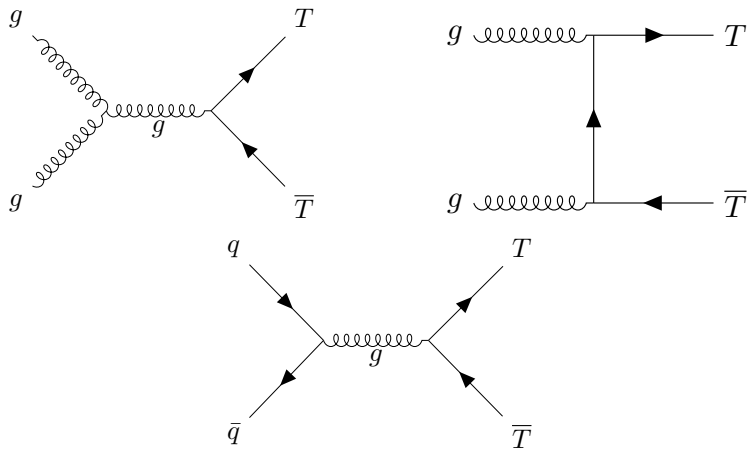

Figure 2.3: Feynman diagrams depicting the possible pair production modes of vector-like T quarks. They were created using the tools presented in Ref. [88].

Section 2.4.2. A detailed overview over vector-like quarks and their possible signatures at the LHC is provided by Refs. [77, 79–81] upon which the following sections are based on.

2.4.1 Phenomenological properties of vector-like quarks

Since vector-like quarks have the same colour charge as their SM counterpart, they are subject to the strong interaction with the same coupling strength as SM quarks. This means that at hadron colliders like the LHC, they can be pair-produced via several strong processes as is depicted in Fig. 2.3. The corresponding production cross section then only depends on the VLQ mass which can be calculated analogous to SM $q\bar{q}$ production. A summary of the predicted cross sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$, calculated at next-to-next-to-leading order (NNLO) with the Top++2.0 program [82–87], can be found in Table 2.1 for various VLQ masses.

Following the idea of partial compositeness, vector-like quarks mix with their SM partners with the highest mixings expected for the partners of third generation quarks. This is motivated by the large Yukawa couplings of third generation SM quarks and suggests that their vector-like partners are the lightest new fermionic resonances with masses of around the TeV scale [69, 70, 80]. As a consequence, most searches at the LHC, including the searches presented in this thesis in Chapters 6 and 7, target such third-generation quark partners and they are the main focus in the remaining section. These include the T and B quarks, with electric charges of $+2e/3$ and $-1e/3$, and the more exotic $X_{5/3}$ and $Y_{-4/3}$ quarks with electric charges of $+5e/3$ and $-4e/3$, respectively.

The mixing terms with SM quarks can be included by expanding the CKM matrix (see Section 2.1.1) which incidentally could also introduce new sources of CP violation [89–92] and thus potentially contribute to a solution of the Baryogenesis problem in the Standard

Signal Mass [GeV]	σ [fb]
$T\overline{T}$ (M(T) = 700 GeV)	$455 \pm ^{19}_{19}$
$T\overline{T}$ (M(T) = 800 GeV)	$196 \pm ^{8.5}_{8.1}$
$T\overline{T}$ (M(T) = 900 GeV)	$90.3 \pm ^{4.0}_{3.8}$
$T\overline{T}$ (M(T) = 1000 GeV)	$44.0 \pm ^{2.1}_{1.9}$
$T\overline{T}$ (M(T) = 1100 GeV)	$22.4 \pm ^{1.1}_{1.0}$
$T\overline{T}$ (M(T) = 1200 GeV)	$11.8 \pm ^{0.64}_{0.56}$
$T\overline{T}$ (M(T) = 1300 GeV)	$6.39 \pm ^{0.37}_{0.32}$
$T\overline{T}$ (M(T) = 1400 GeV)	$3.54 \pm ^{0.22}_{0.19}$
$T\overline{T}$ (M(T) = 1500 GeV)	$2.00 \pm ^{0.14}_{0.12}$
$T\overline{T}$ (M(T) = 1600 GeV)	$1.148 \pm ^{0.087}_{0.072}$
$T\overline{T}$ (M(T) = 1700 GeV)	$0.666 \pm ^{0.056}_{0.045}$
$T\overline{T}$ (M(T) = 1800 GeV)	$0.391 \pm ^{0.037}_{0.029}$

Table 2.1: Predicted cross sections of pair production of vector-like quarks at various mass points calculated at NNLO. T quarks are used as an example in this table but the cross sections for other vector-like quarks (such as B quarks) with the same mass are identical. Uncertainties include contributions from energy scale variations and the uncertainty in the parton distribution function.

Model (see Section 2.2). The mass eigenstates of the SM top quark and its vector-like partner, the T quark, can then be written as

$$\begin{pmatrix} t_{L,R} \\ T_{L,R} \end{pmatrix} = \begin{pmatrix} \cos\theta^u_{L,R} & -\sin\theta^u_{L,R} \\ \sin\theta^u_{L,R} & \cos\theta^u_{L,R} \end{pmatrix} \begin{pmatrix} t'_{L,R} \\ T'_{L,R} \end{pmatrix} \tag{2.14}$$

where the fields t' and T' denote the weak eigenstates and $\theta^u_{L,R}$ are the mixing angles between the SM and vector-like quarks. The mixing terms for partners of the SM b quarks, the B quarks, take a similar form with different mixing angles $\theta^d_{L,R}$.

These mixing angles can depend on the specific model in which vector-like quarks appear and on their transformation properties under the electroweak symmetry group $SU(2)_L \times U(1)_Y$. They are also constrained experimentally by electroweak precision measurements (see Section 2.4.2.2). If there are only $SU(2)_L$ doublets in the scalar sector of the Lagrangian, which is the case in the SM Lagrangian for energies below the compositeness scale, vector-like quarks can appear in one of seven possible electroweak isospin multiplets which are summarised in Table 2.2 [93].

If vector-like quarks appear in doublets or triplets, their masses would be nearly degenerate with a small splitting due to the mixing with third generation SM quarks [79].

Singlets	Doublets	Triplets
$T_{L,R}$, $B_{L,R}$	$\begin{pmatrix} X_{5/3} \\ T \end{pmatrix}_{L,R}$, $\begin{pmatrix} T \\ B \end{pmatrix}_{L,R}$, $\begin{pmatrix} B \\ Y_{-4/3} \end{pmatrix}_{L,R}$	$\begin{pmatrix} X_{5/3} \\ T \\ B \end{pmatrix}_{L,R}$, $\begin{pmatrix} T \\ B \\ Y_{-4/3} \end{pmatrix}_{L,R}$

Table 2.2: Isospin multiplet representations of the electroweak symmetry group $SU(2)_L \times U(1)_Y$ in which vector-like quarks can appear. T and B denote vector-like partners of the SM top and bottom quarks with electric charges of $+2e/3$ and $-1e/3$, respectively, while $X_{5/3}$ and $Y_{-4/3}$ are exotic new fields without a SM correspondence and electric charges of $+5e/3$ and $-4e/3$, respectively. The subscripts L, R indicate that they transform in the same way under left- and right-handed transformations of $SU(2)_L \times U(1)_Y$, thus the label "vector-like".

From this rather small splitting it follows that the decays from one vector-like quark into another are strongly suppressed. The terms in the Lagrangian describing the couplings of vector-like quarks to SM particles then take the following form [79, 93, 94]:

$$\begin{aligned}
\mathcal{L}_W =& -\frac{g_W}{\sqrt{2}}\bar{T}\gamma^\mu(V_{Tb}^L P_L + V_{Tb}^R P_R)bW_\mu^+ + \text{h.c.} \\
& -\frac{g_W}{\sqrt{2}}\bar{b}\gamma^\mu(V_{bT}^L P_L + V_{bT}^R P_R)TW_\mu^+ + \text{h.c.} \\
\mathcal{L}_Z =& -\frac{g_W}{2\cos\theta_W}\bar{t}\gamma^\mu(\pm X_{tT}^L P_L \pm X_{tT}^R P_R)TZ_\mu + \text{h.c.} \\
\mathcal{L}_H =& -\frac{g_W m_T}{2M_W}\bar{t}(Y_{tT}^L P_L + Y_{tT}^R P_R)TH + \text{h.c.}
\end{aligned} \tag{2.15}$$

In these terms, b, t and T denote the two third generation SM quarks and vector-like T quarks, respectively, $P_{L,R}$ are the left- and right-handed projection operators $\frac{1\mp\gamma_5}{2}$, and $V_{Tb,bT}^{L,R}$, $X_{tT}^{L,R}$ and $Y_{tT}^{L,R}$ are coupling coefficients that depend on the mixing angles between the SM quarks and the vector-like quarks. The resulting couplings to left- and right-handed vector-like quarks determine the cross section for the corresponding single production mode (see Fig. 2.4) as well as the partial decay widths to their possible decay modes. As a consequence, the allowed decay modes for the considered third generation quark partners are

- $X_{5/3} \to tW^+$ and $Y_{-4/3} \to bW^-$ for the exotic quark partners (due to electric charge conservation),
- $T \to bW^+$, $T \to tZ$ and $T \to tH$ for the vector-like T quark,
- $B \to tW^-$, $B \to bZ$ and $B \to bH$ for the vector-like B quark.

Model-independent calculations [77, 79, 81, 95] suggest the following, approximate branch-

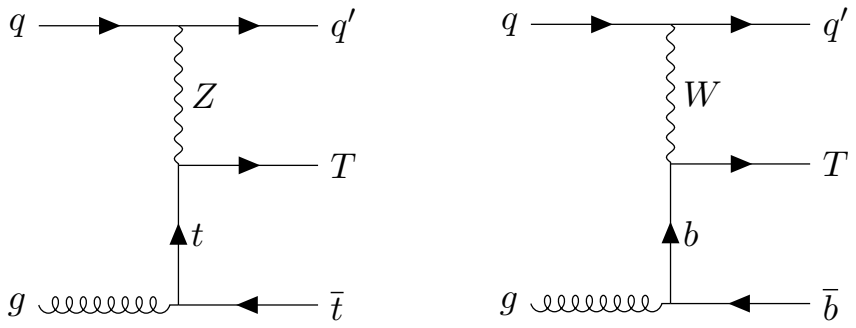

Figure 2.4: Feynman diagrams depicting possible single production processes of a vector-like T quark via the neutral (left) and charged (right) weak interaction. Feynman diagrams created using tools presented in Ref. [88].

ing fractions, depending on the multiplet representations of the T and B quarks under the electroweak symmetry group:

Electroweak isospin singlet T, B: 50% for the $T \to bW$ and $B \to tW$ decay, 25% for the $T \to tZ/tH$ and $B \to bH/bZ$ decays

Electroweak isospin doublet (T,B), ($X_{5/3}$,T): 50% for both the $T \to tZ/tH$ decays, 100% for the $B \to tW$ decays (in case of the (T,B) doublet)

Electroweak isospin doublet (B,$Y_{-4/3}$): 50% for both the $B \to bZ/bH$ decays

For electroweak isospin triplets, vector-like quarks have very similar branching fractions as in the isospin singlet or doublet cases. These are shown, along with an overview over all assumed branching fraction combinations, in Table 2.3. They are also shown as a function of the VLQ mass in Fig. 2.5 where it can be seen that this dependency is becoming negligible for high VLQ masses.

Models incorporating the simplest CH realisation, also referred to as *minimal CH models* [96,97] suggest similar branching fraction combinations. In these models, the strong sector corresponds to a symmetry group $SO(5) \times U(1)$ which is broken down to $SO(4) \times U(1)$. Thus, the branching fraction combinations for the electroweak isospin singlet and doublet scenarios are used as benchmark points in the interpretations of the search results discussed in Chapters 6 and 7.

These branching fractions were calculated assuming that the coupling coefficients in the Z and H interaction terms in Eq. (2.15), $X_{qQ}^{L,R}$ and $Y_{qQ}^{L,R}$, are of equal size, thus producing symmetric branching fractions to the two decay modes. As this, however, does not necessarily have to be the case (e. g. in models that do not incorporate partial compositeness to generate mass terms for SM fermions), the results obtained in this thesis are also interpreted for the entire range of possible branching fractions of T and B quarks.

E.-w. isospin multiplet/branching fractions	$T \to bW/B \to tW$	$T \to tZ/B \to bZ$	$T \to tH/B \to bH$
Singlet T/B	50%	25%	25%
Doublet (T,B)	0%/100%	50%/0%	50%/0%
Doublet $(X_{5/3},T)$	0%/–	50%/–	50%/–
Doublet $(B,Y_{-4/3})$	–/0%	–/50%	–/50%
Triplet $(X_{5/3},T,\ B)$	50%/0%	25%/50%	25%/50%
Triplet $(T,B,Y_{-4/3})$	0%/50%	50%/25%	50%/25%

Table 2.3: Approximate branching fractions for vector-like T and B quarks assuming different multiplet structures. Branching fractions for $X_{5/3}$ and $Y_{-4/3}$ quarks (if present) are always 100% to tW and bW, respectively.

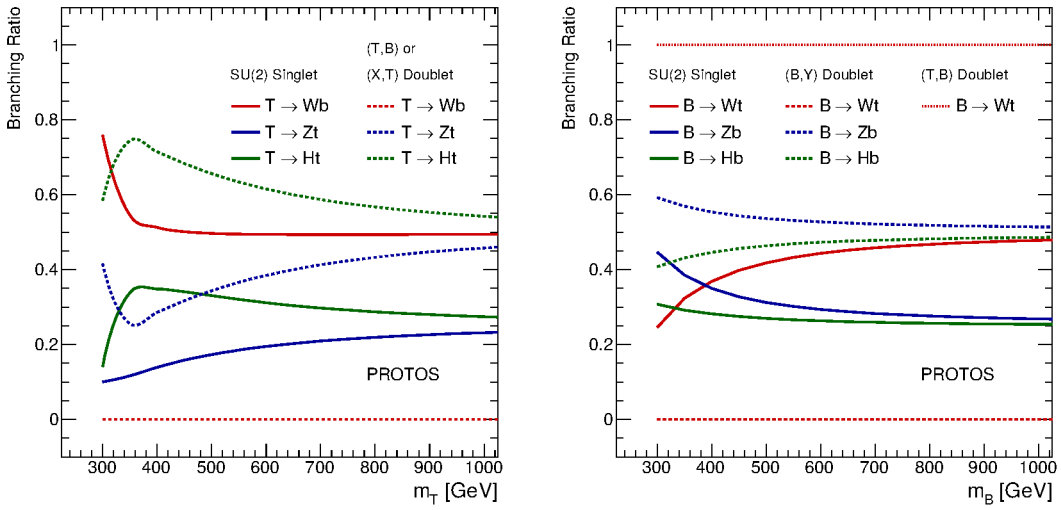

Figure 2.5: Decay modes for vector-like T quarks (left) and B quarks (right) as a function of the VLQ mass for various electroweak isospin multiplet scenarios. Taken from Ref. [98].

2.4.2 Constraints on vector-like quark production

Experimental constraints on parameters determining the VLQ phenomenology can be grouped into two categories, discussed in this section. As no evidence for vector-like quarks has been observed yet in direct searches, upper limits on their production cross section at 95% confidence level (C.L.) and corresponding lower mass bounds were derived which are presented in Section 2.4.2.1. Furthermore, indirect constraints on various parameters in VLQ phenomenology were found also in electroweak precision measurements which are discussed in Section 2.4.2.2.

2.4.2.1 Constraints from direct searches at 8 TeV

Upper bounds on the production cross section of vector-like quarks are obtained mainly from direct searches performed at the LHC with data collected at centre-of-mass energies of $\sqrt{s} = 8\,\text{TeV}$ and 13 TeV. At 8 TeV, upper cross section limits on $T\overline{T}$ and $B\bar{B}$ production at 95% C.L. were derived by both the ATLAS and the CMS Collaborations. Searches with the CMS detector found lower mass bounds ranging from 720 to 920 GeV for vector-like T quarks and from 740 to 900 GeV for vector-like B quarks, depending on the assumed branching fractions of the vector-like quarks [99,100]. The ATLAS Collaboration observed similar bounds for VLQ pair production at 8 TeV, with lower mass limits ranging from 715 to 950 GeV for T quarks and from 575 GeV to 813 GeV for B quarks [101,102].

Single production of vector-like quarks has been studied by the ATLAS collaboration at $\sqrt{s} = 8\,\text{TeV}$ and upper limits at 95% C.L. have been set on the corresponding production cross section [103–106]. Since the cross section for this production mode depends strongly on the couplings of the hypothetical vector-like quark to the SM quarks, the interpretation of these limits always depends on assumptions made for the strengths of these couplings. Depending on these assumptions and the resulting branching fractions of the vector-like quarks, lower limits on VLQ masses of the order $\sim$1 TeV were found. This does, however, not mean that vector-like quarks with masses below these values are excluded in general but only for the respective coupling strength assumptions.

2.4.2.2 Constraints from indirect measurements

Constraints on VLQ parameters can also be derived from measuring *electroweak precision variables* in dedicated collider experiments. A good overview over these precision tests can be found in Ref. [107].

Important observables are the *oblique parameters* $\hat{S}$ and $\hat{T}$ which are form factors encapsulating possible corrections to the self-energies of the SM vector bosons [108,109] and can be measured at e^+e^- colliders like LEP. In the Standard Model, these parameters take the values $\hat{S} = \hat{T} = 0$ so that any measured deviations from zero would imply new

physics beyond the Standard Model. Corrections to these parameters in CH models can in particular arise from loop corrections induced by vector-like quarks [70, 110, 111]. Bands on the $\hat{S}$ and $\hat{T}$ fitted from electroweak precision data [112] imply a rather large difference between the compositeness scale and the Higgs VEV $\xi \lesssim 0.1$ [107] which motivates the masses of new resonances such as vector-like quarks to be at the TeV scale or above (see Section 2.3.2).

Another important observable is the $Zb_L\bar{b}_L$ coupling, which can be affected by a mixing of the b with vector-like third-generation quark partners [113, 114]. Deviations of this coupling from the expected SM value can be constrained at the level 10^{-3} [115, 116] and can be interpreted as constraints on model parameters in various CH models. In the case of minimal CH models, for instance, the measured values seem to disfavour models with only one light VLQ electroweak isospin singlet [107].

Combining both the measurements of the oblique parameters and of the $Zb_L\bar{b}_L$ coupling, limits can also be set on the mixing angles $\sin\theta^{u,d}_{L,R}$ from Eq. (2.14) as a function of the VLQ mass. These limits range from 0.04 and 0.18 for vector-like quarks with masses that have not yet been excluded by direct searches [79]. These bounds can play an important role e. g. in the interpretation of results of searches for single production of vector-like quarks, where the production cross section strongly depends on these mixing angles.

As vector-like quarks would also contribute to loop-mediated cross sections and decay widths of the Higgs boson [117], which can be measured at the LHC, constraints on the VLQ parameter space can also be derived from corresponding measurements. Although such measurements at the LHC are not precise enough to directly discover vector-like quarks [79], they set bounds on the magnitude of the compositeness scale of the order $f \gtrsim 700\,\text{GeV}$ for minimal CH models [118].

3 Experimental setup

One of the main experimental approaches to test the Standard Model, measure its parameters, and find hints at possible extensions is to study particle collisions at dedicated accelerators. Many different types of accelerators exist that differ in their designs and the types of particles they are colliding. In *linear accelerators*, the particles are accelerated on a straight line by a static or dynamic electromagnetic field and brought to collision either with a fixed target or with another accelerated particle. One of the most prominent colliders of this type is the *Stanford Linear Collider* which has been in operation for several decades now.

A different type are accelerators in which the particles are held on a circular trajectory by magnetic fields. This allows for higher collision energies as the acceleration path of the particles can be arbitrarily extended. One of the earliest designs, is the so-called *cyclotron* in which the magnetic field is static so that the radius of the accelerated particles becomes larger with increasing velocity. This concept has been expanded to *synchrotrons* in which the radius of the particles is kept constant, requiring a dynamic magnetic field adapting to the particle's increasing momentum. So-called *storage rings* are constructed in a similar way but are designed to keep the particles at a constant kinetic energy by accelerating them only to bring them to the desired energy and to compensate for eventual energy losses. This facilitates long operating times and a collection of a large amount of data.

The type of particle being accelerated varies from collider to collider and depends on the physical processes one would like to investigate. Electrons and positrons are collided in the Stanford Linear Collider and were also used in the *Large Electron-Positron Collider* (LEP) which operated at the CERN[1] main site on the French-Swiss border close to Geneva. These colliders allow for very precise measurements of SM processes but are limited to relatively low energies as electrons at high energies emit a significant amount of energy in form of synchrotron radiation. Hadron colliders can overcome these limitations and in addition facilitate direct measurements of the strong interaction. This, however, comes at the price of a limited knowledge about the initial state of the colliding particles due to the inner structure of the hadrons. Proton-antiproton collisions were performed at the *Tevatron* collider at Fermilab before it ceased operations in 2011. The *Large Hadron Collider* (LHC) [119] at CERN is the most powerful particle accelerator which performs either proton-proton (pp), lead-lead ($PbPb$) or lead-proton collisions at the highest centre-of-mass energies to date.

[1] European Organization for Nuclear Research

The analysis presented in this thesis uses data from pp collisions produced at the LHC at $\sqrt{s} = 13\,\text{TeV}$ which were recorded by the CMS experiment [120] during the data-taking period in 2015. The aim of this chapter is to provide an overview over the experimental setup, describing the LHC in Section 3.1 and the *CMS detector* in Section 3.2.

3.1 The Large Hadron Collider

As mentioned in the introduction of this chapter, the LHC is a storage ring with a circumference of 27 km which performs proton-proton, lead-lead, and (for brief periods of time) proton-lead collisions. It is described in detail in Ref. [119] and this section briefly summarises its technical details. As the analysis presented in this thesis studies pp collisions, other types of particle interactions are neglected in the following and unless specified otherwise, "collision" always implicitly denotes a pp collision.

The LHC started in 2010 at a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ for pp collisions and continued operations at $\sqrt{s} = 8\,\text{TeV}$ from 2011 on before stopping for a series of major upgrades at the end of 2012. The entire period between 2010 and 2012 is referred to as *LHC Run I* or just *Run I*. It restarted in 2015 and is now operating at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. This current operation cycle is also referred to as *(LHC) Run II*. Heavy ion collisions take place at dedicated runs usually at the end of each year and are currently performed at a centre-of-mass energy of $\sqrt{s} = 5.02\,\text{TeV}$. The protons (or ions) used for collisions are first accelerated by the linear accelerator LINAC 2 to a kinetic energy of 50 MeV before being injected into the *Proton Synchrotron Booster* (PSB) where they are brought to an energy of 1.4 GeV. They further pass the *Proton Synchrotron* (PS) and *Super Proton Synchrotron* (SPS), where they reach kinetic energies of 26 and 450 GeV, respectively, before finally being injected into the LHC. There are four large experiments at the LHC located at collision points of the particle beams:

- The **ALICE** experiment (*A Large Ion Collider Experiment*) [121] which focuses on the analysis of $PbPb$ collisions to study quark-gluon plasmas that are expected to be produced at such large energies.

- The **LHCb** experiment (*Large Hadron Collider beauty*) [122] which is a forward detector to study the properties of b quarks and in particular explores CP violating parameters in the Standard Model, also to look for possible new sources and physics beyond the Standard Model.

- Two multi-purpose detectors, **ATLAS** (*A Toroidal LHC Apparatus*) [123] and **CMS** (*Compact Muon Solenoid*) which are designed for a large variety of measurements and achieved a major success by providing evidence for the Higgs boson in 2012. A more detailed description of the CMS experiment can be found in Section 3.2.

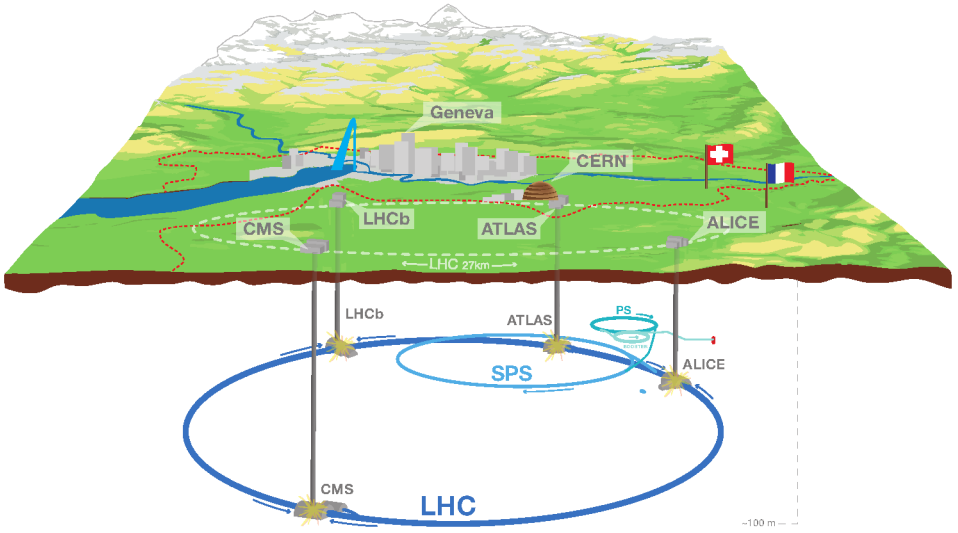

Figure 3.1: The LHC with the four experiments located at it and its pre-accelerator complex. Taken from Ref. [124].

A graphical illustration of the LHC facility with sketches representing the pre-accelerator complex and the four experiments located at it can be found in Fig. 3.1.

Protons are grouped into *bunches* containing approximately 1.15×10^{11} particles and currently up to 2,808 bunches traverse the LHC at the same time. A collision of two proton bunches is also referred to as *bunch crossing*. As the protons are travelling at approximately the speed of light, this results in a time interval or *bunch spacing* of 25 ns and a collision rate of about 40 MHz. The protons travel inside two parallel *beam pipes* in a vacuum with a pressure of $\sim 10^{-13}$ atm in which they are held by about 9600 magnets. Dipole magnets hold the protons on their circular orbit while quadrupole magnets (and magnets of higher multipole order) keep the beams focused, in particular close to the collision points, to increase the probability of bunch interactions. To maintain superconductivity of these magnets, they are cooled down to an operating temperature of 1.9 K ($-271.3°$ C) using superfluid helium-4. The particles are accelerated in eight *radiofrequency cavities* per beam operated at a similar temperature, where the accelerating electromagnetic fields oscillate with a frequency of 400 MHz. This oscillation is necessary to keep the particles at a constant momentum and to avoid a dispersing of the bunches in longitudinal direction. One of the key parameters of the LHC operation is the *instantaneous luminosity* it delivers

to the four main experiments. This is defined as

$$L = \frac{1}{\sigma}\frac{\mathrm{d}N}{\mathrm{d}t} \tag{3.1}$$

where N is the number of interactions of a given process and σ is the corresponding cross section so that the luminosity L provides a measure for the rate of physical processes occurring in the particle collisions. For storage rings like the LHC, the instantaneous luminosity is determined as follows:

$$L = N_b f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{3.2}$$

N_b is the number of bunches circulating in the accelerator, f is the orbit frequency of each bunch, n_1 and n_2 are the number of particles in two colliding bunches, and σ_x and σ_y are the transverse widths of the beams in the plane perpendicular to the beam axis, assuming that the beam profiles follow Gaussian distributions. The quantities $\sigma_{x,y}$ have to be derived in dedicated measurements, so-called *Van der Meer* (VdM) scans [125]. In these scans, the rate of a given observable sensitive to the luminosity is measured as a function of the distance of the beams whilst varying this distance. In CMS, the *Pixel Cluster Counting* method (PCC) [126] is used in which the mean number of pixel clusters per collision detecting a traversing particle is counted. Having determined the values for $\sigma_{x,y}$, the same observable can then be used during operation of the LHC to provide prompt measurements of the instantaneous luminosity. The LHC has been designed to deliver an instantaneous luminosity of $L = 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The peak luminosity for pp collisions during the 2015 data-taking period was measured to be roughly $5.13 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [127] while in the 2016 data taking period, the design luminosity has in fact been surpassed reaching a maximum value of $L = 1.53 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [128].

An important quantity to indicate the total size of a collected data set is the *integrated luminosity* $\mathcal{L}$ over a given time period, defined as

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

Based on its value and using Eq. (3.1) one can then calculate the total number of events of a certain process in a collected data set. The cumulative integrated luminosity over the data-taking period in 2015 is shown in Fig. 3.2. The total integrated luminosity delivered by the LHC corresponds to $\mathcal{L} = 4.22\,\mathrm{fb}^{-1}$ with the one recorded by the CMS experiment being $\mathcal{L} = 3.81\,\mathrm{fb}^{-1}$. The actual data set used for physics analyses in CMS is most of the times smaller since usually only recorded data are used in which the main magnet was fully functional and all subsystems of the detector showed good performance.

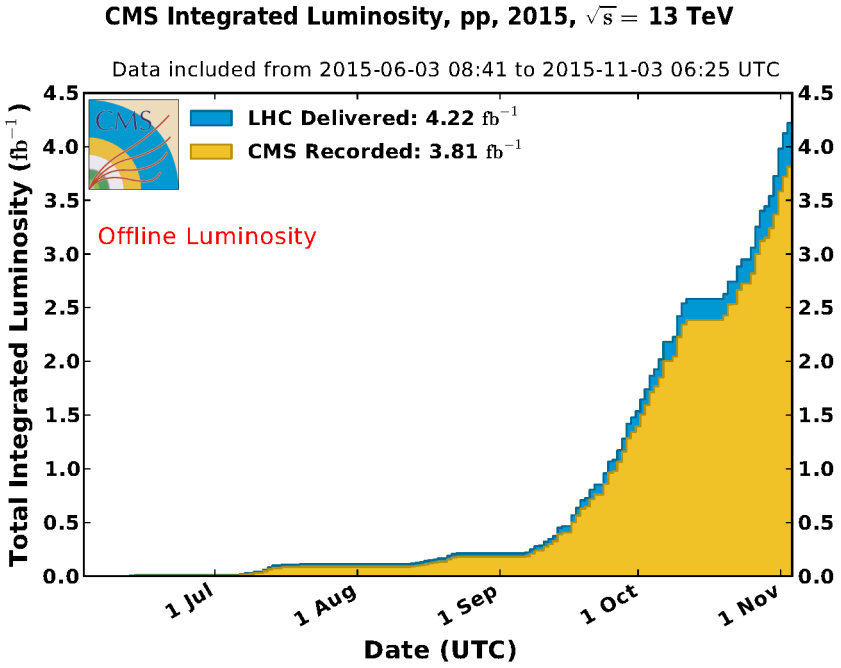

Figure 3.2: Cumulative integrated luminosity over the data-taking period in 2015 as delivered by the LHC (blue) and recorded by the CMS experiment. Taken from Ref. [127].

3.2 The CMS detector

The CMS detector is a multi-purpose experiment designed for a broad range of measurements. These include precision measurements of SM parameters, for example the mass of the top quark and the couplings of the Higgs boson, as well as searches for new particles that would give hints at physics beyond the Standard Model.

A bunch crossing in the CMS detector produces a large quantity of outgoing particles, the entirety of which is denoted as *event*. Physical processes of interest, i. e. with a large momentum transfer or involving rare interaction processes, are referred to as *hard scattering process* and usually occur with a considerably low rate compared to soft elastic or inelastic scattering processes. Additional interactions of other protons within the same bunch crossing are referred to as *in-time pileup interactions* (IT pileup). Contributions in the detector from interactions in previous or subsequent bunch crossings, which occur due to the finite signal decay time in the measuring devices, are referred to as *out-of-time pileup* (OOT pileup).

To make inferences about the initial physical process, it is necessary to measure as many of these particles as possible with a high accuracy. To ensure this, the CMS detector is composed of four different components which are arranged in symmetric layers surrounding the beam line (also called *beam axis*). These form several *subdetectors* that are responsible for different steps in the reconstruction of individual events. They are, from innermost to

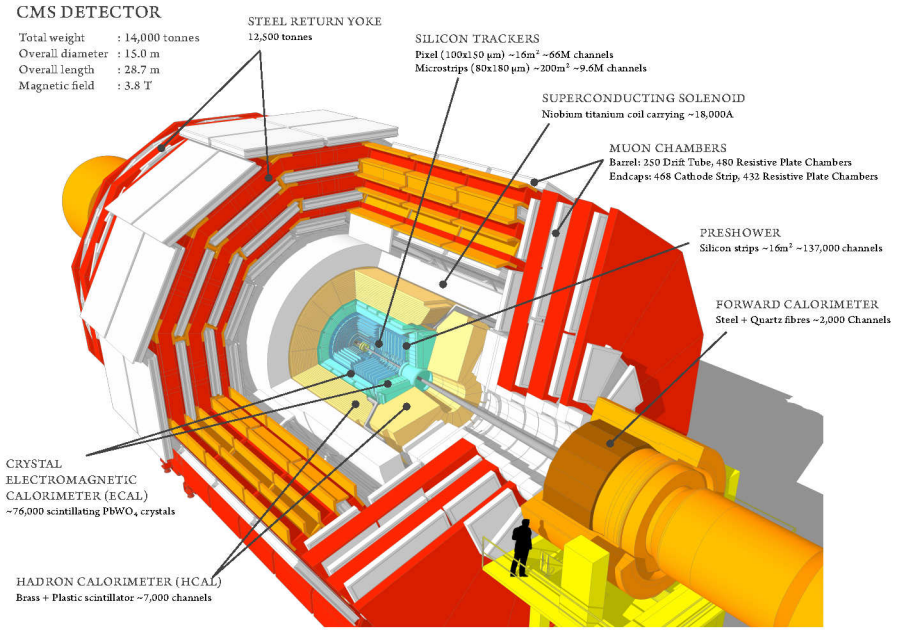

Figure 3.3: A graphical sketch of the CMS detector with all its subsystems. Taken from Ref. [129].

outermost

- the **inner tracking system** which is used to reconstruct the trajectories of charged particles,
- the **electromagnetic calorimeter** (ECAL) which provides precise measurements of the energy of electrons and photons,
- the **hadronic calorimeter** which measures the energy of hadrons,
- the **muon detectors** which detect muons that usually leave the detector before decaying and support the reconstruction of their tracks.

A graphical illustration of the CMS detector is depicted in Fig. 3.3. More detailed descriptions of these subdetectors and their performance can be found in the following sections.

The central feature of the CMS detector is a large, superconducting solenoid magnet which is placed between the hadronic calorimeter and the muon detectors, providing a magnetic field with a field strength of 3.8 T to bend the trajectories of charged particles and measure their momentum . Three layers of iron are located in between the muon chambers forming a return yoke to confine the magnetic field.

This section provides a summary over the functionality of the CMS experiment, a full description can be found in Ref. [120]. It is structured as follows: In Section 3.2.1, the coordinate system used to describe the geometry of the detector and the flight direction of

particles is introduced. Section 3.2.2 contains a description of the tracker while the electromagnetic and hadronic calorimeters are described in Sections 3.2.3 and 3.2.4, respectively. A description of the muon detectors is given in Section 3.2.5 and the Section 3.2.6 provides a summary of the trigger system used in the CMS experiment.

3.2.1 Coordinate system

The nominal interaction point, at which the proton bunches are expected to collide, defines the centre of the CMS detector and the origin of the coordinate system. The x-axis points towards the centre of the LHC ring while the y-axis points vertically upwards towards the surface. The z-axis is parallel to the beam axis in anticlockwise direction, completing a right-handed coordinate system. The (x,y)-plane is also referred to as *transverse plane* and the corresponding component of the momentum of a particle in that plane is defined as the *transverse momentum* p_{T}.

Usually, a cylindrical coordinate system is used in which r defines the distance from the beam axis, ϕ is the azimuthal angle in the (x,y)-plane defined as $\tan\phi = \frac{y}{x}$ and θ is the polar angle measured from the positive z-axis. Normally, instead of θ, the *pseudorapidity* η is used which is defined as $-\ln\tan(\frac{\theta}{2})$ as this is a Lorentz-invariant quantity for particles boosted in z-direction and massless particles. Based on η, each subdetector system can be divided into a *barrel* region spanning roughly the range $|\eta| \lesssim 1$ and an *endcap* region extending to higher values. The angular separation between two particles is usually measured in terms of ΔR, which is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.4}$$

and is a Lorentz-invariant variable, too.

3.2.2 Inner tracking system

The *inner tracking system* or short *tracker* is responsible for the measurement of tracks of charged particles from the pp collision, including electrons, muons, and charged hadrons. It has a total diameter of 2.5 m and is 5.8 m long. As the magnetic field strength is known to a good precision, the curvature of the tracks can be used to determine the transverse momentum of a particle. Another important task of the tracker is to determine the exact positions of a scattering processes within a bunch crossing, the so-called *interaction vertices* among which the position of the hard scattering process is referred to as *primary vertex*. Additional vertices due to decays of intermediate particles such as B-mesons can be reconstructed as well and are called *secondary vertices*. More details about the exact procedure to reconstruct vertices and tracks from the output of the tracker can be found in Section 5.1.

The tracking system is based on a large number of modules that contain small units of silicon semiconductors. Charged particles passing through the material produce electron-hole pairs in these units through ionisation so that they leave a trace of hits which is used to reconstruct the corresponding particle track.

The entire tracker is subdivided into two parts: closest to the beam pipe and the nominal interaction point is the *pixel detector* which is enclosed by the *strip detector*. These two components differ by the size of the smallest silicon unit used to detect particle passages to satisfy the varying requirements to the detector at different distances from the beam axis. The acceptance region of the entire tracking system reaches up to a pseudorapidity of $|\eta| < 2.5$.

The pixel detector is composed of in total 66 million silicon pixels distributed over 1,440 modules, each with a size of $100 \times 150\,\mu\text{m}^2$. This relatively small size ensures a low occupancy between 0.002–0.02% [130] and a high granularity which is necessary as there is an immense track density close to the nominal interaction point (with about 1,000 charged particles per bunch crossing). In the barrel region, these pixel modules are placed on three layers located at radial distances of 4.4 cm, 7.3 cm and 10.2 cm to the beam axis. In the endcap region, two disks at each side of the detector are placed at $|z| = 34.5\,\text{cm}$ and $|z| = 46.5\,\text{cm}$, spanning from $r = 6\,\text{cm}$ to $r = 16\,\text{cm}$. The complete pixel detector has been replaced during the end-of-the-year stop 2016/2017 and is foreseen to be commissioned for the 2017 data-taking period [131]. As upgrade, a fourth pixel layer has been added to further increase the accuracy of the track and pixel reconstruction.

The strip detector consists of a total of 9.6 million strips with varying sizes from $10\,\text{cm} \times 80\,\mu\text{m}$ closer to the beam axis to $25\,\text{cm} \times 180\,\mu\text{m}$ at the outermost point of the tracker. This is possible due to the lower track density further away from the nominal interaction point and helps reducing the costs of the detector. The strip detector ranges from $20\,\text{cm} < r < 116\,\text{cm}$ and has four layers in the so-called *tracker inner barrel* (TIB) and six layers in the *tracker outer barrel* (TOB). It is completed by several disks in the endcap region, three in the so-called *tracker inner disk* (TID) and nine in the *tracker endcap* (TEC) that are located between $60\,\text{cm} < |z| < 280\,\text{cm}$. A schematic view of the inner tracking system with all its subcomponents can be found in Fig. 3.4.

Overall, the CMS tracking system performs exceptionally well, successfully reconstructing charged particle tracks with a $p_\text{T} > 0.9\,\text{GeV}$ with an efficiency of 94% in the barrel region and 85% in the endcap region [130]. In particular, the enormous granularity in the pixel detector combined with a method called *charge sharing*, which exploits the energy spread across multiple pixels, provides a so-called *impact parameter resolution* of $10\,\mu\text{m}$ and $30\,\mu\text{m}$ in the transverse and longitudinal direction, respectively. The *impact parameter* (IP) is defined as the distance between a track and a reference point (e. g. the primary interaction vertex) at the point of closest approach. The total IP measures the distance

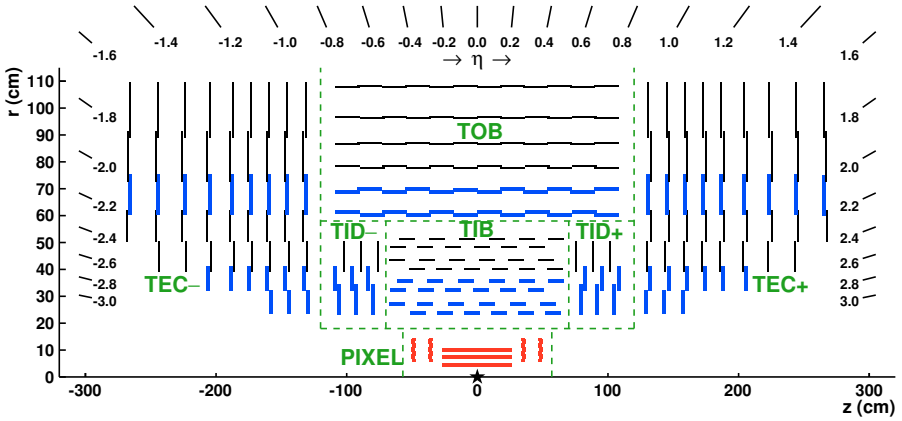

Figure 3.4: A schematic view of the CMS inner tracking system. Taken from Ref. [130].

in three dimensions, while the transverse and longitudinal IPs refer to the distances in the (x,y)-plane and in the z-direction, respectively.

3.2.3 Electromagnetic calorimeter

The ECAL is enclosed around the inner tracker, spanning from $r = 129\,\text{cm}$ to $r = 177\,\text{cm}$ with the main task to detect electrons and photons and determine their energies. This is of particular importance to measure physical processes that involve these particles (such as $H \to \gamma\gamma$) but also to identify neutral pions as they almost exclusively decay to two photons.

It has a homogeneous design and consists of more than 68,000 lead tungstate crystals ($PbWO_4$) that act both as stopping material (through their high density of $8.28\,\text{g}\,\text{cm}^{-3}$) and scintillators at the same time. The short radiation length of 0.89 cm and resulting small Molière radius of 2.2 cm allow for a high granularity and compact design of the ECAL. The length of a crystal is 230 mm which corresponds to 25.8 radiation lengths so that the bulk of a photon's or electron's energy would be deposited inside a crystal.

The ECAL is separated into two blocks: the ECAL barrel (EB) which is made out of 61,200 crystals and uses silicon avalanche photodiodes to read out the scintillating light, and the ECAL endcaps (EE), consisting of 7324 crystals with vacuum phototriodes as scintillation detectors. The EB spans the area $|\eta| < 1.479$ and the EE from $1.653 < |\eta| < 3.0$, meaning that there is a small gap not covered by the ECAL. In the EE, there is an additional preshower detector (PS) ranging from $1.566 < |\eta| < 2.6$ with the main purpose to identify $\pi_0 \to \gamma\gamma$ decays. The structure of the CMS ECAL is schematically illustrated in Fig. 3.5.

The energy resolution of the ECAL has been measured in test beams and can be

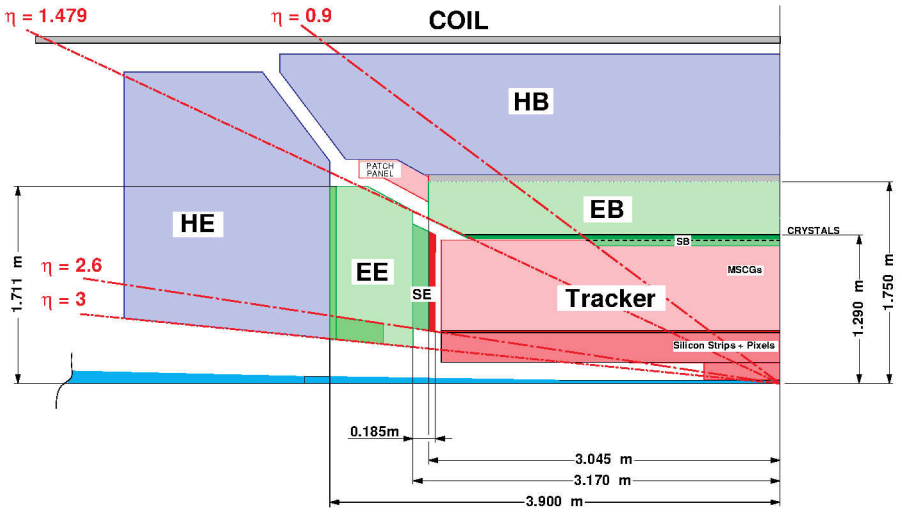

Figure 3.5: A schematic view of the CMS ECAL between the inner tracking system and the HCAL. Taken from Ref. [134].

parametrised by the following formula with the measured values inserted [132]

$$\frac{\sigma}{E} = \frac{12\%}{E(\text{GeV})} + \frac{2.8\%}{\sqrt{E(\text{GeV})}} + 0.3\%. \tag{3.5}$$

For electrons from Z boson decays, the energy resolution was found to be 2% in the central region with $|\eta| < 0.8$ and 2–5% elsewhere, while for simulated photons in $H \to \gamma\gamma$, the energy resolutions were found to be 1.1–2.6% and 2.2–5% in the barrel and endcap regions, respectively [133].

3.2.4 Hadronic calorimeter

The HCAL is a sampling calorimeter composed of large segments of stopping material (brass and steel) interspersed with tiles of plastic scintillators which are read out by photo diodes connected to the scintillators via optical fibres. Its main purpose is to measure the energy of hadrons and absorb as many outgoing particles as possible so that, ideally, muons and neutrinos are the only particles that pass the HCAL. As the latter ones do not produce any measurable signal in the detector and can therefore only be indirectly measured from the total momentum imbalance in an event (so-called *missing transverse momentum*, see Section 5.2), it is quite important that as many particles as possible are absorbed in the HCAL.

It can be divided into four subcomponents: two of them, the hadron barrel (HB) which spans $|\eta| < 1.4$, and the hadron endcap (HE) which spans $1.3 < |\eta| < 3.0$, are still contained inside the solenoid magnet and range from $r = 177\,\text{cm}$ to $r = 295\,\text{cm}$. The hadron outside (HO) is an additional layer of scintillators placed outside the solenoid to

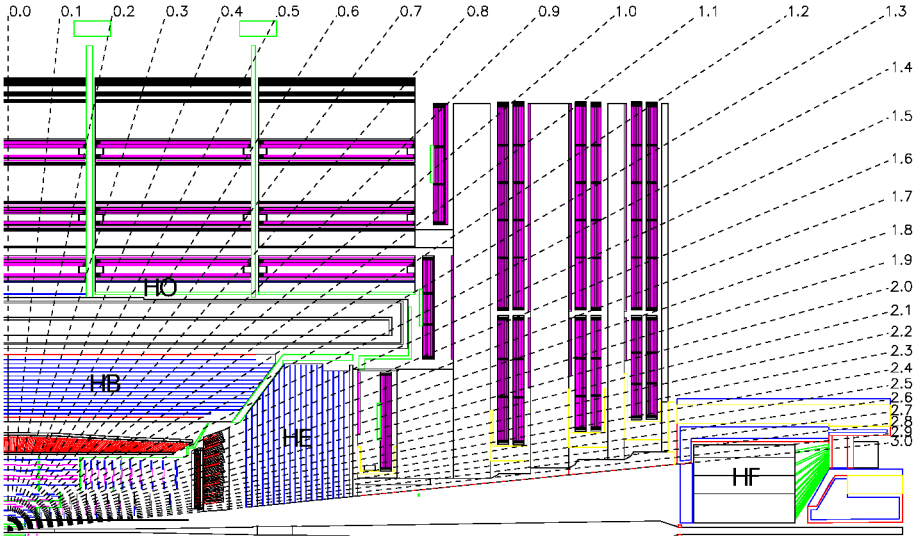

Figure 3.6: A schematic view of the CMS HCAL. Taken from Ref. [120].

measure the remaining shower energy deposited after the HB. This is necessary as the EB and HB do not provide a large enough absorption length to stop all hadronic particles so that the solenoid magnet is employed as additional layer of stopping material. The hadron forward (HF) is located close to the beam axis at a distance of 11.2 m from the nominal interaction point on either side of the detector and covers the pseudorapidity range $3.0 < |\eta| < 5.2$. It ensures the absorption of the bulk of particles produced in the bunch crossing and, as a consequence, receives by far the largest flux of particles. This led to the choice of steel as absorber and quartz fibres as scintillating material as these are more radiation hard materials. A graphical illustration of the HCAL structure is shown in Fig. 3.6.

The total thickness of the HCAL corresponds to 10–15 nuclear interaction lengths λ_I, depending on η. The energy resolution of the HCAL barrel can be parametrised with the equation

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{87.7\%}{\sqrt{E}}\right)^2 + (7.4\%)^2 \tag{3.6}$$

in which the corresponding values were found in test beam measurements for particles with momenta between 2 and 350 GeV [135].

3.2.5 Muon system

The muon system is responsible for the identification of muons and (together with the inner tracker) the measurement of their momenta. It is the outermost of all subdetectors as muons are usually not absorbed by either of the calorimeters and have a decay length far exceeding the dimensions of the CMS detector. The muon system plays an important role as muons are often part of interesting and rare physical processes (e. g. in an $H \to ZZ$

process where both Z bosons decay into muons). Furthermore, it provides a very efficient way to trigger these kinds of events.

The basic principle to detect the passage of a muon is through ionisation of gas contained in multiple so-called *chambers*. These chambers are grouped together to form *stations* and their design and position is chosen to match the geometry of the magnetic field which is different in the barrel and endcap region. The entire muon detector system is placed in between the iron return yoke which is also used as supporting structure to mount the muon chambers.

In the barrel region, where the magnetic field is uniform and has a strength of 0.4 T, *drift tube chambers* (DT) are used covering the area up to $|\eta| < 1.2$. Each chamber consists of one anode wire at the centre and four electrode strips at the walls of the chamber. It is between 1.9 and 4.1 m long, depending on its location and orientation, and has an area of $42 \times 13\,\mathrm{mm}^2$. There are four stations at different radial distances to the beam axis that contain eight chambers to measure the muon coordinate in the (r,ϕ)-plane. The first three stations (from inner to outer) contain four additional chambers to measure the muon coordinate in the z-direction.

As in the endcap region, higher muon fluxes are expected and the magnetic field is not homogeneous, *Cathode strip chambers* (CSC) are used in this region, ranging from $0.9 < |\eta| < 2.4$, as they have shorter drift paths and are therefore better suited to these conditions. The CSCs vary in their size, depending on their position in the detector, having lengths between 1.7 and 3.5 m and spanning angles in ϕ-direction between 10° and 20°. They consist of six layers, each of which contains 80 cathode strips running radially outwards and measuring the muon coordinate in the (r,ϕ)-plane, and several groups of 16 anode wires oriented perpendicular to the strips and also providing a coarse measurement of η. There are also four stations of CSCs at different radial distances to the nominal interaction point.

So-called *resistive plate chambers* (RPC) are mounted in both barrel and endcap, spanning a pseudorapidity range up to $|\eta| < 1.6$. They have an excellent time resolution of passing muons and are therefore mainly used for triggering. An RPC consists of four Bakelite plates with the two outer gaps formed by the plates filled with gas and the inner gap containing a readout strip which picks up the image charge induced by ionised gas molecules. There are six layers of RPCs in the barrel region, two layers for each of the first two DT stations and one for the last two stations, and one layer for each of the first three CSC stations in the endcap region. A graphical representation of the muon system is shown in Fig. 3.7.

The spatial hit resolutions achieved in the muon system is 80-120 μm in the DT system, 40-150 μm in the CSC system and 800-1200 μm in the RPC system. The combined time resolution in all three systems is 3 ns or better and they provide a very high detection

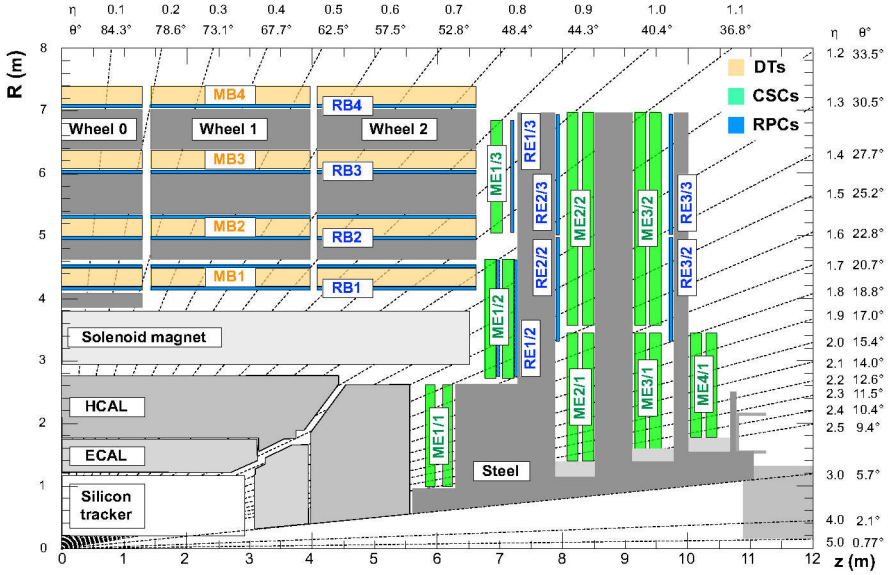

Figure 3.7: A schematic view of the CMS muon system surrounding the inner tracker, ECAL, HCAL and solenoid magnet. Taken from Ref. [136].

efficiency, resulting in an identification efficiency of the correct triggering bunch crossing of 99.5% [136].

3.2.6 Trigger

At the design luminosity of $L = 10^{34}\text{cm}^{-2}\text{s}^{-1}$ with a bunch spacing of 25 ns, the LHC produces pp collisions at a rate larger than 1 GHz [137]. As the bulk of these collisions does not feature any interesting physical processes and the computational storage capacities are limited, a sophisticated *trigger* system is employed to keep only interesting events. A detailed description of the CMS trigger system and its performance can be found in Ref. [137].

There are two stages or *levels* in the process of triggering an event: at level one (L1), events are selected based only on information from the detector subsystems and the interaction rate is reduced to $\sim$ 100 kHz. Events passing the L1 trigger are the input for the second stage, the *high-level trigger* (HLT), which performs a basic *online* event reconstruction in parallel with the data collection. Based on this information, events are selected at a rate of $\sim$ 400 Hz for the *offline* reconstruction, which is usually carried out after the data collection with a larger accuracy, resulting in much longer processing times.

The L1 trigger is completely implemented in custom hardware, mounted at different locations in the CMS detector. It is split into two parts: the *global calorimeter trigger* (GCT) combines information from the ECAL and HCAL (including the HF) and provides candidates for electrons, photons, and jets (both from partons and τ leptons) as well

as a first estimate for missing transverse energy. The *global muon trigger* (GCT) uses information from all subsystems of the muon detector (DTs, CSCs and RPCs) to identify muon candidates using fast pattern recognition and track finding algorithms. The output of both triggers is then processed in the *global trigger* (GT) in which a trigger menu is implemented that rejects or accepts events based on kinematic properties of the object candidates reconstructed in the previous steps. In case of an event being accepted, signals are sent to all detector subsystems to store more information about the event for further evaluation in the HLT trigger. The latency time of the L1 trigger is 3.2 μs during which events are stored in readout buffers for further processing.

The HLT trigger is a software-based trigger running on a single-processor farm with 13,000 CPU cores as of 2012, allowing for a per-event reconstruction time of 175 ms. It consists of a number of so-called *HLT paths*, each of which contains a series of selection steps with increasing complexity aiming at a fast, coarse reconstruction of the whole event. Information from all subdetectors are processed where the most time-consuming steps (like track and vertex reconstruction) are performed as the last steps only if an event already satisfies certain selection requirements. During the processing in the HLT trigger, several physical objects can be reconstructed based on which events are accepted or rejected:

- **Charged tracks** are reconstructed using the Kalman filter algorithm [138] from seeds consisting of two or three hits in the tracking system.
- **Jets** are identified using calorimeter information and reconstructed tracks using the *particle flow* (PF) algorithm (see Section 5.2).
- The **primary interaction vertex** is identified either from jets ("coarse" primary vertex) or, if no primary vertex can be found this way, with a gap-clustering algorithm [130] using tracks reconstructed only from pixel hits.
- **Photons** are reconstructed from ECAL and HCAL information.
- **Electrons** are reconstructed in the same way as photons but with the additional requirement of a matching charged track in the inner tracker.
- **Muons** are reconstructed from hits in the muon system that can be matched to a charged track in the inner tracker.
- Hadronic τ **leptons** are reconstructed similar to jets using calorimeter and tracker information and combining them in the PF algorithm.

Events that are accepted by the HLT trigger form the so-called *primary physics stream* and are stored on disk and eventually sent to the CMS Tier-0 for the offline event reconstruction.

4 Event simulation

In order to measure parameters of the Standard Model or test the existence of new particles in pp collisions, it is essential to compare the observed data with the theoretical prediction of the physical processes under study. This prediction can be expressed e. g. in a differential cross section as a function of a certain observable. One of the main ways to obtain this expectation is to simulate the involved physical processes and the associated rates based on their probabilities and generated random numbers. All necessary steps to do this can be performed using dedicated software frameworks, so called *event generators*, which make use of the *Monte-Carlo technique* (MC) for numerical calculations of the involved integrals so that this step is also referred to as *MC simulation*.

The simulation of an entire pp collision event is usually carried out in two steps listed below and described in more detail in this chapter:

Event generation This first step comprises the simulation of all physical processes within a pp collision up to the final stable particles that can be measured in the detector. This includes the hard scattering process of two partons within a pp collision, additional radiation from the ingoing and outgoing particles, the hadronisation of outgoing partons, decays of unstable particles and additional scattering processes within the same pp collision (*underlying event*). Pileup interactions are usually simulated separately and superimposed over the simulated hard scattering process to get a realistic bunch crossing simulation. Several event generators exist for this purpose which are introduced along with the above steps in Sections 4.1 to 4.4. Excellent introductions into the whole event generation chain can be found e. g. in Refs. [139, 140].

Detector simulation The next step after having generated a full event topology with (almost) only stable particles, the detector response to these particles is simulated using the GEANT 4 [141] software. This includes all subsystems of the CMS detector and additional structures within it (e. g. the beam pipe) and encompasses possible artifacts and noise in these subsystems. This step is described in Section 4.5.

The first step in the event generation is the simulation of the hard scattering process within a pp collision. The total cross section for the process $pp \to X$ can be related to the partonic cross section of the parton-parton scattering $ab \to X$ using the *factorisation*

theorem [142]

$$\sigma_{pp\to cd} = \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b \, f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \, \hat{\sigma}_{ab\to cd}(x_a, x_b, Q^2, \mu_R^2, \mu_F^2) \tag{4.1}$$

where the cross sections $\sigma_{pp\to cd}$ and $\sigma_{ab\to cd}$ can be either inclusive or differential cross sections. The sum runs over all partons a and b that able to participate in the process with $x_{a,b}$ being their fraction at the total proton momentum. $f_{a,b}$ are the *parton distribution functions* (PDF) which provide probability densities to find the partons a, b at momentum fractions x_a, x_b. Q^2 is the total momentum transfer in the process $ab \to cd$ and μ_R^2 is the chosen renormalisation scale at which the partonic scattering process is evaluated. μ_F^2 is introduced as the so-called *factorisation scale* which defines the threshold between the non-perturbative and perturbative regime so that additional emissions below this scale are subsumed into the PDF and emissions above it are calculated in the following simulation chain. Usually, μ_R^2 and μ_F^2 are set to the energy scale of the partonic process given by Q^2 but as they are in principle not fixed, possible variations have to be taken into account as systematic uncertainty (see Section 6.4.2).

The individual components of Eq. (4.1), the partonic cross section and the parton distribution functions, as well as parton showering schemes and underlying event calculations are summarised in the following.

4.1 Matrix element calculation

The partonic cross section for the process $ab \to cd$ is directly proportional to the corresponding matrix element:

$$\hat{\sigma}_{ab\to cd} \propto |\mathcal{M}|^2_{ab,cd} \tag{4.2}$$

This can then be calculated using the Feynman rules to a given order (see also Section 2.1). In practice, the matrix element of a given partonic process has to be calculated numerically which can be done using a variety of so-called *ME generators*.

Some generators that are widely used in the CMS Collaboration to calculate matrix elements are POWHEG [143–146], which provides calculations at next-to-leading order (NLO) precision, and MadGraph (_aMC@NLO) [147, 148] for calculations both at leading order (LO) and NLO. These generators differ, amongst other things, by the matching scheme that they apply to model the transition between the hard and soft energy regimes in a scattering process (see below). Furthermore, POWHEG provides calculations only for a fixed set of physical processes (e. g. $t\bar{t}$ + jets production) while MadGraph is able to incorporate additional processes from external sources. MadGraph is therefore often chosen to simulate processes arising in new theoretical models.

4.2 Parton distribution functions

The PDFs parametrise all effects up to the scale μ_F^2, including both low-energy and perturbative contributions. As the soft part is, however, in the non-perturbative regime, the form of the PDFs has to be determined phenomenologically from experiments. Their dependence on the factorisation scale μ_F^2 is given by the DGLAP[1] evolution equations [149–151] which, the limit of infinite orders in the partonic cross section calculation, should not affect the resulting total pp cross section. Exemplary PDF distributions for two different values of $Q^2 = \mu_F^2$ can be found in Fig. 4.1.

The PDFs used in this analysis were calculated by the NNPDF collaboration [152] using data from the HERA experiment [153] as well as from the ATLAS, CMS and LHCb Collaborations recorded during LHC Run I. The provided PDFs (which are labelled as *NNPDF3.0*) is supplemented by a set of 100 replicas which are obtained by repeating the fit to extract the PDF with different seeds for the initialisation of the random numbers used in the minimisation. These replicas are important in the determination of the systematic uncertainty in the PDFs (see Section 6.4.2).

4.3 Parton showering and hadronisation

An ingoing or outgoing parton from the hard scattering process can radiate off one or more additional gluons that carry away some of its initial momentum. These gluons can themselves produce more radiation or split into quark-antiquark pairs, which again can radiate off more gluons and so on. These radiated particles give rise to a cascade of subsequent coloured particles from the initial parton which eventually stops at the confinement scale Λ_{QCD} after which these partons begin to form hadrons. All radiation products from a sufficiently hard parton usually form particle clusters that can be measured in a detector and reconstructed as so-called *jets*.

In principle, one could calculate this within in the corresponding matrix element by including these radiation processes as higher orders in it. This is, however, practically impossible due to computational limitations and instead, this process is approximated in dedicated *parton showering* (PS) algorithms which follow empirical models.

These algorithms operate by iteratively progressing from the hard scattering scale down to lower energy scales: it starts with the initial parton at a certain energy scale Q_i^2 and radiates off an additional parton at a scale q^2 based on the probability that no such radiation occurs between Q_i^2 and q^2. This probability is given by a *Sudakov form factor* and certain *splitting functions* based on the DGLAP equations. This is repeated until the generated q^2 is below a certain cutoff scale Q_0^2 up to which parton splittings are considered as not resolvable by a detector any longer. The scale q^2 is also referred to

[1] Acronym for Dokshitzer-Gribov-Lipatov-Altarelli-Parisi.

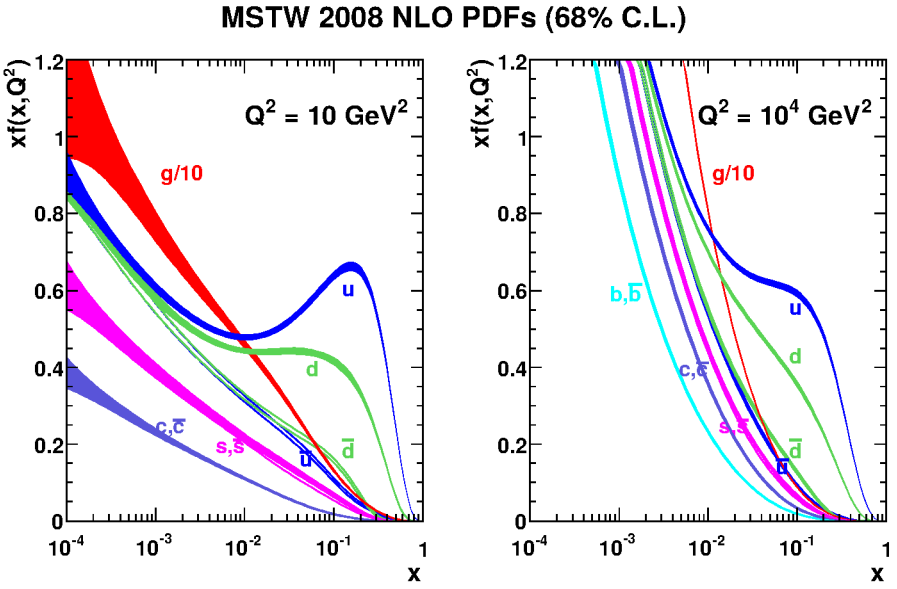

Figure 4.1: Exemplary PDFs as a function of the momentum fraction x for several quark flavours and gluons. The momentum transfer of the hard scattering process is set to $Q^2 = 10\,\mathrm{GeV}^2$ (left) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (right). Taken from Ref. [154].

as *ordering variable* and it is possible to choose a different observable (e. g. the angle of radiated partons) for this instead of q^2. The choice of the ordering variable usually varies in different event generators. Parton showers originating from final state partons are referred to as *final state radiation* (FSR). They can also be simulated for partons in the initial state, so-called *initial state radiation* (ISR), in which case this is done by a backwards evolution from the hard scattering process.

This process eventually stops through the formation of forming mesons and baryons, a process referred to as *hadronisation*. As this occurs exclusively in the non-perturbative regime, it has to be approximated using phenomenological models as well of which the most popular ones are the *Lund string model* [155, 156] and the *cluster model* [157, 158]. Decays of hadrons with a rather short life time are also modeled in most event generators.

An additional difficulty arises if a certain final state topology in a given phase space can be generated both in the ME calculation (by including an appropriate number of radiated partons) and in the PS algorithm. To account for this possibility and avoid a double-counting of these processes, so-called *matching schemes* are applied. For processes simulated at LO, a procedure often used is the *MLM scheme* [159] while for NLO order, one can use the *FxFx merging scheme* [160].

All these processes described above are usually simulated within one software framework that is interfaced to a ME generator. Some of the most popular of these MC generators are listed below:

- PYTHIA [161,162] is used to model parton showering, in which it uses the transverse momentum as ordering variable, and hadronisation based on the Lund String model. It includes underlying event simulation and has also some generic matrix elements stored which can be used to generate $2 \to 2, 3$ processes at LO. Therefore, it can also be used to simulate the entire event chain and is often referred to as "multi-purpose" event generator.

- HERWIG++ [163] is a multi-purpose event generator similar to PYTHIA. It uses angular ordering in the calculation of parton showers and the cluster hadronisation model to simulate the fragmentation of partons. It can be interfaced to a standalone software to simulate the UE and can compute $2 \to 2$ matrix elements at LO.

4.4 Underlying event

A *pp* interaction can, in addition to the hard scattering process, include further parton-parton scattering processes which are referred to as *multiple parton interactions* [164]. These interactions are usually rather soft and can lead to additional jets in the final event topology where these contributions are commonly referred to as *underlying event*. Jets can also be produced by remnants of the scattering proton that do not undergo any further interactions, though these are in most cases almost collinear to the beam axis so that they can usually be ignored. The simulation of the underlying event is usually included in the event generator performing the parton showering and based on empirical models, so-called *tunes*.

4.5 Detector simulation

After the simulation of the entire scattering process including the hadronisation of partons, a set of particles is forwarded to a detector simulation for which in this analysis, GEANT 4 is used. This models the entire geometry of the CMS detector and the response of all subsystems, including possible noise and material interactions such as photon conversions or nuclear interactions. It can furthermore simulate the decay of hadrons with a rather short life time if they have not been decayed in the previous event simulation step (see Section 4.3). The output of this simulation step should be identical to the actual digitised output of the detector and can be passed on to the next step, the event reconstruction, which is discussed in Chapter 5.

5 Event reconstruction and particle identification

Each recorded proton bunch collision produces an enormous amount of data in the form of the outputs of all subdetectors. In order to analyse this data and study the underlying physical processes, it is highly useful to reconstruct candidates for physical objects – such as individual particles or jets – from this data. This chapter provides an overview over some of the most important reconstruction algorithms which are also used in the analysis presented in Chapter 6. It also describes measurements performed to test these algorithms in data and simulation, and derive corrections (so-called *scale factors*) to account for observed differences. This step is of considerable importance as in order to make any statements about the potential presence of new physics, one first needs a good understanding about the used reconstruction algorithms and the validity of the simulation.

This chapter is organized as follows: in Section 5.1, the procedure to reconstruct the tracks of charged particles using information from the inner tracking system is described as well as the reconstruction of the interaction points of pp collisions. Section 5.2 then discusses the particle flow algorithm used to reconstruct candidates and identify particle types for individual particles arising from a bunch crossing. In Sections 5.3 and 5.4, the identification criteria to define muon and electron candidates for the final event analysis are outlined as well as measurements of their efficiency in data and simulation. Section 5.5 provides an overview over the clustering algorithm used to identify jet candidates and calibration methods to match their energy to the initial parton as well as between data and simulation. In the final two sections of this chapter, more sophisticated methods are discussed to reconstruct candidates for elementary particles occurring only in rare processes, namely to identify b quarks in Section 5.7 and heavy particles such as t quarks and heavy bosons in Section 5.8.

5.1 Track and primary vertex reconstruction

5.1.1 General track reconstruction with the CTF algorithm

As charged particles leave a trace of hits in the pixel and strip detector layers of the inner tracking system (described in Section 3.2.2), these hits can be combined to reconstruct their tracks as so-called *inner tracks*. A high reconstruction efficiency and small momentum resolution is crucial for a good performance of the CMS detector as the tracks provide the highest energy resolution for electrons, muons, and charged hadrons. The *track reconstruction* is the first step in the complete event reconstruction upon which many other

reconstruction steps depend. This section provides a short overview over the track reconstruction at the CMS experiment, a detailed overview can be found in Ref. [130].

An important part of the track reconstruction is the identification of hit clusters in the pixel and strip detectors from their electronic output. It is important to maintain a high hit detection efficiency and good hit position resolution while simultaneously minimising the contribution from electronic noise in the readout chips. This is done based on different thresholds for the electronic charge for each pixel and using different algorithms in the pixel and strip detectors which are explained in detail in Ref. [130].

The actual track reconstruction algorithm is based on an extension of the *Kalman filter* algorithm (KF) [138] referred to as *combinatorial track finder* algorithm (CTF) [165–167]. It aims at finding patterns that are compatible with particle trajectories and starts from so-called *seed hits*. It then extrapolates from these seeds based on probability densities of various track variables to estimate the position of the next hit. A final particle trajectory is provided by performing a least-squares fit of the detected hits which returns a χ^2 that can be used to evaluate the quality of the fit.

To reduce the computational complexity of this procedure, as there is usually a huge number of tracker hits resulting in an immense combinatorial multiplicity, the CTF is applied in several iterations. The goal is to progress from the easiest tracks to reconstruct to the most difficult ones so that in each iteration, only a limited number of hits is considered which are then masked for the next iteration, significantly reducing the computational overhead of the entire calculation. While the first iteration aims at reconstructing prompt tracks with $p_\mathrm{T} > 0.8\,\mathrm{GeV}$ that have at least three pixel hits, the number of pixel hits and the assumed p_T are reduced in each following iteration and in the fifth and final iteration, low-p_T tracks originating from outside the so-called *beam spot* are included. The beam spot marks the centre of the estimated *luminous region* and is not reconstructed for each event individually, but as an average over many events. The exact procedure to determine the beam spot is described in Ref. [130] and its position is reconstructed with a precision of approximately $50\,\mu\mathrm{m}$.

Within each iteration, the above procedure is used to find tracks which are then accepted or rejected based on several quality criteria. These vary depending on the current iteration number and include:

- a minimum number of hit layers,
- a maximum number of missed layers,
- a reasonably low χ^2/n_dof,
- low transverse and longitudinal *impact parameter significances* $|d_0^\mathrm{BS}|/\sigma_{d_0}$ and $|d_z^\mathrm{PV}|/\sigma_{d_z}$ with respect to the beam spot and the closest pixel vertex, respectively.

$|d_0^{\mathrm{BS}}|$ is the transverse impact parameter with respect to the beam spot and $|d_z^{\mathrm{PV}}|$ is the longitudinal impact parameter with respect to the closest pixel vertex which is defined as a pixel hit with at least three tracks associated. The corresponding IP significances are the ratios of the impact parameter and their associated uncertainties σ_{d_0,d_z} so that a high IP significance indicates a large probability that the track does not originate from the primary interaction point. The exact requirements to keep a track after each iteration are provided in Ref. [130].

5.1.2 Electron track reconstruction with the GSF algorithm

As electrons can loose energy when traversing the detector by emitting bremsstrahlung, a special tracking algorithm is run to correctly determine their momentum and direction. Seeds for these electron tracks are found based on two different approaches: in the *ECAL-based* approach, seeding hits in the inner tracker are inferred from the positions of so-called superclusters (these are energy deposits in the ECAL corresponding to bremsstrahlung photons, see Section 5.2 for a more detailed description), in the *tracker-based* approach, inner tracks are used as seeds that are compatible with the associated cluster energy from the corresponding block. However, the emission of bremsstrahlung can lead to a small number of hits used in the CTF algorithm and result in a large χ^2 of the corresponding fit, as the CTF algorithm is not suited for significant momentum losses of a track and resulting kinks in its trajectory. Such poorly reconstructed tracks are therefore re-fitted with a Gaussian sum filter (GSF) [168] algorithm to provide seeds in the tracker-based approach as this can better handle energy losses of an electron due to bremsstrahlung. All seeds from both the ECAL-based and the tracker-based approach are then used as input for a final GSF fit. A more detailed description of the track reconstruction for electron candidates, can be found in Ref. [169].

5.1.3 Primary vertex reconstruction

During the 2015 data taking period at $\sqrt{s} = 13\,\mathrm{TeV}$ with a bunch-spacing of 25 ns, on average 10 *pp* scattering processes occurred per bunch crossing. Identifying the hard scattering process amongst these interactions with a high precision is therefore essential to distinguish its associated tracks from the ones originating from IT pileup interactions[1]. This position is referred to as *primary interaction vertex* or, short, *primary vertex* (PV).

As a first step, the positions of all *pp* interactions are attempted to be reconstructed after the track reconstruction step using inner tracks. The tracks used for this reconstruction have to fulfill the following requirements:

[1] In the following, only the contributions of IT pileup are considered as OOT pileup usually just results in a small offset in the CMS subdetectors that is corrected for in all reconstruction algorithms. This IT pileup will therefore be simply referred to as "pileup".

- They have to have transverse impact parameter significance $|d_0|/\sigma_{d_0} < 5$ with respect to the beam spot.

- They have to have at least two hits in the pixel layers and five hits in the strip detectors.

- The fit of the particle trajectory must return $\chi^2 < 20$.

These tracks are then clustered based on the z-coordinate of their point of closest approach to the beam spot with a *deterministic annealing* (DA) algorithm [170] and in each cluster, a vertex is fitted with an *adaptive vertex fitter* (AVF) [171]. In the vertex fit, tracks are assigned weights w_i based on their estimated probability to really originate from the vertex to render the fit more resistant against outliers. The number of degrees of freedom (n_{dof}) in this fit is then defined as the sum of track weights $n_{\mathrm{dof}} = -3 + 2\sum w_i$.

The vertex with the highest sum of transverse momenta of associated tracks $\sum p_{\mathrm{T}}^2$ which also has $n_{\mathrm{dof}} > 4$ is identified as the primary interaction vertex and all other vertices are thus denoted as pileup vertices in the following.

The resolution of the vertex position depends on the number of tracks used to fit its location and on the overall occupancy of the inner tracking system. For vertices which are reconstructed from at least 50 tracks (which is often the case for hard scattering processes, in particular for processes producing many jets), the resolution ranges from $10\,\mu$m to $20\,\mu$m.

If at least three tracks can be reconstructed that correspond to particles from a scattering process, the efficiency of the vertex reconstruction described above is close to 100%. This is, however, not always the case as the track reconstruction is not 100% efficient and many scattering processes can produce uncharged particles or tracks that are outside of the acceptance of the inner tracking system. Consequently, on average only approximately 70% of all *pp* interactions can be reconstructed as a vertex.

Both the track and vertex reconstruction algorithms provide an excellent momentum and position resolution for charged particles, owing to a large degree to the high granularity of the inner tracking system. As an example, it allows to reconstruct the $\mathrm{K}^* \to \mathrm{K}^0_{\mathrm{S}}\pi^- \to \pi^+\pi^-\pi^-$ and $\Xi^- \to \Lambda^0\pi^- \to p^+\pi^-\pi^-$ decays by reconstructing the decay vertices of the intermediate heavy hadrons from the charged daughter particles [172]. The resulting invariant mass spectrum can be seen in Fig. 5.1 which shows clear peaks around the masses of the K^* and Ξ^- resonances, illustrating the high precision achieved in the track and vertex reconstruction.

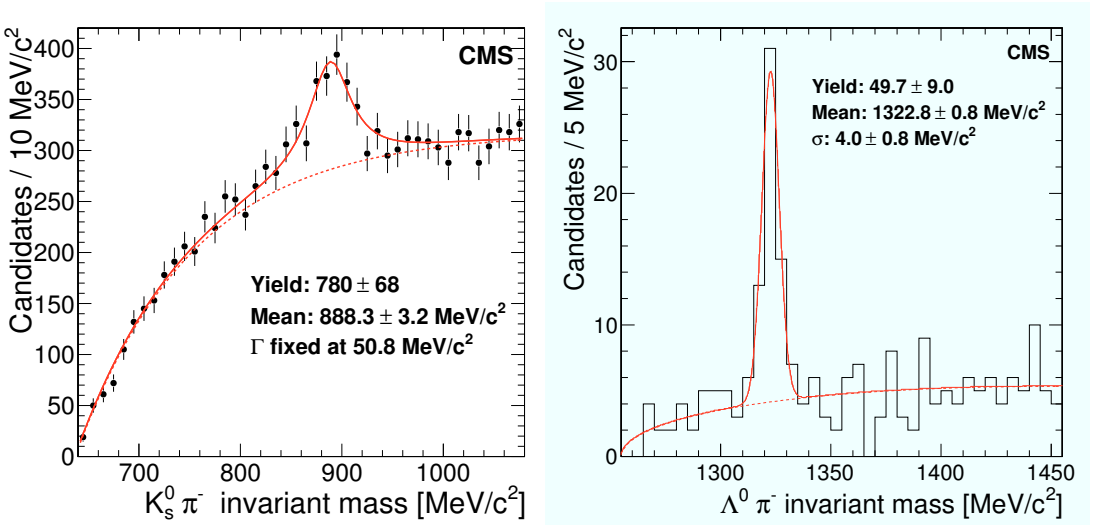

Figure 5.1: Invariant mass plots of the $\pi^+\pi^-\pi^-$ system around the K* resonance (left) and of the $p^+\pi^-\pi^-$ system around the Ξ^- resonance (right). Taken from Ref. [172]

5.2 Particle reconstruction with the particle-flow algorithm

The *particle-flow* (PF) algorithm [173, 174] reconstructs candidates for stable particles that can be used for the following analysis. "Stable", in this context, includes all particle types with life times long enough that they will not decay before leaving the detector or being absorbed by it. These particle candidates can then be directly used (in the case of electrons, muons and photons) to reconstruct the initial physical process or used as input for algorithms to reconstruct objects of higher complexity (such as jets, τ leptons, interaction vertices etc.).

The PF algorithm operates iteratively, reconstructing particle candidates step-by-step by combining information from all subsystems of the CMS detector. It uses as input several basic elements, the most important ones of which are inner tracks and electron tracks reconstructed with the GSF algorithm (see Sections 5.1.1 and 5.1.2) as the momentum measurement from tracks provides in many cases the best resolution for the momentum of a particle candidate.

Another important component are *calorimeter clusters* in the ECAL and HCAL which are primarily used to identify neutral particles such as photons and neutral hadrons and measure their energy and also complement the energy measurement from track information for electrons and charged hadrons. These clusters are found in a two-step algorithm by first finding "seed cells" that correspond to local energy maxima, and then attaching more and more calorimeter cells to these if they measured an energy deposit well above the noise in the detector. In addition to these clusters, so-called *superclusters* are identified in the ECAL from energy deposits in calorimeter cells if their hit pattern is likely caused by

an array of bremsstrahlung photons [169].

Another important basic element are so-called *standalone-muon tracks* which are reconstructed from the hits in the muon chamber alone [175]. Usually, these tracks are only considered if their extrapolated trajectory is compatible with the primary interaction point, in order to reject muons from cosmic rays.

These individual components are then combined in the next step by a so-called *link algorithm* to form several *blocks*, each one consisting of one ore more tracks (either inner tracks or GSF tracks), zero or more (super-)clusters in both ECAL and HCAL and zero or one muon track. Links between tracks and calorimeter clusters are constructed by first extrapolating inner tracks to the calorimeters to find compatible clusters and then extrapolating also tangents of these tracks to find clusters that could originate from bremsstrahlung photons emitted by electrons. In addition to this, tracks that are compatible with an e^+e^- pair from the conversion of a bremsstrahlung or prompt photon are linked to the corresponding blocks along with the calorimeter clusters that these tracks can be extrapolated to.

After linking tracks to clusters, (super-)clusters in the PS, ECAL and HCAL are linked together if they are close to each other in the (η, ϕ)-plane. Inner tracks and muon tracks are then linked if they can be combined in a global fit with a reasonable χ^2 [120].

These blocks are then used to provide candidates for all particle types, referred to as *particle-flow candidates*. They are reconstructed in several iterations, after each of which the blocks used to identify a PF candidate are removed from the list of blocks considered for the next step. By starting with the computationally least expensive steps, this helps in significantly reducing the time to perform all steps and maintain a time-efficient algorithm. A summary of all PF candidates and how they are identified is listed below.

Muons Candidates for *PF muon* candidates are provided by blocks containing an inner track linked to a muon track. In addition to this, there exist three different types of muon candidates and PF muon candidates usually have to be independently identified as at least one of these types: *Standalone muons* are identified from standalone-muon tracks that, as explained above, are reconstructed using only hits in the muon chambers, while *tracker muons* are formed by inner tracks that, if extrapolated, are compatible with at least one hit muon chamber. Finally, *global muons* are identified from standalone-muon tracks and inner tracks that are compatible with each other when being extrapolated. They are then associated with a *global muon track* reconstructed from the combined set of hits in the inner detector and the muon chambers. PF muon candidates usually have to be either global muons or, if they have a p_T below 10 GeV, tracker muons. Their momentum is then set to the momentum of the inner track or, if it exceeds 200 GeV, from the track (either inner track, global muon track or different combinations of inner tracker and muon chamber hits) returning

the smallest χ^2 when being fitted.

Electrons Suitable blocks to obtain *PF electron* candidates are selected if the energy in linked HCAL clusters does not exceed the energy of ECAL clusters by more than 10% and if there is at least one GSF track and in total not more than two tracks in the block. All ECAL clusters linked either to track tangents or contained within superclusters are associated to a PF electron candidate, as well as the GSF track and all tracks of identified photon conversions. To eventually pass the PF algorithm and be returned as a suitable electron candidate, various properties of the corresponding block elements are combined in a multivariate procedure [174]. These variables include, amongst others, the amount of energy radiated off the GSF track and the offset between the extrapolated GSF track and the ECAL cluster location. Finally, the energy and momentum of a PF electron candidate is then calculated by combining all the information from the GSF track, photon conversion tracks, and calorimeter (super-)clusters together. For low-p_{T} and rather central electron candidates, their energy is almost completely determined by the associated GSF track while for high-p_{T} or very forward electron candidates, the energy measured in the ECAL becomes more important.

Charged hadrons *PF charged hadron* candidates are identified from blocks containing inner tracks and clusters in both the ECAL and HCAL in which the momentum of the tracks and the energy deposited in the calorimeters have to be compatible. To make an accurate and reliable comparison, the energy deposits in the ECAL and HCAL have to be re-calibrated using simulated charged hadrons [173] and only inner tracks are considered with a relative uncertainty in their momentum less or equal than the relative uncertainty in the calibrated cluster energy in the calorimeters. The energy and momentum of PF charged hadron candidates is then taken directly from the track using the charged pion hypothesis. At high energies or large pseudorapidities, the momentum resolution of the tracks becomes worse and the four-momentum of the charged hadron is obtained by combining track and calorimetric cluster energy. In the very forward region outside of the tracker acceptance, their energy has to be calculated entirely from the combined ECAL and HCAL energy deposits as no distinction between charged and neutral particles is possible in this region.

Photons and neutral hadrons *Isolated PF photon* candidates are identified from ECAL superclusters with an energy deposit larger than 10 GeV with no inner tracks or other calorimeter clusters in the close vicinity. Similar to PF electron candidates, the energy in linked HCAL clusters must not exceed 10% of the energy of the linked ECAL clusters. Any remaining calorimeter clusters not linked to inner tracks give rise to *non-isolated* PF photon candidates (e. g. from neutral pion decays) or *PF*

neutral hadron candidates. This is also the case for calorimeter clusters associated to PF charged hadron candidates if their energy exceeds the sum of the associated track momenta with a difference larger than the energy resolution of the calorimeter.

Missing transverse momentum and energy As neutrinos couple only to the weak force and almost never interact with other particles, it is in principle impossible to reconstruct their momentum and energy as they do not produce any detectable signal. The sum of all other particle momenta produced in the scattering process, however, should be equal to zero. Therefore, the energy and flight direction of neutrinos can be estimated at least in the transverse plane from the measured imbalance of the total transverse momentum. This quantity, referred to as *missing transverse momentum* is therefore defined as

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_i \vec{p}_{\mathrm{T},i} \tag{5.1}$$

where i runs over all PF candidates and $\vec{p}_{\mathrm{T},i}$ are the corresponding transverse momentum vectors. Often, one is only interested on the magnitude of this vector which is referred to as *missing transverse energy* and defined as

$$\not{E}_{\mathrm{T}} = |\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}| \tag{5.2}$$

Relative lepton isolation A variable called *relative isolation* is defined for a lepton as a function of its p_{T} by the equation

$$I_{\mathrm{PF}}(p_{\mathrm{T}}) = \frac{1}{p_{\mathrm{T}}} \left(\sum_{\mathrm{h}^{\pm}} p_{\mathrm{T}}^{\mathrm{h}^{\pm}} + \sum_{\gamma} p_{\mathrm{T}}^{\gamma} + \sum_{\mathrm{h}^0} p_{\mathrm{T}}^{\mathrm{h}^0} \right) \tag{5.3}$$

where $\mathrm{h}^{\pm}$, γ and h^0 refer to the charged hadron, photon and neutral hadron PF candidates, respectively. This variable is useful to distinguish leptons directly produced in the hard scattering process from soft leptons produced e. g. in b quark decays as the latter ones are usually surrounded by other particles from the hadronisation of the quark.

If after the track selection and calorimeter cluster energy calibration – i. e. when identifying charged hadrons – the total momentum from inner tracks exceeds the total calorimetric energy by more than three standard deviations of the combined uncertainty in these quantities, the PF algorithm tries to identify more PF muon candidates and/or remove tracks that do not correspond to actual charged particles ("fake" tracks). This is done by relaxing the selection criteria for global muons that are not yet identified as PF muon candidates and by tightening once more the criteria for inner tracks.

5.3 Muon identification

Prompt muon candidates (i. e. muons produced directly in the hard scattering process) for the analysis presented in Chapter 6 are selected from the pool of PF muon candidates if they meet certain requirements recommended by the CMS Collaboration [176]. In addition to passing the selection criteria of the PF algorithm, they have to be reconstructed either as a global or tracker muon with more than 80% of the hits in the tracker system being valid hits. Furthermore, a muon candidate has to have a *segment compatibility* of at least 0.451 where this variable provides an approximate probability that a hit pattern in the inner tracking system and in the muon chambers are caused by the same track. If this requirement is not fulfilled, a slightly lower segment compatibility of 0.303 is required in addition to a set of other requirements. These include that the muon candidate must be reconstructed as a global muon, that the fits of the global track and to match the inner track to the standalone-muon track return reasonably low χ^2 values and no kinks in the track are identified due to interactions with tracker material. The efficiency to successfully identify a muon is around 99.5% while the probability to misidentify a hadron or a random pattern of hits (*combinatorial background*) as muon is below 1%.

The efficiency of this identification efficiency can be measured in data and simulation using a so-called *Tag-and-Probe* method [176, 177]. In this procedure, one first chooses a distinct physical process which can be selected with a high purity ("tag"). By choosing a suitable process, it is automatically ensured that events in the selected sample contain the particle for which one would like to calculate the identification efficiency ("probe") with a high probability. The ratio between the "tagged" events and the subset of events that pass the "probe" selection then provides the desired identification efficiency which can be derived in data and simulation.

The efficiency of the muon identification procedure has been measured by the CMS Collaboration using data from pp collisions recorded in 2015 at $\sqrt{s} = 13\,\text{TeV}$ and the resulting scale factors are used in this analysis. This was done by selecting a sample enriched in events from $Z \to \mu^+\mu^-$ decay processes by requiring one "tag" muon. The other muon from the Z boson decay was then used as "probe" muon to perform the measurement. As a further requirement, the invariant mass of the muon pair had to be close to the mass of the Z boson. The exact composition of the data samples before and after applying the "probe" selection was determined by fitting a function with several components to the invariant mass spectrum of the dimuon pair. The ratio of the fitted amplitudes parameters corresponding to actual $Z \to \mu^+\mu^-$ events before and after the "probe" selection provided the identification efficiency.

The scale factors obtained in that measurement, given by the ratio of the identification efficiencies in data and simulation, depend on the p_T and η of the muon and were found to have values between 0.96 and 0.995. As these correction factors are rather small this

indicates a relatively good modelling of muons in the simulation.

5.4 Electron identification

To identify prompt electron candidates for the analysis described in Chapter 6 a set of additional identification criteria [169] are applied on top of the requirements of the PF algorithm (see above). They have to be distinguished both from charged hadrons and non-prompt electrons which are produced in decays of certain hadrons. These include

- The *lateral shower extension* of the associated supercluster $\sigma_{\eta\eta}$ has to be smaller than 0.0101 in the barrel region and 0.0283 in the endcap region. $\sigma_{\eta\eta}$ is defined as the spread of the supercluster in η where ECAL cells are weighted with the logarithm of their energy.

- The angular differences $|\Delta\eta|$ and $|\Delta\phi|$ between the the GSF track and the associated ECAL cluster of the electron candidate have to be lower than 0.0103 (0.00733) and 0.0336 (0.114), respectively, in the barrel (endcap) region.

- The ratio of deposited energies in the HCAL and ECAL has to be lower than 0.0876 in the barrel region and 0.0678 in the endcap region.

- The difference between the inverse of the supercluster energy E_{SC} and the inverse of the momentum p of the GSF track $|1/E_{SC} - 1/p|$ has to be lower than 0.0174 in the barrel region and 0.0898 in the endcap region.

- The transverse and longitudinal impact parameters $|d_0|$ and $|d_z|$ with respect to the primary interaction vertex have to be lower than 0.0118 (0.0739) and 0.373 (0.602), respectively, in the barrel (endcap) region.

- Only two or less layers of the inner tracking system (including pixel and strip detector) may not contain hits associated to the GSF track.

- Electron candidates with tracks identified as originating from photon conversions are rejected. This is determined if the innermost layer of the pixel detector does not contain a hit associated to the track, if the track is part of a two-track system that have the same tangent at the conversion vertex, and if other requirements based on its transverse and longitudinal impact parameters are fulfilled.

These selection requirements result in an identification efficiency for prompt electrons of approximately 80% and in a misidentification probability for charged hadrons or non-prompt electrons (e. g. from B or D meson decays) of approximately 1%.

The efficiency of these identification criteria has been measured as well by the CMS Collaboration with pp collision data recorded in 2015 at $\sqrt{s} = 13\,\text{TeV}$. The same "Tag-and-Probe" method as explained in Section 5.3 (see also Ref. [169]) has been applied with a very similar selection, targeting $Z \to e^+e^-$ events. To select a sample enriched in these events, a trigger requiring a single isolated electron with $p_\text{T} > 24\,\text{GeV}$ where the "Tag" electron has to have an offline p_T of greater than 25 GeV in the barrel and 30 GeV in the endcap and fulfill tighter selection criteria than the ones listed above.

The same procedure to extract the efficiencies by fitting different functions to the invariant dielectron mass spectra is used as described in Section 5.3. The resulting scale factors vary between 0.94 and 1.09 depending on the p_T and $|\eta|$ of the supercluster of the electron candidate.

5.5 Jet clustering and reconstruction

Quarks or gluons emitted in a proton-proton collision cannot be detected as individual particles in the detector as they hadronise soon after their production and create a shower of charged and neutral hadrons. These particle showers can be measured in the detector and reconstructed as *jets* by use of certain clustering algorithms. This is an important step in the event reconstruction to accurately reconstruct the initial scattering process and various algorithms exist for this purpose. The input to such algorithms can be the four-momenta of an arbitrary set of particles. Usually, the PF candidates (see Section 5.2) obtained after the initial event reconstruction steps are used as input (either in simulated or actual data events) but one can also obtain *generator jets* by using four-momenta of simulated particles directly.

A widely used clustering procedure is called *sequential recombination* which is based on the following equations for objects marked by indices i and j:

$$\begin{aligned} d_{ij} &= \min\left(p_{\text{T},i}^{2k}, p_{\text{T},j}^{2k}\right) \frac{\Delta R_{ij}^2}{R_C^2} \\ d_{iB} &= p_{\text{T},i}^{2k} \end{aligned} \tag{5.4}$$

An object can be either a single four-momentum or a cluster of four-momenta created in a previous iteration. ΔR_{ij} is the angular separation between the two objects and k and R_C are pre-defined parameters that influence the shape and size of the reconstructed jets. The algorithm then proceeds in several iterations: in each iteration, the quantities d_{iB} and d_{ij} are calculated for each object or pair of objects, respectively. An object i is now clustered to the object j that has the smallest value for d_{ij} of all considered combinations if $d_{ij} < d_{iB}$. If this is not fulfilled, the object i is denoted as jet and removed from the pool of objects. These steps are performed until there are no more objects left so that

all provided input four-momenta are clustered into jets. This provides a *collinear*- and *infrared*-safe algorithm i. e. the resulting jets are not sensitive to very soft or low-angle radiation from initial partons.

The direction, energy and momentum of any jet is then given by the sum of four-momenta of all clustered particles. Each jet then also has a mass term (the *jet mass*) which should however be treated with caution as this can significantly differ from the mass of the initial parton (see also Section 5.8). Angular distances ΔR to other particles or jets are measured with respect to the jet axis i. e. the three-momentum sum of all its constituents.

The choice of the parameters k and R_C defines the eventual topology of the reconstructed jets. Choosing $k = 1$, in which case the clustering algorithm is referred to as k_T *algorithm* [178,179], results in low-p_T particle candidates being clustered first. In the case $k = 0$, for which the algorithm is called *Cambridge/Aachen algorithm* (C/A) [180, 181], the clustering only depends on the angular separation of the particle candidates. In both cases, the associated particle candidates and thus the shape and p_T of the reconstructed jets depend considerably on the soft constituents in an event. To reconstruct jets whose properties mainly depend on their hard constituents, it is useful to set $k = -1$ in the jet clustering algorithm, which is then referred to as *anti-k_T algorithm* [182]. This produces jets with a nearly circular *cone size* (corresponding to R_C) around high-momentum particles which are usually the hardest particles from the initial scattering process. These often carry the bulk of the momentum of an initial parton and constituents originating from soft radiation from these partons have only little effect on the jet shape and momentum.

There are two types of jets used as standard in CMS analyses and in particular in the analysis presented in Chapter 6 which are both reconstructed from PF candidates using the anti-k_T algorithm with different cone size parameters R_C:

- jets with a cone size parameter $\mathbf{R_C = 0.4}$ are referred to as **AK4 jets** and are well suited to capture all decay products from hard gluons and quarks up to the b quark
- jets with a cone size parameter $\mathbf{R_C = 0.8}$ are referred to as **AK8 jets** and are well suited to capture the decay products from two or more close-by partons, e. g. the two quarks from the decay of a Higgs boson with a large Lorentz boost (see also Section 5.8)

In the analyses presented in Chapter 6 and Chapter 7, both types jets have to be within the central part of the detector with $|\eta| < 2.4$ and AK4 jets must have $p_\mathrm{T} > 30\,\mathrm{GeV}$ while AK8 jets must have $p_\mathrm{T} > 200\,\mathrm{GeV}$. Furthermore, jets of both types have to pass a set final requirements in order to ensure that no isolated electrons, muons or photons are identified as jets since the jet clustering algorithm usually creates individual jets from such particles.

They have to consist of at least two PF candidates and the fraction of energy deposits associated to neutral particles has to be below 99% of the total energy in both the ECAL

and HCAL in order to avoid having a jet from a single particle. Furthermore, they must contain at least one charged particle and the energy fraction of charged particles must be below 99% and above 0% in the ECAL and HCAL, respectively.

After passing these requirements, a final cleaning step is performed to remove AK4 jets that are reconstructed from prompt leptons that are not entirely isolated. This is done by comparing each AK4 jet to prompt muons and electrons that are within $\Delta R < 0.4$ of the jet. For both types of leptons, a p_{T} greater than 20 GeV is required and muons must have $|\eta| < 2.1$ and pass the identification criteria outlined in Section 5.3, while electrons must suffice $|\eta| < 2.4$ and pass the identification criteria described in Section 5.4. If the energy fraction of the jet associated to PF muon candidates is within 10% of the energy of a muon fulfilling these requirements, the four-momentum of the muon is subtracted from the one of the jet and the above selection steps are repeated for the jet. Analogously, if the energy fraction of the jet deposited in the ECAL is within 10% of the energy of an electron meeting the previously described criteria, the same procedure is employed.

One problem in the jet reconstruction is the effect of pileup interactions as they can lead to additional particle candidates being clustered into jets from the hard scattering process that one is usually interested in. To reduce the contribution from additional interactions in the bunch crossing, all particle candidates from with tracks not originating from the primary interaction vertex (see Section 5.1.3) are removed from jets. This is referred to as *charged hadron subtraction* (CHS) and applied to all jets in the analysis presented in Chapter 6. However, as only charged particles from pileup interactions can be removed this way, jets still usually have a contribution from neutral particles from pileup interactions. This is also corrected for as is explained in the following.

5.6 Jet energy corrections

The energy of reconstructed jets usually has a strong dependence on the response of various subsystems of the CMS detector such as the ECAL, as well as on the number of pileup interactions in an event. This can be seen in Fig. 5.2 where the *jet energy response* in simulated events, defined as the ratio of the energies of the initial parton and the reconstructed jet, is shown. In order to get an accurate estimate of the energy of the initial parton which produced the jet, several so-called *jet energy corrections* are applied to reconstructed jets. These corrections are applied in multiple iterations, so-called *levels*, so that first, the jet energy scales are corrected to match the corresponding energies of the initial partons and then, so-called *residual corrections* are applied to jets in the data to match the detector response in simulation. The procedure is described in detail in Ref. [183] and is described in the following.

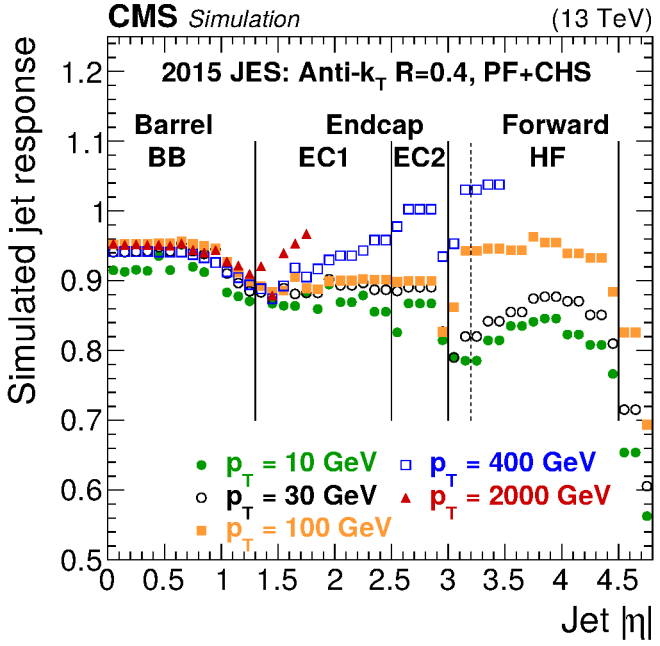

Figure 5.2: Jet response for simulated events at $\sqrt{s} = 13\,\text{TeV}$, defined as ratio of the initial parton energy and reconstructed jet energy before any jet energy corrections. The response is shown as a function of the jet $|\eta|$ for jets with varying p_T. Taken from Ref. [184].

5.6.1 Level-1 pileup corrections

These corrections aim at eliminating any contributions to the jet energy from pileup interactions remaining after the CHS procedure. The correction factor for an individual jet depends on its p_T and η, as well as on its area A and the diffuse offset energy ρ in an event. The jet area A is evaluated by adding a large number of extremely soft so-called *ghost particles* to the event and re-running the jet clustering algorithm. The size of the region in the (η, ϕ)-plane occupied by these soft particles clustered into a jet defines the area for this jet. The diffuse offset energy ρ is calculated by defining a grid of cells in the (η, ϕ)-plane and then computing the median of energies in these cells.

Based on these quantities, correction factors can be determined from simulated dijet events by processing these once with and once without additional pileup interactions. These correction factors are applied to jets both in simulated events and data. An additional scale factor is applied to simulated jets to correct for an offset between data and simulation which is determined from so-called *zero-bias* data streams that are collected with a trigger that randomly selects data events and therefore almost exclusively consists of events with merely inelastic scattering processes.

5.6.2 Level-2 relative & level-3 absolute

The next step of energy corrections is intended to provide a uniform jet response over the whole η-range (level-2) and finally match the absolute jet energy scale to the one from the original partons (level-3). This is derived in simulated multijet events by first matching the initial partons from the hard scattering process to reconstructed jets if they have $\Delta R < 0.2$. Then, a jet response function is defined

$$R_{\text{ptcl}}(p_{\text{T,ptcl}}, \eta_{\text{reco}}) = \frac{\langle p_{\text{T,reco}} \rangle}{\langle p_{\text{T,ptcl}} \rangle}[p_{\text{T,ptcl}}, \eta_{\text{reco}}] \tag{5.5}$$

where "ptcl" and "reco" denote quantities from the simulated particle-level jet and reconstructed jet, respectively. The average transverse momenta $\langle p_{\text{T,reco}} \rangle$ and $\langle p_{\text{T,ptcl}} \rangle$ are calculated over a large number of events in bins of particle-level jet p_{T} and reconstructed jet η. This derived response function is then applied as correction factor to jets in both data and simulation.

These corrections are also propagated to the missing transverse momentum through the following formula

$$\vec{p}^{\,\text{miss}}_{\text{T,corr}} = \vec{p}^{\,\text{miss}}_{\text{T,uncorr}} + \sum_i \vec{p}^{\,i}_{\text{T,uncorr}} - \sum_i \vec{p}^{\,i}_{\text{T,corr}} - \sum_i \vec{\mathcal{O}}^{i}_{\text{RC}} \tag{5.6}$$

where $p_{\text{T,corr}}$ and $p_{\text{T,uncorr}}$ are the corrected and uncorrected jets and i runs over all jets with $p_{\text{T,corr}} > 10\,\text{GeV}$. $\sum_i \vec{\mathcal{O}}^{i}_{\text{RC}}$ provides the average residual level-1 correction due to pileup and so that this does not bias the calculation of the missing transverse momentum.

5.6.3 Level-2 & level-3 residual corrections

Finally, residual corrections to match the jet response in simulation to the one in data are applied in two steps: first, the relative jet response as a function of η is measured in a data sample enriched in dijet events. One of these jets (the "tag" jet) has to be within $|\eta| < 1.3$ while the other one (the "probe" jet) does not have any η requirement. One can then derive corrections using two response functions, one based on the direct comparison of jet transverse momenta and the other one taking into account the hadronic activity in the whole event along with the missing transverse momentum (referred to as MPF method):

$$\begin{aligned} R^{p_{\text{T}}}_{\text{rel}} &= \frac{1 + \langle \mathcal{A} \rangle}{1 - \langle \mathcal{A} \rangle} \\ R^{\text{MPF}}_{\text{rel}} &= \frac{1 + \langle \mathcal{B} \rangle}{1 - \langle \mathcal{B} \rangle} \end{aligned} \tag{5.7}$$

with

$$\begin{aligned} \mathcal{A} &= \frac{p_{\mathrm{T,probe}} - p_{\mathrm{T,tag}}}{p_{\mathrm{T,probe}} + p_{\mathrm{T,tag}}} \\ \mathcal{B} &= \frac{\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \cdot (\vec{p}_{\mathrm{T,tag}}/p_{\mathrm{T,tag}})}{p_{\mathrm{T,probe}} + p_{\mathrm{T,tag}}} \end{aligned} \tag{5.8}$$

where $\langle \mathcal{A} \rangle$ is again measured over many events. From the ratio $R_{\mathrm{rel,MC}}/R_{\mathrm{rel,data}}$ one can then derive scale factors to correct jets in data and the associated missing transverse momentum.

Residual corrections for the absolute jet energy scale as a function of p_{T} are calculated from events containing $Z(\rightarrow e^+e^-/\mu^+\mu^-) + \text{jet}$ or $\gamma + \text{jet}$ processes, making use of the fact that the Z boson mass is well known and the energy measurements of electron, muon and photon candidates are much more precise than jet energy measurements. Therefore, the absolute energy of the associated jet can be calibrated by requiring that it is balanced against the measured energy of the produced boson candidate. As these measurements are usually statistically limited, the calculated energy corrections are extrapolated to jets with higher p_{T} by measuring multijet events so that the absolute jet energy scale is derived in a global fit of all these measurements.

The resulting scale factors for the relative and absolute residual jet energy corrections are shown in Fig. 5.3.

5.6.4 Jet energy resolution (JER)

The *jet energy resolution* (JER) is defined as the spread of the distribution of the ratio $p_{\mathrm{T,reco}}/p_{\mathrm{T,ptcl}}$. This distribution usually follows a Gaussian function, except for very high-p_{T} jets where inactive ECAL regions and particles punching through the HCAL lead to non-Gaussian tails. As this resolution in general differs between data and simulation and is usually worse in data, simulated jets are smeared using the relation

$$p_{\mathrm{T,smeared}} = p_{\mathrm{T,ptcl}} + c(\eta_{\mathrm{reco}}) \cdot (p_{\mathrm{T,reco}} - p_{\mathrm{T,ptcl}}) \tag{5.9}$$

where $c(\eta_{\mathrm{reco}})$ is an η-dependent scale factor that needs to be measured. This correction is also propagated to $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. This measurement is carried out in a similar way as in the derivation of the level-2 and level-3 residual corrections, measuring the dijet balance and the jet energy in $\gamma/Z + \text{jet}$ processes, only this time using the width of the corresponding response function and not its mean.

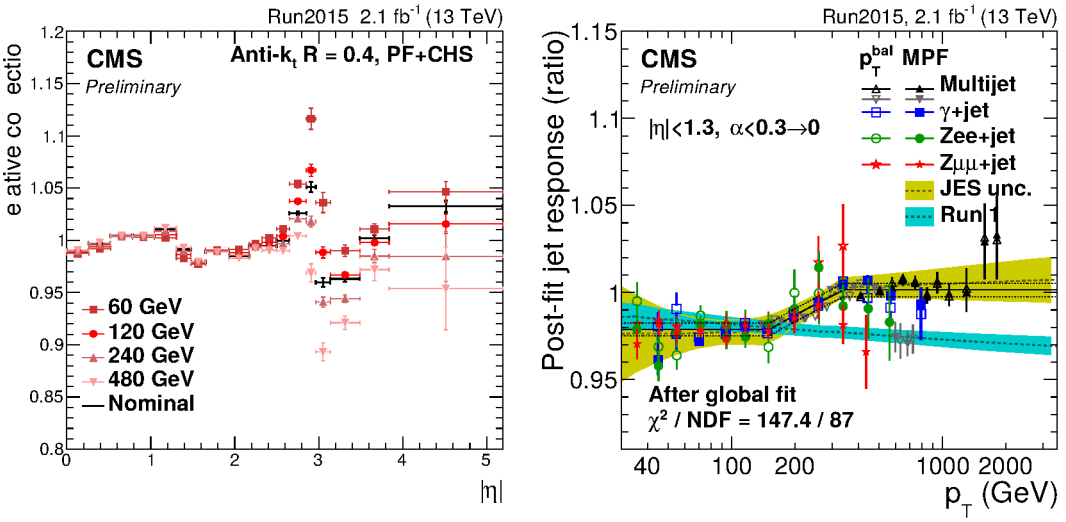

Figure 5.3: Scale factors to correct for differences between data and simulation measured in data collected in 2015 at $\sqrt{s} = 13\,\text{TeV}$. The scale factors for the relative correction as a function of $|\eta|$ are shown on the left for various values of jet p_T. The scale factors for the absolute correction, derived from different measurements, are shown on the right as a function of the p_T of the jet balanced against the reference particle. In the latter distribution, also the global fit to all measurements is shown compared to the fit derived in LHC Run I. Taken from Ref. [184].

5.7 Identification of jets originating from b quarks

5.7.1 The CSV algorithm

Jets originating from gluons or partons of different flavours leave a very similar similar signature and there are no immediate ways to distinguish between them (like there are in the PF algorithm to distinguish e. g. between electrons and photons). There are, however, subtle differences between jets from quarks of different flavours[2] that allow at least to determine the probability that they are produced from the fragmentation of a c or b quark. Such jets are referred to in the following as c or b *jets*, respectively. Identifying b jets is particularly useful as b quarks emerge in many interesting processes (e. g. in the decay of a top quark $t \to bW$). This is due to the fact that these quarks have a relatively long life time so that, depending on their p_T, they usually decay after a few millimeters after having hadronised to form a B meson (sometimes they can even reach the first pixel layer). This produces *displaced tracks* with respect to the primary interaction vertex and the excellent granularity of the pixel detector even allows to reconstruct the locations of the decay of the corresponding B mesons as *secondary vertices* (SV). This information can then be combined in so-called b *tagging algorithms* to help distinguishing b jets from c jets

[2]Techniques to distinguish between jets from quarks and gluons are discussed e. g. in Ref. [185].

and jets produced by light quarks or gluons (so-called *light jets*).

A b tagging algorithm which is widely used in the CMS Collaboration and also in the analyses presented in Chapters 6 and 7 is the so-called *combined secondary vertex* algorithm (CSV) which was developed for LHC Run I [186] and has been modified for LHC Run II [187]. It uses the information of all tracks in a jet and potentially reconstructed secondary vertices and provides a so-called *discriminator value* ranging from zero to one which usually takes high values for b jets and low values for light jets (c jets have values in between and are in general hard to distinguish from b jets). To be considered for the CSV algorithm, tracks must fulfill the following criteria [187]:

- they must have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$,
- the χ^2/n_{dof} returned by the fit in the tracking algorithm must be below 5,
- they must have at least eight hits in the inner tracking system, two of them in the pixel detector,
- they must have a transverse (longitudinal) impact parameter $|d_0|(|d_z|)$ smaller than 2 cm (17 cm),
- the distance between the jet axis and the track at their point of closest approach should be less than 0.07 cm,
- the decay length, defined as the distance between the primary vertex and the point of closest approach between the jet axis and the track, has to be smaller than 5 cm.

These requirements ensure that the considered tracks have a good reconstruction quality and are not too far from the primary interaction vertex e. g. originating from a pileup interaction.

To reconstruct secondary vertices inside a jet, an algorithm referred to as *inclusive secondary vertex finder* (IVF) [188] is used. It pre-selects tracks within a jet with similar requirements as the ones listed above with the main difference that the longitudinal impact parameter $|d_z|$ should be less than 3 cm (other requirements are slightly loosened). As a next steps, clusters of tracks are formed starting from so-called *seed tracks* that have a total impact parameter of at least 50 μm and a total impact parameter significance greater than 1.2[3]. The clustering is based on the minimum distance and the angle between the seed tracks and other tracks. The AVF algorithm is finally used to reconstruct secondary vertices from these track clusters. Secondary vertices are then removed that have a too large *flight distance significance* (FDS), which is defined as the distance between the primary and the secondary vertex divided by the uncertainty in the secondary vertex

[3]The total impact parameter is defined in this context as the distance between the track and the primary vertex at the point of closest approach.

position. If the total FDS is larger than 0.5 or if the FDS in the transverse plane is larger than 2.5, as it is then less likely that the vertex originates from the decay of a b quark. More cleaning steps are then applied based on e. g. the compatibility of tracks with the primary and secondary vertex until a final set of secondary vertices is obtained.

These tracks and secondary vertices are used as input to the CSV algorithm which performs more cleaning steps (e. g. removing tracks compatible with the decay of a K^0_S meson). Several variables related to these input objects (e. g. the flight distance of the secondary vertex or the impact parameter significances of the tracks) are fed into a neural network which provides a discriminator value for each jet. The distribution of this discriminator value for multijet events is shown in Fig. 5.4 where it can be seen that b jets tend to have large values in contrast to jets from other sources. One can then choose a minimum discriminator value above which a jet is labelled as *b-tagged* where the particular choice of the discriminator value allows to determine the selection efficiency for true b jets (also referred to as *b jet efficiency*) and the misidentification rate for jets from other sources. In the analysis presented in Chapter 6, two working points, labelled as *loose* and *medium* [187], are considered which are defined by setting this threshold to values of 0.46 and 0.8, respectively. The corresponding b jet efficiencies they provide are approximately 85% and 70% and the misidentification rates for light jets are approximately 10% and 1%.

5.7.2 B tagging efficiency measurement

The efficiency of the CSV algorithm to identify jets from different quark flavours or from gluons can differ between jets in data and simulation which needs to be taken into account. Corresponding scale factors have been derived in a measurement of the b tagging efficiency using data collected during the LHC data-taking period in 2015 at $\sqrt{s} = 13\,\mathrm{TeV}$ [187].

The b and c jet efficiencies (also referred to as *heavy-flavour efficiency*) have been measured in a data sample enriched in processes involving these quarks by requiring the presence of at least one jet with a muon within $\Delta R < 0.4$ of the jet axis. This is often the case for b and c jets as the corresponding B and D can produce muons in their decays. The b and c efficiencies are then evaluated from the ratio of the number of b-tagged jets containing a muon to the overall number of such jets. To improve the accuracy of this estimation, instead of just counting the b-tagged and untagged jets, a fitting procedure is employed, similar to the one described in Section 5.3 to derive the muon identification scale factor. The b and c efficiencies are then taken from the ratio of the fitted amplitudes of the components describing the b and c contributions before and after the b tagging requirement. The distributions used in the fit are discriminating variables of other b tagging algorithms, e. g. the transverse momentum of the muon relative to the jet axis or the discriminator value of the *jet probability* b tagging algorithm [186].

To evaluate the misidentification rate for light jets (*light-flavour efficiency*), the so-

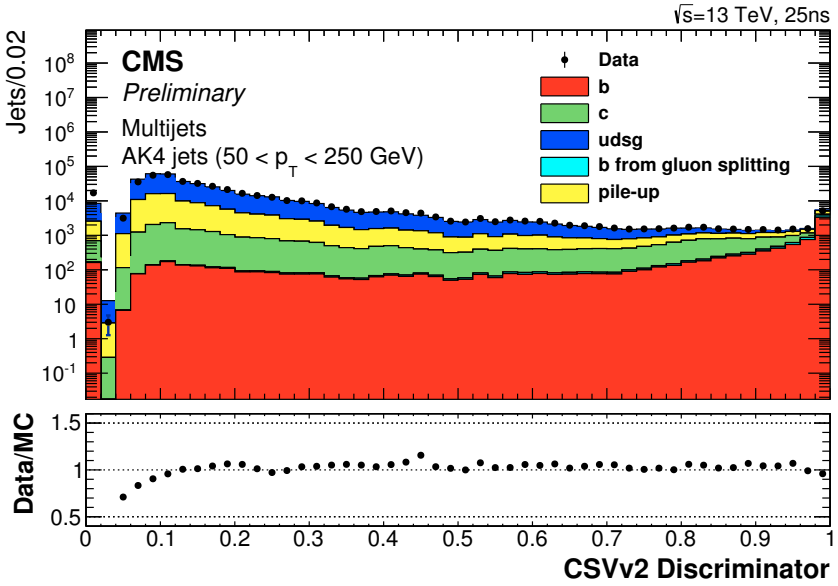

Figure 5.4: Discriminator value returned by the CSV algorithm, observed in an inclusive multijet data sample. The total number of entries in the simulation is normalised to the total data yield. Taken from Ref. [187].

called *negative tag* method [189] is employed which makes use of the *signed* track impact parameters and SV flight distance in a jet. A track is defined to have a *negative* impact parameter, if the scalar product between the jet direction and the vector pointing from the primary vertex to the track at the point of closest approach is negative. Similarly, the flight distance of a secondary vertex is assumed to be negative if the vector from the primary to the secondary vertex points into the opposite direction than the jet. While b and c jets usually produce tracks and secondary vertices for which these parameters are positive, light jets usually have displaced tracks and secondary vertices due to poor-quality tracks or combinatorial effects so that the corresponding parameters have positive and negative values in equal shares. The standard CSV algorithm only selects tracks and secondary vertices with a positive sign for their IP and SV flight direction to reduce the misidentification rate for light jets. Temporarily removing this requirement, however, allows to evaluate the misidentification rate by considering only jets with negative signs for the track impact parameter and SV flight distance parameter values.

Individual jet correction factors f for b, c, and light jets are then extracted from the previously described measurements. These correction factors range from 0.94 to 1 for b and c jets and from 1 to 1.2 for light jets. Scale factors to reweight events in simulation can then be derived from the b tagging efficiencies ϵ for b, c and light jets in simulation, which also depend on the their p_{T} and η, and the individual jet correction factors f with

the formula

$$SF = \frac{\prod_i f_i \epsilon_i \prod_j (1 - f_j \epsilon_j)}{\prod_i \epsilon_i \prod_j (1 - \epsilon_j)} \tag{5.10}$$

where i runs over all tagged jets and j over all untagged jets. These scale factors are applied to AK4 jets and subjets of AK8 jets (see Section 5.8) throughout the analyses presented in Chapters 6 and 7.

5.7.3 Nuclear interaction study

Light jets are often misidentified as b jets by the CSV algorithm if they contain displaced tracks or secondary vertices from any process other than the decay of a B meson. Such processes include inelastic scatterings of hadrons, which are mostly pions, with nucleons in the pixel detector material or the beam pipe. These scattering processes are referred to as *nuclear interactions*. This is a rather frequent process as the accumulated tracker thickness amounts up to 0.5 nuclear interaction lengths λ_I depending on the trajectory of the initial hadron. This can be seen in Fig. 5.5 on the left where the tracker material budget is shown in units of λ_I as a function of the η of a traversing pion. The resulting probability of a pion undergoing a nuclear interaction is shown in Fig. 5.5 as a function of the traversed tracker layers and can reach up to 25%.

Distributions of secondary vertices reconstructed with the IVF algorithm are shown in Fig. 5.6 for approximately 3×10^6 data events recorded during the 2012 data-taking period of the LHC at $\sqrt{s} = 8\,\text{TeV}$. These vertices were reconstructed using the same configuration as described at the beginning of Section 5.7.1, except for the requirement on the flight distance significance which is omitted. A large concentration of secondary vertices is seen close to the nominal interaction region which is mainly caused by decays of short-lived B and D mesons. At a larger radial distance from the beam axis, the material structures of the inner tracking system are visible, including the beam pipe at $r = 3\,\text{cm}$, the three pixel layers at 4 cm, 7 cm and 10 cm, and the pixel endcaps at $|z| = 34.5\,\text{cm}$ and $|z| = 46.5\,\text{cm}$. The distributions are shown after a coordination transformation in a coordinate system which is slightly tilted with respect to the nominal coordinate system described in Section 3.2.1. This is done in order to correct for an observed misalignment of the beam pipe with respect to the inner tracking system [190]. This new coordinate system is obtained by a transformation of the x- and y-coordinates given by $x' = (0.086\,\text{cm}) - 0.0007 \cdot z$ and $y' = (-0.197\,\text{cm}) + 0.0007 \cdot z$. The beam pipe appears as a single line in Fig. 5.6 after this coordinate transformation while the other parts of the inner detector are rather smeared out, indicating that they are not aligned with respect to each other.

The fraction of nuclear interactions in the secondary vertex distributions in Fig. 5.6 is not directly apparent, as they e. g. also contain photon conversions. Their contribution to these distributions is therefore estimated in simulation. In this study, a simulated

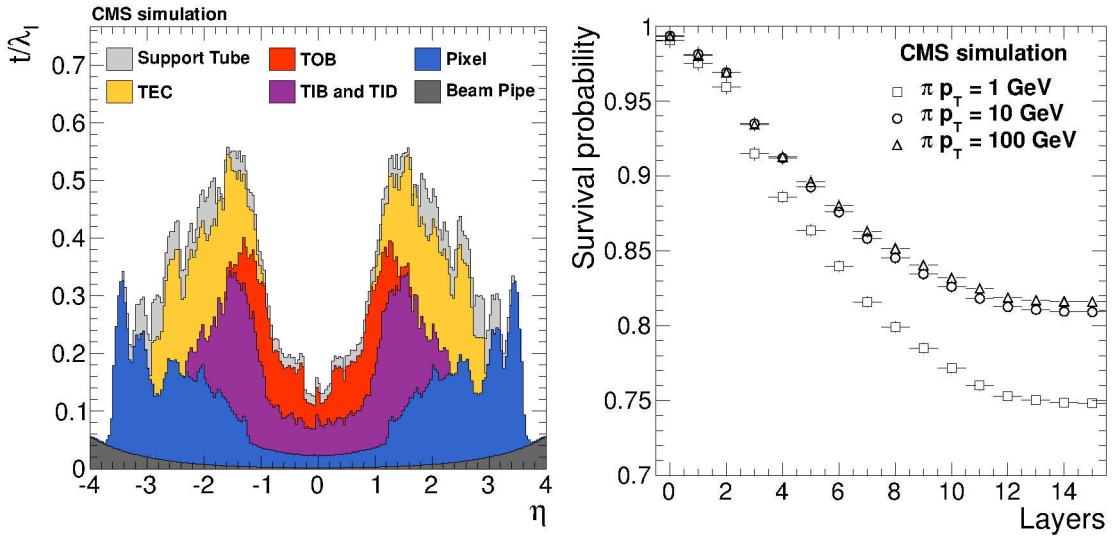

Figure 5.5: Material thickness of the pixel detector system in units of nuclear interaction lengths λ_I as a function of η (left) and probability of a pion undergoing a nuclear interaction as a function of the traversed layers of the tracker (right). Taken from Ref. [130].

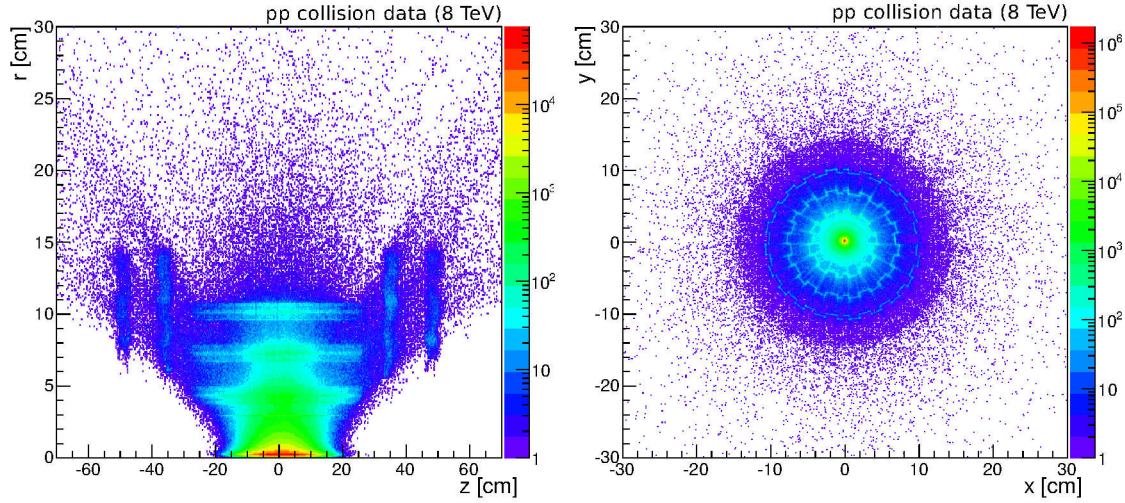

Figure 5.6: Distributions of secondary vertices reconstructed with the IVF algorithm for approximately 3×10^6 data events recorded during the 2012 data-taking period of the LHC at $\sqrt{s} = 8\,\text{TeV}$. These are shown in the (r,z)-plane on the left and in the (x,y)-plane on the right.

sample of approximately 3×10^6 $t\bar{t}$ + jets events is used at $\sqrt{s} = 8\,\text{TeV}$. The processes producing secondary vertices, which also include decays of unstable particles and photon conversions, are determined by first matching reconstructed to simulated tracks. This

matching is based on a comparison of the "true" positions, where a simulated particle traversed a layer of the inner tracker, and the resulting hit position on that layer as simulated by GEANT. Secondary vertices can also be caused by an accidental combination of two completely unrelated tracks which forms the so-called *combinatorial background*. Therefore, the fraction of each contributing process to a vertex is defined as the ratio of the sum of weights of its associated tracks to the sum of weights of all track contributing to the vertex. These weights are the ones determined during the fitting of the reconstructed vertex to mitigate the impact of extreme outliers to the vertex position (see Section 5.1.3). If the fraction of a given process is above 50%, the reconstructed vertex is assigned to this process.

The distribution of secondary vertices in the studied simulated sample is split into several contributing processes, as shown in Fig. 5.7. It can be seen that the majority of secondary vertices close to the nominal interaction point is caused by the decays of unstable particles such as B, D and K^0_S mesons, but also that above 2.5 cm, both nuclear interactions and photon conversions become more important. Vertices from the latter process can also be seen between the pixel layers. This is due to the fact that photon conversions are mostly reconstructed from only two tracks, resulting in a broad resolution of the reconstructed vertex along the flight direction of the initial photon.

An intriguing feature of this distribution is that close to the layers of the pixel detector, not only vertices from material interactions are accumulated but also vertices from decays of unstable particles. The reason for this effect is that the tracks from particles decaying close to a pixel layer do often not produce separate hit clusters but instead a single one. This so-called *cluster merging* then biases the reconstructed vertex position towards the pixel layer. This issue can, however, be resolved in a procedure referred to as *cluster splitting* by exploiting the direction of the jet associated to the vertex to recover separate pixel clusters.

To test the impact of nuclear interactions on the b jet efficiency and the light jet misidentification rate of the CSV algorithm, these parameters are measured in simulated multijet events at $\sqrt{s} = 13\,\text{TeV}$, in which tracks associated to nuclear interactions are excluded from the reconstruction. In a first step, such tracks are identified if they are used to reconstruct a secondary vertex located on either the beam pipe or one of the pixel layers. If they are located on a pixel layer, the tracks are re-fitted using the cluster splitting method previously mentioned. If the secondary vertex remains on the pixel layer or if it was located on the beam pipe in the first place, both the associated tracks and the secondary vertex itself are removed from the collection of tracks considered for the b tagging algorithm.

The performance of the CSV algorithm is measured in multijet events with a very large momentum transfer in the hard scattering process between 3 and 3.5 TeV. This results in a sample of events containing many jets with a high p_T, which is used since improvements

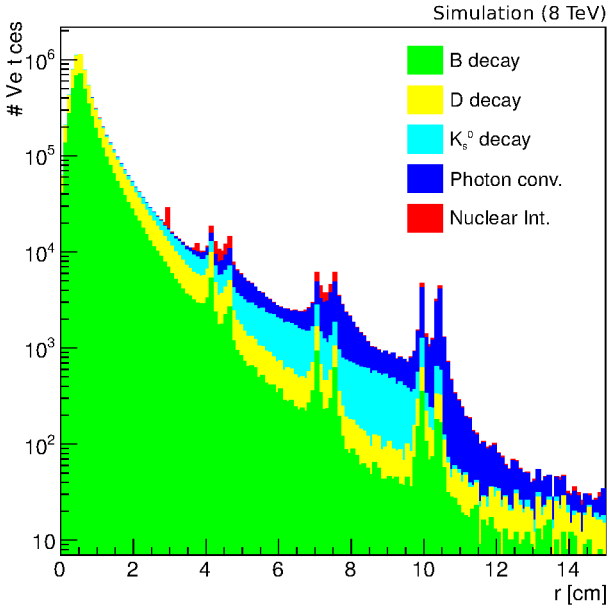

Figure 5.7: Distribution of the number of secondary vertices is shown as a function of r in simulated $t\bar{t}$ + jets events. The individual contributions from B, D and K^0_S meson decays as well as photon conversions and nuclear interactions are shown.

in b tagging from rejecting nuclear interactions are expected particularly in this highly boosted regime. The results are shown in Fig. 5.8, where the b tagging efficiency of the CSV algorithm is plotted against the misidentification rate for light jets[4]. It can be seen that rejecting nuclear interactions by removing tracks from secondary vertices close to the detector material does result in an improved performance of the CSV algorithm over a broad range of the b jet efficiency.

These results are already promising and future implementations of the CSV algorithm could well profit from a nuclear interaction rejection procedure, particularly in the regime of highly boosted b quarks. It is conceivable that such a method will soon be realised, possibly in combination with other techniques to improve the performance of b tagging algorithms in the high-p_T phase space (e. g. methods to identify meson decay chains containing b $\rightarrow$ c $\rightarrow$ s quark decays). This will be of particular importance for future searches for new particles, as new resonances are already excluded up to considerably high masses and b quarks produced in the decays of very massive particles are expected to have exceptionally high transverse momenta.

[4]Both the b jet efficiency and the light jet misidentification rate can be varied by changing the discriminator value threshold to tag jets as b jets.

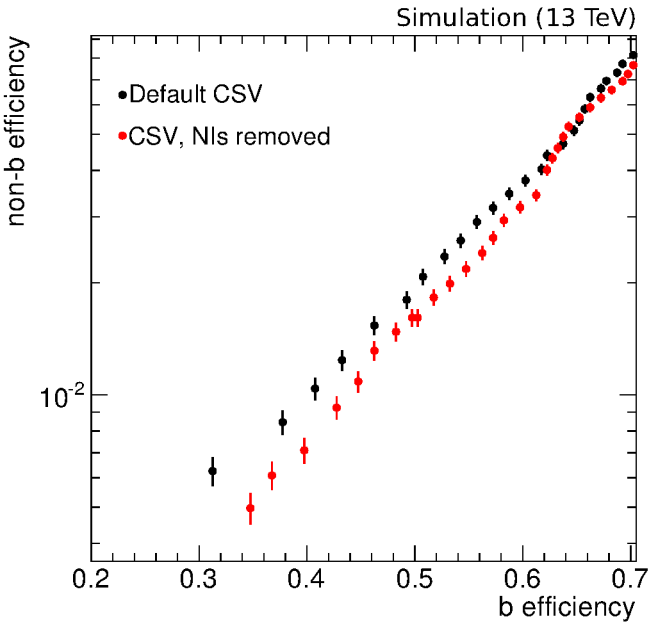

Figure 5.8: The light jet misidentification rate is shown as a function of the b jet efficiency for two different cases: black dots indicate the performance of the default CSV algorithm, red dots denote the CSV performance with cluster splitting applied and nuclear interaction tracks and vertices removed. The efficiencies are derived in multijet events with a very large momentum transfer in the hard scattering process between 3 and 3.5 TeV

5.8 Jet substructure identification

Heavy particles like top quarks, W, Z and Higgs bosons, when they decay exclusively into lighter quarks, can be identified by reconstructing these quarks in separate AK4 jets and then calculating the invariant mass of all these jets. This, however, becomes increasingly difficult as the more boosted the initial heavy particle is. In this case, the decay products of the heavy particle become more and more collinear so that they often cannot be reconstructed as two separate AK4 jets anymore. Instead, another strategy has been developed within the CMS Collaboration which is to reconstruct all decay products of a boosted heavy particle within a single jet (referred to as *heavy-particle jet*) and then look into the inner structure of this jet. These so-called *jet substructure techniques* were first employed during LHC run I at $\sqrt{s} = 8$ TeV [191,192] and since then have been refined for LHC run II [185,193].

The improvement of the reconstruction efficiency for a hadronically decaying heavy particles with large p_{T} by using single jets with a larger cone size can be seen in Fig. 5.9. This shows the efficiencies in simulated events to reconstruct the decay products of a hadronically decaying W boson for two cases, once where a single C/A jet is used with a

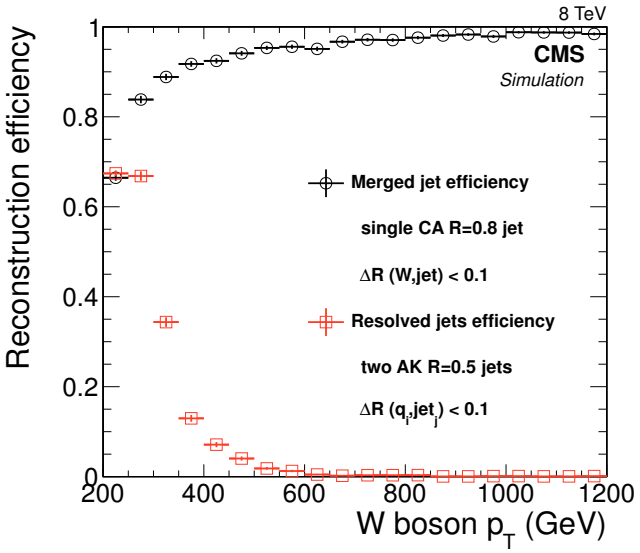

Figure 5.9: Efficiencies to find a reconstructed C/A jet with $R_C = 0.8$ within $\Delta R < 0.1$ of a simulated W boson (black circles) and to find two anti-k_T jets with $R_C = 0.5$ within $\Delta R < 0.1$ of the generated quarks from a W boson decay (red circles). Both efficiencies are shown as a function of the p_T of the generated W boson. Taken from Ref. [192].

cone size of $R_C = 0.8$ and once where two anti-k_T jets with $R_C = 0.5$ are used. It can be seen that for large p_T, in the former case, a single jet with a large cone size can be found in almost 100% of the cases close to the generated W boson while this probability drops to almost 0% if two jets with smaller cone sizes are used.

In order to capture all final state particles after the hadronisation of all these decay products, a cone size R_C larger than 0.4 is chosen for this jet. This cone size can vary depending on the exact strategy adopted, but in most analyses carried out at the CMS experiment (including the one in Chapter 6) it is set to 0.8 so that AK8 jets are reconstructed. Various so-called *grooming techniques* exist that aim at removing soft and wide-angle constituents within these jets so that ideally, only the hard decay products from the initial quarks from the heavy particle decay remain. This allows to use the *groomed jet mass* to discriminate between heavy-particle jets and jets from generic QCD processes (*QCD jets*) as before grooming, their jet mass is strongly influenced by soft radiation and can be quite large as well.

Grooming algorithms that have already been used in analyses at CMS at $\sqrt{s} = 8\,\mathrm{TeV}$ include filtering [194], trimming [195], and pruning [196, 197]. A relatively new algorithm which is used in the analysis presented in Chapter 6 and many other analyses at $\sqrt{s} = 13\,\mathrm{TeV}$ is referred to as *softdrop* algorithm [198]. This algorithm provides a good discrimination power between heavy-particle jets and QCD jets [185] and has the additional advantage that the resulting jet mass can be directly compared to analytic calcula-

tions [199,200]. It also divides jets into up to two so-called *subjets* which for heavy-particle jets very often correspond to the initial quarks from the heavy particle decay.

The algorithm starts by reclustering all constituents of an AK8 jet with the C/A algorithm to sort the order of clustering according to the angular separation of the jet constituents. To divide this jet into subjets, it is iteratively declustered by applying the following steps in each iteration:

1. Reverse the last step of the C/A clustering algorithm, splitting the original jet j into two subjets j_1 and j_2.

2. If j_1 and j_2 satisfy the condition

$$\frac{\min(p_{\mathrm{T},j_1}, p_{\mathrm{T},j_2})}{p_{\mathrm{T},j_1} + p_{\mathrm{T},j_2}} > z_{\mathrm{cut}} \left(\frac{\Delta R(j_1, j_2)}{R_C} \right)^{\beta} \tag{5.11}$$

   define j as the final softdrop jet with the subjets j_1 and j_2. R_C is the cone size parameter of the original jet and z_{cut} and β are pre-defined parameters corresponding to a soft threshold and an angular parameter, respectively.

3. Otherwise, define j as the subjet with the larger p_{T} and repeat the procedure.

4. If j cannot be further declustered, define it as final softdrop jet with itself being the sole subjet.

The parameter z_{cut} is set to 0.1 while β is set to zero, in which case the softdrop grooming algorithm is identical with the *modified mass-drop tagger* [199]. Using the C/A algorithm to recluster the jet allows to remove wide-angle radiation first while the softdrop condition Eq. (5.11) selects two subclusters which most of the time correspond to the hard decay products of the initial quarks. To reduce the impact of pileup interaction, the CHS method (see Section 5.1.3) is applied before performing the grooming algorithm outlined above.

One can now use the jet mass and further selection steps to define *tagging algorithms* that provide a high selection efficiency for processes involving boosted heavy particle decays and a good rejection power for generic multijet processes. Such selection steps include the use of further substructure variables such as:

N-subjettiness The N-subjettiness variable τ_N [201] quantifies the compatibility of a jet with an N-prong structure. Jets originating from the decay of a heavy particle into N daughter particles have low values for τ_N but large values for τ_M where $N \neq M$. The ratio τ_2/τ_1 is then sensitive e. g. to hadronic W boson decays where the two emitted quarks are reconstructed within one AK8 jet.

Subjet b tagging b tagging (see Section 5.7) can be applied to the subjets returned by the softdrop algorithm. This provides a high sensitivity e. g. for $H \to b\bar{b}$ decays or top quark decays.

Tagging algorithms making use of the above methods are used in the analyses presented in Chapter 6 and Chapter 7 and are discussed in detail there.

Jet mass scale and resolution measurement

To evaluate the performance of several grooming algorithms, they are applied to candidate jets for hadronic W boson decays in a data sample enriched in $t\bar{t}$ + jets events [185]. This is achieved by requiring a single lepton, at least one b-tagged AK4 jet, a missing transverse energy $\not{E}_{\mathrm{T}} > 40\,\mathrm{GeV}$ and at least one AK8 jet with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ in addition to further kinematic selection criteria. The CHS method is applied to the AK8 jet which is further groomed with the pruning algorithm. A final selection step is applied by requiring an N-subjettiness ratio for the AK8 jet $\tau_2/\tau_1 < 0.45$ which provides a data sample with a large fraction of $t\bar{t}$ + jets events in which the AK8 jet corresponds to a hadronically decaying W boson.

The jet mass distribution of the AK8 jets in events passing this selection can be found in Fig. 5.10. By fitting a multi-component function including a Gaussian term to this distribution, one can extract the jet mass scale and resolution in data and simulation and compute a scale factor to correct for resulting differences. This measurement was performed in a similar way also for AK8 jets groomed with the softdrop algorithm [202] and found to be consistent with the result obtained for pruned AK8 jets. Both the jet mass scale and resolution found to agree within 2% between data and simulation which is taken into account as systematic uncertainty as explained in Section 6.4.1. The jet mass resolution was found to be compatible between data and simulation within its uncertainty of 10.3% which is taken into account as systematic uncertainty as well.

5.9 Pileup reweighting

Many physical observables can have a dependency on the number of pileup interactions, even after performing steps to mitigate their contributions (e. g. by applying CHS or level-1 jet energy corrections). In order to avoid discrepancies between data and simulation due to differing pileup contributions, a procedure referred to as *pileup reweighting* is applied to match the simulated pileup distribution to the observed one.

The number of pileup interactions in a data event is estimated from the total inelastic cross section σ_{ie} of a proton-proton collision and the approximate instantaneous luminosity per bunch crossing which is averaged over time intervals of $\sim 20\,\mathrm{s}$. This has the advantage over simply comparing the number of reconstructed vertices in data and simulation that this calculation does not depend on the vertex reconstruction efficiency (which itself can depend on many factors that can differ in data and simulation). The cross section σ_{ie} is estimated to a value of 69 mb with a relative uncertainty of 5% while the instantaneous

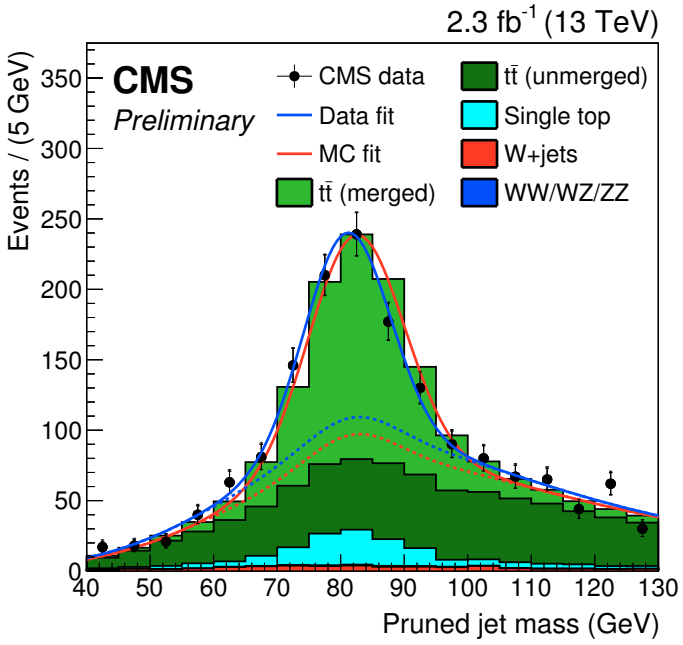

Figure 5.10: Jet mass distribution of AK8 jets passing the selection described in the main text. Solid curves indicate the fit performed to this distribution in data and simulation while the dashed curves show the corresponding fit components denoting the background components. Taken from Ref. [185].

luminosity per bunch crossing can be estimated using the pixel counting method (see Section 3.1).

The first step in the pileup reweighting procedure is then to produce a histogram reflecting an estimated distribution of the number of pileup events over a given data-taking period. This is done by calculating the expected number of pileup interactions per bunch crossing with the method described above and filling the histogram accordingly. A similar histogram can be produced using simulated events but this time filling in the actual number of "true" interactions per event which is naturally known in simulation. When finally performing an analysis and evaluating simulated events, weights can be calculated per event based on the number of "true" interactions in that event and the ratio of the corresponding bins in the two histograms one has produced earlier.

6 Single-lepton search for $T\bar{T} \to tH + X$

Vector-like quarks are motivated by a wide range of hypothetical extensions of the Standard Model (as discussed in Sections 2.3 and 2.4) and offer a very attractive test of such models. In particular, the vector-like partner of the top quark, the T quark, is expected to have a mass around 1 TeV if it is realised in an extension of the Standard Model which resolves the hierarchy problem. This would make its observation possible already with the first data set of pp collisions produced by the LHC in 2015 at $\sqrt{s} = 13\,\text{TeV}$.

This chapter presents a search for pair produced T quarks over a broad mass range with at least one boosted, hadronically decaying Higgs boson in the decay chain of the T quark and one charged lepton in the final state. In Section 6.1, the general analysis strategy is outlined. In Section 6.2, the data sets and simulated samples used to perform this search are introduced. The event selection and categorisation criteria are then presented in Section 6.3 and systematic uncertainties in the simulated samples are discussed in Section 6.4. In Section 6.5, the modelling of background processes is described and Section 6.6 finally contains a discussion and statistical interpretation of the selected data.

The results presented at the end of this chapter have been published by the CMS Collaboration and submitted to the Journal of High Energy Physics as part of the combined search for VLQ pair production presented in Chapter 7 [2]. Some of the figures in that reference are therefore shown in this chapter as well.

6.1 Analysis strategy and overview

Pair production of vector-like T quarks can be the dominant production mode up to very high T quark masses if the couplings of vector-like quarks to SM quarks are small and the single production cross section is low as a consequence (see also Section 2.4). In addition to this, pair produced T quarks with a high mass provide very distinct final states with a large number of high-p_{T} particles for which the expected contribution from SM processes is very low. This production mode is therefore well suited for a discovery of T quarks.

Though the main focus of this search is T quark pair production, the final result of this search is also interpreted for signal models predicting $B\bar{B}$ production, as such events can also result in final state signatures that this search is sensitive to (e. g. for $B\bar{B} \to bHtW$ processes).

Depending on the theoretical model in which T quarks appear and the corresponding branching fraction scenario, a sizable portion of T quarks can decay to a top quark and

a Higgs boson. If, for instance, the T quark is assumed to transform as a doublet under the electroweak gauge group, the branching fraction for the $T \to tH$ decay is 50%. This search is optimised for this decay mode by taking advantage of the characteristic signature of the resulting final state.

As vector-like quarks with masses below 700 GeV have already been excluded by searches at the LHC using the data collected at $\sqrt{s} = 8$ TeV (see also Section 2.4.2), this search only considers T quarks with masses above 700 GeV. In this mass regime, the decay products of the T quark usually have a large Lorentz boost in the $T\overline{T}$ system, resulting in transverse momenta greater than 300 GeV which often leads to the subsequent final state particles being very collimated. To maintain a high sensitivity for $T\overline{T}$ production at such high masses, dedicated analysis techniques are used that were particularly developed for such challenging event topologies.

One lepton is required in the event selection to reduce the background from multijet production which, as a consequence, constitutes only a minor fraction of the final background composition. This lepton often originates from the decay of the top quark produced in association with the Higgs boson. The lepton and the bottom quark from the top quark decay often become very collimated as can be seen in Fig. 6.1 in the left plot. This requires the choice of dedicated triggers and lepton identification criteria in order to maintain a high signal efficiency, which is explained in detail in Section 6.3.1.

The goal in this search is to identify hadronic Higgs bosons decaying to two b quarks as this is the decay mode of the Higgs boson with the largest branching fraction (see Fig. 2.2). The two b quarks from the Higgs boson decay can also become very collimated as is illustrated in the right plot in Fig. 6.1 which shows the distribution of the angular separation ΔR between the decay products of the Higgs boson. This separation often takes values below 0.8 even for the lowest considered mass point. In many cases, it is therefore not possible to reconstruct the b quarks in two separate AK4 jets and instead, they are attempted to be reconstructed in a single AK8 jet.

The softdrop algorithm (see Section 5.8) is used to identify subjets within an AK8 jet and reject events with high-p_{T} AK8 jets arising from other sources e. g. from a high-p_{T} gluon produced in a multijet event. Applying the CSV b tagging algorithm (see Section 5.7) to these subjets allows to identify the two b quarks with a high efficiency while rejecting a considerable portion of background processes. This whole procedure is referred to as H *tagging* and explained in more detail in Section 6.3.3.

Two variables are defined that provide a high sensitivity to $T\overline{T}$ events: H_{T}, which is the sum of transverse momenta of all AK4 jets with $p_{\mathrm{T}} > 30$ GeV in an event, and S_{T}, which is the sum of transverse momenta of the p_{T}-leading lepton, the missing transverse energy $\not{E}_{\mathrm{T}}$ and the H_{T} in an event. These variables usually take very large values for $T\overline{T}$ events as the two T quarks often produce final states with a large number of high-p_{T} particles.

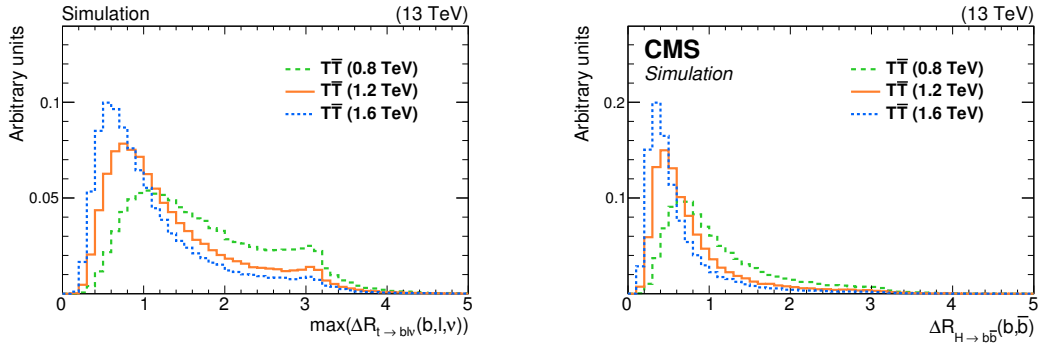

Figure 6.1: Distribution of the maximum ΔR between any two of the decay products of the top quark from a $T \to tH$ decay (left) and ΔR between the b quarks from a Higgs boson decay (right). The right plot is published in Ref. [2].

S_{T} is therefore used to discriminate signal-like from background-like events and for the final statistical interpretation of the data.

The main backgrounds in this search arise from $t\bar{t}$ + jets production with subdominant contributions from multijet and W + jets production. Single top, DY + jets and diboson production are considered as well but constitute only minor backgrounds after the final event selection. All background processes are estimated from MC simulations. The modelling of the main backgrounds from $t\bar{t}$ + jets and W + jets production is evaluated in independent control regions. Discrepancies between the data and simulation are observed in these regions and are corrected using per-event correction factors derived in these control regions. These correction factors are based on H_{T}, the exact procedure for this is explained in more detail in Section 6.5.2.

6.2 Data sets and simulated samples

This section contains information about the data sets and simulated samples that are used in this search. In Section 6.2.1, the data from proton-proton collisions that are analysed in this search are described and in Section 6.2.2, the simulated signal and background samples that are used to compare the data to are outlined.

6.2.1 Data sample

In this search, the full data set recorded with the CMS detector in 2015 is studied with a time interval between two proton bunches of on average 25 ns. Only events passing either a single-electron or a single-muon trigger are considered. The data samples consist of multiple so-called *runs* that correspond to continuous LHC operation cycles and are labelled alphabetically.

Only data events are considered in this search for which all CMS subdetectors and in particular the solenoid magnet were fully operational and none of its sub-systems were malfunctioning. This is ensured in a *data certification* procedure which includes several quality requirements both during the data-taking itself (*online monitoring*) and during the event reconstruction from the detector output (*offline quality management*) [203]. In addition to this, data events are rejected in which the beam spot position was not reconstructed correctly. This affects only events selected by an electron trigger in this search as the beam spot position is not used in the muon trigger. The integrated luminosity corresponding to the data collected with the single-electron trigger is therefore slightly lower than the one collected with the single-muon trigger.

After applying these quality filters, the collected data sets correspond to total integrated luminosities of $\mathcal{L} = 2.62\,\text{fb}^{-1}$ in the muon channel and $\mathcal{L} = 2.53\,\text{fb}^{-1}$ in the electron channel, respectively. The associated uncertainty of this measurement is 2.3% [126] and is propagated to the final result as systematic uncertainty which is explained in Section 6.4. All data sets studied in this search with their corresponding integrated luminosities are listed in Table 6.1.

However, it is possible that data events that are affected by signals not originating from the hard scattering process pass both the trigger selection and the data certification process. Dedicated event filters are therefore used to reject such events which include

- events with spurious signals in the detector caused by beam interactions with residual gas molecules (*beam halo*),
- events with "noise" patterns in the HCAL,
- events in which energy deposited by traversing particles could not be detected due to masked crystals in the ECAL,
- events without at least one reconstructed primary vertex (see Section 5.1.3).

6.2.2 Simulated event samples

Signal and background processes are generated using various MC event generators to simulate the matrix element (ME) of the hard scattering process (see also Section 4.1). W +jets (with $W \to \ell\nu$) and DY+jets (with DY $\to \ell\ell$ and $m(\ell\ell) > 50\,\text{GeV}$) processes are generated with MadGraph 5.2.2.2 [147] at LO. powheg v2 [143–146] is used to simulate $t\bar{t}$ +jets events at NLO order as well as single top quark events associated with a W boson. Single top quark production in the s- and t-channel is simulated with the MadGraph_aMC@NLO generator [148] at NLO. Pythia 8.212 [161, 162] is used for the simulation of W^-W^+, $W^\pm Z$ or ZZ events, which are grouped together in this search and denoted as *diboson* process. Pythia is also used to simulate multijet events produced via the

Data set	Integrated luminosity $\mathcal{L}$
Single electron, run C (Aug 2015)	17.27 pb^{-1}
Single electron, run D (Sep 2015 - Nov 2015)	2512.60 pb^{-1}
Single muon, run C (Aug 2015)	17.27 pb^{-1}
Single muon, run D (Sep 2015 - Nov 2015)	2603.40 pb^{-1}

Table 6.1: Data sets used in this search and their corresponding integrated luminosities.

strong interaction. These are generated inclusively in Pythia but filtered afterwards for certain processes that are more likely to pass the single-lepton selection in this search. This way, multiple samples are provided with the following properties: so-called *soft-lepton* samples contain events in which c or b quarks are produced that decay to either an electron or a muon in the process with a p_{T} of at least 10 and 5 GeV, respectively. *E.-M.-enriched* samples provide events with signatures that are likely to be misidentified as electrons in the reconstruction e. g. $\pi^0 \to \gamma\gamma$ processes in close proximity to charged hadrons. The production cross sections for these processes are calculated by multiplying the corresponding inclusive cross sections with the efficiency ϵ_f of the filters used to select such events.

In the particular phase space of this search, the finite size MC samples can become a problem if only a small number of simulated events pass the event selection, thus leading to an increased statistical uncertainty in the background prediction. Therefore, in order to increase the number of simulated background events in the phase space of this search, for some background processes, several samples are produced that correspond to "bins" of a certain kinematic variable. For the background from $t\bar{t}$ + jets production, for example, three different samples are used: an inclusive sample without any prior generator requirements, one sample in which the invariant mass of the two generated t quarks is required to be between 700 GeV and 1000 GeV and one sample in which the invariant mass of the $t\bar{t}$ system is at least 1000 GeV. For W +jets and DY+jets production, samples are used that are binned according to the generated H_{T}[1] per event. For multijet production, samples are used that are binned in $\hat{p}_{\text{T}}$ which is defined as the momentum transfer in the hard scattering process.

A summary of all simulated background samples along with the event generator used to produce them can be found in Table 6.2. These samples are listed together with the cross sections which were used to derive their normalisation. For some processes, these cross sections were computed in dedicated calculations, in which case the corresponding reference and the order of the matrix element calculation is provided in the table. The cross

[1]The generated H_{T} is defined as the sum of the transverse momenta of all partons from the matrix element calculation.

Data set name	σ/pb	ME generator
Single top, s-channel	3.36 (NLO) [205]	MadGraph_aMC@NLO (NLO)
Single top, t-channel	70.69 (NLO) [206, 207]	MadGraph_aMC@NLO (NLO)
Single top, ass. with W boson	71.2 (approx. NNLO) [205]	powheg v2 (NLO)
$t\bar{t}$ + jets, inclusive production	831.76 (NNLO) [82]	powheg v2 (NLO)
$t\bar{t}$ + jets, m($t\bar{t}$ + jets) $\in [700, 1000]$ GeV	76.605 (NNLO) [82]	powheg v2 (NLO)
$t\bar{t}$ + jets, m($t\bar{t}$ + jets) > 1000 GeV	20.578 (NNLO) [82]	powheg v2 (NLO)
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} \in [80, 120]$ GeV	2758420 * 0.03844 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} \in [120, 170]$ GeV	469797 * 0.05362 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} \in [170, 300]$ GeV	117989 * 0.07335 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} \in [300, 470]$ GeV	7820.25 * 0.10196 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} \in [470, 600]$ GeV	645.528 * 0.12242 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} \in [600, 800]$ GeV	187.109 * 0.13412 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} \in [800, 1000]$ GeV	32.3486 * 0.14552 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(μ)-enriched, $\hat{p}_{\mathrm{T}} > 1000$ GeV	10.4305 * 0.15544 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(e)-enriched, $\hat{p}_{\mathrm{T}} \in [80, 170]$ GeV	3221000 * 0.01183 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(e)-enriched, $\hat{p}_{\mathrm{T}} \in [170, 250]$ GeV	105771 * 0.02492 (ϵ_f)	Pythia 8.212
Multijet, Soft-lep.(e)-enriched, $\hat{p}_{\mathrm{T}} > 250$ GeV	21094.1 * 0.03375 (ϵ_f)	Pythia 8.212
Multijet, E.-M.-enriched, $\hat{p}_{\mathrm{T}} \in [80, 120]$ GeV	2800000 * 0.125 (ϵ_f)	Pythia 8.212
Multijet, E.-M.-enriched, $\hat{p}_{\mathrm{T}} \in [120, 170]$ GeV	477000 * 0.132 (ϵ_f)	Pythia 8.212
Multijet, E.-M.-enriched, $\hat{p}_{\mathrm{T}} \in [170, 300]$ GeV	114000 * 0.165 (ϵ_f)	Pythia 8.212
Multijet, E.-M.-enriched, $\hat{p}_{\mathrm{T}} > 300$ GeV	9000 * 0.15 (ϵ_f)	Pythia 8.212
$W(\to \ell\nu)$+ jets, gen. $H_{\mathrm{T}} \in [100, 200]$ GeV	1345 * 1.21 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$W(\to \ell\nu)$+ jets, gen. $H_{\mathrm{T}} \in [200, 400]$ GeV	359.7 * 1.21 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$W(\to \ell\nu)$+ jets, gen. $H_{\mathrm{T}} \in [400, 600]$ GeV	48.91 * 1.21 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$W(\to \ell\nu)$+ jets, gen. $H_{\mathrm{T}} \in [600, 800]$ GeV	12.05 * 1.21 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$W(\to \ell\nu)$+ jets, gen. $H_{\mathrm{T}} \in [800, 1200]$ GeV	5.501 * 1.21 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$W(\to \ell\nu)$+ jets, gen. $H_{\mathrm{T}} \in [1200, 2500]$ GeV	1.329 * 1.21 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$W(\to \ell\nu)$+ jets, gen. $H_{\mathrm{T}} > 2500$ GeV	0.03216 * 1.21 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$DY(\to \ell\ell)$+ jets*, gen. $H_{\mathrm{T}} \in [100, 200]$ GeV	147.4 * 1.23 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$DY(\to \ell\ell)$+ jets*, gen. $H_{\mathrm{T}} \in [200, 400]$ GeV	40.99 * 1.23 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$DY(\to \ell\ell)$+ jets*, gen. $H_{\mathrm{T}} \in [400, 600]$ GeV	5.678 * 1.23 (k-factor [204])	MadGraph 5.2.2.2 (LO)
$DY(\to \ell\ell)$+ jets*, gen. $H_{\mathrm{T}} > 600$ GeV	2.198 * 1.23 (k-factor [204])	MadGraph 5.2.2.2 (LO)
WW	118.7 (NNLO QCD) [208]	Pythia 8.212
ZZ	16.523 (NLO from MCFM) [209–211]	Pythia 8.212
WZ	47.13 (NLO from MCFM) [209–211]	Pythia 8.212

* With $m(\ell\ell) > 50$ GeV

Table 6.2: A list of all simulated background samples used in this search. The left column indicates the physical process and the middle column lists the corresponding cross sections to which the samples are normalised. If these are based on dedicated calculations and not taken from the MC generator, the corresponding reference is provided. For multijet events, the cross sections are calculated from the inclusive cross section for the respective $\hat{p}_{\mathrm{T}}$ bin and the filter efficiency ϵ_f to select the specific process. For $W(\to \ell\nu)$+ jets and $DY(\to \ell\ell)$+jets events, the LO cross sections returned by MadGraph are provided along with the k-factors to calculate the NLO cross sections.

sections of the remaining processes, for which no reference is provided, were calculated by the ME generator used to produce the corresponding samples. For the W + jets and DY+jets samples, so-called k-factors [204] are then used to calculate the NLO cross section from the LO cross sections returned by the ME generator.

The $T\overline{T}$ and $B\bar{B}$ signal samples are generated for VLQ masses between 700 GeV and 1800 GeV in steps of 100 GeV with the MadGraph event generator. Around 8×10^5 events were simulated per mass point of the T or B quark. The intrinsic branching fractions of the T quarks in the simulated signal samples are 33% for all three decay modes, $T \to bW$, $T \to tH$ and $T \to tZ$. Unless stated otherwise, the signal samples

are reweighted in plots and tables to correspond to a branching fraction of 100% for the $T \to tH$ decay mode in order to illustrate the sensitivity of this search to signal events containing such decay chains. This reweighting is also carried out to interpret the final result for different branching fraction scenarios of the $T\overline{T}$ or $B\bar{B}$ system.

For $T\overline{T}$ production (the procedure is equivalent for $B\bar{B}$ production), this reweighting is done by splitting each signal sample based on generator information into six different samples corresponding to the six possible final states of the $T\overline{T}$ decay chain ($tHtH$, $tHtZ$, $tHbW$, $tZtZ$, $tZbW$ and $bWbW$). These six samples are then re-combined with appropriate weights to obtain the correct event composition for the desired branching fraction assumption. For instance, to obtain a signal sample under the assumption of a branching fraction of 100% for the $T \to tH$ decay, only signal events in the $tHtH$ final state are used which are then scaled by a factor of $1/(0.33 \times 0.33)$.

Pythia is used to simulate parton showering, the hadronisation of quarks and gluons, and the decay of unstable particles for all simulated samples. It is also used to simulate the underlying event using the CUETP8M1 tune [212, 213]. For samples simulated with MadGraph at NLO precision (which include the single top samples in the s- or t-channel), the FxFx merging scheme [160] is used as matching scheme between MadGraph and Pythia, for all other samples generated with MadGraph, the MLM scheme [159] is used. The PDFs used in all samples to calculate the pp scattering cross section were calculated by the NNPDF collaboration [152].

6.3 Event selection

In this section, the event selection for this search is explained. Section 6.3.1 describes the trigger used in this search and the selection requirements for electrons and muons. The preselection is presented in Section 6.3.2. Its purpose is to reject a large fraction of background and to significantly reduce the size of the initial data set for the further analysis. H tagging, which is the main selection requirement to discriminate signal-like from background-like events, is explained in Section 6.3.3. The final event categorisation is summarised in Section 6.3.4, which increases the sensitivity of this search by splitting the remaining data sample after all selection steps into categories with different signal-to-background ratios.

6.3.1 Trigger and lepton selection

Electrons and muons[2] originating from top quark decays are often in close proximity to the nearest jet, as shown previously in Fig. 6.1. The lepton selection applied in this search

[2]In the following, whenever the term "lepton" is used, hadronic τ lepton decays are excluded from the definition and only electrons, muons and leptonic τ lepton decays are included.

is adapted to these final states in order to maintain a large selection efficiency for the $T\overline{T}$ signal. Both the trigger and the offline lepton selection have been optimised for a high signal efficiency in a dedicated study which is summarised in App. B.1.

On trigger level, the presence of at least one electron or muon candidate is required which is reconstructed by the HLT trigger (see Section 3.2.6) without any isolation requirements (see Section 5.2 for the definition of relative isolation). The muon trigger requires a minimum p_{T} of the muon candidate of 45 GeV and it must have $|\eta| < 2.1$. The electron trigger requires an electron candidate with $p_{\mathrm{T}} > 45\,\mathrm{GeV}$ and $|\eta| < 2.4$, and in addition two jets with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 50\,\mathrm{GeV}$, respectively. These additional jets are necessary to keep the trigger rate below the maximum allowed trigger rate since usually more electron candidates are reconstructed by misidentifying jets as electrons. The two triggers are also referred to in the following as μ trigger and e+jets trigger. Events in this search are split into a muon and an electron channel which are treated independently in the statistical analysis in Section 6.6. These are also referred to as μ channel and e channel in figures and tables. If lepton candidates of both flavours are present, so that an event is selected by both triggers, it is assigned to the muon channel.

The lepton candidate with the highest p_{T} in the event is referred to as *primary lepton* in the following. In case of a muon, the primary lepton must have $p_{\mathrm{T}} > 47\,\mathrm{GeV}$ and $|\eta| < 2.1$, and for an electron, $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ and $|\eta| < 2.5$ is required. These requirements ensure that the selected leptons are in a kinematic region where the trigger efficiency is nearly constant. Furthermore, muon candidates are required to pass the identification requirements described in Section 5.3 while electron candidates must meet the criteria described in Section 5.4.

As the lepton candidates are not required to be isolated from nearby hadronic activity, a considerable number of background events can result from multijet production in which a soft lepton is emitted e. g. in a b quark decay. It is possible to distinguish such processes from $t \to b\ell\nu$ processes and reject a significant fraction of multijet events using a combined selection on two variables referred to as "2D isolation requirement" [214]. The first variable is the angular separation $\Delta R(\ell, j)$ between the primary lepton ℓ and the closest jet j. The other variable, $\Delta p_{\mathrm{T}}(\ell, j)$, is the cross product of the primary lepton momentum and the unit vector pointing in the direction of the closest jet. For a lepton from a t quark decay, where the associated b quark is the closest jet in the majority of the cases, one of these variables usually becomes relatively large. Thus, for an event to pass the 2D isolation requirement, either $\Delta R(\ell, j) > 0.4$ or $\Delta p_{\mathrm{T}}(\ell, j) > 40$ GeV is required.

The efficiency for the trigger selection described above has been measured by the CMS Collaboration using a Tag-and-Probe method similar to the measurement of the lepton identification efficiency which is described in Section 5.3. Scale factors were derived in these measurements which are applied in this search to correct for observed differences of

the trigger efficiencies between data and simulation.

The efficiency of the e+jets trigger has been measured in a control sample enriched in events from $t\bar{t}$ + jets production by requiring exactly one electron and one muon candidate [215]. The "tag" events are selected by a muon trigger and very restrictive offline identification criteria are imposed on the electron candidate to select a data sample with only a negligible contribution from non-$t\bar{t}$ events. The trigger efficiency is then defined as the fraction of these events that also pass the e+jets trigger. The efficiency of the e+jets trigger was found to be approximately 96% and the ratio of the inclusive efficiencies in data and simulation $\epsilon_{data}/\epsilon_{MC}$ was measured to be 0.99 ± 0.02. No dependency on the p_{T} or $|\eta|$ of the electron candidate has been observed and this ratio is applied as a scale factor to all events in the electron channel in this search.

The efficiency of the muon trigger used in this search has been measured in a similar way, but in a data sample enriched in DY $\rightarrow \mu\bar{\mu}$ events. Events are selected by a single-muon trigger and by requiring one muon candidate with very restrictive identification criteria. The "probe" muon is then selected using the same identification criteria as described in Section 5.3. The efficiency of the tested muon trigger is then provided by the fraction of events in which the second muon would have been selected by this trigger. It has been found to be approximately 92% in data and approximately 95% in simulation with a slight dependency on the p_{T} and η of the muon. The corresponding scale factors applied to events in the muon channel in this search account for this p_{T} and η dependence.

The efficiency of the 2D isolation requirement has also been measured by selecting a sample enriched in DY $\rightarrow \ell\ell$ events where one of the two leptons has to fulfill very tight selection requirements [215]. The selection efficiency is then determined from the second lepton in the event. The 2D isolation requirement can be interpreted as a requirement on $\Delta p_{\mathrm{T}}(l,j) > 40$ for events with $\Delta R(l,j) < 0.4$. Therefore, the efficiencies of the $\Delta p_{\mathrm{T}}(l,j)$ selection are measured in different bins of $\Delta R(l,j)$ for values below 0.4. These efficiencies have been found to be compatible between data and simulation within the uncertainties of the measurement and no further scale factor is applied in this search.

6.3.2 Preselection

The trigger and lepton selection as well as the 2D isolation requirement described in Section 6.3.1 are combined with additional kinematic selection requirements in the so-called *preselection*. This significantly reduces the number of background events for the further analysis of the data.

$T\overline{T}$ events usually have a large number of AK4 and AK8 jets and therefore, at least two AK8 jets and at least three AK4 jets are required. All jets have to pass the identification criteria outlined in Section 5.5.

The large number of high-p_{T} AK4 jets in $T\overline{T}$ events, along with the high-p_{T} lepton

Variable	e channel	μ channel
p_{T} (prim. lepton)	$> 50.\,\mathrm{GeV}$	$> 47.\,\mathrm{GeV}$
p_{T} (1st AK4 jet)	$> 250.\,\mathrm{GeV}$	—
p_{T} (2nd AK4 jet) p_{T}	$> 70.\,\mathrm{GeV}$	—
2D isolation req.	$\Delta R(l,j) > 0.4$ or $\Delta p_{\mathrm{T}}(l,j) > 40\,\mathrm{GeV}$	
S_{T}	$> 800\,\mathrm{GeV}$	
N(AK8 jets)	≥ 2	
N(AK4 jets)	≥ 3	

Table 6.3: Summary of all selection requirements applied in the preselection.

and a large missing transverse energy arising from the associated neutrino, results in high values for S_{T} which is required to be $S_{\mathrm{T}} > 800\,\mathrm{GeV}$ in the preselection. A summary of all selection criteria in the preselection can be found in Table 6.3.

Distributions of all variables that are used in the preselection can be found in Figs. 6.2 to 6.6. In each of these distributions, all selection criteria listed in Table 6.3 are applied except for the requirement on the variable that is shown. The p_{T} spectra of the primary electron and muon can be found in Fig. 6.4. Distributions for $\Delta R(l,j)$ and $\Delta p_{\mathrm{T}}(l,j)$ can be found in Figs. 6.2 and 6.3 for the muon and electron channel, respectively, where two-dimensional distributions for the multijet background and a $T\overline{T}$ signal sample with a T quark mass of 800 GeV are shown in the top row. The projections of $\Delta R(l,j)$ and $\Delta p_{\mathrm{T}}(l,j)$ are displayed in the bottom row of Figs. 6.2 and 6.3. The large discrepancy between data and simulation in the rejected region below $\Delta R(\ell,j) < 0.4$ and $\Delta p_{\mathrm{T}}(\ell,j) < 40\,\mathrm{GeV}$, points at a poor modelling of the simulated multijet background which is the dominant background process in this region. This effect is accounted for by assigning a systematic uncertainty which is discussed in detail in Section 6.4.

N-1 distributions for the number of AK4 and AK8 jets can be found in Fig. 6.5. The S_{T} variable is shown in Fig. 6.6 in which trends are visible in the ratio between the data and simulated background. This ratio appears to linearly fall towards high S_{T} values. These trends are further scrutinised and discussed in detail in Section 6.5.

The number of remaining expected background and signal events after each selection step are shown in the cutflow tables in Tables 6.4 and 6.5. The signal yield is normalised to the theoretically predicted cross section for each mass point and the intrinsic branching fractions of the samples (33% to each decay mode bW, tZ an tH) are assumed. In total, 16,733 data events are left in the electron channel and 15,735 in the muon channel after the preselection. The expected total background yields are 18,550 and 18,571 events, respectively, while for a $T\overline{T}$ signal with a T quark mass of 800 GeV, 53 and 78 events are expected.

To verify the modelling of the number of pileup interactions in the simulation, the number of reconstructed vertices corresponding to pp interactions is shown in Fig. 6.7.

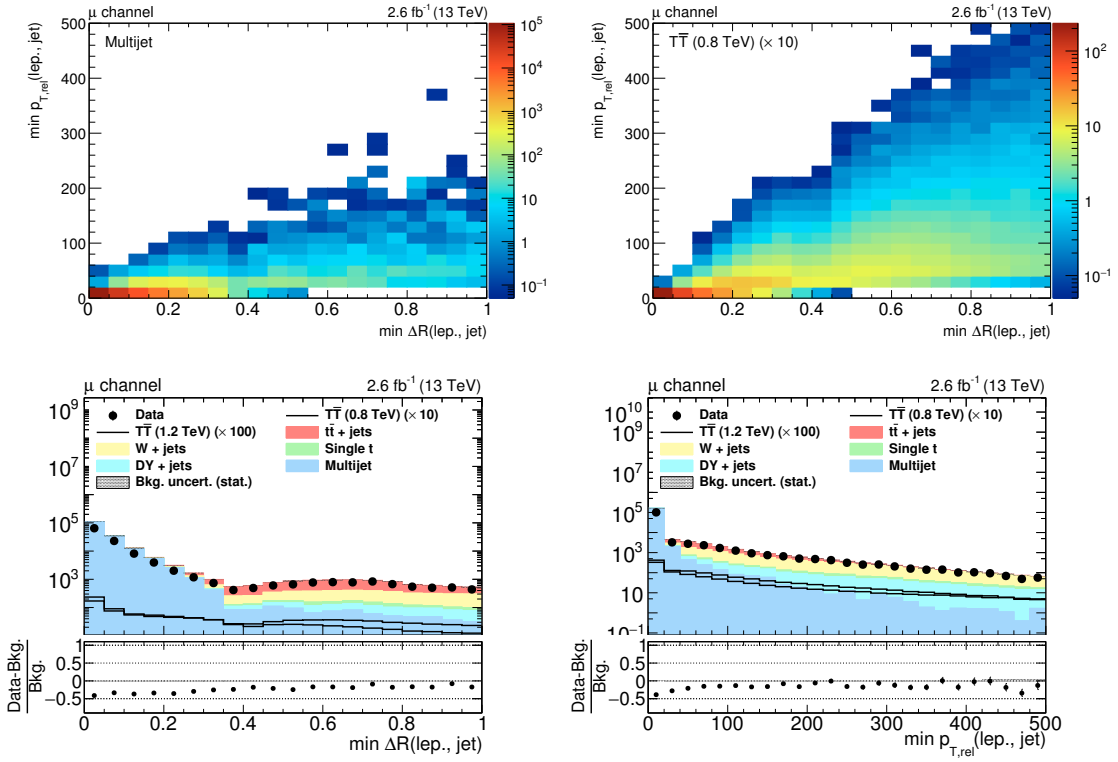

Figure 6.2: Two-dimensional distributions of $\Delta R(\ell, j)$ and $\Delta p_{\mathrm{T}}(\ell, j)$ for multijet (top left) and signal events (top right) for a T quark mass of 1200 GeV in the **muon channel**. All other selection criteria of the preselection are applied for each distribution (N-1). The bottom row shows the individual distributions of $\Delta R(\ell, j)$ (bottom left) and $\Delta p_{\mathrm{T}}(\ell, j)$ (bottom right) including data, before applying the 2D isolation requirement. The signal samples are normalised to their theoretically predicted cross section and the intrinsic branching fractions of the samples (33% to each decay mode bW, tZ an tH) are assumed.

The pileup reweighting procedure described in Section 5.9 is applied and the shapes of the distributions for data and simulation are in good agreement, though an offset in the normalisation of the simulation can be seen. This offset is related to the previously mentioned trends in the S_{T} distribution which are discussed in Section 6.5. Figure 6.8 depicts the number of recorded data events that pass the preselection over the entire data taking period. The average data yield is shown in intervals corresponding to integrated luminosities of approximately $50\,\mathrm{pb}^{-1}$. This yield is found to be approximately constant over the whole data-taking period, indicating a stable and good performance of CMS detector during the data acquisition.

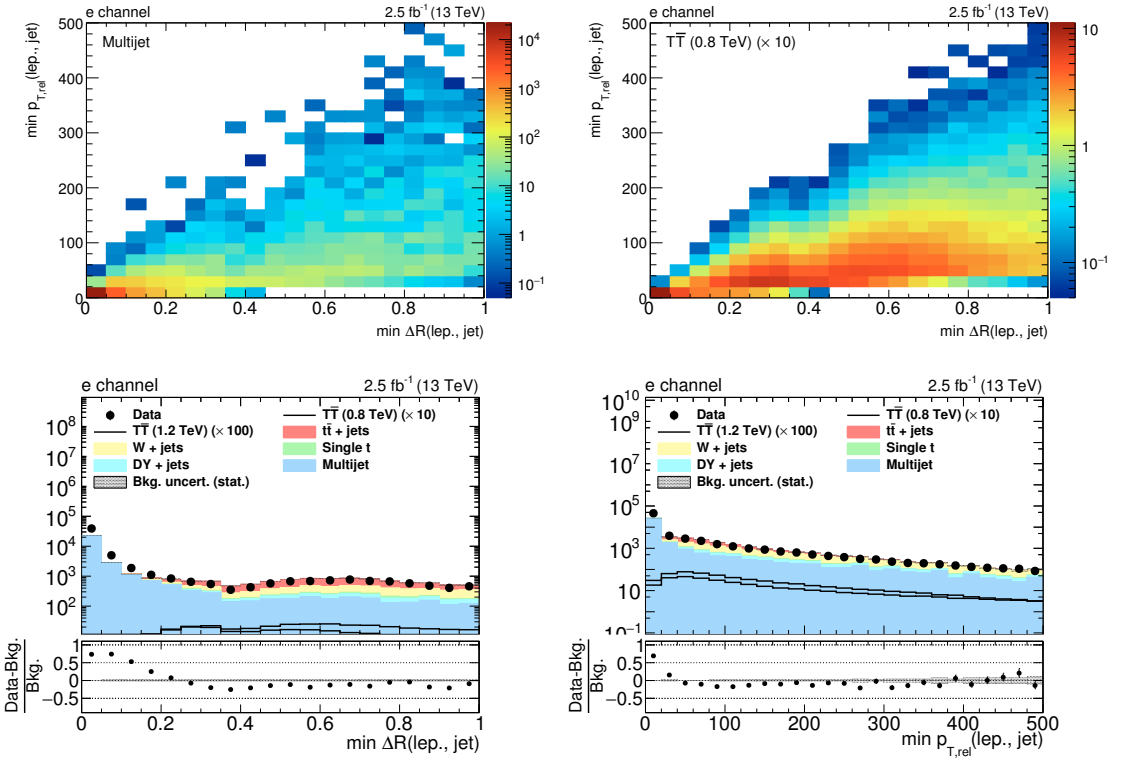

Figure 6.3: Two-dimensional distributions of $\Delta R(\ell, j)$ and $\Delta p_{\mathrm{T}}(\ell, j)$ for multijet (top left) and signal events (top right) for a T quark mass of 1200 GeV in the **electron channel**. All other selection criteria of the preselection are applied for each distribution (N-1). The bottom row shows the individual distributions of $\Delta R(\ell, j)$ (bottom left) and $\Delta p_{\mathrm{T}}(\ell, j)$ (bottom right) including data, before applying the 2D isolation requirement. The signal samples are normalised to their theoretically predicted cross section and the intrinsic branching fractions of the samples (33% for each decay mode) are assumed.

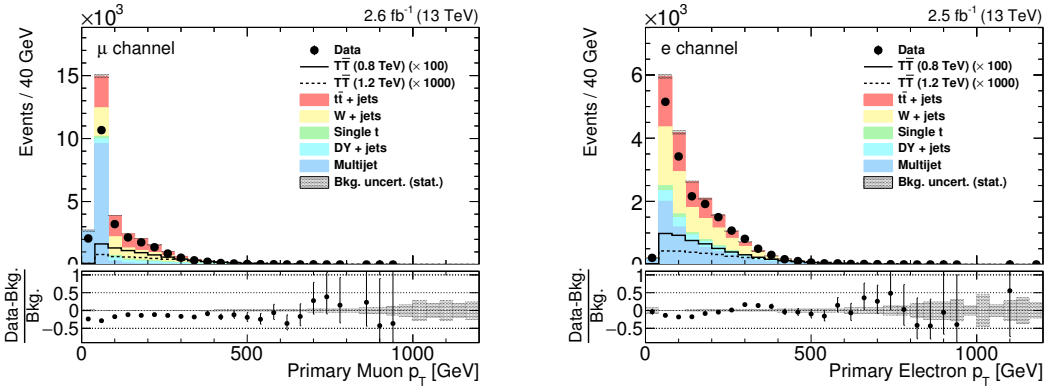

Figure 6.4: N-1 distribution of the p_{T} of the primary muon candidate (left) and of the primary electron candidate (right) before applying the $p_{\mathrm{T}}(\text{prim. lep.}) > 47/50\,\text{GeV}$ requirement. The signal samples are normalised to their theoretically predicted cross section and the intrinsic branching fractions of the samples (33% for each decay mode) are assumed.

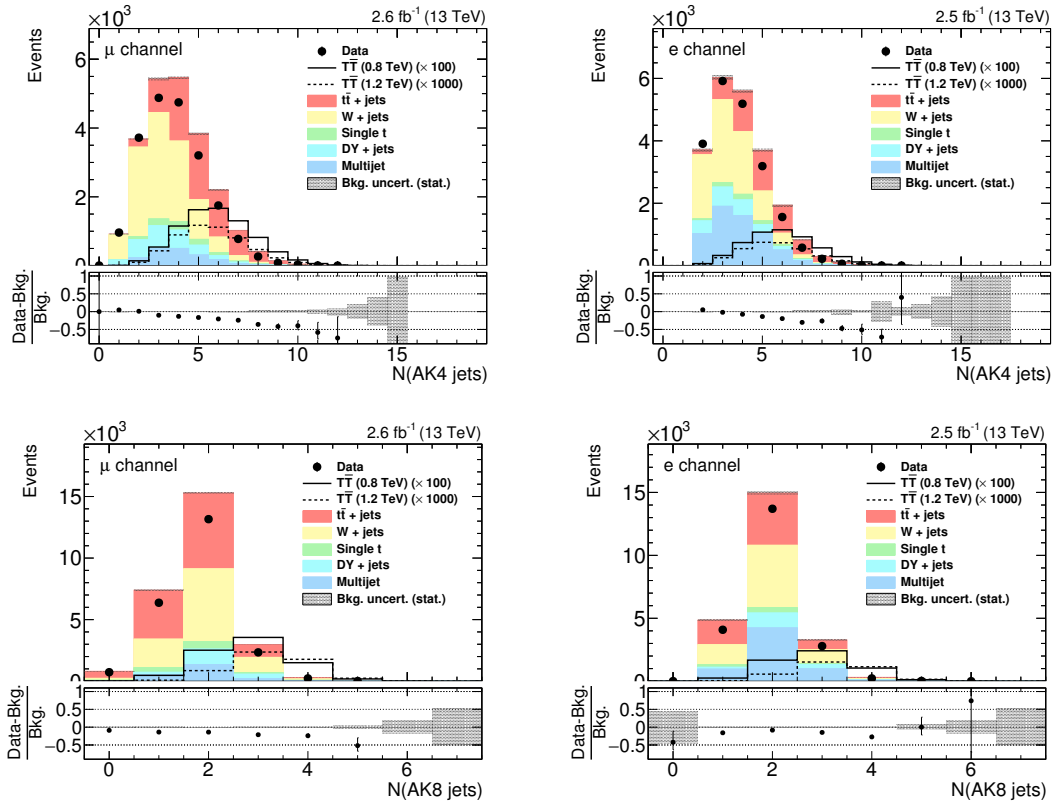

Figure 6.5: N-1 distributions of the number of AK4 jets (top row) before applying the N(AK4 jets) ≥ 3 requirement and AK8 jets (bottom row) before applying the N(AK8 jets) ≥ 2 requirement in the muon (left column) and electron (right column) channel. The signal samples are normalised to their theoretically predicted cross section and the intrinsic branching fractions of the samples (33% for each decay mode) are assumed.

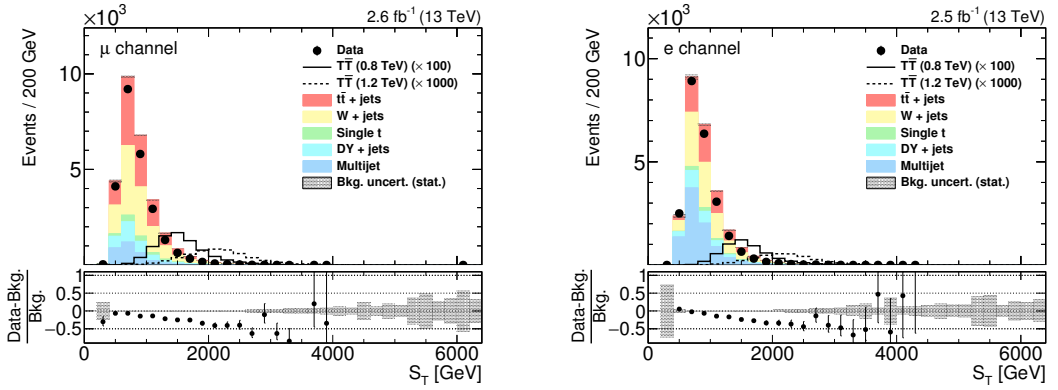

Figure 6.6: N-1 distribution of S_{T} in the muon (left) and electron (right) channel before applying the $S_{\mathrm{T}} > 800\,\mathrm{GeV}$ requirement. The signal samples are normalised to their theoretically predicted cross section and the intrinsic branching fractions of the samples (33% for each decay mode) are assumed.

	no sel.	Trigger	p_T(Electron)	2D iso. req.	p_T(1st AK4 jet)	p_T(2nd AK4 jet)	S_T	N(AK8 jets)	N(AK4 jets)	tot. eff.
$T\bar{T}$ (0.8 TeV)	519.0	76.5	65.1	58.0	56.2	56.0	55.9	53.26	52.59	0.1013
$T\bar{T}$ (1.2 TeV)	31.08	4.634	3.968	3.513	3.492	3.483	3.482	3.437	3.388	0.1090
DY + jets	645789.2	13497.3	11979.9	11509.5	7646.1	5040.9	2123.82	1953.85	1545.87	0.002394
Multijet	1544075449.1	747391.9	172104.2	48884.1	33528.8	25879.9	7269.3	6294.0	5271.8	0.00000341
Single top	390379.7	3156.8	2701.5	2524.5	1694.6	1460.7	822.7	600.33	533.55	0.00137
$t\bar{t}$ + jets	2162099.2	21702.0	18086.0	16016.5	11532.4	10903.7	7242.9	5203.7	5041.2	0.002332
W + jets	5606522.1	58183.4	49413.1	47253.8	34681.0	21341.3	10391.5	8211.0	6158.5	0.001098
Background Sum	1552880239.1	843931.5	254284.6	126188.4	89083.0	64626.6	27850.2	22262.9	18550.9	0.00001195

Table 6.4: Selection cutflow table for the **electron** channel. The first column shows event yields before any selection, the last column shows the total efficiency of the selection defined as the ratio between the second-to-last column and the first column. The signal yields are scaled to their theoretically predicted cross sections (see Table 2.1) and are shown for the intrinsic branching fractions of 33% to each decay mode.

	no sel.	Trigger	p_T(Muon)	2D iso. req.	S_T	N(AK8 jets)	N(AK4 jets)	tot. eff.
$T\bar{T}$ (0.8 TeV)	519.0	147.9	138.3	85.8	84.4	79.2	77.7	0.1496
$T\bar{T}$ (1.2 TeV)	31.08	10.50	9.64	5.466	5.453	5.363	5.259	0.1692
DY + jets	645789.2	130120.7	125485.4	122893.9	2584.66	2323.60	1655.00	0.002563
Multijet	1544075449.1	5920386.2	4719770.2	623527.9	2379.5	1805.0	1563.9	0.00000101
Single top	390379.7	38102.0	35319.4	33305.3	1299.7	842.0	743.4	0.00190
$t\bar{t}$ + jets	2162099.2	202989.2	189645.4	168528.3	12297.7	7600.2	7348.2	0.003399
W + jets	5606522.1	567866.4	536411.6	524883.5	14545.2	10600.0	7261.4	0.001295
Background Sum	1552880239.1	6859464.6	5606632.0	1473138.9	33106.7	23170.8	18571.9	0.00001196

Table 6.5: Selection cutflow table for the **muon** channel. The first column shows event yields before any selection, the last column shows the total efficiency of the selection defined as the ratio between the second-to-last column and the first column. The signal yields are scaled to their theoretically predicted cross sections and (see Table 2.1) are shown for the intrinsic branching fractions of 33% to each decay mode.

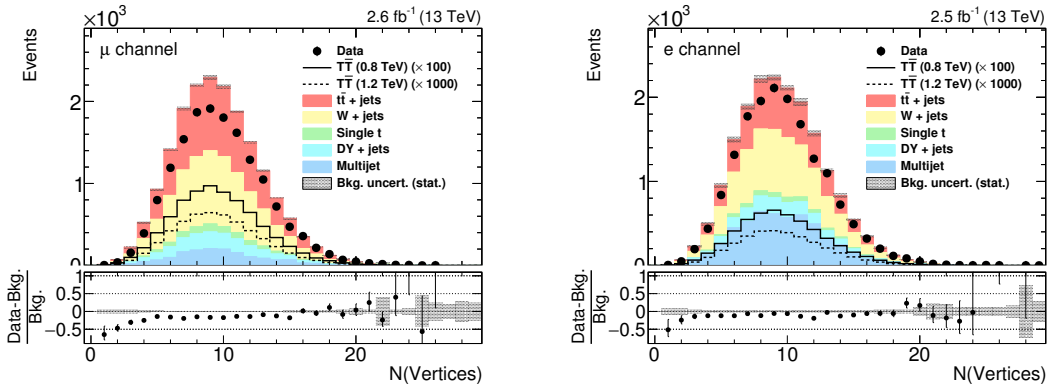

Figure 6.7: Number of reconstructed vertices corresponding to pp interactions in the muon (left) and electron (right) channels after the preselection. The signal samples are normalised to their theoretically predicted cross section and the intrinsic branching fractions of the samples (33% for each decay mode) are assumed.

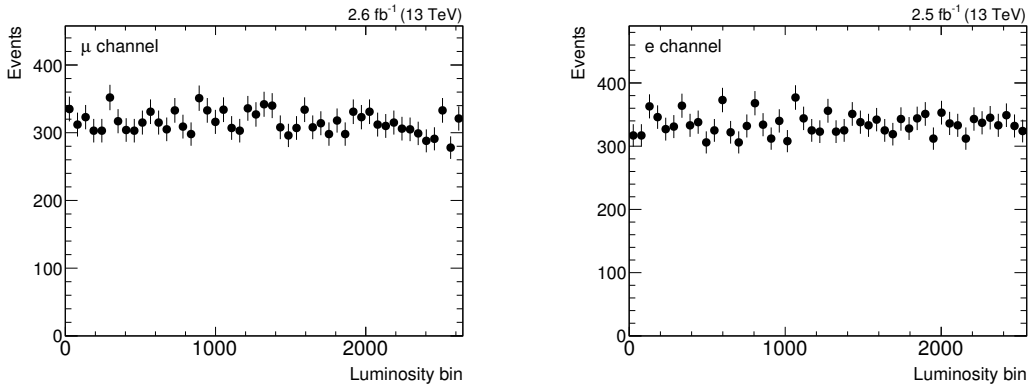

Figure 6.8: Number of data events passing the preselection per integrated luminosity interval, shown in bins of $\sim 50\,\mathrm{pb}^{-1}$.

6.3.3 H tagging

The most important selection technique in this search used to discriminate signal-like from background-like events, as mentioned in Section 6.1, is referred to as H *tagging*. Due to the large expected mass of the T quark, the Higgs boson from the T quark decay has a large Lorentz boost in the $T\overline{T}$ system. As a result, the subsequent b quarks from a $H \to b\bar{b}$ decay become very collimated and can often not be reconstructed in two separate AK4 jets anymore. The purpose of the H tagging algorithm is therefore to collect all decay products of the $H \to b\bar{b}$ decay in a single AK8 jet and then identify individual subjets within this jet corresponding to the two b quarks.

In order to remove constituents that are likely to originate from soft and wide-angle

radiation within the jet, the softdrop algorithm (see Section 5.8) is applied to all AK8 jets in an event. The AK8 jet mass after applying the softdrop algorithm (the *groomed* mass) is then close to the Higgs boson mass for actual $H \to b\bar{b}$ decays. The softdrop algorithm returns up to two subjets which correspond to these hard constituents in actual $H \to b\bar{b}$ decay processes.

So-called *Higgs boson candidate jets* are selected in a first step by requiring the groomed mass of the AK8 jet to be between 60 and 160 GeV and its p_{T} to be greater than 300 GeV. These two requirements already reject a considerable number of background events (e. g. from $gg \to b\bar{b}$ processes) while maintaining very high efficiencies of approximately 80% and 90%, respectively, for $H \to b\bar{b}$ decays where both b quarks are contained in one AK8 jet.

In the next step, the CSV b tagging algorithm (see Section 5.7) is applied to to the subjets returned by the softdrop algorithm, which constitutes the most important component of the H tagging algorithm to discriminate signal from background events. Requiring one of the subjets to be b-tagged greatly reduces the contribution from background processes without top quarks produced in the hard scattering process. In this way, however, a large portion of $t\bar{t}$ + jets events is selected in which only a part of the $t \to bqq'$ decay chain is merged into the AK8 jet as this results in a reconstructed mass lower than the mass of the top quark. Requiring two b-tagged subjets rejects a large portion of these events but also has a relatively low signal acceptance, depending on the chosen working point for the b tagging algorithm (see Section 5.7 for a definition of the considered working points).

A study is then performed to investigate which working point for the b tagging algorithm and which requirement for the number of b-tagged subjets (either one or two) maximises the sensitivity to $T \to tH$ events. This is done by comparing expected limits at 95% C.L. for a number of selection configurations after the preselection in which H-tagged AK8 jets with different subjet b-tagging requirements are compared. The p_{T} and mass requirements previously described are always included in the H-tagging definition in this study.

1 loose b tag At least one H tag is required where at least **one** subjet is b-tagged using the **loose** b-tagging working point.

1 medium b tag At least one H tag is required where at least **one** subjet is b-tagged using the **medium** b-tagging working point.

2 loose b tags At least one H tag is required where **two** subjets are b-tagged using the **loose** b-tagging working point.

2 medium b tags At least one H tag is required where **two** subjets are b-tagged using the **medium** b-tagging working point.

1 loose b tag + 2 loose b tags Events are split into two categories: in the first category,

at least one H-tagged AK8 jet is required where **two** subjets are b-tagged using the **loose** b-tagging working point. In the second category, at least one H-tagged AK8 jet is required with exactly one b-tagged subjet (with the same working point) and **no** H-tagged AK8 jet with two b-tagged subjets.

1 medium b tag + 2 medium b tags Events are split into two categories: in the first category, at least one H-tagged AK8 jet is required where **two** subjets are b-tagged using the **medium** b-tagging working point. In the second category, at least one H-tagged AK8 jet is required with exactly one b-tagged subjet (with the same working point) and **no** H-tagged AK8 jet with two b-tagged subjets.

Expected limits at 95% C.L. are then calculated using asymptotic CLs procedure (see App. A.3).In the limit calculation, normalisation uncertainties of 100% for multijet production and 20% for all other background processes are assigned. For the $T\overline{T}$ signal, a branching fraction of 100% for the $T \to tH$ decay is assumed.

The results of this study are shown in Fig. 6.9, comparing the expected upper cross section limits on $T\overline{T}$ production for T quark masses ranging from 700 to 1800 GeV. One can see that the "1 medium b tag + 2 medium b tags" configuration provides the lowest expected limit on th T quark mass. Consequently, this categorisation is chosen in the further analysis. An H-tagged jet is defined in the following as an AK8 jet that fulfills the mass and p_{T} criteria defined previously and has at least one b-tagged subjet.

Distributions of the variables that are used in the H tagging algorithm can be found in Figs. 6.10 to 6.12 for simulated events after the preselection. For each variable, the distribution is shown for the p_{T}-leading AK8 jet that satisfies all H tagging criteria except for the requirement on the displayed variable. Figure 6.10, for example, shows the p_{T} distributions of the p_{T}-leading AK8 jet with a groomed jet mass between 60 and 160 GeV and either at least one, or exactly two b-tagged subjets. In Figs. 6.11 to 6.12, the corresponding distributions for the groomed AK8 jet mass and the number of b-tagged subjets, in each case for the p_{T}-leading AK8 jet, are shown. The $T\overline{T}$ signal (for a mass of the T quark of 800 GeV) is shown for two cases: the solid curves shows $T\overline{T}$ events where at least one Higgs boson is present in the final state, the dashed curve shows all other $T\overline{T}$ events. The branching fraction assumed for the T quark is the intrinsic branching fraction in the signal samples, 33% to each decay mode.

In the distributions of the groomed jet mass of the p_{T}-leading AK8 jet in Fig. 6.11, it can be seen that signal events with $T \to tH$ decays exhibit a clear peak at 125 GeV. Signal events without a Higgs boson in the decay chain have a less pronounced peak at 90 GeV which is caused by events containing $Z \to b\bar{b}$ decays. The distribution of the number of b-tagged subjets of the p_{T}-leading AK8 jet with a groomed jet mass between 60 and 160 GeV in Fig. 6.12 shows the very high signal-to-background ratio achieved when requiring two b-tagged subjets. These distributions thus illustrate the discrimination power of the variables

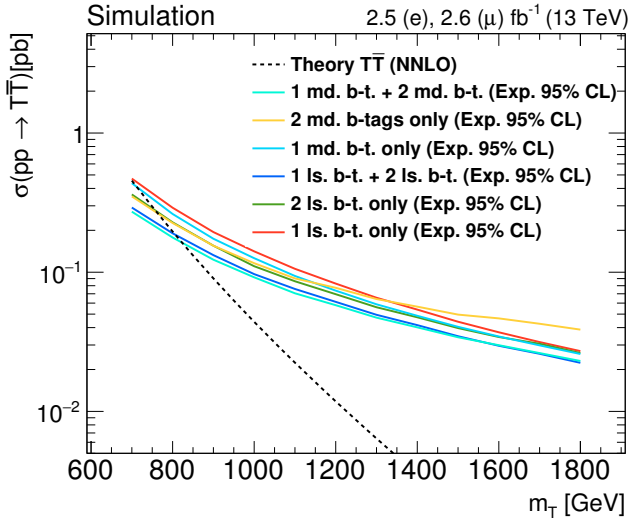

Figure 6.9: Expected upper cross section limits at 95% C.L. for various selection configurations (which are explained in the main text) with varying numbers of b-tagged subjets on the p_{T}-leading Higgs boson candidate and different working points for the b tagging algorithm. The limits are compared to the theoretical cross section for $T\overline{T}$ production. "ls. b.-t." stands for "loose b-tagged subjet" and "md. b.-t." stands for "medium b-tagged subjet".

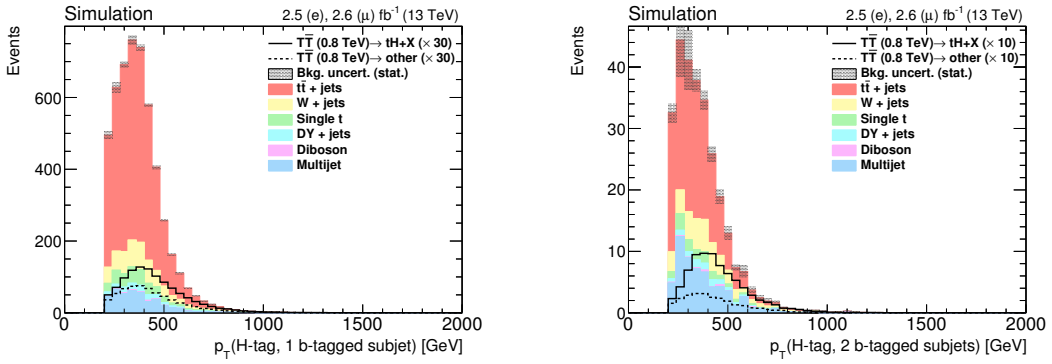

Figure 6.10: p_{T} distributions of the p_{T}-leading Higgs boson candidate jet with M_{jet} in [60, 160 GeV] and at least one (left) or exactly two (right) b-tagged subjets after combining muon and electron channel. The $T\overline{T}$ signal is normalised to the theoretical cross section. The intrinsic branching fractions of the signal samples of 33% to each decay mode are assumed.

used in the H tagging algorithm between $T \to tH$ and background processes.

The efficiency of the H tagging algorithm for Higgs bosons as a function of the generated Higgs boson p_{T} is shown in Fig. 6.13. The denominator for this efficiency is defined as all $H \to b\bar{b}$ decays in which the generated Higgs boson can be matched to a nearby AK8 jet. An AK8 jet and a generated Higgs boson are considered to be matched if they are within

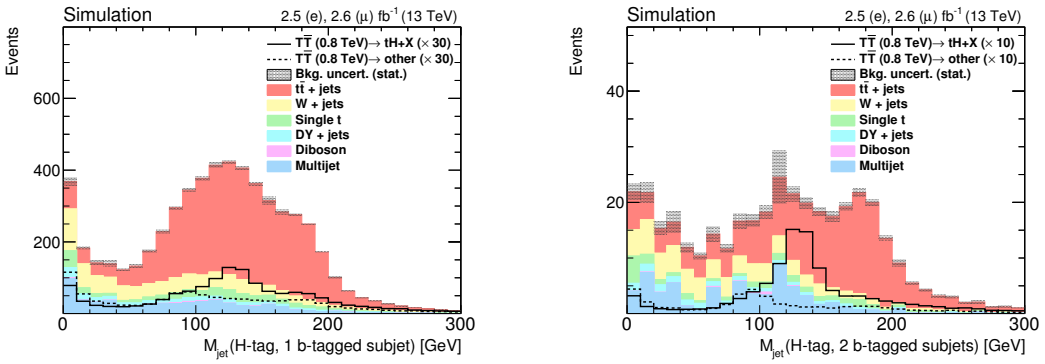

Figure 6.11: Distributions of the jet mass of the p_{T}-leading Higgs boson candidate jet with $p_{\mathrm{T}} > 300\,\mathrm{GeV}$ and at least one (left) or exactly two (right) b-tagged subjets after combining muon and electron channel. The $T\overline{T}$ signal is normalised to the theoretical cross section. The intrinsic branching fractions of the signal samples of 33% to each decay mode are assumed.

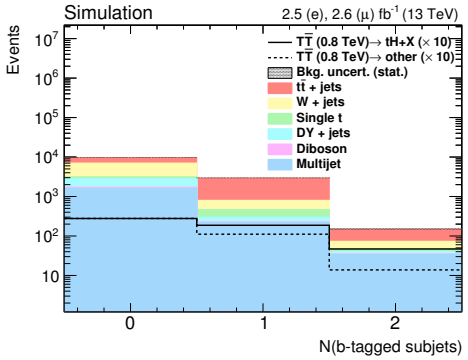

Figure 6.12: Number of b-tagged subjets of the p_{T}-leading Higgs boson candidate jet (M_{jet} in [60, 160 GeV], $p_{\mathrm{T}} > 300\,\mathrm{GeV}$) after combining muon and electron channel. The $T\overline{T}$ signal is normalised to the theoretical cross section. The intrinsic branching fractions of the signal samples of 33% to each decay mode are assumed.

$\Delta R < 0.8$ of each other and if both generated b quarks from the Higgs boson decay are contained within this AK8 jet (applying the same $\Delta R < 0.8$ requirement between the jet axis and each generated b quark). The decreasing efficiency towards high values for p_{T} can be explained with the decreasing b tagging efficiency for higher jet p_{T} values which is, amongst other things, caused by a deteriorating tracking efficiency for pion tracks with a very high p_{T} (see e. g. Ref. [130]).

H tagging identification efficiencies for various processes are listed in Table 6.6 and are determined in simulated signal and background samples after applying the preselection. The efficiencies for specific decay processes in this table are calculated with respect to processes in which the specified decay products are all contained within the AK8 jet that

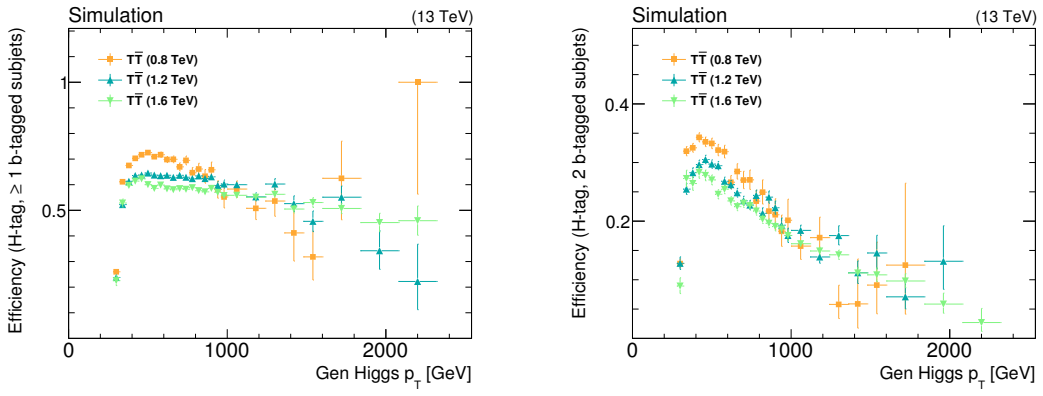

Figure 6.13: H tagging efficiencies as a function of the generated Higgs boson p_{T} for $H \to b\bar{b}$ decays in which both b quarks are merged into the H-tagged AK8 jet. Efficiencies are shown when requiring at least one (left) and exactly two (right) b-tagged subjets in the H-tagging algorithm. Events in the denominator have to pass the preselection described in Section 6.3.2.

is H-tagged (except for the processes that are marked as "inclusive"). The efficiencies of the H tagging algorithm in selecting $H \to b\bar{b}$ decays are approximately 60% and 25%, respectively, when requiring either at least one or exactly two b-tagged subjets. The corresponding misidentification efficiencies for the most important background process from hadronically decaying top quarks are approximately 21% and 1.3%, respectively. The high misidentification efficiency when requiring ≥ 1 b-tagged subjet is caused by top quark decay processes in which only two of the decay products of the top quark are contained within the AK8 jet, one of which being the b quark. This leads to the jet having an invariant mass within the range selected by the H tagging algorithm. Top quark decay processes that are selected by the H-tagging algorithm when requiring two b-tagged subjets have one subjet wrongly identified as b jet which explains the rather low misidentification rate in this case.

It can also be seen that the misidentification rate for processes involving two b quarks (e. g. $Z \to b\bar{b}$ or $g \to b\bar{b}$ decays) is relatively large. This is due to the fact that these efficiencies were measured after the preselection i. e. in a sample of events in which a considerable portion of processes entering the denominator of the efficiency calculation already have a large p_{T}. This results in a relatively large number of such events passing the p_{T} and mass requirements in the H tagging algorithm and explains the high misidentification efficiencies (in particular for $g \to b\bar{b}$ splitting processes). These are, however, relatively rare processes so that the overall contribution from events containing such processes is rather low.

Process	H tag eff. (≥ 1 b tag)	H tag eff. (2 b tags)
$H \to b\bar{b}$	60%	25%
$t \to bq\bar{q}'$ (inclusive)	21%	1.2%
$t \to bq\bar{q}'$ (fully merged)	20%	1.4%
$t \to bq\bar{q}'$ (b + q in jet)	32%	2%
$W \to$ incl.	4%	0.1%
$W \to cs$	6%	0.2%
$W \to ud$	1.5%	$< 0.1\%$
$Z \to b\bar{b}$	55%	24%
$Z \to q\bar{q}$	8%	0.5%
$g \to$ incl.	7%	0.4%
$g \to b\bar{b}$	40%	13%

Table 6.6: H tagging efficiencies for various decay processes. The mass and p_{T} requirements as described in the main text are applied in the H tagging algorithm. Events in the denominator have to pass the preselection described in Section 6.3.2.

6.3.4 Event categorisation

Following the strategy outlined in Section 6.3.3, events passing the preselection are split into three exclusive categories:

- The `H2b` category contains events with at least one H-tagged AK8 jet with exactly two b-tagged subjets.
- The `H1b` category contains events with at least one H-tagged AK8 jet with exactly one b-tagged subjet and no H-tagged AK8 jet with two b-tagged subjets.
- The `0H` category contains events without a H-tagged AK8 jet and with $\not\!\!E_{\mathrm{T}} > 100\,\mathrm{GeV}$.

The $\not\!\!E_{\mathrm{T}}$ requirement in the `0H` category is necessary to reject background events from multijet production which would otherwise constitute a considerable fraction of events in this category. The `0H` category can contain events which fail any of the H tagging criteria i. e. the p_{T} requirement, the mass requirement or the requirement of at least one b-tagged subjet. Distributions of S_{T} for simulated signal and background processes can be found in Fig. 6.14.

The `H2b` category has the highest signal-to-background ratio and is the primary source of the sensitivity in this search. The dominant background process in this category is $t\bar{t}$ +jets production with events containing a hadronically decaying top quark misidentified as a Higgs boson by the H-tagging algorithm. Events from single top, W +jets and DY+ jets production constitute minor contributions in both the `H1b` and `H2b` categories. Additional jets produced in association with the gauge bosons are misidentified by the H tagging

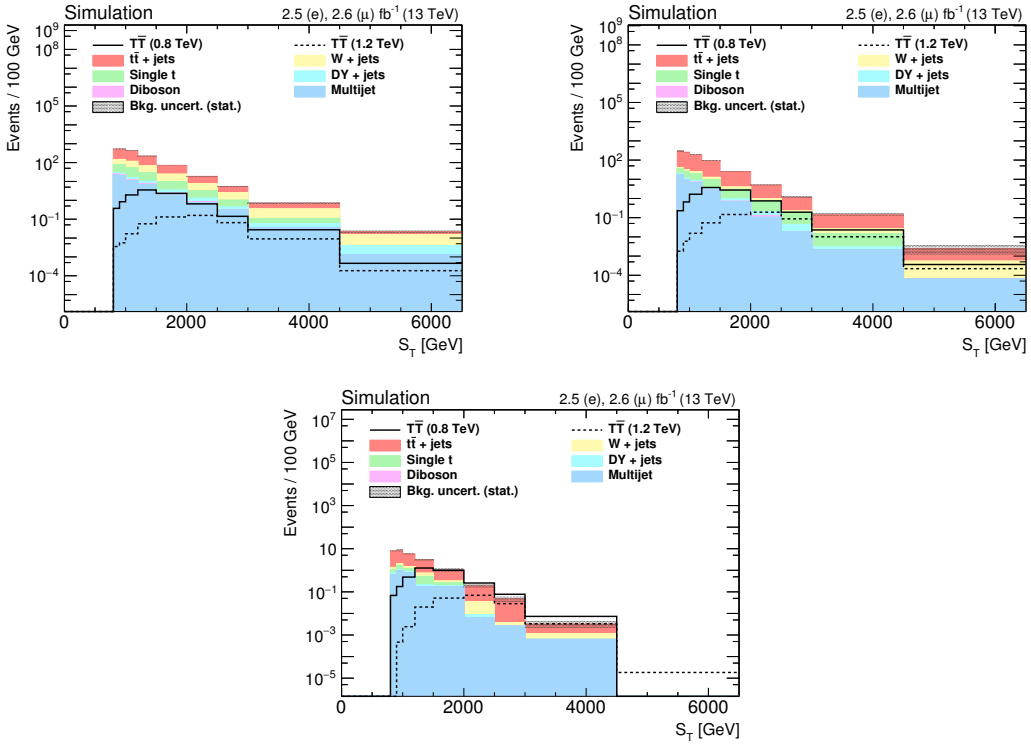

Figure 6.14: Distributions of S_T in the `0H` (top left), `H1b` (top right) and `H2b` (bottom) categories for simulated signal and background events. The $T\overline{T}$ signal is normalised to the corresponding theoretical cross section, assuming a branching fraction of 100% for the $T \to tH$ decay.

algorithm in such processes. A significant fraction of these events are selected due to $g \to b\bar{b}$ processes misidentified by the H tagging algorithm. This is also the main cause of selected multijet events in which the reconstructed leptons either originate from leptonic b or c quark decays that pass the 2D isolation requirement or arise from jets misidentified as leptons, which occurs more frequently in the electron than in the muon channel.

In the `0H` category, the background contributions from $t\bar{t}$ +jets and W +jets production have an approximately equal share. Assuming a branching fraction of 100% for the $T \to tH$ decay mode, the signal-to-background ratio in the `H2b` category is approximately 10 times higher than in the `H1b` category and approximately 30 times higher than in the `0H` category. The `H2b` category and the high-S_T region of the `H1b` category are therefore considered as signal regions while the `0H` category and the low-S_T region of the `H1b` category can be used to fix the normalisation of the background rates and constrain systematic uncertainties in the statistical interpretation of the data.

To further maximise the sensitivity of the event selection for $T\overline{T}$ production, the event selection is optimised by varying several kinematic requirements and evaluating their impact on the expected sensitivity. The details of this study can be found in App. B.2.

Requirement/category	H2b	H1b	0H
N(b-tagged AK4 jets)	≥ 1	≥ 1	≥ 1
N(H tags, 1 b)	≥ 1	$== 0$	$== 0$
N(H tags, 2 b)	—	≥ 1	$== 0$
$\not{E}_{\mathrm{T}}$	—	—	> 100 GeV

Table 6.7: Selection criteria applied in the three event categories after the preselection.

Concluding from this, an additional b-tagged AK4 jet is required in all three categories. This b-tagged AK4 jet must be separated from the nearest H-tagged AK8 jet (if one is present in the event) by requiring a $\Delta R > 0.8$ between the two jets. This is necessary as AK4 jets and AK8 jets are reconstructed independently of each other and therefore frequently overlap so that e. g. the same $t \to bW$ process can result in both a b-tagged AK4 jet and an H-tagged AK8 jet.

The final event selection is then defined by first requiring events passing the preselection to have at least one b-tagged AK4 jet, which has to have a ΔR to the nearest H-tagged AK8 jet of at least 0.8 if one is present. Events that pass this selection are then assigned to one of the three event categories based on the selection requirements summarised in Table 6.7.

Distributions of the number of b-tagged AK4 jets (with $\Delta R > 0.8$ to the nearest H-tagged AK8 jet) and the number of H-tagged AK8 jets with either at least one or exactly two b-tagged subjets after the preselection can be found in Fig. 6.15.

Signal efficiencies for all possible decays of the $T\overline{T}$ system in the three categories are listed in Table 6.8. They are shown for two $T\overline{T}$ signal samples with a high and low T quark mass, respectively. The denominator in this efficiency is defined as the number of signal events with the corresponding $T\overline{T}$ decay before the preselection and the numerator is given by the number of these signal events selected in the three event categories. It can be seen that the selection efficiency in the H2b category is highest for $T\bar{T} \to tHtH$ events and lowest for $T\bar{T} \to bWbW$ while the opposite is the case in the 0H category, thus illustrating a good performance of the H tagging algorithm.

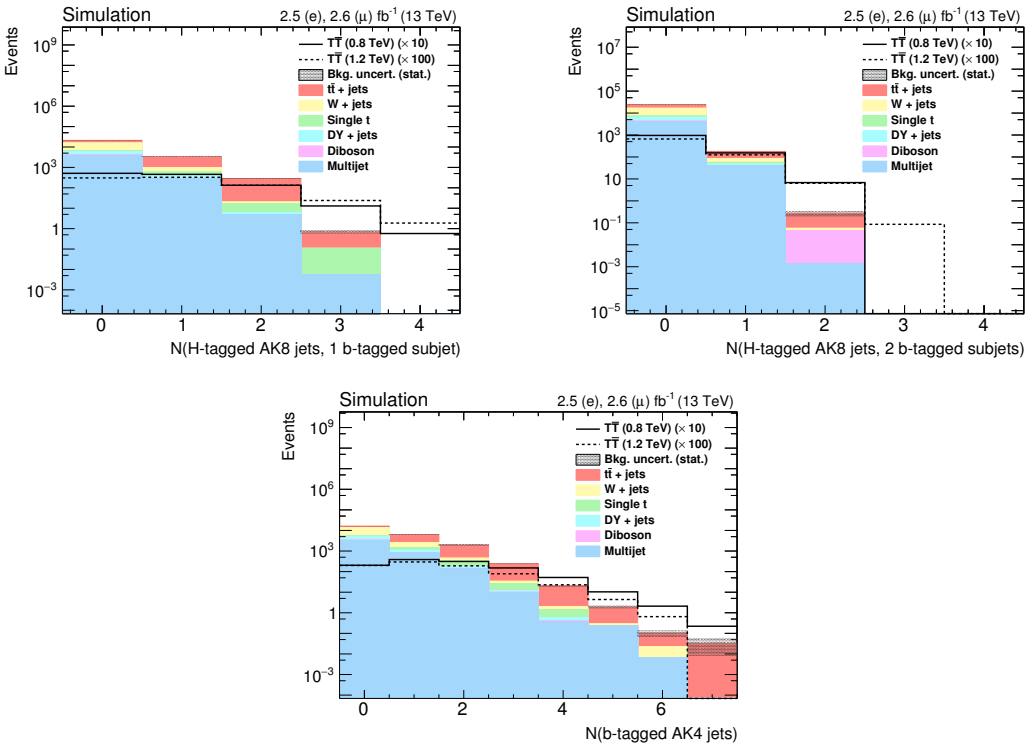

Figure 6.15: Distributions of the number of H-tagged AK8 jets with at least one b-tagged subjet (top left) and with exactly two b-tagged subjets (top right) for simulated events after the preselection. The number of b-tagged AK4 jets with $\Delta R > 0.8$ to the nearest H-tagged AK8 jet is shown on the bottom. The signal is normalised to the theoretically predicted cross section and a branching fraction of 100% is assumed for the $T \to tH$ decay.

Process	0H category	H1b category	H2b category
$T\bar{T}$ (0.8 TeV) $\to$ tHtH	$(6.2 \pm 0.8)\%$	$(6.7 \pm 0.8)\%$	$(2.3 \pm 0.5)\%$
$T\bar{T}$ (0.8 TeV) $\to$ tHtZ	$(6.9 \pm 0.8)\%$	$(5.9 \pm 0.6)\%$	$(1.6 \pm 0.4)\%$
$T\bar{T}$ (0.8 TeV) $\to$ tHbW	$(8.6 \pm 0.8)\%$	$(5.1 \pm 0.5)\%$	$(1.3 \pm 0.3)\%$
$T\bar{T}$ (0.8 TeV) $\to$ tZtZ	$(7.8 \pm 0.8)\%$	$(4.8 \pm 0.5)\%$	$(0.9 \pm 0.2)\%$
$T\bar{T}$ (0.8 TeV) $\to$ tZbW	$(9.1 \pm 0.7)\%$	$(3.8 \pm 0.4)\%$	$(0.5 \pm 0.1)\%$
$T\bar{T}$ (0.8 TeV) $\to$ bWbW	$(10.9 \pm 0.8)\%$	$(2.5 \pm 0.3)\%$	$(0.12 \pm 0.04)\%$
$T\bar{T}$ (1.2 TeV) $\to$ tHtH	$(6.9 \pm 1.0)\%$	$(8.0 \pm 1.1)\%$	$(2.8 \pm 0.6)\%$
$T\bar{T}$ (1.2 TeV) $\to$ tHtZ	$(7.7 \pm 1.0)\%$	$(7.2 \pm 0.8)\%$	$(2.0 \pm 0.4)\%$
$T\bar{T}$ (1.2 TeV) $\to$ tHbW	$(9.3 \pm 1.1)\%$	$(5.7 \pm 0.6)\%$	$(1.5 \pm 0.3)\%$
$T\bar{T}$ (1.2 TeV) $\to$ tZtZ	$(8.8 \pm 1.0)\%$	$(5.8 \pm 0.7)\%$	$(1.0 \pm 0.3)\%$
$T\bar{T}$ (1.2 TeV) $\to$ tZbW	$(10.5 \pm 1.0)\%$	$(4.3 \pm 0.5)\%$	$(0.6 \pm 0.1)\%$
$T\bar{T}$ (1.2 TeV) $\to$ bWbW	$(12.4 \pm 1.0)\%$	$(2.5 \pm 0.3)\%$	$(0.11 \pm 0.05)\%$

Table 6.8: $T\overline{T}$ event selection efficiencies in the three event categories for two T quark mass points, split into the six possible decays of the $T\overline{T}$ system. Efficiencies are calculated with respect to the expected number of events of each decay mode before any selection. Uncertainties include both statistical and systematic uncertainties (which are discussed in Section 6.4) in the simulated samples.

6.4 Systematic uncertainties

The rate of simulated signal and background processes and the shape of corresponding physical observables usually depend on a number of parameters such as the simulated b tagging efficiency (which is corrected to match the one in data) or the renormalisation scale used in the ME calculation. These parameters can carry uncertainties which are referred to as *systematic uncertainties*. These uncertainties have to be taken into account in the statistical interpretation of the final result or when visually inspecting the data.

Systematic uncertainties can either affect the rate of a given process or the shape of physical observables, or both at the same time. Shape uncertainties are accounted for by re-computing the distributions of the affected observables with a given parameter (e. g. the b tagging efficiency scale factor) varied either up or down by one standard deviation of the associated systematic uncertainty. This results in two additional templates per physical process, event category and systematic uncertainty for the given observable, one for the up- and one for the down-variation of the respective uncertainty. These templates can then be used in the statistical evaluation of the data in a technique referred to as *vertical template morphing* (see also App. A) and to create bands in distributions of physical observables to illustrate the impact of systematic uncertainties.

This section contains an overview over the systematic uncertainties which are considered in the evaluation of the background modelling (which is described in Section 6.5.3). All described sources of systematic uncertainties are treated as uncorrelated. A summary of these uncertainties along with approximate ranges for different processes in the three event categories is given at the end of this section in Table 6.9. Additional systematic uncertainties due to the modelling of the $t\bar{t}$ + jets and W + jets background processes are described in Section 6.5.2.

There are two groups of systematic uncertainties considered in this search: uncertainties of the first group, which are referred to as *experimental* systematic uncertainties, are due to uncertainties in the simulated detector response which usually arise from uncertainties in corresponding measurements. Uncertainties of the second group, so-called *theoretical* systematic uncertainties, are due to uncertainties in the theoretical calculations of the underlying physical process (e. g. from missing orders in the calculation of the matrix element). The individual systematic uncertainties are described in the following. Furthermore, this section provides comparisons of the unvaried, so-called *nominal* S_{T} distributions of various processes in the `H2b` category with the templates representing the up- and down-variations by one standard deviation of some of the most important systematic uncertainties.

6.4.1 Experimental uncertainties

Cross Section The normalisation of a given background process is proportional to the corresponding theoretically predicted production cross section, which carries some uncertainty on its own. These uncertainties are outlined in the following:

Multijet production Due to the poor modelling of this process in simulation (as e. g. can be observed in Figs. 6.2 and 6.3) a conservative systematic uncertainty of 100% is assigned in the production cross section of this process. The impact of this uncertainty on the final sensitivity of this search has been tested and found to be negligible as the contribution of multijet events in the final event categories is rather small.

DY+ jets, diboson, single top Production cross sections for these processes were measured at the LHC at $\sqrt{s} = 8\,\text{TeV}$ and $\sqrt{s} = 13\,\text{TeV}$ and good agreement was found between the theoretical predictions and the measured values. In accordance with the uncertainties in these measurements, normalisation uncertainties of 6% for DY+jets, 15% for diboson and 16% for single top production are assigned in this search [216–222].

$t\bar{t}$ + jets, W + jets The modelling of these background processes is evaluated using dedicated control regions which are described in more detail in Section 6.5. Systematic uncertainties due to this modelling are discussed there as well.

Luminosity The integrated luminosity corresponding to the data set used in this search has been measured using the Pixel Counting Method and calibrated in so-called Van der Meer scans during dedicated runs of the LHC (see Section 3.1). The uncertainty in this measurement is 2.3% [126] which is assigned as an uncertainty on the normalisation of all simulated processes (including signal and background).

Trigger The measurement of the selection efficiency scale factors for the triggers used in this search is explained in more detail in Section 6.3.1. For the e +jets trigger, no clear dependency of the measured scale factors on p_T or η was found and an overall scale factor of 0.99 is applied to all simulated events in the electron channel in this search. The uncertainty in the fit to calculate the scale factor, which was found to be 2%, is assigned as an uncertainty on the normalisation of all simulated processes in the electron channel.

For the μ trigger, the corresponding scale factors as well as their associated uncertainties have a dependency on p_T and η of the triggered muon so that both the normalisation and the shape of physical observables are affected. The impact is rather low and the resulting systematic uncertainty is below 1% in all event categories.

Lepton identification The identification efficiency scale factors for both electrons and muons have been measured as described in Sections 5.3 and 5.4. The uncertainties associated in these measurements affect both the event rate and the shapes of the physical observables. Their impact on the event rate ranges from 1–2% depending on the process and category so that this also constitutes a rather small uncertainty.

Jet energy corrections Differences between data and simulation in the detector response for jets are corrected for based on dedicated measurements (see Section 5.6). The associated uncertainties in the jet energy scale and resolution can affect both the event rates and the shapes of physical observables in simulated events. A detailed description of these uncertainties and their sources can be found in Ref. [183].In order to evaluate the impact of the uncertainties in the jet energy scale and resolution, the four-vectors of all both AK4 and AK8 jets are varied simultaneously by one standard deviation of the corresponding systematic uncertainty and all analysis steps are re-executed each time. The correction factors for the jet energy scale and the jet energy resolution are varied independently of each other so that the uncertainties from these two sources are treated as uncorrelated. This can lead to events being rejected or new events being selected in the final event selection. Additionally, this can lead to event migrations from one event category to another. The impact of these systematic uncertainties on the event rates of various processes ranges from 1–5% for jet energy scale uncertainties depending on the category and is below 1% for jet energy resolution uncertainties.

AK8 jet mass In addition to the systematic uncertainties in the energy scale and resolution of AK8 jets, also the groomed jet mass after applying the softdrop algorithm carries an uncertainty. As explained in Section 5.8, the scale and resolution of the groomed jet mass have been measured in data. The systematic uncertainty in the mass scale of AK8 jets is composed of two uncertainties which are added in quadrature: the uncertainty in the jet energy scale and an additional uncertainty of 2% due to the difference between the peaks of the groomed AK8 jet mass in data and simulation. To evaluate its impact, the groomed jet masses of all AK8 jets in an event are simultaneously varied up or down by a factor of 0.02 corresponding to one standard deviation of the this uncertainty. The systematic uncertainty in the jet mass resolution was found to be 10.3% and is evaluated by varying the groomed jet masses of all AK8 jets simultaneously using the equation

$$m_{\text{jet,smeared}} = m_{\text{jet,ptcl}} + u \cdot (m_{\text{jet,reco}} - m_{\text{jet,ptcl}}) \tag{6.1}$$

where u is 1 ± 0.103 depending on whether the up- or down-variation of this uncertainty is evaluated. In both cases, the final event categorisation is re-executed after

varying the corresponding parameter to evaluate the impact of these uncertainties.

The systematic uncertainties in the mass scale and resolution are treated as uncorrelated. For both uncertainties, this can lead to event migrations from the `H1b` and `H2b` categories to the `0H` category and vice versa. Their impact on the event rates in the `H1b` and `H2b` categories ranges from 1–15% for the jet mass scale uncertainty and from 1–12% for the jet mass resolution uncertainty. In the `0H` category, their impact stays below 1% for both sources of uncertainties.

Since the processes of interest in this search are boosted hadronic $H \to b\bar{b}$ decays, the jet mass can also be affected by neutrinos from leptonic bottom quark decays. It is therefore not ensured that the measured jet mass scale and resolution for jets from boosted hadronic W boson decays is equal to the one for jets from $H \to b\bar{b}$ decay chains. Potential differences between jets from $W \to q\bar{q}'$ and $H \to b\bar{b}$ decays have been estimated in a search for heavy resonances decaying to two Higgs bosons [223]. This was performed by evaluating boosted hadronic W and Higgs boson decays simulated with two different hadronisers, Pythia 8 and Herwig++. While for boosted W decays, the scale and resolution of AK8 jets matched to generated W bosons was found to be consistent between Pythia 8 and Herwig++, a significant difference between the two hadronisers was observed for boosted Higgs boson decays, both in the scale and resolution of the jet mass of matched AK8 jets. To account for this effect, a normalisation uncertainty of 10% was assigned corresponding to the different selection efficiency for the H tagging algorithm between the two hadronisers. As the mass window defined in the H tagging algorithm of the search in Ref. [223] is considerably narrower than the one in this search ([105, 135] GeV compared to [60, 160] GeV), a lower normalisation uncertainty of 5% is assigned in this search.

B tagging Uncertainties in the b tagging efficiency and misidentification rate both for AK4 jets and subjets of AK8 jets constitute one of the largest sources of systematic uncertainties in this search since b tagging is one of the key elements in discriminating signal-like from background-like events. Since the heavy-flavour and light-flavour identification efficiencies were measured using two different methods (see Section 5.7), the corresponding systematic uncertainties are treated as uncorrelated. Systematic uncertainties in the derived scale factors for AK4 jets and subjets of AK8 jets are, however, treated as fully correlated as corresponding measurements were performed in a nearly identical data sample. Systematic uncertainties are evaluated by varying the scale factors calculated for each event up and down by one standard deviation of the uncertainty in the measurement.

The resulting up- and down-variations of the S_{T} observable in the `H2b` category can be seen for $t\bar{t}$ + jets and $T\overline{T}$ events in Fig. 6.16. The systematic uncertainty in the

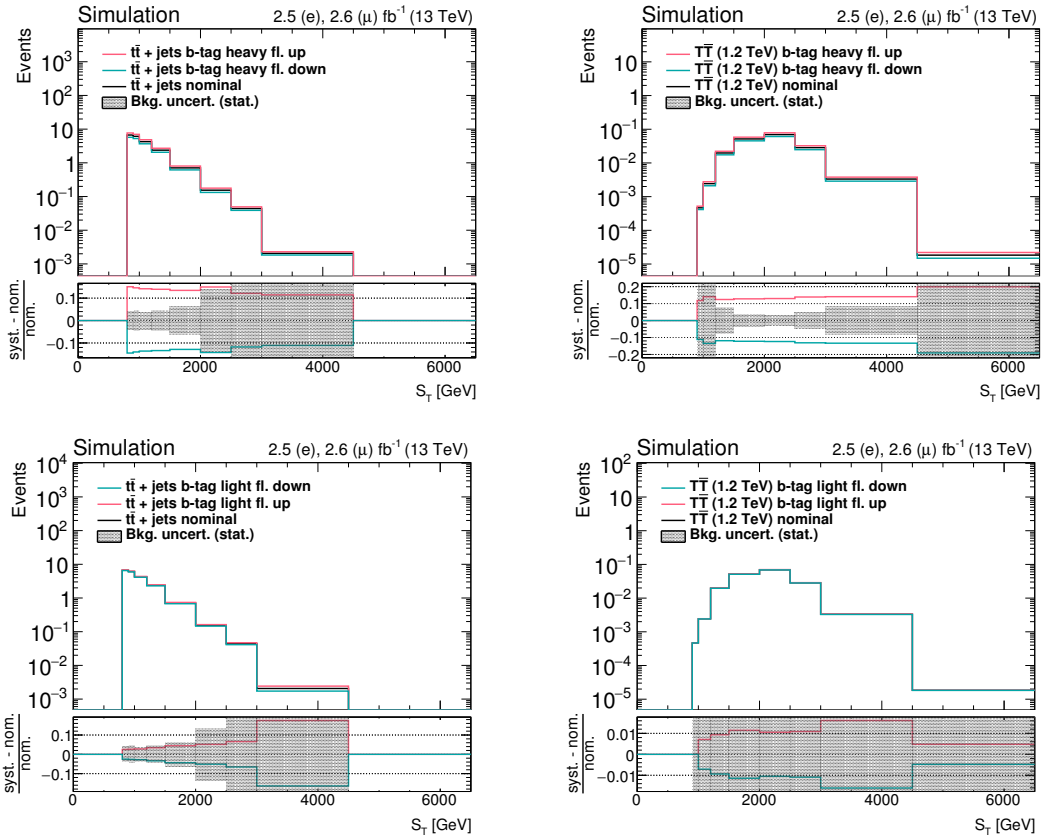

Figure 6.16: Comparisons of the templates for the up- and down-variations corresponding to the systematic uncertainties in the heavy-flavour (top row) and light-flavour b tagging efficiency (bottom row). These are shown in the `H2b` category for the $t\bar{t}$ + jets background (left) and the $T\overline{T}$ signal (right) for a T quark mass of 1200 GeV. The ratio between the individual variations and the nominal distribution per bin is shown in the bottom pads of each distribution.

rate ranges from 3–14% for the heavy-flavour identification efficiency and from 1–3% for the light-flavour identification efficiency.

6.4.2 Theoretical Systematic Uncertainties

In this subsection, systematic uncertainties in the theoretical assumptions that were used to simulate the MC samples are described. These affect both the production cross section as well as the shapes of physical observables of the simulated processes.

The systematic uncertainties in the $\mu_{\mathrm{R/F}}$ energy scales and the PDF set used in this search (which are explained in the following) are treated differently for simulated signal and background samples: for background processes, their impact on the event rate is fully taken into account. For the $T\overline{T}$ signal, they are treated as uncertainties that affect exclusively the shape of physical observables. Their impact on the signal acceptance

after the event categorisation is taken into account which results in an uncertainty in the signal yield in the three categories. The reason for this is that the uncertainties in $\mu_{\mathrm{R/F}}$ scales and PDFs are included in the theoretically predicted cross section when statistically interpreting the final results.

Pileup Reweighting In order to calculate the expected number of interactions per bunch-crossing which is needed to perform pileup reweighting in the simulation (see Section 5.9), an inelastic pp scattering cross section of 69 mb is assumed. The uncertainty in this cross section is 5% which is evaluated by deriving the weights in the pileup reweighting procedure with the inelastic pp scattering cross section varied up or down by 5% and re-computing physical observables with these weights.

$\mu_{\mathrm{R/F}}$ **Scale** A considerable fraction of the systematic uncertainty both in the rate of a process and in the shape of physical observables is due to the choice for the renormalisation and factorisation energy scales $\mu_{\mathrm{R/F}}$. As explained in Chapter 4, the renormalisation scale μ_{R} enters into the partonic cross section and determines e. g. the strength of the running coupling constant α_{s} through the corresponding renormalisation group equations. The factorisation scale μ_{F} is a parameter providing the scale up to which higher-order QCD effects are absorbed into the PDFs in the calculation of pp scattering processes. Both values are usually set to the momentum transfer Q^2 of the simulated hard scattering process. However, there is no a priori constraint on the values for these scales and therefore, possible variations of their values and the impact of these variations on physical observables are taken into account as systematic uncertainties.

This systematic uncertainty is evaluated by considering six variations of the scales μ_{R} and μ_{F}: either simultaneous up or down variations by factors of 2 or 1/2 or independent variations of one scale while fixing the other one[3]. The observables used in this search are re-computed for all six cases by using corresponding event weights that were calculated by the matrix element generators and stored in the MC samples. The final up- and down-variations of an observable are then determined by calculating the envelope of the six variations bin by bin. For the multijet, diboson and single $t + W$ samples, however, these weights were not calculated in the ME simulation so that this uncertainty cannot be evaluated for these processes. It is deemed to be covered, however, by the uncertainties in the normalisation assigned to the simulated processes.

The impact of this uncertainty on the shape of the S_{T} observable is shown in Fig. 6.17 and the associated uncertainty in the event rates for background processes ranges

[3]Variations in opposite directions, i. e. one scale varied up while the other one is varied down, are not considered as they are not deemed physical.

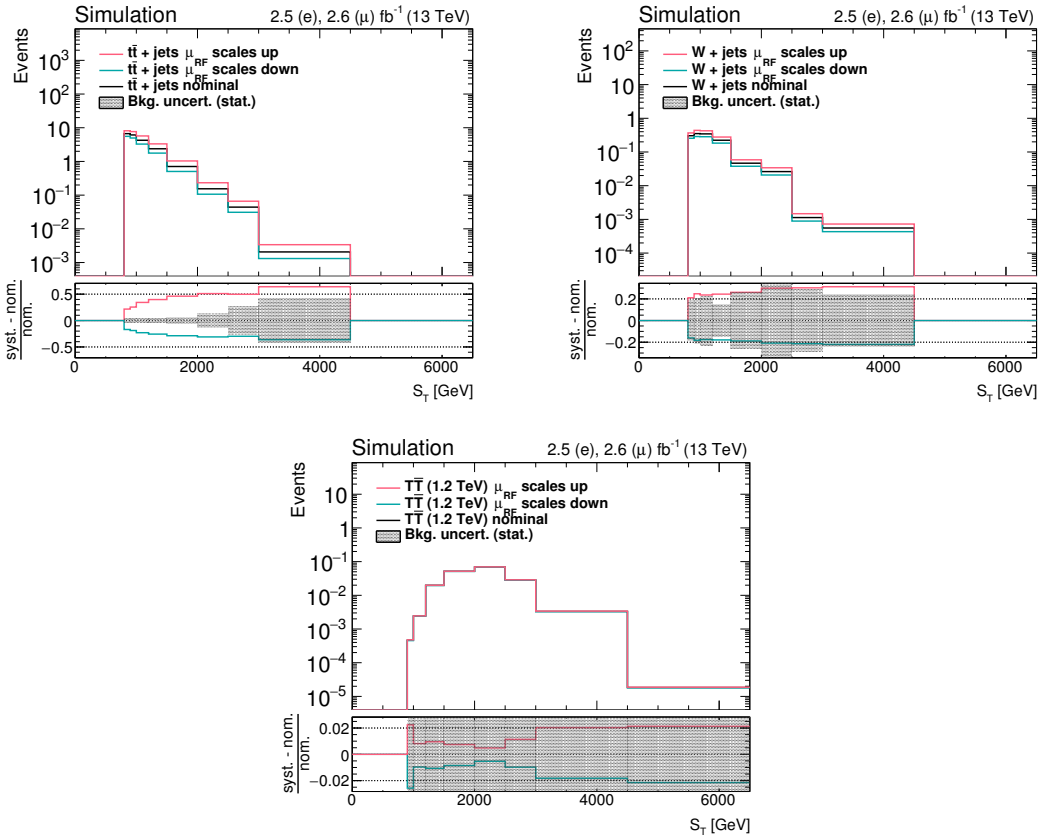

Figure 6.17: Comparisons of the templates for the up- and down-variations corresponding to the systematic uncertainties in the renormalisation and factorisation energy scales $\mu_{\mathrm{R/F}}$. These are shown in the `H2b` category for the $t\bar{t}$ + jets background (left), the W + jets background (right) and the $T\overline{T}$ signal (bottom) for a T quark mass of 1200 GeV. The ratio between the individual variations and the nominal distribution per bin is shown in the bottom pads of each distribution.

from 20–30%. For the signal, the impact on the acceptance is rather small and stays below 1% in all categories.

PDF An important source of systematic uncertainties is due to the form of the PDFs used in the calculation of the *pp* scattering process (see 4.2). The PDFs used in this search were calculated by the NNPDF collaboration [152] and supplemented by a set of 100 replicas. These replicas are obtained by repeating the fit to extract the nominal PDF with different seeds for the initialisation of the random numbers used in the minimisation of the fit.

The replicas are used to determine the systematic uncertainty in the used PDF set. This is done by re-computing physical observables for each of these replicas using corresponding event weights stored in the MC samples. From the resulting

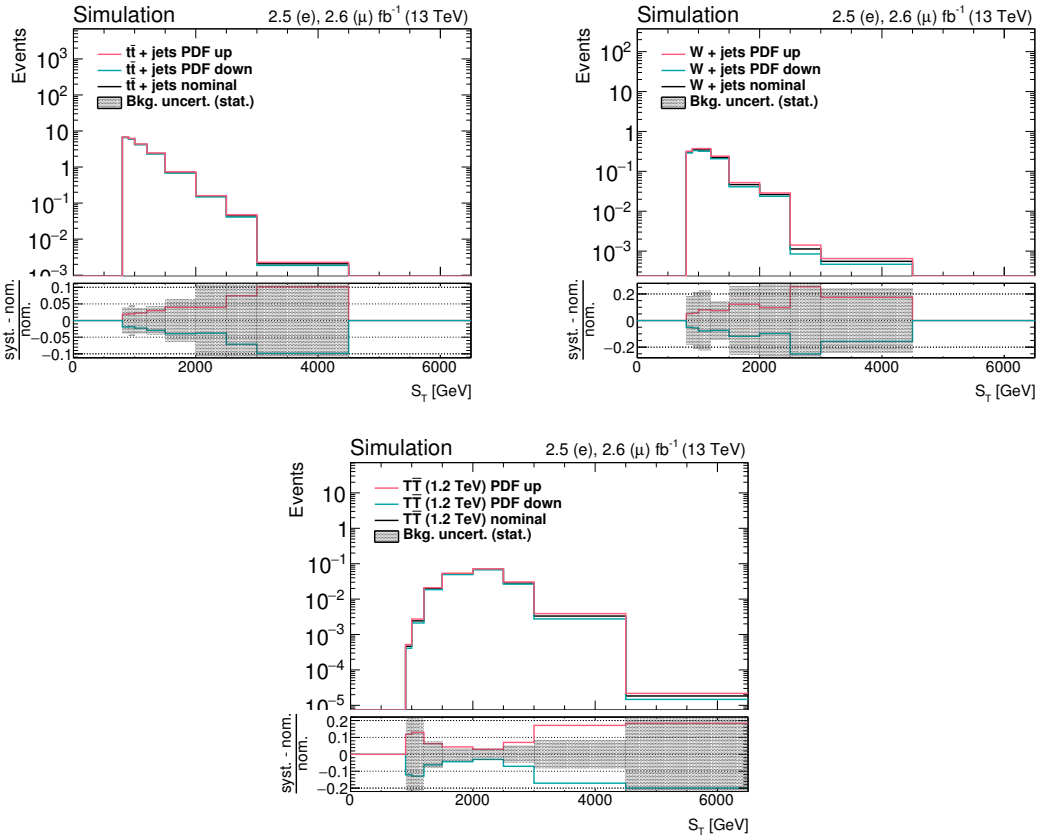

Figure 6.18: Comparisons of the templates for the up- and down-variations corresponding to the systematic uncertainties in the PDF set used. These are shown in the `H2b` category for the $t\bar{t}$ + jets background (left), the W + jets background (right) and the $T\overline{T}$ signal (bottom) for a T quark mass of 1200 GeV. The ratio between the individual variations and the nominal distribution per bin is shown in the bottom pads of each distribution.

100 distributions for each observable, the templates corresponding to the up- and down-variations by one standard deviation of this uncertainty are constructed by calculating the root mean square of all replica distributions per bin of the observable.

The impact of the uncertainty in the PDF ranges from 2–8% in the rate of background processes and from 2–9% in the acceptance of the signal. Their impact on the S_{T} distribution in the `H2b` category can be seen for the $t\bar{t}$ + jets and W + jets backgrounds as well as the $T\overline{T}$ signal for a T quark mass of 1200 GeV in Fig. 6.18.

All systematic uncertainties described in this section along with their approximate impact on the rates of three different processes ($t\bar{t}$ + jets, W + jets and $T\overline{T}$ with a T quark mass of 800 GeV) over the three event categories are summarised in Table 6.9.

Source	Experimental uncertainties		
	$t\bar{t}$ + jets	W + jets	$T\overline{T}$ (M(T) = 800 GeV)
Trigger (e) (R)	2%	2%	2%
Trigger (μ)	$\lesssim$ 1%	$\lesssim$ 1%	$\lesssim$ 1%
e identification eff.	1-2%	1-2%	1-2%
μ identification eff.	1-2%	1-2%	1-2%
Jet energy scale	2–5%	2–4%	$\lesssim$ 1%
Jet energy res.	$\lesssim$ 1%	$\lesssim$ 1%	$\lesssim$ 1%
AK8 jet mass scale	1–8%	0–15%	1–2%
AK8 jet mass res.	0–5%	0–12%	$\approx$ 1%
H tag: W $\to$ H (R)	5%	5%	5%
b tag eff., h.-fl.	4–14%	3–15%	7–13%
b tag eff., l.-fl.	1–3%	3–4%	$\approx$ 1%
Luminosity (R)	2.3%	2.3%	2.3%

Source	Experimental normalisation uncertainties
DY+ jets rate (R)	3%
diboson rate (R)	15%
Single top rate (R)	20%
Multijet rate (R)	100%

Source	Theoretical uncertainties		
	$t\bar{t}$ + jets	W + jets	$T\overline{T}$ (M(T) = 800 GeV)
Pileup reweighting	$\lesssim$ 1%	$\lesssim$ 1%	$\lesssim$ 1%
$\mu_{\mathrm{R/F}}$ scales	33–40%	21–26%	$\lesssim$ 1% (accept.)
PDF	2–6%	4–8%	2–9% (accept.)

Table 6.9: Summary of systematic uncertainties with representative ranges. Approximate values are shown for three processes: $t\bar{t}$ + jets, W + jets and $T\overline{T}$ production (for a T quark mass of 800 GeV). They indicate the range of the systematic uncertainty over the three event categories. The symbol 'R' indicates uncertainties on the normalisation of a process.

6.5 Background modelling and H_{T} reweighting

Standard Model background processes in the three event categories are estimated using simulated samples. As the S_{T} variable is used in the final statistical evaluation of the data, it needs to be verified that this observable is modelled accurately, especially in its tails where potential $T\overline{T}$ production would be visible. This is done using two control regions enriched in events from $t\bar{t}$ + jets and W + jets production which is described in Section 6.5.1. As discrepancies between data and simulation are observed in these regions, a reweighting procedure based on the H_{T} variable is performed which is discussed in Section 6.5.2. The final determination of systematic uncertainties in the background model is outlined in Section 6.5.3.

6.5.1 Control regions

To evaluate the modelling of the two main backgrounds for this analysis, $t\bar{t}$ + jets and W + jets, two control regions are defined after the preselection. A veto on an H-tagged jet is applied in order to be orthogonal to the `H1b` and `H2b` categories, as these are the signal regions in this search (see Section 6.3.4). Furthermore, a missing transverse energy greater than 100 GeV is required in both control regions to reduce the contribution from multijet production in these regions. The control regions are finally defined by the following requirements

- In the $t\bar{t}$ + jets control region, at least two b-tagged AK4 jets are required.
- In the W + jets control region, a veto on a b-tagged AK4 jet is applied.

Events in the $t\bar{t}$ + jets control region therefore constitute a subset of the events in the `0H` category which is, however, not considered as signal region due to its low signal-to-background ratio over the whole range of S_{T}.

$t\bar{t}$ + jets and W + jets production constitute 84% and 71% in their respective control regions so that a large purity in these processes is ensured. The inclusive signal-to-background ratios in these regions are approximately 3% and $< 1\%$, respectively, for a T quark with a mass of 800 GeV assuming the theoretically predicted cross section and a branching fraction of 100% for the $T \to tH$ decay. As the $T\overline{T}$ signal appears primarily in high-S_{T} regions where the background contribution is smaller, potentially present signal could bias assessments of the background model based on the $t\bar{t}$ + jets control region. The impact of the presence of signal events is therefore investigated in dedicated tests which are summarised in App. C.3. These tests include a reweighting of the simulated background processes which is described in detail in Section 6.5.2. Any potential bias due to present signal is found to be negligible compared to other systematic uncertainties and both control regions are considered to provide valid assessments of the background model.

Distributions of S_{T} in these two regions are shown in Fig. 6.19. The uncertainty bands in these plots show both the statistical uncertainty in the MC samples and the total uncertainty after combining the statistical and systematic all uncertainties described in Section 6.4. In both control regions, a trend is visible in the ratio between the data and simulated background distributions. The simulation shows an excess of events which seemingly grows for higher values of S_{T}. This trend appears to be partially covered by the systematic uncertainties in the expected background. This is, however, due only to the very large uncertainty in the event rate associated with the systematic uncertainty in the $\mu_{\mathrm{R/F}}$ energy scales.

To evaluate the modelling of the shape of the S_{T} distribution, it is therefore tested whether this discrepancy is still covered by the systematic uncertainties when only considering the shape-only contributions of systematic uncertainties. This is done by comparing

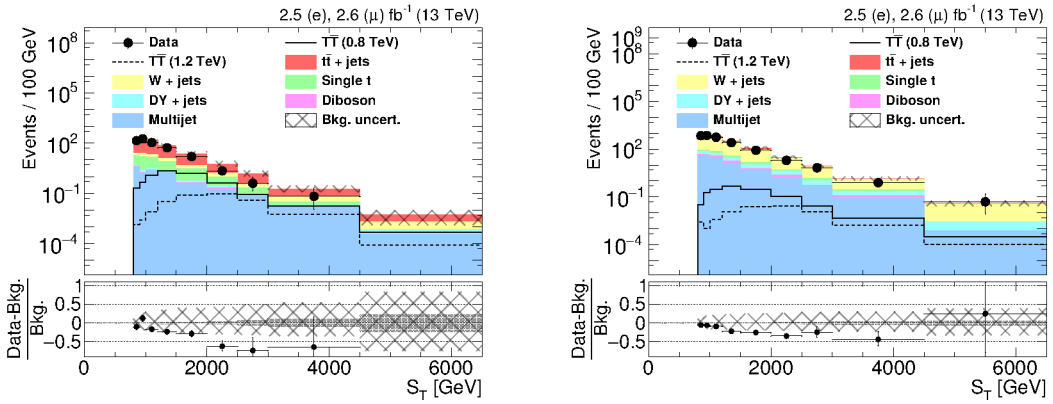

Figure 6.19: S_T distributions in the $t\bar{t}$ +jets (left) and W +jets (right) control regions after combining electron and muon channel. The dark gray band in the ratio plot corresponds to the statistical uncertainty in the MC samples, the hatched band illustrates the total uncertainty after combining statistical and systematic uncertainties in the simulation.

the data and simulated samples after normalising all histograms to unity, including the data distribution as well as the nominal background distribution and the templates for the systematic up- and down-variations. This way, uncertainties in the rate of the background prediction are cancelled out automatically.

Including only experimental systematic uncertainties (see Section 6.4.1), one can see in Fig. 6.20 that their impact on the shape of the S_T distribution is far to small to explain the observed shape discrepancies between data and simulation. Considering only theoretical systematic uncertainties (see Section 6.4.2), which is shown in Fig. 6.21, one observes a larger uncertainty in the shape of the S_T observable, however, this also does not cover the shape discrepancy between data and simulation. The same can be observed when combining both groups of uncertainties and considering them all at once (shown in Fig. 6.22). This indicates that the observed shape discrepancy between the data and the simulation is either due to a systematic effect which is not covered by any of the considered systematic uncertainties, or that some of the considered uncertainties are considerably underestimated.

In addition to these visual assessments, two tests are performed to quantify this mismodelling of the $t\bar{t}$ + jets and W + jets background processes: in the first test, a binned *maximum-likelihood* (MLE) fit of the background-only hypothesis is performed to test how much the simulated background templates would have to varied within the systematic uncertainties to accommodate the observed discrepancies (a more detailed description of this statistical technique can be found in App. A.1). The second test consists of so-called *Goodness-of-fit* (GOF) tests which provide p-values for the hypothesis that an observed data spectrum results from a given background model. The GOF tests used in the following are explained in more detail in App. A.4.

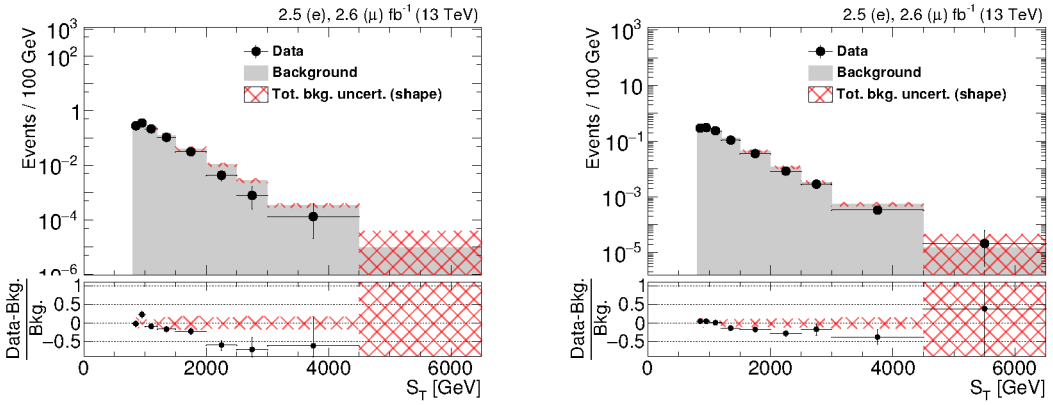

Figure 6.20: S_T distribution in the $t\bar{t}$ +jets (left) and W +jets (right) control regions after combining the electron and muon channels. All histograms, including data, nominal and up and down variation histograms are normalised to unity so that only the shape effects of systematic uncertainties are displayed. The hatched band represents the **experimental** systematic uncertainties in the background prediction.

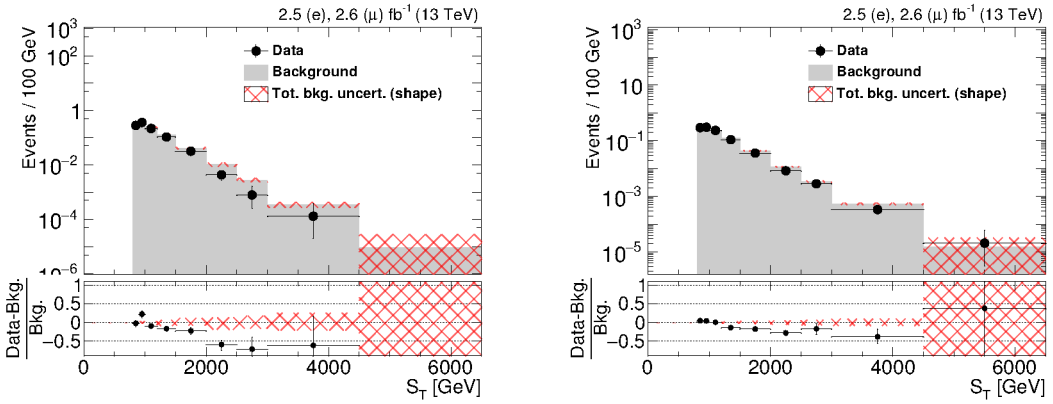

Figure 6.21: S_T distribution in the $t\bar{t}$ +jets (left) and W +jets (right) control regions after combining the electron and muon channels. All histograms, including data, nominal and up and down variation histograms are normalised to unity so that only the shape effects of systematic uncertainties are displayed. The hatched band represents the **theoretical** systematic uncertainties in the background prediction.

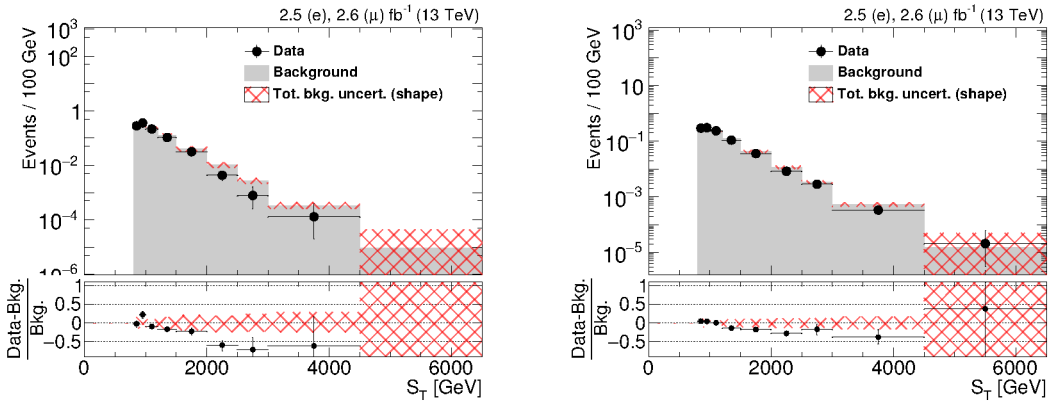

Figure 6.22: S_T distribution in the $t\bar{t}$ +jets (left) and W +jets (right) control regions after combining the electron and muon channels. All histograms, including data, nominal and up and down variation histograms are normalised to unity so that only the shape effects of systematic uncertainties are displayed. The hatched band represents the **total** systematic uncertainties (combining experimental and theoretical uncertainties) in the background prediction.

The MLE fit is performed using the data and MC distributions in the two control regions, split into electron and muon channel, so that four independent distributions are simultaneously fitted. The systematic uncertainties described in Section 6.4 are assigned as nuisance parameters and the final evaluation of the background modelling is based on the comparison of the pre-fit and post-fit values of these nuisance parameters. Additional uncertainties in the rates of the $t\bar{t}$ + jets and W + jets processes are assigned with a flat prior. All other normalisation uncertainties are given log-normal priors while all nuisance parameters corresponding to uncertainties in the shape of the S_T distribution are given Gaussian priors.

The differences between the post-fit values and their prior values, expressed in units of one standard deviation of their prior uncertainties, can be found in Fig. 6.23 for each nuisance parameter. Figure 6.23 also shows the constrained post-fit uncertainties in the same units. It can be seen that some post-fit values are shifted considerably with respect to their prior values, with e. g. the post-fit values for the $t\bar{t}$ + jets rate and the $\mu_{R/F}$ scale uncertainties being more than or close to two standard deviations away from their prior values. This illustrates how the fit attempts to match the shapes of the background distributions to the shape of the data distribution by strongly morphing the templates corresponding to these systematic uncertainties.

The performed GOF tests are a χ^2 *test* and a *Kolmogorov-Smirnov* (KS) test which is particularly sensitive to discrepancies that are correlated between the bins of two distributions. These tests return test statistic values which depend on the compatibility of an observed or simulated spectrum with a given model. They are first performed using the

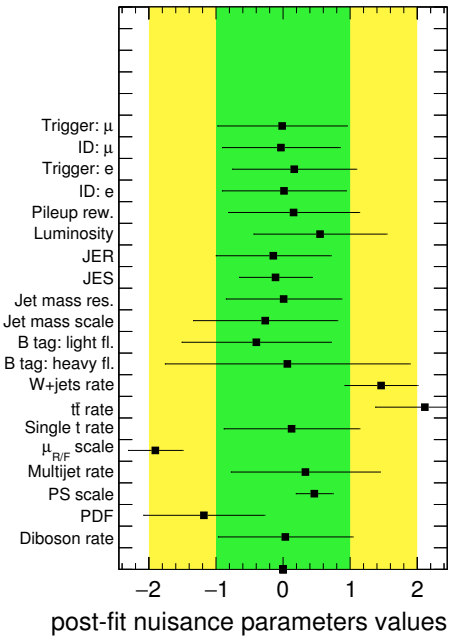

Figure 6.23: Post-fit nuisance parameters values after simultaneously fitting the background-only hypothesis in the $t\bar{t}$ + jets and W+jets control regions. The fit parameters are expressed as deviations from their pre-fit values (corresponding to a value of zero) in units of the corresponding prior uncertainty. The bands representing one and two standard deviations of the prior uncertainty of each parameter are shown in green and yellow, respectively.

data in the two control regions with the simulated background samples used as background model. The returned test statistic value is then compared with a distribution of outcomes of 5000 toy experiments in which pseudo-data are randomly generated from the simulated background samples and compared with the same background model. All systematic uncertainties described in Section 6.4 are taken into account in both tests. The results are shown in Fig. 6.24 where it can be seen that the returned test statistic value for the data is relatively large compared to the outcomes of the toy experiments. The corresponding p-values are 11% for the χ^2 test and 3.4% for the KS test. The results both of the MLE fit and the GOF tests thus indicate a mismodelling of the $t\bar{t}$ + jets and W + jets background processes.

A possible source of the discrepancy in the $t\bar{t}$ + jets control region is a mismodelling of the p_{T} spectra of top quarks in simulated $t\bar{t}$ + jets events which has been observed in measurements of the differential $t\bar{t}$ + jets production cross section in LHC Run I [224, 225]. A similar effect was found again in the first measurements using the 13 TeV data set [226, 227]. In both measurements, a softer p_{T} spectrum has been observed in the data compared to simulated $t\bar{t}$ + jets events.

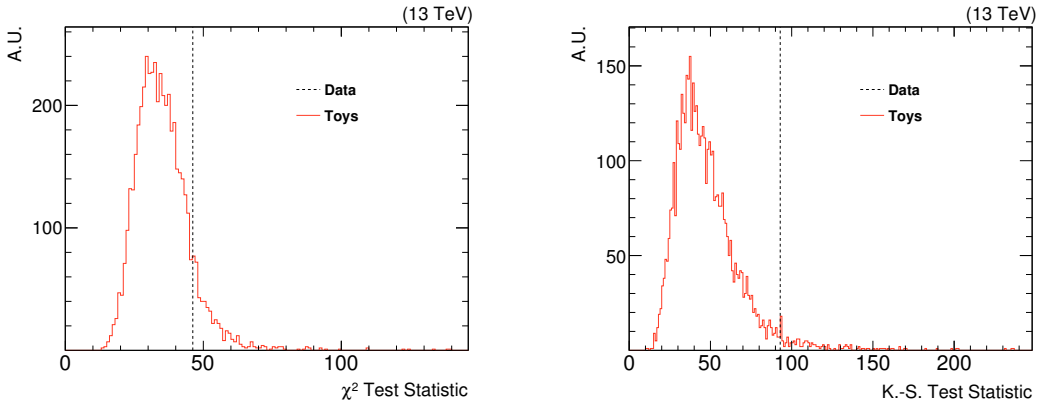

Figure 6.24: Test statistic values for two GOF tests performed using the data (indicated by the dashed line) observed in the $t\bar{t}$ + jets and W + jets control regions and compared to the outcomes of 5000 toy experiments using pseudo-data sampled from the background templates in these two regions. The results of the χ^2 test are shown on the left and the results from the Kolmogorov-Smirnov are shown on the right.

A procedure was developed to correct for this effect in search analyses which is referred to as *top-p_T reweighting*. This procedure is described in App. C.1 and it is tested whether applying this method provides a better modelling of the S_T variable in the $t\bar{t}$ +jets control region. It is found, however, that the top-p_T reweighting has only a minor impact on the shape of S_T so that this cannot account for the observed discrepancies.

6.5.2 H_T reweighting

The components of the S_T variable, which are the primary lepton p_T, $\not{E}_\mathrm{T}$ and H_T, are shown in Fig. 6.25. A significant trend is observed in particular in the H_T distribution in both control regions which is not covered by the band representing the systematic uncertainties in the background expectation.

The mismodelling of the H_T observable can originate from various sources which can be studied by examining various kinematic distributions related to AK4 jets. The number of AK4 jets and the p_T spectra of the first three AK4 jets are shown in Fig. 6.26. One can see that all these observables seem to be slightly overpredicted in the simulation towards higher values. It is therefore difficult to identify a single source for the observed discrepancies, suggesting that they are caused by a combination of several effects.

To account for this mismodelling and to determine an appropriate systematic uncertainty, a reweighting procedure similar to the previously mentioned top-p_T reweighting method is applied to correct the shape and normalisation of the S_T variable. This is referred to as H_T *reweighting* and carried out by deriving event scale factors with a linear

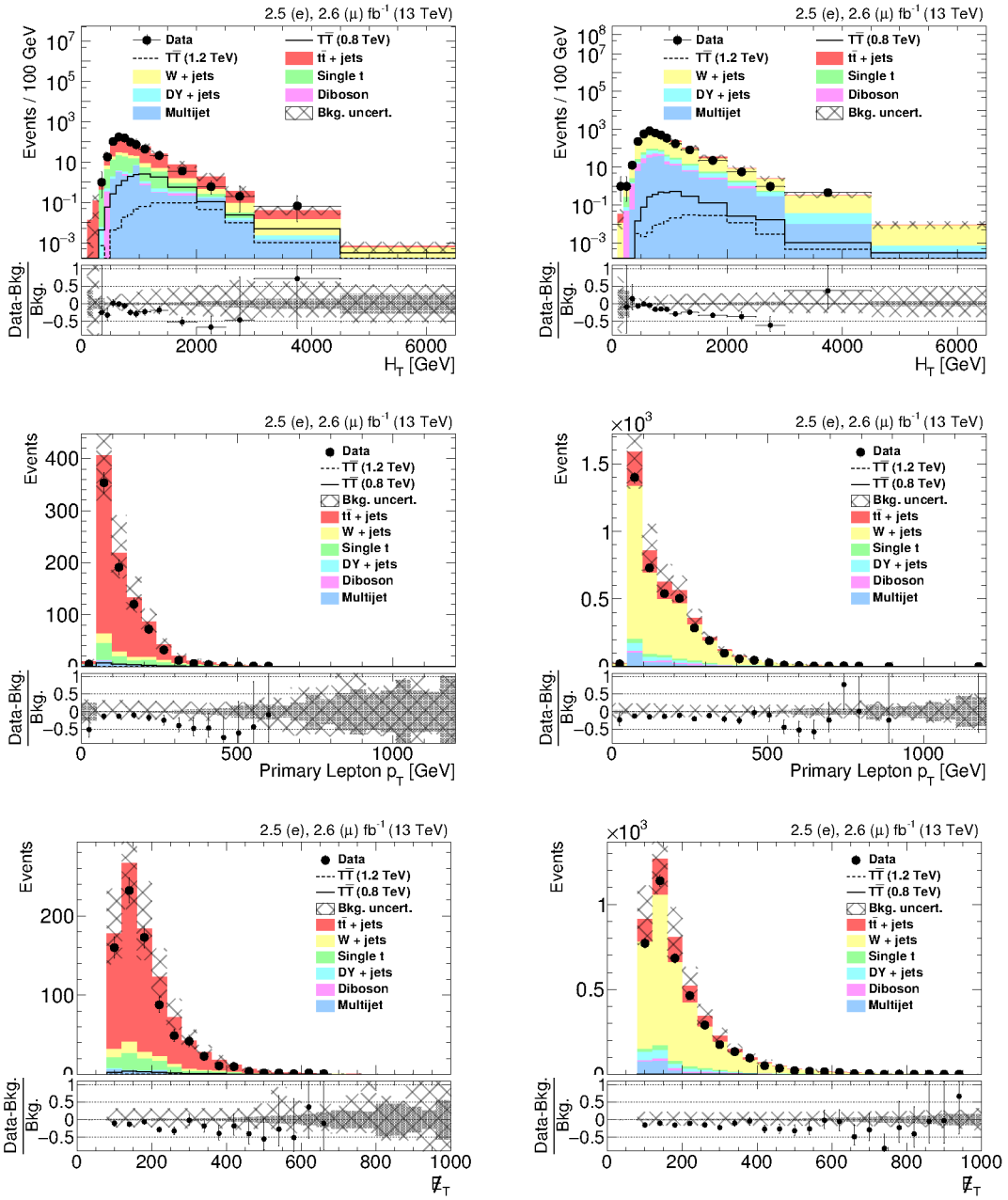

Figure 6.25: Various distributions in the $t\bar{t}$ +jets (left) and W +jets (right) control regions after combining electron and muon channels. From top to bottom H_{T}, p_{T}(prim. lep.) and $\not{E}_{\mathrm{T}}$ are shown. The statistical uncertainty in the simulation is shown as gray band and the total uncertainty, combining the statistical uncertainty and all systematic uncertainties described in Section 6.4, is shown as hatched band.

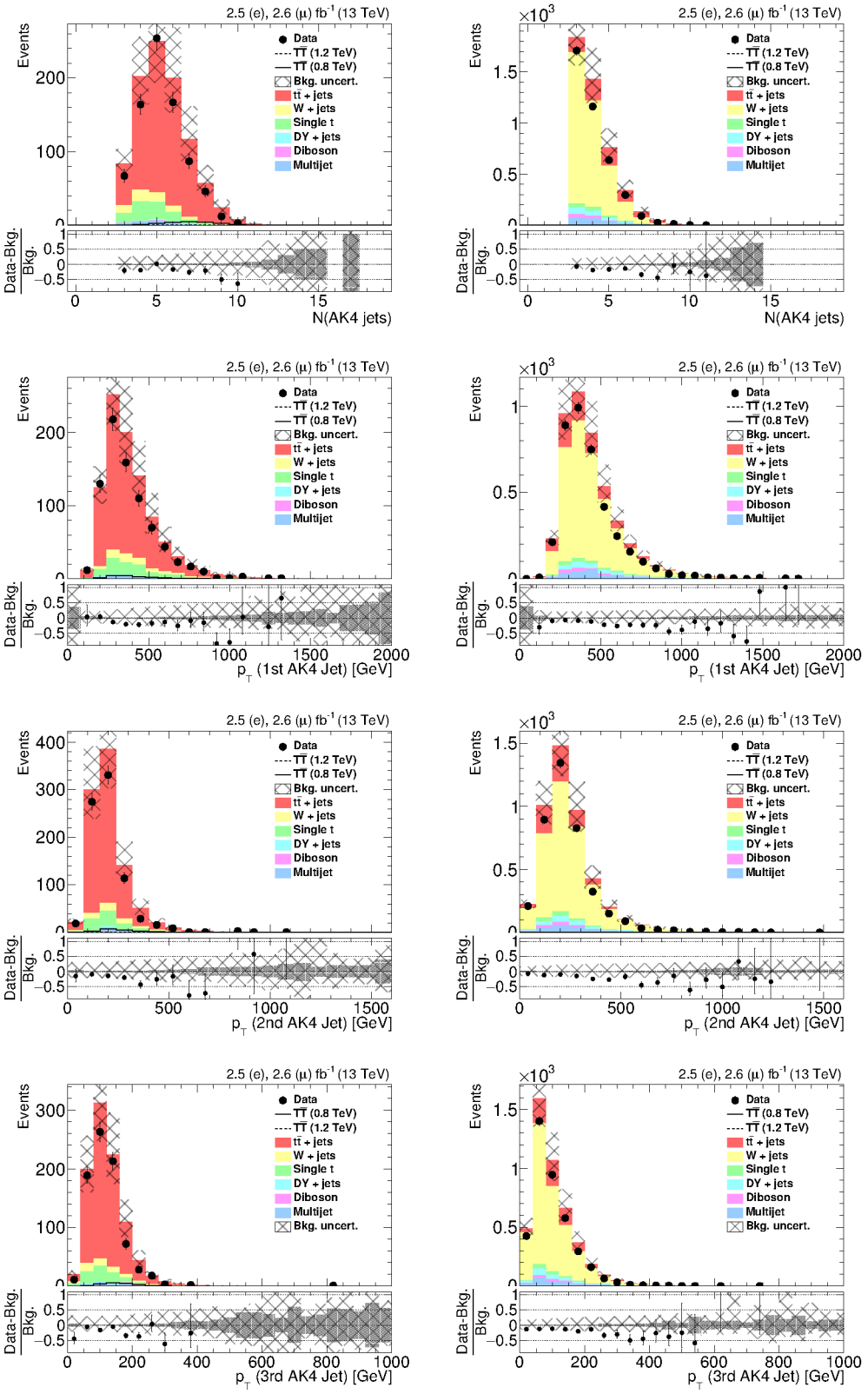

Figure 6.26: Various distributions in the $t\bar{t}$ + jets (left) and W + jets (right) control regions after combining electron and muon channels. From top to bottom the number of AK4 jets and the p_T spectra of the first, second and third AK4 jet (in order of their p_T) are shown. The statistical uncertainty in the simulation is shown as gray band and the total uncertainty, combining the statistical uncertainty and all systematic uncertainties described in Section 6.4, is shown as hatched band.

function of H_{T}

$$\mathrm{SF}(H_{\mathrm{T}}) = a + b \cdot H_{\mathrm{T}}. \tag{6.2}$$

These scale factors are applied to both the $t\bar{t}$ + jets and W + jets background processes using separate parameters a and b for the two processes. To derive these parameters, the function 6.2 is fitted to the H_{T} distributions in the corresponding control regions for the two background processes over the entire H_{T} range.

A linear function is chosen since the trend in the ratio between data and simulation exhibits a linear behaviour. This is also reflected in the results of the fits which return reasonable χ^2/N_{dof} values of order ~ 1 as is shown in the following. The result of fitting Eq. (6.2) in the $t\bar{t}$ + jets control region yields the following parameter values for a and b:

$$\begin{aligned} a_{final} &= 1.12 \pm 0.09 \\ b_{final} &= (-3.1 \pm 0.9) \cdot 10^{-4}\mathrm{GeV}^{-1} \\ \chi^2/N_{\mathrm{dof}} &= 10.4/11 \end{aligned} \tag{6.3}$$

For the W + jets background, the following parameters are found from the fit in the W + jets control region:

$$\begin{aligned} a_{final} &= 1.08 \pm 0.04 \\ b_{final} &= (-2.6 \pm 0.4) \cdot 10^{-4}\mathrm{GeV}^{-1} \\ \chi^2/N_{\mathrm{dof}} &= 16.6/13 \end{aligned} \tag{6.4}$$

Visualisations of the fits in the two control regions can be found in Fig. 6.27. To prevent the scale factor from becoming negative and as there are only very few data events in the last bin in both control regions, the scale factors are determined using function 6.2 only up to an H_{T} value of 3000 GeV. For events with larger H_{T} values, a constant scale factor is applied which corresponds to $\mathrm{SF}(H_{\mathrm{T}} = 3000\ \mathrm{GeV})$.

A systematic uncertainty in this procedure is assigned by defining an unreweighted distribution as up-variation, and a distribution reweighted with the scale factor in quadrature as down-variation of one standard deviation of the uncertainty.

As it is not clear whether the trends in $t\bar{t}$ + jets and W + jets are caused by the same effect or not, this systematic uncertainty is considered to be fully correlated between the two background processes as this provides a more conservative result than treating them as independent of each other.

After applying this H_{T} reweighting procedure, a good agreement between data and MC simulation is found in the S_{T} distributions in the two control regions as can be seen in Fig. 6.28. The impact of the systematic uncertainty in the H_{T} reweighting procedure on the event yields of the $t\bar{t}$ + jets and W + jets background processes ranges from 15 to 20% depending on the event category.

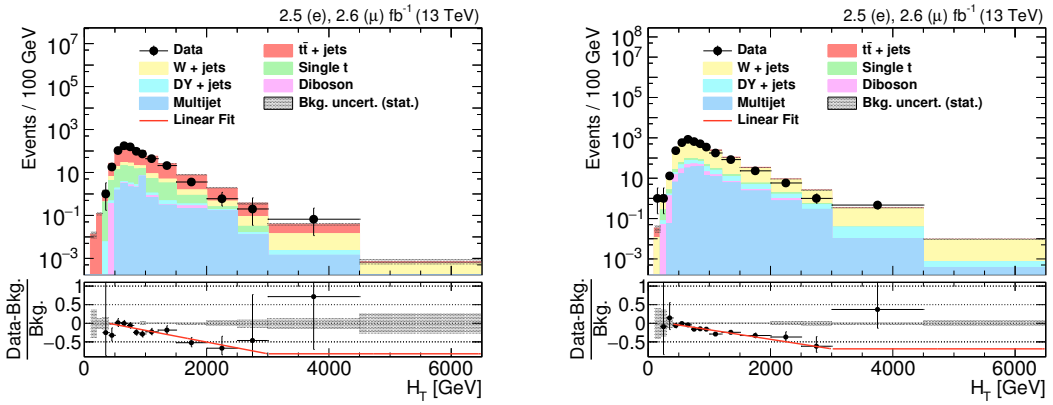

Figure 6.27: H_{T} distributions in the $t\bar{t}$ + jets control region (left) and the W+jets control region (right) in the combined lepton channels. A visualisation of the fit of Eq. (6.2) is shown in the data/MC ratio plot with a constant scale factor for H_{T} values $> 3000\,\mathrm{GeV}$.

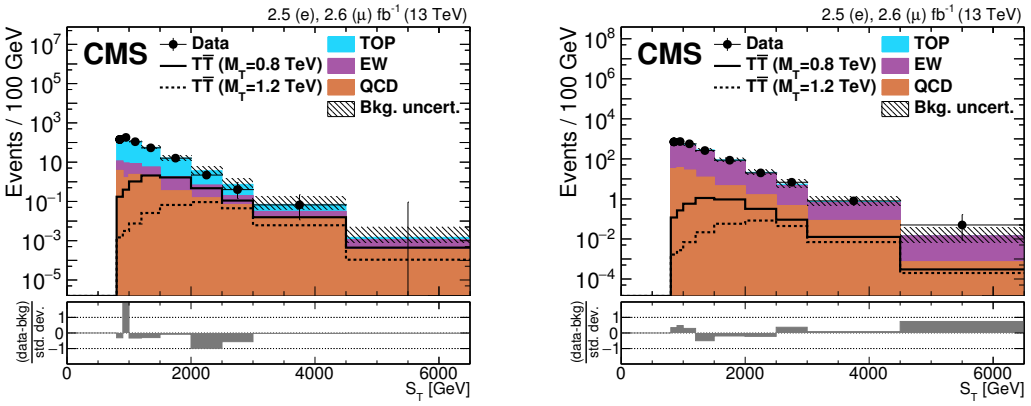

Figure 6.28: Distributions of S_{T} in the $t\bar{t}$ + jets (left) and W + jets (right) control regions after applying the H_{T} reweighting procedure. The $T\overline{T}$ signal is normalized to the theoretical cross section and the electroweak isospin singlet branching fraction scenario is assumed with a branching fraction of 50% for the $T \to bW$ decay and 25% for the $T \to tH$ and $T \to tZ$ decays (see also Section 2.4.1). The bottom pad shows the difference between data and simulated background in each bin in units of one standard deviation of the total uncertainty after combining the statistical uncertainties in the data and the total uncertainty (statistical and systematic) in the simulated background. For illustration purposes, several background processes are grouped together: the "TOP" group incorporates $t\bar{t}$ + jets and single top background processes, "EWK" includes W + jets, DY+ jets and diboson production and "QCD" denotes the background from multijet production. Plots are shown as published in Ref. [2].

To estimate whether potential presence of signal could significantly bias the fit and modify the derived scale factors, various signal injection tests are performed in which the linear fit is performed several times, with different signal hypotheses injected into the data

in the control regions. These studies are summarised in App. C.3. The potential changes are found to be negligible compared to the systematic uncertainty in the H_{T} reweighting procedure.

As the H_{T} reweighting procedure is intended to correct the simulation of the hard scattering process in the high-H_{T} phase space, uncertainties in the $\mu_{\mathrm{R/F}}$ energy scales and the PDFs are not considered in the final statistical evaluation of the data. These uncertainties are, however, sufficiently covered by the H_{T} reweighting uncertainty, especially in the high-S_{T} region. This is shown in Fig. 6.29 which provides a comparison of the templates corresponding to the up- and down-variations by one standard deviation of the systematic uncertainties in the PDFs, $\mu_{\mathrm{R/F}}$ energy scales and the H_{T} reweighting procedure. This shows how the uncertainty in the H_{T} reweighting procedure is the dominant uncertainty in the tails of the S_{T} distribution.

6.5.3 Final estimation of normalisation uncertainties

Remaining differences between data and simulation after the H_{T} reweighting procedure, especially in the low-S_{T} regions, are accounted for by assigning additional normalisation uncertainties to the $t\bar{t}$ + jets and W + jets background processes. These uncertainties are determined from a binned MLE fit of the background-only hypothesis in the two control regions with the H_{T} reweighting procedure applied, similar to the one described in Section 6.5.1. The systematic uncertainty in the H_{T} reweighting procedure is included as nuisance parameter in this fit while the $\mu_{\mathrm{R/F}}$ and PDF uncertainties for the $t\bar{t}$ + jets and W + jets background processes are not considered. Apart from this, all systematic uncertainties described in Section 6.4 are included as nuisance parameters as well.

Additionally, for the $t\bar{t}$ + jets and W + jets background processes, log-normal uncertainties in their normalisations with flat priors are assigned so that the corresponding nuisance parameters can float freely in the fit. The resulting constrained uncertainties in these parameters are then defined as the final systematic uncertainties in the rates of these two background processes.

The results of the MLE fit are shown in Fig. 6.30. It can be seen that after applying the H_{T} reweighting procedure, the post-fit values of all nusiance parameters show a healthy behaviour with no value lying outside of one standard deviation of its prior uncertainty. The resulting post-fit uncertainties in the normalisation parameters for the $t\bar{t}$ + jets and W + jets background processes are 8.7% and 6%, respectively. These uncertainties are then assigned as uncertainties in the normalisation of the corresponding processes in the final statistical interpretation of the results described in Section 6.6.

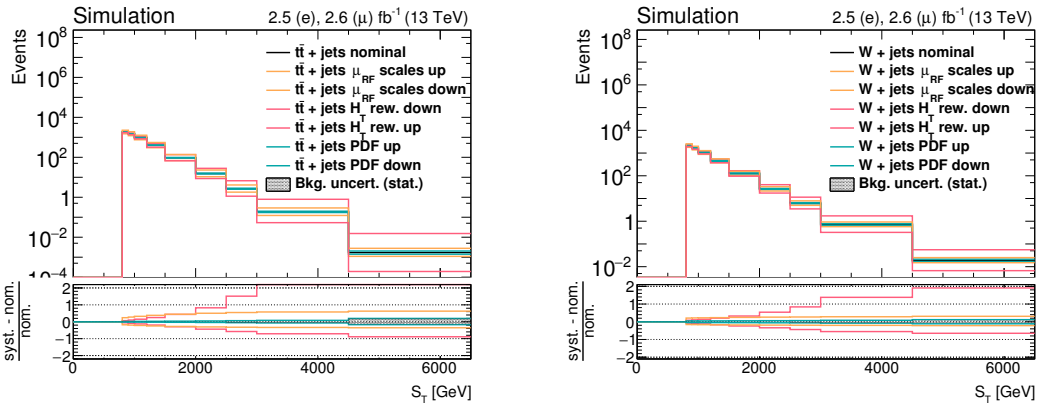

Figure 6.29: Comparisons of the templates for the up- and down-variations of the systematic uncertainties in the H_{T} reweighting procedure, $\mu_{\mathrm{R/F}}$ scales and PDFs. These are shown for the $t\bar{t}$ + jets background (left) and W + jets background (right). The ratio between the individual shifts and the nominal distribution per bin is shown in the bottom pads of each distribution.

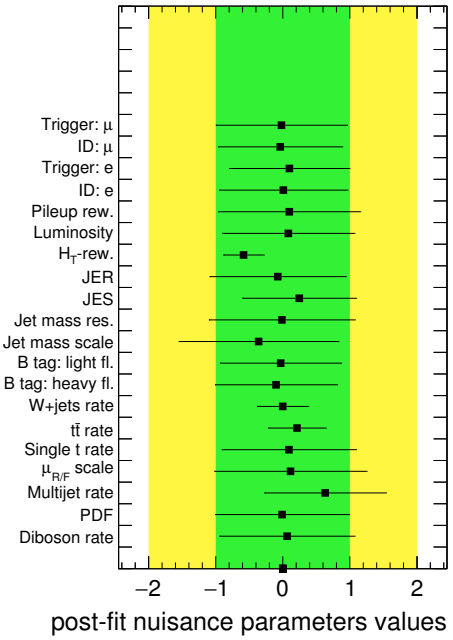

Figure 6.30: Post-fit nuisance parameters values returned by the MLE fit of the background-only hypothesis in the $t\bar{t}$ + jets and W+jets control regions after applying the H_{T} reweighting procedure. The fit parameters are expressed as deviations from their pre-fit values (corresponding to a value of zero) in units of the corresponding prior uncertainty. Not all nuisance parameters affect all background processes at the same time: the systematic uncertainty in the H_{T} reweighting procedure only affects the $t\bar{t}$ + jets and W + jets background processes while the uncertainties in the PDFs and $\mu_{\mathrm{R/F}}$ energy scales are only considered for DY+ jets and single top background processes.

6.6 Results and statistical interpretation

Distributions of S_{T} in the three event categories (see Section 6.3.4) after applying the H_{T} reweighting procedure are shown are shown in Fig. 6.31. The signal in some of these and the following distributions are shown assuming the branching fractions for the scenario that the T quark appears in an electroweak isospin singlet (see also Section 2.4.1). For the interpretation of the results, also the isospin doublet scenario is considered. The isospin singlet scenario corresponds to a branching fraction of 50% for the $T \to bW$ decay and 25% for the $T \to tH$ and $T \to tZ$ decays while for an electroweak isospin doublet, these branching fractions would be 50% to the $T \to tZ$ and $T \to tH$ decays (and 0% for the $T \to bW$ decay).

No significant excess of the data above the SM expectation is observed in either of the three categories which would hint at the presence of $T\overline{T}$ signal.

Distributions of the two variables most sensitive to $T \to tH$ decays that are used in the H-tagging algorithm are shown in Fig. 6.32 for events that pass the preselection after applying the H_{T} reweighting procedure. These are the number of b-tagged subjets for the p_{T}-leading Higgs boson candidate jet (see Section 6.3.3) and the mass of the p_{T}-leading AK8 jet with two b-tagged subjets before the mass requirement in the H-tagging algorithm. The $T\overline{T}$ signal is split into two curves: the solid curve shows $T\overline{T}$ events where at least one Higgs boson is present in the decay chain and the dashed curve shows $T\overline{T}$ events without any $T \to tH$ decays. Good agreement is found between the data and simulated background processes in both of these distributions.

Event yields in all three categories are listed in Table 6.10 for data and simulated processes. The uncertainties in the simulated processes shown in this table represent the total uncertainty in each process after combining statistical and the systematic uncertainties constrained to their post-fit values after applying the MLE fit described in Section 6.5.3.

From the observed data in the three event categories, upper limits on the cross section of $T\overline{T}$ production at 95% C.L. are calculated. These are computed as a Bayesian credible with a lower bound at zero. The details of this method are described in App. A.2. The S_{T} distributions in all three event categories are used as in the corresponding likelihood fit. The electron and muon channel are considered separately, resulting in a total of six independent categories in the final likelihood fit. The binning in the S_{T} distribution is chosen to ensure that the statistical uncertainty in the MC does not exceed 30% in any bin which is necessary to ensure the validity of the Barlow-Beeston light method [228, 229] to treat the statistical uncertainty in the simulation (see also App. A.2). The signal is assigned a flat prior and all systematic uncertainties are included as nuisance parameters with Gaussian priors for uncertainties that affect the shape of the S_{T} distribution and log-normal priors for uncertainties in the normalisation of simulated processes.

Expected and observed upper limits at 95% C.L. on the $T\overline{T}$ production cross section

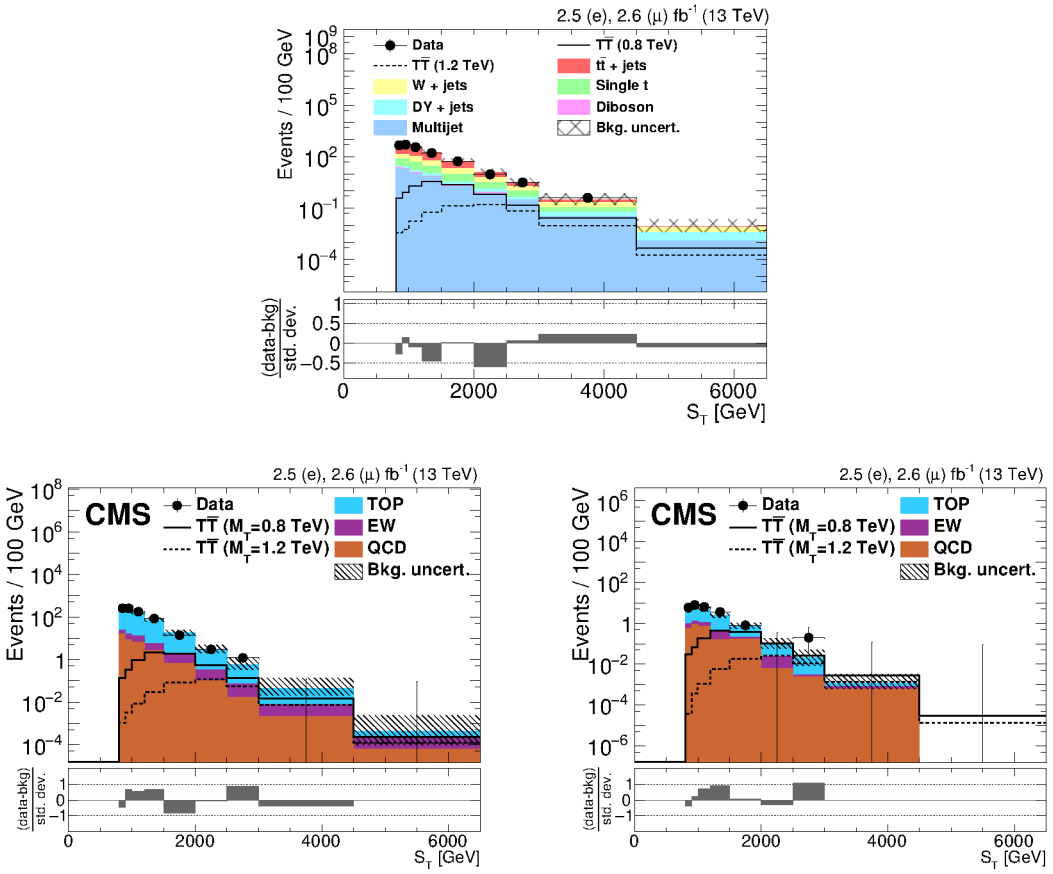

Figure 6.31: Distributions of S_{T} in the `0H` (top), `H1b` (left) and `H2b` (right) categories after combining electron and muon channel. The bottom pad shows the difference between data and simulated background in each bin in units of one standard deviation the total uncertainty after combining the statistical uncertainties in the data and the total uncertainty (statistical and systematic) in the simulated background. The distributions in the `H1b` and `H2b` categories are shown as published in Ref. [2]. The $T\overline{T}$ signal in all distributions is normalised to the theory cross section. In the `0H` category, a branching fraction of 100% is assumed for the $T \to tH$ decay. In the `H1b` and `H2b` categories, the branching fractions for the electroweak isospin singlet scenario (see Section 2.4.1) are assumed. For illustration purposes, several background processes are combined grouped together in the `H1b` and `H2b` categories: the "TOP" group incorporates $t\bar{t}$ + jets and single top background processes, "EWK" includes W + jets, DY + jets and diboson production and "QCD" denotes the background from multijet production.

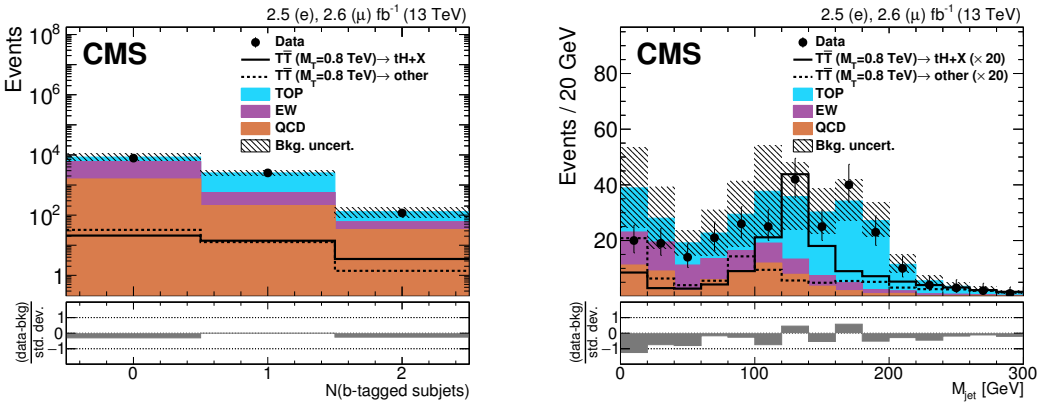

Figure 6.32: Distributions of the number of b-tagged subjets of the p_{T}-leading Higgs boson candidate jet (left), and the groomed jet mass of the p_{T}-leading AK8 jet with $p_{\mathrm{T}} > 300\,\mathrm{GeV}$ and two b-tagged subjets (right). The T quark signal is normalised to the predicted cross section, with the branching fractions corresponding to the electroweak isospin singlet scenario (see Section 2.4.1). The bottom pad shows the difference between data and simulated background in each bin in units of one standard deviation the total uncertainty after combining the statistical uncertainties in the data and the total uncertainty (statistical and systematic) in the simulated background. For illustration purposes, several background processes are combined grouped together: the "TOP" group incorporates $t\bar{t}$ + jets and single top background processes, "EWK" includes W + jets, DY + jets and diboson production and "QCD" denotes the background from multijet production. Plots are shown as published in Ref. [2].

Process	0H category	H1b category	H2b category
$\mathrm{T\bar{T}}$ (0.8 TeV)	45.7 ± 3.6	22.1 ± 2.1	4.6 ± 0.9
$\mathrm{T\bar{T}}$ (1.2 TeV)	3.1 ± 0.3	1.5 ± 0.2	0.32 ± 0.06
$\mathrm{t\bar{t}}$ + jets	1864.3 ± 173.1	1021.8 ± 122.3	28.2 ± 6.4
W + Jets	417.5 ± 31.8	33.2 ± 6.6	2.0 ± 0.7
DY + Jets	33.6 ± 8.6	11.8 ± 3.6	0.5 ± 0.3
Multijet	150.8 ± 143.4	79.5 ± 81.7	7.0 ± 7.5
Single t	272.8 ± 50.2	83.3 ± 18.6	2.9 ± 1.5
Diboson	13.4 ± 4.7	1.1 ± 0.8	0.0 ± 0.1
Total background	2752 ± 342	1230 ± 188	41 ± 13
Data	2556	1204	43

Table 6.10: Event yields in the three event categories after combining electron and muon channels. Uncertainties include statistical and all systematic uncertainties listed in Table 6.9 constrained to their post-fit values after applying the MLE fit described in Section 6.5.3. For the $T\overline{T}$ signal, the theoretically predicted production cross section and a branching fraction of 100% for the $T \to tH$ decay is assumed.

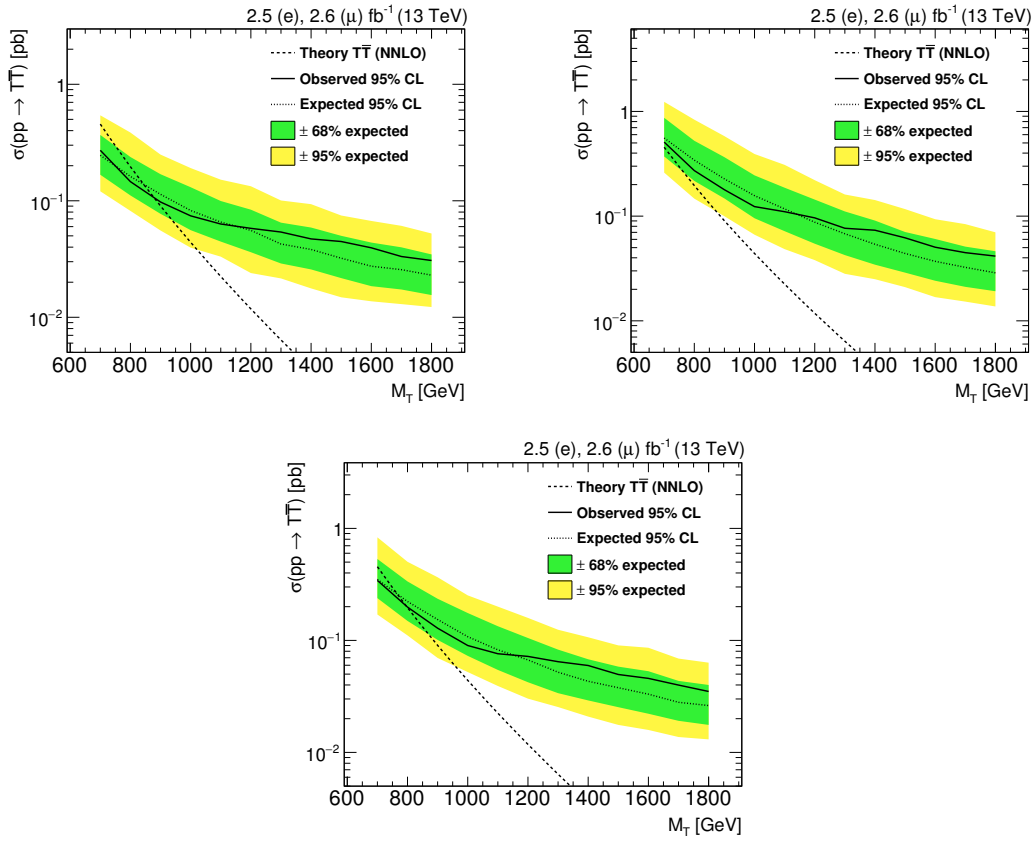

Figure 6.33: Upper 95% C.L. limits on the cross section of T quark pair production assuming a branching fraction of 100% for the $T \to tH$ decay (top left), for the branching fractions of the isospin singlet scenario (top right) and for the branching fractions of the isospin doublet scenario (bottom). The theoretical cross section is displayed as dashed line and is computed at NNLO (see Table 2.1).

are calculated assuming various branching fraction scenarios for the T quark. The three most important benchmark points considered are the branching fractions for the isospin singlet and the doublet models and a branching fraction of 100% for the $T \to tH$ decay. Limits for other branching fraction combinations are obtained by varying the branching fractions to the three decay modes tH, tZ and bW by 20% while respecting the condition that the sum of all three branching fractions is always 100%.

The expected and observed upper cross section limits for the three benchmark points can be found in Fig. 6.33. The seemingly correlated upwards shift of the observed limit above the expected limit for T quark masses greater than 1200 GeV can be explained by one data point in the `H2b` category in the bin between $S_{\mathrm{T}} = 2500\,\mathrm{GeV}$ and $S_{\mathrm{T}} = 3000\,\mathrm{GeV}$. Overall, the observed limit is contained within the band corresponding to one standard deviation of the expected limit, confirming the compatibility of the observed data with the standard model expectation.

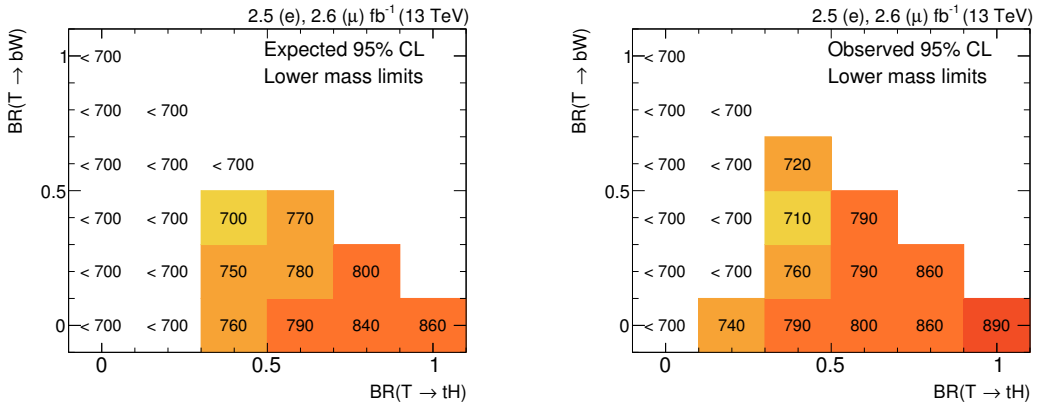

Figure 6.34: Lower 95% C.L. limits on the mass of the T quark for various branching fraction combinations. Expected limits are shown on the left, observed limits on the right. Each point in the plane represents a different branching fraction scenario for the T quark decay where the x-axis denotes the branching fraction of the $T \to tH$ decay and the y-axis the branching fraction of the $T \to bW$ decay. The remainder to one is then the branching fraction for the $T \to tZ$ decay.

These limits are compared to the theoretically predicted cross section for $T\overline{T}$ production calculated at NNLO. As a result, pair production of T quarks can be excluded for T quark masses below 890 GeV (860 GeV expected) assuming a branching fraction of 100% for the $T \to tH$ decay and 790 GeV (770 GeV) for the electroweak isospin doublet scenario. For the isospin singlet scenario, the sensitivity of this search is not large enough to exclude T quarks with masses above 700 GeV.

Lower limits on the mass of the T quark, up to a precision of 10 GeV, for all possible branching fraction scenarios can be found in Fig. 6.34. This search becomes more sensitive the higher the branching fraction of the $T \to tH$ decay becomes. As $T\overline{T}$ production is only considered for T quark masses above 700 GeV, exact mass limits can only be calculated where computed cross section limits and the theoretically predicted cross section intersect at values above 700 GeV.

The observed data distributions can also be interpreted for models that predict $B\bar{B}$ production. This search could be sensitive to this production mode e. g. for decays of the $B\bar{B}$ system to $bHtW$. Upper cross section limits for the electroweak isospin singlet and doublet scenarios (with branching fractions of 50% to tW and 25% to bH and bZ for the isospin singlet scenario and 50% to both bH and bZ for the isospin doublet scenario) are shown in Fig. 6.35. The sensitivity is not yet large enough to exclude $B\bar{B}$ production for any of these scenarios for B quark masses above 700 GeV.

A preliminary version of these results has been published by the CMS Collaboration in a Physics Analysis Summary [1]. The results shown in that reference differ slightly from the ones presented in this chapter as few modifications have been introduced since

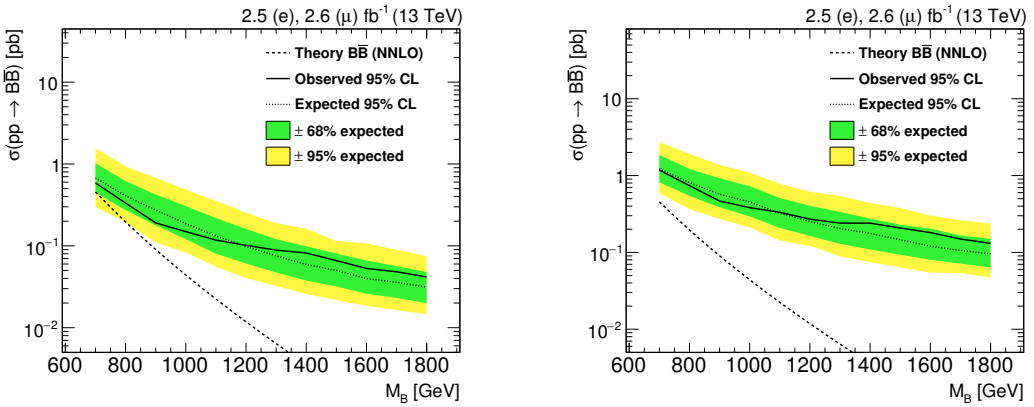

Figure 6.35: Upper 95% C.L. limits on the cross section of B quark pair production assuming the branching fractions for the isospin singlet scenario (left) and for the isospin doublet scenario (right). The theory cross section is shown as dashed line and is computed at NNLO.

then. To improve the sensitivity to branching fraction scenarios for which this search is not sensitive, this search is combined with a complementary search targeting $T \to bW$ decays. This combination and its results are described in more detail in Chapter 7. A more detailed discussion of the results derived in this thesis follow at the end of Chapter 7 where the results of the combined $T\overline{T}$ search are also compared to other results from the CMS and ATLAS Collaborations.

7 Combination with single-lepton search for $T\bar{T} \to bW + X$

To expand the sensitivity of the search presented in Chapter 6 to decay modes of the T quark other than tH, it is combined with a complementary search which was released as preliminary result by the CMS Collaboration in Ref. [230]. The search presented in this reference also targets vector-like $T\overline{T}$ production in the single-lepton final state, but vetoes on H-tagged AK8 jets and uses a technique called W tagging in order to identify hadronic $T \to bW$ decays.

The search presented in Ref. [230] provides the basis for the so-called *boosted W* channel in the combined search while the search presented in Chapter 6 forms the *boosted H* channel. This combined search is also sometimes simply referred to as *combination* in the following. Section 7.1 describes the event selection in the boosted W channel which follows for a large part the selection in [230]. Section 7.2 outlines the combination strategy and describes e. g. the treatment of systematic uncertainties across the two channels and Section 7.3 presents the results of the combined search.

7.1 Event selection in the boosted W channel

The analysis strategy in the boosted W channel has been developed by a different analysis group within the CMS Collaboration and released individually as preliminary result in Ref. [230]. This section provides a short summary of the event selection and analysis techniques used in the boosted W channel to allow a more detailed understanding of the results of the combination.

The purpose of the boosted W channel is to select $T\overline{T}$ events where the T quark has a large branching fraction for the $T \to bW$ decay. Events must pass a single electron or single muon trigger that selects electrons (muons) with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ (20 GeV), $|\eta| < 2.1$ (2.4) that are isolated from nearby hadronic activity.

After the trigger selection, exactly one electron or muon with a $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ and $|\eta| < 2.1$ or 2.4, respectively, is required. An electron must pass several identification criteria based on a multi-variate identification technique while a muon must fulfill the selection criteria described in Section 5.3 and additional requirements to further reduce the contribution from misidentified muons. Furthermore, for both lepton flavours, a *mini-isolation* requirement has to be fulfilled where hadronic activity is measured in a cone around the lepton with a variable cone size depending on the p_{T} of the lepton. Events

with more than one lepton, where the additional leptons pass looser selection criteria, are rejected in order to combine this search with a dedicated multi-lepton search in the future. Three or more AK4 jets are required that must have transverse momenta greater than 300, 150 and 100 GeV, respectively. The missing transverse energy in the event has to surpass 75 GeV and $\Delta R > 1$ between the lepton and the jet with the second highest p_{T} is required. Finally, events must not contain any H-tagged AK8 jet with at least one b-tagged subjet using the same H tagging definition as described in Section 6.3.3. This requirement is important to ensure the orthogonality of the boosted W and boosted H channels which allows for a straightforward combination of the two channels.

To identify boosted $T \to bW$ decays, a technique called W *tagging* is employed which is similar to the H tagging algorithm used in the boosted H channel and aims to identify hadronic boosted W boson decays. In order to be W-tagged, an AK8 jet has to have a pruned mass between 65 and 105 GeV, where pruning [196] is a grooming algorithm similar to the soft-drop algorithm, and must fulfill a N-subjettiness requirement $\tau_2/\tau_1 < 0.6$ (see Section 5.8).

Simulated events are corrected with similar or identical scale factors to match measured efficiencies in the data as events in the boosted H channel, e. g. for b tagging and trigger efficiencies (see the descriptions in Chapter 5). Most of these scale factors were derived from the same measurements. In addition, as similar discrepancies between data and simulation have been observed in control regions in the boosted W channel, events in the boosted W channel are also corrected with the H_{T} reweighting procedure described in Section 6.5.2, using the same parameters.

The background estimation in the boosted W channel relies on MC simulation in the same way as the boosted H channel. To evaluate its modelling, four separate control regions are defined by inverting the ΔR requirement described above and selecting events with a ΔR(lepton, 2nd jet) < 1, and requiring different numbers of b-tagged AK4 jets and W-tagged AK8 jets. Two of these control regions are enriched in $t\bar{t}$ + jets events by requiring either exactly one or ≥ 2 b-tagged AK4 jets, the other two control regions are enriched in W + jets events by requiring exactly zero b-tagged AK4 jets and either zero or ≥ 1 W-tagged AK8 jets, respectively. After applying all corrections to the simulation, including the H_{T}-reweighting, good agreement between data and simulation is observed. Systematic uncertainties in the normalisation of the two background processes are assigned to account for remaining differences in the event yields between data and simulation. Other sources of systematic uncertainties are mostly identical to the ones considered in the boosted H channel described in Section 6.4. More details about these systematic uncertainties, and how correlated uncertainties between the two channels are treated, are presented in Section 7.2.

After passing the selection requirements described above, events are categorised accord-

Sample	**0 W, 0 b**	**0 W, 1 b**	**0 W, 2 b**	**0 W, ≥ 3 b**
$T\overline{T}$ (0.8 TeV)	2.49 ± 0.68	5.3 ± 1.3	3.9 ± 1.0	1.42 ± 0.38
$T\overline{T}$ (1.2 TeV)	0.228 ± 0.064	0.42 ± 0.11	0.261 ± 0.067	0.093 ± 0.025
TOP	103 ± 41	205 ± 78	111 ± 41	16.3 ± 6.8
EWK	460 ± 160	80 ± 30	10.7 ± 4.0	0.59 ± 0.25
QCD	14.1 ± 6.3	6.2 ± 3.7	< 1	< 1
Total Bkg	570 ± 170	292 ± 84	122 ± 41	16.9 ± 6.8
Data	588	288	131	14
Sample	**≥ 1 W, 0 b**	**≥ 1 W, 1 b**	**≥ 1 W, 2 b**	**≥ 1 W, ≥ 3 b**
$T\overline{T}$ (0.8 TeV)	3.26 ± 0.89	6.6 ± 1.7	4.2 ± 1.1	1.02 ± 0.27
$T\overline{T}$ (1.2 TeV)	0.335 ± 0.092	0.52 ± 0.13	0.269 ± 0.070	0.063 ± 0.017
TOP	71 ± 26	111 ± 42	56 ± 20	7.6 ± 3.3
EWK	180 ± 50	29.0 ± 8.4	4.4 ± 2.0	0.25 ± 0.11
QCD	12.6 ± 7.0	3.5 ± 2.6	0.15 ± 0.18	< 1
Total Bkg	263 ± 57	143 ± 43	60 ± 20	7.8 ± 3.3
Data	274	155	45	7

Table 7.1: Data and expected background yields in each category of the boosted W channel after combining electron and muon channels. Uncertainties include statistical and all systematic components summarised in Table 7.2. For the $T\overline{T}$ signal, the theoretically predicted production cross sections are assumed with the singlet branching fraction combination. The labels "TOP", "EWK" and "QCD" denote three groups of background processes. Taken from Ref. [2].

ing to the number of W-tagged AK8 jets and b-tagged AK4 jets passing the medium working point (see Section 5.7). First, events are split into two categories depending on whether they have at least one W-tagged AK8 jet or not. These two categories are then further divided into four subcategories with 0, 1, 2 or ≥ 3 b-tagged AK4 jets, respectively, resulting in a total of eight categories. The background composition differs from category to category with the main background contributions arising from $t\bar{t}$ + jets and W + jets production, similar to the boosted H channel.

For illustration purposes these background processes are combined with some of the minor backgrounds in the following figures and tables, forming in total three groups of backgrounds: the "TOP" group incorporates $t\bar{t}$ +jets and single top background processes, "EWK" includes W + jets, DY+ jets and Diboson production and "QCD" denotes the background from multijet production. The data and predicted background yields in each of these eight categories is shown in Table 7.1. It can be seen that background processes in the EWK group are the dominant contribution in categories with low b tag and W tag multiplicities, while background processes from the TOP group grow more and more important for categories with higher numbers of b tags and W tags. These are also the categories with the highest signal-to-background ratio.

To discriminate signal-like from background-like events, a variable "min[M(ℓ, b)]" is

defined as the smallest invariant mass between the lepton and any of the b-tagged AK4 jets. In categories with zero b-tagged AK4 jets, the minimum invariant mass between the lepton and any AK4 jet, "min[M(l, j)]", is used. For the $t\bar{t}$ + jets background, the distribution of this variable peaks around the t quark mass while for $T\overline{T}$ signal events with the lepton from a $T \to bW$ decay, this distribution rises towards large values for min[M(ℓ, b)]. Distributions of min[M(ℓ, b)] in all eight categories are shown in Fig. 7.1.

7.2 Combination strategy and treatment of systematic uncertainties

In the final statistical interpretation, all categories from the boosted W channel and the `H1b` and `H2b` categories from the boosted H channel (see Section 6.3.4) are combined in a simultaneous likelihood fit. The `0H` category is not considered as this has a large overlap with the boosted W channel and is covered by the corresponding categories.

Combining the boosted W and boosted H channels provides a good coverage of many possible final states of the $T\overline{T}$ system. Each of the two channels provides a good sensitivity for signal models with large branching fractions to tH and bW decay modes, respectively, and serves as a background enriched region to constrain systematic uncertainties in the opposite case. There is also some sensitivity for $T \to tZ$ decays if the Z boson decays hadronically as it can then either be H-tagged, if it decays to $b\bar{b}$, or W-tagged if it decays to any other pair of quarks. Splitting each category into an electron and muon channel results in 16 categories in the boosted W channel and four categories in the boosted H channel, so that there is a total number of 20 categories.

A large number of the systematic uncertainties that affect simulated samples in the boosted H channel are also assigned in the boosted W channel. If not mentioned otherwise, these uncertainties are treated as fully correlated between the two channels as they are considered to originate from the same source and are estimated based on identical or very similar measurements. There are, however, a few differences which are outlined in this section.

Due to different versions of the software used to reconstruct physical objects in the two channels, slightly different b tagging efficiencies are found between them. As a result, the H tagging efficiencies for signal, when requiring at least one b-tagged subjet, differ between the two channels by a factor of approximately 3%. Since the veto on an H-tagged jet in the boosted W channels defines the orthogonality between the two channels, this can lead to a bias of the combined signal efficiency of the two channels. This effect is therefore taken into account by assigning a systematic uncertainty of 3% on the normalisation of the signal in the two channels. As a variation of the H tagging efficiency would lead to opposite changes of the signal yield in the two channels (since, in case of an up-variation,

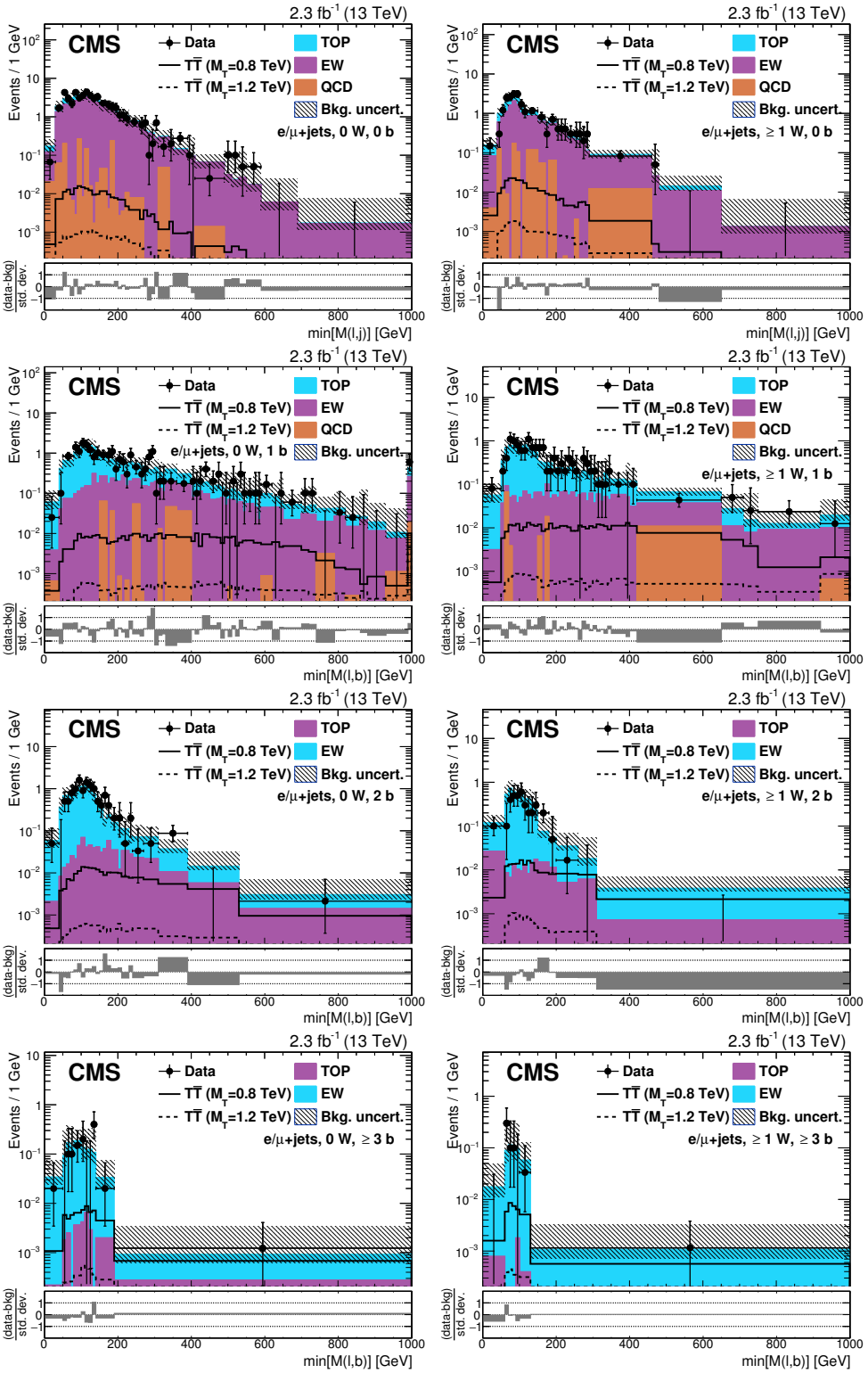

Figure 7.1: The bottom pad shows the difference between data and simulated background in each bin in units of one standard deviation of the total uncertainty after combining the statistical uncertainties in the data and the total uncertainty (statistical and systematic) in the simulated background. For illustration purposes, several background processes are combined grouped together: the "TOP" group incorporates $t\bar{t}$ + jets and single top background processes, "EWK" includes W + jets, DY+ jets and Diboson production and "QCD" denotes the background from multijet production. Taken from Ref. [2].

more signal events would enter the boosted H channel and more would be vetoed in the boosted W channel), this systematic uncertainty is treated as fully anti-correlated between the two channels.

Due to this difference in the H tagging efficiency, it is also possible that data events are selected in both channels at the same time. The number of such "overlapping" data events was explicitly checked and 24 out of in total 2749 events were found to be selected in both channels. As all of these events are in the less sensitive `H1b` category of the boosted H channel, the impact of this overlap is deemed negligible and no further action is taken to prevent it.

In the boosted W channel, an additional systematic uncertainty is assigned due to measured differences in the efficiency of the mini-isolation requirement for leptons between data and simulation. The uncertainty in this measurement is rather small and as a result a systematic uncertainty on the rate of all simulated samples of 1% is assigned in the boosted W channel. The systematic uncertainties in the trigger selection efficiencies are treated as fully correlated even though different triggers are used in the two channels. Since the triggers in the boosted W channel only differ by the additional lepton isolation requirement and a lower p_{T} threshold, the data samples used to measure these efficiencies have a very large overlap meaning that the measurements of the efficiencies for both triggers are not statistically independent.

Furthermore, as the jet mass scale and resolution has been measured to give similar values for jets groomed with the soft-drop and with the pruning algorithms [202], the associated systematic uncertainties are treated as fully correlated across the two channels of the combination. The boosted W channel does, however, not require an additional uncertainty in the groomed mass correction like that applied in the boosted Higgs channel.

The selection efficiency of the τ_2/τ_1 requirement used in the W tagging algorithm has also been measured in data and compared with simulation [185] to derive scale factors. The uncertainty in this measurement is propagated to the observables in the boosted W channels, affecting both their shapes and rates. An additional systematic uncertainty is assigned when extrapolating the measured scale factors to AK8 jets that have a larger p_{T} ($> 300\,\mathrm{GeV}$) than the jets in the sample used for the measurement. This uncertainty is derived by comparing selection efficiencies between simulated samples that were generated with the Pythia and Herwig++ hadronisers and also affects both the event rate and the shape of physical observables.

As mentioned in the previous section, systematic uncertainties in the normalisation of the $t\bar{t}$ + jets and W + jets background processes are estimated from control regions and assigned in the final statistical interpretation. For minor background processes coming from DY + jets, single top and Diboson production, the same systematic uncertainties in their rate as in the boosted H channel (described in Section 6.4.2) are assumed. All these

normalisation uncertainties are treated as fully correlated between the two channels.

To estimate the background from multijet production, simulated samples are used that were produced with the MadGraph generator and have sufficient information stored to estimate systematic uncertainties in the $\mu_{\mathrm{R/F}}$ energy scales and PDFs used to generate these samples. As these uncertainties are rather large, no additional systematic uncertainty in the rate of multijet production is assumed in the boosted W channel.

As the $t\bar{t}$ + jets and W + jets samples in the boosted W channel are corrected with the same H_{T} reweighting procedure used in the boosted H channel (see Section 6.5.2), the systematic uncertainty associated with this procedure is also assigned in the boosted W channel. As the observed trends are most likely due to the same effect in the two search channels, this uncertainty is treated as fully correlated between the two channels.

Finally, systematic uncertainties due to the $\mu_{\mathrm{R/F}}$ energy scales and PDFs used in the generation of DY+jets, single top and all signal samples are treated in the same way as for the boosted H channel (see Section 6.4.2) and are correlated between the two channels. For backgrounds from $t\bar{t}$ + jets and W + jets production, no uncertainty is assigned to account for variations in the PDFs used to simulate these samples as this is assumed to be covered by the large systematic uncertainty in the H_{T} reweighting procedure. The systematic uncertainty in the $\mu_{\mathrm{R/F}}$ energy scales is, however, assigned as its effect on the min[M(ℓ, b)] variable is of the same order as the uncertainty in the H_{T} reweighting procedure.

A summary of all systematic uncertainties and how they are treated between the two channels can be found in Table 7.2. In order to test there are no differences in the treatment of systematic uncertainties between the two channels, an MLE fit of the background-only hypothesis is performed simultaneously in all 20 categories (see App. A.1 for a description of this method). Systematic uncertainties that affect only the rates of physical processes are assigned log-normal priors in this study while all other systematic uncertainties are assigned Gaussian priors. The results of this test can be found in App. E which confirm the validity of the treatment of systematic uncertainties in this combination.

7.3 Results and statistical interpretation

No significant excess above the standard model background expectation is observed in any of the categories in the boosted W and boosted H channels. Upper cross section limits on $T\overline{T}$ production at 95% C.L. are calculated as Bayesian intervals (see App. A.2) with a lower bound at zero by performing a simultaneous binned likelihood fit of the min[M(ℓ, b)] variable in the boosted W categories and of the S_{T} variable in the boosted H channel to the data. The statistical treatment resembles the one described in Section 6.6: A flat prior is assigned to the signal cross section and systematic uncertainties are treated as nuisance parameters that are marginalised in the likelihood fit. Uncertainties affecting only the rates

Source	Experimental uncertainties		
	boosted W	boosted H	Correlated
Luminosity	✓	✓	✓
W tagging: τ_2/τ_1	✓	—	—
W tagging: τ_2/τ_1 p_{T}-dep.	✓	—	—
H tag: W $\to$ H	—	✓	—
Trigger eff.	✓	✓	✓
Lepton identification eff.	✓	✓	✓
AK8 jet mass scale	✓	✓	✓
AK8 jet mass res.	✓	✓	✓
Jet energy scale	✓	✓	✓
Jet energy res.	✓	✓	✓
b tag eff., h.-fl.	✓	✓	✓
b tag eff., l.-fl.	✓	✓	✓
b tagging: H tag eff. diff.	✓*	✓*	✓*
Source	**Normalisation uncertainties from control regions**		
	boosted W	boosted H	Correlated
$t\bar{t}$ + jets rate	✓	✓	✓
W + jets rate	✓	✓	✓
Source	**Experimental normalisation uncertainties**		
	boosted W	boosted H	Correlated
DY+ jets rate	✓	✓	✓
Diboson rate	✓	✓	✓
Single top rate	✓	✓	✓
Multijet rate	—	✓	—
Source	**Theoretical uncertainties**		
	boosted W	boosted H	Correlated
H_{T} reweighting	✓	✓	✓
Pileup reweighting	✓	✓	✓
PDF	✓	✓	✓
$\mu_{\mathrm{R/F}}$ scales	✓	✓	✓

*correlated to heavy-flavour b tagging uncertainty in boosted H channel, anti-correlated in boosted W channel

Table 7.2: Summary of correlated and uncorrelated uncertainties in the combination of the boosted W and boosted H channels. Not all uncertainties are applied to all samples within a category.

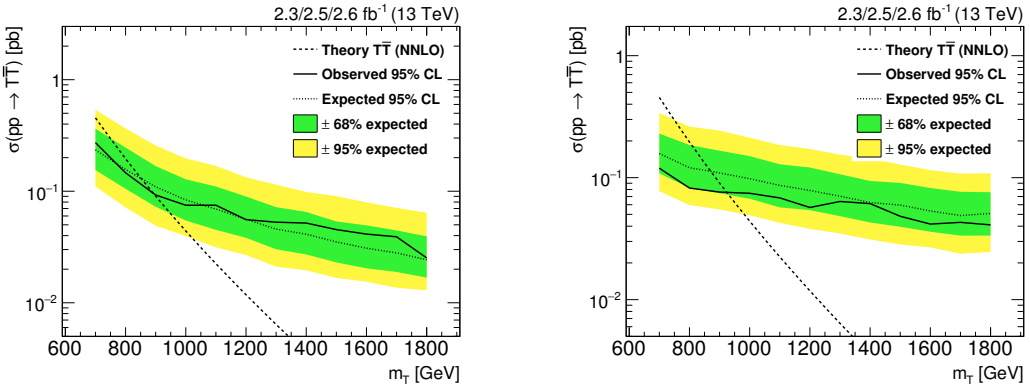

Figure 7.2: The expected and observed upper limits at 95% C.L. on the cross section of $T\overline{T}$ production assuming a branching fraction of 100% for the $T \to tH$ (left) and the $T \to bW$ (right) decay mode in the combination of the boosted W and boosted H channels. The limits are compared to the theoretically predicted cross section for $T\overline{T}$ production which is shown as dashed line.

of physical processes are assigned log-normal priors while all other systematic uncertainties are assigned Gaussian priors.

Upper limits on the cross section of $T\overline{T}$ production at 95% C.L. are shown in Figs. 7.2 and 7.3 for a few benchmark points of the T quark branching fraction. Assuming a branching fraction of 100% either for the $T \to bW$ decay or for the $T \to tH$ decay, T quarks can be excluded up to masses of 910 GeV (870 GeV expected) or 890 GeV (860 GeV), respectively. For the branching fractions corresponding to the electroweak isospin singlet or doublet scenario (see Section 2.4.1), the observed (expected) lower mass limits are 860 GeV (790 GeV) or 830 GeV (780 GeV), respectively. These branching fraction combinations are 50% to bW and 25% to both tZ and tH for in the isospin singlet scenario and 50% to both tZ and tH (and 0% to bW) for the isospin doublet scenario.

Expected and observed lower limits on the mass of the T quark for various possible branching fractions are shown in Fig. 7.4. Comparing these limits with the corresponding branching fraction scans for the individual channels of the combination, one can clearly see how the two channels complement each other so that, particularly, the sensitivity for $T\overline{T}$ production with sizable branching fractions to bW is improved.

The combination of the two search channels provides a significant improvement over the individual results of the boosted H channel (which are equal to the ones discussed in Section 6.6) for a wide range of assumed branching fractions of the T quark. For very high branching fractions for the $T \to tH$ decay close to 100%, the lower mass limits are completely driven by the boosted H channel alone (and equivalently by the boosted W channel for very high branching fractions for the $T \to bW$ decay).

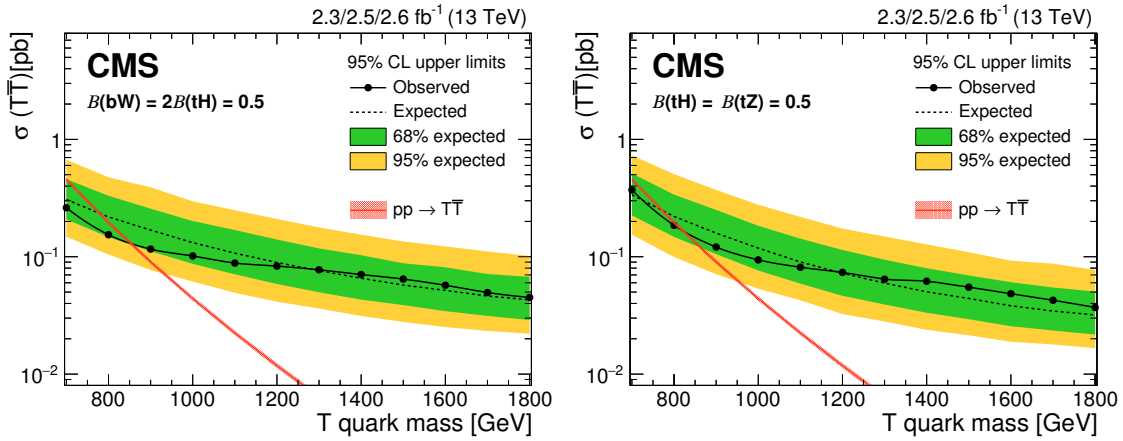

Figure 7.3: The expected and observed upper limits at 95% C.L. on the cross section of $T\overline{T}$ production for the electroweak isospin singlet (left) and doublet scenario (right) after combining the boosted W and boosted H channels. The limits are compared to the theoretically predicted cross section for $T\overline{T}$ production (calculated at NNLO) which is shown as red line, with the uncertainties in the PDFs and renormalisation and factorisation scales indicated by the shaded area. Published in Ref. [2].

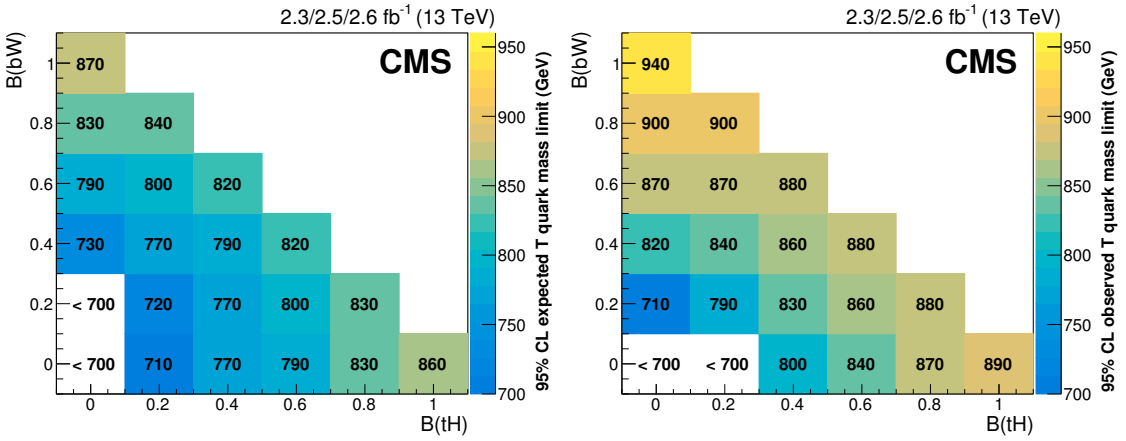

Figure 7.4: Lower 95% C.L. limits on the mass of the T quark for various branching fraction combinations obtained from the combination of the boosted W and boosted H channels. Expected limits are shown on the left, observed limits on the right. Each point in the plane represents a different branching fraction scenario for the T quark decay where the x-axis denotes the branching fraction of the $T \to tH$ decay and the y-axis the branching fraction of the $T \to bW$ decay. The remainder to one is then the branching fraction for the $T \to tZ$ decay. Published in Ref. [2].

Upper limits on the $B\bar{B}$ production cross section for the electroweak isospin singlet and doublet models can be found in Fig. 7.5 and expected and observed lower mass limits for a variety of branching fractions of the B quark are shown in Fig. 7.6. One can see

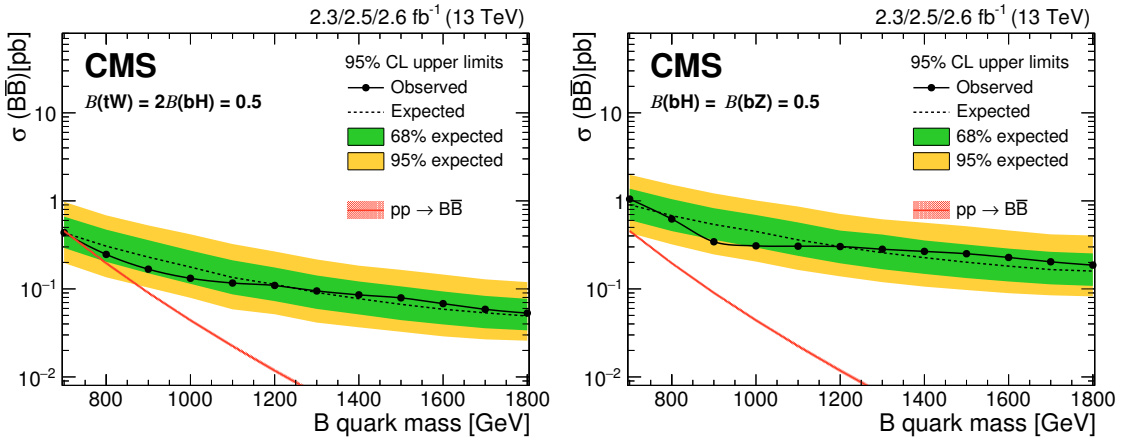

Figure 7.5: The expected and observed upper limits at 95% C.L. on the cross section of $B\bar{B}$ production for the electroweak isospin singlet (left) and doublet scenario (right) after combining the boosted W and boosted H channels. The limits are compared to the theoretically predicted cross section for $B\bar{B}$ production (calculated at NNLO) which is shown as red line, with the uncertainties in the PDFs and renormalisation and factorisation scales indicated by the shaded area. Published in Ref. [2].

that the sensitivity to this signal model is considerably enhanced by combining the two search channels. For the isospin singlet scenario, B quarks with masses below 730 GeV (720 GeV expected) are excluded while for the isospin doublet scenario, the sensitivity of the combined search is not high enough to exclude B quarks with masses above 700 GeV. The corresponding branching fractions are equivalent to the branching fractions to decay modes with the same heavy bosons of T quarks appearing in the same electroweak isospin multiplet (see also Section 2.4.1). This results in exclusion limits for B quark masses above 700 GeV for a significant number of branching fraction combinations, while in the boosted H channel alone, B quarks with masses above 700 GeV cannot be excluded.

The results presented in this chapter have been published by the CMS Collaboration and submitted to the Journal of High Energy Physics [2]. A comparison of these results to previous searches at $\sqrt{s} = 8$ TeV as well as to more recent results at 13 TeV follows in the next section.

7.4 Comparison to other VLQ searches

For $T\overline{T}$ quark production with large branching fractions of the T quark to tH, the observed and expected limits calculated in this search come close to the results of previous searches performed at the LHC.

For the branching fractions assuming that the T quark appears in an electroweak isospin singlet, the CMS Collaboration excluded T quarks with masses approximately

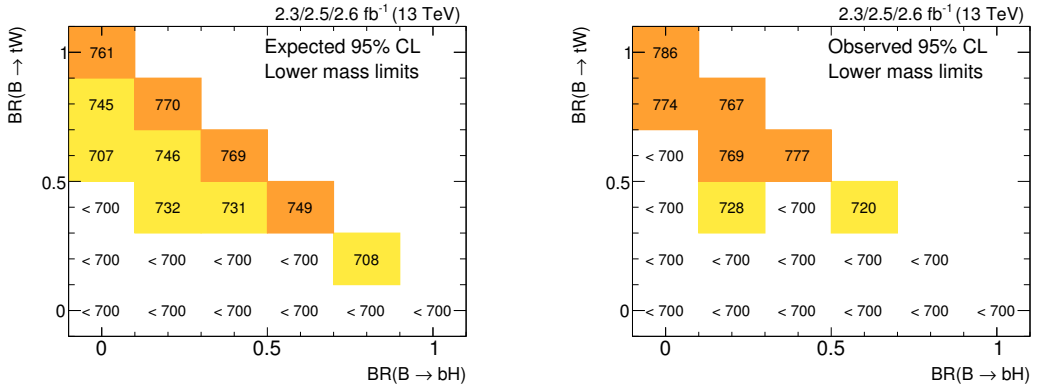

Figure 7.6: Lower 95% C.L. limits on the mass of the B quark for various branching fraction combinations obtained from the combination of the boosted W and boosted H channels. Expected limits are shown on the left, observed limits on the right. Each point in the plane represents a different branching fraction scenario for the B quark decay where the x-axis denotes the branching fraction of the $B \to bH$ decay and the y-axis the branching fraction of the $B \to tW$ decay. The remainder to one is then the branching fraction for the $B \to bZ$ decay.

below 730 GeV (800 GeV expected) at $\sqrt{s} = 8$ TeV, in a data set corresponding to $\mathcal{L} = 19.7\,\text{fb}^{-1}$ [99]. The lower mass limits for a T quark appearing in an isospin doublet found in the same search are 760 GeV (820 GeV expected). The observed limits on the T quark mass are extended in this search to 860 GeV and 830 GeV while the expected sensitivity is slightly lower than in the previous searches.

Assuming a branching fraction of 100% for the $T \to tH$ decay, an observed (expected) lower limit on the T quark mass at 95% C.L. was set to 770 GeV (840 GeV) by the CMS Collaboration at $\sqrt{s} = 8$ TeV with a data set corresponding to $\mathcal{L} = 19.7\,\text{fb}^{-1}$ [99]. This result is even surpassed by the result of this search which sets a lower mass limit of 890 GeV (860 GeV). A search carried out by the ATLAS collaboration at $\sqrt{s} = 8$ TeV studied a data set corresponding to $\mathcal{L} = 20.3\,\text{fb}^{-1}$ and quotes an observed (expected) lower mass limit of 950 (885) GeV [101] which is just slightly stronger than the result of this search. The relatively high limit on the T quark mass obtained in this search is primarily caused by a significant increase in the theoretical cross section for $T\overline{T}$ production from 8 TeV to 13 TeV e. g. by more than a factor of 10 for the M(T) = 900 GeV mass point [99].

Comparing the lower mass limits for vector-like B quarks with the results from the 8 TeV analyses both from ATLAS and CMS [100–102] shows that the 13 TeV results do not yet reach the sensitivity previously achieved: while in this search, lower limits on the B quark mass can only be set for few branching ratio combinations ranging from 720 to 790 GeV (710 to 780 GeV expected), the 8 TeV search from CMS excludes B quark masses up to values ranging from 740 to 900 GeV (710 to 890 GeV expected) for all possible branching

fraction scenarios of the B quark. This is, however, expected as this search has not been optimised for $B\bar{B}$ production so that dedicated searches for this signal mode would likely be able to compete with the 8 TeV results as well.

More recent results on pair production using pp collision data from the 2016 data-taking period at the LHC were released by the ATLAS collaboration. Several analyses were performed that were optimised for different decay modes of vector-like T quarks and used data sets corresponding to integrated luminosities between $\mathcal{L} = 13.3\,\text{fb}^{-1}$ [231–233] and $\mathcal{L} = 36.1\,\text{fb}^{-1}$ [234]. These searches set lower mass limits for the T quark of 1200 GeV (1160 GeV expected) for a branching fraction of 100% for the $T \to tH$ decay and 1160 GeV and 1020 GeV (1110 GeV and 960 GeV expected) for the branching fraction combination corresponding to the isospin singlet and doublet models, respectively. These limits are significantly stronger than the ones derived in the combination in this thesis as they are based on a much larger data sample. Including the full data set recorded in 2016 would therefore allow to exclude pair production of vector-like T quarks in a similar mass regime also with this search.

Exclusion limits on VLQ production were also set for other types of vector-like quarks besides T and B quarks and for other production modes. At $\sqrt{s} = 13\,\text{TeV}$, the CMS Collaboration excluded pair production of $X_{5/3}$ quarks for masses below 1160 GeV if the $X_{5/3}$ quarks are assumed to have only left-handed couplings and below 1100 GeV for exclusive right-handed couplings [235] (see also Section 2.4.1).

The single production mode of vector-like quarks has been studied as well in direct searches at the LHC. No evidence for this process has been observed yet and upper limits on the production cross section at 95% C.L. were derived by the ATLAS Collaboration at $\sqrt{s} = 8\,\text{TeV}$ [103–106] and by both the ATLAS and CMS Collaborations at $\sqrt{s} = 13\,\text{TeV}$ [202, 236–239]. A recent measurement by the CMS Collaboration using the LHC data set recorded in 2016 excludes single production of vector-like T quarks decaying to tZ for various assumptions of their couplings to SM quarks [240]. Lower limits are set on the T quark mass, depending on the assumed couplings to SM quarks, ranging from 1200 to 1400 GeV assuming it appears in an electroweak isospin singlet and from 850 to 950 GeV if it appears in an isospin doublet. The single production mode will become increasingly important in future searches to maintain a high sensitivity for very heavy vector-like quarks as then, the corresponding production cross section can possibly become considerably larger than the pair production cross section.

Based on the results of this search and the ones that were obtained with the data set recorded in 2016, it is most likely that the sensitivity to vector-like quarks will be significantly improved over the next few years. With the data set that the LHC is expected to deliver after the upgrade to the High Luminosity LHC, it is therefore well possible that this sensitivity can be expanded to VLQ masses around 1.5 TeV and maybe even beyond.

8 Conclusions

This thesis presents a search for pair-produced vector-like T quarks which was carried out using proton-proton collision data delivered by the LHC in 2015 at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. These data were recorded with the CMS detector and correspond to an integrated luminosity up to $\mathcal{L} = 2.6\,\text{fb}^{-1}$.

Vector-like T quarks appear in many extensions of the Standard Model in which the Higgs boson is a bound state of an unknown strong dynamics and their observation would give deep insights into the structure of physics beyond the Standard Model. In order to solve the problem of the unnaturally low electroweak scale, they must have masses of order 1 TeV which would make them observable at the LHC at $\sqrt{s} = 13\,\text{TeV}$. They are assumed to decay via three decay modes, bW, tZ and tH, with branching fractions depending on their transformation properties under the electroweak symmetry group. The search presented in this thesis was optimised for the $T \to tH$ decay by using dedicated substructure analysis techniques.

At least one muon or electron was required in the trigger selection step to suppress the contribution from multijet background processes. More kinematic requirements were imposed in an offline preselection to select events with a large number of high-p_T jets as such final states are characteristic for $T\overline{T}$ events. To this end, an observable denoted as S_T, defined as the p_T sum of the lepton, the missing transverse energy and the p_T of all AK4 jets, was employed.

Higgs bosons emitted in the decay of a T quark often possess a high transverse momentum due to the large mass difference between the two particles. The dominant decay mode of the Higgs boson, $H \to b\bar{b}$, was targeted and AK8 jets were used to capture all its decay products within a single jet to maximise the sensitivity to the presence of such high-p_T Higgs bosons. The softdrop grooming algorithm was used to remove soft and wide-angle radiation products within these AK8 jets and reconstruct subjets corresponding to the two b quarks from the Higgs boson decay. AK8 jets were then defined as H-tagged if they had at least one b-tagged subjet and a groomed jet mass close to the Higgs boson mass.

The two dominant SM background processes, $t\bar{t}$ + jets and W + jets production, were estimated from MC simulations and the corresponding modelling of the S_T observable was evaluated in two control regions. The observed S_T distributions showed trends in the ratio between data and simulation, exhibiting an overestimation of the data in the simulation towards high S_T values. These trends were found to be related to a mismodelling of the H_T variable, defined as the sum of the transverse momenta of all AK4 jets in an event. A

procedure denoted as H_{T} reweighting was introduced to correct for this mismodelling by reweighting simulated events based on their H_{T} using a linear function.

The events passing the preselection were split into three orthogonal event categories. Events with no H-tagged jet were sorted into the `0H` category, while events with at least one H-tagged jet were assigned into the `H1b` or `H2b` categories, depending on whether they had an H-tagged jet with one or two b-tagged subjets.

No significant excess of the data above the SM expectation was observed in either of the three event categories. Upper cross section limits at 95% C.L. were set by performing a simultaneous likelihood fit of the signal+background hypothesis to the data in all three event categories using the S_{T} variable. Systematic uncertainties in the simulation, of which the dominant contribution was found to be the uncertainty in the H_{T} reweighting procedure, were considered and marginalised using a Bayesian method. The resulting upper cross section limits were interpreted as lower limits on the T quark mass by comparing them to the theoretically predicted $T\overline{T}$ production cross section. Assuming a branching fraction of 100% for the $T \to tH$ decay, T quarks with masses below 890 GeV were excluded. Furthermore, a lower mass limit of 790 GeV was set assuming the electroweak isospin doublet scenario, in which the T quark has branching fractions of 50% for the $T \to tZ$ and $T \to tH$ decay modes.

The search was then combined with a complementary search targeting $T \to bW$ decays in the single-lepton final state which used substructure techniques to identify boosted hadronic W boson decays. These two searches were considered as individual channels of a global combination and labelled boosted H and boosted W channel in the combined search.

No excess of the data over the SM expectation was observed in neither the boosted W nor the boosted H channel and upper cross section limits at 95% C.L. were derived in a Bayesian calculation similar to the one described above. This was done by performing a simultaneous likelihood fit of the signal+background hypothesis using a variable referred to as $\min[\mathrm{M}(\ell, b)]$ in all categories of the boosted W channel and the S_{T} variable in the `H1b` and `H2b` categories of the boosted H channel (with the `0H` category not considered as it was covered by the boosted W channel). Systematic uncertainties were correlated between the two channels where appropriate, including the systematic uncertainty in the H_{T} reweighting procedure, and marginalised in the final likelihood fit.

These cross section limits were compared with the theoretical prediction for $T\overline{T}$ production for all possible branching fraction combinations for the T quark decay. Lower limits on the T quark mass were set to values ranging from 710 to 940 GeV, depending on the assumed branching fraction combination. A lower mass limit of 860 GeV was found for the electroweak isospin singlet branching fraction scenario, corresponding to a branching fraction of 50% for the $T \to bW$ decay and 25% to the $T \to tZ$ and $T \to tH$ decays.

For the isospin doublet assumption, masses below 830 GeV were excluded. The data were also interpreted assuming a signal model with vector-like B quark pair production, which were excluded for masses below 730 GeV assuming the isospin singlet branching fraction scenario.

These are amongst the first results derived for vector-like quark pair production at $\sqrt{s} = 13\,\text{TeV}$. The observed lower limits on the mass of vector-like T quarks are comparable to the ones obtained by the CMS experiment with a data set corresponding to $\mathcal{L} = 19.7\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$ and at some points even exceed them. The much larger data set, which the LHC is expected to deliver in the next few years, will certainly increase the sensitivity of this search to vector-like T quark production to masses above 1 TeV. Indeed, other searches that studied the data set collected in 2016 already indicate that vector-like quarks are excluded up to masses just above this scale [231–235, 240]. The higher these lower mass limits become, however, the more difficult will it be to observe vector-like quarks in the pair production channel as the corresponding cross section shows a steeply falling behaviour towards larger masses. Therefore, single production searches will grow increasingly important in order to be sensitive to vector-like quarks in the very high mass region of $\sim 2\,\text{TeV}$ at the LHC.

In any case, as naturalness suggests that vector-like quarks should have masses around the TeV scale, the next few years might very well shed light onto the question, whether the laws of nature follow a natural theory, or not.

Appendix

A Statistical Methods

In this section, some of the statistical methods employed in this thesis are explained. Section A.1 describes the maximum-likelihood method to estimate certain parameters which is mainly used to assess the modelling of the background in this thesis given a set of systematic uncertainties. Section A.2 describes limit estimation using Bayesian inference which is used for the statistical interpretation of the data in the final event categories in this thesis. In Section A.3, limit determination with the *asymptotic* $\mathrm{CL_s}$ *method* [241] is explained which is a very fast method but only valid under certain conditions and therefore used mainly for studies of expected sensitivities in this thesis. All these methods are performed in this thesis using the THETA [242] framework. A detailed review about all statistical methods used in this thesis can be found e. g. in Ref. [243].

A.1 Maximum-likelihood method

The maximum-likelihood method is a statistical tool to estimate the true values for a set of unknown parameters given a sample of measured data. It is based on the likelihood function

$$L(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i|\boldsymbol{\theta}) \tag{A.1}$$

where $\boldsymbol{\theta}$ is the set of parameters that one would like to estimate (they define what is commonly referred to as *model*), $\mathbf{x} = \{x_1, \ldots, x_N\}$ is a set of measured data (usually x_i gives the data yield in bin i of an observable with N bins) and $f(x_i, \boldsymbol{\theta})$ is a probability density function (p.d.f.) according to which the data $\mathbf{x}$ are distributed.

This p.d.f. is often assumed to follow a Poisson distribution

$$f(x_i|\lambda_i(\boldsymbol{\theta})) = \frac{\lambda_i(\boldsymbol{\theta})^{x_i} e^{-\lambda_i(\boldsymbol{\theta})}}{x_i!} \tag{A.2}$$

in each bin where $\lambda_i(\boldsymbol{\theta})$ is the expectation value in bin i which depends on the parameters of the model. An estimate for the parameters $\boldsymbol{\theta}$ is then given by finding the corresponding values that solve

$$\frac{\partial \ln L(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \tag{A.3}$$

to maximise the likelihood $L(\boldsymbol{\theta})$ which is usually done numerically.

The likelihood function can be expanded by an additional term

$$L(\mathbf{x}|\boldsymbol{\theta}(\boldsymbol{\nu}),\boldsymbol{\nu}) = \prod_{i=1}^{N} f(x_i|\boldsymbol{\theta}(\boldsymbol{\nu}))D(\boldsymbol{\nu}) = \prod_{i=1}^{N}\prod_{j=1}^{M} f(x_i|\boldsymbol{\theta}(\nu_j))d(\nu_j|\hat{\nu}_j,\sigma_{\nu_j}) \tag{A.4}$$

to incorporate a set of M systematic uncertainties in the parameters $\boldsymbol{\theta}$. The parameters $\boldsymbol{\nu}$ are also referred to as *nuisance parameters* so that each nuisance parameter j corresponds to an individual source of systematic uncertainty. The additional terms $D(\boldsymbol{\nu}) = \prod_j d(\nu_j|\hat{\nu}_j,\sigma_{\nu_j})$ follow a given probability distribution with mean $\hat{\nu}_j$ and width σ_{ν_j}. ν_j is a free parameter in the maximisation of the likelihood function which influences the parameters $\boldsymbol{\theta}(\boldsymbol{\nu})$ through a given response function (which is explained further below). The parametrisation of this response function in ν_j is usually chosen such that its mean value and standard deviation are $\hat{\nu}_j = 0$ and $\sigma_{\nu_j} = 1$, respectively, so that $\nu_j = 0$ corresponds to the nominal value for $\boldsymbol{\theta}$ and the values $\nu_j = \pm 1$ correspond to one standard deviation of the systematic uncertainty j.

For uncertainties that only affect the rate of certain background processes but not the shape of physical observables (*rate uncertainties*), log-normal terms $d(\nu_j|\hat{\nu}_j,\sigma_{\nu_j})$ are included in $D(\boldsymbol{\nu})$. Uncertainties that can have an impact on both, shape and rate, of physical observables (*shape uncertainties*) are incorporated by including Gaussian terms. The likelihood function $L(\mathbf{x}|\boldsymbol{\theta}(\boldsymbol{\nu}),\boldsymbol{\nu})$ defined in Eq. (A.4) is not only used in the MLE method but has a broad range of applications e. g. when calculating Bayesian intervals (see App. A.2).

In the cases in which the MLE method is used in this thesis, the parameters $\boldsymbol{\theta}(\boldsymbol{\nu})$ correspond to the predicted background yields in the bins of a given observable which depend on a set of systematic uncertainties parametrised by $\boldsymbol{\nu}$. This prediction is obtained by producing templates of the desired observable using MC generators (this is described in more detail in Chapter 4). As only a finite number of MC events can be simulated, there is an additional *statistical uncertainty* in the model $\boldsymbol{\theta}$ in addition to the systematic uncertainties. This statistical uncertainty is equivalent of the uncertainty of a Poisson distribution with the mean corresponding to the number of simulated MC events in each bin. It can be treated in the same way like other systematic uncertainties with the difference that there is one parameter ν_j per bin of the observable as statistical fluctuations are not correlated between bins. Variations of other systematic uncertainties, however, are usually correlated across all bins of the observable so that there is only one parameter ν_j per systematic uncertainty, independent of the number of bins. If the number of simulated events is large enough, the distribution of the variation parameter ν_j can be approximated by a Gaussian distribution which is referred to as *Barlow-Beeston light* method [228, 229]. This condition is ensured by choosing a proper binning for the physical observable under consideration.

The previously mentioned response function can be trivially derived for rate uncertainties since a variation of the parameter ν_j corresponds to a correlated shift of the rate across all bins of the observable. For shape uncertainties, in order to approximate the response function, a method called *vertical template morphing* is employed. In this procedure, the physical observable under consideration is re-computed twice for each shape uncertainty, varying the parameter affected by the uncertainty (e. g. a scale factor for a given selection efficiency) up and down by one standard deviation of the systematic uncertainty. The resulting templates correspond to the values $\nu_j = \pm 1$ of the parameter associated with this systematic uncertainty in the likelihood function in Eq. (A.4). The response function for the parameter $\boldsymbol{\theta}(\nu_j)$ is then given by interpolating between $\nu_j = -1$, $\nu_j = 0$ and $\nu_j = +1$ with a cubic function and by linearly extrapolating beyond $\nu_j = \pm 1$.

The main application of this method in this thesis is to assess the modelling of simulated background processes for certain physical observables. This is done by including systematic uncertainties using the vertical template morphing outlined above and evaluating the shifts of the parameters ν_j from Eq. (A.4) and their post-fit uncertainties. Shifts of these parameters within one or two standard deviations of their prior values are expected and indicate a good modelling while extreme shifts beyond two standard deviations hint at some problems of the background model.

A.2 Bayesian limit setting

After having collected data and performed an event selection in order to have a sample where it would be possible to observe effects of new physic models, one usually wants to make some quantitative statement about the possibility of such signal models to exist. In order to do this, one can construct a hypothesis $H(\mu)$ as a function of the signal strength μ so that a hypothesis that the presence of signal would correspond to $\mu > 0$ and the null-hypothesis (or background-only hypothesis) corresponds to $\mu = 0$. In the context of high-energy physics, μ is usually proportional to the production cross section of the signal process. The goal of any statistical inference is either to reject a given hypothesis (usually the null-hypothesis when deciding whether or not to claim a discovery) or derive so-called *confidence intervals* that contain the true value for μ at a certain *confidence level* (C.L.). There are two main approaches to carry out these two tasks, the frequentist and the Bayesian approach. In the analyses in Chapters 6 and 7, the latter approach is adopted which is therefore described in more detail in the following.

The central concept in Bayesian inference is summarised in Bayes' theorem

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{A.5}$$

which gives a relation for the conditional probabilities of two events A and B given their

total probabilities $P(A)$ and $P(B)$.

This can be translated into a statistical test of the parameter μ and gives

$$p(\mu|\mathbf{x}) = \frac{L(\mathbf{x}|\mu)\pi(\mu)}{\int L(\mathbf{x}|\mu')\pi(\mu')\,\mathrm{d}\mu'} \tag{A.6}$$

where $p(\mu|\mathbf{x})$ is the *posterior probability* as a function of the parameter μ, $L(\mathbf{x}|\mu)$ is the likelihood function depending on the signal strength μ (constructed in a similar way as in Eq. (A.1)) and $\pi(\mu)$ is the *prior probability* as a function of μ. The denominator is needed to normalise the posterior probability to values between 0 and 1. In the Bayesian approach, the posterior probability $p(\mu|\mathbf{x})$ can be thought of as *degree of belief* that the signal strength takes a certain value.

The prior probability $\pi(\mu)$ encodes the knowledge in the signal strength parameter μ before the statistical test. This is usually assumed to have the values

$$\begin{aligned} \pi(\mu) &= 0 \text{ for } \mu < 0 \\ \pi(\mu) &= 1 \text{ for } \mu \geq 0 \end{aligned} \tag{A.7}$$

which is commonly referred to as *flat prior*.

Expected background processes and systematic uncertainties can be incorporated in the likelihood function in a similar manner as described in Section A.1 by defining

$$L(\mathbf{x}|\mu) = \int L(\mathbf{x}|\mu(\boldsymbol{\nu}), \boldsymbol{\theta}(\boldsymbol{\nu}))\pi(\boldsymbol{\nu})\,\mathrm{d}\boldsymbol{\nu} \tag{A.8}$$

which is called the *marginal likelihood*. $\boldsymbol{\nu}$ is a set of nuisance parameters corresponding to systematic uncertainties which can influence the signal strength $\mu(\boldsymbol{\nu})$ and the background parameters $\boldsymbol{\theta}(\boldsymbol{\nu})$. $\pi(\boldsymbol{\nu})$ is the prior probability distribution function for the parameters $\boldsymbol{\nu}$ that can be constructed equivalently to the parameter $D(\boldsymbol{\nu})$ from Eq. (A.4). Statistical uncertainties in MC templates are usually also incorporated in $\boldsymbol{\nu}$ using the Barlow-Beeston light method [228, 229].

Bayesian credible intervals are then computed at a given confidence level α by calculating the integral of the posterior probability

$$1 - \alpha = \int_{\mu_0}^{\mu_1} p(\mu|\mathbf{x})\,\mathrm{d}\mu \tag{A.9}$$

In the context of high energy physics, upper limits, e. g. on the cross section of a certain process, are usually reported at 95% C.L. by setting $1 - \alpha = 0.95$ and $\mu_0 = 0$, leaving the upper bound μ_1 as free parameter. If the upper limit on a parameter is below the theoretically predicted value, the corresponding signal model is considered excluded at this confidence level.

The value computed using the measured data sample is referred to as *observed limit*.

In order to test the sensitivity of a given selection or in order to put this observed limit into context and compare it to one's expectation, *expected limits* can be calculated using pseudo-data sampled from the null-hypothesis in a so-called *toy experiment*. In the search presented in Chapter 6, 1000 toy experiments are run when computing limit intervals using the Bayesian method while in the the search presented in Chapter 7, 5000 toy experiments are performed. The median value from all these experiments is quoted as the central value for the expected upper cross section limit and the intervals that contain 68% and 95% of all calculated values are presented as bands that correspond to one and two standard deviations, respectively, from the central value.

Integrals of the form Eq. (A.9) are usually calculated numerically using a method called *Markov Chain Monte Carlo* (MCMC) in which the integral is approximated by a random walk process with a large number of steps. In each step, the function value at a certain point of the parameter space is calculated and added to the integral and the next point is randomly chosen based on the underlying probability distribution. For this calculations, usually 10000 iterations per Markov Chain are undertaken. As the MCMC method gives only an approximate value for the integral which can fluctuate around the true value, ten Markov Chains are run and each time an observed limit is calculated (where the median of the individual results is reported as result) while for toy experiments, only one Markov Chain is calculated.

A.3 Asymptotic $\mathrm{CL_s}$ limit setting

Applying the Bayesian method described in App. A.2 to derive upper cross section limits can result in a high computational overhead as performing many toy experiments or calculating integrals with the MCMC method usually takes a significant amount of time. If one is instead rather interested in obtaining fast and only approximate results, one can use so-called *asymptotic methods* to perform these calculations. This section provides an overview over the method, a detailed description can be found in Ref. [241].

These methods are based on *frequentist* methods and rely on the profile likelihood ratio

$$\lambda_b(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\nu}}})}{L(\hat{\mu}, \hat{\boldsymbol{\nu}})} \tag{A.10}$$

where L is the likelihood function defined in Eq. (A.4), μ is the signal strength for the process of interest (so that $\mu = 0$ corresponds to the background-only hypothesis) and $\boldsymbol{\nu}$ is a set of nuisance parameters. $\hat{\mu}$ and $\hat{\boldsymbol{\nu}}$ are the values for which L has its maximum while $\hat{\hat{\boldsymbol{\nu}}}$ is the set of nuisance parameter values that maximise L for a given μ.

One can then define a test statistic $q_\mu = -2 \ln \lambda_b(\mu)$ for $\mu \geq \hat{\mu}$ and $q_\mu = 0$ for $\mu < \hat{\mu}$ to evaluate the desired confidence intervals. In the limit of large statistics, the profile

likelihood ratio is normally distributed and this test statistic becomes approximately

$$-2\ln\lambda_b(\mu) \approx \frac{(\mu-\hat{\mu})^2}{\sigma^2} \tag{A.11}$$

with $\hat{\mu}$ and σ being the mean and standard deviation, respectively, of the Gaussian curve of the profile likelihood ratio. The standard deviation σ can be calculated from the covariance matrix of the estimators for μ and the nuisance parameters $\boldsymbol{\nu}$ Upper limits on the parameter μ at a given confidence level $1-\alpha$ can then directly be calculated from this test statistic setting q_μ to

$$q_\mu = \hat{\mu} + \sigma \cdot \Phi^{-1}(1-\alpha) \tag{A.12}$$

where $\Phi(x)$ is the cumulative distribution of the standard Gaussian with zero mean and unit variance. The upper limit at 95% C.L., for instance, corresponds to the value of μ for which $q_\mu = \hat{\mu} + 1.68\sigma$. One can also derive expected upper exclusion limits by defining a so-called *Asimov* pseudo-data set. In this, the pseudo-data yields in every bin of the considered observable are set to their expectation values given by the considered background model and q_μ is calculated for this data set.

When performing a search with a signal model s and a background model b in which one would like to quote an upper exclusion limit on the signal+background hypothesis, it is useful to quote the $CL_s = \frac{CL_{s+b}}{CL_b}$ confidence interval [244, 245]. This way, one can avoid setting too strong limits in case of a downwards fluctuation of the data below the expected background.

A.4 Goodness-of-fit tests

The compatibility of an observed data spectrum with a given model can be quantified using so-called *goodness-of-fit* (GOF) tests. The χ^2 test statistic is defined as [246]

$$-2\ln\lambda_{\chi^2} = 2\sum_i f_i - d_i + d_i\ln(d_i/f_i) \tag{A.13}$$

where d_i and f_i correspond to the observed and expected yields, respectively, in the bins i of an observable and it is assumed that f_i follows a Poisson distribution. This test statistic follows a χ^2 distribution in the limit of large data samples.

Another goodness-of-fit test is the *Kolmogorov-Smirnov* (KS) test [247]. It is based on the empirical distribution function

$$F_N(y) = \frac{1}{N}\sum_{i=1}^{N} A(y, x_i) \tag{A.14}$$

of a free parameter y and for an observable with N bins with observed values x_i in those

bins. $A(y, x_i)$ is equal to 1 if $x_i \leq y$ and equal to 0 otherwise.

The KS test then returns the test statistic

$$\lambda_{\mathrm{KS}} = \sup_y |F_{N,d}(y) - F_{N,b}(y)| \tag{A.15}$$

where $F_{N,d}$ and $F_{N,b}$ are the empirical distribution functions for the observed data and expected background distributions and $\sup_y$ gives the maximum value over the entire y-range.

Both tests can be combined with the MLE method described in App. A.1 to first compute the background expectation with nuisance parameters set to their maximum-likelihood estimates and then evaluate the respective test statistic. These tests are performed for the observed data spectrum and for a large number of toy experiments with pseudo-data sampled from the background prediction. The ratio

$$p = \frac{\sum_{i=\lambda_d}^{\infty} \lambda_i}{\sum_{i=0}^{\infty} \lambda_i} \tag{A.16}$$

with the returned test statistic value for the data λ_d and for the toy experiments λ_d then provides the p-value for the observed data spectrum in light of the background expectation.

B Optimisation studies for the search for $T\bar{T} \to tH + X$

B.1 Lepton selection optimisation study

In order to maximise the trigger and offline lepton selection (see Section 6.3.1) with respect to a high selection efficiency of $T\overline{T}$ production, an optimisation study has been performed by comparing the signal efficiency of various selection configurations for two T quark mass points. The selection configurations that are tested for the muon channel are defined in the following and summarised in Table B.1.

Non-isolated muon The trigger and offline requirements as well as the 2D isolation requirement are applied as described in Section 6.3.1.

Isolated Muon Events are selected by a trigger requiring an isolated muon candidate with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and by requiring a reconstructed offline muon candidate with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ and a relative isolation of less than 0.15 (see Section 5.2).

In the electron channel, three setups were tested in total which are defined as:

Non-isolated electron The trigger and offline requirements along with the 2D isolation requirement are applied as described in Section 6.3.1.

Non-isolated high-p_{T} electron Events need to pass a trigger that requires an electron candidate with a p_{T} of at least 105 GeV without any requirements on its relative isolation. The 2D isolation requirement as described in Section 6.3.1 is applied to reduce the contribution from multijet background events.

Isolated electron Events are selected by a trigger requiring an isolated electron candidate with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ and $|\eta| < 2.1$. Offline, the electron must have $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ and a relative isolation of < 0.0987 or 0.0902, depending on whether it is reconstructed in the barrel part of the detector or in endcap.

In order to compare the signal efficiencies for $T\overline{T}$ production for all these selection combinations in a phase space close to the final selection of this search, additional kinematic selection requirements are imposed. $S_{\mathrm{T}} > 800\,\mathrm{GeV}$ is required as well as at least three AK4 jets and two AK8 jets, one of which has to have a p_{T} of at least 300 GeV. Furthermore, at least one generated electron for the electron selection and at least one generated muon

Table B.1: Selection requirements for the trigger and lepton optimisation study. The first block lists the baseline selection that all events entering in the denominator have to pass while the second block provides the individual selection criteria for the selection configurations studied.

Selection/setup	Non-iso. μ	Iso. μ	Non-iso. e	Non-iso. high-p_{T} e	Iso. e
Generator req.	≥ 1 μ		≥ 1 e		
S_{T}	> 800 GeV				
N(AK4 jets)	≥ 3				
N(AK8 jets)	≥ 2				
AK8 jet p_{T}	> 300 GeV				
Min. lepton p_{T} (Trigger)	45 GeV	20 GeV	45 GeV	105 GeV	27 GeV
Min. lepton p_{T} (Offline)	47 GeV	40 GeV	50 GeV	115 GeV	40 GeV
Add. jet p_{T}	—	—	200 (250) GeV, 50 (75) GeV *	—	—
Max. rel. isolation	—	0.15	—	—	0.0987 (0.0902)**
2D Isolation Criteria	$\Delta R(l,j) > 0.4$ or $\Delta p_{\mathrm{T}}(l,j) >$ 40 GeV	—	$\Delta R(l,j) > 0.4$ or $\Delta p_{\mathrm{T}}(l,j) >$ 40 GeV	$\Delta R(l,j) > 0.4$ or $\Delta p_{\mathrm{T}}(l,j) >$ 40 GeV	—

* Online (offline) p_{T} requirements on the p_{T}-leading and -subleading AK4 jet .

** Relative isolation requirements for electrons reconstructed in the barrel (endcap) part of the detector.

for the muon selection is required. These generated leptons are required to originate from t quark decays as these are the main target in the search in Chapter 6.

Table B.1 lists these general selection requirements as well as the individual criteria for all tested trigger and offline selection setups. In the calculation of efficiencies for the different selection setups, the denominator is given by events that pass the general selection requirements so that the fraction of these events passing the additional individual selection steps yield the corresponding efficiency.

Figure B.1 shows the selection efficiency of all considered selection configurations for two mass points of the T quark as a function of the primary lepton p_{T}. Inclusive selection efficiencies for several mass points are summarised in Table B.2. It can be seen that in the muon channel, the efficiency of the non-isolated selection is higher for both the high and the low mass point compared to the isolated selection. For the electron channel, this is only the case for the high-mass signal, but the advantage of the isolation selection for the low-mass hypothesis is only small. Although the high-p_{T} electron selection provides by far the best efficiency for electrons with p_{T} above 115 GeV, the low-p_{T} non-isolated electron selection has the higher inclusive efficiency (as can be seen in Table B.2) as a large portion of $T\overline{T}$ events produce leptons with p_{T} values below 115 GeV.

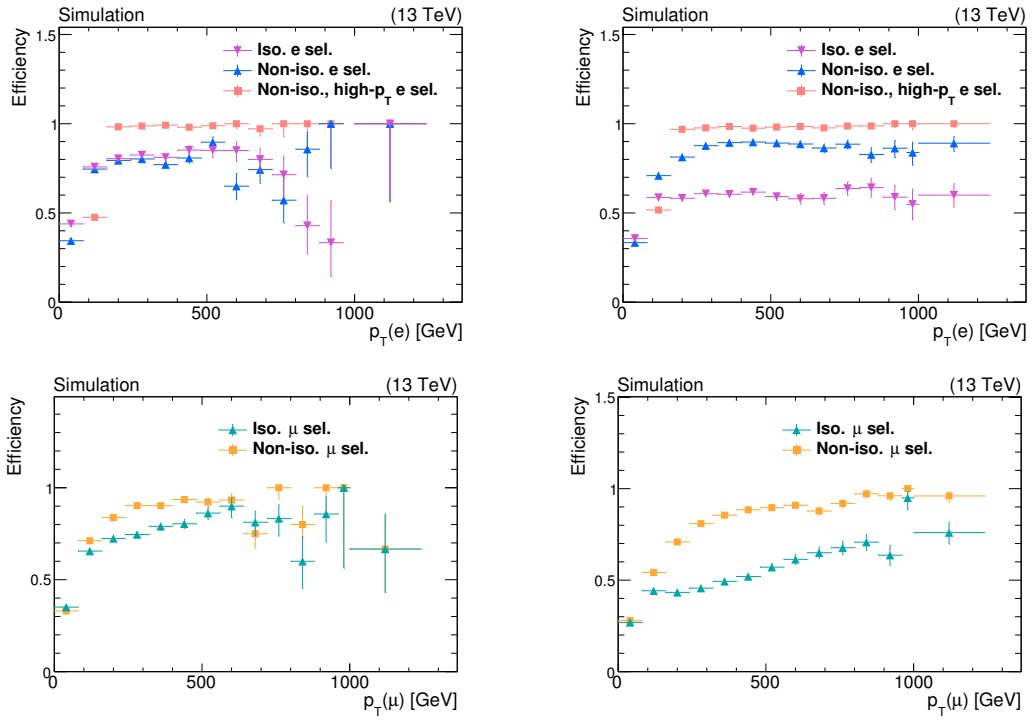

Figure B.1: Selection efficiencies for different selection setups described in the main text. Efficiencies are shown for a T quark mass of 800 GeV (left) and 1600 GeV (right) in the electron (top row) and muon (bottom row) channel.

Table B.2: Inclusive signal efficiencies for two mass points of the T quark (800 and 1600 GeV) for different trigger and lepton selections described in the main text. The efficiencies are calculated with respect to the number of events that pass the general selection outlined in the main text. The lepton in these denominator events always originate from a top quark decay in a $T \to tH, tZ$ process.

Selection	M(T) = 800 GeV	M(T) = 1600 GeV
Non-Iso. Muon	60.8	62.9
Iso. Muon	56.2	43.2
Non-Iso. High-p_{T} Electron	45.4	65.4
Non-Iso. Electron	60.2	70.7
Iso. Electron	64.9	53.8

B.2 Selection optimisation study

This section describes the optimisation of the event selection after the preselection and after defining the `H1b` and `H2b` categories described in Section 6.3.4. Additional requirements on various kinematic variables are tested and their impact on the sensitivity with respect to $T\overline{T}$ production is evaluated by calculating expected upper cross section limits at 95% C.L.. The selection providing the highest lower mass limit on the T quark mass is considered to be the best selection.

The tested selection requirements are:

- the number of AK4 or AK8 jets,
- the value for $\not{E}_{\mathrm{T}}$,
- the p_{T} of the primary lepton,
- the p_{T} of the leading/subleading AK4 jet,
- the number of b-tagged AK4 jets.

When testing the sensitivity of additional b tagging requirements on AK4 jets, a ΔR of at least 0.8 is required between a b-tagged AK4 jet and the nearest H-tagged AK8 jet. This is necessary in order to avoid double-counting a b-tagged AK4 jet also as H-tagged AK8 jet since the two types of jets are reconstructed independent of each other and can therefore overlap.

Events have to pass the preselection described in Section 6.3.2 and are categorised into `H1b` and `H2b` categories (and split into an electron and muon channel) for this study. Expected limits at 95% C.L. are then calculated using the asymptotic $\mathrm{CL_s}$ procedure (see App. A.3) using events that pass an additional selection requirement from the list of tested selections. A rate uncertainty of 100% is assigned to multijet production and of 20% to all other background processes. No other systematic uncertainties are considered for this study. For the $T\overline{T}$ signal, a branching fraction of 100% for the $T \to tH$ decay is assumed to maximise the sensitivity for this decay mode.

The results are shown in Fig. B.2. It can be seen that only the requirement of at least one additional b-tagged AK4 jet provides a significant improvement of the selection. Requiring at least four AK4 jets or at least three AK8 jets would improve the sensitivity only marginally and is not considered in this analysis.

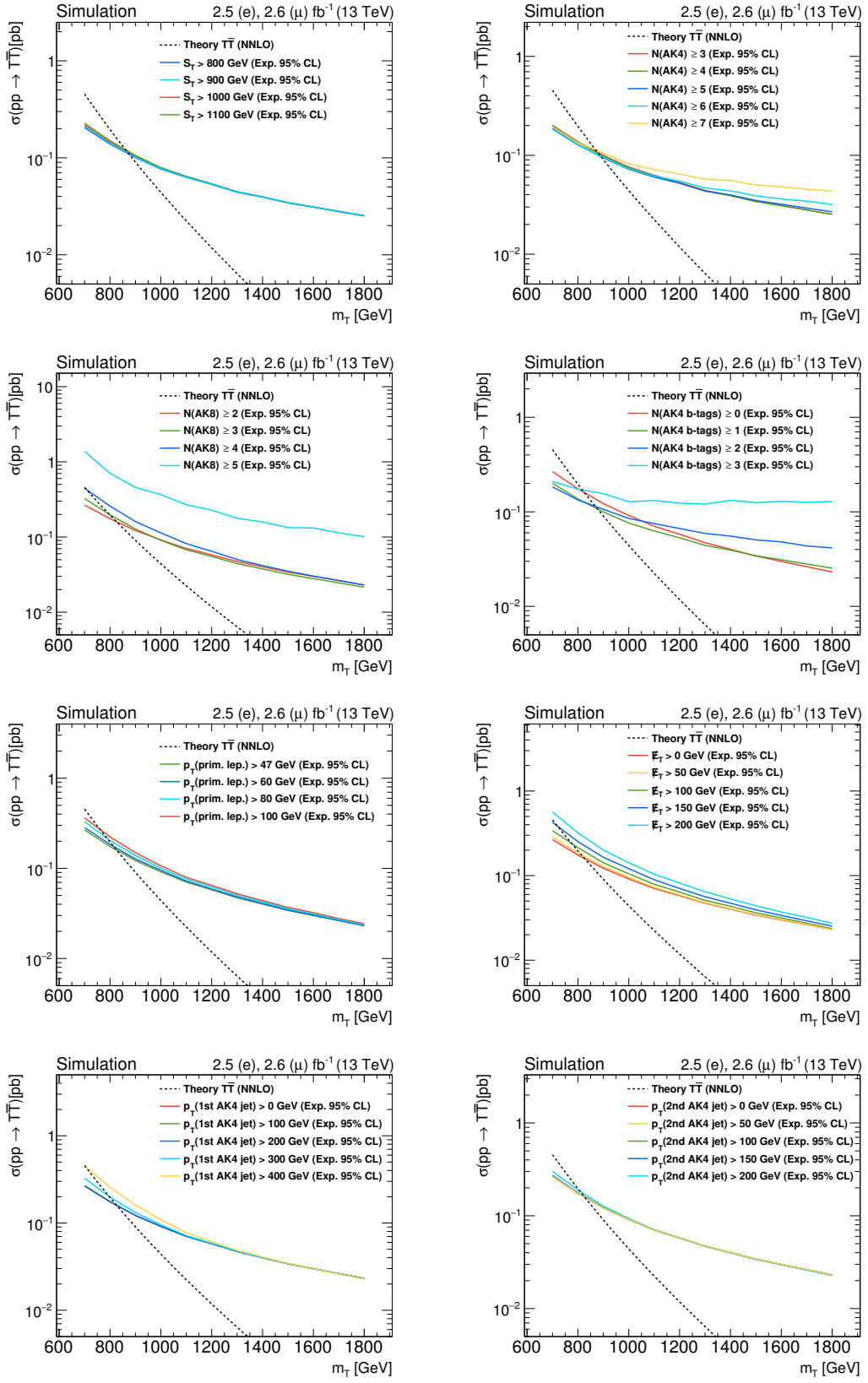

Figure B.2: Expected sensitivities when varying various kinematic variables. From top left to bottom right: S_{T}, N(AK4 jets), N(AK8 jets), N(b-tagged AK4 jets), p_{T}(prim.lep.), $\not\!\!E_{\mathrm{T}}$, p_{T}(1st AK4 jet), p_{T}(2nd AK4 jet).

C Studies on the background modelling in the search for $T\bar{T} \to tH + X$

C.1 Top-p_{T} reweighting study

A reweighting procedure referred to as *top-p_T reweighting* has been developed in the scope of differential cross section measurements for $t\bar{t}$ + jets production [224–227] to account for observed differences between the top quark p_{T} spectra in data and simulation. It is tested whether this reweighting procedure can improve the mismodelling of the S_{T} distribution discussed in Section 6.5.1 as S_{T} is clearly correlated with the top quark transverse momenta in $t\bar{t}$ + jets events.

In this procedure, simulated $t\bar{t}$ + jets events are scaled down according to correction factors that are calculated per event with the following equation:

$$\mathrm{SF}(p_{\mathrm{T}}) = e^{a-b\cdot p_{\mathrm{T}}(t_1)} \times e^{a-b\cdot p_{\mathrm{T}}(t_2)} \tag{C.1}$$

where t_1 and t_2 refer to the two generated top quarks. The parameters a and b were found to be $a = 0.0615$ and $b = 0.0005$. The upper uncertainty assigned to this procedure corresponds to an unscaled distribution while the lower uncertainty is determined by applying the correction factor in quadrature. A comparison of the templates corresponding to the upper and lower uncertainties in the top-p_{T} reweighting procedure with the nominal distribution in the $t\bar{t}$ + jets control region is shown in Fig. C.1.

Performing the top-p_{T} reweighting procedure results in the distributions for S_{T} in the $t\bar{t}$ + jets control region shown in Fig. C.2. On the left, one can see the actual distribution of the data and the corresponding expected background yield. On the right, the S_{T} distribution with all histograms normalised to unity are shown to allow for a shape-only comparison of the data and simulation as described previously. The uncertainty band in both cases includes only the uncertainty in the top-p_{T} reweighting procedure and the uncertainty in the $\mu_{\mathrm{R/F}}$ energy scales as this constitutes the dominant source of the uncertainty in the background prediction.

One can see that applying top-p_{T} reweighting reduces the overall rate of the $t\bar{t}$ + jets background but has only a limited effect on the shape of the S_{T} distribution. This suggests that a mismodelling of the p_{T} spectrum of simulated top quarks is most likely not the only source of the discrepancies observed in the S_{T} distributions in the $t\bar{t}$ + jets control regions.

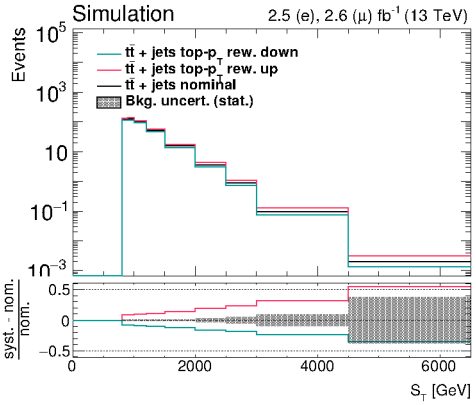

Figure C.1: Comparison of the nominal S_{T} distribution with the templates corresponding to upper and lower variations by one standard deviation of the top-p_{T} reweighting uncertainty. The $t\bar{t}$ + jets background process is shown in the corresponding control region.

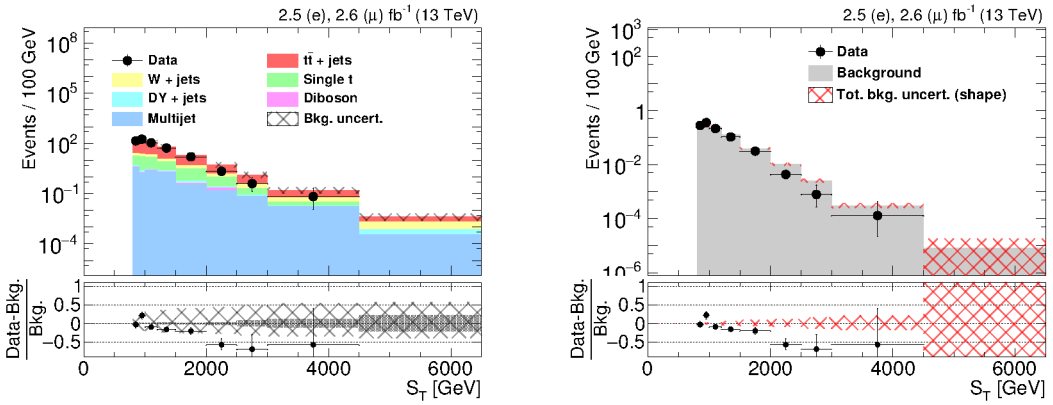

Figure C.2: S_{T} in the $t\bar{t}$ + jets control region after combining the electron and muon channels with top-p_{T} reweighting applied. The left plot shows the distributions with absolute event yields, the right plot with all histograms normalised to unity to evaluate the shapes of the data and MC histograms with shape-only systematic uncertainties. Only the systematic uncertainties in the $\mu_{\mathrm{R/F}}$ scales and the top-p_{T} reweighting procedure are taken into account.

C.2 Control distributions after H_T reweighting

This section contains control distributions in Figs. C.3 to C.5 for a number of kinematic variables in the $t\bar{t}$ + jets and W + jets control regions (defined in Section 6.5.1) after applying the H_T reweighting method described in Section 6.5.2. Overall, good agreement between data and simulated background processes is observed.

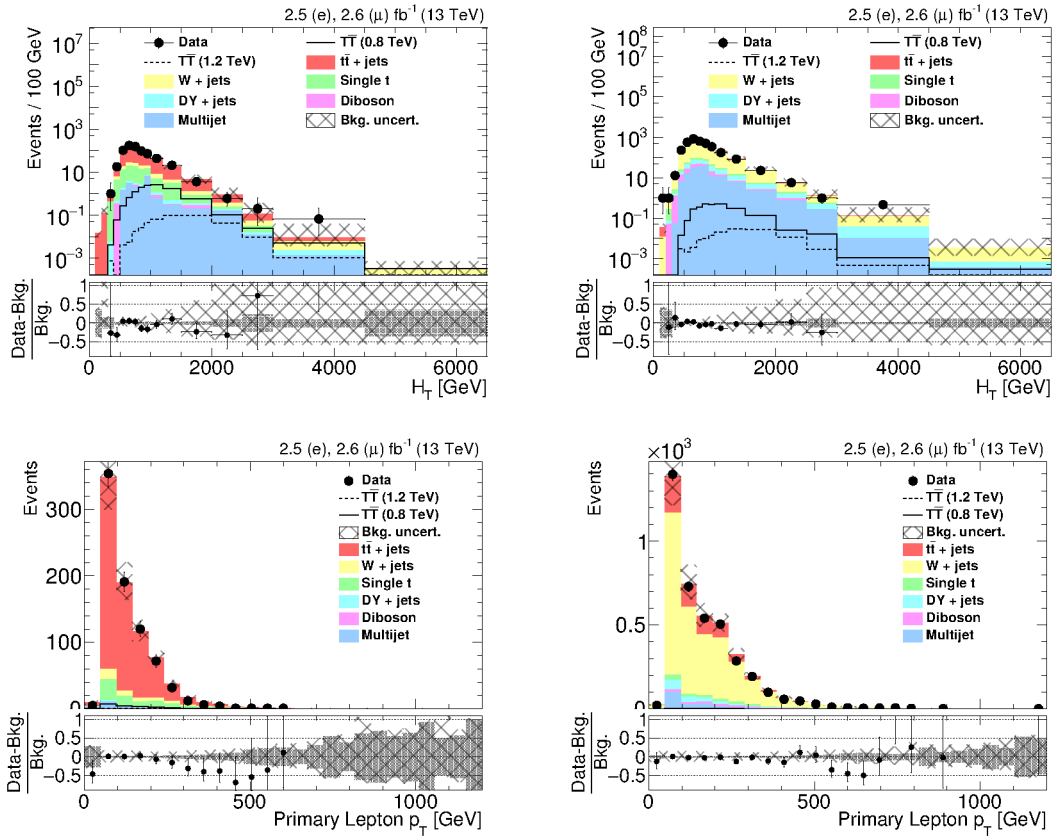

Figure C.3: Distributions in the $t\bar{t}$ + jets (left) and W + jets (right) control regions after combining electron and muon channels and applying the H_T reweighting procedure. H_T (top row) and p_T (primary lepton) (bottom row) are shown. The statistical uncertainty in the simulation is shown as gray band and the total uncertainty (combining statistical and systematic components) is shown as hatched band. The systematic component includes the uncertainty in the H_T reweighting procedure.

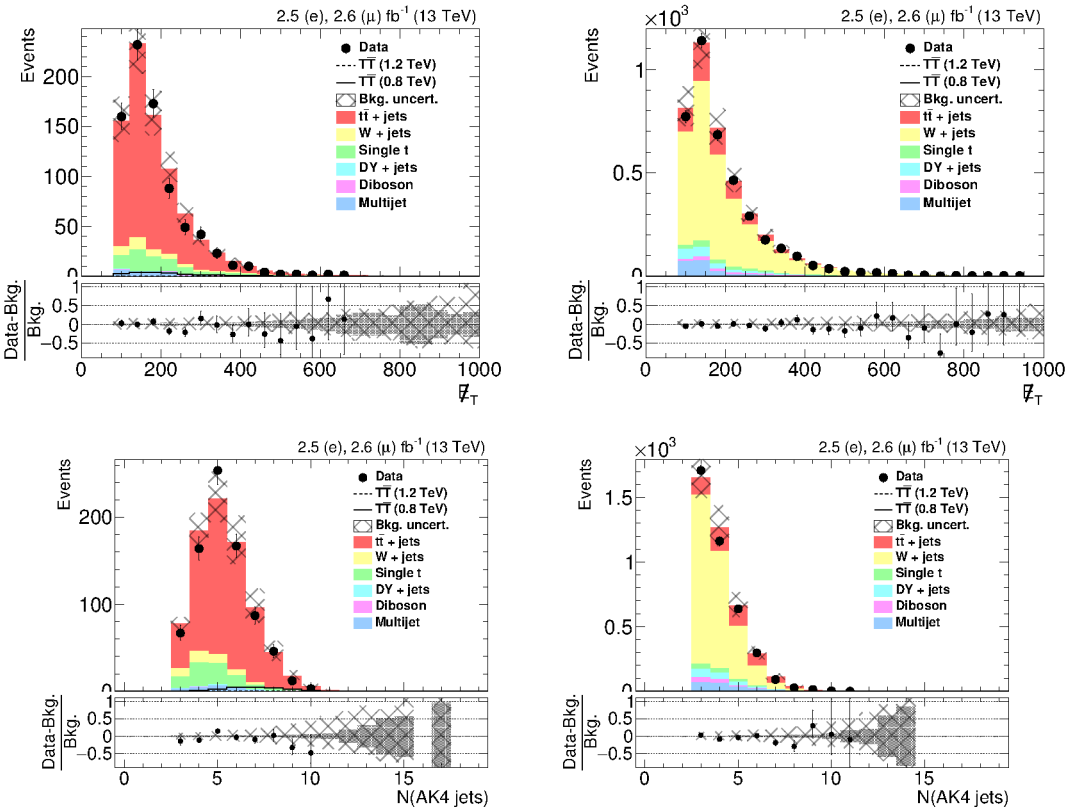

Figure C.4: Distributions in the $t\bar{t}$ + jets (left) and W + jets (right) control regions after combining electron and muon channels and applying the H_T reweighting procedure. $\not{E}_T$ (top row) and the number of AK4 jets (bottom row) are shown. The statistical uncertainty in the simulation is shown as gray band and the total uncertainty (combining statistical and systematic components) is shown as hatched band. The systematic component includes the uncertainty in the H_T reweighting procedure.

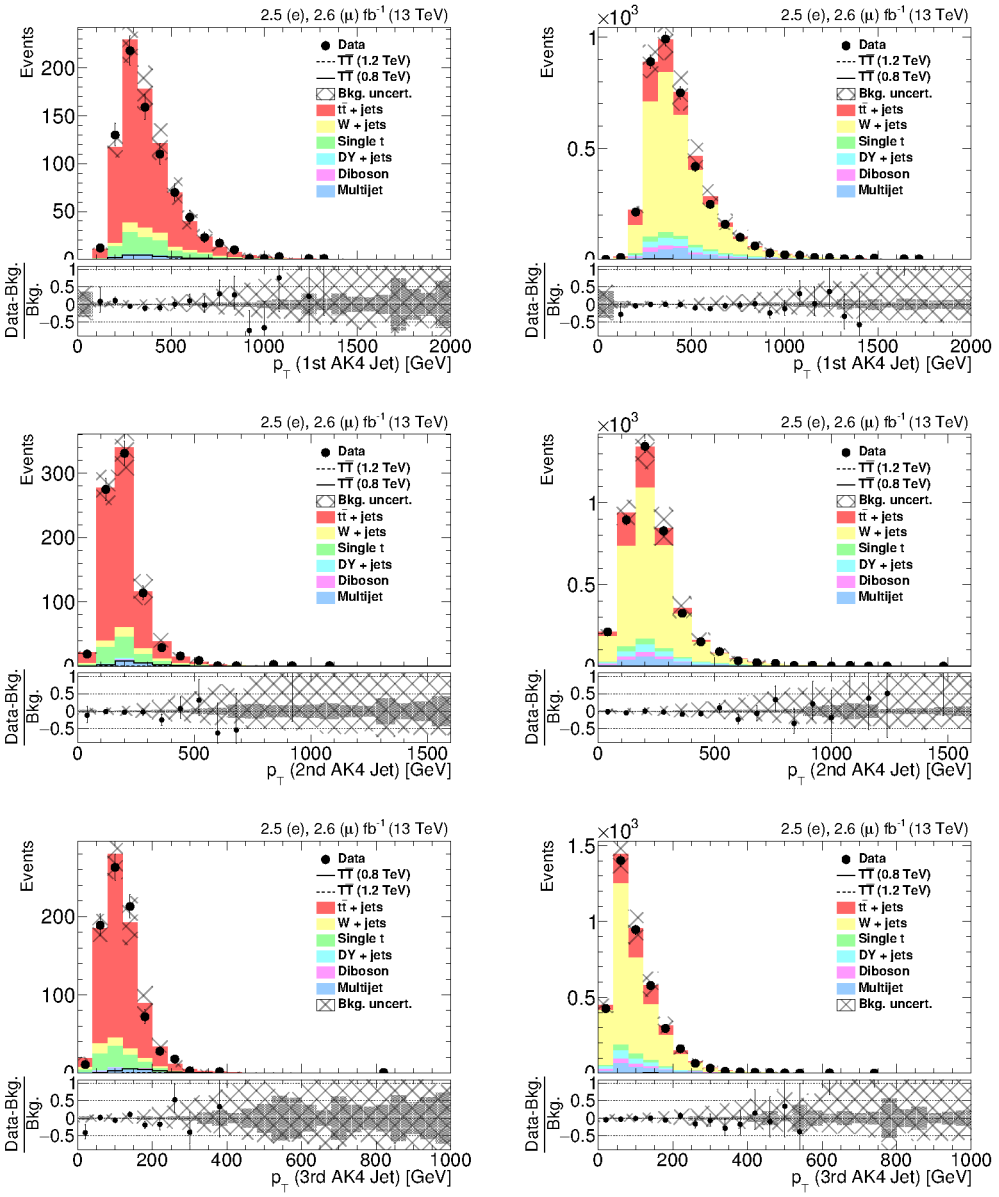

Figure C.5: Various distributions in the $t\bar{t}$ +jets (left) and W +jets (right) control regions after combining electron and muon channels and applying the H_T reweighting procedure. From top to bottom the p_T spectra of the first, second and third AK4 jet (in order of their p_T) are shown. The statistical uncertainty in the simulation is shown as gray band and the total uncertainty (combining statistical and systematic components) is shown as hatched band. The systematic component includes the uncertainty in the H_T reweighting procedure.

C.3 Signal injection bias test

In order to test the potential bias of the linear fit described in Section 6.5.2 in the presence of signal, the fit is executed again multiple times with various signal hypotheses injected into the data. The resulting impact on the derived parameters for the H_{T} reweighting is then compared to the original parameters without any signal injected. A sizable signal-to-background ratio is found only in the $t\bar{t}$ + jets control region while this ratio is $< 1\%$ over the whole H_{T} range in the W + jets control region. Therefore, these tests are only performed for the $t\bar{t}$ + jets reweighting parameters as potential signal is unlikely to have any effect on the W + jets reweighting parameters.

These tests are performed by adding signal templates to the data distribution in the $t\bar{t}$ +jets control region that correspond to $T\overline{T}$ or $B\bar{B}$ signal models with different branching fraction scenarios. The resulting distributions can be found in Figs. C.6 to C.8[1]: Figs. C.6 and C.7 show the H_{T} distributions in the $t\bar{t}$ + jets control region with $T\overline{T}$ signal injected assuming $\mathcal{B}(T \to tH) = 100\%$ and $\mathcal{B}(T \to bW) = 100\%$, respectively, for three different mass points. Figure C.7 shows the same distribution with a $B\bar{B}$ signal injected assuming $\mathcal{B}(\mathrm{B} \to \mathrm{tW}) = 100\%$.

The $\mathcal{B}(T \to tH) = 100\%$ hypothesis is considered as this is the signal hypothesis to which the search in Chapter 6 is most sensitive to. In the $\mathcal{B}(T \to bW) = 100\%$ scenario as well as for $B\bar{B}$ production with a branching fraction of 100% for the $\mathrm{B} \to \mathrm{tW}$ decay[2], a large portion of the signal could fail the H tag veto applied in the $t\bar{t}$ + jets control region, leading to a large portion of signal events entering that region. It is assumed that the results of testing these three scenarios sufficiently cover other branching fraction combinations.

The corresponding fit parameters a and b (using the function Eq. (6.2) described in Section 6.5.2) are listed in Table C.1 compared with the actual parameters used in this analysis (see also the values 6.3 and 6.4 in Section 6.5.2). One can see that there are small differences for the lowest considered mass point while these vanish for higher mass points. Overall, the observed differences are within the uncertainties of the corresponding fits and negligible compared to the large systematic uncertainty applied to the H_{T} reweighting procedure.

[1] In the following, the symbol $\mathcal{B}$is used to indicate the branching fraction of a given decay mode.
[2] This is equivalent to the production of two X quarks, see also Section 2.4.1.

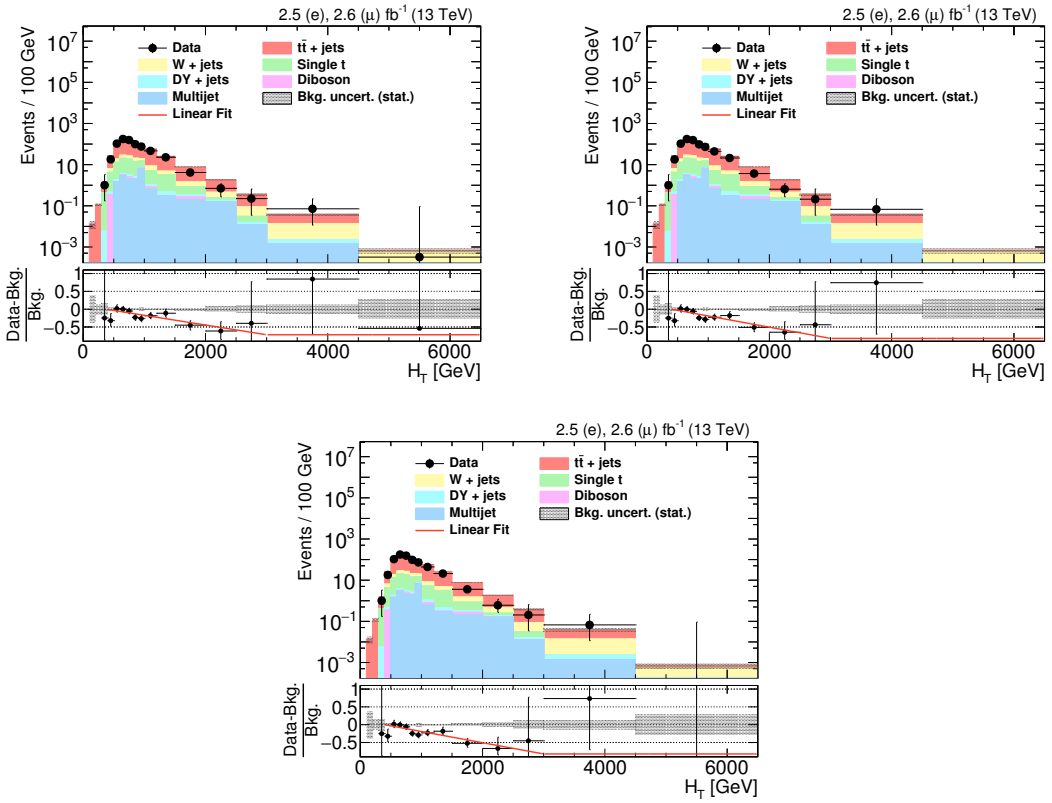

Figure C.6: Comparison of the H_{T} distributions for three different $T\overline{T}$ hypotheses injected (assuming $\mathcal{B}(T \to tH) = 100\%$): Signal with $\mathrm{M}(T) = 800\,\mathrm{GeV}$ injected (top left), with $\mathrm{M}(T) = 1200\,\mathrm{GeV}$ injected (top right) and with $\mathrm{M}(T) = 1600\,\mathrm{GeV}$ injected (bottom).

Table C.1: Fit parameters without and with signal samples for various mass points and branching fraction assumptions injected. The first line shows the fit parameters without any signal injected which correspond to the values 6.3 and 6.4 used in Chapters 6 and 7. The parameter a denotes the constant term in the linear function, b is the coefficient in the linear term.

Signal hypo./params.	a	b ($\cdot 10^{-4}\mathrm{GeV}^{-1}$)
no sig. inj.	1.12 ± 0.09	(-3.1 ± 0.9)
M(T) = 800 GeV, $\mathcal{B}(T \to tH) = 100\%$	1.10 ± 0.09	(-2.7 ± 0.9)
M(T) = 1200 GeV, $\mathcal{B}(T \to tH) = 100\%$	1.12 ± 0.09	(-3.1 ± 0.9)
M(T) = 1600 GeV, $\mathcal{B}(T \to tH) = 100\%$	1.12 ± 0.09	(-3.1 ± 0.9)
M(T) = 800 GeV, $\mathcal{B}(T \to bW) = 100\%$	1.11 ± 0.09	(-2.8 ± 0.9)
M(T) = 1200 GeV, $\mathcal{B}(T \to bW) = 100\%$	1.12 ± 0.09	(-3.1 ± 0.9)
M(T) = 1600 GeV, $\mathcal{B}(T \to bW) = 100\%$	1.12 ± 0.09	(-3.1 ± 0.9)
M(B) = 800 GeV, $\mathcal{B}$(B →tW) = 100%	1.10 ± 0.09	(-2.6 ± 0.9)
M(B) = 1200 GeV, $\mathcal{B}$(B →tW) = 100%	1.12 ± 0.09	(-3.1 ± 0.9)
M(B) = 1600 GeV, $\mathcal{B}$(B →tW) = 100%	1.12 ± 0.09	(-3.1 ± 0.9)

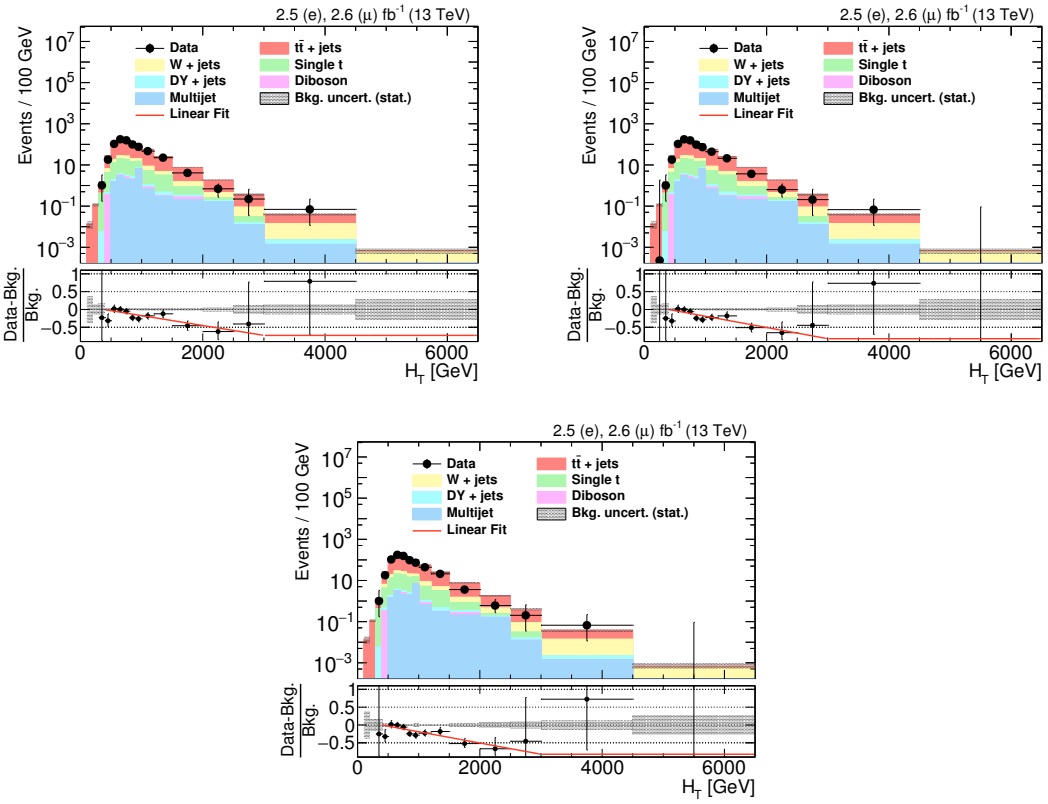

Figure C.7: Comparison of the H_{T} distributions for three different $T\overline{T}$ hypotheses injected (assuming $\mathcal{B}(T \to bW) = 100\%$): Signal with $\mathrm{M}(T) = 800\,\mathrm{GeV}$ injected (top left), with $\mathrm{M}(T) = 1200\,\mathrm{GeV}$ injected (top right) and with $\mathrm{M}(T) = 1600\,\mathrm{GeV}$ injected (bottom).

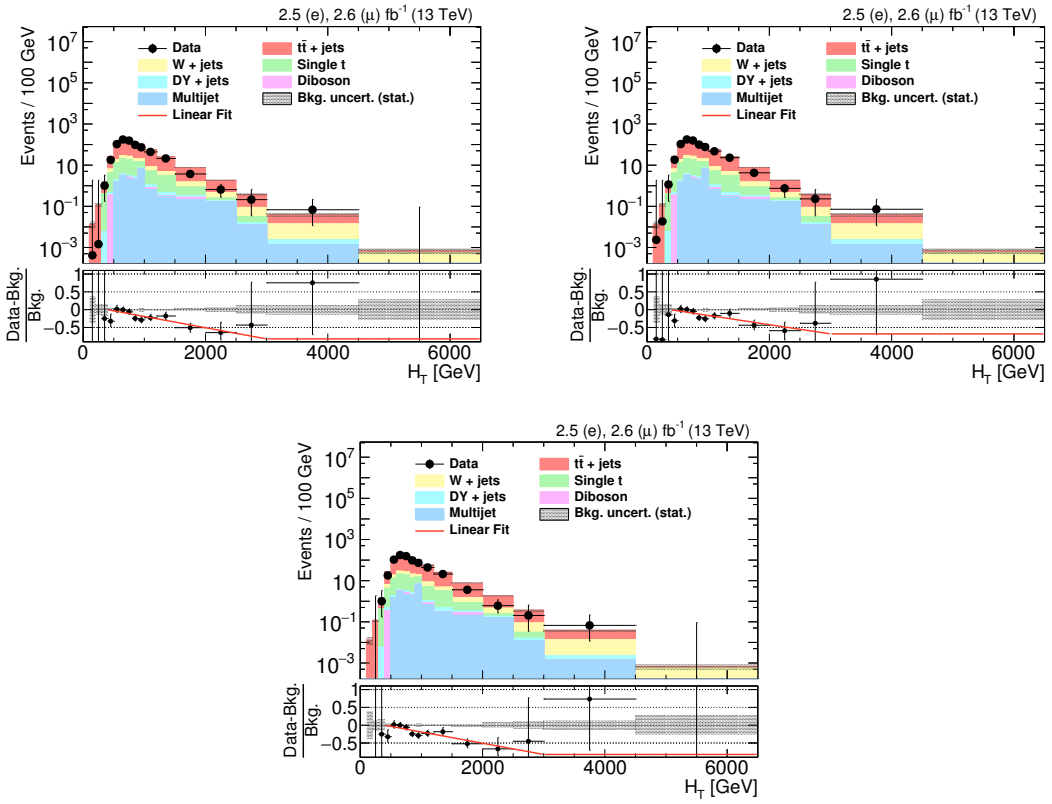

Figure C.8: Comparison of the H_{T} distributions for three different $B\bar{B}$ hypotheses injected (assuming $\mathcal{B}(\mathrm{B} \to \mathrm{tW}) = 100\%$): Signal with $\mathrm{M}(T) = 800\,\mathrm{GeV}$ injected (top left), with $\mathrm{M}(T) = 1200\,\mathrm{GeV}$ injected (top right) and with $\mathrm{M}(T) = 1600\,\mathrm{GeV}$ injected (bottom).

D List of prior uncertainties in the search for $T\bar{T} \to tH + X$

Prior systematic uncertainties in the rates of physical processes for all categories used in the statistical interpretation of the data described in Section 6.6 are shown in Tables D.1 to D.6.

Table D.1: Summary of systematic uncertainties in the `0H` **category** in the **electron** channel.

proc./ns. par.	PDF	μ_{RF} scale var.	b tag eff., h.-fl.	b tag eff., l.-fl.	AK8 jet mass sc.	AK8 jet mass res.	JEC	JER	H_{T} rew.	Pileup rew.	e id. eff.
Z + jets	$^{+6.55}_{-6.33}$	$^{+21.10}_{-15.91}$	$^{+2.64}_{-2.63}$	$^{+3.27}_{-3.26}$	$^{-0.11}_{-0.30}$	$^{-0.17}_{-0.18}$	$^{+2.16}_{-1.39}$	$^{+2.18}_{-2.32}$	± 0.0	$^{+1.44}_{-1.65}$	$^{+1.28}_{-1.28}$
Diboson	—	—	$^{+3.53}_{-3.52}$	$^{+1.19}_{-1.19}$	± 0.0	± 0.0	$^{+0.02}_{-2.89}$	$^{+3.13}_{-0.00}$	± 0.0	$^{+6.87}_{-6.24}$	$^{+2.64}_{-2.64}$
Multijet	—	—	$^{+2.89}_{-2.87}$	$^{+1.88}_{-1.88}$	$^{-0.15}_{+0.07}$	$^{+0.05}_{-0.17}$	$^{+2.30}_{-2.88}$	$^{+1.99}_{+0.61}$	± 0.0	$^{-0.81}_{+1.79}$	$^{+1.52}_{-1.52}$
Single t	$^{+0.75}_{-0.76}$	$^{+2.41}_{-3.15}$	$^{+3.92}_{-3.87}$	$^{+0.61}_{-0.61}$	$^{+0.12}_{-1.03}$	$^{-0.21}_{-0.71}$	$^{+1.01}_{-1.41}$	$^{+0.59}_{-0.51}$	± 0.0	$^{+0.41}_{-0.23}$	$^{+1.15}_{-1.15}$
$\mathrm{t\bar{t}}$ + jets	—	—	$^{+4.31}_{-4.24}$	$^{+0.71}_{-0.71}$	$^{+0.17}_{-0.73}$	$^{+0.08}_{-0.49}$	$^{+2.06}_{-2.48}$	$^{+0.46}_{-0.34}$	$^{+18.53}_{-14.17}$	$^{+0.34}_{-0.39}$	$^{+1.02}_{-1.02}$
W + jets	—	—	$^{+2.91}_{-2.90}$	$^{+2.70}_{-2.69}$	$^{-0.08}_{-0.48}$	$^{-0.07}_{-0.33}$	$^{+1.59}_{-3.71}$	$^{+0.49}_{-0.52}$	$^{+17.85}_{-14.09}$	$^{+0.48}_{-0.58}$	$^{+1.35}_{-1.35}$
$\mathrm{T\bar{T}}$ M(T)=0.8 TeV	$^{+2.94}_{-2.95}$	$^{+0.91}_{-0.90}$	$^{+7.64}_{-7.40}$	$^{+0.82}_{-0.82}$	$^{+0.96}_{-1.57}$	$^{+0.28}_{-0.70}$	$^{+0.58}_{-0.83}$	$^{+0.17}_{-0.78}$	± 0.0	$^{+0.46}_{-0.64}$	$^{+1.05}_{-1.05}$
$\mathrm{T\bar{T}}$ M(T)=1.2 TeV	$^{+4.62}_{-4.65}$	$^{+0.73}_{-0.74}$	$^{+8.37}_{-8.09}$	$^{+0.90}_{-0.90}$	$^{+0.65}_{-0.91}$	$^{+0.69}_{-1.10}$	$^{-0.16}_{+0.33}$	$^{+0.15}_{-0.35}$	± 0.0	$^{+0.03}_{-0.29}$	$^{+1.11}_{-1.11}$
$\mathrm{T\bar{T}}$ M(T)=1.6 TeV	$^{+6.53}_{-6.46}$	$^{+0.56}_{-0.56}$	$^{+9.72}_{-9.35}$	$^{+1.16}_{-1.15}$	$^{+0.51}_{-1.71}$	$^{+0.24}_{-0.58}$	$^{-0.00}_{+0.04}$	$^{+0.23}_{-0.54}$	± 0.0	$^{+0.85}_{-0.81}$	$^{+1.15}_{-1.15}$

Table D.2: Summary of systematic uncertainties in the `H1b` **category** in the **electron** channel.

proc./ns. par.	PDF	μ_{RF} scale var.	b tag eff., h.-fl.	b tag eff., l.-fl.	AK8 jet mass sc.	AK8 jet mass res.	JEC	JER	H_{T} rew.	Pileup rew.	e id. eff.
Z + jets	$^{+7.02}_{-6.75}$	$^{+21.91}_{-16.47}$	$^{+7.26}_{-7.02}$	$^{+4.12}_{-4.00}$	$^{+0.62}_{+7.84}$	$^{+1.10}_{+5.70}$	$^{+2.99}_{-3.78}$	$^{+0.20}_{-0.08}$	± 0.0	$^{-0.43}_{+0.37}$	$^{+1.37}_{-1.37}$
Diboson	—	—	$^{+5.86}_{-5.81}$	$^{+5.21}_{-5.20}$	± 0.0	± 0.0	$^{+0.01}_{-0.01}$	$^{+0.01}_{-0.01}$	± 0.0	$^{+19.90}_{-21.41}$	$^{+1.03}_{-1.03}$
Multijet	—	—	$^{+7.96}_{-7.68}$	$^{+2.46}_{-2.40}$	$^{-0.59}_{+0.82}$	$^{-2.18}_{+6.58}$	$^{+1.98}_{-1.17}$	$^{-0.10}_{+0.16}$	± 0.0	$^{-1.26}_{+0.92}$	$^{+1.09}_{-1.09}$
Single t	$^{+1.20}_{-1.20}$	$^{+5.25}_{-5.09}$	$^{+6.97}_{-6.73}$	$^{+0.82}_{-0.82}$	$^{-0.15}_{+3.71}$	$^{+0.43}_{+3.40}$	$^{+1.35}_{-4.47}$	$^{-0.14}_{-0.39}$	± 0.0	$^{+0.97}_{-0.54}$	$^{+1.15}_{-1.15}$
$\mathrm{t\bar{t}}$ + jets	—	—	$^{+6.93}_{-6.71}$	$^{+1.01}_{-1.01}$	$^{-0.80}_{+2.54}$	$^{-0.41}_{+1.77}$	$^{+3.17}_{-3.89}$	$^{-0.10}_{+0.09}$	$^{+18.61}_{-14.53}$	$^{-0.14}_{+0.14}$	$^{+1.04}_{-1.04}$
W + jets	—	—	$^{+8.06}_{-7.78}$	$^{+4.20}_{-4.08}$	$^{+2.34}_{+12.05}$	$^{+2.11}_{+7.95}$	$^{+2.60}_{-4.01}$	$^{-0.03}_{+0.14}$	$^{+19.14}_{-15.23}$	$^{+1.06}_{-1.24}$	$^{+1.39}_{-1.39}$
$\mathrm{T\bar{T}}$ M(T)=0.8 TeV	$^{+3.06}_{-3.04}$	$^{+0.90}_{-0.89}$	$^{+9.14}_{-8.77}$	$^{+0.97}_{-0.97}$	$^{-0.93}_{+1.46}$	$^{-0.32}_{+0.86}$	$^{+0.57}_{-1.15}$	$^{-0.16}_{+0.10}$	± 0.0	$^{-0.27}_{+0.23}$	$^{+0.99}_{-0.99}$
$\mathrm{T\bar{T}}$ M(T)=1.2 TeV	$^{+4.32}_{-4.30}$	$^{+0.71}_{-0.70}$	$^{+9.53}_{-9.12}$	$^{+1.15}_{-1.14}$	$^{-0.39}_{+0.64}$	$^{-0.59}_{+0.79}$	$^{+0.18}_{-0.77}$	$^{-0.18}_{+0.01}$	± 0.0	$^{-0.35}_{+0.54}$	$^{+1.14}_{-1.14}$
$\mathrm{T\bar{T}}$ M(T)=1.6 TeV	$^{+6.14}_{-5.99}$	$^{+0.50}_{-0.49}$	$^{+9.98}_{-9.54}$	$^{+1.31}_{-1.31}$	$^{-0.36}_{+1.58}$	$^{-0.10}_{+0.38}$	$^{+0.06}_{-0.45}$	$^{-0.06}_{+0.01}$	± 0.0	$^{+0.38}_{-0.34}$	$^{+1.22}_{-1.22}$

Table D.3: Summary of systematic uncertainties in `H2b` **category** in the **electron** channel.

proc./ns. par.	PDF	μ_{RF} scale var.	b tag eff., h.-fl.	b tag eff., l.-fl.	AK8 jet mass sc.	AK8 jet mass res.	JEC	JER	H_T rew.	Pileup rew.	e id. eff.
Z + jets	$^{+9.05}_{-8.75}$	$^{+22.82}_{-16.96}$	$^{+16.50}_{-15.72}$	$^{+5.40}_{-5.14}$	$^{+4.24}_{+5.23}$	$^{+4.59}_{+7.99}$	$^{+0.47}_{-7.95}$	± 0.0	± 0.0	$^{+4.32}_{-4.57}$	$^{+1.38}_{-1.38}$
Diboson	—	—	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$	$^{+nan}_{+nan}$
Multijet	—	—	$^{+12.98}_{-12.44}$	$^{+4.52}_{-4.50}$	$^{+2.73}_{+8.52}$	$^{+0.00}_{+12.02}$	$^{+0.06}_{-12.57}$	$^{-12.47}_{+0.01}$	± 0.0	$^{+2.18}_{+0.26}$	$^{+0.92}_{-0.92}$
Single t	$^{+0.63}_{-1.05}$	$^{+8.24}_{-6.80}$	$^{+17.20}_{-16.38}$	$^{+1.82}_{-1.82}$	$^{-1.21}_{+6.71}$	$^{+0.00}_{+2.36}$	$^{-0.04}_{-3.82}$	$^{+4.21}_{-3.58}$	± 0.0	$^{-3.20}_{+3.38}$	$^{+1.76}_{-1.76}$
$t\bar{t}$ + jets	—	—	$^{+14.36}_{-13.79}$	$^{+3.11}_{-3.09}$	$^{-1.01}_{+8.27}$	$^{+0.75}_{+5.12}$	$^{+2.74}_{-5.16}$	$^{+0.03}_{-0.01}$	$^{+20.35}_{-15.65}$	$^{+1.26}_{-1.55}$	$^{+1.10}_{-1.10}$
W + jets	—	—	$^{+14.43}_{-13.89}$	$^{+3.68}_{-3.62}$	$^{+11.23}_{+14.59}$	$^{+10.86}_{+4.20}$	$^{+2.35}_{-3.18}$	$^{+0.00}_{+0.44}$	$^{+20.55}_{-16.17}$	$^{+0.75}_{+2.34}$	$^{+1.65}_{-1.65}$
$T\bar{T}$ M(T)=0.8 TeV	$^{+3.90}_{-3.81}$	$^{+1.00}_{-1.01}$	$^{+13.39}_{-12.74}$	$^{+1.20}_{-1.19}$	$^{-0.68}_{+1.39}$	$^{-0.42}_{+0.76}$	$^{+0.94}_{-0.60}$	$^{+0.01}_{+0.01}$	± 0.0	$^{-0.16}_{+0.10}$	$^{+1.00}_{-1.00}$
$T\bar{T}$ M(T)=1.2 TeV	$^{+4.90}_{-4.90}$	$^{+0.66}_{-0.68}$	$^{+12.74}_{-12.14}$	$^{+1.12}_{-1.12}$	$^{-0.58}_{+0.82}$	$^{-0.35}_{+0.95}$	$^{+0.61}_{-0.52}$	$^{-0.26}_{+0.11}$	± 0.0	$^{+0.15}_{+0.27}$	$^{+1.02}_{-1.02}$
$T\bar{T}$ M(T)=1.6 TeV	$^{+9.39}_{-9.12}$	$^{+0.71}_{-0.72}$	$^{+13.49}_{-12.79}$	$^{+1.34}_{-1.33}$	$^{-0.59}_{+1.00}$	$^{-0.28}_{+0.29}$	$^{-0.11}_{-0.66}$	$^{+0.01}_{-0.00}$	± 0.0	$^{-0.09}_{+0.68}$	$^{+1.16}_{-1.16}$

Table D.4: Summary of systematic uncertainties in the `OH` **category** in the **muon** channel.

proc./ns. par.	PDF	μ_{RF} scale var.	b tag eff., h.-fl.	b tag eff., l.-fl.	AK8 jet mass sc.	AK8 jet mass res.	JEC	JER	H_T rew.	Pileup rew.	μ id. eff.	μ trg. eff.
Z + jets	$^{+6.27}_{-6.10}$	$^{+20.07}_{-15.26}$	$^{+2.70}_{-2.68}$	$^{+3.30}_{-3.29}$	$^{-0.20}_{-0.38}$	$^{-0.28}_{-0.53}$	$^{+1.94}_{-3.48}$	$^{+2.44}_{-2.86}$	± 0.0	$^{+0.05}_{-0.35}$	$^{+1.56}_{-1.56}$	$^{+0.85}_{-0.85}$
Diboson	—	—	$^{+2.97}_{-2.96}$	$^{+3.10}_{-3.10}$	$^{+0.00}_{-4.20}$	± 0.0	$^{+3.54}_{-0.04}$	± 0.0	± 0.0	$^{+6.19}_{-3.72}$	$^{+1.69}_{-1.69}$	$^{+0.92}_{-0.92}$
Multijet	—	—	$^{+3.55}_{-3.54}$	$^{+0.98}_{-0.98}$	$^{+0.01}_{-0.27}$	$^{+0.14}_{-0.22}$	$^{+2.63}_{-3.97}$	$^{+0.54}_{-0.32}$	± 0.0	$^{+0.09}_{-0.87}$	$^{+1.41}_{-1.41}$	$^{+0.76}_{-0.76}$
Single t	$^{+0.91}_{-0.91}$	$^{+1.78}_{-2.57}$	$^{+3.85}_{-3.80}$	$^{+0.49}_{-0.49}$	$^{+0.16}_{+0.09}$	$^{+0.36}_{-0.27}$	$^{+2.37}_{-2.53}$	$^{+0.37}_{-0.34}$	± 0.0	$^{-0.63}_{+0.97}$	$^{+1.55}_{-1.55}$	$^{+0.84}_{-0.84}$
$t\bar{t}$ + jets	—	—	$^{+4.20}_{-4.14}$	$^{+0.70}_{-0.70}$	$^{+0.19}_{-0.71}$	$^{+0.11}_{-0.46}$	$^{+2.15}_{-2.75}$	$^{+0.33}_{-0.39}$	$^{+15.39}_{-11.97}$	$^{+0.07}_{-0.07}$	$^{+1.46}_{-1.46}$	$^{+0.79}_{-0.79}$
W + jets	—	—	$^{+2.90}_{-2.89}$	$^{+2.65}_{-2.65}$	$^{-0.02}_{-0.30}$	$^{-0.05}_{-0.22}$	$^{+2.10}_{-3.37}$	$^{+0.56}_{-0.45}$	$^{+16.58}_{-13.14}$	$^{+0.29}_{-0.27}$	$^{+1.57}_{-1.57}$	$^{+0.86}_{-0.86}$
$T\bar{T}$ M(T)=0.8 TeV	$^{+2.37}_{-2.46}$	$^{+0.81}_{-0.79}$	$^{+7.33}_{-7.10}$	$^{+0.83}_{-0.83}$	$^{+0.58}_{-1.17}$	$^{+0.47}_{-0.77}$	$^{-0.60}_{-0.17}$	$^{+0.39}_{-0.22}$	± 0.0	$^{+0.26}_{+0.02}$	$^{+1.46}_{-1.46}$	$^{+0.79}_{-0.79}$
$T\bar{T}$ M(T)=1.2 TeV	$^{+3.75}_{-3.83}$	$^{+0.65}_{-0.65}$	$^{+8.41}_{-8.12}$	$^{+0.88}_{-0.88}$	$^{+0.81}_{-1.71}$	$^{+0.38}_{-1.01}$	$^{-0.36}_{-0.03}$	$^{+0.37}_{-0.29}$	± 0.0	$^{+0.15}_{-0.50}$	$^{+1.59}_{-1.59}$	$^{+0.86}_{-0.86}$
$T\bar{T}$ M(T)=1.6 TeV	$^{+5.93}_{-5.72}$	$^{+0.57}_{-0.57}$	$^{+9.57}_{-9.20}$	$^{+1.05}_{-1.05}$	$^{+0.33}_{-0.97}$	$^{+0.14}_{-0.24}$	$^{-0.12}_{+0.21}$	$^{+0.25}_{-0.45}$	± 0.0	$^{+0.52}_{-0.68}$	$^{+1.68}_{-1.68}$	$^{+0.90}_{-0.90}$

Table D.5: Summary of systematic uncertainties in the `H1b` **category** in the **muon** channel.

proc./ns. par.	PDF	μ_{RF} scale var.	b tag eff., h.-fl.	b tag eff., l.-fl.	AK8 jet mass sc.	AK8 jet mass res.	JEC	JER	H_T rew.	Pileup rew.	μ id. eff.	μ trg. eff.
Z + jets	$^{+7.13}_{-6.91}$	$^{+21.93}_{-16.44}$	$^{+7.98}_{-7.70}$	$^{+3.80}_{-3.70}$	$^{+2.60}_{+7.63}$	$^{+2.24}_{+6.71}$	$^{+2.48}_{-4.23}$	$^{-0.13}_{-0.58}$	± 0.0	$^{+0.05}_{-0.55}$	$^{+1.65}_{-1.65}$	$^{+0.90}_{-0.90}$
Diboson	—	—	$^{+9.07}_{-8.83}$	$^{+4.20}_{-4.20}$	$^{+0.00}_{+138.88}$	± 0.0	$^{+0.04}_{+0.33}$	$^{+0.01}_{-0.01}$	± 0.0	$^{-5.50}_{+0.69}$	$^{+1.76}_{-1.76}$	$^{+0.94}_{-0.94}$
Multijet	—	—	$^{+6.80}_{-6.60}$	$^{+1.46}_{-1.45}$	$^{-0.22}_{+21.78}$	$^{+0.29}_{+0.61}$	$^{-0.08}_{-2.61}$	± 0.0	± 0.0	$^{-9.55}_{+9.57}$	$^{+1.19}_{-1.19}$	$^{+0.64}_{-0.64}$
Single t	$^{+1.12}_{-1.12}$	$^{+4.90}_{-4.14}$	$^{+7.10}_{-6.87}$	$^{+0.95}_{-0.95}$	$^{-0.94}_{+2.03}$	$^{-1.25}_{+2.11}$	$^{+2.86}_{-3.25}$	$^{+0.15}_{-0.04}$	± 0.0	$^{+0.96}_{-1.73}$	$^{+1.59}_{-1.59}$	$^{+0.86}_{-0.86}$
$t\bar{t}$ + jets	—	—	$^{+6.91}_{-6.69}$	$^{+0.95}_{-0.95}$	$^{-0.81}_{+2.74}$	$^{-0.51}_{+1.78}$	$^{+3.10}_{-3.94}$	$^{-0.05}_{+0.02}$	$^{+16.14}_{-12.79}$	$^{-0.05}_{+0.07}$	$^{+1.57}_{-1.57}$	$^{+0.85}_{-0.85}$
W + jets	—	—	$^{+8.21}_{-7.93}$	$^{+4.19}_{-4.07}$	$^{+1.41}_{+10.37}$	$^{+2.47}_{+7.98}$	$^{+2.71}_{-3.22}$	$^{-0.05}_{+0.09}$	$^{+17.99}_{-14.36}$	$^{+0.50}_{-0.76}$	$^{+1.54}_{-1.54}$	$^{+0.84}_{-0.84}$
$T\bar{T}$ M(T)=0.8 TeV	$^{+3.04}_{-3.01}$	$^{+0.95}_{-0.91}$	$^{+8.89}_{-8.53}$	$^{+0.93}_{-0.92}$	$^{-0.74}_{+1.23}$	$^{-0.61}_{+1.06}$	$^{+0.75}_{-0.96}$	$^{-0.03}_{+0.07}$	± 0.0	$^{+0.50}_{-0.62}$	$^{+1.53}_{-1.53}$	$^{+0.82}_{-0.82}$
$T\bar{T}$ M(T)=1.2 TeV	$^{+3.45}_{-3.49}$	$^{+0.65}_{-0.66}$	$^{+9.30}_{-8.91}$	$^{+1.06}_{-1.06}$	$^{-0.57}_{+1.31}$	$^{-0.19}_{+0.72}$	$^{+0.47}_{-0.26}$	$^{-0.02}_{+0.05}$	± 0.0	$^{-0.48}_{+0.44}$	$^{+1.64}_{-1.64}$	$^{+0.88}_{-0.88}$
$T\bar{T}$ M(T)=1.6 TeV	$^{+5.04}_{-5.12}$	$^{+0.47}_{-0.48}$	$^{+9.98}_{-9.54}$	$^{+1.22}_{-1.21}$	$^{-0.42}_{+0.82}$	$^{-0.12}_{+0.17}$	$^{+0.18}_{-0.48}$	$^{-0.06}_{+0.10}$	± 0.0	$^{-0.13}_{+0.10}$	$^{+1.72}_{-1.72}$	$^{+0.92}_{-0.92}$

Table D.6: Summary of systematic uncertainties in the `H2b` **category** in the **muon** channel.

proc./ns. par.	PDF	μ_{RF} scale var.	b tag eff., h.-fl.	b tag eff., l.-fl.	AK8 jet mass sc.	AK8 jet mass res.	JEC	JER	H_T rew.	Pileup rew.	μ id. eff.	μ trg. eff.
Z + jets	$^{+13.95}_{-13.65}$	$^{+23.21}_{-17.40}$	$^{+14.62}_{-14.04}$	$^{+4.54}_{-4.47}$	$^{+8.80}_{+28.15}$	$^{+3.61}_{+35.67}$	$^{+2.91}_{-7.24}$	± 0.0	± 0.0	$^{+1.67}_{+0.74}$	$^{+1.67}_{-1.67}$	$^{+0.90}_{-0.90}$
Diboson	—	—	$^{+23.79}_{-21.88}$	$^{+0.26}_{-0.26}$	± 0.0	± 0.0	$^{+0.01}_{-0.01}$	± 0.0	± 0.0	$^{-22.78}_{+22.27}$	$^{+2.09}_{-2.09}$	$^{+1.31}_{-1.31}$
Multijet	—	—	$^{+13.40}_{-12.83}$	$^{+1.02}_{-1.01}$	$^{-0.06}_{+8.50}$	$^{-0.01}_{+1.37}$	$^{+11.33}_{-3.49}$	± 0.0	± 0.0	$^{-4.25}_{+8.11}$	$^{+1.19}_{-1.19}$	$^{+0.66}_{-0.66}$
Single t	$^{+1.54}_{-1.54}$	$^{+8.09}_{-7.67}$	$^{+16.88}_{-16.15}$	$^{+2.89}_{-2.89}$	$^{-2.78}_{-3.35}$	$^{-4.78}_{+2.99}$	$^{+0.71}_{+1.53}$	± 0.0	± 0.0	$^{-4.06}_{+3.74}$	$^{+1.64}_{-1.64}$	$^{+0.90}_{-0.90}$
$t\bar{t}$ + jets	—	—	$^{+14.31}_{-13.73}$	$^{+3.00}_{-2.97}$	$^{-1.78}_{+7.93}$	$^{+0.30}_{+4.99}$	$^{+4.09}_{-4.37}$	$^{-0.02}_{-0.02}$	$^{+19.28}_{-14.73}$	$^{-0.10}_{-0.05}$	$^{+1.39}_{-1.39}$	$^{+0.75}_{-0.75}$
W + jets	—	—	$^{+14.56}_{-13.96}$	$^{+3.90}_{-3.86}$	$^{+1.32}_{+10.89}$	$^{-0.62}_{+12.23}$	$^{+0.84}_{-3.36}$	$^{-0.00}_{+1.33}$	$^{+20.60}_{-16.02}$	$^{+0.50}_{+0.32}$	$^{+1.50}_{-1.50}$	$^{+0.84}_{-0.84}$
$T\bar{T}$ M(T)=0.8 TeV	$^{+3.80}_{-3.76}$	$^{+0.90}_{-0.90}$	$^{+13.01}_{-12.40}$	$^{+0.71}_{-0.71}$	$^{-0.56}_{+1.42}$	$^{-0.78}_{+0.88}$	$^{+0.64}_{-0.77}$	$^{-0.08}_{+0.01}$	± 0.0	$^{+0.24}_{+0.18}$	$^{+1.44}_{-1.44}$	$^{+0.78}_{-0.78}$
$T\bar{T}$ M(T)=1.2 TeV	$^{+5.36}_{-5.37}$	$^{+0.87}_{-0.89}$	$^{+13.32}_{-12.68}$	$^{+1.10}_{-1.10}$	$^{-0.52}_{+1.02}$	$^{-0.28}_{+0.51}$	$^{+0.15}_{-0.61}$	$^{-0.04}_{+0.06}$	± 0.0	$^{+0.45}_{-0.56}$	$^{+1.62}_{-1.62}$	$^{+0.87}_{-0.87}$
$T\bar{T}$ M(T)=1.6 TeV	$^{+6.90}_{-7.02}$	$^{+0.48}_{-0.49}$	$^{+13.25}_{-12.59}$	$^{+1.10}_{-1.10}$	$^{+0.19}_{+0.49}$	$^{+0.14}_{+0.14}$	$^{+0.07}_{-0.18}$	$^{-0.00}_{+0.02}$	± 0.0	$^{-0.38}_{+0.07}$	$^{+1.69}_{-1.69}$	$^{+0.91}_{-0.91}$

E Systematic uncertainty study in the combination

To confirm that the treatment of systematic uncertainties (described in Section 7.2) in the combination of the two VLQ search channels is valid, a maximum likelihood fit of the background-only hypothesis (see App. A.1) is performed by fitting the corresponding observables, $\min[\mathrm{M}(\ell, b)]$ in the boosted W channel and S_{T} in the boosted H channel, in all 20 event categories simultaneously. Systematic uncertainties that affect only the rates of physical processes are assigned log-normal priors in this study while all other systematic uncertainties are assigned Gaussian priors.

The resulting variations of the post-fit values of the nuisance parameters in units of one standard deviation of their prior uncertainties are displayed in Fig. E.1. All variations are within one standard deviation of the prior uncertainty, pointing at a healthy behaviour of the MLE fit. As further cross-check, the resulting correlation matrix after the MLE fit is inspected to see whether there exist any unphysical correlations. Except for a few, most likely accidental correlations, this is not the case and further validates the treatment of systematic uncertainties between the two search channels.

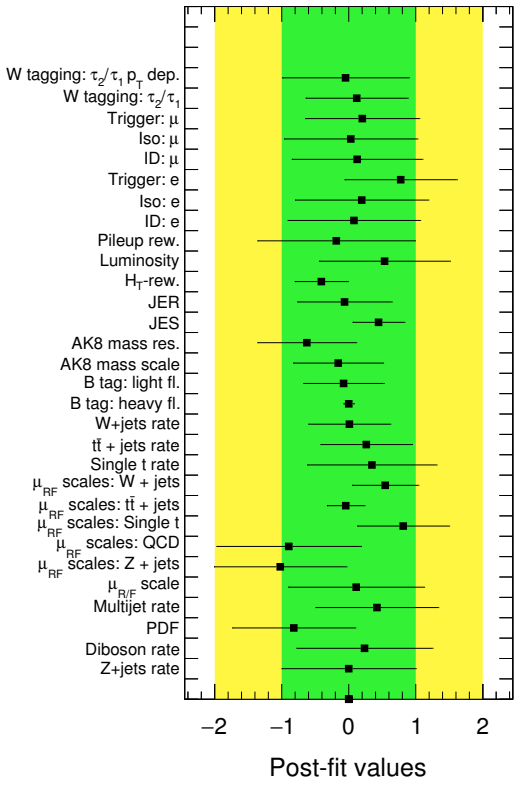

Figure E.1: Post-fit values of the nuisance parameters of the background model after fitting the background-only hypothesis in all categories from the boosted W and boosted H channels simultaneously. The post-fit values are expressed as deviations from their pre-fit values in units of the corresponding prior uncertainty. The uncertainties in these values, indicated by solid lines, represent the constrained uncertainties after the MLE fit. The bands representing one and two standard deviations of the prior uncertainty of each parameter are shown in green and yellow, respectively.

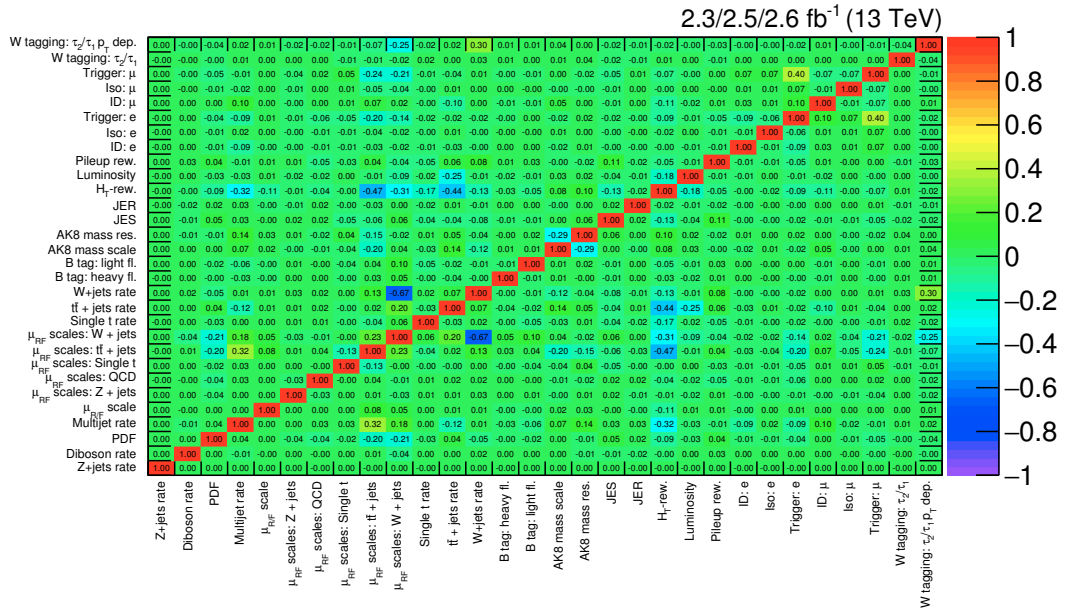

Figure E.2: Correlation matrix for all nuisance parameters considered in the maximum-likelihood fit. A value of zero indicates that the corresponding nuisance parameters are found to be uncorrelated in the fit while values of +1 (-1) denote that they are fully correlated (anti-correlated) in the fit.

Bibliography

[1] CMS Collaboration, "Search for vector-like quark pair production in final states with leptons and boosted Higgs bosons at $\sqrt{s} = 13$ TeV", CMS-PAS-B2G-16-011, 2016.

[2] CMS Collaboration, "Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure techniques at sqrt(s) = 13 TeV", *submitted to: JHEP* (2017) arXiv:1706.03408.

[3] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588. doi:10.1016/0029-5582(61)90469-2.

[4] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266. doi:10.1103/PhysRevLett.19.1264.

[5] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367–377.

[6] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133. doi:10.1016/0031-9163(64)91136-9.

[7] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509. doi:10.1103/PhysRevLett.13.508.

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

[9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587. doi:10.1103/PhysRevLett.13.585.

[10] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons", *Phys. Rev.* **145** (1966) 1156–1163. doi:10.1103/PhysRev.145.1156.

[11] T. W. B. Kibble, "Symmetry breaking in nonAbelian gauge theories", *Phys. Rev.* **155** (1967) 1554–1561. doi:10.1103/PhysRev.155.1554.

[12] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **06** (2013) 081, arXiv:1303.4571. doi:10.1007/JHEP06(2013)081.

[13] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021.

[14] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

[15] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory". 1995.

[16] S. Weinberg, "The Quantum theory of fields. Vol. 1: Foundations". Cambridge University Press, 2005.

[17] SNO Collaboration, "Measurement of the rate of nu/e + d –¿ p + p + e- interactions produced by B-8 solar neutrinos at the Sudbury Neutrino Observatory", *Phys. Atom. Nucl.* **65** (2002) 2147–2155. [Yad. Fiz.65,2210(2002)]. doi:10.1134/1.1530292.

[18] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, arXiv:hep-ex/9807003. doi:10.1103/PhysRevLett.81.1562.

[19] Daya Bay Collaboration, "Observation of electron-antineutrino disappearance at Daya Bay", *Phys. Rev. Lett.* **108** (2012) 171803, arXiv:1203.1669. doi:10.1103/PhysRevLett.108.171803.

[20] T2K Collaboration, "Evidence of Electron Neutrino Appearance in a Muon Neutrino Beam", *Phys. Rev. D* **88** (2013), no. 3, 032002, arXiv:1304.0841. doi:10.1103/PhysRevD.88.032002.

[21] OPERA Collaboration, "Observation of a first ν_τ candidate in the OPERA experiment in the CNGS beam", *Phys. Lett. B* **691** (2010) 138–145, arXiv:1006.1623. doi:10.1016/j.physletb.2010.06.022.

[22] Cush, "Standard Model of Elementary Particles.svg", *Wikipedia* (2017) https://en.wikipedia.org/wiki/Standard_Model – accessed Apr. 18, 2017.

[23] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531–533. [,648(1963)]. doi:10.1103/PhysRevLett.10.531.

[24] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657. doi:10.1143/PTP.49.652.

[25] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys. C* **40** (2016), no. 10, 100001. doi:10.1088/1674-1137/40/10/100001.

[26] Y. Nambu, "Quasiparticles and Gauge Invariance in the Theory of Superconductivity", *Phys. Rev.* **117** (1960) 648–663. doi:10.1103/PhysRev.117.648.

[27] J. Goldstone, "Field Theories with Superconductor Solutions", *Nuovo Cim.* **19** (1961) 154–164. doi:10.1007/BF02812722.

[28] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries", *Phys. Rev.* **127** (1962) 965–970. doi:10.1103/PhysRev.127.965.

[29] S. Dittmaier et al., "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", arXiv:1201.3084. doi:10.5170/CERN-2012-002.

[30] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints", *Phys. Rept.* **405** (2005) 279–390, arXiv:hep-ph/0404175. doi:10.1016/j.physrep.2004.08.031.

[31] D. D. Cline and I. S. on Sources of Dark Matter in the Universe, "Proceedings of the 1st International Symposium on Sources of Dark Matter in the Universe : 16-18 February 1994, Bel Air, California". World Scientific, 1995.

[32] J. Silk et al., "Particle Dark Matter: Observations, Models and Searches". Cambridge Univ. Press, Cambridge, 2010.

[33] G. Bertone and D. Merritt, "Dark matter dynamics and indirect detection", *Mod. Phys. Lett. A* **20** (2005) 1021, arXiv:astro-ph/0504422. doi:10.1142/S0217732305017391.

[34] A. D. Sakharov, "Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe", *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35. [Usp. Fiz. Nauk161,61(1991)]. doi:10.1070/PU1991v034n05ABEH002497.

[35] A. G. Cohen, D. B. Kaplan, and A. E. Nelson, "Progress in electroweak baryogenesis", *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 27–70, arXiv:hep-ph/9302210. doi:10.1146/annurev.ns.43.120193.000331.

[36] G. Hooft, "Recent developments in gauge theories". NATO ASI series: Physics. Plenum Press, 1980.

[37] G. F. R. Ellis, U. Kirchner, and S. J. Stoeger, William R., "Multiverses and physical cosmology", *Mon. Not. Roy. Astron. Soc.* **347** (2004) 921–936, arXiv:astro-ph/0305292. doi:10.1111/j.1365-2966.2004.07261.x.

[38] B. Carter, "Large number coincidences and the anthropic principle in cosmology", *IAU Symp.* **63** (1974) 291.

[39] S. P. Martin, "A Supersymmetry primer", arXiv:hep-ph/9709356. [Adv. Ser. Direct. High Energy Phys.18,1(1998)]. doi:10.1142/9789812839657_0001, 10.1142/9789814307505_0001.

[40] J. L. Feng, "Naturalness and the Status of Supersymmetry", *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 351–382, arXiv:1302.6587. doi:10.1146/annurev-nucl-102010-130447.

[41] M. J. Dugan, H. Georgi, and D. B. Kaplan, "Anatomy of a Composite Higgs Model", *Nucl. Phys. B* **254** (1985) 299–326. doi:10.1016/0550-3213(85)90221-4.

[42] S. Dimopoulos and J. Preskill, "Massless Composites With Massive Constituents", *Nucl. Phys. B* **199** (1982) 206–222. doi:10.1016/0550-3213(82)90345-5.

[43] D. B. Kaplan, H. Georgi, and S. Dimopoulos, "Composite Higgs Scalars", *Phys. Lett. B* **136** (1984) 187–190. doi:10.1016/0370-2693(84)91178-X.

[44] T. Banks, "CONSTRAINTS ON SU(2) x U(1) BREAKING BY VACUUM MISALIGNMENT", *Nucl. Phys. B* **243** (1984) 125–130. doi:10.1016/0550-3213(84)90389-4.

[45] H. Georgi, D. B. Kaplan, and P. Galison, "Calculation of the Composite Higgs Mass", *Phys. Lett. B* **143** (1984) 152–154. doi:10.1016/0370-2693(84)90823-2.

[46] H. Georgi and D. B. Kaplan, "Composite Higgs and Custodial SU(2)", *Phys. Lett. B* **145** (1984) 216–220. doi:10.1016/0370-2693(84)90341-1.

[47] D. B. Kaplan and H. Georgi, "SU(2) x U(1) Breaking by Vacuum Misalignment", *Phys. Lett. B* **136** (1984) 183–186. doi:10.1016/0370-2693(84)91177-8.

[48] B. Bellazzini, C. Csáki, and J. Serra, "Composite Higgses", *Eur. Phys. J. C* **74** (2014), no. 5, 2766, arXiv:1401.2457. doi:10.1140/epjc/s10052-014-2766-x.

[49] R. Contino, "The Higgs as a Composite Nambu-Goldstone Boson", in *Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, pp. 235–306. 2011. arXiv:1005.4269.

[50] G. Panico and A. Wulzer, "The Composite Nambu-Goldstone Higgs", *Lect. Notes Phys.* **913** (2016) pp.1–316, arXiv:1506.01961. doi:10.1007/978-3-319-22617-0.

[51] R. Foadi, M. T. Frandsen, and F. Sannino, "125 GeV Higgs boson from a not so light technicolor scalar", *Phys. Rev. D* **87** (2013), no. 9, 095001, arXiv:1211.1083. doi:10.1103/PhysRevD.87.095001.

[52] J. Galloway, J. A. Evans, M. A. Luty et al., "Minimal Conformal Technicolor and Precision Electroweak Tests", *JHEP* **10** (2010) 086, arXiv:1001.1361. doi:10.1007/JHEP10(2010)086.

[53] N. Arkani-Hamed, A. G. Cohen, E. Katz et al., "The Minimal moose for a little Higgs", *JHEP* **08** (2002) 021, arXiv:hep-ph/0206020. doi:10.1088/1126-6708/2002/08/021.

[54] N. Arkani-Hamed, A. G. Cohen, E. Katz et al., "The Littlest Higgs", *JHEP* **07** (2002) 034, arXiv:hep-ph/0206021. doi:10.1088/1126-6708/2002/07/034.

[55] I. Low, W. Skiba, and D. Tucker-Smith, "Little Higgses from an antisymmetric condensate", *Phys. Rev. D* **66** (2002) 072001, arXiv:hep-ph/0207243. doi:10.1103/PhysRevD.66.072001.

[56] D. E. Kaplan and M. Schmaltz, "The Little Higgs from a simple group", *JHEP* **10** (2003) 039, arXiv:hep-ph/0302049. doi:10.1088/1126-6708/2003/10/039.

[57] S. Chang and J. G. Wacker, "Little Higgs and custodial SU(2)", *Phys. Rev. D* **69** (2004) 035002, arXiv:hep-ph/0303001. doi:10.1103/PhysRevD.69.035002.

[58] W. Skiba and J. Terning, "A Simple model of two little Higgses", *Phys. Rev. D* **68** (2003) 075001, arXiv:hep-ph/0305302. doi:10.1103/PhysRevD.68.075001.

[59] M. Schmaltz, "The Simplest little Higgs", *JHEP* **08** (2004) 056, arXiv:hep-ph/0407143. doi:10.1088/1126-6708/2004/08/056.

[60] M. Schmaltz, D. Stolarski, and J. Thaler, "The Bestest Little Higgs", *JHEP* **09** (2010) 018, arXiv:1006.1356. doi:10.1007/JHEP09(2010)018.

[61] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, "Electroweak symmetry breaking from dimensional deconstruction", *Phys. Lett. B* **513** (2001) 232–240, arXiv:hep-ph/0105239. doi:10.1016/S0370-2693(01)00741-9.

[62] R. Contino, T. Kramer, M. Son et al., "Warped/composite phenomenology simplified", *JHEP* **05** (2007) 074, arXiv:hep-ph/0612180. doi:10.1088/1126-6708/2007/05/074.

[63] G. Panico and A. Wulzer, "The Discrete Composite Higgs Model", *JHEP* **09** (2011) 135, arXiv:1106.2719. doi:10.1007/JHEP09(2011)135.

[64] S. De Curtis, M. Redi, and A. Tesi, "The 4D Composite Higgs", *JHEP* **04** (2012) 042, arXiv:1110.1613. doi:10.1007/JHEP04(2012)042.

[65] K. Agashe, R. Contino, and A. Pomarol, "The Minimal composite Higgs model", *Nucl. Phys. B* **719** (2005) 165–187, arXiv:hep-ph/0412089. doi:10.1016/j.nuclphysb.2005.04.035.

[66] D. B. Kaplan, "Flavor at SSC energies: A New mechanism for dynamically generated fermion masses", *Nucl. Phys. B* **365** (1991) 259–278. doi:10.1016/S0550-3213(05)80021-5.

[67] S. J. Huber and Q. Shafi, "Fermion masses, mixings and proton decay in a Randall-Sundrum model", *Phys. Lett. B* **498** (2001) 256–262, arXiv:hep-ph/0010195. doi:10.1016/S0370-2693(00)01399-X.

[68] K. Agashe, G. Perez, and A. Soni, "Flavor structure of warped extra dimension models", *Phys. Rev. D* **71** (2005) 016002, arXiv:hep-ph/0408134. doi:10.1103/PhysRevD.71.016002.

[69] R. Contino, L. Da Rold, and A. Pomarol, "Light custodians in natural composite Higgs models", *Phys. Rev. D* **75** (2007) 055014, arXiv:hep-ph/0612048. doi:10.1103/PhysRevD.75.055014.

[70] O. Matsedonskyi, G. Panico, and A. Wulzer, "Light Top Partners for a Light Composite Higgs", *JHEP* **01** (2013) 164, arXiv:1204.6333. doi:10.1007/JHEP01(2013)164.

[71] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370–3373, arXiv:hep-ph/9905221. doi:10.1103/PhysRevLett.83.3370.

[72] R. Sundrum, "Tasi 2004 lectures: To the fifth dimension and back", in *Theoretical Advanced Study Institute in Elementary Particle Physics: Many Dimensions of String Theory (TASI 2005) Boulder, Colorado, June 5-July 1, 2005*, pp. 585–630. 2005. arXiv:hep-th/0508134. [,585(2005)].

[73] C. Csaki, "TASI lectures on extra dimensions and branes", in *From fields to strings: Circumnavigating theoretical physics. Ian Kogan memorial collection (3 volume set)*, pp. 605–698. 2004. arXiv:hep-ph/0404096. [,967(2004)].

[74] J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity", *Int. J. Theor. Phys.* **38** (1999) 1113–1133, arXiv:hep-th/9711200. [Adv. Theor. Math. Phys.2,231(1998)]. doi:10.1023/A:1026654312961.

[75] 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, arXiv:1209.1101. doi:10.1103/PhysRevLett.109.241802.

[76] A. Djouadi and A. Lenz, "Sealing the fate of a fourth generation of fermions", *Phys. Lett. B* **715** (2012) 310–314, arXiv:1204.1252. doi:10.1016/j.physletb.2012.07.060.

[77] F. del Aguila, L. Ametller, G. L. Kane et al., "Vector Like Fermion and Standard Higgs Production at Hadron Colliders", *Nucl. Phys. B* **334** (1990) 1–23. doi:10.1016/0550-3213(90)90655-W.

[78] A. Falkowski, D. M. Straub, and A. Vicente, "Vector-like leptons: Higgs decays and collider phenomenology", *JHEP* **05** (2014) 092, arXiv:1312.5329. doi:10.1007/JHEP05(2014)092.

[79] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer et al., "Handbook of vectorlike quarks: Mixing and single production", *Phys. Rev. D* **88** (2013), no. 9, 094010, arXiv:1306.0572. doi:10.1103/PhysRevD.88.094010.

[80] J. A. Aguilar-Saavedra, "Mixing with vector-like quarks: constraints and expectations", *EPJ Web Conf.* **60** (2013) 16012, arXiv:1306.4432. doi:10.1051/epjconf/20136016012.

[81] J. A. Aguilar-Saavedra, "Identifying top partners at LHC", *JHEP* **11** (2009) 030, arXiv:0907.3155. doi:10.1088/1126-6708/2009/11/030.

[82] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014) 2930, arXiv:1112.5675. doi:10.1016/j.cpc.2014.06.021.

[83] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004. doi:10.1103/PhysRevLett.110.252004.

[84] M. Czakon and A. Mitov, "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction", *JHEP* **01** (2013) 080. doi:10.1007/JHEP01(2013)080.

[85] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels", *JHEP* **12** (2012) 054. doi:10.1007/JHEP12(2012)054.

[86] P. Bärnreuther, M. Czakon, and A. Mitov, "Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$", *Phys. Rev. Lett.* **109** (2012) 132001. doi:10.1103/PhysRevLett.109.132001.

[87] M. Cacciari, M. Czakon, M. Mangano et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Phys. Lett. B* **710** (2012) 612. doi:10.1016/j.physletb.2012.03.013.

[88] J. Ellis, "TikZ-Feynman: Feynman diagrams with TikZ", *Comput. Phys. Commun.* **210** (2017) 103–123, arXiv:1601.05437. doi:10.1016/j.cpc.2016.08.019.

[89] Y. Nir and D. J. Silverman, "Z Mediated Flavor Changing Neutral Currents and Their Implications for CP Asymmetries in B^0 Decays", *Phys. Rev. D* **42** (1990) 1477–1484. doi:10.1103/PhysRevD.42.1477.

[90] G. C. Branco, T. Morozumi, P. A. Parada et al., "CP asymmetries in B0 decays in the presence of flavor changing neutral currents", *Phys. Rev. D* **48** (1993) 1167–1175. doi:10.1103/PhysRevD.48.1167.

[91] F. del Aguila, J. A. Aguilar-Saavedra, and G. C. Branco, "CP violation from new quarks in the chiral limit", *Nucl. Phys. B* **510** (1998) 39–60, arXiv:hep-ph/9703410. doi:10.1016/S0550-3213(97)00708-6, 10.1016/S0550-3213(98)81003-1.

[92] G. Barenboim, F. J. Botella, G. C. Branco et al., "How sensitive to FCNC can B0 CP asymmetries be?", *Phys. Lett. B* **422** (1998) 277–286, arXiv:hep-ph/9709369. doi:10.1016/S0370-2693(97)01515-3.

[93] F. del Aguila, M. Perez-Victoria, and J. Santiago, "Observable contributions of new exotic quarks to quark mixing", *JHEP* **09** (2000) 011, arXiv:hep-ph/0007316. doi:10.1088/1126-6708/2000/09/011.

[94] F. del Aguila, M. Perez-Victoria, and J. Santiago, "Effective description of quark mixing", *Phys. Lett. B* **492** (2000) 98–106, arXiv:hep-ph/0007160. doi:10.1016/S0370-2693(00)01071-6.

[95] A. Atre, G. Azuelos, M. Carena et al., "Model-Independent Searches for New Quarks at the LHC", *JHEP* **08** (2011) 080, arXiv:1102.1987. doi:10.1007/JHEP08(2011)080.

[96] A. De Simone, O. Matsedonskyi, R. Rattazzi et al., "A First Top Partner Hunter's Guide", *JHEP* **04** (2013) 004, arXiv:1211.5663. doi:10.1007/JHEP04(2013)004.

[97] O. Matsedonskyi, G. Panico, and A. Wulzer, "On the Interpretation of Top Partners Searches", *JHEP* **12** (2014) 097, arXiv:1409.0100. doi:10.1007/JHEP12(2014)097.

[98] ATLAS Collaboration, "Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **08** (2015) 105, arXiv:1505.04306. doi:10.1007/JHEP08(2015)105.

[99] CMS Collaboration, "Search for vector-like charge 2/3 T quarks in proton-proton collisions at sqrt(s) = 8 TeV", *Phys. Rev. D* **93** (2016), no. 1, 012003, arXiv:1509.04177. doi:10.1103/PhysRevD.93.012003.

[100] CMS Collaboration, "Search for pair-produced vectorlike B quarks in proton-proton collisions at $\sqrt{s} = 8$ TeV", arXiv:1507.07129.

[101] ATLAS Collaboration, "Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **08** (2015) 105, arXiv:1505.04306. doi:10.1007/JHEP08(2015)105.

[102] ATLAS Collaboration, "Search for vector-like B quarks in events with one isolated lepton, missing transverse momentum and jets at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Phys. Rev. D* **91** (2015), no. 11, 112011, arXiv:1503.05425. doi:10.1103/PhysRevD.91.112011.

[103] ATLAS Collaboration, "Search for single production of a vector-like quark via a heavy gluon in the $4b$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett. B* **758** (2016) 249–268, arXiv:1602.06034. doi:10.1016/j.physletb.2016.04.061.

[104] ATLAS Collaboration, "Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Eur. Phys. J. C* **76** (2016), no. 8, 442, arXiv:1602.05606. doi:10.1140/epjc/s10052-016-4281-8.

[105] ATLAS Collaboration, "Search for the production of single vector-like and excited quarks in the Wt final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **02** (2016) 110, arXiv:1510.02664. doi:10.1007/JHEP02(2016)110.

[106] ATLAS Collaboration, "Search for pair and single production of new heavy quarks that decay to a Z boson and a third-generation quark in pp collisions at $\sqrt{s} = 8$

TeV with the ATLAS detector", *JHEP* **11** (2014) 104, arXiv:1409.5500. doi:10.1007/JHEP11(2014)104.

[107] C. Grojean, O. Matsedonskyi, and G. Panico, "Light top partners and precision physics", *JHEP* **10** (2013) 160, arXiv:1306.4655. doi:10.1007/JHEP10(2013)160.

[108] M. E. Peskin and T. Takeuchi, "Estimation of oblique electroweak corrections", *Phys. Rev. D* **46** (1992) 381–409. doi:10.1103/PhysRevD.46.381.

[109] R. Barbieri, A. Pomarol, R. Rattazzi et al., "Electroweak symmetry breaking after LEP-1 and LEP-2", *Nucl. Phys. B* **703** (2004) 127–146, arXiv:hep-ph/0405040. doi:10.1016/j.nuclphysb.2004.10.014.

[110] G. F. Giudice, C. Grojean, A. Pomarol et al., "The Strongly-Interacting Light Higgs", *JHEP* **06** (2007) 045, arXiv:hep-ph/0703164. doi:10.1088/1126-6708/2007/06/045.

[111] L. Lavoura and J. P. Silva, "The Oblique corrections from vector - like singlet and doublet quarks", *Phys. Rev. D* **47** (1993) 2046–2057. doi:10.1103/PhysRevD.47.2046.

[112] M. Baak, M. Goebel, J. Haller et al., "The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC", *Eur. Phys. J. C* **72** (2012) 2205, arXiv:1209.2716. doi:10.1140/epjc/s10052-012-2205-9.

[113] G. Cacciapaglia, C. Csaki, G. Marandella et al., "The Minimal Set of Electroweak Precision Parameters", *Phys. Rev. D* **74** (2006) 033011, arXiv:hep-ph/0604111. doi:10.1103/PhysRevD.74.033011.

[114] R. S. Chivukula, S. B. Selipsky, and E. H. Simmons, "Nonoblique effects in the Z b anti-b vertex from ETC dynamics", *Phys. Rev. Lett.* **69** (1992) 575–577, arXiv:hep-ph/9204214. doi:10.1103/PhysRevLett.69.575.

[115] B. Batell, S. Gori, and L.-T. Wang, "Higgs Couplings and Precision Electroweak Data", *JHEP* **01** (2013) 139, arXiv:1209.6382. doi:10.1007/JHEP01(2013)139.

[116] D. Guadagnoli and G. Isidori, "$B(B_s \to \mu^+\mu^-)$ as an electroweak precision test", *Phys. Lett. B* **724** (2013) 63–67, arXiv:1302.3909. doi:10.1016/j.physletb.2013.05.054.

[117] J. Kearney, A. Pierce, and N. Weiner, "Vectorlike Fermions and Higgs Couplings", *Phys. Rev. D* **86** (2012) 113005, arXiv:1207.7062. doi:10.1103/PhysRevD.86.113005.

[118] A. Falkowski, F. Riva, and A. Urbano, "Higgs at last", *JHEP* **11** (2013) 111, arXiv:1303.1812. doi:10.1007/JHEP11(2013)111.

[119] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (2008), no. 08, S08001.

[120] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[121] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002.

[122] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005. doi:10.1088/1748-0221/3/08/S08005.

[123] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003.

[124] Service graphique, CERN, "Overall view of the LHC. Vue d'ensemble du LHC", OPEN-PHO-CHART-2014-006, Jun, 2014. General Photo.

[125] S. van der Meer, "Calibration of the effective beam height in the ISR", CERN-ISR-PO-68-31. ISR-PO-68-31, 1968.

[126] CMS Collaboration, "CMS luminosity measurement for the 2015 data-taking period", CMS-PAS-LUM-15-001, 2017.

[127] CMS Collaboration, "Public CMS Luminosity Information", 2015. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2015_Proton_Proton_13_TeV_Collis.

[128] CMS Collaboration, "Public CMS Luminosity Information", 2016. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016_Proton_Proton_13_TeV_Collis.

[129] T. Sakuma and T. McCauley, "Detector and event visualization with SketchUp at the CMS experiment", CMS-CR-2013-379. arXiv:1311.4942, Oct, 2013. Comments: 5 pages, 6 figures, Proceedings for CHEP 2013, 20th International Conference on Computing in High Energy and Nuclear Physics.

[130] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, arXiv:1405.6569. doi:10.1088/1748-0221/9/10/P10009.

[131] CMS Collaboration, A. Dominguez, D. Abbaneo, K. Arndt et al., "CMS Technical Design Report for the Pixel Detector Upgrade", CERN-LHCC-2012-016. CMS-TDR-11, Sep, 2012. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

[132] P. Adzic et al., "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST* **2** (2007) P04004. doi:10.1088/1748-0221/2/04/P04004.

[133] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) P09009, arXiv:1306.2016. [JINST8,9009(2013)]. doi:10.1088/1748-0221/8/09/P09009.

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

[135] USCMS, ECAL/HCAL Collaboration, "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c", *Eur. Phys. J. C* **60** (2009) 359–373. [Erratum: Eur. Phys. J.C61,353(2009)]. doi:10.1140/epjc/s10052-009-0959-5, 10.1140/epjc/s10052-009-1024-0.

[136] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC", *JINST* **8** (2013) P11002, arXiv:1306.6905. doi:10.1088/1748-0221/8/11/P11002.

[137] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, arXiv:1609.02366. doi:10.1088/1748-0221/12/01/P01020.

[138] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth. A* **262** (1987) 444–450. doi:10.1016/0168-9002(87)90887-4.

[139] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, arXiv:1101.2599. doi:10.1016/j.physrep.2011.03.005.

[140] P. Skands, "Introduction to QCD", in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012): Boulder, Colorado, June 4-29, 2012*, pp. 341–420. 2013. arXiv:1207.2389.

[141] GEANT4 Collaboration, "GEANT4: A simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.

[142] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, arXiv:hep-ph/0409313. doi:10.1142/9789814503266_0001.

[143] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, arXiv:hep-ph/0409146. doi:10.1088/1126-6708/2004/11/040.

[144] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, arXiv:0709.2092. doi:10.1088/1126-6708/2007/11/070.

[145] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, arXiv:1002.2581. doi:10.1007/JHEP06(2010)043.

[146] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, arXiv:0707.3088. doi:10.1088/1126-6708/2007/09/126.

[147] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5: going beyond", *JHEP* **06** (2011) 1. doi:10.1007/JHEP06(2011)128.

[148] J. Alwall, R. Frederix, S. Frixione et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, arXiv:1405.0301. doi:10.1007/JHEP07(2014)079.

[149] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298–318. doi:10.1016/0550-3213(77)90384-4.

[150] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450. [Yad. Fiz.15,781(1972)].

[151] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641–653. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[152] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, arXiv:1410.8849. doi:10.1007/JHEP04(2015)040.

[153] "HERA - A Proposal for a Large Electron Proton Colliding Beam Facility at DESY", DESY-HERA-81-10, 1981.

[154] A. D. Martin, W. J. Stirling, R. S. Thorne et al., "Parton distributions for the LHC", *Eur. Phys. J. C* **63** (2009) 189–285, arXiv:0901.0002. doi:10.1140/epjc/s10052-009-1072-5.

[155] T. Sjöstrand, "Jet fragmentation of multiparton configurations in a string framework", *Nuclear Physics B* **248** (1984), no. 2, 469 – 502. doi:http://dx.doi.org/10.1016/0550-3213(84)90607-2.

[156] B. Andersson, G. Gustafson, G. Ingelman et al., "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145. doi:10.1016/0370-1573(83)90080-7.

[157] G. Marchesini and B. R. Webber, "Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation", *Nucl. Phys. B* **310** (1988) 461–526. doi:10.1016/0550-3213(88)90089-2.

[158] B. R. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference", *Nucl. Phys. B* **238** (1984) 492–528. doi:10.1016/0550-3213(84)90333-X.

[159] M. L. Mangano, M. Moretti, F. Piccinini et al., "Matching Matrix Elements and Shower Evolution for Top-Quark Production in Hadronic Collisions", *JHEP* **01** (2007) 013, arXiv:hep-ph/0611129. doi:10.1088/1126-6708/2007/01/013.

[160] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, arXiv:1209.6215. doi:10.1007/JHEP12(2012)061.

[161] T. Sjöstrand, S. Mrenna and P. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175. doi:10.1088/1126-6708/2006/05/026.

[162] T. Sjöstrand, S. Ask, J. R. Christiansen et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, arXiv:1410.3012. doi:10.1016/j.cpc.2015.01.024.

[163] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639–707, arXiv:0803.0883. doi:10.1140/epjc/s10052-008-0798-9.

[164] T. Sjöstrand and M. van Zijl, "A Multiple Interaction Model for the Event Structure in Hadron Collisions", *Phys. Rev. D* **36** (1987) 2019. doi:10.1103/PhysRevD.36.2019.

[165] P. Billoir, "Progressive track recognition with a Kalman like fitting procedure", *Comput. Phys. Commun.* **57** (1989) 390–394. doi:10.1016/0010-4655(89)90249-X.

[166] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the Kalman filtering method", *Nucl. Instrum. Meth. A* **294** (1990) 219–228. doi:10.1016/0168-9002(90)91835-Y.

[167] R. Mankel, "A Concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system", *Nucl. Instrum. Meth. A* **395** (1997) 169–184. doi:10.1016/S0168-9002(97)00705-5.

[168] W. Adam, R. Fruhwirth, A. Strandlie et al., "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC", *eConf C* **0303241** (2003) TULT009, arXiv:physics/0306087. [J. Phys.G31,N9(2005)]. doi:10.1088/0954-3899/31/9/N01.

[169] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, arXiv:1502.02701. doi:10.1088/1748-0221/10/06/P06005.

[170] K. Rose, "Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems", in *Proceedings of the IEEE*, pp. 2210–2239. 1998.

[171] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys. G* **34** (2007) N343. doi:10.1088/0954-3899/34/12/N01.

[172] CMS Collaboration, "CMS Tracking Performance Results from early LHC Operation", *Eur. Phys. J. C* **70** (2010) 1165–1192, arXiv:1007.1988. doi:10.1140/epjc/s10052-010-1491-3.

[173] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS-PAS-PFT-09-001, 2009.

[174] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", CMS-PAS-PFT-10-001, 2010.

[175] CMS Collaboration, "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *JINST* **5** (2010) T03022, arXiv:0911.4994. doi:10.1088/1748-0221/5/03/T03022.

[176] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, arXiv:1206.4071. doi:10.1088/1748-0221/7/10/P10002.

[177] CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **01** (2011) 080, arXiv:1012.2466. doi:10.1007/JHEP01(2011)080.

[178] S. Catani, Y. Dokshitzer, M. Seymour et al., "Longitudinally-invariant k_T-clustering algorithms for hadron-hadron collisions", *Nuclear Physics B* **406** (1993), no. 1, 187 – 224. doi:http://dx.doi.org/10.1016/0550-3213(93)90166-M.

[179] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (1993) 3160–3166, arXiv:hep-ph/9305266. doi:10.1103/PhysRevD.48.3160.

[180] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 270–279. 1998. arXiv:hep-ph/9907280.

[181] Y. L. Dokshitzer, G. D. Leder, S. Moretti et al., "Better jet clustering algorithms", *JHEP* **08** (1997) 001, arXiv:hep-ph/9707323. doi:10.1088/1126-6708/1997/08/001.

[182] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063.

[183] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, arXiv:1607.03663. doi:10.1088/1748-0221/12/02/P02014.

[184] "Jet energy scale and resolution performances with 13TeV data", CMS-DP-2016-020, Jun, 2016.

[185] CMS Collaboration, "Jet algorithms performance in 13 TeV data", CMS-PAS-JME-16-003, 2017.

[186] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, arXiv:1211.4462. doi:10.1088/1748-0221/8/04/P04013.

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

[188] CMS Collaboration, "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV", *JHEP* **03** (2011) 136, arXiv:1102.3194. doi:10.1007/JHEP03(2011)136.

[189] CMS Collaboration, "Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events", CMS-PAS-BTV-13-001, 2013.

[190] CMS Collaboration, "Studies of Tracker Material", CMS-PAS-TRK-10-003, 2010.

[191] CMS Collaboration, "Studies of jet mass in dijet and W/Z + jet events", *JHEP* **05** (2013) 090, arXiv:1303.4811. doi:10.1007/JHEP05(2013)090.

[192] CMS Collaboration, "Identification techniques for highly boosted W bosons that decay into hadrons", *JHEP* **12** (2014) 017, arXiv:1410.4227. doi:10.1007/JHEP12(2014)017.

[193] CMS Collaboration, "Top Tagging with New Approaches", CMS-PAS-JME-15-002, 2016.

[194] 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, arXiv:0802.2470. doi:10.1103/PhysRevLett.100.242001.

[195] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming", *JHEP* **02** (2010) 084, arXiv:0912.1342. doi:10.1007/JHEP02(2010)084.

[196] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Techniques for improved heavy particle searches with jet substructure", *Phys. Rev. D* **80** (2009) 051501, arXiv:0903.5081. doi:10.1103/PhysRevD.80.051501.

[197] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", *Phys. Rev. D* **81** (2010) 094023, arXiv:0912.0033. doi:10.1103/PhysRevD.81.094023.

[198] A. J. Larkoski, S. Marzani, G. Soyez et al., "Soft Drop", *JHEP* **05** (2014) 146, arXiv:1402.2657. doi:10.1007/JHEP05(2014)146.

[199] M. Dasgupta, A. Fregoso, S. Marzani et al., "Towards an understanding of jet substructure", *JHEP* **09** (2013) 029, arXiv:1307.0007. doi:10.1007/JHEP09(2013)029.

[200] M. Dasgupta, A. Fregoso, S. Marzani et al., "Jet substructure with analytical methods", *Eur. Phys. J. C* **73** (2013), no. 11, 2623, arXiv:1307.0013. doi:10.1140/epjc/s10052-013-2623-3.

[201] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* **02** (2012) 093, arXiv:1108.2701. doi:10.1007/JHEP02(2012)093.

[202] CMS Collaboration, "Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state", *submitted to: Phys. Lett. B* (2016) arXiv:1612.00999.

[203] CMS DQM Collaboration, "CMS Run Registry: Data certification bookkeeping and publication system", *J. Phys. Conf. Ser.* **331** (2011) 042038. doi:10.1088/1742-6596/331/4/042038.

[204] R. Gavin, Y. Li, F. Petriello et al., "FEWZ 3.1 guide", Manual FEWZManual_3.1.pdf, 2012.

[205] N. Kidonakis, "Top Quark Production", in *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013*, pp. 139–168. 2014. arXiv:1311.0283.

[206] M. Aliev, H. Lacker, U. Langenfeld et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034–1046, arXiv:1007.1327. doi:10.1016/j.cpc.2010.12.040.

[207] P. Kant, O. M. Kind, T. Kintscher et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", *Comput. Phys. Commun.* **191** (2015) 74–89, arXiv:1406.4403. doi:10.1016/j.cpc.2015.02.001.

[208] T. Gehrmann, M. Grazzini, S. Kallweit et al., "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD", *Phys. Rev. Lett.* **113** (2014), no. 21, 212001, arXiv:1408.5243. doi:10.1103/PhysRevLett.113.212001.

[209] J. M. Campbell and R. K. Ellis, "An Update on vector boson pair production at hadron colliders", *Phys. Rev. D* **60** (1999) 113006, arXiv:hep-ph/9905386. doi:10.1103/PhysRevD.60.113006.

[210] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, arXiv:1105.0020. doi:10.1007/JHEP07(2011)018.

[211] J. M. Campbell, R. K. Ellis, and W. T. Giele, "A Multi-Threaded Version of MCFM", *Eur. Phys. J. C* **75** (2015), no. 6, 246, arXiv:1503.06182. doi:10.1140/epjc/s10052-015-3461-2.

[212] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016), no. 3, 155, arXiv:1512.00815. doi:10.1140/epjc/s10052-016-3988-x.

[213] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C* **74** (2014), no. 8, 3024, arXiv:1404.5630. doi:10.1140/epjc/s10052-014-3024-y.

[214] CMS Collaboration, "Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. Lett.* **111** (2013), no. 21, 211804, arXiv:1309.2030. doi:10.1103/PhysRevLett.111.211804, 10.1103/PhysRevLett.112.119903.

[215] CMS Collaboration, "Search for $\mathrm{t\bar{t}}$ resonances in highly-boosted lepton+jets and fully hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:1704.03366.

[216] CMS Collaboration, "Measurements of differential production cross sections for a Z boson in association with jets in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **04** (2017) 022, arXiv:1611.03844. doi:10.1007/JHEP04(2017)022.

[217] CMS Collaboration, "Measurement of the $\mathrm{W^+W^-}$ cross section in pp collisions at $\sqrt{s} = 8$ TeV and limits on anomalous gauge couplings", *Eur. Phys. J. C* **76** (2016), no. 7, 401, arXiv:1507.03268. doi:10.1140/epjc/s10052-016-4219-1.

[218] CMS Collaboration, "Measurement of the ZZ production cross section and $\mathrm{Z} \to \ell^+\ell^-\ell'^+\ell'^-$ branching fraction in pp collisions at $\sqrt{s}$=13 TeV", *Phys. Lett. B* **763** (2016) 280, arXiv:1607.08834. doi:10.1016/j.physletb.2016.10.054.

[219] CMS Collaboration, "Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* (2016) arXiv:1607.06943. (submitted).

[220] CMS Collaboration, "Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV", *submitted to: Phys. Lett. B* (2016) arXiv:1610.00678.

[221] CMS Collaboration, "Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s}$ =8 TeV", *Phys. Rev. Lett.* **112** (2014), no. 23, 231802, arXiv:1401.2942. doi:10.1103/PhysRevLett.112.231802.

[222] CMS Collaboration, "Search for s channel single top quark production in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **09** (2016) 027, arXiv:1603.02555. doi:10.1007/JHEP09(2016)027.

[223] CMS Collaboration, "Search for heavy resonances decaying to a pair of Higgs bosons in four b quark final state in proton-proton collisions at sqrt(s)=13 TeV", CMS-PAS-B2G-16-008, 2016.

[224] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in *pp* colisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **73** (2013), no. 3, 2339, arXiv:1211.2220. doi:10.1140/epjc/s10052-013-2339-4.

[225] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **75** (2015), no. 11, 542, arXiv:1505.04480. doi:10.1140/epjc/s10052-015-3709-x.

[226] CMS Collaboration, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV", *submitted to: Phys. Rev. D* (2016) arXiv:1610.04191.

[227] CMS Collaboration, "Measurement of the differential cross section for $\mathrm{t\bar{t}}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV", CMS-PAS-TOP-16-011, 2016.

[228] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* (1993) 219. doi:10.1016/0010-4655(93)90005-W.

[229] 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) 115, arXiv:1103.0354.

[230] CMS Collaboration, "Search for pair production of vector-like T quarks in the lepton plus jets final state", CMS-PAS-B2G-16-002, 2016.

[231] ATLAS Collaboration, "Search for pair production of vector-like top partners in events with exactly one lepton and large missing transverse momentum in $\sqrt{s} = 13$ TeV *pp* collisions with the ATLAS detector", ATLAS-CONF-2016-101, Sep, 2016.

[232] ATLAS Collaboration, "Search for new phenomena in $t\bar{t}$ final states with additional heavy-flavour jets in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", ATLAS-CONF-2016-104, Sep, 2016.

[233] ATLAS Collaboration, "Search for pair production of heavy vector-like quarks decaying to high-p_T W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector", ATLAS-CONF-2016-102, Sep, 2016.

[234] ATLAS Collaboration, "Search for pair production of vector-like top quarks in events with one lepton and an invisibly decaying Z boson in $\sqrt{s} = 13$ TeV pp collisions at the ATLAS detector", ATLAS-CONF-2017-015, Mar, 2017.

[235] CMS Collaboration, "Search for heavy vector-like quarks decaying to same-sign dileptons", CMS-PAS-B2G-16-019, 2017.

[236] ATLAS Collaboration, "Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", ATLAS-CONF-2016-072, Aug, 2016.

[237] CMS Collaboration, "Search for single production of vector-like quarks decaying into a b quark and a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:1701.08328.

[238] CMS Collaboration, "Search for single production of vector-like quarks decaying to a Z boson and a top or a bottom quark in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:1701.07409.

[239] CMS Collaboration, "Search for electroweak production of a vector-like quark decaying to a top quark and a Higgs boson using boosted topologies in fully hadronic final states", arXiv:1612.05336.

[240] CMS Collaboration, "Search for single production of vector-like quarks decaying to a Z boson and a top or a bottom quark in proton-proton collisions at 13 TeV", CMS-PAS-B2G-17-007, 2017.

[241] G. Cowan, K. Cranmer, E. Gross et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)]. doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z.

[242] T. Müller, J. Ott, and J. Wagner-Kuhr, "theta – a framework for template-based modeling and inference". http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto/index.html, 2012.

[243] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys. C* **40** (2016), no. 10, 100001. doi:10.1088/1674-1137/40/10/100001.

[244] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys. G* **28** (2002) 2693–2704. [,11(2002)]. doi:10.1088/0954-3899/28/10/313.

[245] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435–443, arXiv:hep-ex/9902006. doi:10.1016/S0168-9002(99)00498-2.

[246] S. Baker and R. D. Cousins, "Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms", *Nucl. Instrum. Meth.* **221** (1984) 437–442. doi:10.1016/0167-5087(84)90016-4.

[247] I. Chakravarti, R. Laha, and J. Roy, "Handbook of methods of applied statistics". Number Bd. 1 in Wiley series in probability and mathematical statistics. Wiley, 1967.

Acknowledgements

I would like to thank all the people who supported me in many different ways during my PhD studies. Without their help, it would not have been possible for me to accomplish this work.

Firstly, I would like to thank my supervisor, Johannes Haller, and Andreas Meyer who kindly reviewed my thesis. I also want to express my gratitude to the other members of my disputation committee, Caren Hagner, Gudrid Moortgat-Pick and Alexander Schmidt. In spite of their many other obligations, they found some time in their calendar and allowed me to have my disputation before the summer break.

I would like to thank Jochen Ott, who had supervised me for the first two years of my PhD and taught me a lot about particle physics, and Alexander Schmidt, who took over after Jochen left our working group. It was their invaluable support and effort that made it possible for me to complete my thesis and even publish a paper about my work. I also want to thank Thomas Peiffer for reading large parts of my thesis, supporting me in the preparation of my disputation and helping me out countless times whenever I had any questions.

Completing this enormous amount of work would have been a lot more difficult and much less pleasant without my fellow PhD and master students. They helped me out so many times when I had problems and created the relaxed and cheerful atmosphere that made me enjoy every day in the office. In this sense, I want to thank Heiner Tholen, Mareike Meyer, Simon Kurz, and Emanuele Usai, who have accompanied me from the very beginning of my PhD, Anna Benecke and Dennis Schwarz, who joined our group more recently, and all present and former members of our group, especially Teresa Lenz for helping me towards the end of my thesis.

Finally, I would like to say thank you to all my friends outside of physics, to my family, who have supported me through all my life, and to Eileen, for everything. Thank you!