Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Search for heavy resonances decaying to a top quark and a W boson with the CMS experiment

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

Vorgelegt von

Alexander Fröhlich

Hamburg

2021

Gutachter der Dissertation:	Prof. Dr. Johannes Haller Prof. Dr. Christian Schwanenberger
Zusammensetzung der Prüfungskommission:	Prof. Dr. Johannes Haller Prof. Dr. Christian Schwanenberger Prof. Dr. Jochen Liske Prof. Dr. Elisabetta Gallo Dr. Kai Schmidt-Hoberg
Vorsitzender der Prüfungskommission:	Prof. Dr. Jochen Liske
Datum der Disputation:	03.02.2022
Vorsitzender des Fach-Promotionsausschusses Physik:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs Physik:	Prof. Dr. Günter Sigl
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Abstract

A search for a heavy resonance decaying into a top quark and a W boson using proton-proton collision data recorded by the CMS experiment at a center-of-mass energy of 13 TeV in the years 2016, 2017 and 2018 is presented. The data analyzed correspond to a total integrated luminosity of $138\,\mathrm{fb}^{-1}$. The search is performed in the final state with an isolated electron or muon and missing transverse momentum from the W boson decay, and a single high energetic jet reconstructing the top quark decay. The jet capturing the boosted top quark decay products is identified using a top tagging technique based on jet clustering with an adaptive distance parameter and inherent jet grooming. This technique is used for the first time in a search for new physics with LHC data. The single production of an excited bottom quark b^* is used as a benchmark for the analysis. A combination with an analysis of the all-hadronic final state is performed. The models with left-handed, right-handed and vector-like couplings of the b^* quark are excluded at 95% confidence level up to b^* quark masses of 3.0, 3.0 and 3.2 TeV, respectively. The results are also interpreted in the context of models with a vector-like B quark. For these models, upper limits on the production cross section of the vector-like B quark are set between 1.6 and 0.007 pb.

Zusammenfassung

Eine Suche nach einer schweren Resonanz, die in ein Top-Quark und ein W-Boson zerfällt, wird präsentiert. Der analysierte Datensatz wurde vom CMS experiment in Proton-Proton Kollisionen bei einer Schwerpunktsenergie von 13 TeV in den Jahren 2016, 2017 und 2018 aufgezeichnet. Die Daten entsprechend einer integrierten Luminosität von $138\,\mathrm{fb}^{-1}$. Die Suche wird im Endzustand mit einem isolierten Elektron oder Muon und fehlendem Transversalimpuls vom W-Boson-Zerfall, und einem hochenergetischen Jet vom Top-Quark-Zerfall durchgeführt. Der Jet aus dem geboosteten Top-Quark-Zerfall wird mit Hilfe einer Top-Quark-Identifikationstechnik identifiziert, die auf einer Jet-Clusterbildung mit adaptiven Distanz-Parameter und inhärentem Jet-Grooming basiert. Diese Technik wird hier zum ersten Mal in einer Suche nach neuer Physik mit LHC Daten angewendet. Die Einzelproduktion von einem angeregten Bottom-Quark b^* wird verwendet, um die Sensitivität der Analyse zu quantifizieren. Eine statistische Kombination mit einer Analyse des vollhadronischen Endzustandes wird durchgeführt. Die Modelle mit linkshändiger, rechtshändiger und vektorartiger Kopplung des b^*-Quarks können bei einem Konfidenzintervall von 95% bis zu b^*-Quark-Massen von 3.0, 3.0 bzw. 3.2 TeV ausgeschlossen werden. Die Ergebnisse der Analyse werden darüber hinaus im Kontext von Modellen mit vektorarigen B Quarks interpretiert. Für diese Modelle werden Ausschlussgrenzen auf den Produktionswirkungsquerschnitt eines vektorarigen B Quarks zwischen 1.6 und 0.007 pb ermittelt.

List of own contributions

Search for a heavy resonance decaying into a top quark and a W boson in the lepton+jets final state

I was the principle analyzer of the search and performed all analysis steps in the ℓ+jets channel:

- the development and optimization of the analysis strategy,
- the efficiency studies of the HOTVR algorithm,
- the development of the data driven background estimation,
- the statistical interpretation of the results.

The search was published in Ref. [1] and has been submitted to JHEP. As part of the publication, I performed a statistical combination of this analysis with an analysis of the all-hadronic final state [2] in collaboration with the main analyzer of the all-hadronic analysis. As contact person for the analysis during the CMS-internal review, my responsibilities included:

- the presentation of regular updates in working group meetings,
- writing and editing of the analysis documentation,
- the preapproval and approval presentations,
- managing the communication with the review committee during the internal

review,

- writing and editing of the paper draft.

The analysis of the lepton+jets final state, as well as the statistical combination are presented in Chapter 9.

I supervised a bachelor's thesis [3], related to my analysis. The results of the studies of the bachelor's thesis are discussed in the context of Section 9.8.

CMS Pixel Online Monitoring

In April 2018 I joined the newly formed CMS Tracker Online Monitoring group, where I contributed to the development of the online monitoring tools for the data taking in 2018 and for the upcoming LHC Run 3. My tasks included:

- the design and development of a database used to store information related to the monitoring of the pixel detector,

- the integration of the pixel monitoring database with the tracker monitoring tools,

- the development of stored SQL queries (views) to quickly access various information from tracker monitoring and detector control system (DCS) databases.

In 2020 I became database contact for the Tracker Online Monitoring group, where I administrated the tracker monitoring database and accounts, and reviewed updates and additions to the database. An overview of the pixel online monitoring tools is given in Chapter 5.

During the 2018 data taking run of the LHC, I spent nine months on a research stay at CERN, where I worked with the pixel operation team. In this time, I took part in the operation of the pixel detector and in measurements on the detector.

Contents

1. Introduction

The Standard Model of particle physics describes the elementary building blocks of our universe and their fundamental interactions. It has withstood extensive experimental tests and validations over many orders of magnitude in energy. Despite its remarkable success, there are a number of known shortcomings to the theory, indicating the need for extensions to the standard model. For example, the standard model does not describe the gravitational interaction, neither the abundance of matter over antimatter in the universe, nor the small observed Higgs boson mass compared to expectations from quantum loop corrections. It is also missing a suitable candidate particle to describe dark matter.

The many proposed extensions to the standard model include models of compositeness or extra dimensions, which predict the existence of excited quark states with masses at the order of TeV. The large mass of the third generation quarks suggest a particularly strong coupling of excited quarks to the top and bottom quark. Evidence for these models could manifest as a heavy resonance decaying into a top quark and a W boson.

In this thesis, a search for such a heavy resonance is presented. Data recorded by the CMS experiment in pp collisions at a center-of-mass energy of 13 TeV in the years 2016, 2017 and 2018, corresponding to an integrated luminosity of $138\,\mathrm{fb}^{-1}$ are analyzed. The lepton+jets final state is used, where the top quark decays hadronically and the W boson decays into a muon or electron and the corresponding neutrino. The results have been published in Ref. [1]. The top quark decay is reconstructed using a single jet with adaptive radius and is identified using jet substructure techniques. For this purpose, the Heavy Object Tagger with Variable R is employed, which achieves a stable selection efficiency over a large range of top quark momenta. A model with an excited bottom quark b^* is used to quantify the sensitivity of the analysis. Three different scenarios for the coupling of the b^* quark are considered: purely left- or

right-handed, and vector-like.

Searches for such a b^* quark decaying to a top quark and a W boson were already performed at the LHC by the ATLAS and CMS Collaborations at center-of-mass energies of 7 [4] and 8 TeV [5, 6] excluding the b^* models up to b^* quark masses of 1.4 TeV, 1.4 TeV and 1.5 TeV for the left-handed, right-handed and vector-like models respectively. At a center-of-mass energy of 13 TeV, a search was performed by the CMS Collaboration in the all-hadronic final state [2]. That search was able set limits on the b^* production cross section for b^* quark masses between 1.4–4.0 TeV and exclude the left-handed, right-handed and vector-like b^* models up to masses of 2.6 TeV, 2.8 TeV and 3.1 TeV, respectively. A statistical combination of the presented analysis and the analysis of the all-hadronic final state is performed. With the lower thresholds of the triggers used in the presented lepton+jets analysis and the stable performance of the HOTVR [7] algorithm over a wide range of top quark momenta, the range of probed b^* quark masses can be extended down to 700 GeV, while maintaining good sensitivity at high b^* quark masses.

This thesis is structured as follows: A description of the Standard Model and possible extensions are given in Chapter 2. There, the studied signal models are also introduced. In Chapter 3, jets are introduced and a description of jet clustering and jet substructure algorithms is given. An overview of the LHC and the CMS detector is given in Chapter 4. Chapter 5 gives an overview of the online monitoring of the CMS pixel detector. In Chapter 6, the event reconstruction in CMS is outlined. An efficiency study of the HOTVR algorithm used in this work is presented in Chapter 8. The search for a heavy resonance decaying to a top quark and a W boson is detailed in Chapter 9. Finally, the thesis is concluded in Chapter 10.

2. Theory

The description of elementary particles and their interactions via three of the four known fundamental forces is provided by the Standard Model of elementary particle physics (SM). The SM is an incredibly successful theory, predicting new particles that were later discovered experimentally and withstanding many precision tests from various experiments and over a large range of energy scales. In the following, a brief description of the SM is given, followed by a discussion of the physics of inelastic proton-proton (pp) scattering in particle colliders. These sections are based on information from Refs. [8–10], unless stated otherwise. Then, the simulation of high-energy proton-proton interactions with Monte Carlo event generators is discussed. Finally, some shortcomings of the SM and possible extensions to the SM are discussed, and two models of heavy resonances decaying into a top quark and a W boson are introduced.

2.1. The Standard Model of Elementary Particle Physics

The SM aims at describing the most fundamental building blocks of matter, the elementary particles, and their interactions. The elementary particles are categorized into fermions with half-integer spin and bosons with integer spin. The SM describes the interactions between these particles via three fundamental forces in terms of quantum field theories. These forces are the electromagnetic force and the strong and weak nuclear forces. They are mediated by the exchange of bosons. The electromagnetic force is mediated by the photon (γ), the strong force is mediated by the gluons (g), and the weak force is mediated by the $W^{\pm}$ and Z bosons. The fourth known fundamental force of nature, gravity, is not described in the SM, because it

has not yet been successfully formulated in terms of a quantum field theory and its impact on elementary particles is about 10^{36} times smaller than e.g. the EM force.

All fermions with negative chirality (left-handed) have a weak isospin $T = 1/2$ and are grouped into left-handed isospin doublets with $T_3 = \pm 1/2$, where T_3 denotes the third component of the weak isospin. Fermions with positive chirality (right-handed) form isospin singlets with $T = T_3 = 0$. The fermions can be further grouped into quarks and leptons, depending on their charges and thus on the interactions they exhibit. Quarks come in six different quark flavors called up (u), down (d), charm (c), strange (s), top (t), and bottom (b). The quark isospin doublets are arranged as

$$\begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix}_{\mathrm{L}}, \begin{pmatrix} \mathrm{c} \\ \mathrm{s} \end{pmatrix}_{\mathrm{L}}, \begin{pmatrix} \mathrm{t} \\ \mathrm{b} \end{pmatrix}_{\mathrm{L}}. \tag{2.1}$$

The quarks in the upper row are called up-type quarks and have electric charge of $Q = \frac{2}{3}e$, where e denotes the elementary electric charge, and weak isospin with $T_3 = +1/2$. The quarks in the lower row are analogously called down-type quarks. They have electric charge of $Q = -\frac{1}{3}e$ and weak isospin with $T_3 = -1/2$. The corresponding right-handed quarks form isospin singlets with $T_3 = 0$, as described above. All quarks carry one of three color charges, making them the only fermions in the SM to interact via the strong force. Leptons on the other hand do not carry color charge. The lepton isospin doublets are arranged as

$$\begin{pmatrix} \nu_{\mathrm{e}} \\ \mathrm{e} \end{pmatrix}_{\mathrm{L}}, \begin{pmatrix} \nu_{\mu} \\ \mu \end{pmatrix}_{\mathrm{L}}, \begin{pmatrix} \nu_{\tau} \\ \tau \end{pmatrix}_{\mathrm{L}}. \tag{2.2}$$

In the lower row are the charged leptons, the electron (e), the muon (μ), and the tauon (τ), carrying an electric charge of $Q = 1e$ and weak isospin with $T_3 = -1/2$. In the top row are the corresponding lepton neutrinos. They are electrically neutral and have a weak isospin with $T_3 = +1/2$. The neutrinos can hence only interact via the weak force. Right-handed charged leptons form isospin singlets as described above. There are no right-handed neutrinos in the SM, because neutrinos are assumed to be massless. For each particle comes a corresponding antiparticle with the same mass, but inverted signs on all other quantum numbers. Right-handed antifermions form isospin doublets and left-handed antifermions form isospin singlets.

The interactions in the SM are formulated as quantum field theories. Each interaction postulates invariance of the SM Lagrangian under transformations of an underlying

local gauge group.

The strong force is formulated in the theory of quantum chromodynamics (QCD). It describes the interaction of color charged particles. QCD postulates an underlying symmetry of the Lagrangian under transformations of the $SU(3)_C$ gauge group. The C denotes color charge, which is conserved in the strong interaction. Quarks are assigned to the fundamental representation of the $SU(3)_C$, introducing a new quantum number of color charge with three degrees of freedom, called red, green and blue. Analogously, antiquarks are assigned to the antifundamental representation. Since only color neutral states are invariant under transformations of the $SU(3)_C$ gauge group, quarks cannot be observed freely. They form bound states called hadrons. The simplest possible combinations are quark-antiquark pairs called mesons and (anti)quark triplets called (anti)baryons. The generators of the $SU(3)_C$ gauge group are eight non-commuting 3×3 Hermitian matrices, called Gell-Mann matrices. Directly related to the number of generators is the number of gauge bosons. In the case of QCD these are the eight massless gluons. The gluons themselves carry a combination of color and anticolor charge and can thus interact with each other. The gluon self-interaction is reflected in the non-Abelian nature of the underlying gauge group and leads to color confinement and asymptotic freedom. Color confinement describes the property of QCD that the coupling strength α_s increases with decreasing energy scales q^2, i.e. with larger distances or smaller momentum transfer between the interacting particles. At sufficiently large distance, it becomes energetically more favorable to form a new quark-antiquark pair from the energy in the gluon field potential. Another consequence of this is that perturbation theory collapses at small q^2 and processes like the formation of hadrons cannot be calculated analytically. Conversely, α_s becomes small at large q^2, allowing the treatment of color charged particles as quasi-free particles. This property is called asymptotic freedom.

Historically, electromagnetism was the first interaction to be formulated as a quantum field theory in the theory of quantum electrodynamics (QED). There, invariance under local phase rotations, represented by the $U(1)_Q$ group is postulated. The Q denotes the electric charge, which is conserved in the electromagnetic interaction. Only one gauge boson is required in QED, which is the massless photon. Since the photon itself is not electrically charged, there is no self-interaction of the photon in QED. Consequently, the coupling constant α decreases with smaller q^2. This ensures that higher order processes are suppressed and allows the calculation of QED processes in perturbation theory.

Similar to QCD and QED, the theory of the weak interaction is described by a gauge theory. The underlying symmetry is of the non-Abelian $\mathrm{SU(2)_L}$ group. The conserved charge of the weak interaction is the weak isospin T. Hence, the weak interaction only affects fermions with left-handed chirality (and antifermions with right-handed chirality), making it maximally parity violating. This is indicated by the subscript L. Because of this property of the weak interaction, left-handed fermions have to be arranged in isospin doublets, while right-handed fermions form isospin singlets, as described above. The generators of the $\mathrm{SU(2)_L}$ group are three non-commuting 2×2 Hermitian matrices. The corresponding three gauge bosons are $\mathrm{W}_{1,2,3}$. The $\mathrm{W}_{1,2}$ mediate the charged current weak interaction, which describes transitions of fermions within an isospin doublet. The corresponding physical states $\mathrm{W}^{\pm}$ are obtained as linear combinations of the $\mathrm{W}_{1,2}$,

$$\mathrm{W}^{\pm} = \frac{1}{\sqrt{2}} \left(\mathrm{W}_1 \mp i \mathrm{W}_2 \right) . \tag{2.3}$$

It was found experimentally that the charged weak interaction allowed transitions between quarks from different generations. This phenomenon was explained by Cabibbo, Kobayashi and Maskawa by proposing the quark states in the weak interaction are obtained as a mixture of the quark mass eigenstates [11, 12]. The mixing of the quark mass eigenstates is described by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix} , \tag{2.4}$$

where the primed quarks denote the states in the weak interaction. The probability for a transition between an up-type quark i and a down-type quark j is directly given by $\left|V_{ij}\right|^2$. The magnitudes of V_{ij} are determined experimentally as [13]

$$\begin{aligned} V_{\mathrm{CKM}} &= \begin{pmatrix} \left|V_{\mathrm{ud}}\right| & \left|V_{\mathrm{us}}\right| & \left|V_{\mathrm{ub}}\right| \\ \left|V_{\mathrm{cd}}\right| & \left|V_{\mathrm{cs}}\right| & \left|V_{\mathrm{cb}}\right| \\ \left|V_{\mathrm{td}}\right| & \left|V_{\mathrm{ts}}\right| & \left|V_{\mathrm{tb}}\right| \end{pmatrix} \\ &= \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix} . \end{aligned} \tag{2.5}$$

The third boson of the weak interaction, W_3, represents a neutral current weak interaction. Like $\mathrm{W}_{1,2}$ bosons, the W_3 boson couples only to the isospin doublets. It

was however found experimentally, that the observed Z boson also couples to isospin singlets, indicating another interaction of the Z boson. Furthermore, the bosons of the weak interaction have to be massless, to ensure invariance of the Lagrangian. However, the $W^{\pm}$ and Z bosons were experimentally found to be massive.

The issue of massive gauge bosons can be solved with a mechanism called spontaneous symmetry breaking. This mechanism, proposed by Brout, Englert and Higgs in 1964 [14–16], allows introducing mass terms for gauge bosons into the Lagrangian by spontaneously breaking the local gauge symmetry with an additional scalar field ϕ, also called the Higgs field, that has a non-zero vacuum expectation value v.

In the late 1960s it was realized by Glashow, Salam and Weinberg [17–19] that the electromagnetic and the weak interaction could both be described within a unified theory of electroweak interaction. This electroweak theory postulates a symmetry of the Lagrangian under transformations of the $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ group. The subscript Y indicates a different conserved charge than in QED. This new charge is called the hypercharge $Y = Q - T_3$, where Q is the electric charge and T_3 is the third component of the weak isospin. Four gauge bosons arise from the required gauge invariance, two charged bosons $\mathrm{W_{1,2}}$, and two neutral bosons $\mathrm{W_3}$ and B. A complex scalar field ϕ of the form

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.6}$$

is introduced. The field forms an $\mathrm{SU(2)_L}$ doublet with a weak hypercharge $Y = +1/2$. The potential $V(\phi)$ is chosen of the form

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \frac{\lambda}{2} \left(\phi^\dagger \phi\right)^2 , \tag{2.7}$$

with $\mu^2 > 0$ and $\lambda > 0$. If ϕ obtains a vacuum expectation value of the form

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} , \tag{2.8}$$

where $v = \sqrt{\mu^2/\lambda}$, the electroweak symmetry is spontaneously broken. Expanding the Lagrangian around the vacuum state of ϕ, new terms are introduced that give mass to three of the four electroweak gauge bosons, while leaving the last one massless. Additionally, a new scalar boson, called the Higgs boson H, with mass $m_H = \sqrt{2}\mu$ arises from excitations of the vacuum state.

The physical states of the SM bosons are identified as linear combinations of the electroweak bosons. The $W^{\pm}$ states are identified analogously to Eq. (2.3). The Z and γ are identified as

$$\begin{pmatrix} \mathrm{Z} \\ \gamma \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} \mathrm{W}_3 \\ \mathrm{B} \end{pmatrix} , \tag{2.9}$$

where θ_w is the weak mixing angle, also called Weinberg angle, with

$$\cos\theta_w = \frac{g}{\sqrt{g^2 + g'^2}} , \quad \sin\theta_w = \frac{g'}{\sqrt{g^2 + g'^2}} . \tag{2.10}$$

Here, g and g' are the $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$ coupling constants. The boson masses can be expressed as

$$m_{\mathrm{W}} = g\frac{v}{2} , \quad m_{\mathrm{Z}} = \sqrt{g^2 + g'^2}\frac{v}{2} . \tag{2.11}$$

Using Eq. (2.10) and Eq. (2.11), it can be seen that the two boson masses are not independent, but connected via the weak mixing angle,

$$m_{\mathrm{W}} = m_{\mathrm{Z}} \cos\theta_w . \tag{2.12}$$

The electroweak theory is thus able to describe the observed massive gauge bosons of the weak interaction. It also allows the Z boson to couple to right-handed fermions via their hypercharge, solving the above described discrepancy of the weak theory with experimental observations. However, with the electroweak Lagrangian it is now forbidden by global gauge invariance to include fermion mass terms directly, because left- and right-handed components of the fermion fields carry different hypercharges. Fortunately, it is possible to incorporate mass terms for the fermions via a coupling to the Higgs field, called Yukawa coupling. The fermion masses are then given by

$$m_f = \frac{1}{\sqrt{2}} y_f v , \tag{2.13}$$

where y_f is the Yukawa coupling strength of the Higgs field to a fermion f. This mechanism also introduces terms describing the Higgs-fermion couplings with a coupling strength proportional to the fermion mass.

The theory of the electroweak interaction together with the QCD are embedded in the SM Lagrangian with an underlying symmetry of the $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$

group. With the discovery of the Higgs boson by the ATLAS and CMS collaborations in 2012 [20, 21], the last missing particle of the SM has been discovered and an era of precision tests of the SM has begun. Despite its incredible success, the SM is not a complete theory and leaves some important phenomena unresolved. These shortcomings and some possible extensions to the SM are discussed in Section 2.4. The top quark plays an important role in searches for new physics beyond the standard model, so in the following the top quark will be introduced in more detail and the properties of its decay will be discussed.

2.1.1. The Top Quark

The top quark is, with a mass of $m_{\mathrm{t}} = (173.34 \pm 0.76)\,\mathrm{GeV}$ [22], the heaviest known elementary particle. It was first discovered in 1995 by the CDF and D0 collaborations at the Tevatron collider [23, 24]. The large mass of the t quark sets it apart from the other quarks in the SM, because it results in an extremely short lifetime of $\mathcal{O}\left(10^{-25}\,\mathrm{s}\right)$, which is about an order of magnitude smaller than the typical time scale of the hadronization. This means, the t quark decays before it can form hadronic states, allowing it to be treated as a free quark. The large value of m_{t} also translates to the strongest coupling to the Higgs field of all SM particles. Therefore, the t gives the dominant contributions to the corrections of Higgs mass and plays an important role in the electroweak symmetry breaking.

The most important production processes of t quarks at proton-proton colliders is the top-antitop pair production ($\mathrm{t\bar{t}}$) via the strong interaction and the single top quark production[1] via the electroweak interaction. Tree level Feynman diagrams for the $\mathrm{t\bar{t}}$ and single t production are shown in Fig. 2.1 and Fig. 2.2, respectively.

The top quark decays via the weak interaction. According to the CKM matrix (Eq. (2.5)), the dominant decay process is $\mathrm{t} \to \mathrm{bW}$ with a branching fraction close to unity, while the decays into dW and sW are suppressed. Depending on the subsequent decay of the W boson, the final state of a top quark decay is either called hadronic ($\mathrm{t} \to \mathrm{bW} \to \mathrm{bq\bar{q}'}$) or leptonic ($\mathrm{t} \to \mathrm{bW} \to \mathrm{b}\ell\nu$). In the case of $\mathrm{t\bar{t}}$ production, this leads to three possible decay modes: the all-hadronic deacy, where both top quarks decay hadronically, the ℓ+jets decay mode, where one top quark

[1] The analogous production of single anti top quarks is implied, if not stated otherwise.

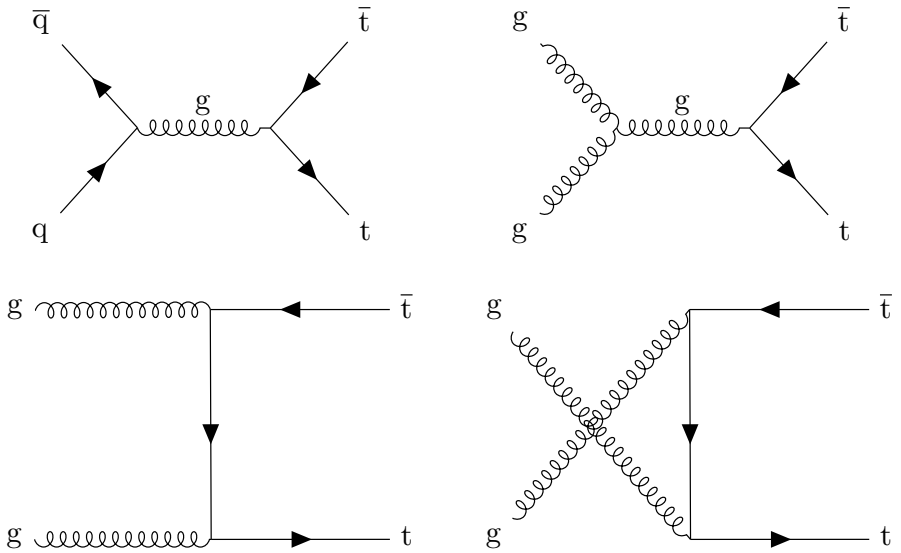

Figure 2.1.: Tree level Feynman diagrams of top quark anti-quark pair production. The $t\bar{t}$ production via quark-antiquark annihilation is shown in the top left. The top right shows the $t\bar{t}$ production via gluon-gluon fusion in the s-channel. The bottom row shows the $t\bar{t}$ production via gluon-gluon fusion in the t-channel (bottom left) and u-channel (bottom right).

decays hadronically and one top quark decays leptonically, and the dilepton decay mode, where both top quarks decay leptonically.

2.2. The Physics of Proton-Proton Collisions

Inelastic scattering experiments play an important role in the research of elementary particles. Scattering experiments at particle colliders allow to produce and study elementary particles and measure the interactions predicted by the SM in a controlled environment. In order to produce massive particles in a collider experiment, the center-of-mass energy $\sqrt{s}$ must be greater than the sum of the masses of the particles produced. Here, s describes the total energy and momentum of the colliding initial-state particles

$$s = (P_1 + P_2)^2 \,, \tag{2.14}$$

where P_i are the four-momenta of the initial-state particles. In this thesis, data from proton-proton (pp) collisions are analyzed. Since protons are not elementary, the structure of the proton needs to be considered in high-energy pp collisions. This

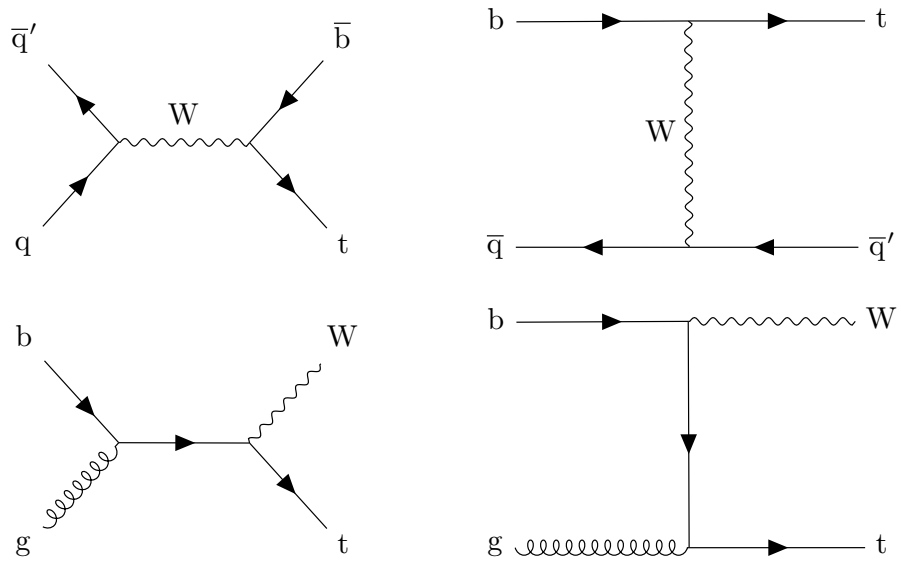

Figure 2.2.: Tree level Feynman diagrams of single t production. The top row shows the *s*- (top left) and *t*-channel (top right) of single t production. The bottom row shows the single t production in association with a W boson in the *s*- (bottom left) and *t*-channel (bottom right). Diagrams for single $\overline{\text{t}}$ production are constructed analogously.

section discusses the physics of deep inelastic pp scattering in the context of the parton model.

The proton is a baryon composed of two up quarks and one down quark (uud). The quarks within the proton are constantly interacting with each other via the strong nuclear force by exchanging gluons. The gluons can also interact with each other, creating more gluons and quark-antiquark pairs from vacuum fluctuations. The three main constituents, called valence quarks, together with the fluctuating cloud of gluons and quark-antiquark pairs (sea quarks) within the proton are summarized as partons. Each parton carries a fraction x of the protons momentum. The probability density function of x for each type of parton is described in the parton distribution function (PDF). The PDFs depend on the energy scale q^2. In Fig. 2.3, PDFs at $q^2 = 10\,\text{GeV}^2$ and $q^2 = 1000\,\text{GeV}^2$ are shown, obtained by the NNPDF collaboration [25] from a global fit of QCD calculations at next-to-next-to-leading order (NNLO) perturbation theory to various experimental results. At high values of x, the valence quarks are dominantly found, while at lower values of x, the gluons and sea quarks dominate.

Because of the asymptotic freedom of QCD, high energy pp collisions can be described

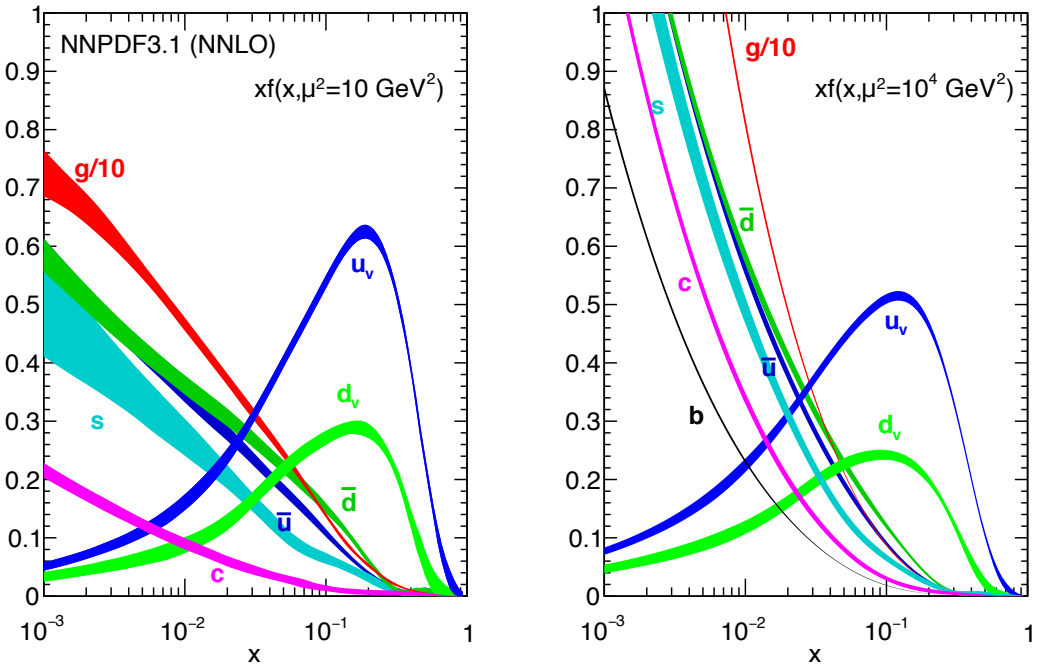

Figure 2.3.: Parton distribution functions f times x at NNLO as a function of x obtained by the NNPDF collaboration for $q^2 = 10\,\mathrm{GeV}^2$ (left) and $q^2 = 1000\,\mathrm{GeV}^2$ (right). Taken from Ref. [25].

as the scattering of quasi-free partons. This allows the factorization of the cross section of a process in pp collisions $\sigma_{\mathrm{pp}\to\mathrm{X}}$ into the cross section of the hard interaction $\sigma_{ij\to\mathrm{X}}$ and the PDFs [26, 27]:

$$\sigma_{\mathrm{pp}\to\mathrm{X}} = \sum_{i,j} \iint f_i(x_1, \mu_f^2) f_j(x_2, \mu_f^2) \sigma_{ij\to\mathrm{X}} \,\mathrm{d}x_1 \,\mathrm{d}x_2 \ , \tag{2.15}$$

where μ_f is the factorization scale. The cross section of the hard interaction $\sigma_{ij\to\mathrm{X}}$ depends on the factorization scale μ_f and the renormalization scale μ_r. A typical choice in calculations of the pp cross section is $\mu_f = \mu_r = q^2$.

In general, the momentum fractions of the interacting partons are not equal, resulting in a boost of the center-of-mass system of the interaction in longitudinal direction (z-axis). Because the momentum of the initial particles is unknown, the momentum of the final state particles cannot be determined. However, since the initial momentum in the transverse (x-y) plane is negligible, the transverse momentum $p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}$ measured in the detector can be assumed to originate from the interaction process,

where p_x and p_y are the momentum components in the x and y direction, respectively. Additionally, the pseduorapidity η, defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{2.16}$$

is introduced to replace the angle from the z-axis θ, because unlike in θ differences in η are conserved under Lorentz-transformations along the longitudinal direction. Using η and the azimuthal angle ϕ, measured in the transverse plane, the Lorentz-invariant angular distance ΔR between two objects becomes

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}\,. \tag{2.17}$$

2.3. Simulation of Proton-Proton Interactions

The simulations of high energy particle physics processes are an important tool in the analysis of collision data. They provide predictions for the distributions of event observables to compare to experimental data. By allowing comparisons of reconstructed objects to the underlying particle level objects (simulation truth) simulations are very useful in the study of new analysis techniques and detector designs. Due to their complexity and large number of particles in the final state, high-energy pp collision events are simulated using Monte Carlo (MC) event generators to provide a prediction for the analyzed data. In the presented analysis, the MadGraph5_aMC@NLO [28], PYTHIA 8.2 [29] and POWHEG v2 [30–33] event generators are used. In the following, a brief description of the event generation process is given. A comprehensive summary of MC event generators can be found e.g. in Ref. [34].

The event generation is performed in multiple separate steps. First, the hard partonic interaction is simulated according to the leading-oder (LO) or next-to-leading-order (NLO) matrix element of the given process folded with the PDFs, and the phase space. In the second step, the particle showers initiated by the final state particles of the hard interaction, as well as radiations in the initial or final state are simulated following the parton shower model. The simulation of the parton shower is performed using PYTHIA. Hard interaction processes simulated with other event generators

are matched to the parton showers using the MLM [35] or FxFx [36] algorithms for matrix element calculations at LO or NLO, respectively. The parton shower simulation proceeds until a cutoff scale of about 1 GeV is reached. Below this cutoff scale, hadronization processes set in. In PYTHIA, the hadronization is simulated following the Lund string model [37]. The potential between two color charged particles is assumed to increase linearly with their distance. When the distance between the particles becomes large enough, a new $q\bar{q}$ pair is formed. With this step, all the stable final state particles of the main interaction are simulated. The event description at this level is called the particle level or generator level.

In addition to the main interaction, other processes contributing to the event description are modeled. These processes include the hadronization of the beam remnants, multiparton interactions, as well as their associated initial and final state radiation processes. They are simulated using a phenomenological model, which is tuned using sets of experimentally determined parameters [38, 39]. Other pp collisions can occur in the same or adjacent bunch crossing, also contributing to the number of particles observed in the detector. These additional events, referred to as pileup events, are simulated using PYTHIA and are superimposed on the main interaction event.

Finally, to obtain an event description matching the experimental data, the interactions of stable particles with a detector are simulated. The detector simulation is performed in GEANT4 [40].

2.4. Physics Beyond the Standard Model

Experimental observations and theoretical motivated problems require explanations left open by the SM. In the following a selection of those open questions, as well as possible extensions to the SM are discussed.

2.4.1. Shortcomings of the Standard Model

The most obvious shortcoming of the SM is the missing description of gravity. Because gravity is so weak compared to the other fundamental forces of nature, it

is sufficient to assume the description of the SM valid up to an energy scale where gravity becomes non-negligible, which is considered to happen at the Planck scale $\Lambda_{\text{Planck}} \sim 10^{19}\,\text{GeV}$. This, however, introduces another problem called the hierarchy problem. The Higgs boson mass receives quantum loop corrections of the order of at least the cut-off scale squared. In order to cancel these large corrections, the bare Higgs mass needs to be finely tuned to a precision of $\sim (m_H/\Lambda)^2$ to obtain the measured Higgs boson mass, which is considered unnatural.

Another problem arises from the obvious asymmetry between matter and antimatter in the universe. Only CP-violating processes can result in an imbalance of matter over antimatter. However, the only CP-violating process in the SM, introduced in the quark mixing via the CKM matrix, is not strong enough to explain the amount of matter observed in the universe.

Furthermore, the SM relies on a considerably large number of free parameters that need to be determined experimentally, including the fermion masses and the quark mixing angles. There also seems to be a similarity between quarks and leptons, suggesting an underlying connection between the two.

Apart from the theoretically motivated problems, also a number of experimental results cannot be explained by the SM. The most prominent of these is the experimental evidence for dark matter in astrophysical and cosmological observations, like measurements of galaxy rotation curves [41], gravitational lensing [42] and the cosmic microwave background [43, 44]. Although evidence for dark matter is observed via its gravitational interaction, it could be explained by weakly interacting massive particles. However, no such dark matter candidate is included in the SM.

Deficits in measurements of the solar neutrino flux suggested an oscillation of neutrino mass eigenstates. Evidence for neutrino oscillations was found experimentally [45], demonstrating that neutrinos are not massless in contrast to the SM assumption.

Furthermore, some precision measurements of SM predictions have recently shown deviations from the SM, indicating possible new mechanisms not described by the SM. Measurements of rare B meson decays to leptons show tensions with the SM prediction at up to 3.1 standard deviations [46]. The combination of different measurements of the anomalous magnetic moment of the muon a_μ results in a deviation from SM calculations of 4.2 standard deviations [47].

2.4.2. Theories Beyond the Standard Model

The above described arguments point towards new physics beyond the standard model. Various different models have been proposed to address these arguments. A selection of prominent types of theories beyond the SM (BSM) are introduced in the following.

Motivated by the success of the electroweak unification, theories aiming at unifying all three SM interactions into a so called grand unified theory (GUT) have been proposed [48]. In these models, the three interactions of the SM are embedded within a higher symmetry group. This new symmetry is broken below a certain energy scale, referred to as the GUT scale, yielding the interactions described by the SM.

The striking symmetry between quarks and leptons in the SM and the large number of free parameters related to the fermions motivate compositeness models [49], where the SM particles are not considered elementary particles, but consist of new, more fundamental particles. As a consequence of compositeness excited states of SM fermions would arise. Models including a composite Higgs boson [50] are also able to address the problem of the small Higgs boson mass, as it would then only receive corrections up to the scale on which the Higgs constituents are confined.

Another approach to address the hierarchy problem is introducing compactified or warped extra dimensions. An example are Randall-Sundrum models [51], where a bulk with one extra dimension is assumed. The SM physics is confined to a four dimensional brane. The mediator particle of gravity (graviton) can move along the warped extra dimension and its probability function has a maximum located at a second brane, separated by the SM brane in the extra dimension. The probability function of the graviton drops exponentially when moving towards the SM brane, explaining why gravity is much weaker than the other forces.

Although the scale of the above described models is generally not in reach of the LHC experiments, the models would give rise to phenomena observable at the TeV scale. In the following, two general models are introduced that describe the interactions of new particles that can manifest as a consequence of the above described theories. These new particles would be observable in pp collisions as a heavy resonance decaying to a t quark and a W boson [52, 53].

2.4.3. Excited Bottom Quarks

Compositeness models predict that SM particles are built from a set of more fundamental elementary particles. A consequence of this is the existence of excited states of SM particles, only possible for compound objects. Third generation quarks are of particular interest in searches for excited quarks, as the large masses of the third generation quarks suggest a strong coupling to excited quarks. Such couplings can be realized in Randall-Sundrum [54, 55] or composite Higgs [56, 57] models.

This thesis focuses on excited bottom quarks (b^*). In pp collisions, the dominant production process is $\mathrm{bg} \rightarrow \mathrm{b}^*$. The most general Lagrangian describing this process [53, 58, 59] is given by

$$\mathcal{L} = \frac{g_s}{2\Lambda} G_{\mu\nu} \bar{\mathrm{b}} \sigma^{\mu\nu} (\kappa_L P_L + \kappa_R P_R) \mathrm{b}^* + \mathrm{h.c.}\,, \tag{2.18}$$

where g_s is the strong coupling constant, Λ is the compositeness scale, $G_{\mu\nu}$ is the gluon field strength tensor and $\sigma^{\mu\nu}$ are the Pauli spin matrices. The free parameters κ_L and κ_R determine the coupling strength to the left-handed and right-handed chirality projections P_L and P_R, respectively. The compositeness scale is chosen to be the b^* mass. Possible decay channels of the b^* quark are bg, bH, bZ and tW. The branching fractions of each decay are shown in Fig. 2.4 as a function of the b^* quark mass. For b^* quark masses $m_{\mathrm{b}^*} > 700\,\mathrm{GeV}$, the decay into tW becomes the dominant process with the branching fraction approaching almost 40%. The leading-oder Feynman diagram for this process is shown in Fig. 2.5.

The weak coupling of the b^* is described by

$$\mathcal{L} = \frac{g}{\sqrt{2}} \mathrm{W}^+{}_\mu \bar{\mathrm{t}} \gamma^\mu (f_L P_L + f_R P_R) \mathrm{b}^* + \mathrm{h.c.}\,, \tag{2.19}$$

where g is the weak coupling constant, γ^μ are the gamma matrices, and f_l and f_r are free parameters determining the left- and right-handed coupling strength. The single production of a b^* quark via the weak interaction is also possible. The final state of this production mode includes an additional forward jet. This produces the same signature as a vector-like B quark, which is discussed in the next section.

Three choices of coupling parameters are considered: purely left-handed (LH) ($f_L = \kappa_L = 1$, $f_R = \kappa_R = 0$), purely right-handed (RH) ($f_L = \kappa_L = 0$, $f_R = \kappa_R = 1$), and

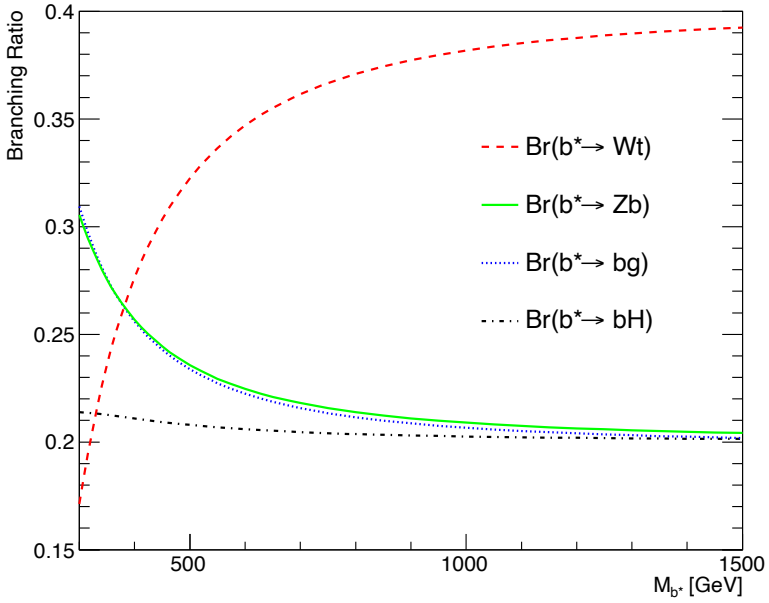

Figure 2.4.: Branching fractions of b^* to tW, bg, bZ and b H, as a function of the b^* mass. Taken from [53]. Labels were adjusted to match the notation of the thesis.

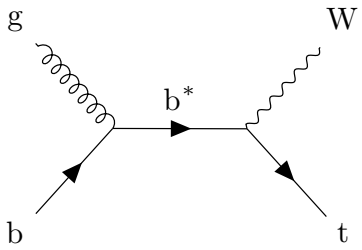

Figure 2.5.: Feynman diagram of the process $bg \to b^* \to tW$.

vector-like (VL) ($f_L = \kappa_L = 1$, $f_R = \kappa_R = 1$).

Previous searches for a b^* decaying to tW were conducted at the LHC by ATLAS at 7 TeV [4] and 8 TeV [5], and by CMS at 8 TeV [6]. No evidence for b^* quarks was found. The CMS analysis set the strongest limits of theses analyses on the LH, RH and VL models at b^* quark masses of up to 1.4 TeV, 1.4 TeV and 1.5 TeV at 95% confidence level (CL), respectively. While the ATLAS search at 7 TeV targeted resolved b^* decays, the two searches at 8 TeV started to implement techniques to identify and reconstruct boosted t and W decays, improving the signal sensitivity particularly towards higher b^* quark masses. The first search for $b^* \to tW$ at 13 TeV was performed by CMS using 137 fb^{-1} of pp collision data [2]. This search

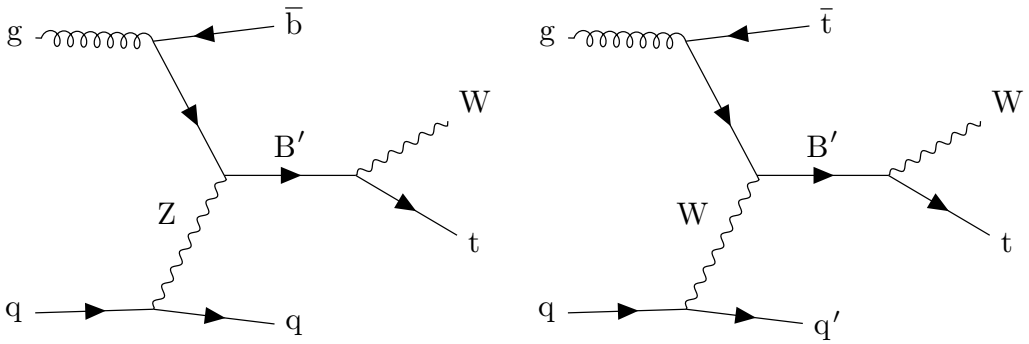

Figure 2.6.: Feynman diagram of the process qg → B → tW + q in association with a b quark (left) and t quark (right).

targets boosted W and t decays in the fully hadronic final state. It is sensitive to b^* quark masses above $m_{b^*} \geq 1.4\,\text{TeV}$ and improves the previous mass exclusion limits up to 2.6 TeV, 2.8 TeV and 3.1 TeV at 95% CL for the LH, RH and VL models, respectively.

2.4.4. Vector-Like B Quarks

While the presence of a fourth generation of chiral quarks has been excluded by the discovery of the Higgs boson [60, 61], quarks where the left- and right-handed chiral components couple to the electroweak interaction are still a possibility, because Dirac mass terms for such quarks would be invariant under the SM gauge group and contributions to the Higgs mass corrections would cancel. These vector-like quarks (VLQ) appear in many extensions of the SM, e.g. composite Higgs models [62–64], extra dimension models [65, 66] or GUTs [67].

The phenomenology of VLQs at pp collider experiments is studied using a model independent framework [68–70]. The VLQs appear in singlets or multiplets of two third-generation partners B and T with electrical charges $-\frac{1}{3}e$ and $\frac{2}{3}e$, and two exotic quarks X and Y with electrical charges $\frac{5}{3}e$ and $-\frac{4}{3}e$, respectively. In this thesis, a B VLQ is considered in a reinterpretation of the search results. Vector-like B quarks are singly produced via the weak interaction, as shown in Fig. 2.6. The production occurs either via the exchange of a Z boson, where an associated b quark is produced, or via the exchange of a W boson, where an assoicated t quark is produced. These associated particles generally have smaller values of p_T than the decay products of

the B quark. The initial state quark is scattered under smaller angles and can be detected in the forward region of the detector. This forward quark can be used to discriminate B events from SM background events.

In a search by the CMS collaboration using $35.9\,\mathrm{fb}^{-1}$ of pp collision data at $\sqrt{s} = 13\,\mathrm{TeV}$ recorded in the year 2016 [71], limits on the production cross section of a B quark in association with a b or t quark are set between 0.6–0.04 pb, and 0.35–0.03 pb, respectively, for B quark masses between 0.7–1.8 TeV. In that search, the lepton+jets final state of the tW decay mode is targeted. Both, boosted and resolved decays of the t quark and W boson are considered and the presence of an additional jet in the forward region of the detector is used to discriminate signal events from SM background events.

3. Jets

The quark confinement and subsequent hadronization of quarks and gluons described in Section 2.1 leads to the formation of collimated cascades of hadrons in high energetic pp collisions, called jets. Measurements of these jets allow to draw conclusions about the kinematics of the initiating particles. The jet four-momenta are reconstructed using dedicated jet clustering algorithms. Beyond the jet four-momentum, observables describing the substructure of the jet can be reconstructed. Using the jet substructure, it is possible to distinguish jets initiated by light quarks, gluon jets and jets containing the decay products of hadronically decaying heavy particles, i.e. t quarks and W, Z or H bosons.

This chapter introduces jet clustering algorithms and a selection of jet substructure observables. Since the identification of jets containing the decay of t quarks (t jets) is an important aspect of the analysis presented in this thesis, the techniques are discussed in the context of t jet identification.

3.1. Jet Clustering Algorithms

The task of jet clustering algorithms is to combine the cascade of hadrons and other particles created during the hadronization into a single object. In order to ensure that perturbative QCD calculations of the jets are finite, the clustering algorithms must be infrared and collinear (IRC) safe, i.e. the resulting jet must not change under the emission of infinitely soft particles or collinear splitting of a particle within the jet.

In the following, a set of IRC safe jet finding algorithms, called sequential clustering

algorithms, are introduced. The jet clustering algorithms used in this analysis are implemented within the FastJet framework [72]. Since the jet clustering can be applied to any set of objects with four-momenta, these objects will be referred to as pseudojets, in line with the FastJet terminology.

3.1.1. Sequential Clustering

Sequential clustering algorithms start with a list of pseudojets that are iteratively paired, until all objects are considered jets. The clustering is based on the distance measure $d_{i,j}$ between a pair of pseudojets i and j and the distance of a pseudojet i to the beam axis $d_{i,\mathrm{B}}$. They are defined as

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

$$d_{i,\mathrm{B}} = p_{\mathrm{T},i}^{2k}\,, \tag{3.2}$$

where $\Delta R_{i,j} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ is the distance between pseudojets i and j in the rapidity-azimuth plane and p_{T} is the transverse momentum of the pseudojet. The distance parameter R defines the jet radius and the parameter k determines clustering behavior of the algorithm.

For each pair of pseudojets, $d_{i,j}$ and $d_{i,\mathrm{B}}$ are calculated and the smallest $d_{i,j}$ and $d_{i,\mathrm{B}}$ of all pairs are identified. If $d_{i,j} < d_{i,\mathrm{B}}$, the pseudojets i and j are replaced by a single new pseudojet that is the sum of the four-momenta of i and j. If $d_{i,\mathrm{B}} < d_{i,j}$, the pseudojet i is stored as a jet and removed from the list of pseudojets. This procedure is repeated, until no pseudojets are left.

The order in which the pseudojets are combined is determined by the choice of k. The k_{T} algorithm [73,74] uses $k = 1$. Here, soft particles are combined first, resulting in more irregular shaped jets. In the anit-k_{T} algorithm [75] with $k = -1$, priority is given to the hardest pseudojets. Jets clustered with the anti-k_{T} algorithm have a more uniform, almost circular shape. At the LHC experiments, these jets are most commonly used. Finally, for $k = 0$ the Cambridge-Aachen (CA) algorithm [76,77] is obtained, where the order of clustering is determined purely by the geometrical distance between the pseudojets. This feature makes CA jets very suitable for many

jet substructure techniques, because the geometrical hierarchy within the jet can reveal underlying dynamics like the presence of subjets.

3.1.2. Varialbe R

The jet finding algorithms described above operate with a fixed distance parameter R. Hereby, the optimal choice of R depends on the use case of the jets. For jets used for t tagging, the choice of R directly influences the identification efficiency as a function of p_{T}, because the decay products of the t quark decay are more collimated with increasing p_{T} of the t quark. If R is too small, the decay of the t quark is only partially reconstructed within the jet. On the other hand, if R is chosen too large, the influence of non-perturbative effects on the jet substructure observables can degrade the discrimination power of the tagger.

The Variable R (VR) algorithm [78] proposes a solution to this by introducing a p_{T} dependent distance parameter

$$R_{\mathrm{eff}} = \frac{\rho}{p_{\mathrm{T}}}\,, \tag{3.3}$$

replacing the constant R in Eq. (3.1). The parameter ρ controls the slope of R_{eff} and has to be chosen according to the physics use case of the jets. Boundaries on R_{eff} can be introduced with

$$R_{\mathrm{eff}} = \begin{cases} R_{\mathrm{min}} & \text{if } \rho/p_{\mathrm{T}} < R_{\mathrm{min}}\,, \\ R_{\mathrm{max}} & \text{if } \rho/p_{\mathrm{T}} > R_{\mathrm{max}}\,, \\ \rho/p_{\mathrm{T}} & \text{else,} \end{cases} \tag{3.4}$$

to avoid too large or too small jets. Like for the conventional sequential clustering algorithms, the choice of k determines the clustering behavior of the VR algorithm.

3.2. Jet Substructure Observables

Jets containing the boosted hadronic decay products of heavy particles, like t quarks or W, Z and H bosons, feature distinct characteristics that can be used to distinguish

them from each other and from jets originating from light quarks or gluons (QCD jets). Jet substructure observables play a crucial role as discriminators in tagging algorithms. In the following, a selection of jet substructure algorithms are introduced, with a focus on the use for t tagging. A comprehensive summary of jet substructure techniques can be found in Ref. [79].

3.2.1. Jet Mass

The most important quantity in the identification of heavy particle initiated jets is the jet mass m_{jet}, defined as

$$m_{\text{jet}}^2 = \left(\sum_i P_i \right)^2 , \tag{3.5}$$

where P_i is the four-momentum of the jet constituent i. For a jet containing the hadronic decay of a heavy particle, m_{jet} is sensitive to the mass of the initiating particle and thus makes a strong discriminator against other processes. For QCD jets on the other hand, the mass of the initiating parton is negligible and the jet mass is dominated by radiation and hadronization effects, resulting in a Sudakov peak that shifts in m_{jet} with p_{T} of the jet [80].

3.2.2. Jet Grooming

Jet grooming techniques aim at removing soft wide-angle radiation from jets to obtain a better description of jet substructure observables. Two different grooming algorithms are introduced in the following.

The soft drop algorithm [81] is a generalized version of the modified Mass Drop Tagger [80]. The algorithm starts by reverting the last clustering step and obtaining the last two pseudojets that were combined. The soft drop condition

$$\frac{\min(p_{\text{T},i}, p_{\text{T},j})}{p_{\text{T},i} + p_{\text{T},j}} > z_{\text{cut}} \left(\frac{\Delta R_{i,j}}{R} \right)^{\beta} \tag{3.6}$$

is checked for the two pseudojets i and j. Here, p_{T} denotes the transverse momenta

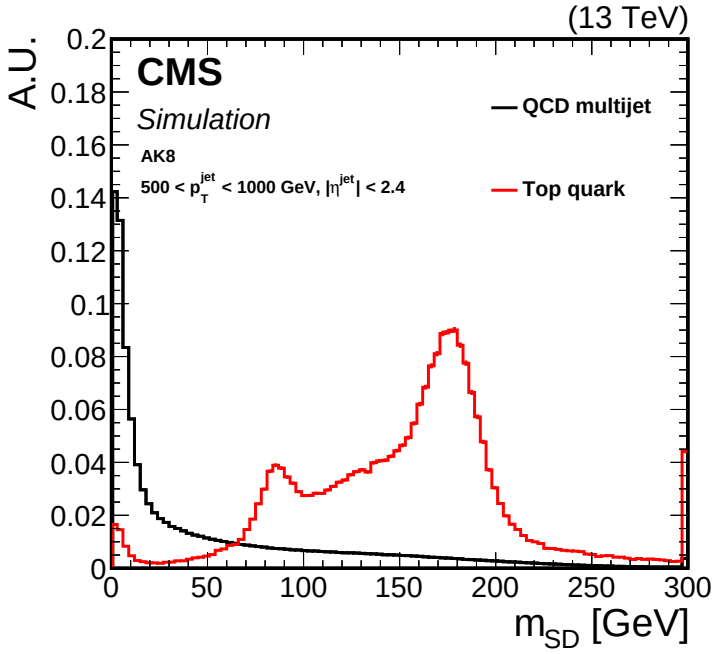

Figure 3.1.: Distribution of the soft drop mass m_{SD} of jets clustered with the anti-k_{T} algorithm and $R = 0.8$ from $\mathrm{t\bar{t}}$ (red) and QCD multijet (black) simulation. All jets have $500\,\mathrm{GeV} < p_{\mathrm{T}} < 1000\,\mathrm{GeV}$ and $|\eta| < 2.4$. Taken from Ref. [82].

of the pseudojets, $\Delta R_{i,j}$ is the distance in the rapidity-azimuth plane, and R is the distance parameter of the jet algorithm. The threshold parameter z_{cut} determines the strength of the grooming and the angular exponent parameter β controls the angular dependency of the criterion. For $\beta = 0$, the modified Mass Drop Tagger condition is obtained. If the soft drop condition is not fulfilled, the pseudojet with lower p_{T} is discarded and the remaining pseudojet is declared the jet. The procedure is repeated until the soft drop condition is fulfilled.

Soft drop utilizes jets clustered with the CA algorithm, because the emissions in the parton shower are following approximately an angular ordering [81]. For jets clustered with different algorithms, the jet constituents are reclustered with the CA algorithm first before applying the algorithm and obtaining groomed substructure observables.

The jet mass of a soft drop groomed jet m_{SD} improves the discrimination power of the plain jet mass m_{jet}, because the removal of the soft radiation suppresses the Sudakov peak structure of the jet mass distribution [83]. For t jets, m_{SD} peaks

around the t quark mass, while for QCD jets the m_{SD} distribution peaks at values close to zero. A comparison of the m_{SD} distributions of t jets (red) and QCD jets (black) is shown in Fig. 3.1 for jets clustered with the anti-k_{T} algorithm with $R = 0.8$. To obtain m_{SD}, the jet constituents were reclustered using the CA algorithm before applying the soft drop algorithm. The soft drop parameters are chosen as $z_{\mathrm{cut}} = 0.1$ and $\beta = 0$. The t quark distribution features a prominent peak around the t quark mass, arising from jets that capture the complete hadronic decay. The shoulder towards lower m_{SD} of the t mass peak and the second peak around the W boson mass arise from jets, where only two of the three decay products are merged into the jet. Similarly, the peak at zero constitutes of jets where only one quark was merged into the jet, thus showing same signature as QCD jets.

The mass jump algorithm [84] takes a different approach to jet grooming. Instead of reverting the clustering process and discarding pseudojets, the mass jump algorithm is applied during the jet clustering process. The clustering follows the sequential clustering approach detailed in Section 3.1.1. Here, all pseudojets begin labeled as active and in each iteration, only active psudojets are considered. If $d_{i,\mathrm{B}}$ is found the smallest distance measure, the pseudojet is not declared a jet, but instead labeled passive. Additionally, if the mass of the combined pseudojets m_{ij} is found to be above a threshold $m_{ij} > \mu$, the mass jump criterion

$$\theta \cdot m_{ij} > \max(m_i, m_j) \tag{3.7}$$

is checked, where m denotes the invariant mass of the pseudojets. The threshold parameter θ contols the strength of the condition. If a mass jump is found, the pseudojets i and j are labeled passive. The mass jump condition is also checked for each active pseudojet i and passive pseudojet k, where $d_{i,k} < d_{i,j}$. If no mass jump was found, pseudojets i and j are combined. The clustering is finished, if no active jets are left. All passive jets are then labelled jets. A collection of groomed jets can be obtained by dropping all passive jets that did not fulfill the mass jump condition.

3.2.3. N-subjettiness

The N-subjettiness τ_N [85, 86] is a measure for the compatibility of a jet consisting of N or less subjets. It is defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T},k} \min \left(\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}, \right) , \tag{3.8}$$

where $\Delta R_{j,k}$ is the distance between the jet constituent k and subjet axis j in the pseudorapidity-azimuth plane, and $p_{\mathrm{T},k}$ is the transverse momentum of the jet constituent k. The factor d_0 is used to normalize τ_N to the interval $(0, 1)$, using

$$d_0 = \sum_k p_{\mathrm{T},k} R \, , \tag{3.9}$$

where R is the distance parameter of the jet algorithm used for clustering the jet. The subjet axes are found by minimizing τ_N.

The value of τ_N becomes small, if all jet constituents are aligned with the N subjet axes. Generally, τ_N alone does not offer much discrimination power between jets from an N-pronged decay and QCD jets. A better discrimination is achieved using the ratio of two N-subjettiness values, τ_N/τ_M, where $N > M$. For t jet identification, the ratio τ_3/τ_2 gives good discrimination power, because the typical three-prong structure of a t jet results generally in a significantly larger τ_2 than τ_3. In Fig. 3.2, distributions of τ_2 (top left), τ_3 (top right) and τ_3/τ_2 (bottom) are shown for simulated $\mathrm{t\bar{t}}$ events (red) and QCD multijet events (blue). It can clearly be seen that the ratio τ_3/τ_2 offers a significantly better discrimination power compared to the single values τ_2 or τ_3.

3.2.4. Heavy Object Tagger with Variable R

Based on the VR algorithm described above, the Heavy Object Tagger with Variable R [7] (HOTVR) introduces a mass jump [84] inspired criterion during the jet clustering. This addresses a problem of the original VR algorithm, which is the clustering of additional radiation and thus artificially increasing the jet p_{T} [78]. Additionally, the mass jump criterion allows the identification subjets during the jet clustering.

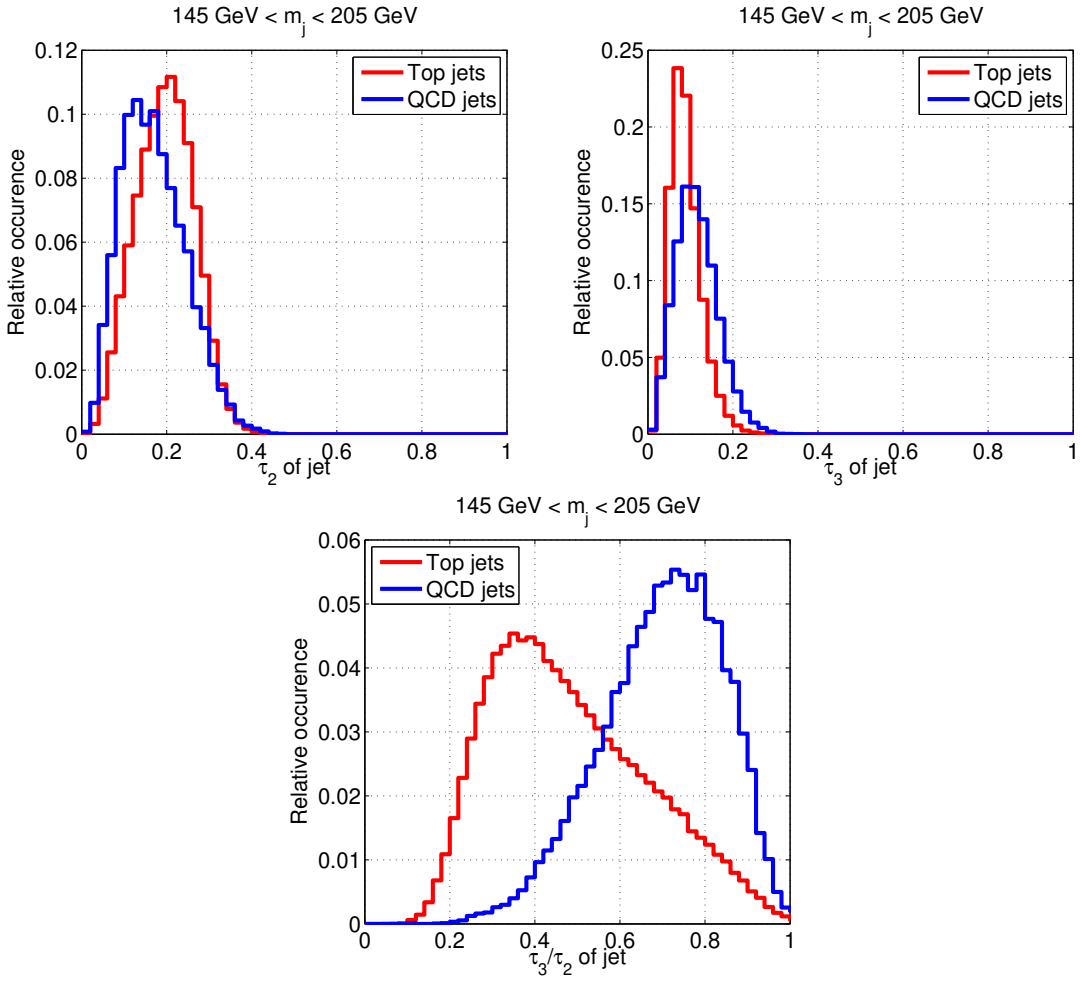

Figure 3.2.: Distributions of τ_2 (top left), τ_3 (top right) and τ_3/τ_2 (bottom) for a top jet (red) and a QCD jet (blue). Taken from Ref. [85].

The HOTVR algorithm modifies the sequential clustering procedure described in Section 3.1.1 by requiring the mass jump condition Eq. (3.7) when the combined mass of the pseudojets i and j, m_{ij}, is above a threshold $m_{ij} > \mu$. If the mass jump condition is not fulfilled, the lighter pseudojet is removed from the list of pseudojets. If a mass jump is found, the p_T of both pseudojets is compared to the parameter $p_{\mathrm{T,sub}}$ and the respective pseudojet is removed from the input list, if $p_\mathrm{T} < p_{\mathrm{T,sub}}$. If both pseudojets fulfill this criterion, they are combined to a single pseudojet. The original pseudojets are stored as subjets of the new pseudojet. HOTVR jets are clustered in CA mode, i.e. $k = 0$.

In the presented analysis, the HOTVR parameters are set to the default values for t tagging mode [7], which are summarized in Table 3.1.

Parameter	R_{min}	R_{min}	ρ	μ	θ	$p_{\text{T,sub}}$
Value	0.1	1.5	600 GeV	30 GeV	0.7	30 GeV

Table 3.1.: Parameters of the HOTVR algorithm in the t tagging mode [7].

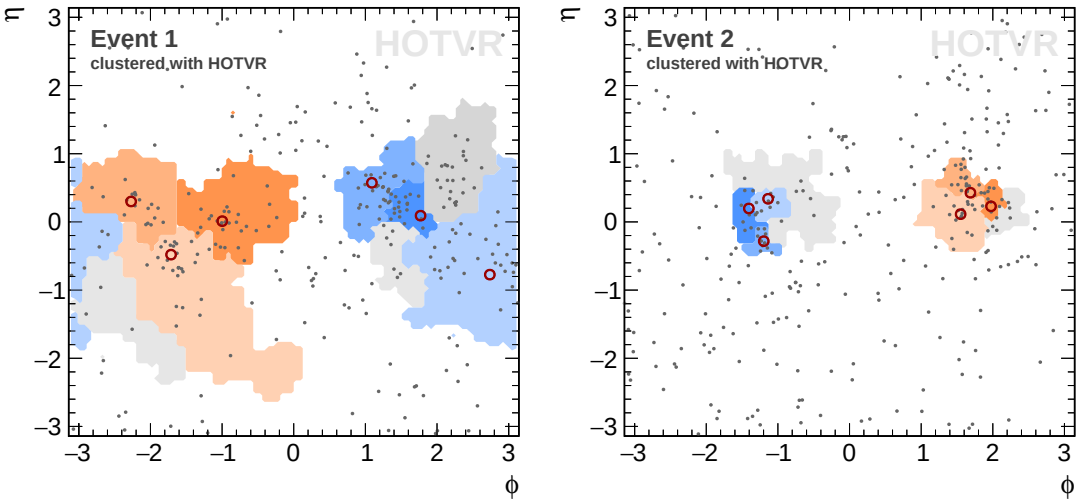

Figure 3.3.: Jet displays showing the jet area of the two jets leading in p_{T} from two simulated $\text{t}\bar{\text{t}}$ events as blue and orange shaded areas. The jets were clustered with the HOTVR algorithm. The different shades depict the jet areas of the subjets. Event 1 (left) shows an event with low-p_{T} t quarks and Event 2 (right) shows an event with high-p_{T} t quarks. The input of the clustering algorithm is shown as grey dots. For illustration purposes, the three quarks from the hadronic t decay are shown as red circles. The grey areas show pseudojets rejected by the mass jump condition. Taken from Ref. [7].

In Fig. 3.3, jets from low-p_{T} (left) and high-p_{T} (right) simulated $\text{t}\bar{\text{t}}$ events clustered with the HOTVR algorithm are visualized. The jet areas of the two leading jets in p_{T} are drawn in the pseudorapidity-azimuth plane as orange and blue colored areas. The grey area shows the regions rejected by the mass jump condition. The dynamic adaptation of R_{eff} can be seen in the different jet sizes obtained by HOTVR between the two events. In both cases, the HOTVR jets reproduce the hadronic top decays with their subjets matching the three hadronic decay products, depicted by the red circles.

4. Experimental Setup

The analysis presented in this thesis is based on pp collision data recorded with the CMS detector at the LHC in the years 2016–2018. A description of the LHC, summarizing the experimental conditions, and an overview of the CMS detector are given in this chapter.

4.1. The Large Hadron Collider

The LHC [87] is a superconducting synchrotron and storage ring with a circumference of 26.7 km, designed to collide protons or heavy ions. It is located at the European Organization for Nuclear Research (Conseil Européen pour la Recherche Nucléaire - CERN) near Geneva, Switzerland. In the following, the LHC operation is detailed assuming pp collision mode.

Protons are obtained from the ionization of hydrogen molecules and then prepared and accelerated in a chain of pre-accelerators before they are injected into the LHC ring. A schematic overview of the full CERN accelerator complex is shown in Fig. 4.1. The first accelerator in this chain is the linear accelerator Linac 2. After this, the protons are injected into the Proton Synchrotron Booster, followed by the Proton Synchrotron and finally the Super Proton Synchrotron, where they reach an energy of 450 GeV. From the Super Proton Synchrotron, the protons are injected into the two counter-rotating beam pipes of the LHC. Due to the different sizes of the accelerators, it takes several cycles until the LHC ring is fully filled and the beam energy can be ramped up to the operational energy. After the ramp up, stable beam conditions are declared and the proton beams are collided at dedicated interaction points along the ring. The proton beams can be stored for several hours in the LHC ring, before they

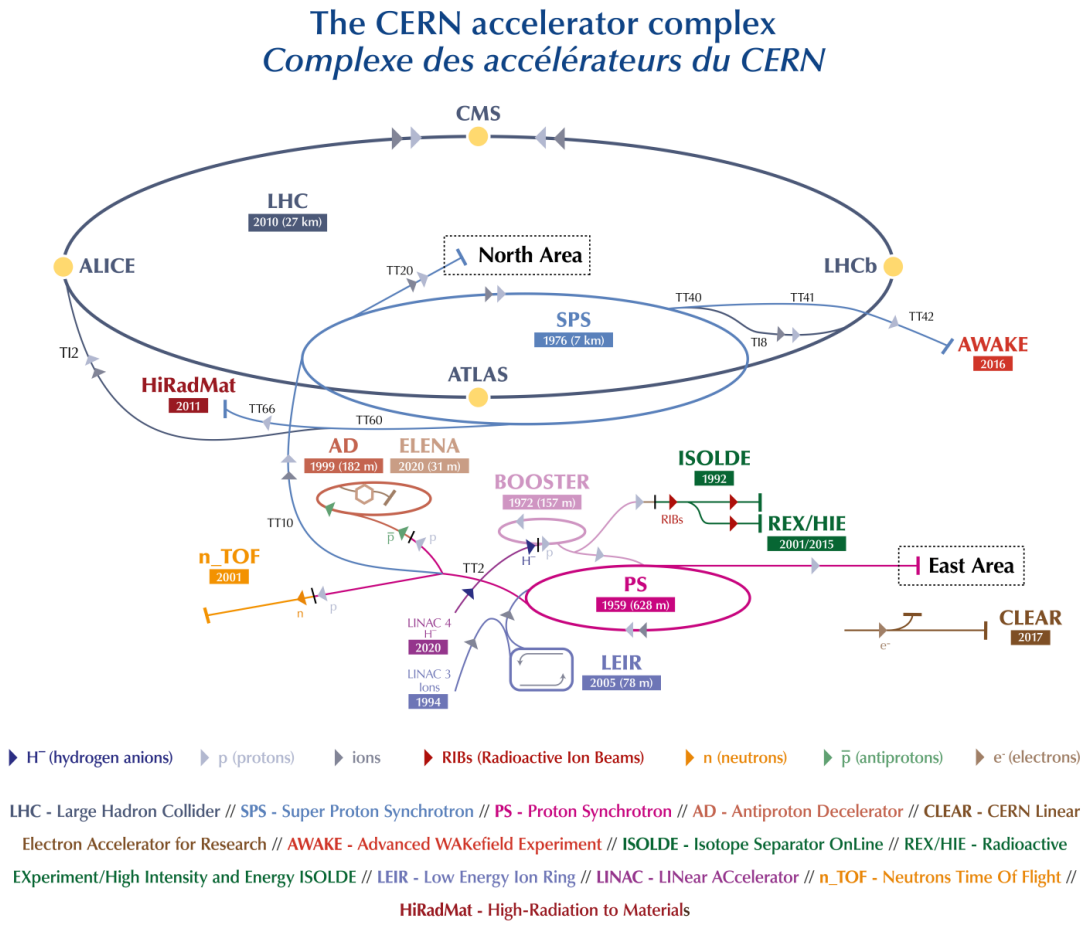

Figure 4.1.: The CERN accelerator complex. Taken from Ref. [88].

are dumped and the LHC is filled again.

The LHC was built as a *discovery machine*, allowing the possible discovery of new physics, as well as precise tests of the SM at unprecedented energies. It was designed to operate at a center-of-mass energy of 14 TeV, however during the two data taking runs that were completed to date, the design center-of-mass energy was not yet reached. In the first run from 2010 to 2012, Run 1, the LHC was operated at a center-of-mass energy of 7 and 8 TeV. The energy was increased to 13 TeV in Run 2, which lasted from 2015 to 2018. In addition to the high center-of-mass energies, another critical parameter of the LHC is its high instantaneous luminosity $\mathcal{L}$. It is a measure of the event rate per unit cross section

$$\mathcal{L} = \frac{dN}{dt}\frac{1}{\sigma}. \tag{4.1}$$

The instantaneous luminosity is dependent on a number of proton beam parameters and can be expressed as

$$\mathcal{L} = N_b f_r \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}\,, \tag{4.2}$$

where N_b is the number of bunches per beam, f_r is the revolution frequency of the bunches, n_1 and n_2 are the numbers of protons in the colliding bunches, and σ_x and σ_y are the transverse beam sizes in x and y direction. The LHC was designed for a luminosity of $\mathcal{L} = 1 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and has already surpassed this design value, reaching an instantaneous luminosity as large as $\mathcal{L} = 2.2 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in 2018.

Integrating Eq. (4.1) over the measurement time yields the expected number of recorded events per unit cross-section:

$$N = \sigma \int \mathcal{L}\,\mathrm{d}t = \sigma L\,. \tag{4.3}$$

Hence, the size of a dataset is usually measured using the recorded integrated luminosity L. The evolution of the integrated luminosity delivered by the LHC to the CMS experiment as a function of time for the years 2010–2018 is shown in Fig. 4.2.

The particle collisions are recorded by the LHC experiments ALICE (A Large Ion Collider Experiment), ATLAS (A Toroidal LHC Apparatus), CMS (Compact Muon Solenoid), and LHCb (LHC beauty), located at four interaction points along the LHC ring. The ATLAS and CMS experiments utilize multipurpose detectors to study a broad field of physics, such as precision measurements of the SM and searches for new physics beyond. The LHCb experiment is focused on measurements of b quark physics. The ALICE experiment examines heavy-ion collisions to study the quark-gluon plasma.

4.2. The Compact Muon Solenoid

The CMS experiment [90, 91] is a multipurpose detector located at the interaction point 5 of the LHC in Cessy, France. The detector is 21.6 m long, with a diameter of 14.6 m. It consists of several sub-detector layers arranged in a central barrel part and two end cap parts, covering in total a solid angle of almost 4π. Each sub-detector system is dedicated to the measurement of different particle types and observables.

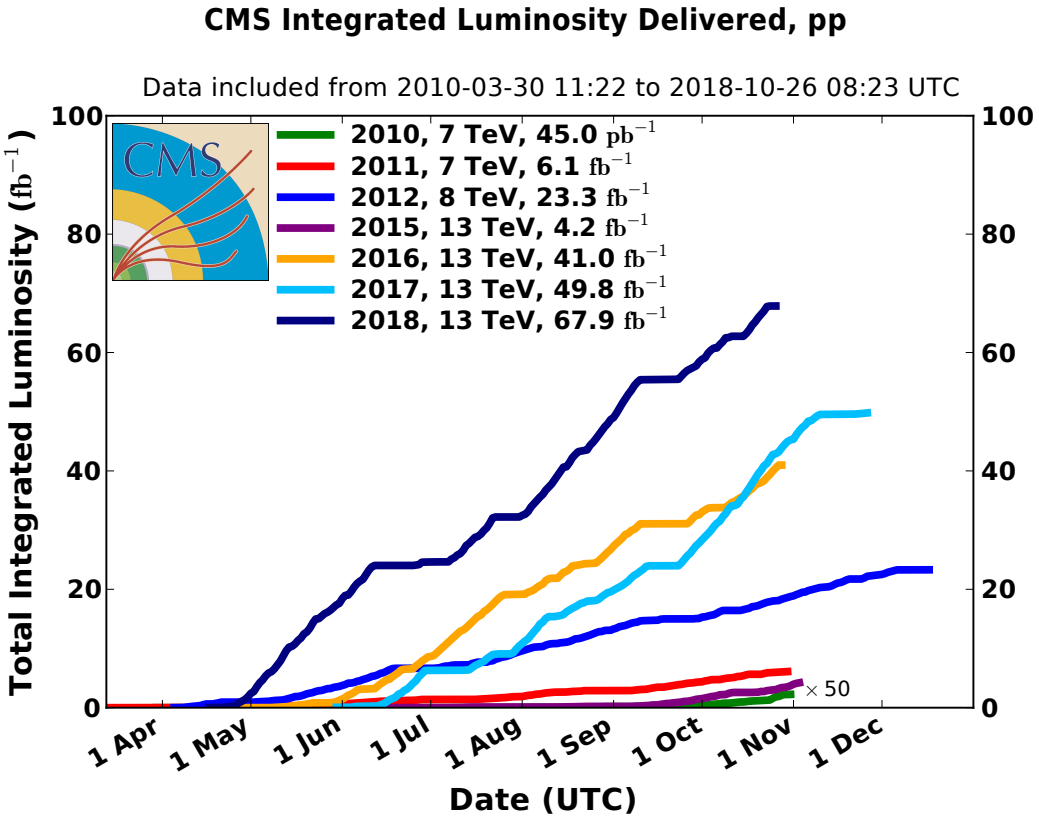

Figure 4.2.: Cumulative luminosity delivered by the LHC to the CMS experiment during stable beams conditions for pp collisions as a function of time for the years 2010–2018. Taken from Ref. [89].

The first system that particles emerging from the collision point traverse is the inner tracking system, where the trajectories of charged particles are measured. Built around the inner tracking system are the calorimeters, dedicated to stop the particles and measure their energy. A compact design was chosen, where the tracking system and the calorimeters are located within the coil of the large solenoid magnet. The superconducting solenoid provides a powerful magnetic field parallel to the beam axis to bend the trajectories of charged particle, enabling the charge and momentum measurement. Muon detectors are located in the outer part of the detector interlaced with the magnets iron return yoke to detect muons, which are able to traverse the calorimeters. A schematic overview of the CMS detector is shown in Fig. 4.3. Each sub-system will be discussed in the following sections.

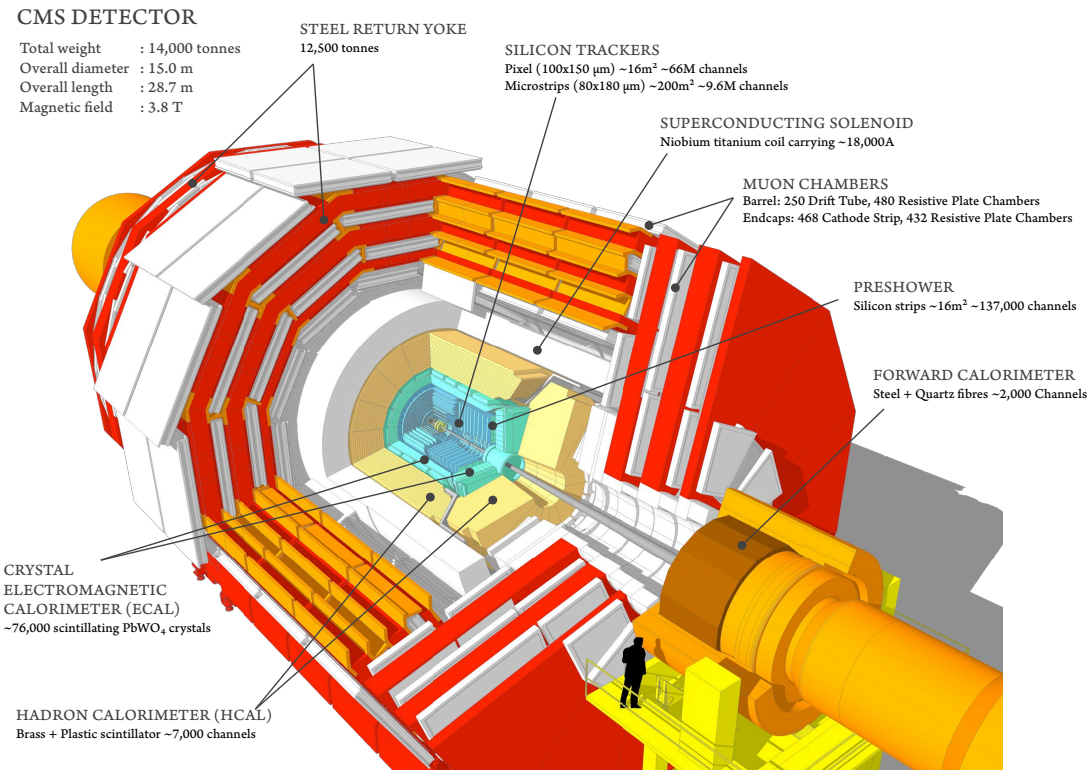

Figure 4.3.: Cutaway view of the CMS detector. Taken from Ref. [92].

4.2.1. Coordinate System

The CMS experiment uses a right-handed coordinate system that has the origin at the nominal collision point in the center of the detector. The x-axis points inward toward the center of the LHC ring, the y-axis points upward, and the z-axis points along the beam axis in counterclockwise direction. It is convenient, given the detector geometry, to use spherical coordinates, using the radial distance r from the z-axis, the azimuthal angle ϕ measured in the x-y-plane starting at the x-axis, and the polar angle θ measured from the z-axis.

4.2.2. Inner Tracking System

Built directly around the interaction point is the inner tracking system, used to measure the trajectories of charged particles. The trajectories are used as input to determine the particles charge signs and momenta, and to identify primary and

secondary interaction vertices. The inner tracking system is 5.8 m long, 2.5 m in diameter and covers a pseudorapidity of $|\eta| < 2.5$. It consists of two sub-systems, the pixel detector and the strips detector, both exclusively utilizing silicon semiconductor detectors. A sketch of the CMS tracker layout is shown in Fig. 4.4. The inner tracking system achieves a p_{T} resolution of 1% for charged particles with $p_{\mathrm{T}} < 20\,\mathrm{GeV}$ [93]. The momentum resolution then degrades with increasing p_{T}.

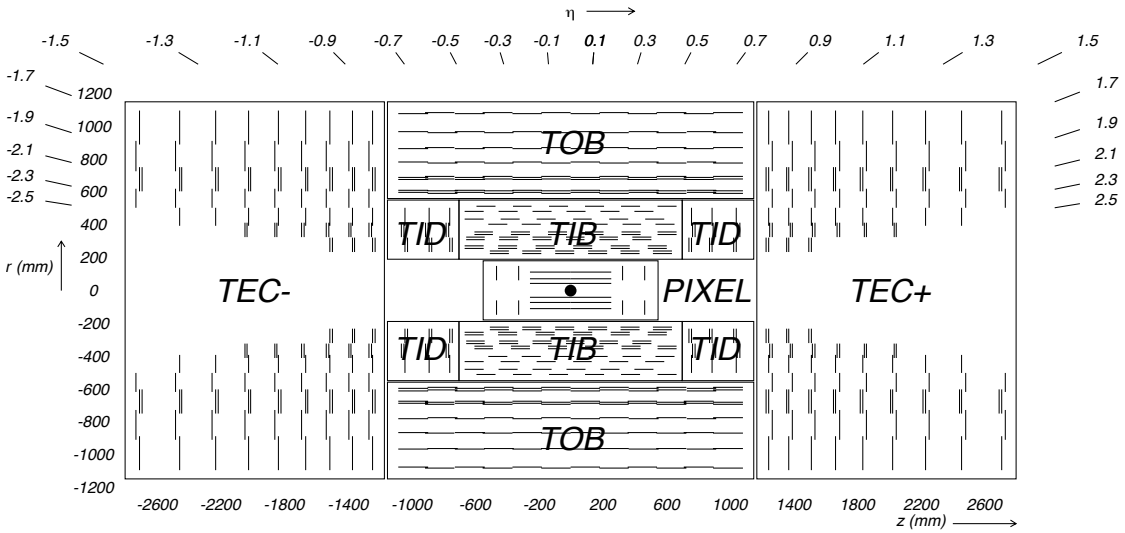

Figure 4.4.: Schematic cross section through the CMS tracker. The figure shows the Phase-0 pixel detector layout. Taken from Ref. [90].

Closest to the interaction point is the pixel detector using silicon pixel sensors with a size of 100 µm × 150 µm. The first detector layout, called Phase-0 detector, features a barrel part (BPIX) of three cylindrical layers at a distance of 4.4, 7.3 and 10.2 cm from the beam pipe, complemented by two disks on the end caps (FPIX) at distances of 34.5 and 46.5 cm in $|z|$ from the interaction point, extending from 6 to 16 cm in r.

In the year-end technical stop between the 2016 and 2017 data-taking runs, the pixel detector was upgraded to the Phase-1 design [94,95] to ensure full performance at the increased instantaneous luminosity and hence higher particle densities and greater number of overlapping collision events expected in the following runs. The Phase-1 pixel detector consists of four layers at 3.0, 6.8, 10.2 and 16.0 cm radii and three disks with a range of 4.6 to 16.1 cm in r and a distance of 29.1, 39.6 and 51.6 cm in $|z|$ from the interaction point. A comparison of the pixel detector layout before and after the upgrade is shown in Fig. 4.5. Additionally, a new read out chip for the pixel sensors was developed to cope with the higher instantaneous luminosity in the

2017 and 2018 runs. The higher power consumption due to the increased number of modules in the pixel detector was addressed by a new DC-DC converter based powering system. The C_6F_{14} liquid cooling system was upgraded to an evaporative CO_2 cooling system.

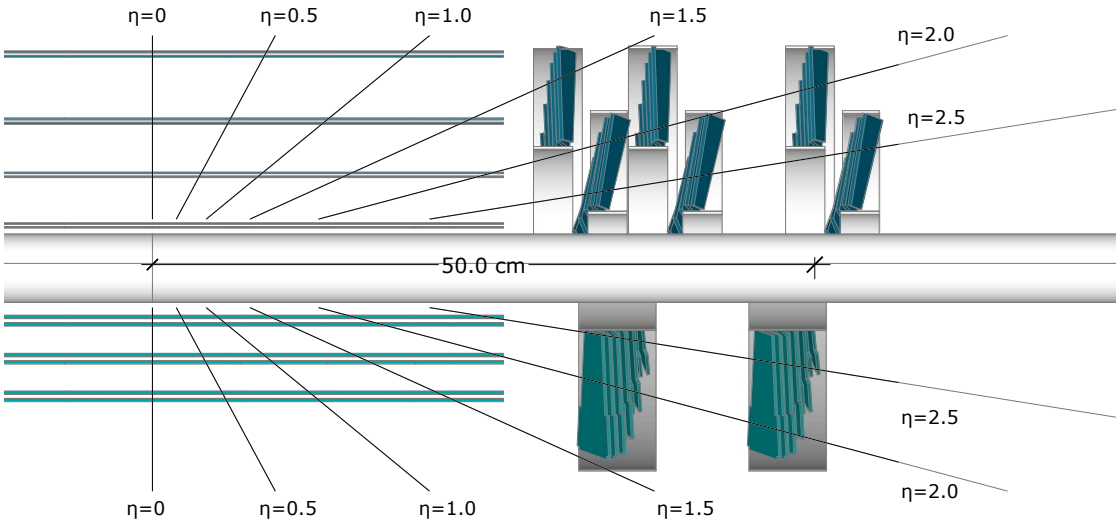

Figure 4.5.: A comparison of the Phase-0 (bottom) and Phase-1 (top) pixel detector layouts. Taken from Ref. [92]

Surrounding the pixel detector is the strip detector. It is arranged in two inner parts, the Tracker Inner Barrel (TIB) and the Tracker Inner Disks (TID), and two outer parts, the Tracker Outer Barrel (TOB) and the Tracker EndCaps (TEC). The inner systems extend from 20 to 55 cm in r, consisting of four layers in TIB and three disks in TID. Both systems feature silicon micro-strip sensors with a thickness of 320 µm. Enclosing TIB and TID is TOB with six layers, extending from 55 to 110 cm in r and covering the region within $|z| < 110$ cm. At this distance from the interaction point, the track density is lower, allowing to use strip sensors with a thickness of 500 µm. Complementing TOB is TEC with 9 discs on each side, extending from 120 to 280 cm in $|z|$. Each disc carries up to 7 rings of strip detectors with a thickness of 320 µm on the inner 4 rings and 500 µm on the outer 3 rings.

4.2.3. Electromagnetic Calorimeter

Surrounding the inner tracking detector is the electromagnetic calorimeter (ECAL), a homogeneous calorimeter designed to measure the energy of electrons and photons. It is made up of lead tungstate ($PbWO_4$) crystal scintillators, arranged in a central barrel part (EB) complemented by two endcaps (EE). An overview of the ECAL is shown in Fig. 4.6. The high density of the $PbWO_4$ results in a short radiation length of $X_0 = 0.89\,\text{cm}$ and a small Moliére radius of $R_M = 2.2\,\text{cm}$, allowing for a high granularity and still compact design. Additionally, the crystals have a short scintillation decay time, where about 80% of the light is emitted in 25 ns.

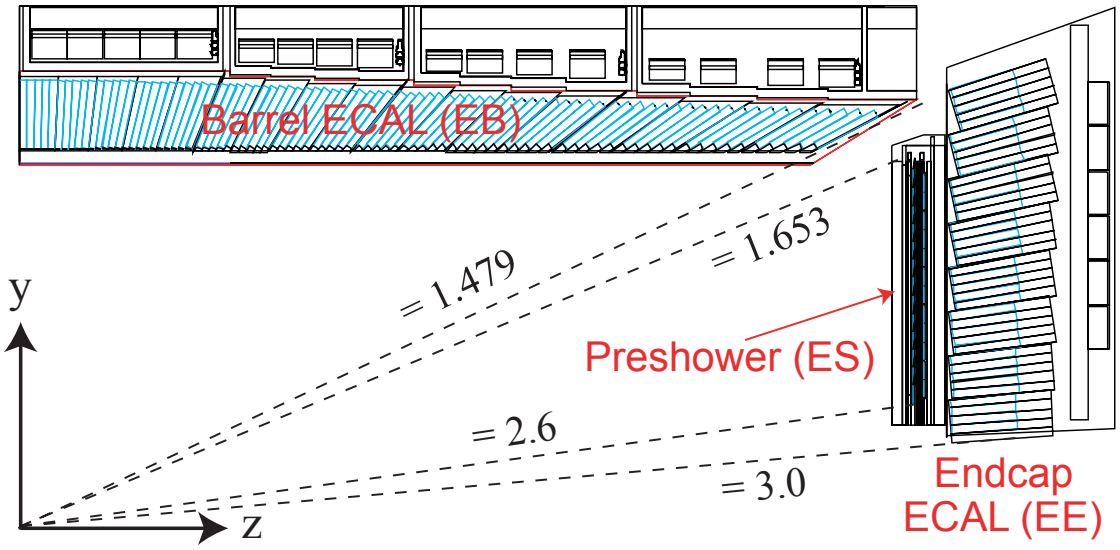

Figure 4.6.: Cross section trough one quarter the ECAL. Taken from Ref. [91].

The EB covers the pseudorapidity range $|\eta| < 1.479$ and is equipped with crystals with a length of 230 mm ($25.8\,X_0$) and a cross section of $22\,\text{mm} \times 22\,\text{mm}$ at the front face, mounted at a radius of 129 cm. The EE consist of 220 mm long ($24.7\,X_0$) crystals with a front face cross section of $28.62\,\text{mm} \times 28.62\,\text{mm}$. They are placed at a distance of 3.14 m along z from the interaction point and cover the pseudorapidity range $1.479 < |\eta| < 3.0$.

Located in front of the EE are the preshower detectors (ES), which are used to identify neutral pions and improve the discrimination of electrons against minimum ionizing particles. The ES are sampling calorimeters, made up of two absorber layers interlaced with silicon strip sensors. They cover the pseudorapidity range $1.653 < |\eta| < 2.6$ and have a thickness of 20 cm, corresponding to $3\,X_0$.

The resolution of the ECAL was measured using electrons at a test beam setup. It can be parametrized as a function of energy using three terms [90]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\mathrm{GeV}}} \oplus \frac{12\%}{E/\mathrm{GeV}} \oplus 0.3\% \,. \tag{4.4}$$

The $\oplus$ denotes that the terms are added in quadrature. The first term describes stochastic effects in the shower development, the second term describes electronic noise, and the third term describes constant contributions arising from calibration errors and non-linearity in the readout elements.

4.2.4. Hadronic Calorimeter

The hadron calorimeter (HCAL) is the final sub-system inside the magnet coil, enclosing the inner tracker and ECAL. It is a sampling calorimeter made of alternating layers of brass or stainless steel absorber and active plastic scintillators. In combination with the ECAL, the HCAL is used for the measurement of hadron jets and missing transverse energy. The layout of the HCAL is shown in Fig. 4.7, it consists of four separate sections, a barrel section (HB) inside the solenoid coil, complemented by an outer section (HO) located outside the magnet coil, as well as endcap (HE) and forward (HF) sections.

The HB extends from 177.7 to 287.65 cm in r and covers the pseudorapidity range $|\eta| < 1.3$. It consists of 17 active layers, with the first layer starting directly behind the ECAL. The first and last absorber layers are made of stainless steel, the rest of brass. At $|\eta| = 0$, the total absorber thickness amounts to 5.82 interaction lengths (λ_I) and increases with $1/\sin\theta$ to $10.6\,\lambda_I$ at $|\eta| = 1.3$. The ECAL material in front of HB adds about $1.1\,\lambda_I$. Outside of the magnet coil, covering the region $|\eta| < 1.26$ is the HO. It is composed of scintillator layers and utilizes the magnet coil as absorber material, aiming to measure particle showers that have passed the HB and the magnet coil and ensures an absorber thickness of at least $11.8\,\lambda_I$ in the entire barrel section. Inserted into the ends of the solenoid magnet and overlapping with the HB are the HE, which cover $1.3 < |\eta| < 3.0$. Each consists of 19 active plastic scintillator layers, interspersed with brass absorbers. The total absorber thickness, including the material of the ECAL is $10.6\,\lambda_I$. In the very forward region, covering $2.9 < |\eta| < 5.2$, are the HF. The harsh radiation environment so close to the beam pipe played a

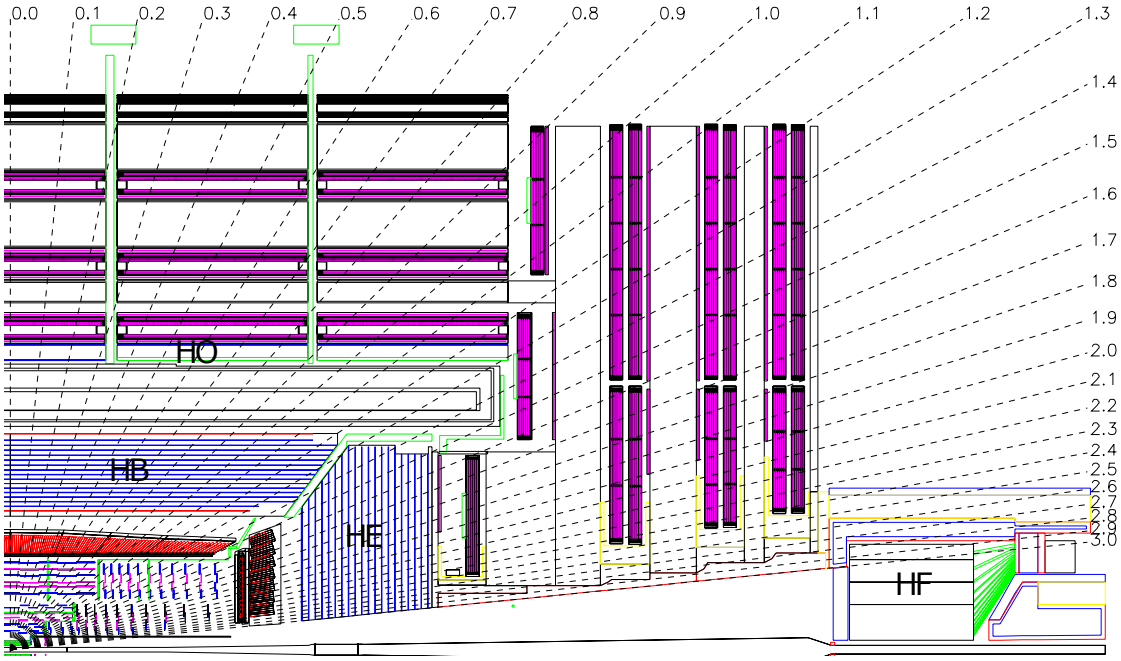

Figure 4.7.: Cross section trough one quarter the HCAL. Taken from Ref. [90].

critical role in the design of this sub-system. Hence, the HF calorimeters are made of grooved steel absorbers with radiation hard quartz fibers inserted in the grooves as active medium and an absorber thickness of about $10\,\lambda_I$.

The resolution of the HCAL, including the ECAL, was studied at a test beam experiment and determined to [96]

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E/\mathrm{GeV}}} \oplus 4.5\%\,, \tag{4.5}$$

following the parametrization in Eq. (4.4).

4.2.5. Solenoid Magnet

The CMS detector features a superconducting solenoid magnet with an inner diameter of 5.9 m and a length of 12.9 m that houses the inner tracking system and calorimeters. The magnet coil is wound with four layers of NbTi and provides a homogeneous magnetic field of 3.8 T on its inside. The magnetic field bends the trajectories of charged particles and allows the a precise momentum reconstruction. The magnetic flux is returned through iron yokes on the outside.

4.2.6. Muon System

The outermost sub-system of the CMS detector, housed in the return yoke, is the muon system. It is dedicated to identify muons, which can pass through the whole detector without considerable energy loss. The large area covered by the muon detectors and the varying radiation environments played a key role in the choice of detector technology. Three different types of gaseous detectors are used, each optimized to operate in a different section of the detector. In the barrel region, where the magnetic field is homogeneous and muon rates are low, drift tubes (DT) are used. They are arranged into 4 stations, alternating with the return yoke layers and cover the pseudorapidity range $|\eta| < 1.2$. In the endcap regions, where the magnetic field is non-uniform and the particle flux is higher, cathode strips changers (CSC) are used. They as well are arranged in 4 stations, embedded between the return yoke layers and cover the pseudorapidity range $0.9 < |\eta| < 2.4$. A system of resistive plate chambers (RPC) complements the DT and CSC detectors in the range $|\eta| < 1.6$. This system is used as an independent second input for the muon trigger system and allows an allocation of the correct bunch crossing. The RPCs are arranged in 6 layers in the barrel and 2 in the endcaps. A layout of the muon system is shown in Fig. 4.8.

Information from the muon detectors is used in combination with information from the inner tracking system to improve the p_{T} resolution for muons with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$. For muons with p_{T} up to 1 TeV, a p_{T} resolution of less than 7% is achieved [97].

4.2.7. Trigger System

At the nominal bunch spacing of 25 ns and design luminosity, there are $\mathcal{O}(10^9)$ pp interactions occurring per second. It is technically not feasible to permanently store all collision events. A two level trigger system is used to reduce the event rate to a manageable rate of about 1 kHz [98].

The detector data is held in a buffer for 3.2 µs. During that time the signal travels from the detector electronics to the Level-1 Trigger (L1) and back, leaving less than 1 µs for the L1 to form a decision. The L1 decides to keep an event based on data from the calorimeters and muon system and is able to reduce the event rate to about

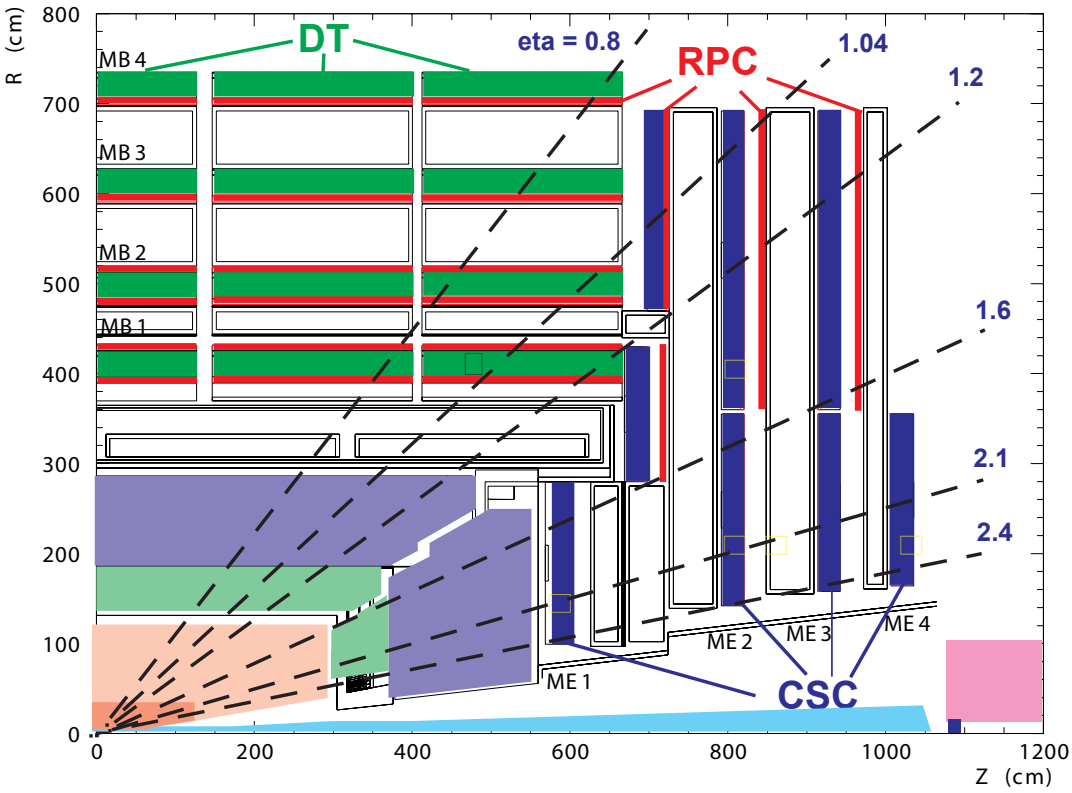

Figure 4.8.: Cross section trough one quarter the muon system. Taken from Ref. [91].

100 kHz. After an event passes the L1 trigger, the full detector data is read out and transferred to the High-Level trigger (HLT). The HLT is a software based trigger that runs on a processor farm and forms a decision based on the full detector data, allowing for a more dedicated analysis.

5. Online Monitoring of the CMS Pixel Detector

The pixel detector upgrade outlined in Section 4.2.2 introduced a number of new technologies to the detector, including a new pixel detector module design and powering scheme. In Fig. 5.1, a sketch of the new pixel detector modules for the BPIX layers 2-4 is shown. The design of the modules in other parts of the pixel detector vary only slightly. The upgraded modules each feature 16 read out chips (ROC) that are bump-bonded to the silicon sensors. The ROCs are connected to the high density interconnect with token bit manager (TBM), which distributes power and communicates with the ROCs. To cope with the increased power consumption of the upgraded pixel detector, a powering system utilizing DC-DC converters is employed [94, 95].

During the operation of the new pixel detector in 2017 it was found that the TBM could become stuck due to a single event latch-up [99]. These events are caused by single ionizing particles corrupting an unprotected register of the TBM. In order to recover a stuck TBM, a power reset has to be performed. During operations in 2017, this was performed by power-cycling, i.e. disabling and enabling, the respective DC-DC converter. However, in October 2017, DC-DC converters started malfunctioning during power-cycling and could not be enabled again. All DC-DC converters were extracted in the year-end technical stop between 2017 and 2018, and afterwards characterized. Of the extracted converters, 65 were found malfunctioning and 333 showed an increased power consumption in the disabled state [100]. Further investigations by the chip designers found that the malfunction was caused by a transistor that was not sufficiently shielded against radiation damage induced leakage current [99]. As a consequence, the power reset procedure for stuck TBMs was changed in 2018 and the power-cycling was performed via the power supply channels. In the long shutdown after the 2018 run, all DC-DC converters were replaced with

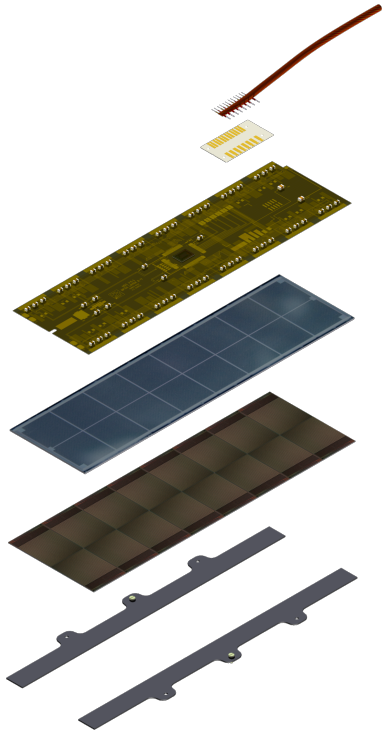

Figure 5.1.: Exploded view of a module from the BPIX layers 2-4 of the upgraded pixel detector. From top to bottom: signal and power cable with connector, high density interconnect with token bit manager, silicon sensors, readout chip, and base strip. Taken from Ref. [94].

an upgraded version of the chip that addressed the discovered issue.

The investigation of the DC-DC malfunctions lead to the formation of the Tracker Online Monitoring group with the task to monitor the tracking system and its auxiliary systems, collect information about known issues and provide tools for detector monitoring to the tracker operation teams. The goal is to be able to catch issues in the detector early and provide the information necessary to investigate and resolve these issues, while minimizing the downtime during data taking.

The tracker monitoring group provides online platforms, that allow quick access to detector status overviews created from information gathered by the detector control system and the tracker operations group. Additionally, the monitoring group deploys monitoring databases, where information gathered by the group is stored. In the following, the design of the pixel monitoring database is detailed, which was developed by the author of this text. Afterwards, the two monitoring platforms Tracker Online Monitor and Online Monitoring System are introduced and their

interactions with the pixel monitoring database are discussed.

5.1. Pixel Monitoring Database

The pixel monitoring database is an Oracle database implemented in the central CMS database system. In the core of the pixel monitoring database is a mapping of the CMS pixel detector, including the powering and data acquisition (DAQ) systems, which allows to directly link information gathered by the monitoring group to the respective detector components. Each component in this mapping is assigned a name, following the naming conventions of the actual detector. The name encodes information about the type and the position of the component. This information is also stored in separate tables in the database and linked to the component name. Links between tables in the database are established using foreign key relations, where a unique ID of one table is referenced in the other table. The name, position and type of component will not change, unless the detector design is changed, i.e. the detector is upgraded or replaced. However, single components can be replaced and the gathered information about the replaced component becomes invalid for the new component. To encode this behavior in the database, each name is linked to a so called *part*. This part reflects the physical component occupying the position encoded by the name. The link between a part and a name is assigned an interval of validity (IOV), i.e. a time interval in which this link is true. Information gathered by the monitoring group about a detector component, like calibration results and problem reports are directly linked to the respective part in the database. If a detector component is replaced, the IOV for this part ends and a new part is linked to the respective name, representing the new component. An illustration of the relation between parts and names is shown in Fig. 5.2.

The complex relations between the different components of the CMS pixel detector are implemented in the database as hierarchical one-to-many relationships. Three different relation hierarchies are implemented, where each component is linked to its direct parent component in the respective hierarchy. The three relationship types correspond to the DAQ path and the auxiliary power and high voltage power systems. Utilizing these relation trees, it is for example possible to easily identify components affected by a malfunctioning component further upstream in a relation hierarchy.

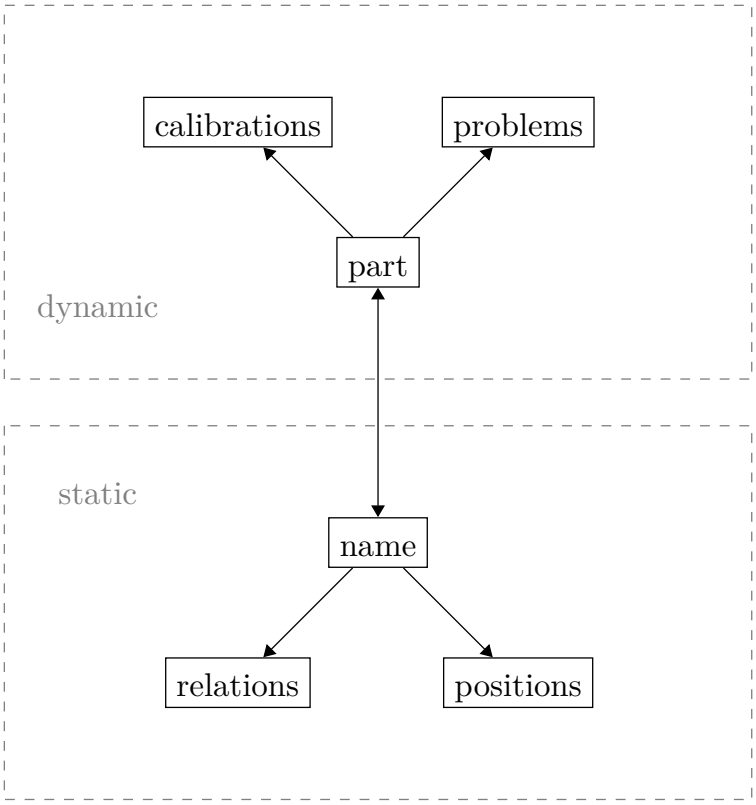

Figure 5.2.: An illustration of the concept of the detector map realized in the pixel monitoring database. The rectangles represent tables and the arrows indicate relations between objects within the database.

As part of the monitoring of the pixel detector, regular calibrations are performed by the tracker monitoring and pixel operations groups, to ensure an efficient operation of the detector. For each calibration, a new run is registered in the DAQ system and the measurements are written to a file system. In order to organize these results and allow quick access for further analysis, they have to be stored in the database. An illustration of the database layout implemented in the pixel monitoring database to hold the calibration data is shown in Fig. 5.3. For each calibration run an entry is created in the database and the type of calibration, as well as the detector configurations are stored. The calibration results are then read from file and stored in dedicated tables in the database and linked to the corresponding run and detector components. Using this layout, the list of stored calibration types can easily be expanded in the future by adding new tables that can store the results in the calibration specific format and reference the calibration runs and parts tables.

Similar to the calibration results, problems reported by the pixel operations group are recorded in the database with a link to the affected detector component. Problem

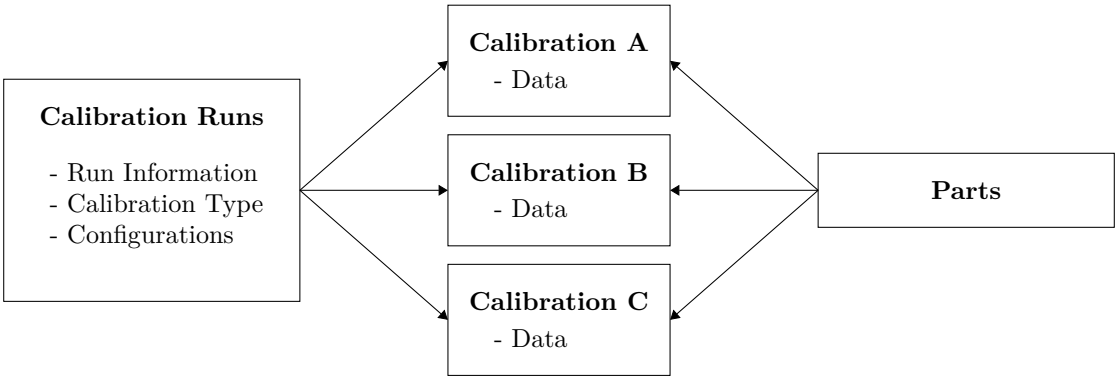

Figure 5.3.: A diagram illustrating the relationship of tables storing calibration results in the pixel monitoring database. The rectangles represent tables and the arrows indicate relations between objects within the database.

reports are entered and updated via an interface on the TOM. For each known problem, a short description, as well as an IOV is stored, starting when the problem first occurred and ending when it is resolved.

5.2. Online Monitoring Tools

Two platforms are used and maintained by the Tracker Online Monitoring group. The first one is the Tracker Online Monitor (TOM), which offers customizable overview plots for a variety of monitoring observables, as well as the possibility to interact with the monitoring database, by adding or updating entries. The second one, the Online Monitoring System [101] (OMS), is a platform developed for all CMS sub-systems to be used during the LHC run 3, which allows the visualization of information stored in databases. Both systems rely on data gathered by the Tracker Online Monitoring group that is stored in different databases.

The TOM is a web based application offering a number of monitoring tools that interact with various databases to provide a central platform for the Tracker Online Monitoring project. Detector overview plots for important observables, e.g. the temperature within the detector or the current drawn by the detector, are created using so called *probes*. These probes obtain the information from the database via SQL queries, and then process and convert the returned data into plots that can be displayed on the TOM. The probes can be customized via an web interface in the

TOM, where the time range or the parts of the detector for which the information is queried can be specified. Probes can be combined, which allows to view information from different databases within a single overview. Additionally, the TOM features an interface to the pixel monitoring database, where new problematic detector components can be reported and existing reports can be viewed and updated. The information is entered via a form on the TOM. It is then automatically processed by the TOM and inserted into the database, where a link to the affected detector part is established. Using the relations between detector components stored in the database, it is possible to quickly identify the components potentially affected by a malfunctioning component further upstream in the relation hierarchy.

The second monitoring tool is the OMS, which is a web based platform developed by the CMS collaboration as a monitoring platform for all CMS sub-systems starting in the run 3 of the LHC. The OMS is a successor of the Web Based Monitoring tool [102]. It features a two-layer design, with an aggregation layer and a presentation layer that communicate via a representational state transfer application programming interface (RESTful API). The aggregation layer is responsible to fetch data from multiple databases, using so called endpoints that obtain the data via specialized SQL queries. The presentation layer features customizable workspaces for all CMS sub-systems, where summary tables and plots can be created from data obtained from the aggregation layer.

The two systems offer very similar, yet complementary functionalities. The OMS is optimized for fast access to data and thus allows quick access to overview plots. However, the fast access time comes partly from stricter constrains on the SQL queries performed in the endpoints of the aggregation layer. The TOM on the other hand does not require these constrains, which allows to perform more complex queries and additional processing once the data is obtained from the database. Additionally, the TOM offers not only read access to data, but also an interface to update and insert new data into the monitoring database.

6. Event Reconstruction

The collision information recorded by the CMS detector are interpreted and converted into a physics object based event description in a process called event reconstruction. A particle flow (PF) reconstruction approach is chosen, where information from all sub-detector systems are combined to reconstruct and identify different final state particles. The list of PF particles is used as a basis for the physics objects used in this analysis.

In this chapter, the PF algorithm implemented by CMS is described, followed by a detailed definition of the specific physics objects used in the presented analysis.

6.1. Particle Flow

The event reconstruction from detector level information is performed using the PF [93] algorithm. Final state particles are reconstructed and identified by correlating information from all sub-detectors instead of relying on single systems to reconstruct specific particles. The high technical demands of a PF reconstruction approach on the detector, requiring fine spatial granularity in all subsystems, hermeticity of the calorimeter systems and dedicated muon detectors, are met in the CMS detector, which allowed the first successful application of a PF reconstruction at a hadron collider.

In the first step of the PF so called PF elements are reconstructed from each sub system. Particle trajectories (tracks) are reconstructed iteratively from hits in the tracking system and the muon chambers using a Kalman filtering based combinatorial track finding algorithm [103]. The track reconstruction starts with tight quality

criteria on the track seeds, the track fit χ^2, and compatibility of the track candidates with originating from a primary vertex. In each subsequent iteration, the previously used hits are removed from the input list of the track finder and the quality criteria are relaxed in favor of more complex and time-consuming seeding, filtering, and track finding algorithms, thus achieving a high efficiency while keeping the misidentification rate low.

Hits in the calorimeter cells are grouped into calorimeter clusters using a dedicated clustering algorithm specifically developed for PF reconstruction. The calorimeter clusters are build separately for each sub-detector, except for the HF calorimeters, where no clustering is performed. Clusters are seeded using cells with the energies above a given threshold that form a local energy maximum. Then, topological clusters are formed by iteratively adding neighboring cells with energy above twice the noise level to the cluster. Finally, the calorimeter clusters are reconstructed from topological clusters using an expectation-maximization algorithm, where a Gaussian-mixture model is used to determine the position and energy of each cluster.

In the next step, the above described PF elements from different sub-detectors are connected using a link algorithm to form PF blocks that are then used to reconstruct stable particles. A link between a track in the inner tracking system and a calorimeter cluster is established if the track can be extrapolated to within the area of the calorimeter cluster. In case of ambiguous links, only the link with the smallest distance is kept. Bremsstrahlung photons emitted by electrons are accounted for by linking ECAL clusters that coincide with tangents extrapolated from inner tracks at each tracker layer. Additional tracks originating from photon conversion within the tracker are identified and linked using a dedicated conversion finder [104]. Links between different calorimeter clusters are established based on the cluster position, again resolving ambiguities by choosing the link with the smallest distance. Multiple inner tracks are linked if they share a common interaction vertex to account for nuclear-interactions within the tracker material. Finally, inner tracks are linked to tracks in the muon systems, if their track parameters are compatible within uncertainty, or to single hits in the muon detectors, if the hit matches the extrapolated inner track.

Stable particles are identified and reconstructed iteratively in each block and the PF elements used are removed after each step. First, muons are reconstructed solely from inner tracks with a link to the muon systems, using additional PF elements to

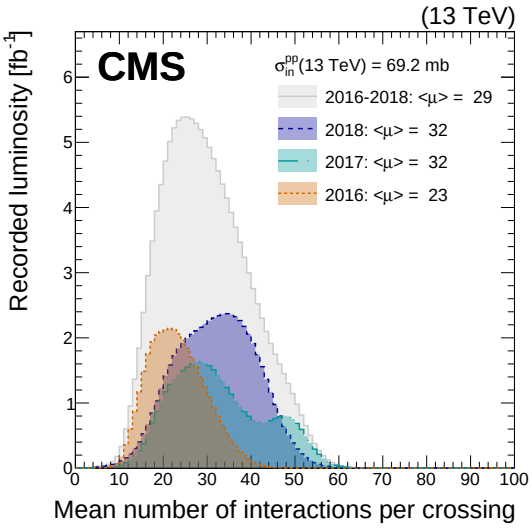

Figure 6.1.: Recorded luminosity by CMS as a function of the mean number of inelastic pp interactions per bunch crossing. The average number of inelastic pp interactions per bunch crossing is given in the legend. Taken from Ref. [106].

impose stringent identification criteria. Electrons and photons are reconstructed next. Electrons are identified by ECAL clusters with a link to an inner track, while photons are identified by trackless ECAL clusters. Other PF elements, like additional tracks and HCAL clusters are again used to determine variables to improve discrimination against other particles. Finally, neutral and charged hadrons are identified by HCAL clusters. Charged hadron identification requires a link to an inner track, while neutral hadrons use trackless HCAL clusters.

6.2. Primary Vertices

The primary vertex (PV) describes the position of an inelastic pp interaction. Due to the high instantaneous luminosity, it is very likely that multiple pp interactions occur alongside a hard scattering process during a single bunch crossing (pileup). A dedicated PV reconstruction is performed in CMS [105] that is robust against the high pileup conditions at the LHC. The recorded luminosity as a function mean number of inelastic pp interactions per bunch crossing during Run 2 is shown in Fig. 6.1. In this period, there was an average of 29 interactions per bunch crossing. The additional pileup interactions lead to the presence of multiple PV in each event.

The PV reconstruction utilizes tracks identified by the PF algorithm. Additional quality criteria are imposed on the tracks. The significance of the impact parameter relative to the beam spot is required to be less than 5. The track must have least 5 hits in the inner tracker, of which at least two must be located in the pixel detector, and the normalized χ^2 of the track fit must be less than 20. The selected tracks are then clustered using a deterministic annealing algorithm [107] to identify vertex candidates. Finally, an adaptive vertex fitter [108] is used to obtain the vertex parameters for each vertex candidate with more than one associated track. Each track is assigned a weight w_i reflecting the likelihood of the track originating from the vertex candidate. The number of degrees of freedom in the fit, given by

$$n_{\text{dof}} = -3 + 2 \sum_{i=1}^{\#\text{tracks}} w_i \,, \tag{6.1}$$

is connect to the number of tracks compatible with arising from the PV and is used as an indicator for true pp interactions.

All PVs considered in the following are required to have $n_{\text{dof}} \geq 4$ and be within $\sqrt{x^2 + y^2} < 2\,\text{cm}$ and $|z| < 24\,\text{cm}$ of the nominal interaction point. The PV with the largest sum of transverse momenta squared p_{T}^2 of the associated track-level physics objects is considered the main interaction vertex or leading vertex, while all remaining PV are considered as pileup vertices.

6.3. Electrons

The electron reconstruction in CMS [109] is fully integrated into the PF algorithm. Dedicated algorithms are used to account for radiative losses due to bremsstrahlung and subsequent photon conversion.

Electron candidates are built using ECAL clusters and inner tracks. The ECAL clusters are combined to superclusters, to include energy deposits from secondary particles created by interactions of the electron in the inner tracker. Starting from a seed cluster, ECAL clusters within a certain area in the $\eta - \phi$ plane are aggregated into the supercluster. Electron tracks are identified either starting from a supercluster, or from inner tracks. The supercluster-seeded method selects PF track seeds compatible

with the supercluster position extrapolated from the energy of the supercluster. This method has the best performance for high energetic, isolated electrons. The tracker-based approach is designed to recover reconstruction efficiency for non-isolated, soft, or forward electrons. This method selects PF tracks compatible with an ECAL cluster. All tracks selected in both methods are refitted using a Gaussian sum filter based fit to determine the electron track parameters [110], taking into account radiative losses due to bremsstrahlung. The refitted tracks are combined with compatible ECAL clusters or superclusters using the PF link algorithm.

Electrons are identified from the reconstructed candidates with additional identification (ID) criteria. In the following, a brief description of the electron identification variables and the cut-based IDs used in this analysis is given. A detailed description of the electron IDs is given in Ref. [109].

The aim of electron IDs is to select electrons produced at the main interaction vertex (prompt) and reject background sources, such as photon conversions, misidentified hadrons and secondary electrons from weak decays of b or c quarks. The variables used in the IDs describe the electron isolation, the shape of the electromagnetic shower in the ECAL, and the quality of the electron track and supercluster.

- The combined isolation

$$I_{\text{combined}} = I_{\text{ch}} + \max(0, I_{\text{n}} + I_{\gamma} - I_{\text{PU}}) \tag{6.2}$$

  of an electron is determined using the sum of p_{T} of charged hadrons (I_{ch}), photons (I_{γ}), and neutral hadrons (I_{n}) within a cone of $\Delta R = 0.3$ around the electron. The contribution of pileup is estimated in I_{PU} using the median energy per unit area ρ in the isolation cone. For the electron ID, the relative combined isolation $I_{\text{combined}}/E_{\text{T}}$ is used, where $E_{\text{T}} = \sqrt{m^2 + p_{\text{T}}^2}$ is the electron's transverse energy.

- The ratio between the energy deposited in the HCAL within a cone of $\Delta R = 0.15$ around the supercluster and the electron candidate energy, H/E, is expected to be small for prompt electrons.

- The distribution of the electromagnetic shower in η is taken into account in $\sigma_{i\eta i\eta}$. Electron signals result in small values of $\sigma_{i\eta i\eta}$.

- The compatibility of the supercluster and the electron track is measured using the difference between the energy of the supercluster and the track momentum $|1/E - 1/p|$, and the distance between track and supercluster $|\Delta\eta|$ and $|\Delta\phi|$.

- The number of missing hits in the innermost tracker layers are an indication for an electron arising from photon conversion. Electron candidates must have not more than one missing hit and pass a photon conversion veto.

Different working points (WP) of the electron ID are obtained, by varying the thresholds on each variable. The electron ID WPs are defined by the average signal efficiency. The loose and tight WPs are used in the presented analysis. They correspond to a signal efficiency of 90% and 70%, respectively.

In addition to the electron ID, all electrons considered in the analysis are required to have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$.

Differences in reconstruction and identification efficiencies between data and simulation are corrected using correction factors provided centrally by the CMS collaboration as functions of p_{T} and η of the electron.

6.4. Muons

Muons can pass through the calorimeters and leave the detector without substantial energy loss, due to their mass. Hence, the muon reconstruction in CMS [97] relies on the measurement of muon tracks.

Three types of muon candidates are reconstructed, depending on the sub-detector information used. *Tracker muons* are reconstructed from tracks in the inner tracker that can be extrapolated to at least one hit in the muon system. *Standalone muons* are built from muon tracks measured in the muon detectors. If the track of a standalone muon is compatible with a track in the inner tracker, a *global muon* is reconstructed from the combined track. Tracker muons sharing the same track with a global muon are merged into a single candidate.

All muon candidates have to pass additional ID criteria. The loose and tight WPs of the muon ID are used in this analysis. The only requirement of the loose WP is that the muon candidate must be reconstructed as either a tracker or a global muon. The tight WP requires the muon candidate to be reconstructed as both a tracker and a global muon, as well as a number of criteria, detailed in the following:

- The inner track must use hits from at least six layers, including at least one hit in the pixel detector.

- The extrapolation of the inner track must be compatible with at least two hits in the muon systems.

- The global track fit, using hits in the inner tracker and muon detectors, must have a goodness-of-fit per degrees of freedom $\chi^2/\text{dof} < 10$.

- The muon candidate must be compatible with arising from the main interaction vertex. This is determined from the impact parameters with respect to the leading vertex, i.e. by the transverse (d_0) and longitudinal (d_z) distances from the leading vertex, where $d_0 < 0.2\,\text{cm}$ and $d_z < 0.5\,\text{cm}$ is required.

The reconstruction and identification efficiency for muons is $> 99\%$ for the loose ID and between $95 - 99\%$ for the tight ID.

An isolation criterion is applied to suppress the selection of muons within jets as expected e.g. from weak decays. Similar to electrons, the relative muon isolation is defined as the sum of p_T of PF particles in a cone of $\Delta R < 0.4$ around the muon, relative to the muon p_T

$$I_\text{rel} = \frac{I_\text{ch} + \max(0, I_\text{n} + I_\gamma - 0.5 I_\text{PU,ch})}{p_\text{T}^\mu}\,, \tag{6.3}$$

where $I_\text{PU,ch}$ is the contribution of pileup estimated from charged hadrons originating from pileup vertices. The tight isolation WP is used in this analysis, where $I_\text{rel} < 0.15$. This corresponds to a selection efficiency of 95% for isolated muons.

Finally, all muons considered in the presented analysis are required to have $p_\text{T} > 30\,\text{GeV}$ and $|\eta| < 2.4$.

Differences in identification and isolation efficiencies between data and simulation are corrected using correction factors provided centrally by the CMS collaboration as a function of p_{T} and η of the muon.

6.5. Jets

The jet clustering in CMS is performed using the clustering algorithms implemented in the FastJet [72] software library with all PF particles as input. In the presented analysis, jets clustered with either the anti-k_{T} using a distance parameter of $R = 0.4$ (AK4 jets), or the HOTVR algorithm (HOTVR jets) are used. HOTVR jets are clustered using the default parameters for the top-tagging mode given in Ref. [7]. A detailed description of these algorithms is given in Chapter 3.

After applying the jet calibrations described in Section 6.5.2, additional requirements are placed on the jets. All AK4 jets considered in this analysis must fulfill the tight working point of the jet ID [111] and have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. HOTVR jets are required to have $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ and $|\eta| < 2.5$.

6.5.1. Pileup Mitigation

Pileup interactions from the same or neighboring bunch crossings deposit additional energy in the detector that influences the measurement of physics objects from the hard interaction process. Various techniques are used in CMS to mitigate the influence of pileup from the jet reconstruction. In the following, the charged-hadron subtraction (CHS) [112] and pileup per particle identification (PUPPI) [113] techniques used in this analysis are discussed. A detailed study and comparison of these techniques in CMS is given in Ref. [106].

The CHS algorithm identifies charged particles associated to pileup vertices by using tracking information. A charged particle is considered to originate from pileup, if the associated track was used in the pileup vertex fit. All charged particles originating from pileup vertices are removed from the event before the jet clustering. Neutral particles are unaffected by this procedure. Their contribution is mitigated using jet

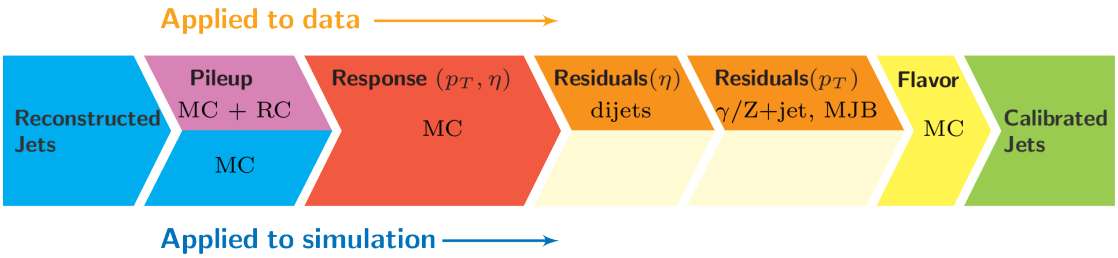

Figure 6.2.: A sketch of the consecutive stages of the jet energy correction in CMS, for data (top) and MC simulation (bottom). Taken from Ref. [116]

area based pileup subtraction [114, 115] after the jet clustering. These corrections are part of the jet calibration in CMS described in Section 6.5.2. In this analysis, the CHS algorithm is used to mitigate the effect of pileup on AK4 jets.

The PUPPI algorithm uses local particle distributions, event pileup properties and charged particle tracks to identify particles originating from pileup. Charged particles are identified similar to the CHS method. Then, a discriminator α_i is calculated for each neutral particle i in an event using the p_{T} and angular distance ΔR of nearby charged particles from the leading vertex. The resulting distribution of α is compared to the distribution calculated from the charged hadrons originating from pileup vertices in the event. The difference of α_i to the expected mean value is translated to a weight between 0 and 1, where particles originating from pileup are likely to obtain weights close to 0. Charged particles originating from pileup vertices are assigned a weight of 0, while charged particles from the main interaction vertex are assigned a weight of 1. The momentum of each PF particle is scaled using these weights. Particles with PUPPI weights < 0.01 are entirely removed from the event. In this analysis, the HOTVR jets are clustered using the PUPPI-weighted PF particles as input.

6.5.2. Jet Energy Calibration

In CMS, the jet energy calibration (JEC) is performed in a factorized approach [116, 117] to ensure agreement between true and measured jet energy scale (JES) and jet energy resolution (JER). A schematic overview of the JEC procedures in CMS is shown in Fig. 6.2.

In the first step, the L1 offset corrections, the effect of neutral pileup is mitigated

using the hybrid jet area method. A correction is obtained as a function of the reconstructed jet p_{T}, η, effective jet area and the average energy per unit area ρ in the event. The L1 offset corrections are not applied to jets clustered from the PUPPI-weighted PF particles, because the contribution from pileup is already removed on the individual particles. In the next step, the relative and absolute detector response is corrected with the L2 relative and L3 absolute corrections. These corrections are derived from simulation as a function of p_{T} and η of the jet. Residual detector response effects not covered by the simulation are addressed by the L2L3 residual corrections. The η-dependence of the relative detector response is measured in a sample of dijet events selected from data, where one jet is in the well described barrel region and the other jet is used to measure the detector response as a function of p_{T} and η. Similarly, the p_{T}-dependence of the absolute detector response is measured in γ+jets and $\mathrm{Z}(\rightarrow \mu\mu/\mathrm{ee})$+jets events, where a jet balancing the γ /Z is used to measure the detector response as a function of p_{T}. Additionally, any source of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ in those events is assumed to arise from miscalibrated jets and is considered in the residual corrections. The L2L3 residual corrections are only applied to jets in data. Due to different experimental conditions, the JEC are determined separately for the 2016, 2017 and 2018 datasets [117].

The above described corrections are centrally provided by the CMS collaboration for AK4 jets. The HOTVR subjets are calibrated, by applying the same corrections as for AK4 jets clustered with PUPPI-weighted PF particles. The contribution from pileup is mitigated using PUPPI and no further L1 offset corrections are applied. The L2L3 corrections are stable with respect to the jet size [116] and can thus be assumed for HOTVR subjets. The corrected HOTVR jet four-momentum is obtained as the sum of its corrected subjet four-momenta. The response of HOTVR jets is measured as the average ratio of the reconstructed p_{T} on detector level $p_{\mathrm{T}}^{\mathrm{rec}}$ and the simulated p_{T} on generator level $p_{\mathrm{T}}^{\mathrm{gen}}$ as a function of particle level jet p_{T}. It is shown in Fig. 6.3 for corrected and uncorrected HOTVR jets. The measurement was performed in simulated $\mathrm{t\bar{t}}$ events. Particle level jets were matched to detector level jets by selecting the particle level jet within $\Delta R < R_{\mathrm{eff}}$ with the smallest angular distance to the detector level jet, where R_{eff} is the effective distance parameter of the detector level jet. The response of calibrated HOTVR jets is found compatible with unity within $< 1\%$. The procedure was verified in data from a $\mathrm{t\bar{t}}$ enriched sample.

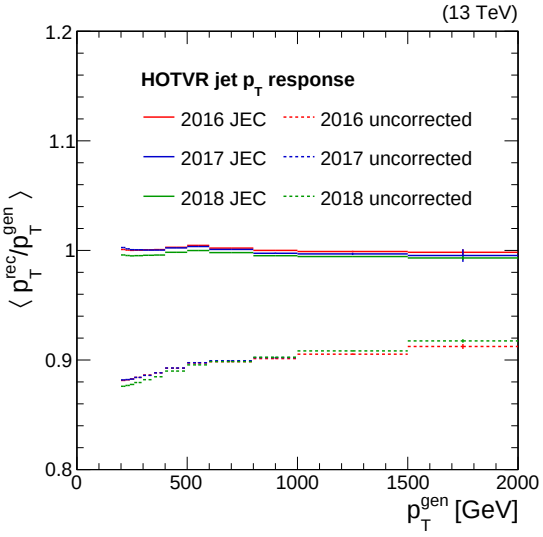

Figure 6.3.: Jet p_{T} response as a function of simulated jets on generator level p_{T} for HOTVR jets before (dashed) and after (solid) jet energy corrections, which are applied to the HOTVR subjets. The p_{T} response was measured using $\mathrm{t\bar{t}}$ MC simulation.

6.5.3. Identification of b Jets

Hadrons containing b quarks have a comparatively long lifetime of $\sim 1.5\,\mathrm{ps}$ due to the suppressed decay by the off diagonal CKM matrix elements, and are generally highly boosted, leading to displaced tracks a few mm from the PV. A secondary vertex can be reconstructed using the displaced tracks [118], which is an important input to distinguish jets originating from b quarks (b jets) from other jets.

In this thesis, the DeepJet [119] algorithm is used to identify b jets. The algorithm uses a deep neural network with input parameters from PF particles clustered into the jets and secondary vertices associated to the jet. Jets are classified as originating from b, c, or light quarks and gluons. The classifier output is a number between 0 and 1 for each class, related to the probability of originating from that respective class. Jets can be identified as b jets by imposing a requirement on the corresponding classifier output node.

The DeepJet classification is applied to all AK4 jets in the event. The medium WP is chosen to identify b jets in this analysis, which corresponds to a misidentification rate of 1% for light jets.

6.5.4. Identification of Boosted Top Quark Decays

The targeted final state of this analysis features a hadronically decaying t quark, which has a distinct three-pronged signature. Due to the high mass of the heavy resonance, the t quark is boosted in the laboratory rest frame and the three prongs of the t are collimated. At a sufficiently large Lorentz boost, it is possible to reconstruct the t decay within a single jet. Using jet substructure techniques it is possible to separate these t jets from a background from QCD multijet production.

Various t tagging algorithms are used in CMS for this purpose [82]. In the choice of a suitable tagging algorithm for this analysis, a focus was put on the stability of the tagging efficiency as a function of the p_{T} of the t candidate. Detailed studies of the tagging performance as a function of the t quark p_{T} are discussed in Chapter 8. As a result of these studies the HOTVR tagger has been chosen for this analysis, where a HOTVR jet is considered t-tagged, given the following criteria [7]:

- the fractional p_{T} of the leading subjet with respect to the jet $f_{p_{\mathrm{T}}} = \frac{p_{\mathrm{T}}^{\mathrm{s1}}}{p_{\mathrm{T}}^{\mathrm{jet}}} < 0.8$,
- the number of subjets $N_{\mathrm{sub}} \geq 3$,
- the jet mass $140 < m_{\mathrm{jet}} < 220\,\mathrm{GeV}$,
- the minimum mass of pairs of subjets $\min(m_{\mathrm{pair}}) = \min(\sqrt{(P_i + P_j)^2}) > 50\,\mathrm{GeV}$

Additionally, a requirement on the N-subjettiness ratio $\tau_3/\tau_2 < 0.56$ is introduced to further increase the discrimination power.

6.6. Missing Transverse Momentum

The initial p_{T} in pp collisions is negligible. Hence, a global p_{T} imbalance in the event is an indicator that energy escaped the detector undetected. This missing energy can be reconstructed using the missing transverse momentum [120], defined as the

negative vectorial sum of p_{T} of all PF particles,

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_i \vec{p}_{\mathrm{T},i}\,. \tag{6.4}$$

The magnitude of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$. The only SM particle that can escape the CMS detector completely undetected is the neutrino. Therefore $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is sensitive to the presence of a neutrino in the event and can be used to reconstruct the neutrino four-momentum given kinematic constraints. Other sources of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ are detector noise and miscalibrated objects. The latter is accounted for by propagating the jet calibration applied to the AK4 jet collection to the calculation of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. Events with high values of $p_{\mathrm{T}}^{\mathrm{miss}}$ from failed reconstruction or malfunctioning detector parts are removed using dedicated event filters [120].

7. Datasets and Simulated Samples

The analysis uses pp collision data recorded with the CMS detector at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ in the years 2016, 2017 and 2018. The total dataset corresponds to an integrated luminosity of $138\,\text{fb}^{-1}$ ($35.9\,\text{fb}^{-1}$ [121] (2016) + $41.5\,\text{fb}^{-1}$ [122] (2017) + $59.7\,\text{fb}^{-1}$ [123] (2018)). Only data taking periods certified for physics analyses are used, ensuring that all triggers, the event reconstruction and all sub-detectors show expected performance.

Since the targeted final state contains an isolated charged lepton, the single electron and single muon primary datasets are used. These datasets include events recorded by HLT triggers, which are seeded by single electron or single muon L1 triggers. For the years 2016 and 2017, the single photon primary dataset is used in addition to the single electron dataset to recover events due to trigger inefficiencies for high-p_T electrons. In 2018, the single electron and single photon primary datasets are merged into a single dataset. The trigger requirements for the used triggers in the three years are summarized in Table 7.1.

Simulations are used for the estimation of SM backgrounds and to derive the

Year	**e+jets**	**μ+jets**
2016	$p_{\text{T,e}} > 27\,\text{GeV} \vee E_\gamma > 175\,\text{GeV}$	$p_{\text{T},\mu} > 24\,\text{GeV}$
2017	$p_{\text{T,e}} > 35\,\text{GeV} \vee E_\gamma > 200\,\text{GeV}$	$p_{\text{T},\mu} > 27\,\text{GeV}$
2018	$p_{\text{T,e}} > 32\,\text{GeV} \vee E_\gamma > 200\,\text{GeV}$	$p_{\text{T},\mu} > 24\,\text{GeV}$

Table 7.1.: Trigger requirements for the 2016, 2017 and 2018 datasets. The first column indicates the dataset. The second and third columns give the trigger requirements for the e+jets and μ+jets final states, respectively. Here, $p_{\text{T,e}}$ and $p_{\text{T},\mu}$ denote the transverse momentum of the e- and μ candidate, respectively, and E_γ denotes the energy of the photon candidate.

signal sensitivity. The above described triggers are emulated in the simulation. Differences between the trigger selection efficiencies in data and simulation are corrected using scale factors. Muon trigger scale factors are provided centrally by the CMS Collaboration [124]. Electron trigger scale factors are measured using a tag-and-probe method in a dileptonic $t\bar{t}$ enriched control region. The measurement is described in Ref. [125].

The signal samples of singly produced b^* quarks are simulated at LO using the MadGraph5_aMC@NLO event generator. The five-flavour scheme is used in the simulation of the initial state, i.e. using b quark PDFs instead of simulating the preceding gluon splitting $g \to b\bar{b}$ explicitly. Only the decay mode $b^* \to tW$ is considered. Mass states of the b^* quark ranging from 0.7 to 4.2 TeV are simulated for the LH and RH parameter sets. The VL b^* samples are obtained as the sum of the respective LH and RH samples. A summary of the b^* signal samples is given in Table 7.2. The production cross sections in that table refer to the LH and RH parameter sets. For the VL case, the production cross section is twice as large.

The simulation of singly produced vector-like B quark signal samples are performed with MadGraph5_aMC@NLO at LO. Separate samples for the B VLQ production in association with a b or t quark are simulated. For each production mode, mass states ranging from 0.7 to 1.8 TeV are simulated. The simulated B samples are summarized in Table 7.3. Two different theory predictions for the B+b production cross section are considered. In the first case, a model with a VLQ doublet is assumed. The second case assumes a model with VLQ singlet. For the B+t production, both models result in the same cross section.

The SM production of $t\bar{t}$ in the ℓ+jets final state makes up the dominant background in the presented search. The other $t\bar{t}$ final states are considered as well, but their contributions are comparatively small. Samples of SM $t\bar{t}$ production are simulated using POWHEG v2 at NLO. For the 2016 dataset, a single sample inclusive in the $t\bar{t}$ final state is used, while for the 2017 and 2018 datasets separate samples for each final state are used. The cross section is adjusted to NNLO accuracy in perturbative QCD, including resummation of next-to-next-to-leading-logarithmic soft gluon terms, obtained with the Top++ 2.0 program [126]. The t quark p_T spectrum is found to be different between data and simulation [127, 128]. To account for this mismodelling, many analyses by the CMS Collaboration adapted a correction scheme known as t-p_T reweighting. In this analysis a correction based on that scheme is applied to the

M_{b^*} [GeV]	σ [pb]
$\mathrm{pp} \to \mathrm{b}^* \to \mathrm{tW}$ (cf. Fig. 2.5)	
700	32.8
800	17.1
900	9.46
1000	5.93
1100	3.23
1200	1.98
1400	0.81
1600	0.35
1800	0.17
2000	0.081
2200	0.041
2400	0.022
2600	0.012
2800	$6.7 \cdot 10^{-3}$
3000	$3.8 \cdot 10^{-3}$
3200	$1.9 \cdot 10^{-3}$
3400	$1.1 \cdot 10^{-3}$
3600	$0.62 \cdot 10^{-3}$
3800	$0.36 \cdot 10^{-3}$
4000	$0.22 \cdot 10^{-3}$
4200	$0.13 \cdot 10^{-3}$

Table 7.2.: Summary of the simulated b^* signal samples. For each sample, the b^* quark mass is shown in the first column, and the production cross section is given in the second column. All samples were simulated using the MadGraph5_aMC@NLO event generator.

M_{B} [GeV]	σ doublet [pb]	σ singlet [pb]
$\mathrm{pp} \to \mathrm{B} + \mathrm{b} + \mathrm{q}$, $\mathrm{B} \to \mathrm{tW}$ (cf. Fig. 2.6 left)		
700	0.32	0.068
800	0.16	0.035
900	0.086	0.020
1000	0.049	0.011
1100	0.029	0.0067
1200	0.018	0.0042
1300	0.011	0.0026
1400	0.0071	0.0017
1700	0.0021	0.000 51
1800	0.0014	0.000 35
$\mathrm{pp} \to \mathrm{B} + \mathrm{t} + \mathrm{q}$, $\mathrm{B} \to \mathrm{tW}$ (cf. Fig. 2.6 right)		
700	0.015	0.015
800	0.0081	0.0081
900	0.0046	0.0046
1000	0.0028	0.0028
1100	0.0017	0.0017
1200	0.0011	0.0011
1300	0.000 69	0.000 69
1400	0.000 46	0.000 46
1500	0.000 30	0.000 30
1600	0.000 21	0.000 21
1700	0.000 14	0.000 14
1800	0.000 099	0.000 099

Table 7.3.: Summary of the simulated B signal samples. For each sample, the B mass is shown in the first column, and the production cross sections for the doublet and singlet models are given in the second and third column, respectively. All samples were simulated using the MadGraph5_aMC@NLO event generator. The top part of the table lists the B samples with an associated b quark, the bottom half with an associated t quark.

t quark p_{T} in simulation following the procedure described in Ref. [2]. Scale factors are calculated for each event using

$$SF(p_{\mathrm{T}}) = e^{\alpha - \beta p_{\mathrm{T}}} \,. \tag{7.1}$$

The parameters are initialized to $\alpha = 0.0615$ and $\beta = 0.0005\,\mathrm{GeV}^{-1}$ and the appropriate factor will be determined later in a fit to data. A weight is obtained as

$$\omega = \sqrt{SF(\mathrm{t})SF(\bar{\mathrm{t}})}\,, \tag{7.2}$$

where $SF(\mathrm{t})$ and $SF(\bar{\mathrm{t}})$ are the scale factors obtained using the p_{T} of the generator level t and $\bar{\mathrm{t}}$ quark, respectively.

The SM single production of t quarks in association with W bosons constitutes an irreducible background, because it results in the same signature as the signal. Single t production in the t and s channels are also considered, however, their contributions are comparatively small. Separate samples for the three production modes are simulated. The tW- and t-channel samples are generated with POWHEG v2 at NLO and the s-channel samples are produced with MadGraph5_aMC@NLO at LO. The cross sections of the t and s channels are adjusted to NLO predictions obtained with HATHOR [129]. The tW-channel cross section is adjusted to NNLO approximate calculations taken from Refs. [130, 131].

Finally, samples for the production of W and Z bosons in association with jets (W/Z+jets) and diboson (WW, WZ, ZZ) production are generated to be used in the data-driven background estimation. The W/Z+jets samples are simulated with MadGraph5 at LO and diboson samples are simulated with PYTHIA 8.2 at LO. The boson p_{T} spectrum in the W/Z+jets is corrected to account for NLO QCD and electroweak contributions according to Ref. [132]. A summary of the background samples used in the analysis is given in Table 7.4.

Because of changing experimental conditions between the years, data and simulation are categorized by year and dedicated corrections are applied separately. The separate distributions from all three years are combined to derive the final result. The most severe change on the detector side was the upgrade of the pixel detector between 2016 and 2017, described in Section 4.2.2. Also smaller changes in the sub-detector calibrations and alignment, as well as damaged or degrading detector components

Process	**Generator**	σ [pb]
$t\bar{t}$, inclusive *	POWHEG	831.76
$t\bar{t}$, all-hadronic †	POWHEG	380.11
$t\bar{t}$, ℓ+jets †	POWHEG	364.31
$t\bar{t}$, dilepton †	POWHEG	87.33
single t (tW), t, leptonic	POWHEG	19.57
single t (tW), $\bar{t}$, leptonic	POWHEG	19.57
single t (t-ch), t	POWHEG	136.02
single t (t-ch), $\bar{t}$	POWHEG	80.95
single t (s-ch), leptonic decays	MadGraph5_aMC@NLO	3.36
W+jets ($\rightarrow \ell\nu$) $100 < H_T < 200$	MadGraph5_aMC@NLO	1363
W+jets ($\rightarrow \ell\nu$) $200 < H_T < 400$	MadGraph5_aMC@NLO	416
W+jets ($\rightarrow \ell\nu$) $400 < H_T < 600$	MadGraph5_aMC@NLO	58.88
W+jets ($\rightarrow \ell\nu$) $600 < H_T < 800$	MadGraph5_aMC@NLO	12.99
W+jets ($\rightarrow \ell\nu$) $800 < H_T < 1200$	MadGraph5_aMC@NLO	5.467
W+jets ($\rightarrow \ell\nu$) $1200 < H_T < 2500$	MadGraph5_aMC@NLO	1.107
W+jets ($\rightarrow \ell\nu$) $2500 < H_T < \text{Inf}$	MadGraph5_aMC@NLO	$7.8 \cdot 10^{-3}$
Z+jets ($\rightarrow \ell\ell$) $100 < H_T < 200$	MadGraph5_aMC@NLO	161.50
Z+jets ($\rightarrow \ell\ell$) $200 < H_T < 400$	MadGraph5_aMC@NLO	48.61
Z+jets ($\rightarrow \ell\ell$) $400 < H_T < 600$	MadGraph5_aMC@NLO	6.89
Z+jets ($\rightarrow \ell\ell$) $600 < H_T < 800$	MadGraph5_aMC@NLO	1.77
Z+jets ($\rightarrow \ell\ell$) $800 < H_T < 1200$	MadGraph5_aMC@NLO	0.81
Z+jets ($\rightarrow \ell\ell$) $1200 < H_T < 2500$	MadGraph5_aMC@NLO	0.19
Z+jets ($\rightarrow \ell\ell$) $2500 < H_T < \text{Inf}$	MadGraph5_aMC@NLO	$3.5 \cdot 10^{-3}$
WW	PYTHIA	75.88
WZ	PYTHIA	27.54
ZZ	PYTHIA	12.16

Table 7.4.: Summary of the SM background samples used. The simulated process is given in the first column. The MC event generator used for the simulation of the hard interaction process is given in the second column, and cross section is given in the third column. The samples marked with an $*$ are only used for the year 2016, and samples marked with a † are only used in the years 2017 and 2018.

affected the detector performance differently in each year. Additionally, the LHC beam conditions varied for each data taking period. While for example the number of simultaneous interactions per beam crossing was on average 23 in 2016, it increased to 32 in the years 2017 and 2018 (see Fig. 6.1), increasing the instantaneous luminosity, but also the pileup conditions.

In the years 2016 and 2017 a timing shift of some ECAL cells was not properly propagated to the L1 trigger, leading to a wrong association of trigger objects to the previous bunch crossing. Due to a L1 trigger rule where no two consecutive bunch crossings can fire the trigger, events with a significant amount of energy in the ECAL in the affected region $2.0 < \eta < 3.0$ can be falsely rejected. To account for this effect in simulation, event weights based on the probability of such a trigger pre-firing occurence are applied [133]. The effect of the trigger pre-firing on the presented analysis was checked and found to be small.

During the data-taking of the 2018 dataset, two modules in the HCAL end cap failed. The missing HCAL measurements in the affected region cause a miscalibration of jets and an increased rate of misidentified electrons within the affected area enclosed in $-3.0 < \eta < -1.3$ and $-1.57 < \phi < -0.87$. Events having jets or electrons within this region are rejected in the affected data-taking period. Simulated events with jets or electrons within the region are weighted, to match the expected yield of the unaffected dataset.

For the 2016 simulations, the underlying event tune CUETP8M1 [38] is used for all samples except the $\mathrm{t\bar{t}}$ samples, where the CUETP8M2T4 [134] tune is used. The PDFs are taken from the NNPDF3.0 [135] sets. In the simulation of the 2017 and 2018 samples, the CP5 [38] tune and the NNPDF3.1 [25] PDFs are used. The pileup distribution in simulation is corrected using a minimum bias cross section of 69.2 mb [136] to match the distribution observed in data. The simulated events are weighted using

$$\omega_{\mathcal{L}} = \frac{N_{\mathrm{sim}}}{\sigma_i \mathcal{L}_{\mathrm{int}}}\,, \tag{7.3}$$

to obtain the expected number of events in a dataset with integrated luminosity $\mathcal{L}_{\mathrm{int}}$. Here, N_{sim} is the number of simulated events and σ_i is the production cross section of the simulated process.

8. Efficiency Studies of the HOTVR Algorithm

In the context of the presented analysis, the t tagging performance of HOTVR is studied. For this, the top tagging efficiency is measured in a simulation of b^* signal events and compared to the misidentification rate, measured in simulated QCD multijet events.

The performance of HOTVR in t tagging mode is compared to a t tagging method commonly used in CMS, using anti-k_T jets with radius $R = 0.8$ (AK8 jets) and soft drop mass m_{SD}. The parameters of the soft drop algorithm are set to $z_{cut} = 0.1$ and $\beta = 0$. The soft drop based tagger requires $110\,\text{GeV} < m_{SD} < 220\,\text{GeV}$ and a requirement on τ_3/τ_2 based on the given working point. Both jet collections were clustered using the PUPPI weighted PF particles.

Events with at least one jet with $p_T > 200\,\text{GeV}$ and $|\eta| < 2.5$ are selected. In the signal sample, hadronically decaying t quarks with $p_T < 200\,\text{GeV}$ are identified on particle level and matched to detector level jets. A jet is considered matched, when the angular distance ΔR between the jet axis and the t quark is $\Delta R < 0.4$. In the QCD multijet simulation, particle level jets with $p_T > 30\,\text{GeV}$, clustered with the anit-k_T algorithm with $R = 0.4$, are matched to the particle level jets, using $\Delta R < 0.4$ between the jet axes. The selection is the same for HOTVR and AK8 jets.

The signal efficiency as measured in the b^* sample is defined as

$$\varepsilon_S = \frac{N_{\text{matched}}}{N_{\text{t,had.}}}\,, \tag{8.1}$$

where $N_{t_{had}}$ is the total number of hadronically decaying t quarks in the sample

and N_{matched} is the number of t tagged jets matched to the particle level t quark. Similarly, the misidentification rate is measured in a sample of simulated QCD multijet events as

$$\varepsilon_{\text{B}} = \frac{N_{\text{matched}}}{N_{\text{gen. jet}}}\,, \tag{8.2}$$

where $N_{\text{gen. jet}}$ is the number of particle level jets and N_{matched} is the number of t tagged jets matched to a particle level jet.

The τ_3/τ_2 criterion in both taggers is scanned from 1 to 0 and for each point, the signal and background efficiency is calculated. The efficiencies are measured in six different kinematic regions, depending on the p_{T} of the matched particle level object, i.e. the t quark (ε_{S}) and the particle level jet (ε_{B}). The results of the scan are shown as Receiver Operating Characteristics (ROC) curves in figure 8.1 for the six different regions. The HOTVR algorithm (red) shows a significantly better performance in the first region ($200\,\text{GeV} < p_{\text{T}} < 400\,\text{GeV}$). This is due to the variable radius parameter of HOTVR compared to the fixed radius parameter of the AK8 jets. In this region, the decay products of the t quark decay are less collimated, as shown in Fig. 8.2. Here, the maximum angular distance between the three decay products of a hadronic t quark decay $(\text{b}, \text{q}, \overline{\text{q}}')$ ΔR_{max} is shown as a function of the t quark p_{T}. A jet with a radius parameter smaller than ΔR_{max} can not reconstruct the full decay. The median value of ΔR_{max} as a function of p_{T} is shown as a solid orange line and the interval containing 68% of the distribution around the median is indicated by the area enclosed in the dashed orange lines. For comparison, the function $600\,\text{GeV}/p_{\text{T}}$ is drawn as a dotted purple line. The median of ΔR_{max} in this region is ~ 1.5. The HOTVR jets obtain large radus parameters of $R_{\text{eff}} \approx R_{\text{max}}$ in this region and are thus able to reconstruct the less collimated t quark decay more consistently than the AK8 jets with fixed radius parameter. In the next three regions, covering $400\,\text{GeV} < p_{\text{T}} < 1000\,\text{GeV}$ in steps of 200 GeV, the performance of both taggers is very comparable. In this region ΔR_{max} is closer to the radius paramter of the AK8 jets, making this an approriate choice for this kinematic region. HOTVR jets obtain an effective radius parameter of a similar order as the AK8 radius parameter. In the final two regions, $1000\,\text{GeV} < p_{\text{T}} < 1500\,\text{GeV}$ and $1500\,\text{GeV} < p_{\text{T}} < 2000\,\text{GeV}$, the performance of the soft drop based tagger starts to slightly deteriorate, while HOTVR maintains a stable performance. Here, the median ΔR_{max} is almost half the size of the AK8 radius parameter. Additional radiation and non-pertubative effects due to the larger jet size compared to the t decay increase the measured jet mass of the AK8 jets, leading to a smaller signal efficiency. The HOTVR jets on the other

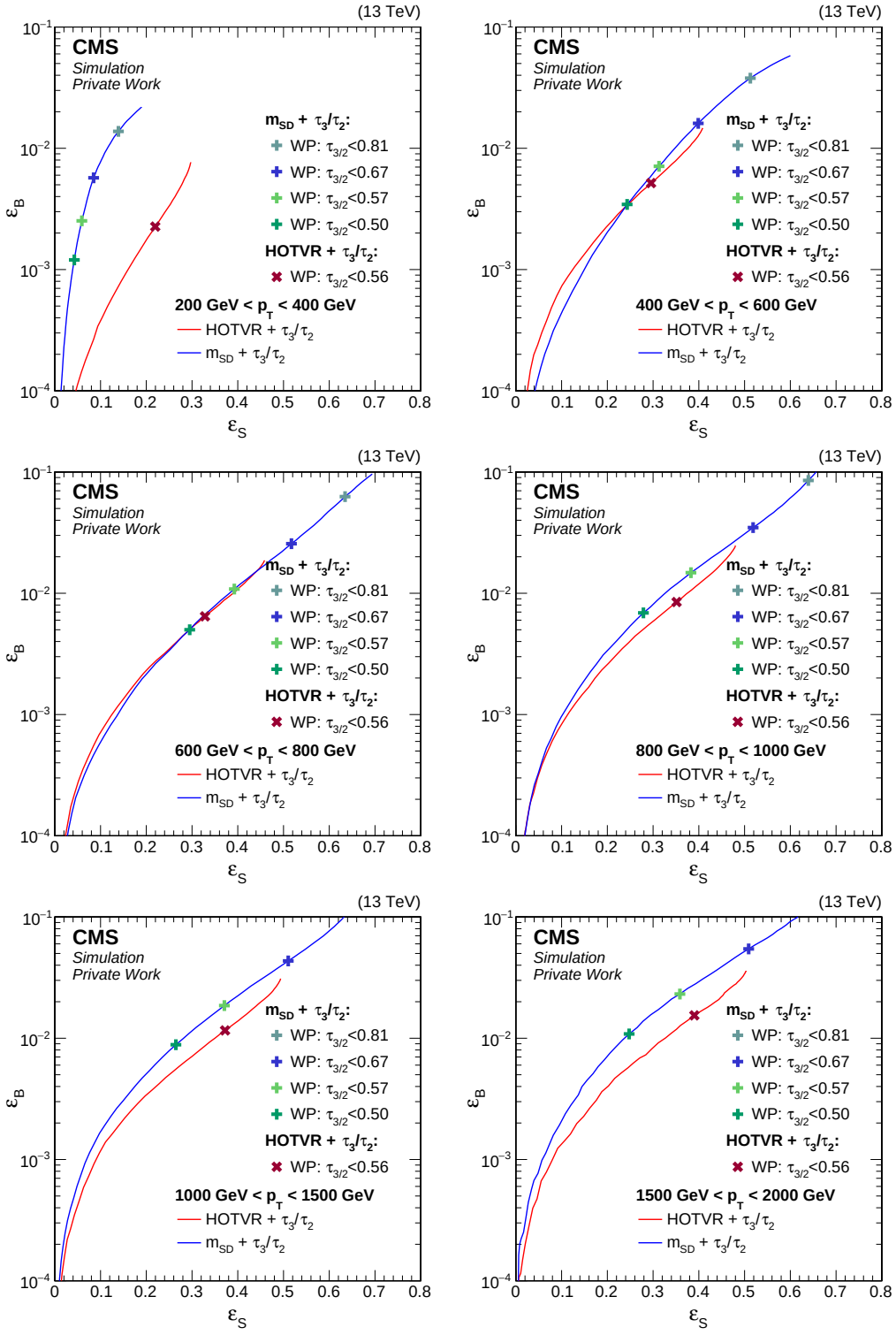

Figure 8.1.: ROC curves showing the t tagging performance with HOTVR (red) and a soft drop mass based tagger (blue) obtained by a scan of τ_3/τ_2 in different p_T regions. Commonly used working points (WP) are shown as markers. The ROC curves are measured for six different regions of p_T of the matched particle level object.

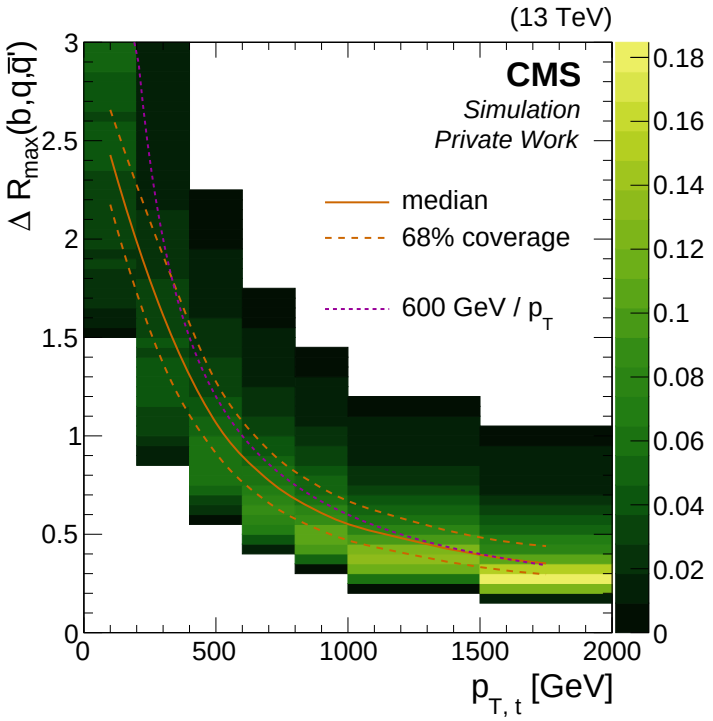

Figure 8.2.: Distribution of the maximum angular distance in the $(\eta - \phi)$-plane ΔR_{max} between the decay products of a hadronic t decay as a function of the t quark p_{T}. Each slice in p_{T} is normalized to 1. The median as a function of p_{T} is drawn as a solid line and the 68% interval around the median is indicated by dashed lines. Additionally, the function $600\,\text{GeV}/p_{\text{T}}$ is drawn as a dotted line.

side obtain an effective radius parameter close to the median value of ΔR_{max} and can thus maintain the signal efficiency.

The efficiency measurements in Fig. 8.1 demonstrate the power of the t tagging approach utilizing variable R jets, as implemented in the HOTVR. The increased sensitivity towards lower transverse momenta allows to cover a wider range of signal masses with a single approach, while maintaining a similar signal efficiency to comparable t tagging methods with fixed radius parameter. This makes the HOTVR tagger a very suitable choice for the analysis presented in the following chapter.

9. Search for a Heavy Resonances in the tW Decay Mode

In this chapter, a search for a heavy resonance decaying into a t quark and a W boson is presented. The described analysis is performed for a final state with an isolated electron or muon, missing transverse momentum, and jets (ℓ+jets). The analysis targets the $W \to \ell\nu$ and $t \to bW \to bq\overline{q}'$ decay mode. The event selection is introduced in Section 9.1, followed by a description of the reconstruction of the invariant tW mass and the final event categorization in Sections 9.2 and 9.3. The estimation of the expected backgrounds is detailed in Section 9.4. The considered sources of systematic uncertainties are summarized in Section 9.5, followed by the statistical interpretation of the analysis for a b^* signal in Section 9.6. The presented analysis is combined with an analysis targeting the all-hadronic final state [2]. The statistical combination of the two analyses is described in Section 9.7. Finally, an interpretation of the results in the context of models with a vector-like B quark is presented in Section 9.8.

The analysis results of the ℓ+jets and the combination with the all-hadronic analysis was published in Ref. [1].

9.1. Event Selection

An event selection is applied, to reduce the number of SM background events and to enhance the signal purity. The decay of $b^* \to tW$, where $W \to \ell\nu$ and $t \to bW \to bq\overline{q}'$ is targeted. Events are selected that match the general signal event signature. A sketch of the expected signature is shown in Fig. 9.1. Due

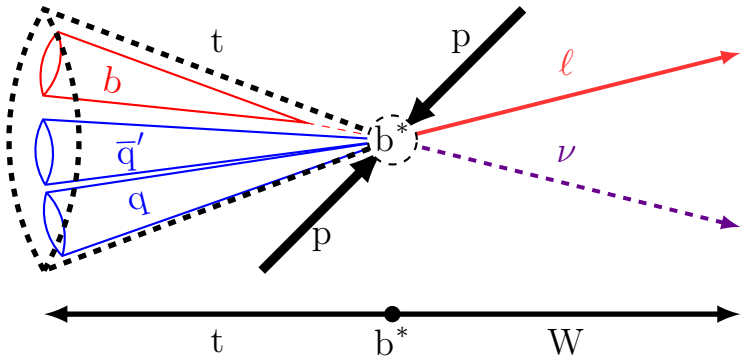

Figure 9.1.: Sketch of the expected event signature of a $pp \to b^* \to tW$ event. The t and W are back-to-back in the laboratory system and receive a large Lorentz boost. The t quark decay is very collimated and can thus be reconstructed into a single jet, indicated by the dashed cone. The neutrino from the leptonic W boson decay escapes the detector undetected and results in $\vec{p}_T^{\,miss}$.

to the high mass of the b^* quark, it is produced nearly at rest. The two decay products, the t quark and the W boson, receive a substantial Lorentz boost and are back-to-back in the laboratory system. The subsequent decay of the W boson results in an isolated lepton and missing transverse momentum $\vec{p}_T^{\,miss}$ from the neutrino escaping the detector. Due to the boost of the W boson, the lepton and $\vec{p}_T^{\,miss}$ are close together in ϕ. The boosted decay of the t quark results in a single jet with high p_T that can be identified using t tagging methods.

An example of the targeted event signature is shown in the event display in Fig. 9.2, taken from the b^* simulation. A high energy jet, arising from the boosted t quark decay, is seen on one side of the detector and opposite of that jet is an isolated muon from the leptonic W decay. The $\vec{p}_T^{\,miss}$ points in the same direction in ϕ as the muon.

The event selection starts with the events selected by the triggers described in Chapter 7. On top of the trigger selection, stricter requirements on the lepton candidates are imposed after full event reconstruction. Events with exactly one isolated lepton with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.4$ fulfilling the tight ID criterion are selected. Events with additional leptons with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ fulfilling at least the loose ID are discarded. In addition to fulfilling the lepton isolation criteria, described in Chapter 6, the lepton is required not to overlap with any AK4 jet by imposing $\Delta R(\ell, \text{jet}) > 0.4$. This criterion enhances the isolation criterion especially

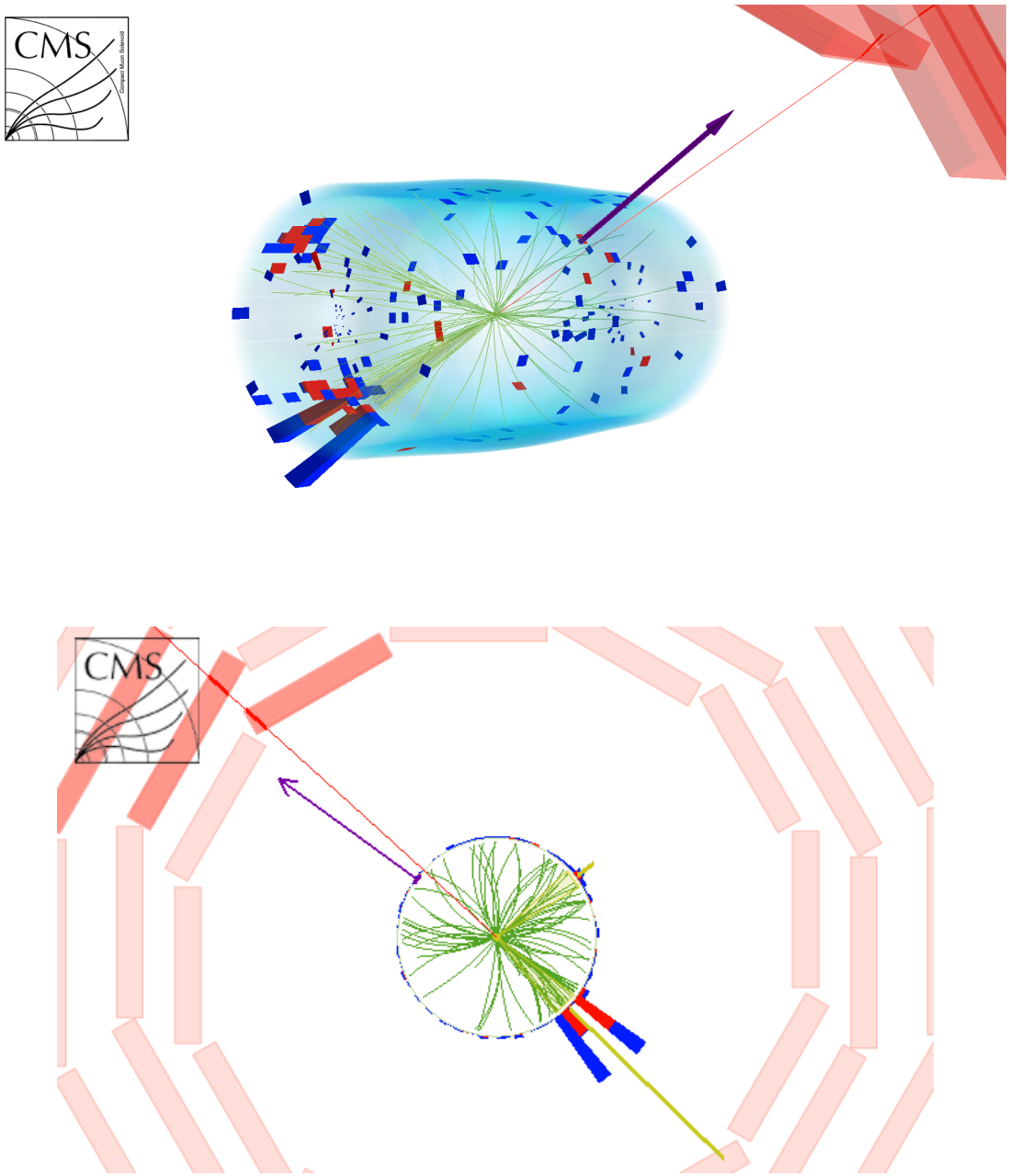

Figure 9.2.: Event display of a simulated $pp \to b^* \to tW$ event, where $W \to \mu\nu$ and $t \to bW \to bq\bar{q}'$. The upper image shows a 3D view of the event and the lower image shows the projection of the same event along the beam axis. The green lines show reconstructed inner tracks and the blue and red towers indicate energy deposits in the ECAL and HCAL, respectively. The height of the towers corresponds to the amount of energy deposited. The red line shows the muon trajectory and the light red areas indicate hits in the muon system. The purple arrow points in the direction of $\vec{p}_T^{\,miss}$ in ϕ and its length corresponds to the magnitude of $\vec{p}_T^{\,miss}$. The light blue area in the upper image shows the location of the CMS ECAL.

for electrons and suppresses the contribution from QCD multijet events. The presence of a neutrino in the expected final state is accounted for by requiring $p_{\mathrm{T}}^{\mathrm{miss}} > 50\,\mathrm{GeV}$. Additionally, since the neutrino and the charged lepton originate both from the boosted decay of the W boson, $\Delta\phi(\ell, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < \pi/2$ is required. Two global event observables, H_{T} and S_{T}, are used to further suppress SM backgrounds. The H_{T} observable is defined as the scalar p_{T} sum of all AK4 jets in the event $H_{\mathrm{T}} = \sum p_{\mathrm{T,jet}}$. It gives a measure of the hadronic activity in the event. For the signal, H_{T} peaks at roughly half the b^* quark mass M_{b^*}, while the spectrum obtained for SM processes falls with increasing H_{T}. A requirement of $H_{\mathrm{T}} > 200\,\mathrm{GeV}$ is imposed. Similarly, S_{T} is defined as the sum of H_{T}, the lepton p_{T} and $p_{\mathrm{T}}^{\mathrm{miss}}$. It is a measure for the overall activity in the event and results in a peak at roughly the b^* quark mass M_{b^*} for the signal, while falling exponentially for background events. A requirement of $S_{\mathrm{T}} > 400\,\mathrm{GeV}$ is imposed. Finally, exactly one HOTVR jet with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ and $|\eta| < 2.5$ fulfilling the HOTVR t-tagging criteria is required, reflecting the presence of a boosted hadronically decaying t quark in the event.

The selection efficiency for signal and background is illustrated in the upper plot in Fig. 9.3 for each step of the event selection. The efficiency is defined as the fraction of events passing a selection step with respect to the dataset obtained from the trigger selection. Events from processes without a t quark in the final state, i.e. W/Z+jets and diboson production, are summarized into non-top backgrounds. The signal efficiency is shown for three exemplary b^* signal samples with $m_{\mathrm{b}^*} = 0.7\,\mathrm{TeV}$, $2.4\,\mathrm{TeV}$ and $4.0\,\mathrm{TeV}$. The drop in signal efficiency after the requirement of one isolated lepton comes mainly from the exclusion of the full leptonic decay mode ($\mathrm{b}^* \to \mathrm{tW} \to \mathrm{b}\ell\nu\ell\nu$), which amounts to about 30% of the events after the trigger selection. The largest impact on the signal efficiency comes from the requirement of a t-tagged HOTVR jet. This is due to the signal efficiency of the t tagging algorithm of about 40%, as well as the rejection of events with the $\mathrm{t} \to \mathrm{b}\ell\nu$ final state. The lower plot of Fig. 9.3 shows the ratio of the total number of signal and background events after each step of the event selection. With each selection, the signal over background ratio improves. Here, it can be seen more clearly that the t tag requirement greatly improves the signal over background ratio.

Distributions of important observables after the event selection are shown in Figs. 9.4 and 9.5. The data-to-background ratio is shown in the bottom panel of each plot. An overall good agreement is seen in all distributions. Even though the non-top backgrounds are strongly suppressed by the event selection, a large fraction of the

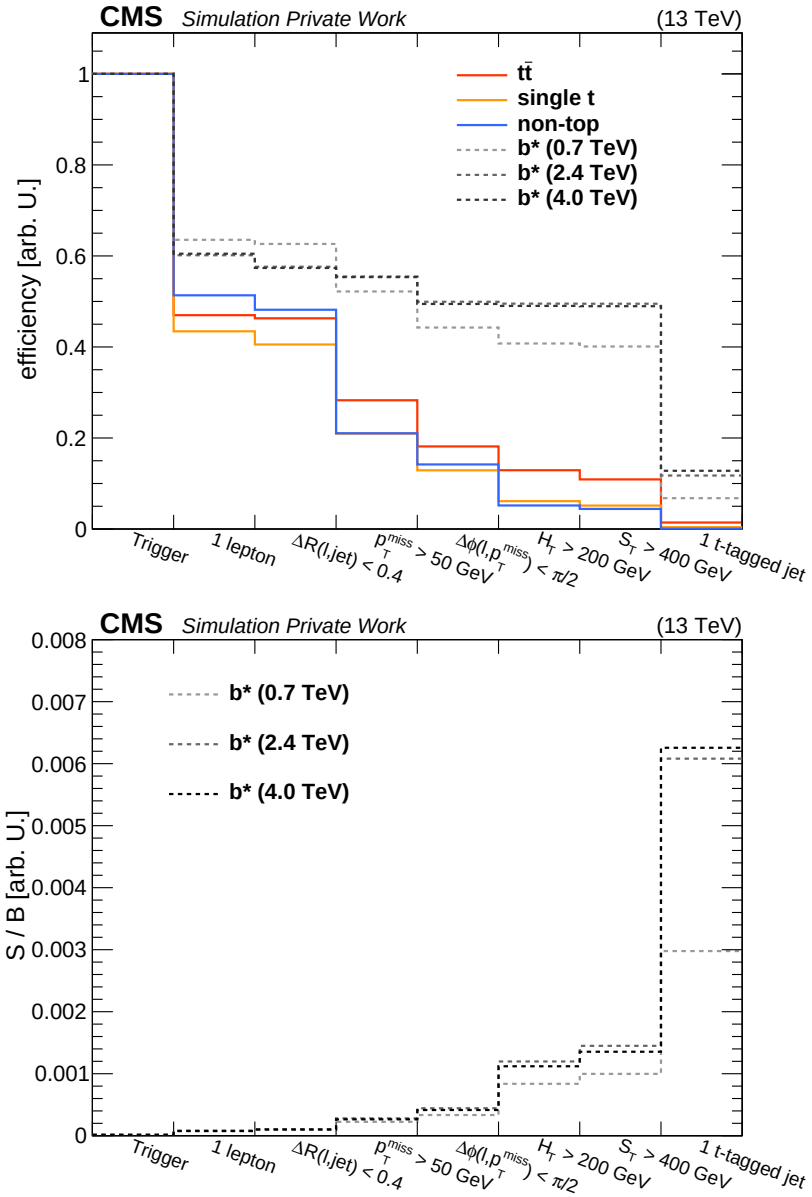

Figure 9.3.: Cutflow diagrams of the event selection. The total selection efficiency with respect to the dataset obtained with the respective trigger selection (upper) and the signal over background ratio (lower) is shown for each step in the event selection. The selection efficiency is shown separately for the simulation of the $t\bar{t}$ (red), single t (yellow) and non-top backgrounds (blue) as solid lines. Three exemplary b^* signal samples with $m_{b^*} = 0.7\,\text{TeV}$, $2.4\,\text{TeV}$ and $4.0\,\text{TeV}$ are shown as dashed lines.

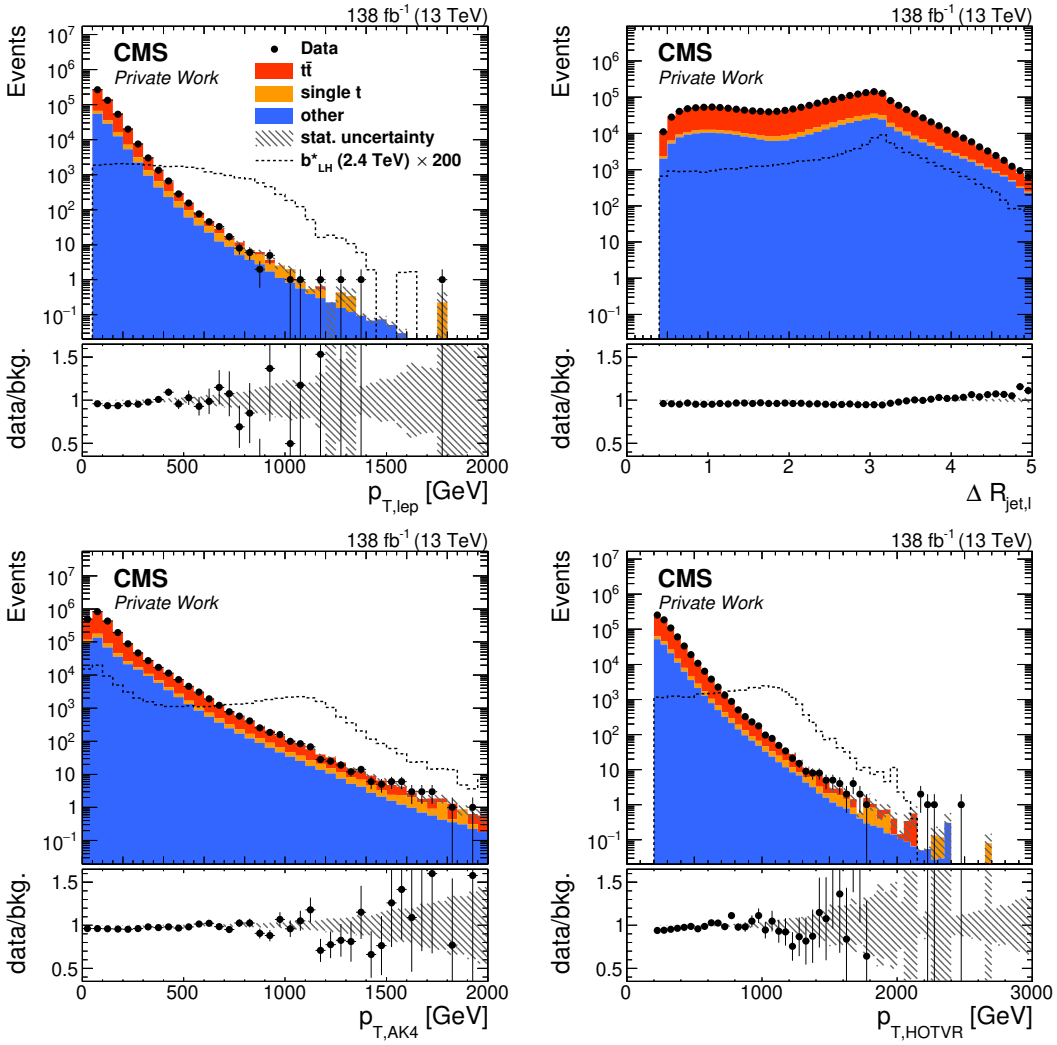

Figure 9.4.: Distributions of (left-to-right, top-to-bottom): p_{T} of the lepton, $\Delta R(\text{AK4 jet}, \ell)$, p_{T} of the AK4 jets, p_{T} of the HOTVR jet after the event selection. Data are shown as black markers, the error bars show the statistical uncertainty. The expected SM background is drawn as stacked filled areas and the hatched area indicates the statistical uncertainty. The expected signal for a b^* quark with $m_{\mathrm{b}^*} = 2.4\,\mathrm{TeV}$ is shown as a dashed line. The signal is scaled up by a factor of 200 for illustration purposes. The bottom panel in each plot shows the data-to-background ratio. The error bars correspond to the statistical uncertainty on the data and the gray area shows the statistical uncertainty on the background.

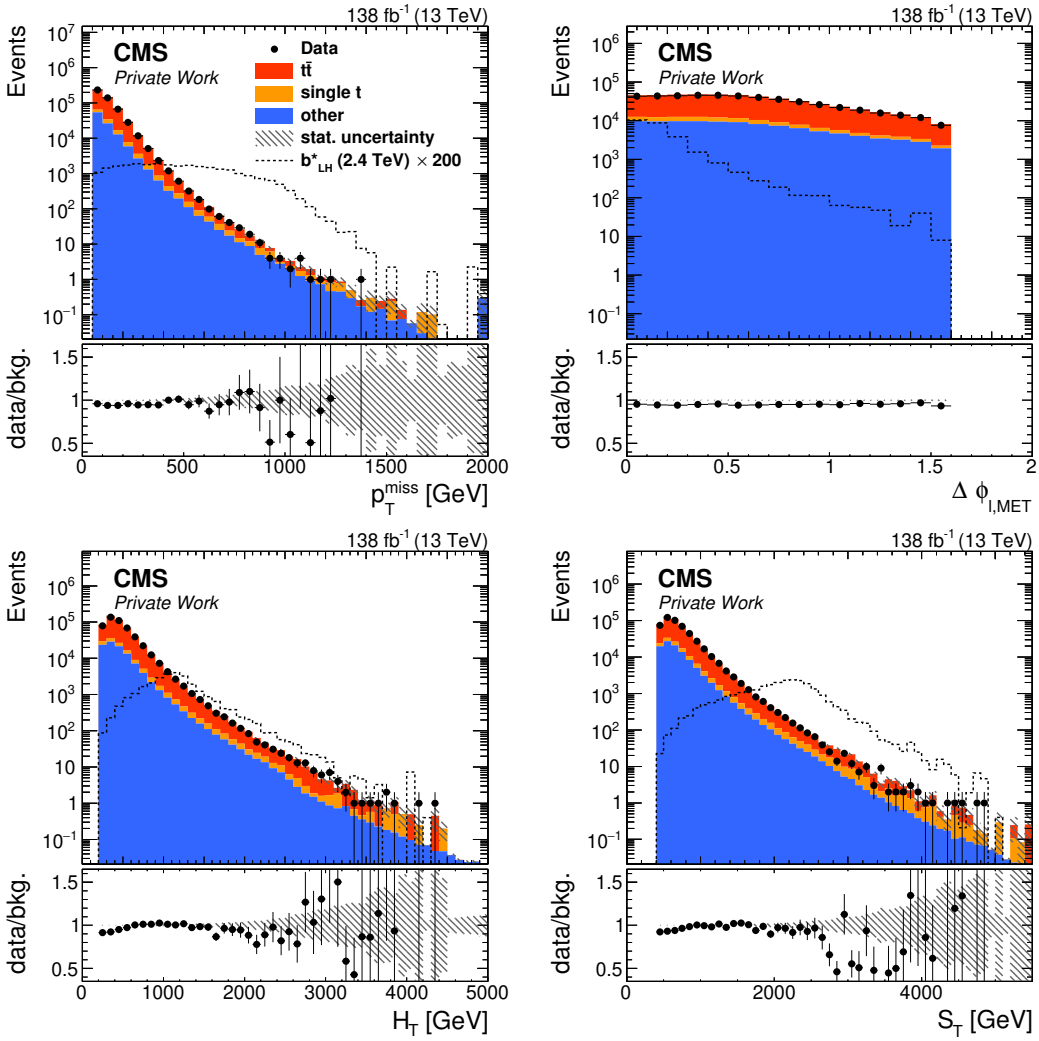

Figure 9.5.: Distributions of (left-to-right, top-to-bottom): $p_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(\ell, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$, H_{T}, S_{T} after the event selection. Data are shown as black markers, the error bars show the statistical uncertainty. The expected SM background is drawn as stacked filled areas and the hatched area indicates the statistical uncertainty. The expected signal for a b^* quark with $m_{\mathrm{b}^*} = 2.4\,\mathrm{TeV}$ is shown as a dashed line. The signal is scaled up by a factor of 200 for illustration purposes. The bottom panel in each plot shows the data-to-background ratio. The error bars correspond to the statistical uncertainty on the data and the gray area shows the statistical uncertainty on the background.

expected SM background still consists of these processes, due to their comparatively large cross sections. In Section 9.3, the categorization of the events passing the event selection is described, where a further separation of the non-top backgrounds will be achieved.

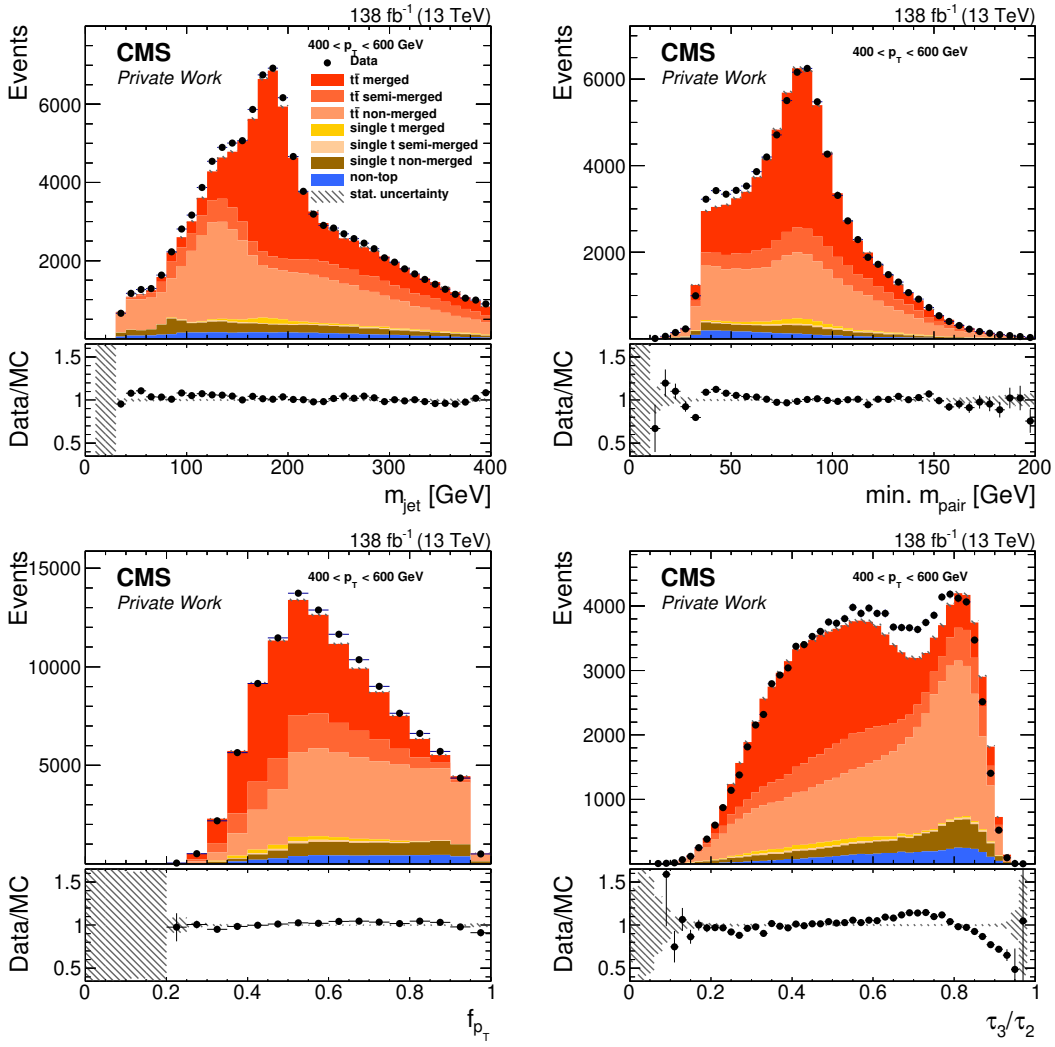

Figure 9.6.: Distributions of (left-to-right, top-to-bottom): HOTVR jet mass, minimum m_{pair}, $f_{p_{\text{T}}}$, τ_3/τ_2 in a $\text{t}\bar{\text{t}}$ enriched region for HOTVR jets with $400 < p_{\text{T}} < 600\,\text{GeV}$. Data are shown as black markers, the error bars show the statistical uncertainty. The expected SM background is drawn as stacked filled areas and the hatched area indicates the statistical uncertainty. The bottom panel in each plot shows the data-to-background ratio. The error bars correspond to the statistical uncertainty on the data and the gray area shows the statistical uncertainty on the background.

In Fig. 9.6, distributions of the HOTVR substructure variables used for t tagging are shown. They are measured using events passing the event selection described above

without the t tagging requirement. Additionally, at least two b tagged AK4 jets are required, to supress the non-top backgrounds. The distributions are shown for HOTVR jets with $400 < p_{\mathrm{T}} < 600\,\mathrm{GeV}$, the distributions for $200 < p_{\mathrm{T}} < 400\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 600\,\mathrm{GeV}$ are shown in Appendix A. The $\mathrm{t\bar{t}}$ and single t events are each separated into three categories, depending on the number of generator level quarks from the hadronic t quark decay that are matched to the HOTVR jet. A generator level quark is considered matched to the HOTVR jets if its angular distance to the jet axis is smaller than the effective distance parameter of the jet $\Delta R < R_{\mathrm{eff}} = \rho/p_{\mathrm{T}}$. If all three generator level quarks are matched to the HOTVR jet, it is considered a merged t jet. If instead only two generator quarks are matched to the HOTVR jet, it is considered a semi-merged t jet, and if one or no generator level quarks are matched to it, the jet is considered non-merged. The regions selected by the t tagging criteria described in Section 6.5.4 are dominantly populated by merged t jets. A deviation between data and simulation is seen at high values of τ_3/τ_2. This is a known mismodeling of the simulation and is also seen by other analyses [82]. All other distributions are well described by the simulation. Dedicated efficiency measurements are performed by the CMS Collaboration [82, 137], to correct for the effect of this mismodeling on the t tagging efficiency.

9.2. Reconstruction of the tW System

The invariant mass of the tW system $M_{\mathrm{tW}} = \sqrt{(p_{\mathrm{t}}^{\mu} + p_{\mathrm{W}}^{\mu})^2}$ is sensitive to the signal. The signal is expected to peak at the respective mass of the heavy resonance in the M_{tW} spectrum, while the SM processes feature a falling distribution.

The tW system is reconstructed by summing the four momenta of the t quark and W boson candidates. The t-tagged HOTVR jet is used as the t quark candidate and the W boson candidate is built by combining the four momenta of the isolated charged lepton and the neutrino. Since the neutrino itself can not be measured with the CMS detector, it has to be reconstructed from other available information. A commonly used approach within the CMS collaboration is followed. Assuming that the W boson is produced on its mass shell, the neutrino can be reconstructed by

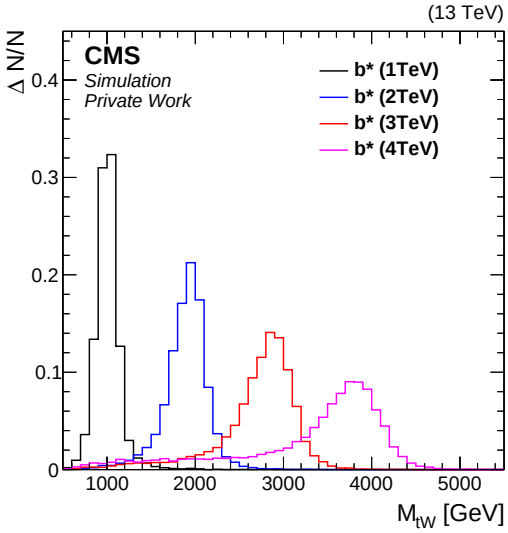

Figure 9.7.: Invariant mass of the tW system for b* quark samples with $M_{\mathrm{b}^*} = 1\,\mathrm{TeV}$, 2 TeV, 3 TeV and 4 TeV. Each distribution is normalized such that the integral is one.

solving the quadratic equation

$$M_{\mathrm{W}}^2 = (p_{\mathrm{W}}^{\mu})^2 = (p_{\ell}^{\mu} + p_{\nu}^{\mu})^2 \tag{9.1}$$

for the z component of the neutrino four-momentum $p_{z,\ell}$. The solutions for this equation are given by

$$p_{z,\nu}^{\pm} = \frac{a p_{z,\ell}}{p_{\mathrm{T},\ell}^2} \pm \sqrt{\frac{a^2 p_{z,\ell}^2}{p_{\mathrm{T},\ell}^4} - \frac{E_{z,\ell}^2 p_{\mathrm{T},\nu}^2 - a^2}{p_{\mathrm{T},\ell}^2}}\,. \tag{9.2}$$

with $a = \frac{M_{\mathrm{W}}^2}{2} + p_{\mathrm{T},\nu} p_{\mathrm{T},\ell} \cos(\Delta\phi)$. The transverse momentum of the neutrino is obtained by assuming that the neutrino is the only source of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ in the event, i.e.

$$\vec{p}_{\mathrm{T}}^{\,\nu} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\,. \tag{9.3}$$

Equation (9.2) can have zero, one or two real solutions. If no real solution is found, the real parts of the complex solutions are taken. In case two solutions are found, the solution that minimizes $\left|p_{z,\nu} - p_{z,\mu}\right|$ is chosen. Using the reconstructed neutrino candidate, the tW system can be reconstructed. A distribution of the reconstructed M_{tW} for four different b* quark signal samples is shown in Fig. 9.7.

A χ^2-like estimator X is calculated based on the reconstructed tW system to determine

how signal like the event is, using

$$X^2 = \left(\frac{\Delta\phi_{\mathrm{tW}} - \pi}{\sigma_{\Delta\phi_{\mathrm{tW}}}} \right)^2 + \left(\frac{A_{p_{\mathrm{T}}}}{\sigma_{A_{p_{\mathrm{T}}}}} \right)^2 . \tag{9.4}$$

The first term resembles the back-to-back signature of the signal. It becomes small, if the azimuthal distance between the t quark and the W boson candidates $\Delta\phi_{\mathrm{tW}}$ is close to π. The second term utilizes momentum conservation. The t quark and the W boson in signal events originate from the decay of a heavy particle, so their transverse momenta should be of the same size. Hence, the second term becomes small, if the p_{T} asymmetry

$$A_{p_{\mathrm{T}}} = \frac{p_{\mathrm{T}}^{\mathrm{t}} - p_{\mathrm{T}}^{\mathrm{W}}}{p_{\mathrm{T}}^{\mathrm{t}} + p_{\mathrm{T}}^{\mathrm{W}}} \tag{9.5}$$

is close to zero. The width parameters $\sigma_{\Delta\phi_{\mathrm{tW}}}$ and $\sigma_{A_{p_{\mathrm{T}}}}$ are estimated from fits to the distributions of $\Delta\phi_{\mathrm{tW}}$ and $A_{p_{\mathrm{T}}}$ in signal events where the reconstructed t quark, lepton and neutrino were matched to generator level objects. For the matching, an angular distance between the generator level and detector level objects of $\Delta R < 0.4$ is required. The sum of all signal samples is used in the fit. The results are shown in figure 9.8. Note, that the distributions are only approximately Gaussian in the peak region. A more dedicated, but computationally more expensive method to obtain a discriminator evaluating the likelihood of a given value based on Lorentzian or Crystal Ball functions was tested, but did not improve the sensitivity of the discriminator significantly.

9.3. Event Categorization

Events passing the event selection are sorted into one of three categories according to the number of b-tagged AK4 jets. The distribution of the number of b-tagged AK4 jets after the event selection is shown in Fig. 9.9. The categories are defined by having either zero ("0b"), one ("1b"), or more than one b-tagged jets ("2b"). A summary of the event categorization, together with the obtained M_{tW} distributions is shown in Fig. 9.10. All three categories result in a very similar overall shape. The different background compositions between the categories allow to define a dedicated signal region, as well as control regions for the dominant and subdominant SM backgrounds.

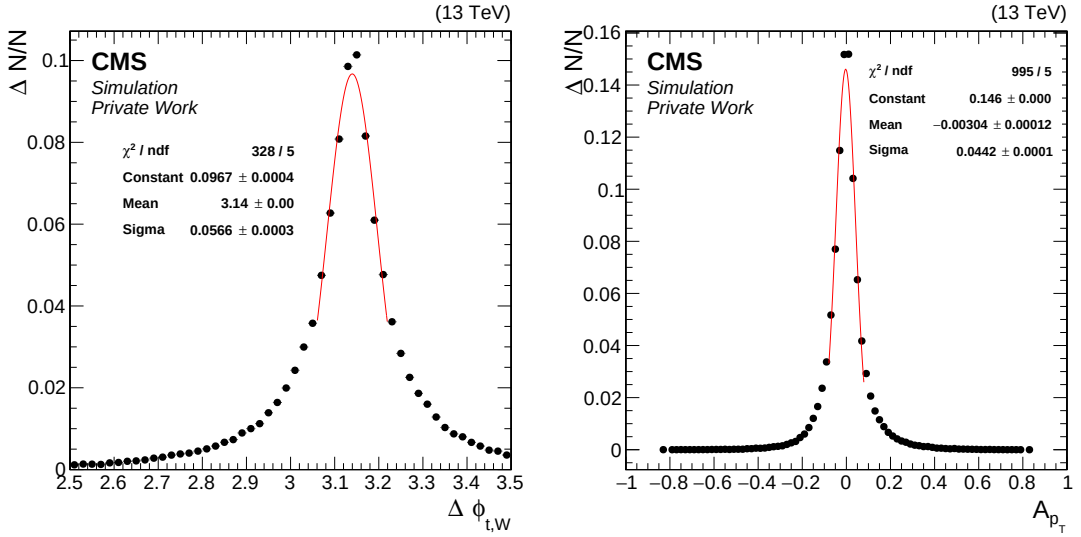

Figure 9.8.: Distributions of $\Delta\phi_{\mathrm{tW}}$ (left) and $A_{p_{\mathrm{T}}}$ (right). Each distribution is normalized to one. A Gaussian is fitted to the peak region of each distribution, to estimate the widths $\sigma_{\Delta\phi_{\mathrm{tW}}}$ and $\sigma_{A_{p_{\mathrm{T}}}}$. The best fit parameters are given on each plot and used in Eq. (9.4).

Due to the requirement of a t-tagged HOTVR jet, but a veto on b-tagged jets, the 0b category is populated dominantly by SM background events containing no t quarks (non-top background), where the HOTVR jet was misidentified as a t jet. The W/Z+jets and diboson production constitute the dominant processes in this region. The 0b region is used to derive the non-top background from data in section Section 9.4. The 2b category is dominated by SM $\mathrm{t\bar{t}}$ production and is used to constrain systematic uncertainties associated with the modeling of this background. Even though the b^* signal results in two b quarks in the final state, one from the t decay and one from the gluon splitting in the initial state, only a small fraction of signal events end up in the 2b category. This happens because the spectator b quark generally has too small values of p_{T} or is too forward in the detector to be identified as a b jet. The highest signal sensitivity is achieved in the 1b category. To further increase the sensitivity in this category, two additional selection criteria are imposed. The b-tagged AK4 jet is required to have an angular distance to the isolated lepton $\Delta R(\ell, \text{b-jet}) > 2.0$, and the χ^2-like estimator must have a value smaller than 20. Distributions of the χ^2-like estimator (left) and $\Delta R(\ell, \text{b-jet})$ (right) are shown in Fig. 9.11. The first requirement suppresses the contribution from $\mathrm{t\bar{t}}$ production, since the b-tagged jet has an equal probability to originate from either t quark, resulting in the two peaks in the right plot. The isolation criterion reduces the peak at low $\Delta R(\ell, \text{b-jet})$, but a large number of $\mathrm{t\bar{t}}$ events can still be removed

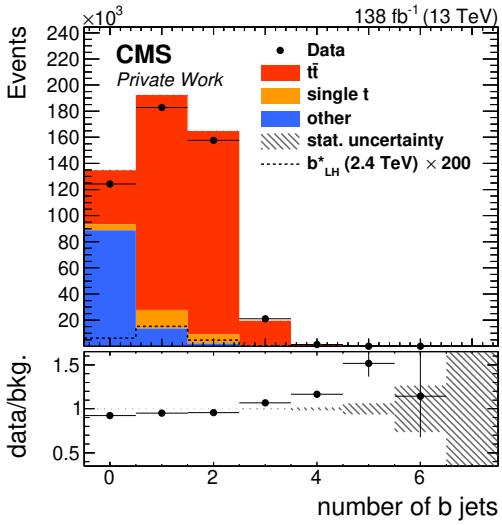

Figure 9.9.: Distribution of the number of b-tagged AK4 jets after the event selection. For better visualization, the signal is scaled up by a factor of 200. The gray area in the bottom pannel corresponds to the statistical uncertainty on the background prediction.

by this requirement. For signal, the shape of $\Delta R(\ell, \text{b-jet})$ depends on m_{b^*} and the requirement was chosen to keep the selection inclusive for all considered mass states. The distribution of the χ^2-like estimator peaks at low values for signal and SM backgrounds. However, since the signal distribution falls faster towards higher values of X^2, the selection increases the signal sensitivity.

9.4. Background Estimation

The dominant backgrounds in the signal region (1b), $\text{t}\bar{\text{t}}$ and single t production are estimated using the simulated samples. Scale factors are applied to correct for different selection efficiencies for the lepton, and t- and b-jets between data and simulation. The modeling of the dominant $\text{t}\bar{\text{t}}$ backgrounds is constrained using the dedicated 2b control region in a fit to data.

Although the data and simulation are in overall good agreement, a slight trend in the data to simulation ratio is visible in the M_{tW} distributions shown in Fig. 9.10. This trend is most pronounced in the 0b category, indicating a disagreement between data and the simulation of the non-top backgrounds, where the t jet was misidentified.

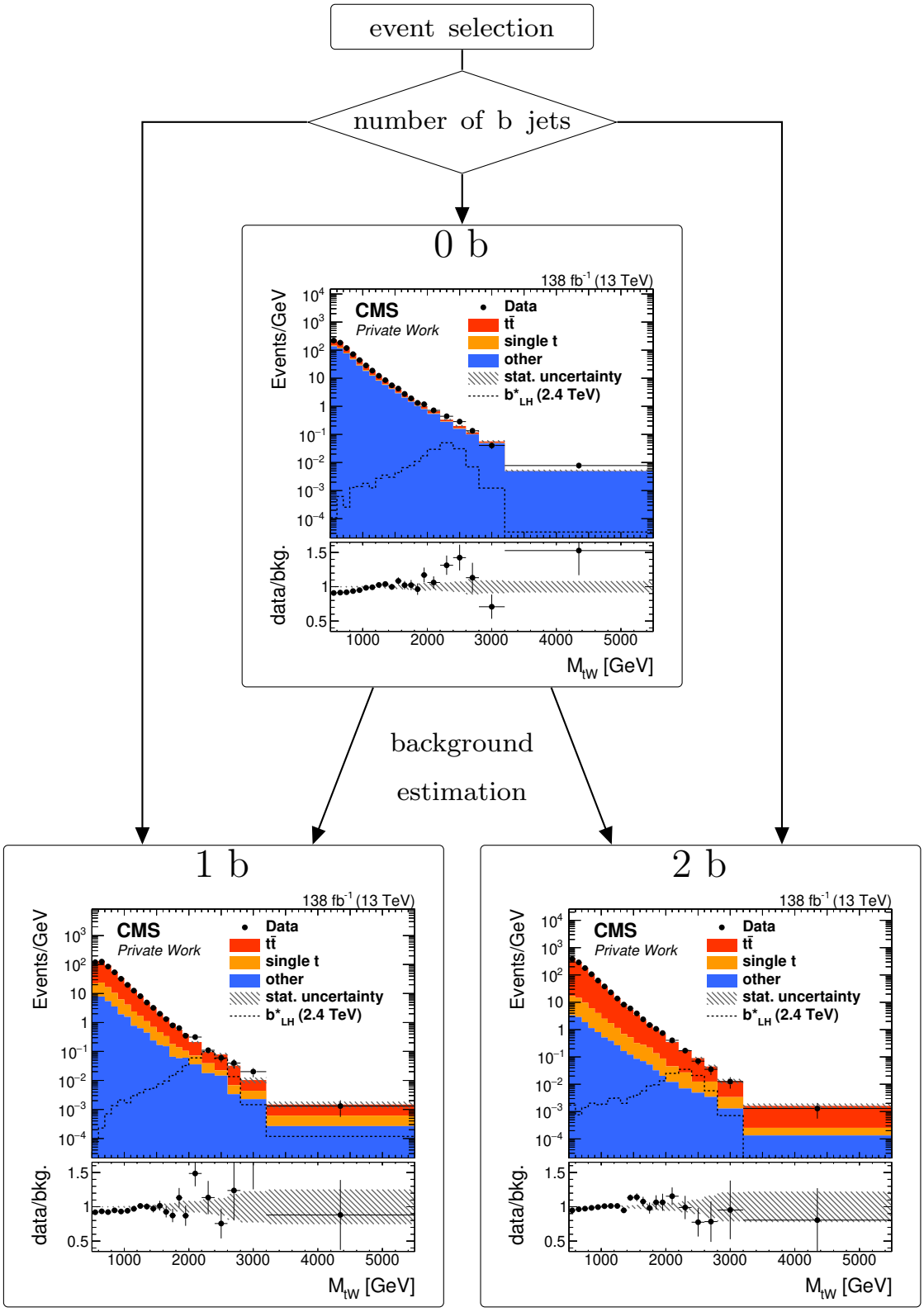

Figure 9.10.: Overview of the event categorization. The plots show the distributions of M_{tW} in the 0b (top), 1b (bottom left) and 2b (bottom right) categories.

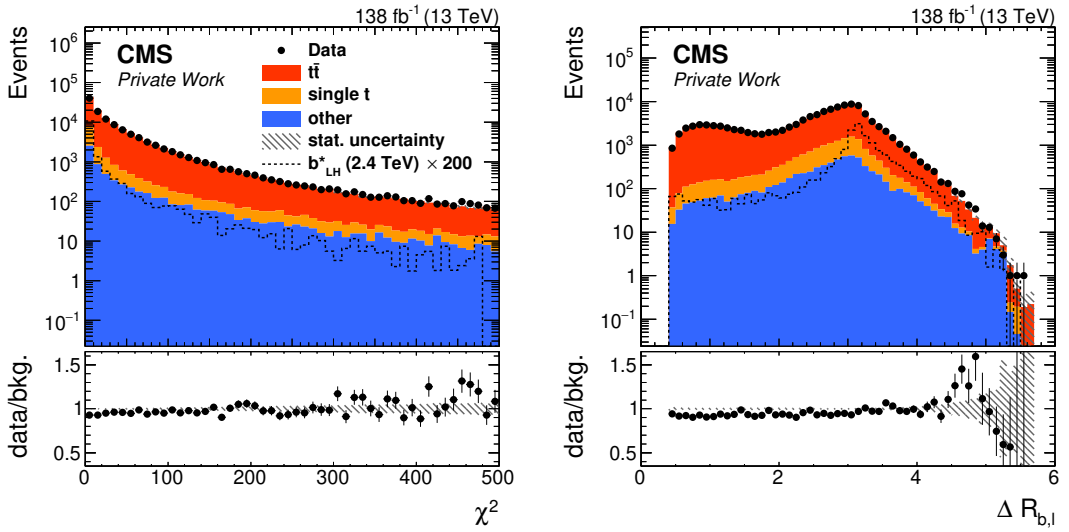

Figure 9.11.: Distributions of the χ^2 estimator (left) and $\Delta R(\ell, \text{b-jet})$ (right) in the 1b category. The signal is scaled up by a factor of 200 to enhance visibility. The gray area in the bottom pannel corresponds to the statistical uncertainty on the background prediction.

The simulation of the misidentification rate of t jets relies on the modeling of jet substructure in light quark and gluon jets, which is not described perfectly. The exact misidentification rate depends on the composition of light quark and gluon jets in the selected phase space, such that a general correction factor can not be given. Hence, contributions from non-top backgrounds with a misidentified t jet in the 1b and 2b categories are estimated in a data-driven approach using the 0b category. The α-ratio method [138] is used to extrapolate the shape and normalization of the non-top backgrounds into the 1b and 2b regions.

The α-ratio is calculated as

$$\alpha_i = \frac{N_{\text{destination},i}}{N_{\text{control},i}}\,, \tag{9.6}$$

where N is the number of weighted non-top events from simulation of bin i in M_{tW} from the 0b control and destination region (1b or 2b). The resulting distributions are shown in Fig. 9.12 as black markers. The distribution in the 1b category (left) shows a turn-on behavior at lower values of M_{tW} due to the additional $\Delta R(\ell, \text{b-jet})$ and X^2 requirements in the signal region. The distribution for the 2b (right) is mostly flat. In order to obtain a transfer function $\alpha(M_{\text{tW}})$ unaffected of statistical fluctuations in the simulated samples, the α distributions are fitted. A Gaussian function plus a constant are used to model the distributions. The systematic uncertainty arising from the choice of the parametrization of $\alpha(M_{\text{tW}})$ is estimated from a fit to the α

distribution using a Landau function. The total uncertainty on the transfer function is obtained as the difference between the two functions added in quadrature to the statistical uncertainties of the fits.

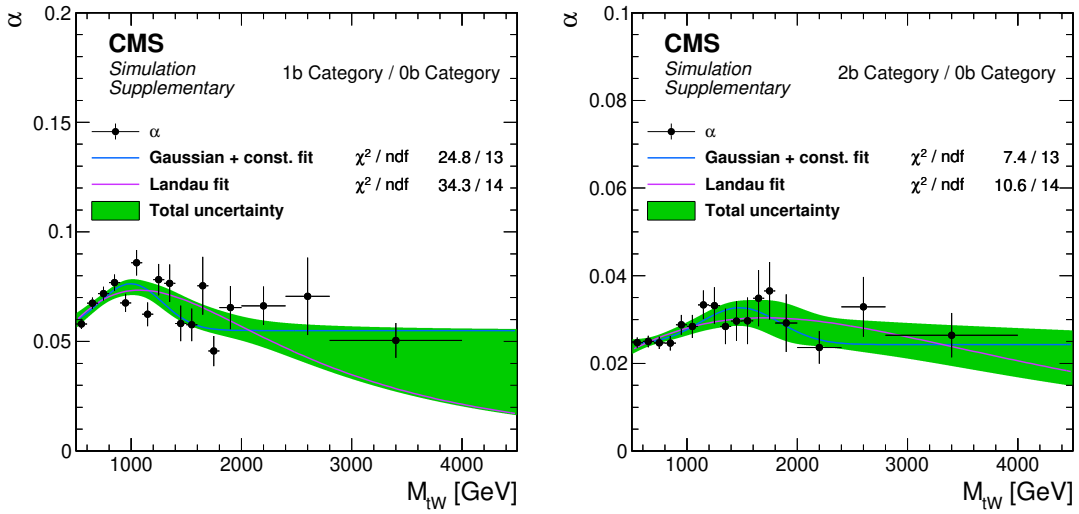

Figure 9.12.: Distribution of the α-ratio as a function of M_{tW} for the 1b (left) and 2b (right) categories. The ratio is fitted using two different parametrizations to obtain a transfer function, depicted as solid lines. The total uncertainty on the transfer function is shown as colored area. Published in the supplementary material [139] of Ref. [1].

The contributions of the non-top backgrounds in the 1b and 2b categories are estimated from data using the obtained functions $\alpha(M_{\mathrm{tW}})$. In order to keep the background estimation in the 1b and 2b categories statistically independent, the data in the 0b category is split randomly into two subsets. Two thirds of the data are used to estimate the non-top backgrounds in the 1b category and one third of the data are used in the 2b category. The prediction of the non-top backgrounds in each region are then obtained by assigning a weight $\omega = \alpha(M_{\mathrm{tW}})/f$ to each event in one of the subsets of the 0b category, using the transfer function $\alpha(M_{\mathrm{tW}})$ to extrapolate into the 1b or 2b categories. Here, f is the fraction of events in the subset of the 0b category. The statistical uncertainty from the non-top background obtained from data in the 0b category is negligible compared to other uncertainties, so the exact choice of f is not important. The small fraction of $\mathrm{t\bar{t}}$ and single t events in the 0b category are accounted for, by subtracting their contributions, as obtained from the simulation, from the data in the 0b category.

The non-top background estimation procedure is validated using a signal depleted validation region in the 1b category, obtained by changing the χ^2 requirement to

$50 < \chi^2 < 150$. The background estimation procedure is performed in the same way as described above for the 1b validation region. The α-ratio distribution and transfer function are shown in Fig. 9.13. Due to the different χ^2 requirement, the turn-on behavior observed in the 1b signal region is no longer visible. The distributions of M_{tW} in the 1b validation region using the non-top prediction obtained from simulation (left) and from the background estimation (right) are shown in Fig. 9.14. A good agreement between data and background prediction is observed in both cases. However, the data-driven background estimation shows less statistical fluctuations than the simulation.

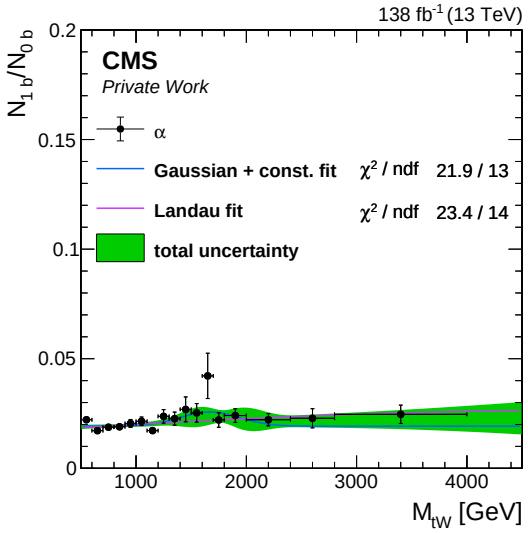

Figure 9.13.: Distribution of the α-ratio as a function of M_{tW} for the 1b validation region. The ratio is fitted using two different parametrizations to obtain a transfer function, depicted as solid lines. The total uncertainty on the transfer function is shown as colored area.

9.5. Systematic Uncertainties

Several systematic uncertainties are considered in the analysis, affecting shape and normalization of the final M_{tW} distributions in the 1b and 2b categories. In the following, a detailed description of the considered sources of systematic uncertainties is given.

- The total integrated luminosity of the analyzed dataset is assigned an uncertainty of 1.6% [121–123]. The integrated luminosity is used to calculate the

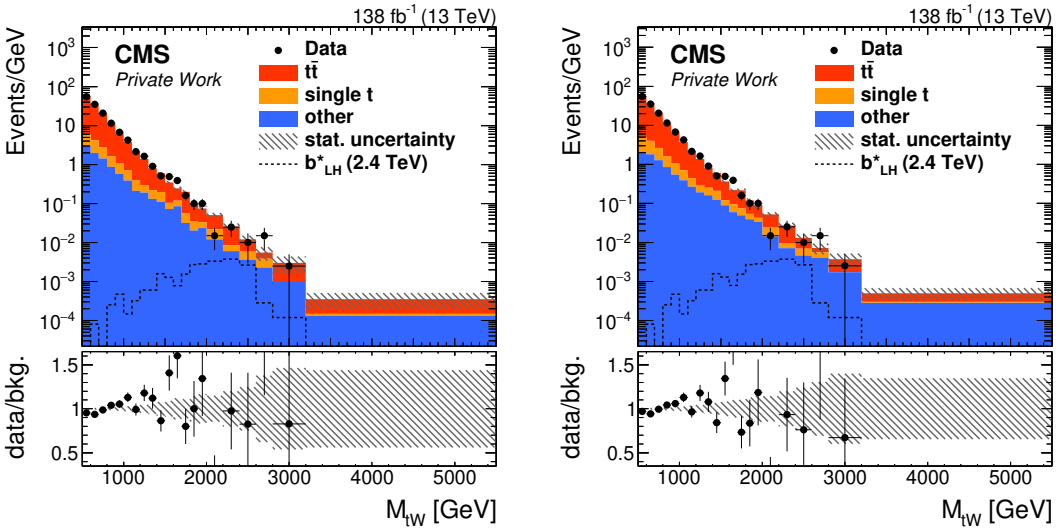

Figure 9.14.: Distributions of M_{tW} in the 1b validation region using the prediction of the non-top backgrounds from simulation (left) and the data driven estimation (right). The gray area in the bottom pannel corresponds to the statistical uncertainty on the background prediction.

luminosity weights in Eq. (7.3) and affects only the normalization of the M_{tW} distributions.

- The $\mathrm{t\bar{t}}$ and single t production cross sections are assigned normalization uncertainties of 20% and 30%, respectively. These account for uncertainties due to missing higher orders, estimated by halving and doubling the renormalization μ_r and factorization μ_f scales in the corresponding simulations, and for uncertainties from the normalization to the NNLO and NLO predictions.

- The uncertainty from the choice of PDFs is estimated by calculating the signal and background predictions in each bin of the M_{tW} distribution using sets of varied PDF replicas [140]. The standard deviation of these predictions is used to construct shape variations. For the simulations of $\mathrm{t\bar{t}}$, single t, and B production, 100 replicas of the NNPDF sets are used. For the b^* signal samples, the PDF replicas used are listed in Table 9.1. In case of the signal samples, the distributions are normalized to the respective cross sections, such that only acceptance effects are considered.

- The minimum bias cross section of 69.2 mb is assigned an uncertainty of $\pm 5\%$ [136]. From this, the uncertainty on the the pileup distribution is estimated

and propagated to the event yields in the simulation.

- The jet energy scale (JES) and jet energy resolution (JER) corrections are varied within their uncertainties to determine their effect on the shapes of the M_{tW} distributions. The varied M_{tW} distributions are smoothed using a linear fit to the ratio between the varied and nominal distributions to suppress effects from statistical fluctuations in the alternative distributions.

- Uncertainties in the electron and muon identification and reconstruction are estimated by varying the corresponding efficiency corrections within their uncertainties.

- Trigger efficiency corrections are varied within their uncertainties.

- The transfer functions $\alpha(M_{\mathrm{tW}})$ used for the background estimation of the non-top backgrounds are varied within their uncertainties, as discussed in Section 9.4.

- Differences in the t tagging efficiency between data and simulation are accounted for by data-to-simulation scale factors [82, 137]. These scale factors are varied within their uncertainties. The HOTVR t tagging scale factors are split into fully merged, partially merged and non-merged cases, depending on how many partons from the t quark decay can be matched to the HOTVR jet. These scale factors are uncorrelated and are varied independently to obtain the effect on the M_{tW} distributions for each variation.

- The b tagging scale factors are handled similar to the ones from t tagging. They are split into efficiencies for b, c and light jets [118]. The uncertainties for b and c jets are taken as correlated and varied simultaneously, while the uncertainties for light jets are taken as uncorrelated and varied independently.

- The uncertainty arising from the modeling of the t quark p_{T} spectrum is obtained from a fit to data simultaneously with the statistical interpretation of the results. The parameter β (see Eq. (7.1)), affecting the shape of the p_{T} distribution, is constrained by this fit. The templates needed for this fit are obtained by changing β by $\pm 50\%$. The parameter α only affects the normalization of the $\mathrm{t\bar{t}}$ background and is accounted for in the 20% uncertainty

Signal Sample	PDF Set Name	Number of Replicas
2016 1.2 TeV $\leq M_{\mathrm{b}^*} \leq$ 3 TeV	MMHT2014lo68cl	51
2016 1.2 TeV $\geq M_{\mathrm{b}^*} \| \| M_{\mathrm{b}^*} \geq$ 3 TeV	PDF4LHC15_nnlo_30_pdfas	33
2017	PDF4LHC15_nnlo_30_pdfas	33
2018	PDF4LHC15_nnlo_30_pdfas	33

Table 9.1.: PDF sets used to evaluate the uncertainty arising from the choice of PDF in b^* signal samples. The first column indicates the b^* samples. The name of the PDF set used to evaluate the uncertainty on the choice of PDFs is given in the second column. The third gives the number of PDF replicas in the given set.

on the $\mathrm{t\bar{t}}$ production cross section.

A summary of all considered sources is given in Table 9.2. Some of these uncertainties are assumed to be fully correlated throughout all three years, these uncertainties are given in the top part of the table. The contributions of uncertainties without year-to-year correlations are calculated for each year independently.

9.6. Statistical Interpretation

The distributions of M_{tW} in the 1b (left) and 2b (right) including all systematic uncertainties are shown in Fig. 9.15. The shaded area in the bottom panel of the plots corresponds to the total uncertainty of the predicted background. The largest source of uncertainty arises from the modeling of the $\mathrm{t\bar{t}}$ prediction. In order to obtain the final background prediction and to constrain the uncertainty on the $\mathrm{t\bar{t}}$ modeling a simultaneous binned maximum likelihood template fit of the M_{tW} distribution of the expected background in the 1b and 2b categories to data is performed. The systematic uncertainties discussed in Section 9.5 are accounted for by nuisance parameters. A log-normal prior distribution is used for normalization uncertainties. Uncertainties affecting the shape of the M_{tW} distributions are treated using a template morphing approach using Gaussian priors [141]. Statistical uncertainties are accounted for using a simplified version of the Barlow-Beeston method [142], where for each bin a single nuisance parameter with a Gaussian prior is added to reflect the statistical uncertainty in this bin. The compatibility of the signal hypothesis with data is tested, by performing a signal + background fit, where the signal rate r is allowed to float unconstrained.

Source	Uncertainty	Samples	Impact (up/down)
$t\bar{t}$ cross section*	± 20%	$t\bar{t}$	+5.1 / -5.0%
Single t cross section*	± 30%	single t	-4.2 / +4.8%
Luminosity*	± 1.6%	$t\bar{t}$, single t, signal	−1.4 / +1.5%
t p_T reweighting*	Shape	$t\bar{t}$	-5.0 / +5.1%
PDF*	Shape	$t\bar{t}$, single t, signal	-4.1 / +4.6%
Background estimation (1b)	Shape	non-top (from data)	-5.3 / +6.9%
Background estimation (2b)	Shape	non-top (from data)	-0.3 / -0.4%
Pileup*	Shape	$t\bar{t}$, single t, signal	-0.4 / +0.5%
JES*	Shape	$t\bar{t}$, single t, signal	-1.3 / +2.3%
JER*	Shape	$t\bar{t}$, single t, signal	+0.0 / +0.4%
Trigger prefiring	Shape	$t\bar{t}$, single t, signal	+0.1 / -0.0%
Muon identification	Shape	$t\bar{t}$, single t, signal	+0.4 / -0.1%
Muon isolation	Shape	$t\bar{t}$, single t, signal	+0.3 / -0.1%
Muon trigger	Shape	$t\bar{t}$, single t, signal	-0.1 / +0.4%
Electron identification	Shape	$t\bar{t}$, single t, signal	+0.4 / -0.4%
Electron reconstruction	Shape	$t\bar{t}$, single t, signal	+0.2 / -0.0%
Electron trigger	Shape	$t\bar{t}$, single t, signal	+0.3 / -0.0%
t tagging (fully merged)	Shape	$t\bar{t}$, single t, signal	-1.2 / +1.5%
t tagging (partially merged)	Shape	$t\bar{t}$, single t, signal	-0.7 / +0.8%
t tagging (non-merged)	Shape	$t\bar{t}$, single t, signal	-0.0 / +0.2%
b tagging (b,c)	Shape	$t\bar{t}$, single t, signal	-3.6 / +4.0%
b tagging (u,d,s,g)	Shape	$t\bar{t}$, single t, signal	+0.7 / -0.6%

Table 9.2.: Summary of all considered systematic uncertainties affecting the M_{tW} distributions in the 1b and 2b categories. The source of the uncertainty is given in the first column. The second column indicates if the uncertainty results in a change of normalization or the shape of the M_{tW} distribution. The third column gives the samples affected by the given source ouf uncertainty and the fourth column shows the impact of the uncertainty on the signal rate, when varying it up or down by one standard deviation from its post-fit value of the fit to the combination with the all-hadronic analysis, estimated for a left-handed b^* signal with $M_{b^*} = 2.4\,\text{TeV}$. Uncertainties given in the top part of the table are taken to be fully correlated across the three years. Uncertainties marked with an asterisk (∗) are fully correlated between the all-hadronic and ℓ+jets channels. Taken from Ref. [1].

The best-fit values of the nuisance parameters after the background-only (black) and signal + background (gray) fits to data are shown in Fig. 9.16. The pre-fit values are all initialized to one and the green and yellow areas indicate the $\pm1\sigma$ and $\pm2\sigma$ pre-fit uncertainty, respectively. The error bars correspond to the post-fit uncertainties. All nuisance parameters agree with their pre-fit values witin 1 standard deviation. The parameters affecting the $t\bar{t}$ modeling, i.e. the $t\bar{t}$ cross section and β parameter of the t p_T reweighting function, are constrained by the data in the 2b category, reducing

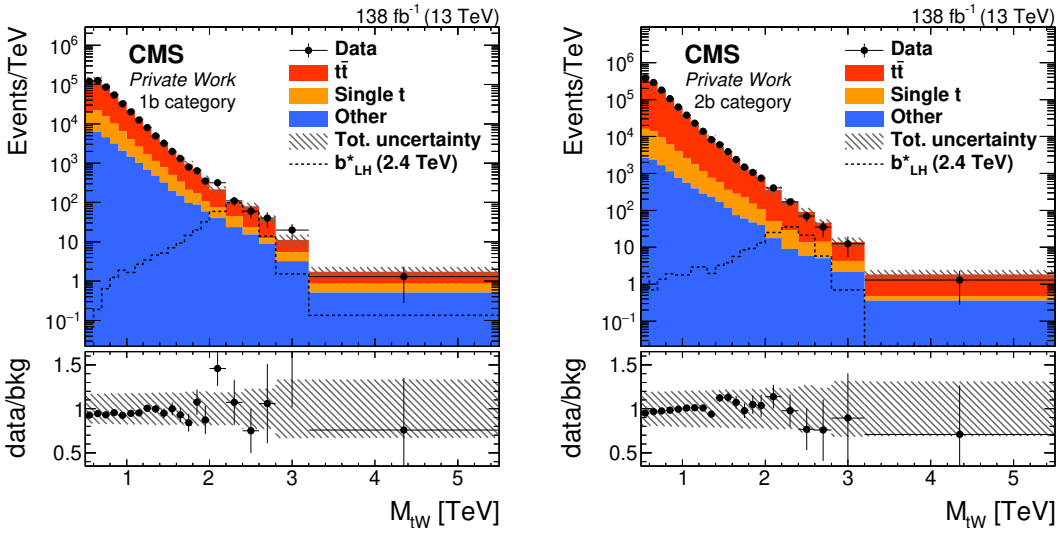

Figure 9.15.: Pre-fit distributions of M_{tW} in the 1b (left) and 2b (right) categories. The gray area in the bottom pannel corresponds to the total uncertainty on the background prediction.

the post-fit uncertainties on these parameters. The background-only and signal + background fit show the same best-fit values, indicating that no strong correlation between the nuisance parameters and the signal rate is present. Additionally, signal injection tests were performed to verify the statistical model. Three different b* signal samples with m_{b^*}=1.0, 2.0 and 3.0 TeV are used in these tests and for each sample 500 pseudo-data sets with a signal injected were created. Another pseudo-data set is created, where no signal was injected. Fits of the signal+background hypothesis are performed to each pseudo-data set. The distributions of the fitted signal strength is found compatible with the injected signal strength for all tested samples.

The post-fit M_{tW} distributions of the background-only fit in the 1b (left) and 2b (right) categories are shown in Fig. 9.17. The uncertainties shown on the plots are the total uncertainties of the post-fit nuisance parameters including correlation effects between the nuisance parameters. A very good agreement between data and prediction is observed and no significant excess of data over the predicted background is found.

Upper limits on the cross section of a singly produced b* decaying to tW are set at 95% confidence level (CL) using an asymptotic approximation of the profile likelihood test statistic [143]. The resulting observed and expected limits are shown in Fig. 9.18. Expected and observed limits are in agreement within two standard

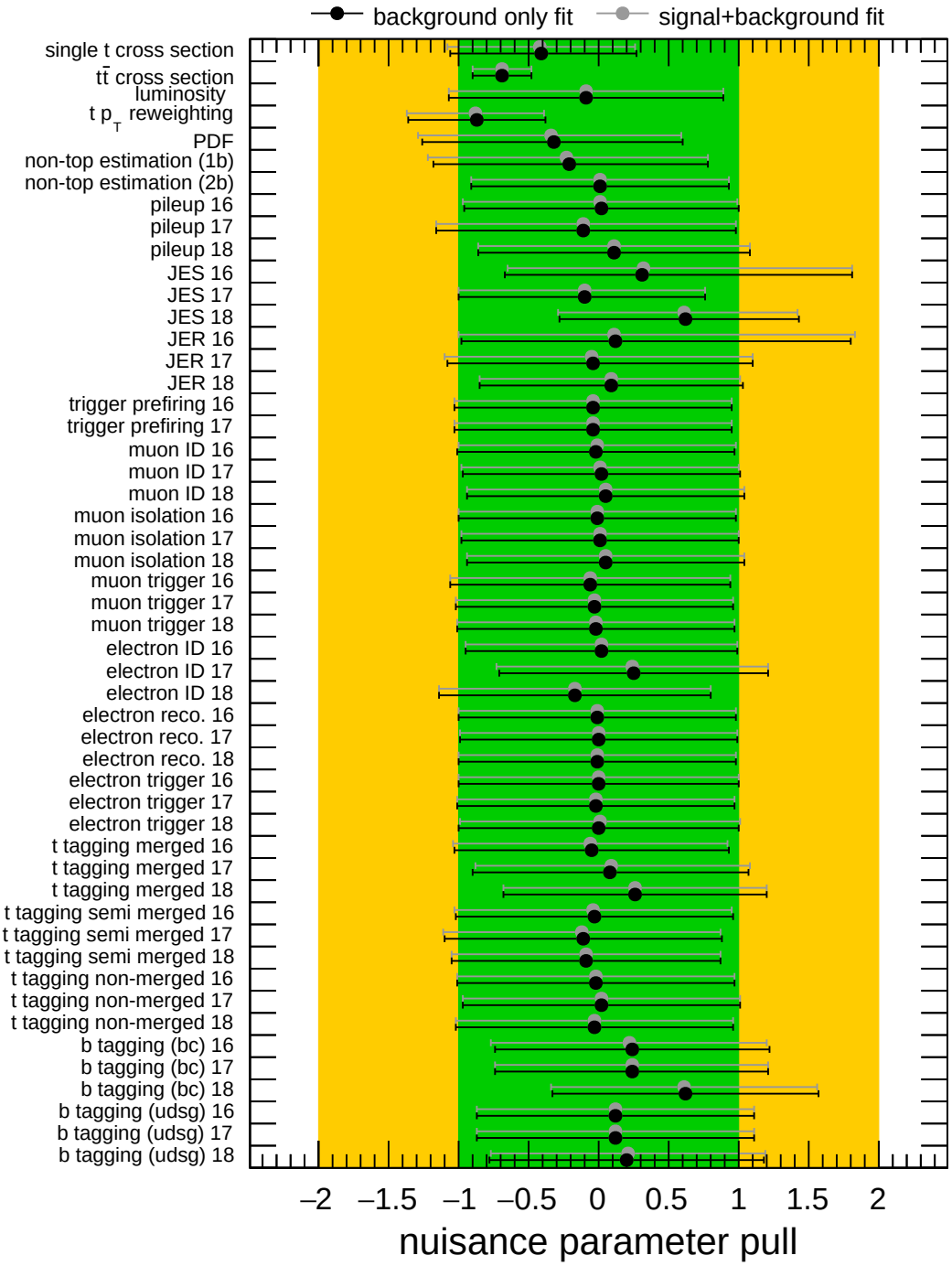

Figure 9.16.: Post-fit pulls of the nuisance parameters for the background only (black) and signal+background fit (gray) with respect to their prefit values. The sample with $m_{\mathrm{b}^*} = 2.4\,\mathrm{TeV}$ and VL chirality is used for the shown signal+background fit. The error bars correspond to the post-fit uncertainties. The green and yellow bands indicate the $\pm 1\sigma$ and $\pm 2\sigma$ prefit uncertainties.

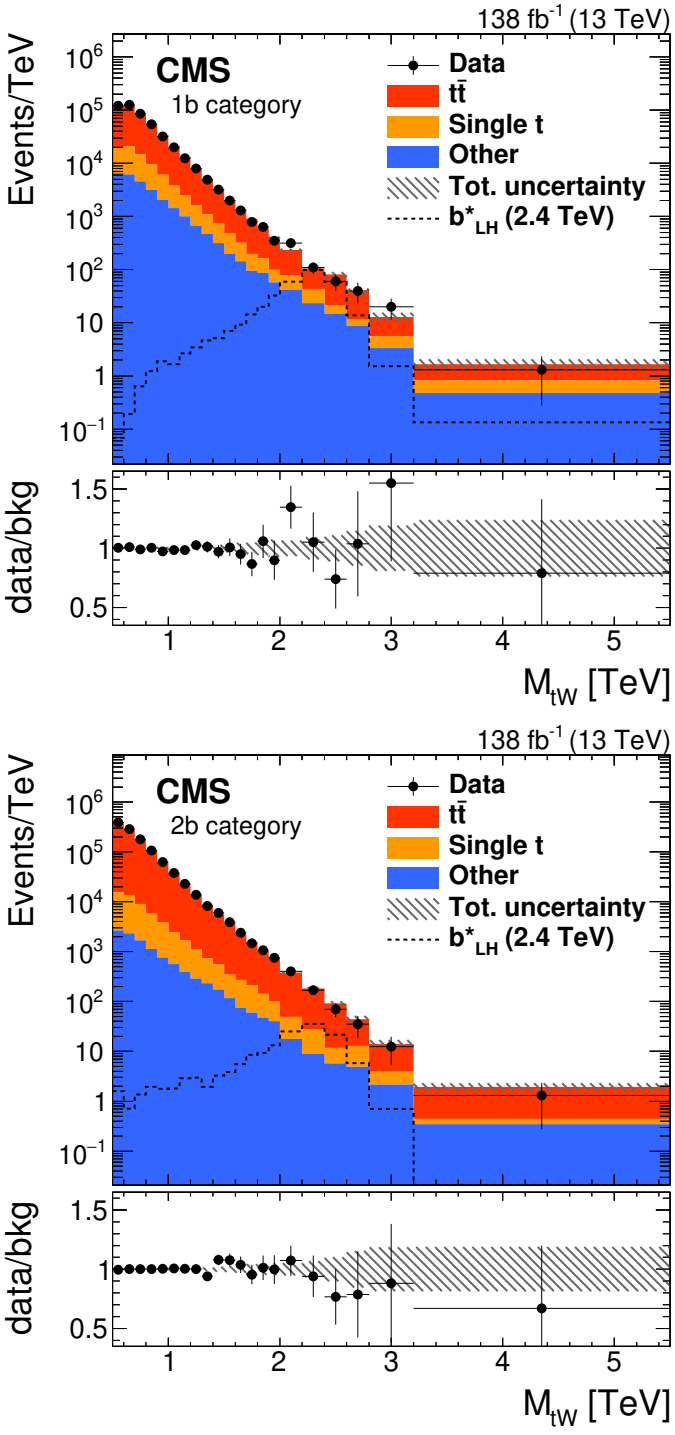

Figure 9.17.: Distributions of M_{tW} in the 1b (left) and 2b (right) categories after the background-only fit to data. The gray area in the bottom pannel corresponds to the total uncertainty on the background prediction. Published in [1].

deviations. The larges deviations between expected and observed limits are at b^* masses of 1.0 TeV and 2.2 TeV. The limit at 1.0 TeV is stronger than the expected limit with a significance of 1.8 standard deviations, resulting from a deficit in data in two bins around $M_{tW} = 1.0$ TeV in the 1b category. At 2.2 TeV the limit is weaker with a significance of 1.6 standard deviations. Mass exclusion limits on the b^* mass are set by comparing the upper cross section limit with the theoretical cross section of $pp \to b^* \to tW$. The observed (expected) mass limits are found to be 2.85 TeV (2.93 TeV), 2.95 TeV (3.04 TeV) and 3.17 TeV (3.23 TeV) for the LH, RH and VL models, respectively.

9.7. Combination with the All-Hadronic Decay Channel

A statistical combination of the presented analysis with the analysis in the all-hadronic final state published by the CMS collaboration in Ref. [2] is performed. In the following, a brief overview of the all-hadronic channel is given and the statistical combination is described.

9.7.1. Analysis of the All-Hadronic Decay Channel

The analysis of the all-hadronic final state targets an event signature with two high-energy jets that are back-to-back in the detector. One of the jets is required to be t-tagged, and the other is required to be W-tagged. Anti-k_T jets with a radius parameter of $R = 0.8$, and $p_T > 400$ GeV and $|\eta| < 2.4$ are used. A jet is considered t-tagged, if the soft drop mass is between $105\,\text{GeV} < m_{SD} < 220\,\text{GeV}$, the N-subjettiness ratio $\tau_3/\tau_2 < 0.65$, and one of the subjets, identified by the soft drop algorithm, has to be tagged as a b jet using the DeepCSV algorithm [118]. Similarly, a jet is considered W-tagged, if the soft drop mass is between $65\,\text{GeV} < m_{SD} < 105\,\text{GeV}$, and the N-subjettiness ratio $\tau_2/\tau_1 < 0.4$ in 2016 or $\tau_2/\tau_1 < 0.45$ in 2017 and 2018. Additionally, the two jets are required to have an azimuthal separation of $|\Delta\phi| > \pi/2$ and in rapidity of $|\Delta y| < 1.6$. Finally, events containing charged leptons with $p_T > 30$ GeV and $|\eta| < 2.4$ fulfilling the loose ID

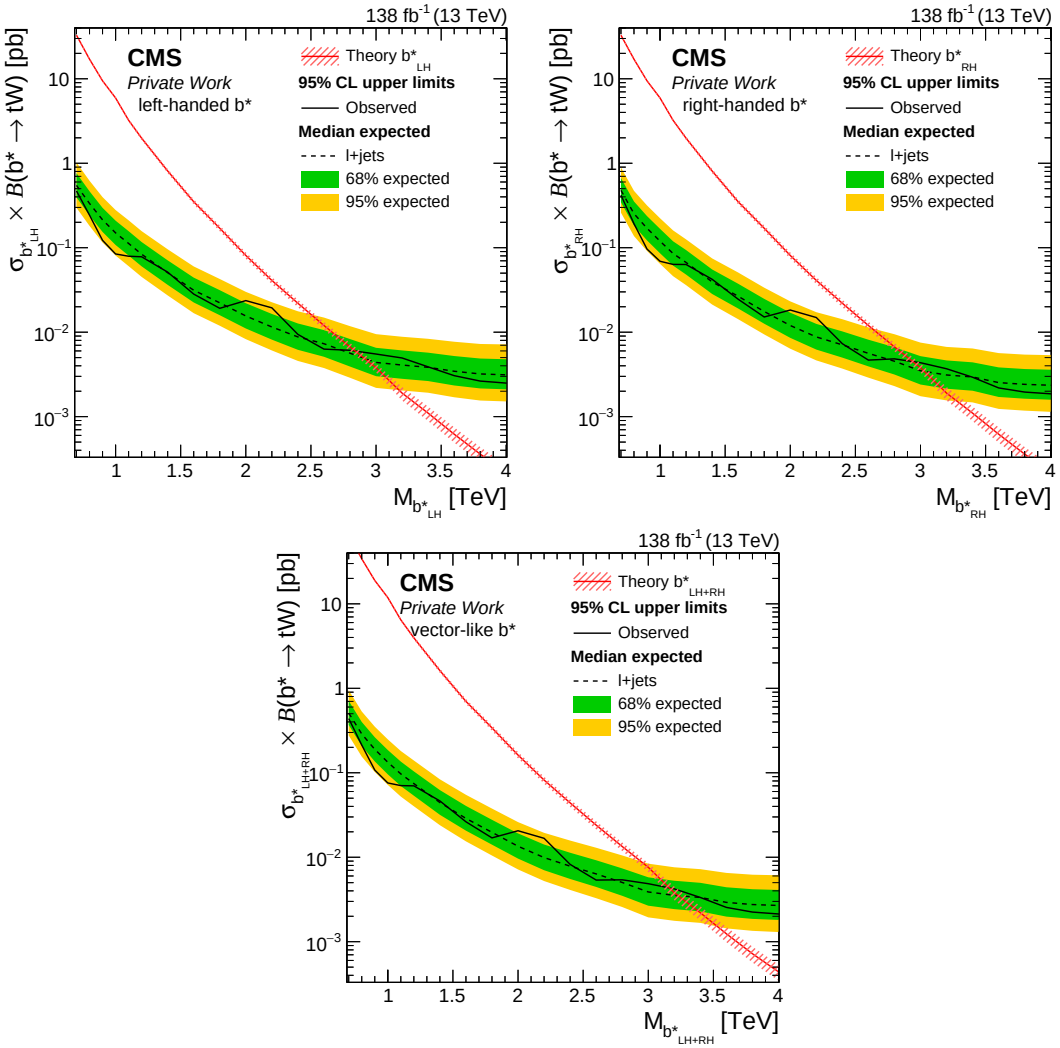

Figure 9.18.: Upper cross section limits on the process pp $\rightarrow$ b* $\rightarrow$ tW for the left-handed (top left), right-handed (top right) and vector-like (bottom) benchmark scenario at 95% CL obtained from the analysis of the ℓ+jets final state. The expected and observed limits are shown as dashed and solid lines, respectively. The green and yellow bands indicate the 68% and 95% confidence intervals on the expected limits. The theoretically predicted cross section is shown as a red line, where the uncertainties due to missing higher orders are depicted by shaded areas.

criteria are rejected to avoid double counting of events that also appear in the ℓ+jets channel.

Control regions of the $\mathrm{t\bar{t}}$ and QCD multijet backgrounds are defined by having different jet tags. Events fulfilling the W tag, but not the t tag requirement are sorted into the QCD multijet control region. Events, where both jets can be t tagged are sorted into the $\mathrm{t\bar{t}}$ control region.

Due to the high cross section, the QCD multijet production constitues the dominant background in the analysis. The contribution of this background is estimated from data. The ratio between signal and control region of the two-dimensional (m_{t}, M_{tW}) distributions is calculated in data $R^{\mathrm{data}}_{\mathrm{P/F}}(m_{\mathrm{t}}, M_{\mathrm{tW}})$ and simulation of QCD multijet production $R^{\mathrm{MC}}_{\mathrm{P/F}}(m_{\mathrm{t}}, M_{\mathrm{tW}})$. The data-to-simulation ratio $R_{\mathrm{ratio}}(m_{\mathrm{t}}, M_{\mathrm{tW}})$ is taken and parametrized using polynomials in the $(m_{\mathrm{t}}, M_{\mathrm{tW}})$ plane. The prediction for QCD multijet production in each bin i in the signal region is then given as

$$n^{\mathrm{QCD}}_{\mathrm{P}}(i) = n^{\mathrm{QCD}}_{\mathrm{F}}(i) \cdot R^{\mathrm{MC}}_{\mathrm{P/F}}(m_{\mathrm{t}}, M_{\mathrm{tW}}) \cdot R_{\mathrm{ratio}}(m_{\mathrm{t}}, M_{\mathrm{tW}})\,. \tag{9.7}$$

Similarly to the ℓ+jets analysis, the $\mathrm{t\bar{t}}$ prediction is taken from simulation and constrained in a dedicated control region. Due to the higher sensitivity of the 2b category of the ℓ+jets analysis, the $\mathrm{t\bar{t}}$ control region of the all-hadronic analysis is not considered in the statistical combination.

9.7.2. Statistical Combination

The 1b and 2b categories of the ℓ+jets analysis are fit simultaneously with the signal and QCD multijet control regions of the all-hadronic analysis to data in a binned maximum likelihood fit, as described in Section 9.6. Systematic uncertainties shared between the two analyses are considered fully correlated. These uncertainties are marked with an asterisk ($*$) in Table 9.2.

Again, no significant deviations from the expected background are observed and upper limits on the cross section of $\mathrm{pp} \to \mathrm{b}^* \to \mathrm{tW}$ for the considered models are set at 95% CL. The resulting observed and expected limits are shown in Fig. 9.19. The

individual expected cross section limits from the all-hadronic (magenta) and ℓ+jets (blue) analyses are shown as colored dashed lines. The all-hadronic analysis is not sensitive to b^* masses below 1.4 TeV, due to the significantly higher trigger thresholds of the all-hadronic trigger selection. Thus the ℓ+jets analysis sets unique constraints in the range 0.7–1.4 TeV, resulting in a step in the combined limits at $M_{b^*} = 1.4\,\text{TeV}$. For b^* masses above 1.4 TeV, the individual analyses show comparable sensitivity.

The largest deviation of the observed limit is found at $M_{b^*} = 2.2\,\text{TeV}$, corresponding to a local significance of 2.2 standard deviations. However, the global significance is smaller than one standard deviation. Using the theoretical cross sections of the considered models, mass exclusion limits are set. The observed (expected) mass limits are 3.0 TeV (3.1 TeV), 3.0 TeV (3.2 TeV) and 3.2 TeV (3.4 TeV) for the LH, RH and VL models, respectively. These are the most stringent limits on these models to date.

Compared to the results of the previous analyses at 8 TeV by the ATLAS [5] and CMS [6] collaborations, the new expected mass exclusion limits are better by more than a factor of 2. This improvement results on the one hand from the larger dataset and higher center-of-mass energy, but also from improvements of analysis techniques, especially the identification of boosted hadronic decays of the W boson and t quark.

9.8. Interpretation as Vector-Like Quark Signal

The searched event signature is very similar to the signature expected for the production of a vector-like B quark decaying to tW. The B is produced in association with either a b quark or a t quark via the electroweak interaction (cf. Fig. 2.6). The final state features an additional jet with large rapidity (forward jet) originating from the scattered initial state quark. The associated b/t quark are generally produced with low p_{T} compared to the decay products of the B.

The associated b has generally a too small p_{T} to pass the b jet selection, so B events with an associated b quark still populate mostly the 1b category. The associated t quark has a too small p_{T} to be sufficiently boosted and identified as t jet. However,

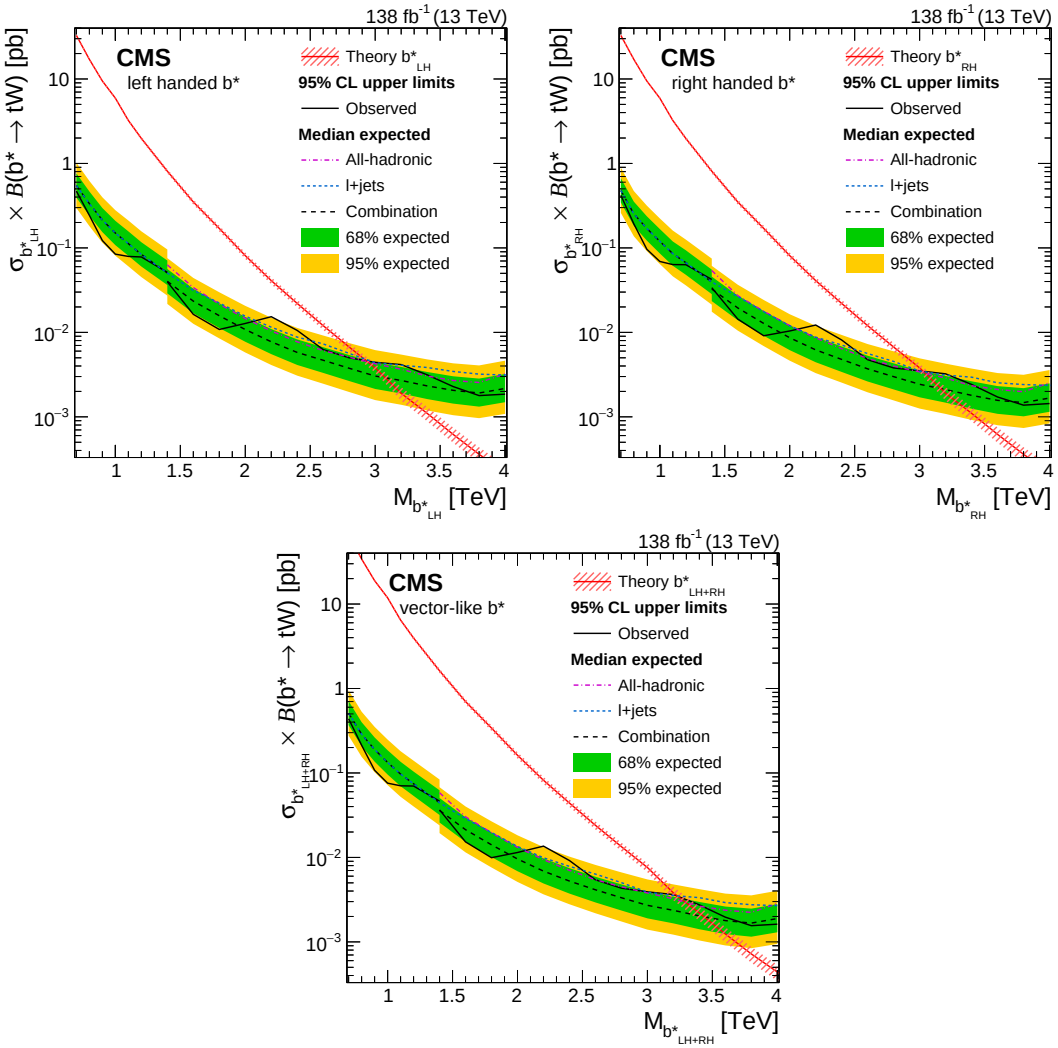

Figure 9.19.: Upper cross section limits on the process pp $\rightarrow$ b* $\rightarrow$ tW for the left-handed (top left), right-handed (top right) and vector-like (bottom) benchmark scenario at 95% CL obtained from the statistical combination of the all-hadronic and ℓ+jets channels. The expected and observed limits are shown as dashed and solid black lines, respectively. The colored dashed lines show the individual median expected limits of the all-hadronic (magenta) and ℓ+jets (blue) channels. The green and yellow bands indicate the 68% and 95% confidence intervals on the expected limits of the combination. The theoretically predicted cross section is shown as a red line, where the uncertainties due to missing higher orders are depicted by shaded areas. Published in Ref. [1].

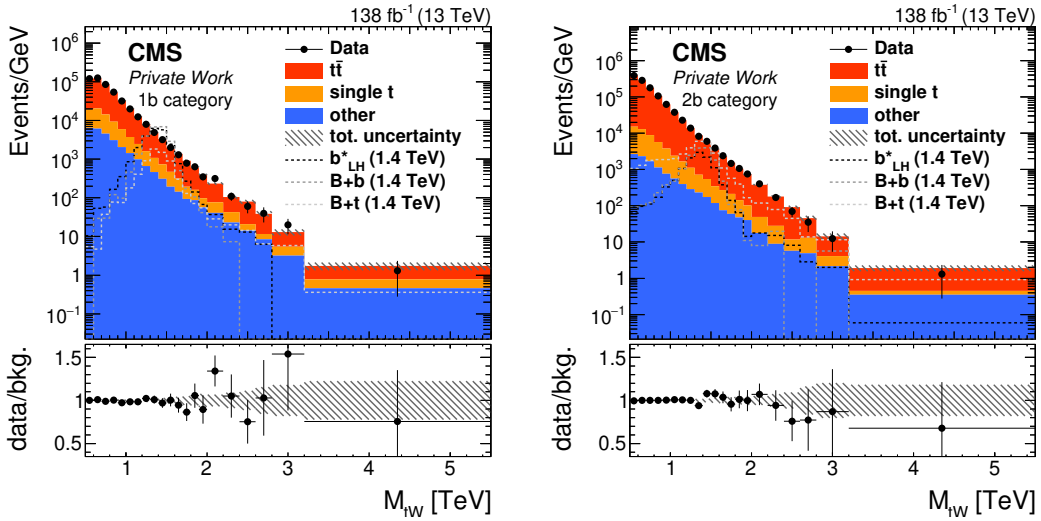

Figure 9.20.: Distributions of M_{tW} in the 1b (left) and 2b (right) categories after the fit to data. Three different signals with a mass of 1.4 TeV are shown as dashed lines: b* (black), B + b (dark grey) and B + t (light grey). All signals are scaled to a production cross section of 1 pb. The gray area in the bottom pannel corresponds to the total uncertainty on the background prediction.

the b quark from the decay of the associated t quark can receive enough p_{T} to be selected as a second b jet. Hence, B events with an associated t quark are more likely to be selected in the 2b category, lowering the sensitivity of the ℓ+jets analysis for this production mode. The post-fit M_{tW} distribution with the b* signal and both B signals for a signal mass of 1.4 TeV is shown in Fig. 9.20. The b* signal is drawn as a black dashed line while the signal for a the B with an associated b/t quark are drawn as dark grey and light grey lines, respectively. It can be seen that the b* and B + b distributions are very similar and they are selected dominantly in the 1b category, while the B + t events are suppressed in this category and are dominant in the 2b category.

The results of the analysis are also interpreted in the context of the production of a vector-like B by fitting the signal+background hypothesis again using the B samples as signal hypotheses. The fit is performed in the ℓ+jets and all-hadronic channels simultaneously. All systematic uncertainties are taken into account. Upper limits on the production cross section of single B production in association with a b/t quark in the B $\rightarrow$ tW decay channel are set at 95% CL over a B mass range from 0.7 to 1.8 TeV. The expected and observed limits are shown in Fig. 9.21. The observed upper limits range from 0.63 to 0.007 pb for the B+b production mode,

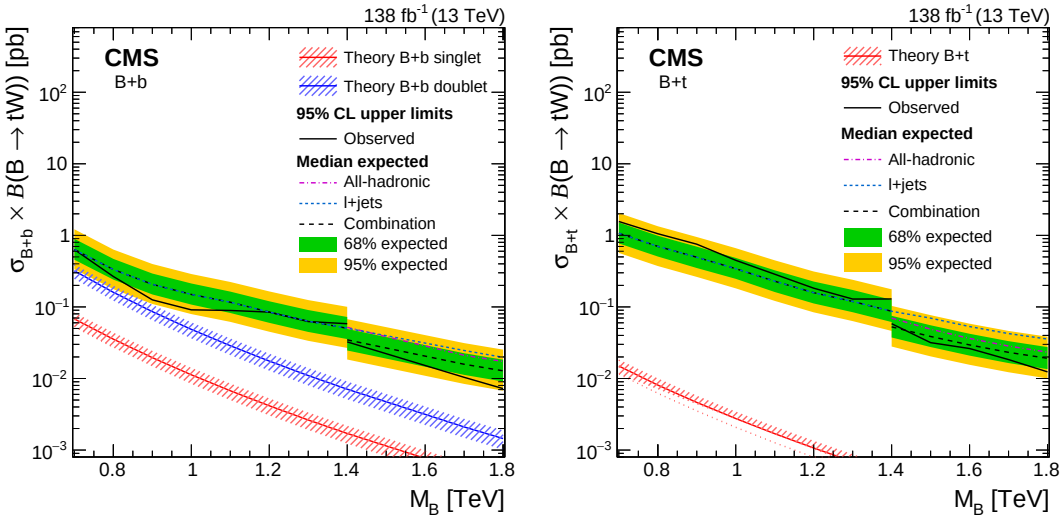

Figure 9.21.: Upper limits on the B production cross section with an associated b quark (left) or an associated t quark (right). The expected and observed limits are shown as dashed and solid lines, respectively. The green and yellow bands indicate the 68% and 95% confidence intervals on the expected limits. The theoretically predicted cross sections are shown as a red and blue lines, where the uncertainties due to missing higher orders are depicted by shaded areas. Published in Ref. [1].

and from 1.6 to 0.01 pb for the B+t production. The obtained cross section limits are compared to the theory predictions from Ref. [144]. Models with a VLQ singlet or VLQ doublet are considered in case of B+b production. For B+t production, both models result in the same production cross section. Because of the small cross sections for these processes, no limits on the B mass are set.

The upper cross section limits can be compared to a prior search for singly produced vector-like B quarks by the CMS collaboration using the 2016 dataset [71]. For the B+b production mode, the expected limits are 8–29% better at low B quark masses, where the ℓ+jets channel sets unique limits, and up to 75% better at the highest considered B quark mass. For the B+t production mode, the ℓ+jets channel is less sensitive, due to the additional b jet from the decay of the associated t quark, leading to weaker limits at low B quark masses compared to the prior search. At high B quark masses, this search improves the upper limits by up to 50%.

The presented analysis was not specifically optimized for the vector-like B quark signal. One possible optimization would be to include a selection based on the presence of an additional forward jet. A study of a possible improvement to the

signal sensitivity in the ℓ+jets channel using a selection of the forward jet has been conducted in the scope of a bachelor's thesis [3], that was supervised by the author of this thesis. The study is based on the 2016 dataset only. The event selection in this study follows the one described in Section 9.1. An additional event categorization is introduced, based on the presence of a forward jet. The forward jet is defined as an AK4 jet within $2.4 < |\eta| < 4.0$. The PUPPI algorithm was used to mitigate the influence of pileup on the forward jets. Events with and without a forward jet are then sorted into separate categories, similar to Section 9.3. The categories featuring a forward jet are considered signal categories, while the categories without a forward jet are used as control regions for the SM backgrounds. The background estimation follows the procedure described in Section 9.4. It was found that the additional forward jet selection could improve the sensitivity by up to 25% for a B+b signal. By including an additional requirement on the χ^2-like estimator in the 2b category for events with a forward jet, the sensitivity to a B+t signal could be improved by up to 60%.

10. Conclusion

In this thesis, a search for a heavy resonance decaying to tW in the ℓ+jets final state was presented. The analysis is based on data of pp collisions at a center-of-mass energy of 13 TeV recorded with the CMS detector in the years 2016, 2017 and 2018. The full dataset analyzed corresponds to an integrated luminosity of $138\,\mathrm{fb}^{-1}$.

The HOTVR algorithm is used for the first time in a data analysis at the LHC. The variable radius parameter of HOTVR allows to efficiently identify jets from the boosted hadronic decay of t quarks over a wide range of t momenta. In combination with the low trigger thresholds of isolated lepton triggers, signal masses down to 0.7 TeV can be probed, while keeping the best possible sensitivity at high masses, up to 4 TeV. The background arising from misidentified t jets is estimated from data. The modeling of the dominant $\mathrm{t\bar{t}}$ background is constrained using a dedicated control region in a fit to data.

A statistical combination with another analysis targeting the all-hadronic final state has been performed. The all-hadronic and ℓ+jets channels are fit to data simultaneously. No significant excess of data over the predicted SM background is observed. Exclusion limits on the b^* benchmark models are set, excluding the left-handed, right-handed and vector-like models up to masses of 2.95 TeV, 3.03 TeV and 3.22 TeV, respectively. These constitute the strongest limits on these models to date. In the range from 0.7 to 1.4 TeV, the presented analysis of the ℓ+jets final state is able to set unique limits on the b^* production cross section, due to its extended sensitivity to lower signal masses compared to the all-hadronic analysis.

The results of the statistical combination are also interpreted in the context of a vector-like B quark. The electroweak production processes with an associated b or t quark are considered. Upper limits are set on the cross sections of these processes, which improve the results of a previous search for almost all resonance masses.

The results presented in this thesis have been published by the CMS Collaboration in Ref. [1].

In the upcoming LHC Run 3 another $250\,\mathrm{fb}^{-1}$ of data are expected to be collected by the CMS Collaboration. Additionally, the center-of-mass energy will be increased up to 13.6 TeV. With the increased energy and larger dataset the presented results can however only be improved slightly. Further improvements can be gained by including the final state with a hadronically decaying W boson and leptonically decaying t quark in the ℓ+jets channel. Using jet substructure techniques, the hadronic decay of the W can be identified and the t quark can be reconstructed similarly to the W boson reconstruction in the presented analysis. However, because the charged lepton in this final state is not isolated, the contribution from the QCD multijet background can no longer be neglected, which makes the background estimation more challenging. Finally, the additional categorization based on the presence of a forward jet showed promising improvements of the sensitivity for the vector-like B quark.

A. HOTVR Control Distributions

In Fig. A.1 and Fig. A.2, distributions of the HOTVR substructure variables used for t tagging are shown. They are measured using events passing the event selection described in Section 9.1, without the t tagging requirement. Additionally, at least two b tagged AK4 jets are required, to supress the non-top backgrounds. The distributions are shown jets with for $200 < p_{\mathrm{T}} < 400\,\mathrm{GeV}$ (Fig. A.1) and $p_{\mathrm{T}} > 600\,\mathrm{GeV}$ (Fig. A.2). The $\mathrm{t\bar{t}}$ and single t events are each separated into three categories, depending on the number of generator level quarks from the hadronic t decay that are reconstructed within the jet. The generator level quarks are matched to the HOTVR jets using the effective distance parameter $R_{\mathrm{eff}} = \rho/p_{\mathrm{T}}$. If all three generator level quarks are matched to the HOTVR jet, it is considered a merged t jet. If instead only two generator quarks are matched to the HOTVR jet, it is considered a semi-merged t jet, and if one or no generator level quarks are matched to it, the jet is considered non-merged.

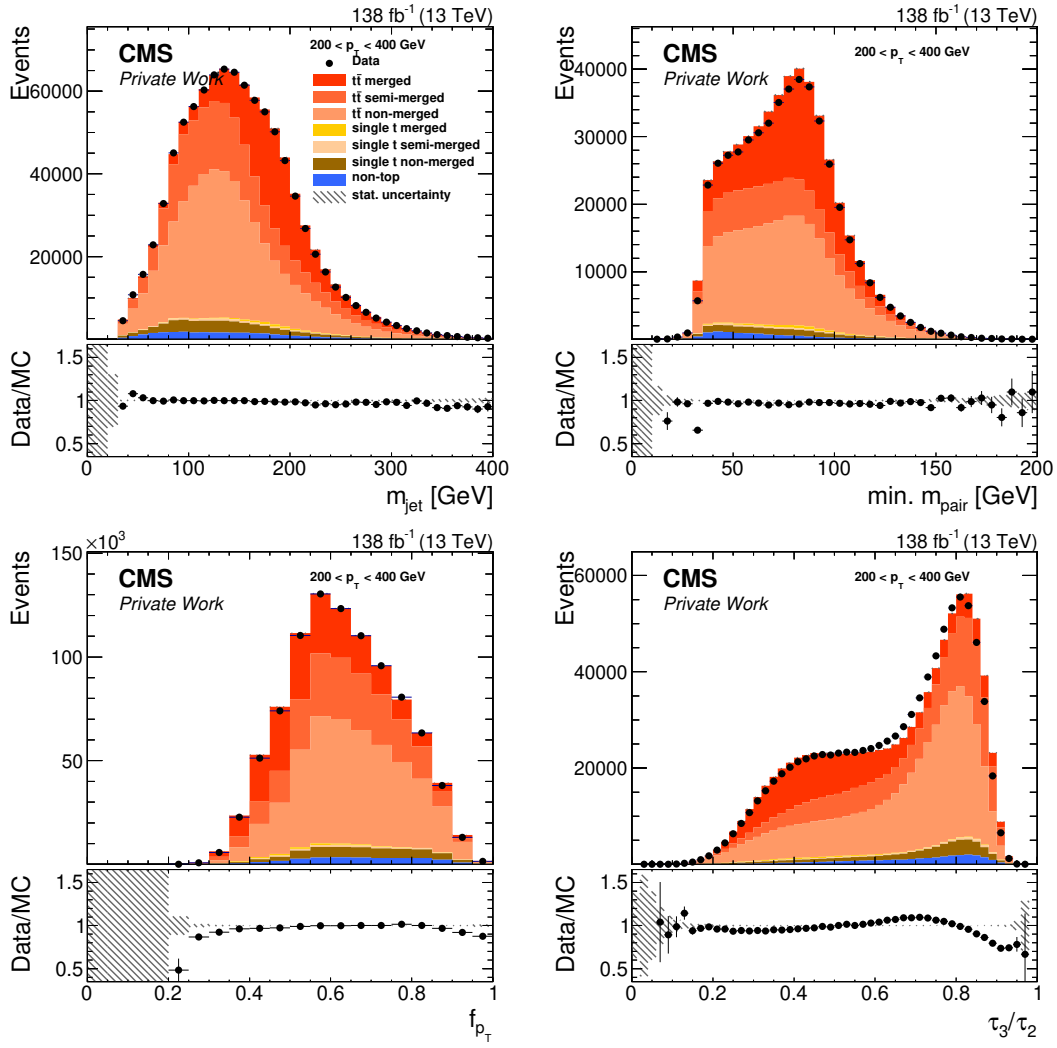

Figure A.1.: Distributions of (left-to-right, top-to-bottom): HOTVR jet mass, minimum m_{pair}, $f_{p_{\text{T}}}$, τ_3/τ_2 in a $\text{t}\bar{\text{t}}$ enriched region for HOTVR jets with $200 < p_{\text{T}} < 400\,\text{GeV}$. Data is shown as black markers, the error bars show the statistical uncertainty. The Expected SM background is drawn as stacked filled areas and the hatched area indicates the statistical uncertainty. The bottom panel in each plot shows the data-to-background ratio. The error bars correspond to the statistical uncertainty on the data and the gray area shows the statistical uncertainty on the background.

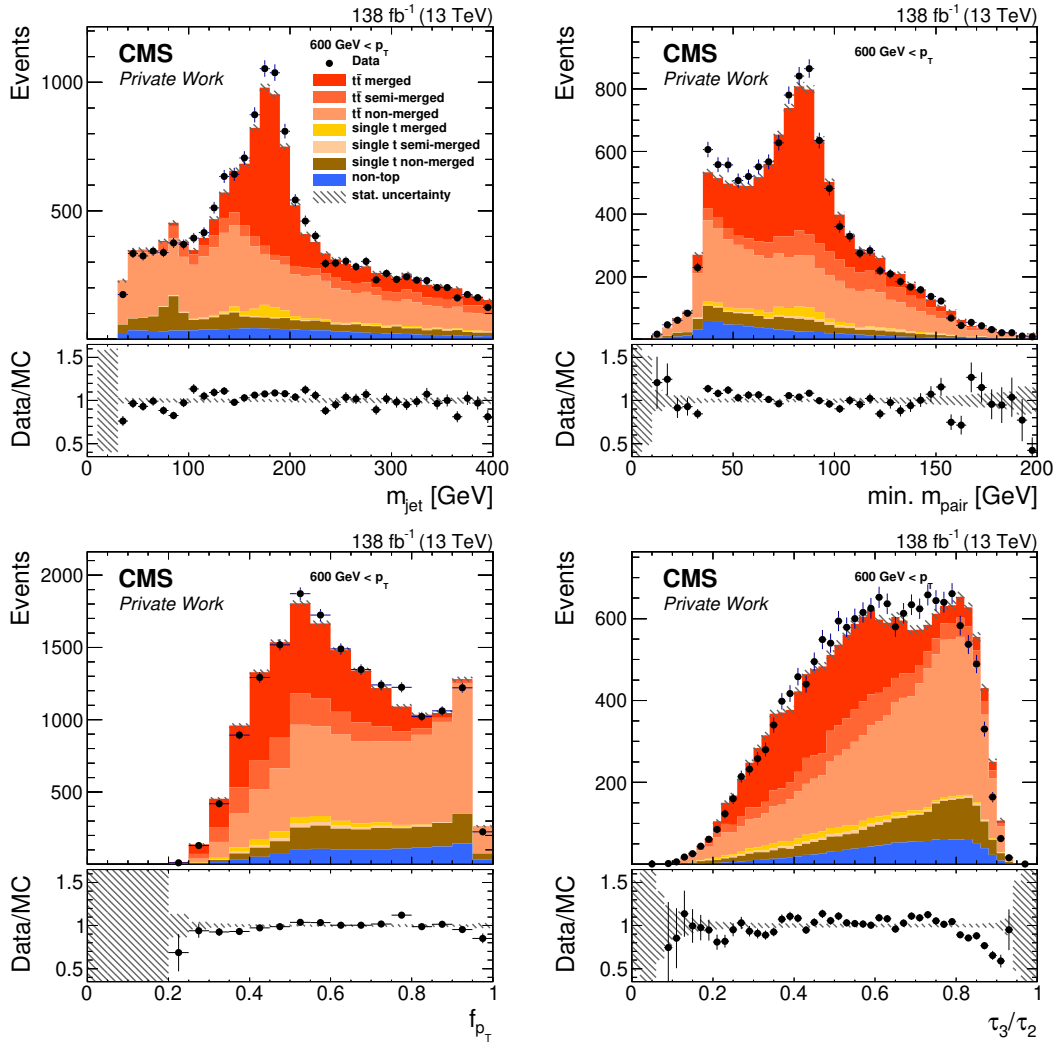

Figure A.2.: Distributions of (left-to-right, top-to-bottom): HOTVR jet mass, minimum m_{pair}, $f_{p_{\text{T}}}$, τ_3/τ_2 in a $\text{t}\bar{\text{t}}$ enriched region for HOTVR jets with $p_{\text{T}} > 600\,\text{GeV}$. Data is shown as black markers, the error bars show the statistical uncertainty. The Expected SM background is drawn as stacked filled areas and the hatched area indicates the statistical uncertainty. The bottom panel in each plot shows the data-to-background ratio. The error bars correspond to the statistical uncertainty on the data and the gray area shows the statistical uncertainty on the background.

Bibliography

[1] CMS Collaboration, "Search for a heavy resonance decaying into a top quark and a W boson in the lepton+jets final state at $\sqrt{s}$= 13 TeV", arXiv:2111.10216. Submitted to JHEP.

[2] CMS Collaboration, "Search for a heavy resonance decaying to a top quark and a W boson at $\sqrt{s} = 13$ TeV in the fully hadronic final state", arXiv:2104.12853.

[3] H. Koser, "Search for singly produced vector-like bottom quarks with the cms detector", Bachelor's thesis, Universität Hamburg, 2021.

[4] ATLAS Collaboration, "Search for single b^*-quark production with the ATLAS detector at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* **721** (2013) 171–189, doi:10.1016/j.physletb.2013.03.016, arXiv:1301.1583.

[5] 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, doi:10.1007/JHEP02(2016)110, arXiv:1510.02664.

[6] CMS Collaboration, "Search for the production of an excited bottom quark decaying to tW in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JHEP* **01** (2016) 166, doi:10.1007/JHEP01(2016)166, arXiv:1509.08141.

[7] T. Lapsien, R. Kogler, and J. Haller, "A new tagger for hadronically decaying heavy particles at the LHC", *Eur. Phys. J. C* **76** (2016), no. 11, 600, doi:10.1140/epjc/s10052-016-4443-8, arXiv:1606.04961.

[8] D. Griffiths, "Introduction to elementary particles". 2008. ISBN 978-3-527-40601-2.

[9] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory". Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.

[10] M. Thomson, "Modern particle physics". Cambridge University Press, New York, 2013. ISBN 978-1-107-03426-6.

[11] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531–533, doi:10.1103/PhysRevLett.10.531.

[12] 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.

[13] Particle Data Group Collaboration, "Review of Particle Physics", *PTEP* **2020** (2020), no. 8, 083C01, doi:10.1093/ptep/ptaa104.

[14] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133, doi:10.1016/0031-9163(64)91136-9.

[15] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[16] 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.

[17] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[18] A. Salam and J. C. Ward, "Electromagnetic and weak interactions", *Phys. Lett.* **13** (1964) 168–171, doi:10.1016/0031-9163(64)90711-5.

[19] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[20] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[21] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[22] ATLAS, CDF, CMS, D0 Collaboration, "First combination of Tevatron and LHC measurements of the top-quark mass", arXiv:1403.4427.

[23] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

[24] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.

[25] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017), no. 10, 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[26] R. K. Ellis et al., "Perturbation Theory and the Parton Model in QCD", *Nucl. Phys. B* **152** (1979) 285–329, doi:10.1016/0550-3213(79)90105-6.

[27] 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, doi:10.1142/9789814503266_0001, arXiv:hep-ph/0409313.

[28] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[29] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys.*

Commun. **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[30] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[31] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[32] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[33] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[34] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[35] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[36] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[37] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.

[38] CMS Collaboration, "Event generator tunes obtained from underlying event

and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[39] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020), no. 1, 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[40] GEANT4 Collaboration, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[41] E. Corbelli and P. Salucci, "The Extended Rotation Curve and the Dark Matter Halo of M33", *Mon. Not. Roy. Astron. Soc.* **311** (2000) 441–447, doi:10.1046/j.1365-8711.2000.03075.x, arXiv:astro-ph/9909252.

[42] V. Trimble, "Existence and Nature of Dark Matter in the Universe", *Ann. Rev. Astron. Astrophys.* **25** (1987) 425–472, doi:10.1146/annurev.aa.25.090187.002233.

[43] WMAP Collaboration, "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation", *Astrophys. J. Suppl.* **192** (2011) 18, doi:10.1088/0067-0049/192/2/18, arXiv:1001.4538.

[44] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters", *Astron. Astrophys.* **594** (2016) A13, doi:10.1051/0004-6361/201525830, arXiv:1502.01589.

[45] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.

[46] LHCb Collaboration, "Test of lepton universality in beauty-quark decays", arXiv:2103.11769.

[47] Muon g-2 Collaboration, "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm", *Phys. Rev. Lett.* **126** (2021), no. 14, 141801, doi:10.1103/PhysRevLett.126.141801, arXiv:2104.03281.

[48] H. Georgi and S. L. Glashow, "Unity of All Elementary Particle Forces", *Phys. Rev. Lett.* **32** (1974) 438–441, doi:10.1103/PhysRevLett.32.438.

[49] H. Harari, "Composite Models for Quarks and Leptons", *Phys. Rept.* **104** (1984) 159, doi:10.1016/0370-1573(84)90207-2.

[50] 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.

[51] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[52] T. M. P. Tait and C. P. Yuan, "Single top quark production as a window to physics beyond the standard model", *Phys. Rev. D* **63** (2000) 014018, doi:10.1103/PhysRevD.63.014018, arXiv:hep-ph/0007298.

[53] J. Nutter, R. Schwienhorst, D. G. E. Walker, and J.-H. Yu, "Single Top Production as a Probe of B-prime Quarks", *Phys. Rev. D* **86** (2012) 094006, doi:10.1103/PhysRevD.86.094006, arXiv:1207.5179.

[54] C. Cheung, A. L. Fitzpatrick, and L. Randall, "Sequestering CP Violation and GIM-Violation with Warped Extra Dimensions", *JHEP* **01** (2008) 069, doi:10.1088/1126-6708/2008/01/069, arXiv:0711.4421.

[55] A. L. Fitzpatrick, G. Perez, and L. Randall, "Flavor anarchy in a Randall-Sundrum model with 5D minimal flavor violation and a low Kaluza-Klein scale", *Phys. Rev. Lett.* **100** (2008) 171604, doi:10.1103/PhysRevLett.100.171604, arXiv:0710.1869.

[56] N. Vignaroli, "Discovering the composite Higgs through the decay of a heavy fermion", *JHEP* **07** (2012) 158, doi:10.1007/JHEP07(2012)158, arXiv:1204.0468.

[57] N. Vignaroli, "Δ F=1 constraints on composite Higgs models with LR parity",

Phys. Rev. D **86** (2012) 115011, doi:10.1103/PhysRevD.86.115011, arXiv:1204.0478.

[58] A. De Rujula, L. Maiani, and R. Petronzio, "Search for Excited Quarks", *Phys. Lett. B* **140** (1984) 253–258, doi:10.1016/0370-2693(84)90930-4.

[59] U. Baur, I. Hinchliffe, and D. Zeppenfeld, "Excited Quark Production at Hadron Colliders", *Int. J. Mod. Phys. A* **2** (1987) 1285, doi:10.1142/S0217751X87000661.

[60] A. Djouadi and A. Lenz, "Sealing the fate of a fourth generation of fermions", *Phys. Lett. B* **715** (2012) 310–314, doi:10.1016/j.physletb.2012.07.060, arXiv:1204.1252.

[61] O. Eberhardt et al., "Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations", *Phys. Rev. Lett.* **109** (2012) 241802, doi:10.1103/PhysRevLett.109.241802, arXiv:1209.1101.

[62] G. Dissertori, E. Furlan, F. Moortgat, and P. Nef, "Discovery potential of top-partners in a realistic composite Higgs model with early LHC data", *JHEP* **09** (2010) 019, doi:10.1007/JHEP09(2010)019, arXiv:1005.4414.

[63] O. Matsedonskyi, G. Panico, and A. Wulzer, "Light Top Partners for a Light Composite Higgs", *JHEP* **01** (2013) 164, doi:10.1007/JHEP01(2013)164, arXiv:1204.6333.

[64] B. Bellazzini, C. Csáki, and J. Serra, "Composite Higgses", *Eur. Phys. J. C* **74** (2014), no. 5, 2766, doi:10.1140/epjc/s10052-014-2766-x, arXiv:1401.2457.

[65] I. Antoniadis, K. Benakli, and M. Quiros, "Finite Higgs mass without supersymmetry", *New J. Phys.* **3** (2001) 20, doi:10.1088/1367-2630/3/1/320, arXiv:hep-th/0108005.

[66] R. Contino, L. Da Rold, and A. Pomarol, "Light custodians in natural

composite Higgs models", *Phys. Rev. D* **75** (2007) 055014, doi:10.1103/PhysRevD.75.055014, arXiv:hep-ph/0612048.

[67] J. L. Hewett and T. G. Rizzo, "Low-Energy Phenomenology of Superstring Inspired E(6) Models", *Phys. Rept.* **183** (1989) 193, doi:10.1016/0370-1573(89)90071-9.

[68] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria, "Handbook of vectorlike quarks: Mixing and single production", *Phys. Rev. D* **88** (2013), no. 9, 094010, doi:10.1103/PhysRevD.88.094010, arXiv:1306.0572.

[69] M. Buchkremer, G. Cacciapaglia, A. Deandrea, and L. Panizzi, "Model Independent Framework for Searches of Top Partners", *Nucl. Phys. B* **876** (2013) 376–417, doi:10.1016/j.nuclphysb.2013.08.010, arXiv:1305.4172.

[70] A. De Simone, O. Matsedonskyi, R. Rattazzi, and A. Wulzer, "A First Top Partner Hunter's Guide", *JHEP* **04** (2013) 004, doi:10.1007/JHEP04(2013)004, arXiv:1211.5663.

[71] CMS Collaboration, "Search for single production of vector-like quarks decaying to a top quark and a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **79** (2019) 90, doi:10.1140/epjc/s10052-019-6556-3, arXiv:1809.08597.

[72] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[73] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.

[74] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron

collisions", *Nucl. Phys. B* **406** (1993) 187–224, doi:10.1016/0550-3213(93)90166-M.

[75] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[76] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

[77] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*. 1998. arXiv:hep-ph/9907280.

[78] D. Krohn, J. Thaler, and L.-T. Wang, "Jets with Variable R", *JHEP* **06** (2009) 059, doi:10.1088/1126-6708/2009/06/059, arXiv:0903.0392.

[79] R. Kogler, "Advances in Jet Substructure at the LHC". Springer International Publishing, 2021. doi:10.1007/978-3-030-72858-8, ISBN 978-3-030-72858-8.

[80] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure", *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.

[81] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[82] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *JINST* **15** (2020), no. 06, P06005, doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.

[83] CMS Collaboration, "Measurements of the differential jet cross section as a function of the jet mass in dijet events from proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2018) 113, doi:10.1007/JHEP11(2018)113, arXiv:1807.05974.

[84] M. Stoll, "Vetoed jet clustering: The mass-jump algorithm", *JHEP* **04** (2015) 111, doi:10.1007/JHEP04(2015)111, arXiv:1410.4637.

[85] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.

[86] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* **02** (2012) 093, doi:10.1007/JHEP02(2012)093, arXiv:1108.2701.

[87] L. Evans and P. Bryant, eds., "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[88] E. Mobs, "The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019", 2019. https://cds.cern.ch/record/2684277, General Photo.

[89] CMS Collaboration, "CMS Luminosity - Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, Accessed: 18.01.2021.

[90] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[91] CMS Collaboration, G. L. Bayatian et al., "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software", 2006.

[92] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment", *J. Phys. Conf. Ser.* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032, arXiv:1311.4942.

[93] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[94] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade", 2012. doi:10.2172/1151650.

[95] CMS Tracker Group Collaboration, "The CMS Phase-1 Pixel Detector Upgrade", *JINST* **16** (2021), no. 02, P02027, doi:10.1088/1748-0221/16/02/P02027, arXiv:2012.14304.

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

[97] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[98] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[99] CMS Collaboration, "The CMS Phase-1 pixel detector - experience and lessons learned from two years of operation", *JINST* **14** (2019), no. 07, C07008, doi:10.1088/1748-0221/14/07/C07008.

[100] CMS Collaboration Collaboration, "Phase-I Pixel ROC Thresholds and DC-DC Converter Characterization in 2018", 2020. https://cds.cern.ch/record/2714878.

[101] J.-M. André et al., "Presentation layer of CMS Online Monitoring System", *EPJ Web Conf.* **214** (2019) 01044, doi:10.1051/epjconf/201921401044.

[102] W. Badgett et al., "CMS OnlineWeb-Based Monitoring", *Phys. Procedia* **37** (2012) 1869–1875, doi:10.1016/j.phpro.2012.02.510.

[103] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track reconstruction in the CMS tracker", 2005.

[104] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) =

8 TeV", *JINST* **10** (2015), no. 08, P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[105] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[106] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020), no. 09, P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.

[107] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *IEEE Proc.* **86** (1998), no. 11, 2210–2239, doi:10.1109/5.726788.

[108] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys. G* **34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.

[109] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *JINST* **16** (2021), no. 05, P05014, doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.

[110] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC", *eConf* **C0303241** (2003) TULT009, doi:10.1088/0954-3899/31/9/N01, arXiv:physics/0306087.

[111] CMS Collaboration, "Jet algorithms performance in 13 TeV data", 2017.

[112] CMS Collaboration Collaboration, "Pileup Removal Algorithms", 2014. http://cds.cern.ch/record/1751454.

[113] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[114] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett.*

B **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[115] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.

[116] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[117] CMS Collaboration Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018", 2020. https://cds.cern.ch/record/2715872.

[118] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[119] CMS Collaboration Collaboration, "Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector", 2018. https://cds.cern.ch/record/2627468.

[120] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019), no. 07, P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[121] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.

[122] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[123] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking

period at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.

[124] CMS Collaboration, "Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **16** (2021) P07001, doi:10.1088/1748-0221/16/07/P07001, arXiv:2102.04790.

[125] A. Reimers, "Searching for Heavy Particles Coupled to Top Quarks with CMS". PhD thesis, Hamburg U., Hamburg, 2020. doi:10.3204/PUBDB-2020-02589.

[126] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[127] 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", *Phys. Rev. D* **95** (2017), no. 9, 092001, doi:10.1103/PhysRevD.95.092001, arXiv:1610.04191.

[128] CMS Collaboration, "Measurements of $\mathrm{t\bar{t}}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons", *JHEP* **02** (2019) 149, doi:10.1007/JHEP02(2019)149, arXiv:1811.06625.

[129] M. Aliev et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034–1046, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.

[130] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-", *Phys. Rev. D* **82** (2010) 054018, doi:10.1103/PhysRevD.82.054018, arXiv:1005.4451.

[131] N. Kidonakis, "Top Quark Production", in *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*. 2013. arXiv:1311.0283. doi:10.3204/DESY-PROC-2013-03/Kidonakis.

[132] S. Kallweit et al., "NLO QCD+EW predictions for V + jets including off-shell

vector-boson decays and multijet merging", *JHEP* **04** (2016) 021, doi:10.1007/JHEP04(2016)021, arXiv:1511.08692.

[133] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **15** (2020), no. 10, P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[134] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", 2016.

[135] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[136] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2018) 161, doi:10.1007/JHEP07(2018)161, arXiv:1802.02613.

[137] D. Schwarz, "Measurements of jet mass, top quark mass and top tagging efficiencies in decays of boosted top quarks at CMS". PhD thesis, Hamburg U., 2020.

[138] CMS Collaboration, "Search for massive resonances decaying into WW, WZ or ZZ bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **03** (2017) 162, doi:10.1007/JHEP03(2017)162, arXiv:1612.09159.

[139] CMS Collaboration, "Supplementary Material of CMS-B2G-20-010". http://cms-results.web.cern.ch/cms-results/public-results/publications/B2G-20-010/.

[140] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.

[141] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", in *PHYSTAT 2011*. 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.

[142] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219–228, doi:10.1016/0010-4655(93)90005-W.

[143] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[144] A. Carvalho et al., "Single production of vector-like quarks with large width at the Large Hadron Collider", *Phys. Rev. D* **98** (2018) 015029, doi:10.1103/PhysRevD.98.015029, arXiv:1805.06402.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen bedanken, die mich während der Erstellung dieser Arbeit und während meines Studiums unterstützt haben.

Zunächst möchte ich Johannes Haller dafür danken, dass er mir diese Arbeit in seiner Arbeitsgruppe ermöglicht hat. Vielen Dank auch für die vielen Diskussionen, Tipps und Anregungen zu meiner Forschung, sowie die vielen Gelegenheiten, meine Ergebnisse auf Konferenzen und Workshops zu präsentieren.

Ich danke Christian Schwanenberger für die Übernahme des Zweitgutachtens dieser Arbeit. Weiter bedanke ich mich bei allen Mitgliedern der Prüfungskommission, Jochen Liske, Johannes Haller, Christian Schwanenberger, Elisabetta Gallo und Kai Schmidt-Hoberg, für die Beurteilung meiner Disputation.

Roman Kogler möchte ich für die zahlreichen Diskussionen und Ratschläge, und für die tatkräftige Unterstützung bei der Veröffentlichung dieser Analyse, sowie für das Lesen dieser Arbeit danken.

Klaas Padeken, Benedikt Vormwald, Jory Sonneveld und allen Anderen vom Pixel Operations Team möchte ich für die herzliche Aufnahme ins Team, die vielen Erfahrungen und die schöne Zeit am CERN danken.

I'd like to thank all members of our working group for the nice working atmosphere, lots of coffee on the sofa and many Friday beers. I'd like to specially thank my office mates Andrea Malara, Arne Reimers and Irene Zoi, as well as my DPG flatmate and Maria-Laach collegue Dennis Schwarz for all the fun time! It was a real pleasure to be part of the working group.

Ganz besonders möchte ich Christina für die mentale Unterstützung in den letzten

Jahren danken. Du warst immer für mich da und hast mich wieder aufgemuntert, wenn es mal nicht so gut lief.

Zum Schluss möchte ich ganz herzlich meinen Freunden und meiner Familie, und ganz besonders meinen Eltern für die Unterstützung und Motivation während der letzten Jahre danken. Ich konnte mich immer darauf verlassen bei euch ein offenes Ohr und einen guten Rat zu finden.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Ort, Datum

Alexander Fröhlich