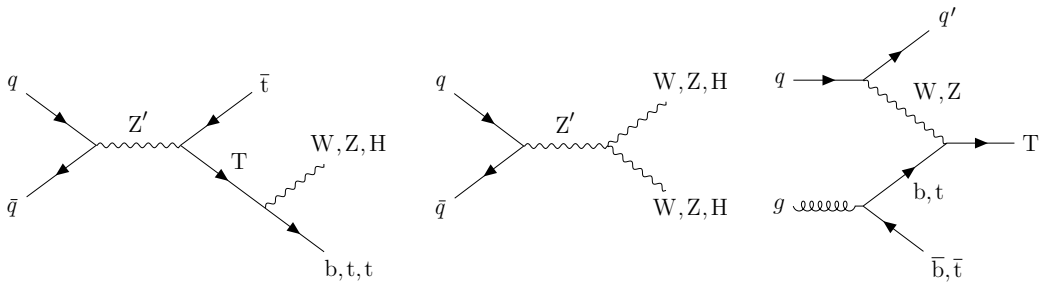
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Searches for new heavy bosons and vector-like quarks with the CMS experiment at $\sqrt{s} = 13\,\text{TeV}$ and novel pileup mitigation techniques

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

vorgelegt von
Anna Benecke

Hamburg 2020



Gutachter/innen der Dissertation:	Dr. Andreas Hinzmann Prof. Dr. Johannes Haller Prof. Dr. Ariel G. Schwartzman
Zusammensetzung der Prüfungskommission:	Prof. Dr. Elisabetta Gallo Prof. Dr. Joahnnes Haller Dr. Andreas Hinzmann Prof. Dr. Dieter Horns Dr. Kai Schmidt-Hoberg
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Dieter Horns
Datum der Disputation:	15.12.2020
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang Hansen
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Abstract

In this thesis two searches for new physics are presented, involving a heavy resonance Z' and a vector-like quark T. Both new particles are predicted by theories beyond the standard model that address the smallness of the Higgs boson mass compared to the Planck scale. The searches are based on proton-proton collision data at a center of mass energy of 13 TeV recorded with the CMS experiment at the LHC.

In the first analysis, a Z' decaying to Tt is searched for in the lepton+jets final state for the first time. Data corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$ are analyzed. The three T decay channels Ht, Zt, and Wb are taken into account. The analysis is performed in the highly Lorentz-boosted regime where substructure techniques are used to identify the heavy bosons. This search leads to the most stringent upper cross section limits on a $Z' \to tT$ resonance to date. A heavy gluon can be excluded between a mass of 1.5 and 2.3 TeV if $M_T = 1.2\,\text{TeV}$, and between 2.0 and 2.5 TeV if $M_T = 1.5\,\text{TeV}$.

The second analysis considers a singly produced vector-like T that decays into Ht in the lepton+jets final state. While vector-like Ts have been excluded up to a mass of 1.3 TeV, additional decay channels can weaken this bound considerably. The analysis is designed to achieve sensitivity for resonances with masses down to 600 GeV. The T is reconstructed using three jets originating from the fragmentation of b quarks, the lepton and missing transverse momentum. A resonant structure on a smoothly falling background is searched for. The analysis is carried out using data corresponding to $137.2\,\text{fb}^{-1}$. The signal region is still blinded, because the analysis is still undergoing the CMS internal review. But the expected sensitivity corresponds to a significance of five standard deviations for a possible signal with mass of 650 GeV, visible in the all-hadronic final state.

The analysis of proton-proton collision data is impaired by particles that originate from additional proton-proton interactions during the same bunch crossing (pileup). Pileup mitigation is important for analyses of data with high instantaneous luminosities, which have been reached during Run 2 (2016-2018). Even higher levels of pileup are expected in future data acquisition periods. The performance of pileup mitigation techniques, including the novel pileup per particle identification (PUPPI) algorithm, is studied for up to 70 simultaneous collisions per bunch crossing. In addition, the validation in data and improvements of the PUPPI algorithm are shown. Following these studies, PUPPI has become the default pileup mitigation technique in CMS.

In order to improve the data quality and sensitivity of the CMS experiment further, the CMS pixel detector was upgraded in 2016/2017 and a new powering system with

DC-DC converters was installed. A failure of several DC-DC converters during the data taking in 2017 caused losses in the data quality. The systematic analysis of this failure presented here led to a change in the operation of the pixel detector in 2018, preventing losses due to broken DC-DC converters during data acquisition.

This thesis presents significant advancements in the search for Z' and T, pileup mitigation techniques ensuring future p p collision physics at the LHC and a failure analysis of the pixel powering system that led to successful acquisition of $80\,\mathrm{fb}^{-1}$ in 2018.

Zusammenfassung

In dieser Arbeit werden zunächst zwei Suchen nach einer schweren Resonanz Z' und vektorartigen Quarks T mit dem CMS Experiment präsentiert. Die zwei neuen Teilchen werden von Theorien vorhergesagt, die das Standard Modell erweitern und sich mit der im Vergleich zur Planck-Skala niedrigen Masse des Higgs Bosons befassen. Die Suchen basieren auf Proton-Proton-Kollisionsdaten bei einer Schwerpunktsenergie von 13 TeV, die mit dem CMS Experiment am LHC aufgezeichnet wurden.

In der ersten Analyse wird nach einem Z' gesucht, welches in Tt zerfällt. Die Suche wird zum ersten Mal im Endzustand mit einem Lepton und Jets durchgeführt und analysiert Daten, die einer integrierten Luminosität von $35.9\,\mathrm{fb}^{-1}$ entsprechen. Dabei werden die drei Zerfallskanäle des Ts, Ht, Zt und Wb, berücksichtigt. Die schweren Bosonen in dieser Analyse werden mit Substrukturtechniken identifiziert, da diese einen hohen Loorentz-Boost haben. Das Ergebnis sind die niedrigsten bisher ermittelten oberen Grenzen auf den Produktionswirkungsquerschnitt einer Resonanz $\mathrm{Z}' \to \mathrm{tT}$. Ein schweres Gluon kann ausgeschlossen werden mit einer Masse zwischen 1.5 und 2.3 TeV, wenn $M_\mathrm{T} = 1.2\,\mathrm{TeV}$ ist, und zwischen 2.0 und 2.5 TeV, wenn $M_\mathrm{T} = 1.5\,\mathrm{TeV}$ ist.

Die zweite Analyse betrachtet ein einzeln produziertes vektorartiges T, welches in Ht zerfällt. Der betrachtete Endzustand enthält ebenfalls ein Lepton und Jets. Vektorartige Ts sind bis zu einer Masse von 1.3 TeV ausgeschlossen. Dieses Massenlimit wird jedoch signifikant abgeschächt, falls es mehr als drei mögliche Zerfallskanäle geben sollte. Die Analyse wurde daraufhin optimiert Sensitivität bis zu einer Masse von 600 GeV zu erreichen und sucht nach einer resonanten Struktur auf einem monoton abfallenden Untergrund. Das T wird mithilfe von drei Jets aus der Fragmentierung der b Quarks, dem Lepton und fehlender transversaler Energie rekonstruiert. Insgesamt werden $137.2\,\mathrm{fb}^{-1}$ an Daten analysiert. Die Daten in der Signalregion werden noch nicht betrachtet, da sich diese Analyse noch im CMS internen Review befindet. Es wird aber gezeigt, dass die erwartete Sensitivität für ein mögliches Signal mit einer Masse von 650 GeV, welches im Endzustand mit Jets sichtbar ist, einer Signifikanz von fünf Standardabweichungen entspricht.

Die Analyse von Proton-Proton Kollisionsdaten beeinträchtigt durch Teilchen, die bei zusätzlichen Proton-Proton Interaktionen während einer Strahlkreuzung passieren (Pileup). Die Verminderung des Effekts von Pileup ist besonders wichtig, wenn Daten bei hoher instantaner Luminosität analysiert werden, welche bereits während Run 2 (2016-2018) erreicht wurden. Noch höhere Pileup-Szenarios werden für zukünftige Datennahmen erwartet. Es wird eine Studie mit allen in der CMS Kollaboration genutzten Techniken zur Minderung dieses Effekts präsentiert um die Performance

für bis zu 70 gleichzeitige Kollisionen pro Strahlkreuzung zu vergleichen. Dies schließt auch eine Validierung des neuartigen pileup per particle identification (PUPPI) Algorithmus in Daten mit ein. Zusätzlich wird eine Verbesserung des PUPPI Algorithmus präsentiert, aufgrund welcher der Algorithmus von nun an der Standard in der CMS Kollaboration ist.

Um die Datenqualität und die Sensitivität des CMS Experimentes zu verbessern wurde der CMS Pixeldetektor während des Jahreswechsel 2016/2017 erneuert und eine neue Stromversorgung mit Hilfe von DC-DC Konvertern eingebaut. Das Fehlverhalten vieler DC-DC Konverter während der Datennahme in 2017 führte zu einer verminderten Datenqualität. In dieser Arbeit wird eine systematische Analyse des Fehlverhaltens beschrieben, dessen Ergebnisse zu einem anderen Betriebsmodus in 2018 führten. Dadurch ist in 2018 kein DC-DC Konverter ausgefallen.

Diese Arbeit präsentiert signifikante Fortschritte in der Suche nach Z' und T, der Verminderung des Effektes von Pileup, welches zukünftige Physik mit p p Kollisionen am LHC sicherstellt und einer Fehleranalyse der Stromversorgung des Pixeldetektors, welche zu einer erfolgreichen Datenahme von $80\,\mathrm{fb}^{-1}$ in 2018 geführt hat.

List of Own Contributions

Search for a heavy spin-1 resonance decaying into a vector-like quark and a top quark

I have been the principle analyzer for this publication. My contributions include the following points.

- the study of beyond the standard model signals and simulation with MadGraph,
- production of signal simulation with a new underlying theory model,
- development of the search strategy including the classification of large-cone jets originating from H, W/Z and t quarks and the reconstruction of the invariant mass of the heavy resonance,
- measurements of the efficiencies and misidentification rates of the classification of large-cone jets,
- derivation of data-to-simulation correction factors for the classification of large-cone jets, and
- the statistical interpretation of the results, including the calculation of the predicted production cross section of the signal.

I was the contact person, responsible for the for the CMS-internal review and peer-review of the journal, including:

- regular updates in working group meetings,
- the preapproval and approval presentations of this analysis,
- writing and editing of the paper draft, and
- responses to the questions from the journal.

This effort resulted in a publication in EPJC [1]. In addition, I supervised a bachelor thesis [2] on this topic which main goal was to improve the sensitivity for the Wb decay mode.
Under the supervision of Prof. Dr. Johannes Haller, Dr. Andreas Hinzmann and Dr. Roman Kogler.

The CMS Pixel Detector Upgrade

From October 2017 to April 2018 I was part of the pixel operation team and responsible for the operation, calibration and commissioning of the barrel pixel detector. This includes the following tasks.

- I was part of the extraction and reinstallation team of the pixel detector during the year-end-technical-stop 2017/2018.
- As part of the commissioning team, I performed calibrations and measurements of the noise thresholds.
- During this period I performed measurements on the detector while it was in operation during data taking periods.
- I measured the current-voltage characteristic curves of around 800 DC-DC converters that were extracted from the detector.
- I categorized all 1200 DC-DCs into normal operating, malfunction and broken, based on a statistical analysis of all current-voltage characteristic curves.

This categorisation was necessary to understand the dimension of the problem and to investigate the reason for the DC-DC failure. This statistical analysis and the threshold measurement resulted in a detector performance note published by CMS [3].
Under the supervision of Dr. Stefanos Leontsinis, in collaboration with Dr. Klaas Padeken, Dr. Jory Sonnefeld and Dr. Benedikt Vormwald.

Pileup Mitigation Techniques

I performed detailed performance studies of the pileup mitigation techniques used within the CMS Collaboration. My contributions include the following points.

- The analysis of the key variables entering the pileup mitigation algorithm.

- The comparison of pileup mitigation algorithms in terms of efficiency, purity and pileup jet rate.

- A study of the correlation between the number of vertices and the number of interactions simulated.

All these studies entered the first pileup mitigation techniques paper with 2016 data published in JINST [4]. I was the contact person of this publication and responsible for the CMS-internal review and peer-review of the journal, including:

- the coordination of all different analyses that are part of the paper,

- regular updates in working group meetings,

- the preapproval and approval presentations of this analysis,

- writing and editing of the paper draft, and

- responses to the questions from the journal.

With this publication the new pileup mitigation algorithm, called PUPPI, was commissioned and will be the default pileup mitigation algorithm in use by the CMS Collaboration for future operations of the LHC. Based on the commissioning, I improved the PUPPI algorithm in terms of jet energy resolution, resulting in a decrease by more than 20% in several detector regions. For improvements of the PUPPI algorithm I supervised Ksenia de Leo to analyse the new tunes I produced.
Under the supervision of Dr. Andreas Hinzmann.

Studies on Decorrelated Taggers in the Context of a Search for Diboson Resonances

I was responsible for the choice of the classification technique used in the diboson resonance search in the all-hadronic final state. This includes the following tasks.

- I performed a study comparing all the available classification techniques of hadronically decaying high Lorentz-boosted objects.

- I have investigated the decorrelation methods important for the multidimensional fit to search for a new heavy resonance decaying into a pair of bosons.

- With these studies it was decided which classification technique and decorrelation technique is used in the analysis and how the prioritisation of the categories is done which resulted in an improvement of up to 30%.

Under the supervision of Dr. Andreas Hinzmann, in collaboration with Dr. Clemens Lange and Dr. Jennifer Ngadiuba.

Search for a Singly-produced Vector-like T Quark decaying into Ht

I have been the principle analyzer for this search. My contributions include the following points.

- development of the strategy to reconstruct the invariant mass of the vector-like quark,
- development of the strategy to select data and optimize the sensitivity of the analysis,
- a study to decide on the parametrisation of the background including F-tests, goodness-of-fit test, signal injection test,
- the statistical interpretation of the results and the combination of all three years (2016 – 2018).

I was the contact person, responsible for the internal review including:

- regular updates in working group meetings,
- and the preapproval presentation of this analysis.

this analysis is still blinded in the signal region since it is in the internal review of the CMS Collaboration. Under the supervision of Dr. Roman Kogler, in collaboration with Arne Reimers.

Contents

1 | Introduction

The standard model of particle physics describes three of the four known fundamental forces in nature. Furthermore, it characterises all the fundamental particles that we know to this day and their interactions. With the enormous amount of measurements at lepton-lepton, lepton-hadron and hadron-hadron colliders, the standard model is tested and confirmed to high precision.

Despite its numerous successes, the standard model can not explain all of our observations, such as the experimental evidence of dark matter, non-zero neutrino masses or the smallness of the Higgs boson mass compared to the Planck scale. In order to overcome the shortcomings of the standard model, many theories beyond the standard model have been developed to extend it. Numerous searches and measurements are performed at the Large Hadron Collider (LHC) experiments ATLAS, CMS, LHCb and ALICE to find signatures predicted by theories beyond the standard model.

In order to perform a search for new physics phenomena with proton-proton collision data with the CMS experiment, three main steps are necessary:

- measurement and calibration of the electronic signals in the detector,
- reconstruction and identification of all particles originating from the collisions of interest,
- discrimination between the events of interest (signal events) and standard model background events together with the statistical analysis of them.

This thesis presents new developments and improvements on all three topics, together with two searches for new physics.

While it is not clear how new physics will manifest itself, new bosonic resonances, here called Z', and an extension of the quark sector are predicted by theories, which address the unnatural small mass of the Higgs boson [5–10]. In such scenarios the bosonic resonance can decay into a pair of standard model particles, a standard model particle and a new heavy quark or into a pair of two new heavy quarks. The probability to decay in one or the other depends on the free parameters of the theory, e.g. the mass of the new resonance and the new quarks, mixing angles and coupling strengths.

At the LHC several searches for heavy resonances and new heavy quarks have been performed in various final states. No evidence has been found for the existence of either a Z' nor a heavy quark and stringent limits on the possible mass of the new particles have been set. However, regions of parameter space exist of standard model extensions, which have not been covered by current searches. In order to probe unexplored final states and extend the reach of the LHC, two searches with the CMS experiment are presented in this thesis, involving a heavy resonance Z' and vector-like quarks T.

This thesis presents the first search for a heavy spin-1 resonance Z' decaying into a top quark and a new quark T in the lepton+jets final state at a center of mass energy of 13 TeV. The data set analysed was collected with the CMS experiment at the LHC in 2016 and corresponds to an integrated luminosity of $35.9\,\text{fb}^{-1}$. Carried out in the boosted regime, this analysis identifies the boosted H, Z and W boson of the T decay by subjet b-tagging and substructure variables. All three standard decay modes of the T (Ht, Zt, Wb) are captured in this analysis, leading to the most stringent exclusion limits of $Z' \to Tt$ signals in the context of composite Higgs [6] and heavy gluon models [5].

While the $Z' \to tT$ search focuses on masses above 1 TeV of the Z' and T, also lighter resonance masses are allowed and can be relevant in certain models. For example, Ts with masses below 1.3 TeV are excluded by searches for single and pair production [11–25], but these searches take only into account the three common T decays (Ht, Zt, Wb). When including an additional exotic decay into a new scalar particle a ($T \to at$) [26], the low mass region is not excluded by current searches. This thesis presents a search for a singly-produced heavy top quark partner T decaying into Ht in the lepton+jets channel focusing on the low mass region of the T. The data set analysed were collected with the CMS experiment and correspond to an integrated luminosity of $137.2\,\text{fb}^{-1}$. This analysis is the first one carried out in the lepton+jets channel in the resolved regime and uses the identification of the three particle showers originating from the fragmentation of b quarks. It also serves as an orthogonal search region to an all-hadronic analysis, which observes a significant discrepancy between data and standard model expectation between 600 GeV and 700 GeV [20]. This analysis is still blinded, but the expected sensitivity corresponds to a large of five standard deviations for a possible signal with mass of 650 GeV, visible in the all-hadronic final state.

In order to further increase the data quality and the sensitivity of the analyses performed, new developments and improvements on the detector and analyses techniques are needed. Three topics are presented in this thesis in addition to the searches to improve the sensitivity: the challenges of the CMS pixel detector upgrade, a study comparing the performances of the different pileup mitigation techniques, and a study

of new machine-learning based identification techniques and their correlations.

The innermost part of the CMS detector consists of the tracking systems, that compromise the silicon pixel detector and the strips detector. Due to high radiation damage the CMS pixel detector performance started to degrade in 2016 and was therefore exchanged in 2016/2017. The new pixel detector has a novel powering system that was never tested before in the unique conditions of the LHC. This thesis present the challenges observed with the new powering system and a systematic analysis of the failures.

At the LHC strongly focused bunches containing 10^{11} protons each are collided. Therefore, at each bunch crossing not only one collision (the collision of interest) happens but several. Particles from these additional soft collisions are called pileup particles and they can add energy to the physics objects measured from the collision of interest. In order to resolve the particle interaction of interest traditional methods use the tracking system to identify the origin of each particle [27, 28]. The disadvantage of this method is that it can not act on neutral pileup and no mitigation is done outside the tracking volume. This thesis presents the validation of a novel technique which is based on the local environment of each particle in the event. This technique assigns a weight between 0 and 1 to each particle based on the likelihood that it originates from the collision of interest. This has the advantage that it is applicable to neutral particles and can furthermore be calculated outside the tracker volume. An important prerequisite to any pileup mitigation technique is the identification of the origin of charged particles, i.e. pileup or main interaction, and a precise tracking system within the detector is needed.

An efficient tracking detector is not only necessary to mitigate the effects of pileup, but also to identify particle showers originating from the fragmentation of b quarks, or to identify highly Lorentz-boosted particles. The identification of these particles is of special interest when reconstructing heavy resonances at the TeV range. The first search presented in this thesis, $\mathrm{Z}' \to \mathrm{tT}$, uses substructure techniques to identify the heavy bosons, e.g. the decay of a Higgs boson into two quarks, where the Higgs boson has a high momentum, and improvements on the identification would result in a higher sensitivity in this search. The high momentum of the heavy boson results in a small angular separation of the decay products, which makes it difficult to identify them individually. Instead, these decays are reconstructed by large-radius jets, containing the merged showers of particles. In order to differentiate between large-radius jets, e.g. signal and background jets, the substructure is analysed. The substructure of large-radius jets is impaired by pileup particles and the influence of different pileup mitigation techniques on substructure variables is presented. In addition, the performance of

recently developed substructure techniques is studied. Most of these techniques have the disadvantage that they shape the mass distributions of objects from light quarks and gluons in such a way that the mass distribution becomes similar to the signal mass distribution, reducing the discrimination power in this variable. The change in the background shape results from a correlation between the momentum and the mass of a particle shower. In order to avoid such a behaviour dedicated decorrelation methods are needed. A two-dimensional decorrelation method is explained and studied in this thesis.

In the following the standard model of particle physics is described in Chap. 2. The theoretical models the searches are based on, which extend the standard model, are described in Chap. 3. The description of the detector together with the upgrade of the CMS pixel detector is presented in Chap. 4. Afterwards, the reconstruction of an event measured with the CMS detector as well as improvements in identifying Lorentz-boosted particle decays are described in Chap. 5. A comparison of the pileup mitigation techniques together with the latest improvements is presented in Chap. 6. The two searches for new physics are presented in Chap. 7 and Chap. 8. The thesis concludes with a summary in Chap. 9.

2 | The Standard Model of Particle Physics

The formulation of a relativistic quantum field theory allowed to build a mathematical model that describes all of the elementary particles known to date and three out of four known forces in nature. This mathematical model is called the standard model of particle physics, which is outstandingly successfully in describing the experimental data. In this chapter the basic concepts of the standard model are explained. In addition, the physics of proton-proton collisions is discussed. This chapter is based on Ref. [29–31].

2.1 The Particle Content of the Standard Model

The standard model of particle physics incorporates 12 matter particles with spin 1/2, which are also called *fermions* and the mediators of the forces with integer spin, also called *bosons*. The fermions are further split into leptons and quarks with each group having three generations. Each generation consist of two particles. In the case of leptons it is either an electron (e), a muon (μ) or a tau (τ) together with its corresponding neutrino. In case of the quarks it is one up-type quark (u, c, t) together with one down-type quark (d, s, b). Considering the bosons, there are massless bosons, like the photon (γ) and the gluon (g), and massive bosons, like the W and the Z boson.

Each particle can be described and differentiated by its quantum numbers: spin s, mass m, electromagnetic charge e, weak isospin T_3 and color charge C. The antiparticle of a particle has the same mass but opposite charges. The quantum numbers also indicate how and if they are interacting with the corresponding force: the electromagnetic force (electromagnetic charge), the weak force (weak isospin) and the strong force (color charge).

Leptons have an electromagnetic charge of $1e$ for the electron, muon and tau, while the neutrino is neutral. They are color neutral and have a weak isospin of $T_3 = \pm\frac{1}{2}$ (left-handed leptons) or $T_3 = 0$ (right handed leptons). Left-handed leptons appear in

doublets

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \tag{2.1}$$

where the neutrinos have an isospin of $T_3 = +\frac{1}{2}$ and the electron, muon and tau have an isospin of $T_3 = -\frac{1}{2}$. Right handed leptons appear in singlets

$$\nu_{eR}, e_R, \nu_{\mu R}, \mu_R, \nu_{\tau R}, \tau_R. \tag{2.2}$$

Quarks have an electromagnetic charge of $+\frac{2e}{3}$ and $-\frac{1e}{3}$ for up-type and down-type quarks, respectively. They carry one out of three color charges: red, blue or green. Similar to the leptons, left-handed quarks have a weak isospin of $\pm\frac{1}{2}$ and are represented in doublets

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L, \tag{2.3}$$

where the up-type quarks have $T_3 = +\frac{1}{2}$ and the down-type quarks have $T_3 = -\frac{1}{2}$. Right-handed quarks have a weak isospin of $T_3 = 0$ and are represented in singlets

$$u_R, d_R, c_R, s_R, t_R, b_R. \tag{2.4}$$

Bosons have an electromagnetic charge of $\pm 1e$ or 0 for the $W^\pm$ or the Z, γ and g, respectively. The gluon is the only boson that carries a color charge, hence it can interact with itself. The W bosons carry a weak isospin of $T_3 = \pm 1$, all others have $T_3 = 0$.

All together the standard model (SM) is a quantum field theory constructed by requiring local invariance under the gauge groups

$$U(1)_Y \otimes SU(2)_L \otimes SU(3)_C, \tag{2.5}$$

where $U(1)_Y$ represents the electromagnetic interaction with the hypercharge Y, $SU(2)_L$ represents the weak interaction which couples only to left-handed particles and $SU(3)_C$ represents strong force with the color charge C. Each gauge group and its corresponding interaction will be described in the following. Natural units will be used, which means $\hbar = c = 1$, resulting in $[\text{mass}] = [\text{energy}] = [\text{momentum}] = \frac{1}{[\text{length}]} = \frac{1}{[\text{time}]}$.

2.1.1 The Electroweak Sector

The combination of gauge groups $U(1)_Y \otimes SU(2)_L$ describes the unification of the electromagnetic and weak forces. Both forces and the corresponding gauge bosons

arise from the symmetry breaking of $U(1)_Y \otimes SU(2)_L$ to $U(1)_{\text{em}}$. The charge of the gauge group $SU(2)_L$ is the weak isospin T_3 and the corresponding gauge bosons are labelled W_1, W_2, W_3. The charge of the gauge group $U(1)_Y$ is the hypercharge Y with the gauge boson B. The electroweak unification was first described by Salam, Glashow, and Weinberg [32, 33].

The physical $\text{W}^\pm$ bosons can be identified by a linear combination of W_1 and W_2,

$$\text{W}^\pm = \frac{1}{\sqrt{2}}(\text{W}_1 \mp \text{W}_2). \tag{2.6}$$

The $\text{W}^\pm$ bosons are responsible for weak charged current interactions and have a mass of $80.379 \pm 0.012\,\text{GeV}$ [29]. They couple only to left-handed particles and right-handed anti-particles, i.e. particles with a weak isospin of $T_3 = \pm\frac{1}{2}$.

The W_3 and the B represent the Z boson and the photon γ,

$$\begin{pmatrix} \text{Z} \\ \gamma \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} = \begin{pmatrix} B \\ \text{W}_3 \end{pmatrix}, \tag{2.7}$$

where θ_W is the Weinberg angle or weak mixing angle.

The photon is the mediator of the electromagnetic force and it is massless. It is electrically neutral and interacts only with charged particles, i.e. e, μ, τ and quarks. The Z boson, with a mass of $91.1876 \pm 0.0021\,\text{GeV}$ [29], is responsible for weak neutral current interactions. Unlike the $\text{W}^\pm$ bosons, the Z bosons couples to left- and right-handed particles but with different strengths. While the Z boson is electrically neutral too, it also interacts with neutrinos.

Through the electroweak unification T_3 and Y are related by the electromagnetic charge Q,

$$Q = T_3 + \frac{1}{2}Y. \tag{2.8}$$

The electromagnetic and weak force behave as one force above the unification energy, but they give rise to very different effects in the day-to-day life. The weak interaction is responsible for the interaction during radioactive decays, while the electromagnetic interaction is responsible for atomic and molecular bindings. The weak force is the only force that allows the conversion from a charged lepton into its corresponding neutrino with no mixing between generations. The mixing is prohibited by lepton number conservation, where each generation has its own lepton number $L_e = 1$, $L_\mu = 1$ and $L_\tau = 1$ for the generations involving the e, μ and τ, respectively, and zero otherwise. All three numbers must be conserved in each process. In addition, recent measurements show that neutrinos are able to oscillate between the different generations since they

are massive [29].

Quarks can change between generations through interactions with the W bosons. Each process involving a generation change is suppressed as described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [29]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix}. \tag{2.9}$$

The square of each element $|V_{ij}|^2$ gives the probability for the transition from one quark into another.

2.1.2 Quantum Chromodynamics

The last gauge group of Eq. 2.5 represents the strong force with eight massless gluons as mediators, called Quantum Chromodynamics (QCD). Gluons themselves have a color charge which enables self-interaction. They couple to quarks the only other colored particles in the SM. The coupling strength of the strong force increases with the distance. As a result quarks can not appear freely but only in bound, color neutral states, called hadrons. The process in which a quark decays to a hadron is called hadronization and is the reason why these particles are measured in the detector as particle spray rather than individual particles. The top (t) quark is the only particle that decays before it hadronizes due to its short lifetime. At small distances the coupling strength is rather small which means that quarks can be assumed as nearly free which is called the asymptotic freedom.

2.1.3 Electroweak Symmetry Breaking

The SM Lagrangian itself is not invariant if direct mass terms of the electroweak bosons or of the fermions would be included. This is in strong contradiction to the measured masses of the electroweak bosons and the fermions. Therefore, Brout, Englert, and Higgs postulated the mechanism of electroweak symmetry breaking [34, 35]. A new complex scalar field is predicted and represented by an electro weak isospin doublet and hypercharge $Y = 1$

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix}, \tag{2.10}$$

with $\Phi^+ = (\Phi_1 + i\Phi_2)/\sqrt{2}$ and $\Phi^0 = (\Phi_3 + i\Phi_4)/\sqrt{2}$. One component of this doublet is electromagnetic charged, while the other one is neutral, which is important for a neutral

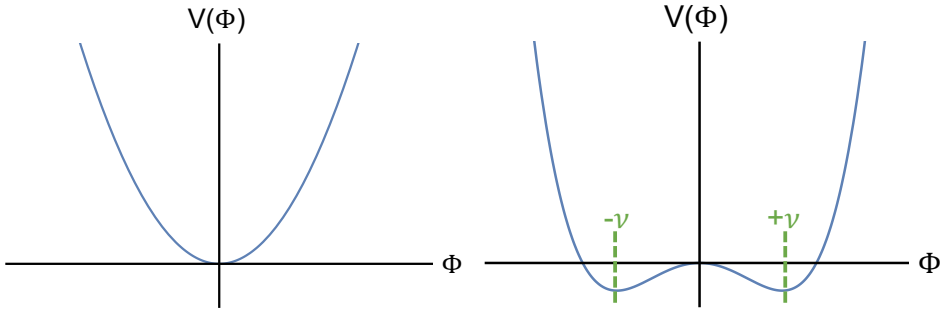


Figure 2.1: The potential of Eq. (2.11) can have two shapes depending on the choice of the parameter μ^2 and λ. For (left) $\mu^2 > 0$ and $\lambda > 0$ the potential has a parabolic form and for (right) $\mu^2 < 0$ and $\lambda > 0$ it has the shape of a Mexican hat.

vacuum. It is the only spin-0 field in the SM. The potential of this mechanism needs to be renormalisable and invariant under $U(1)_Y \otimes SU(2)_L$, leading to the following form

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{2.11}$$

with the two new parameters μ and λ. The term $\lambda(\Phi^\dagger\Phi)^2$ describes the selfinteraction among the scalar field. The parameter λ needs to be greater zero to have a stable vacuum, i.e. the potential is positive at high values of $|\Phi|$. The parameter μ^2 can be either positive or negative. While $\mu^2 > 0$ results in a parabolic potential with one minimum, $\mu^2 < 0$ results in a potential with the shape of a Mexican hat as shown in Fig. 2.1. The former one will preserve the symmetry and is therefore not suitable. The later results in a set of minima at minima at $\Phi\Phi^\dagger \equiv \frac{v}{2} = -\frac{\mu^2}{2\lambda}$, with v known as the vacuum expectation value.

A possible ground state, e.g. the neutral one, is then given by

$$\Phi_0 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix}. \tag{2.12}$$

An expansion about the vacuum given in Eq. (2.12), lead to the excitation interpreted as the real Higgs boson, $h(x)$,

$$\Phi(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix}. \tag{2.13}$$

Inserting Eq. 2.13 into Eq. 2.11 leads to

$$V(\Phi) = -\mu^2 h^2 + \lambda \nu h^3 + \frac{1}{4}\lambda h^4, \tag{2.14}$$

with $-\mu^2 h^2$ being the mass term of the Higgs boson, and the other terms representing the selfinteraction of the Higgs boson.

After the symmetry breaking three of the four degrees of freedom are used to the masses of the electroweak gauge bosons. The masses are predicted by $M_W = \frac{1}{2} v g_W$ and $M_Z = \frac{1}{2} v \sqrt{g_W^2 + (e/\cos\theta_W)^2}$ with the vacuum expectation value of $v = \frac{\sqrt{2} M_W^2}{8 g_W^2} = \frac{1}{\sqrt{\sqrt{2} G_F^0}} \simeq 246\,\text{GeV}$.

The masses of the fermions are not predicted by the mechanism, but are given by

$$M_{f_i} = \frac{1}{\sqrt{2}} v y_i, \tag{2.15}$$

where y_i is the Yukawa coupling between the Higgs boson and the fermion f_i.

The last missing piece of the SM, the Brout-Englert-Higgs (BEH) mechanism [34, 35], was experimentally confirmed by the detection of the Higgs (H) boson in 2012 [36, 37].

2.2 Physics of Proton-Proton Collisions

The SM is tested to high precision at lepton-lepton, lepton-hadron and hadron-hadron collider. However, some questions like the experimental evidence for dark matter or the light mass of the Higgs boson remain. Many theories exist that solve these unanswered questions and are validated or falsified at the Large Hadron Collider (LHC), the most powerful hadron-hadron collider to date. The data collected with the CMS experiment at the LHC are typically compared to simulated events which predict the SM. In order to have precise predictions of the SM it is important to simulated each part of the interaction correctly. Since protons are no elementary particles, it is necessary to understand their inner structure. In this chapter the structure of the proton as well as the basic concepts of event simulation are described.

2.2.1 The Structure of the Proton

The constituent of the protons are grouped into the primary elementary particles, *the valence quarks* and the secondary elementary particles, the *sea quarks*. The group of valence quarks consist of two up quarks and one down quark. These valence quarks are constantly interacting through the strong force (gluons) with each other. Vacuum fluctuations create and annihilate quark-anti-quark pairs through the strong force, which build the sea quarks. A constituent of a proton is called *parton*. When two proton collide not the proton itself interacted but a parton from each proton interacts.

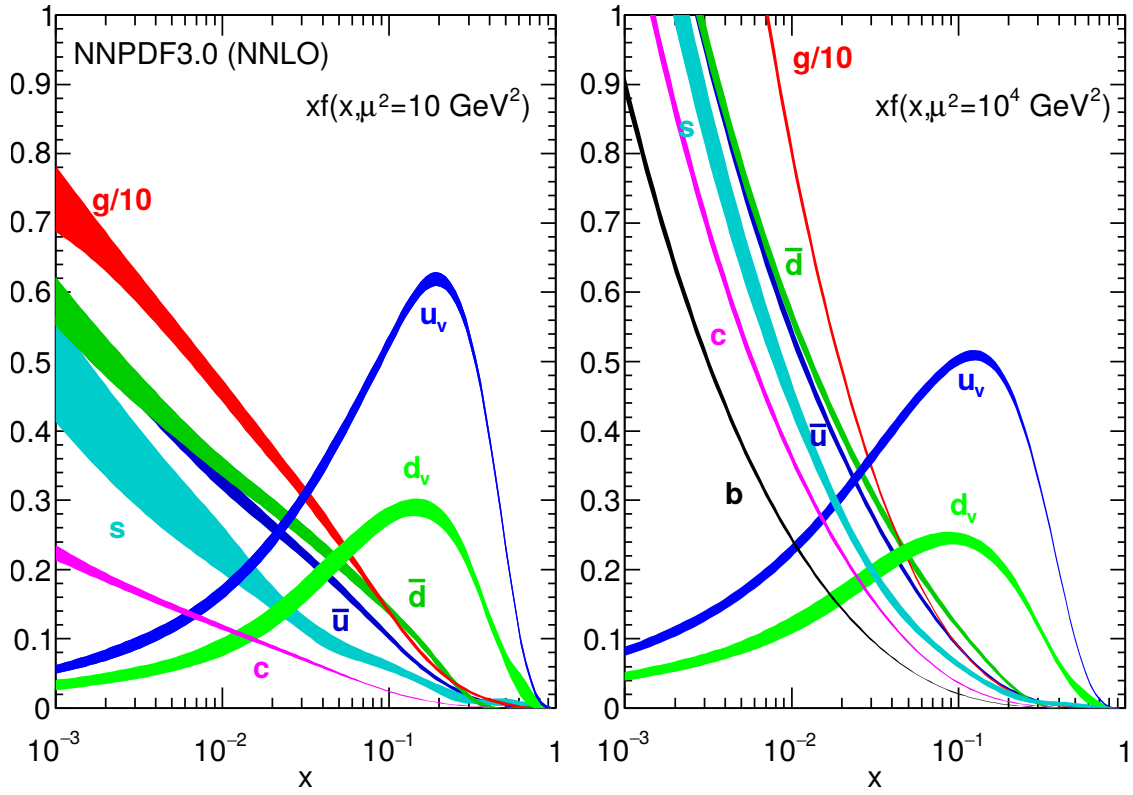


Figure 2.2: Parton density functions at next-to-next-to-leading order for a momentum transfer of (left) $q^2 = 10\,\text{GeV}$ and (right) $q^2 = 1000\,\text{GeV}$ as a function of x. Taken from Ref. [29].

The structure of the proton and the distribution of the proton is precisely measured in deep inelastic scattering experiments, e.g. an electron beam scattered at a proton target. The momentum that is transferred in such a process is called q^2 and is a measure for the resolution with which the structure can be analysed.

The momentum of a proton is split between all partons. The variable x indicates which fraction of the momentum is given to a specific parton. Therefore, the center-of-mass energy of a proton-proton collision is scaled by the two x values of the interacting partons

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}. \tag{2.16}$$

Each of the partons has a distinct Parton Density Function (PDF), which is the probability distribution to find a parton with a certain x value at a given momentum transfer q. An example of the PDF obtained from a global fit to available measurements by the NNPDF Collaboration [38] is shown in Fig. 2.2. It shows that at low values of x the interaction is happening most likely with gluons and sea quarks and at high with the valence quarks.

As a consequence of the internal structure of the proton and the parton interaction is that the parton-parton interaction might be Lorentz-boosted along the beam direction, due to the fact that the x value of both partons is most likely not equal. However, the transverse momentum $\mathrm{p_T} = \sqrt{p_x^2 + p_y^2}$ vanish to good approximation. Therefore, the z component of the momentum of final state particles is unknown but the sum of all transverse momentum of final state particles must be zero due to momentum conservation.

The cross section of a certain process in proton-proton collisions is given by a convolution of PDFs f_i and f_j and the cross section of the parton interaction $\hat{\sigma}_{ij}$

$$\sigma_{pp} = \sum_{i,j} \int\int dx_1 dx_2 f_i(x_1, Q^2) f_j(x_2, Q^2) \hat{\sigma}_{ij}(x_1, x_2, Q^2), \tag{2.17}$$

where i and j indicate all parton flavors possible for this process.

2.2.2 Event Simulation

The simulation of the SM prediction involves several steps and need to be correctly calculated and validated by experimental data. Analysis like Ref. [39] and [40] measure the discrepancy of the simulated SM prediction to the data, which is used for further developments of event generators. The SM prediction is obtained from Monte-Carlo event generators such as POWHEG [41–43], MadGraph [44], Herwig [45] or pythia 8.1 [46]. The procedure of generating a certain physics process involves different parts of a proton-proton collision, such as the hard interaction of the partons, the showering and hadronization and the additional activity such as additional proton-proton interaction during one collision (pileup) or additional parton interaction (underlying event). The first step, the hard interaction, is done based on the PDFs and the cross section of a certain process by either of the event generators depending on the needs of each process. While processes like QCD multijet production at leading order (LO) are typically produced with the LO event generator pythia 8.1, processes like $\mathrm{t\bar{t}}$ production are generated to next-to-leading order (NLO) precision with MadGraph. The next step, the showering, is done by Herwig or pythia 8.1 while the final state particles are matched to the showering with either the FxFx [47] or MLM [48] algorithm. The last step is to simulate the underlying event activity. A sophisticated tuning of this step is need, which is done with measurements of variables that are sensitive to the underlying event, e.g. a new underlying event tune for pythia 8.1 was developed recently and is used in simulations of the year 2017 and 2018 [49–51].

After the modeling of the physics process itself it is needed to simulated the be-

haviour of such a process in the detector. GEANT4 [52] is used to simulate the CMS detector and the interaction of particles with it.

3 | Extending the Standard Model of Particle Physics

Particle physics focuses not only on the properties of the known elementary particles and their interactions, but also on the exploration of new theories that explain every observation in the subatomic world the SM is not able to predict.

The last missing piece of the SM, the BEH mechanism [34, 35], was experimentally confirmed by the detection of the H boson in 2012 [36, 37]. All elementary particles that interact with the H boson acquire their mass through the BEH mechanism. The SM is well tested and confirmed by many experiments from the LHC, Tevatron and LEP collider in high energy physics [53, 54]. However, the SM has shortcomings and can, for instance, not explain astrophysical observations suggesting the existence of new dark matter particles [55], nor can it explain non-zero measured neutrino masses [56]. An other unanswered question of the SM is the so called hierarchy problem [57], which refers to the mass of the H boson. The measured mass of the H boson m_{H} consists of the bare mass m_{bare} plus its scale-dependent corrections [58]:

$$m_{\mathrm{H}}^2 = m_{\mathrm{bare}}^2 - \frac{3y_t^2}{8\pi^2} \cdot \Lambda_{\text{cut-off}}^2 + \frac{\alpha_W^2}{16 \cdot \pi^2} \cdot \Lambda_{\text{cut-off}}^2 + ..., \tag{3.1}$$

where y_t is the Yukawa coupling between the H boson and the t quark, α_W is the weak coupling constant and $\Lambda_{\text{cut-off}}$ is the scale up to which the SM is valid. If the scale $\Lambda_{\text{cut-off}}^2$ is set to the Planck scale[1] [59], it results in corrections of the order of $10^{38}\,\mathrm{GeV}^2$. Given a measured $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ [29], the bare mass must be chosen very carefully to cancel out the large corrections. This is referred to as hierarchy or fine-tuning problem [57]. However, if new physics is at the TeV scale, hence $\Lambda_{\text{cut-off}} \sim \mathcal{O}(\mathrm{TeV})$, the difference between the measured mass and its corrections gets reduced.

Many theories of new physics have been developed to extend the SM and alleviate the unnaturally large corrections but also to address other shortcomings of the SM. This thesis will focus on the validation or exclusion of theories that address the

[1] The Planck scale is the scale at which the gravitational force starts to become relevant. Assuming the SM is the only valid theory, the Planck scale would be the scale to which the SM is valid.

hierarchy problem, like Composite Higgs and extra dimension models.

Composite Higgs models [6–8] aim at solving the hierarchy problem by introducing a new strongly interacting sector at the TeV scale. As a result, the H boson discovered in 2012 would not be an elementary particle but a bound state. The spontaneous symmetry breaking of the new force, originating from the new strongly interacting sector, would make it possible to detect a composite H boson at 125 GeV. Composite Higgs models introduce new heavy spin-1 resonances and additional heavy partners of the third generation quarks.

Extra dimension models [9, 10] predict additional spatial dimensions. While the SM particles participate only in the four known dimensions, gravity can interact also in all other dimensions. This would explain why gravity in our four dimensional space is much weaker than the SM forces. The consequences of extra dimension models are Kaluza-Klein excitations, which are higher mass realisations of SM particles resulting from field dynamics in the extra dimensions. Kaluza-Klein excitations can manifest themselves for example in a heavy gluon [5].

Both of the theories introduced above have new heavy bosonic resonances and an extension of the quark sector in common. The new bosonic resonance have either an electric charge e of $0e$, $+1e$ or $-1e$ and has similar properties as the SM bosons Z, W and g. This thesis focuses on neutral bosonic resonances, here called Z'. Just like the SM bosons, the heavy resonance is mainly produced via $qq \to \mathrm{Z}'$ (Drell-Yan production). Depending on the model an other possible production mode is the vector boson fusion (e.g. $qq \to \mathrm{WW}\,qq \to \mathrm{Z}'qq$), which has normally a lower cross section compared to the Drell-Yan production.

The new resonance Z' can either decay into a pair of SM particles q, a pair of new heavy quarks Q or a mixture of a third generation quark q_3 and a new heavy partner of the third generation quarks Q_3. While the exact branching ratio ($\mathcal{B}$) of the individual decay channels depends on the model assumed, in general the preferred decay mode depends on two parameters: the mass of Z', $M_{\mathrm{Z}'}$, and the mass of Q_3, M_{Q_3}.

If

$$M_{\mathrm{Z}'} < M_{q_3} + M_{Q_3}, \tag{3.2}$$

a decay into a pair of SM particles is favoured as depicted in Fig. 3.1. As the couplings between the heavy resonance and the SM particles are typically enhanced for heavy SM particles in order to address the hierarchy problem, searches with t quarks in the final state are most sensitive. Searches for a leptophobic Z' model [60] and a Kaluza-Klein excitation [10, 61] decaying into $\mathrm{t\bar{t}}$ have been performed by the CMS [62, 63] and ATLAS Collaboration [64, 65], leading to a mass limit on the resonance mass from

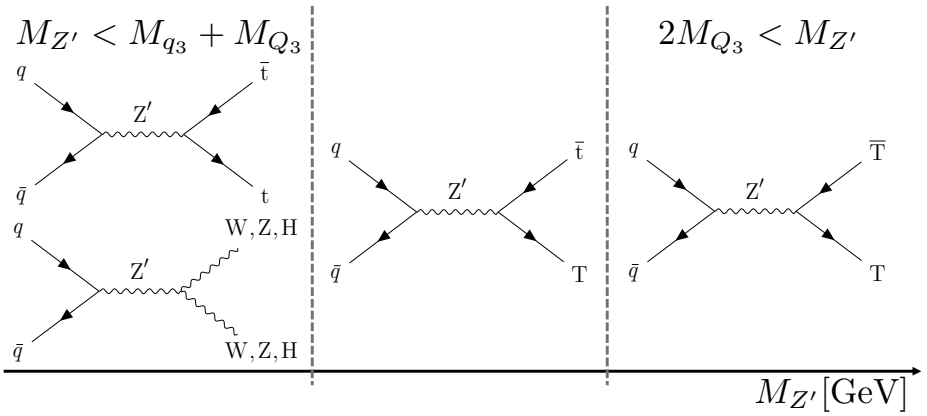


Figure 3.1: Schematic sketch of the decay with the dominating branching fraction as a function of the Z′ mass. The branching fraction depends on the mass of third generation quark, M_{q_3}, and the mass of the new heavy top partner, M_{Q_3}. The decay into a pair of SM particles, e.g. $t\bar{t}$ or a pair of the heavy bosons, is enhanced if $M_{Z'} < M_{q_3} + M_{Q_3}$. Above this threshold the decay into a third generation quark and a new heavy top quark partner, e.g. tT, becomes kinematically possible. If $M_{Z'} > 2M_{Q_3}$ the decay into a pair of new heavy top quark partners, e.g. TT, becomes possible. The exact branching fractions depend on the model parameters assumed. Diagrams that can be obtained by charge conjugation are omitted.

3.8 to 6.65 TeV, depending on the width of the resonance. The heavy resonance Z′ can also decay into a pair of heavy bosons H, W or Z. Both the CMS [66–72] and ATLAS Collaborations [73–82] have made an immense effort in this field, which resulted in the most stringent upper limits on the cross section for resonances decaying into a pair of heavy bosons, where a heavy resonance with a mass of up to 3.8 TeV for the HVT B model [83] is excluded.

However, if $M_{Z'}$ is greater than the sum of M_{q_3} and M_{Q_3} (Fig. 3.1 middle Feynman diagram), a decay mode involving at least one new heavy quark Q_3 in the final state is favoured. Due to the high mass of Q_3 the H boson coupling to this new particle would be the strongest. This is the reason why a fourth generation of chiral quarks would lead to an enhancement in the H boson signal strength, especially for $H \to \gamma\gamma$, which is excluded with 5.3 standard deviations by recent H measurements [84]. Therefore, Q_3 must be non-chiral (vector-like) [85, 86]. The left-handed and right-handed components of vector-like particles transform in the same way under the electroweak symmetry group. This results in direct mass terms, i.e. no mass generation via the BEH mechanism. Signal strength measurements of the H boson would not be affected by the presence of these new vector-like particles. Depending on the exact model assumed, between two and ten new vector-like quarks (VLQ)s are introduced with various

Table 3.1: Summary of the current mass exclusion limits for different possible decays of the heavy bosonic resonance Z' and different productions of the T. Note: Each mass limits is derived for a specific theory model.Decay channels with a * are considered in this thesis.

decay mode	mass exclusion limit	searches
$Z' \to ll$	5.1 TeV	[88–92]
$Z' \to qq$	2.7 TeV	[93–96]
$Z' \to VV$ *	3.8 TeV	[66–82]
$Z' \to t\bar{t}$	$3.8 - 4.55$ TeV	[62–65]
$Z' \to Tt$ *	$1.5\,\text{TeV} - 2.3\,\text{TeV}$ for $M_T = 1.2\,\text{TeV}$ $2.0 - 2.5\,\text{TeV}$ for $M_T = 1.5\,\text{TeV}$	[1, 21, 87]
$Z' \to TT$	no results public	
$gg \to TT$	1.3 TeV	[11–16, 18, 19, 24, 25]
$qq \to TX$ *	1 TeV	[14, 17, 20–23]

charges of $2e/3$, $-1e/3$, $5e/3$ and $-4e/3$ [85, 86]. This thesis will focus on the VLQ T, with an electric charge of $2e/3$, predicted in all models. Three decay modes of the T are possible: Ht, Zt or Wb. Two representations of the VLQs are considered, leading to different $\mathcal{B}$ configurations: either $\mathcal{B}(T \to Wb) = 2 \cdot \mathcal{B}(T \to H/Zt)$ (singlet, e.g. Ref. [5]) or $\mathcal{B}(T \to Ht) = \mathcal{B}(T \to Zt)$ (doublet, e.g. Ref. [6]). Only three searches for a heavy resonance decaying into a vector-like T and a t quark exist [1, 21, 87]. In addition to the production of a T via a new heavy resonance (e.g. $Z' \to Tt$), SM production modes like gluon-gluon fusion ($gg \to g \to TT$) or electroweak production (e.g. $qg \to q'Wb\bar{b} \to Tq'\bar{b}$) are possible. Searches performed for single and pair production of VLQs have set lower bounds of around $M_T > 1\,\text{TeV}$ and $M_T > 1.3\,\text{TeV}$, respectively, on the mass of the T, under the assumption that the T is a narrow (1%) resonance [11–25]. A summary of the exciting results for the different decay modes of the Z' and the T are shown in Tab. 3.1.

4 | Experimental Setup

The data analysed in this thesis are pp collisions happened at the LHC and are collected with the Compact Muon Solenoid (CMS) experiment. The CMS experiment recorded data from 2015 to 2018 at a center-of-mass energy of 13 TeV, while two data sets are analysed in this thesis. The first search considers the data set collected in 2016 corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$ and the second search considers the full Run 2 data set (2016-2018) corresponding to $137.2\,\text{fb}^{-1}$. This Chapter summarizes the experimental setup and gives an overview of the CMS pixel detector upgrade in 2016/2017. Due to the new powering system, this upgrade resulted in challenges and a reduced data quality in 2017. A statistical analysis of the failure is presented, which led to a different operation mode in 2018 where no failure occurred.

4.1 The Large Hadron Collider

The LHC [97] is the most powerful particle collider to date located at Conseil Européen pour la Recherche Nucléaire (CERN). It has a circumference of 26.7 km and can collide protons or heavy ions. Since this thesis focuses on data collected from pp collision the following chapter will focus on this mode.

The protons are obtained from the ionisation of hydrogen molecules, which are accelerated in a chain of different linear and circular accelerates to an energy of 450 GeV before injected into the LHC. The LHC itself is designed to accelerate the protons to up to 7 TeV, which would result in a center-of-mass energy of 14 TeV. At the start of the LHC particles were collided with an energy of 7 and 8 TeV in 2010-2011 and 2012, respectively. After an upgrade phase in 2013 and 2014 the LHC started again in 2015 and collided particles until 2018 with a center-of-mass energy of 13 TeV. The next time period of data taking is planned for 2022 and targets the design center-of-mass energy of 14 TeV.

Around the LHC ring there are four interaction points where the particles can collide: LHCB, A Large Ion Collider Experiment (ALICE), A Toroidal LHC ApparatuS (ATLAS) and CMS. While ATLAS and CMS are multi-purpose detectors, LHCB is

optimised in analysing B-physics data and ALICE was designed to understand the quark-gluon plasma originating from heavy ion collisions.

The LHC produces a lot of different physics processes, while the interesting and/or new phenomena are rare. The number of event of a certain process N is defined as

$$N = \sigma \int \mathcal{L} dt, \tag{4.1}$$

where σ is the cross section of a process, which is a measure for the probability of a process and $\int \mathcal{L} dt$ is the integrated luminosity, which is the amount of data collected. Since the cross section of a process is given by calculations, more integrated luminosity (more data) is needed to increase the number of events of a rare process. The LHC is constantly increase the instantaneous luminosity $\mathcal{L}$ by e.g. reducing the size of the beams

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

where N_b is the number of bunches, which is limited by the current induced by the bunches, f_{rev} is the frequency of the collisions, which is limited by the circumference, $n_{1/2}$ is the number of protons in a bunch and $\sigma_{x/y}$ is a measure for the width of the beams. The LHC was designed to deliver an instantaneous luminosity of $10^{34} s^{-1} cm^{-2}$, while it already exceeded this value in 2018.

4.2 The Compact Muon Solenoid Detector

The CMS detector [98] is a multi-purpose detector, which means that it is designed to analyses a variety of (new) physics processes. It has an onion-like structure of subsystems around the beam pipe. A schematic sketch of the different subdetector systems that are wrapped around the beam pipe is shown in Fig. 4.1. Starting from the interaction point the first layer is the tracking system, which is needed to measure the particles momentum, track their path and determine the charge-sign. The next layers are the electromagnetic calorimeter (ECAL), which is designed to stop photons and electron and to measure their energy, and the hadronic calorimeter (HCAL), which measured the energy of neutral and charged hadrons. After that the impressive solenoid with a magnetic field of 3.8 T is constructed to bend the path of charged particles, which makes is possible to determine their momentum. The last layer are the muon chambers, which are used to measure the momentum of muons. In order to increase the precision of the energy and momentum measurement of each individual subsystem, an algorithm is linking the information of the different subsystems. This algorithm is described in

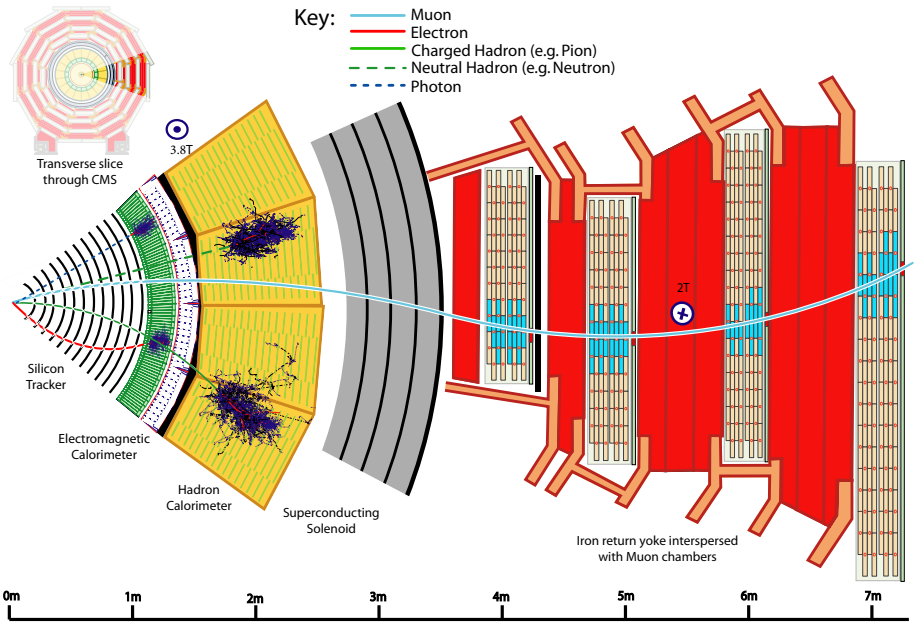


Figure 4.1: Schematic sketch of a transverse CMS detector slice. It is shown which particles interacts with the different subdetector systems. Taken from Ref. [27].

Chap. 5. This section is based on Refs. [98, 99] unless stated otherwise.

4.2.1 The Coordinate System

The coordinate system of the CMS experiment can be described in a Cartesian coordinate system, where the x-axis point toward the ring center, the y-axis point upwards and the z-axis point along the beam direction. However, due to the cylinder symmetry of the detector a conversion into spherical coordinates is useful. In the spherical coordinate system the radial distance from the interaction point is given by r, the azimuthal angle ϕ is the angle in the x-y plane enclosed with the x-axis and the polar angle θ is the scattering angle from the z-axis. Differences in θ are not invariant under a Lorentz-transformation, therefore the polar angle is transformed to the pseudo-rapidity η

$$\eta = -\ln(\tan\frac{\theta}{2}). \tag{4.3}$$

Differences in η are invariant under a Lorentz-transformation. The angular distance ΔR between two objects in the detector is given by

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}, \tag{4.4}$$

which is also invariant under Lorentz-transformation.

4.3 The Tracking System

The tracking systems is the subdetector closet to the interaction point. It is needed to determine the charge-sign, measure the momentum and track the paths of particles to identify the position of the interaction. Since spray of hadronizing particles consist of around 65% of charged particles, the tracking system is responsible for around 65% of the energy measurement of these sprays. The tracking system is divided into two parts: the pixel detectors and the strips detector. In general each system is split into the central part with $|\eta| < 1.3$ called *barrel* and the endcaps with $1.3 < |\eta| < 2.5$ called *forward*. The inner part, the pixel detector is in place since 2010 with three layers in the barrel and two endcaps, but it suffered in 2016 from radiation damage leading to a reduced hit efficiency. Therefore it was exchange at the end of 2016 which led to challenges with the new powering system during the data taking in 2017, which is presented in the next Sec. 4.4. An important design limit is the amount of dead material, e.g. support structure and power cables, that is in front of the calorimeters. With the upgraded pixel detector the amount of dead material is significantly reduced. The resolution of the momentum measurement is 1% for particles with $p_T < 20\,\text{GeV}$ and increases with increasing p_T of the particles [27].

4.4 The CMS Pixel Detector Upgrade: An Overview

The first pixel detector (also called phase 0) is the most inner instrumented part of the CMS detector and is in place since 2010, the begin of data taking at the LHC. This detector part is crucial to reconstruct the primary and secondary vertices as well as to reconstruct the momenta of charged particles. Since the pixel detector is closest to the interaction point it suffers most from radiation damage leading to a reduced hit reconstruction efficiency over time. The original pixel detector was designed to cope with instantaneous luminosities of $1 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ [100].

This design value was reached already in 2016 and it was planned to achieve a two times higher instantaneous luminosity in 2017 and 2018. Fig. 4.2 (left) shows the hit efficiency as a function of the instantaneous luminosity in 2016. The CMS Pixel detector is split into two regimes: barrel (BPix) and forward (FPix) part. The barrel part are the layers parallel to the beam pipe and the forward part are the discs placed perpendicular to the beam pipe. BPix Layer 1, the layer closest to the interaction point, demonstrates an inefficiency of around 6% at $1.5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$, while it was expected to reach an inefficiency of around 16% at $2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ [100].

Therefore, a new pixel detector (called phase 1 detector in the following) was de-

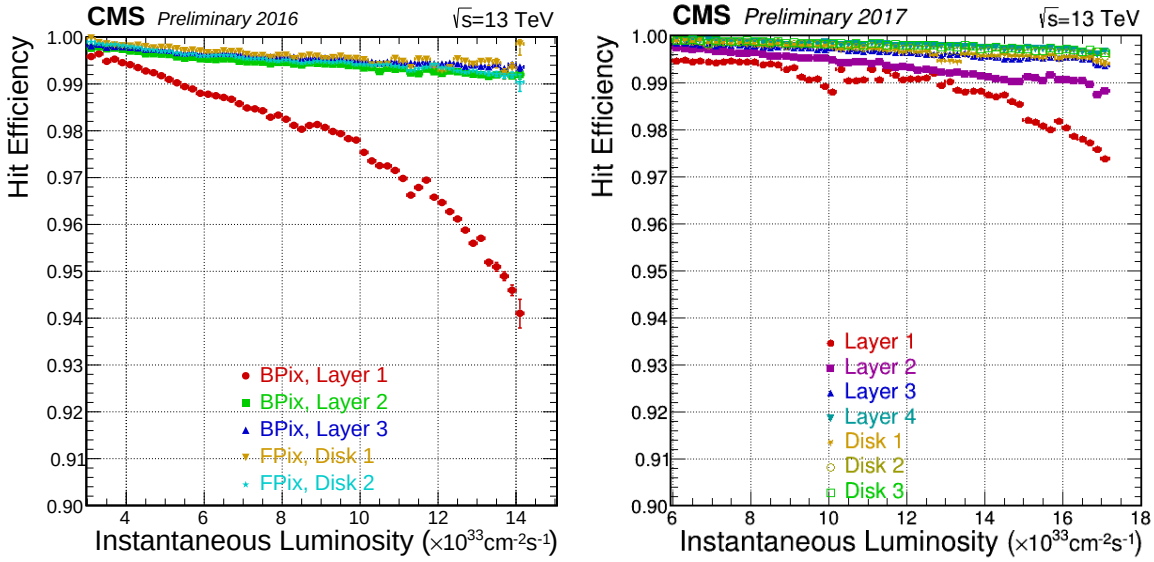


Figure 4.2: Hit reconstruction efficiency as function of the instantaneous luminosity for the original pixel detector in 2016 (left) and the phase 1 pixel detector in 2017 (right). Taken from Ref. [101].

signed and installed in the extended-years-end-technical-stop of 2016/2017. In order to meet the ambitious schedule for installation, it was agreed to keep a similar structure as before and to reuse the off-detector services, such as cooling, power cables and fibers. This also led to the fact that the impact on the following data acquisition was very small.

The phase 1 pixel detector has four instead of three layers in $|\eta| < 1.3$ and an extended coverage by a third disk in the forward region resulting in tracks up to $|\eta| \sim 2.7$ compared to 2016 with $|\eta| \sim 2.5$ [102]. A comparison of the two geometries is depicted in Fig. 4.3.

The four hit coverage, combined with moving the service structure outside the tracker volume ($|\eta| \sim 4$), leads to a more robust tracking and vertex assignment. Moreover, moving the first layer closer to the beam pipe (from 4.4 cm to 2.9 cm) results in a better reconstruction of the secondary vertices especially important for b-tagging.

In addition, a new readout chip [103] with increased buffer size and double bandwidth has been developed to recover the hit efficiency at even harsher instantaneous luminosity conditions than in 2016. The recovered hit efficiency in 2017 can be seen in Fig. 4.2 (right).

This chapter presents the challenges that occurred with the phase-1 pixel detector, the analysis done to investigate it and the commissioning of the reinstalled phase-1 detector. A brief description of the module layout together with a discussion of the changes of the powering system is given in Sec. 4.4.1. During the operation of the phase

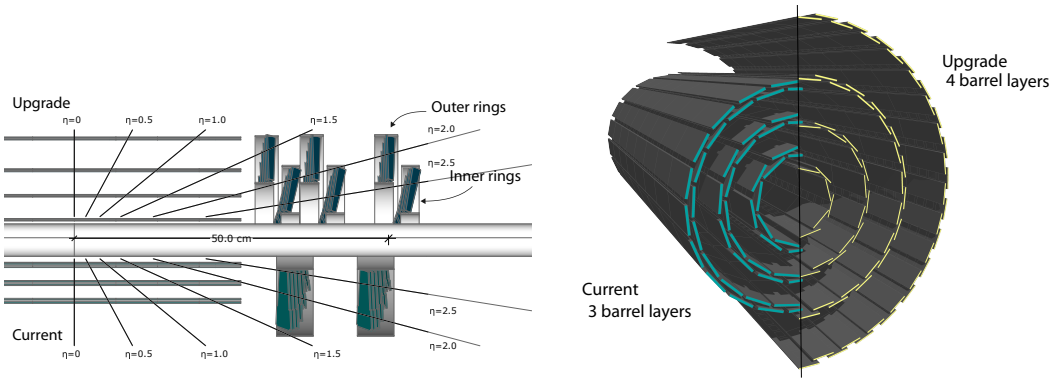


Figure 4.3: Schematic sketch of the original pixel detector (refereed to as current) and the phase 1 pixel detector (refer ed to as upgrade). Taken from Ref. [100].

1 detector in 2017 several problems arising from the modules and the new powering system were encountered, which are described in Sec. 4.4.2 followed by the investigation and activities done to solve the issues (Sec. 4.4.3).

4.4.1 The New Powering System

A schematic sketch of one module of the phase 1 detector is shown in Fig. 4.4 for BPix Layers 2-4. The other modules (BPix Layer 1 and FPix) have only minor differences in the layout. From top to bottom the cable, the High Density Interconnect with token bit manager (TBM), the silicon sensor, the readout chips and the base strips are displayed. Each module has 16 readout chips while the communication and transfer of data is managed by the TBM [104]. Two different voltages, 2.4 V (analog) and 3.3 V (digital) are needed to operate the module.

The semiconductor sensor are bump-bonded to the 16 readout chips. The readout chips themselves are wire-bonded along the two long sides of the modules extending 2 mm beyond the semiconductor sensors to the TBM. The TBM sends a token bit to the first readout chip in a row to collected the data from each of them in a circle. After the token bit went through all readout chips of a module it arrives at the TBM again where it is send with a header and trailer to the data acquisitions system of the pixel detector.

The phase 1 pixel detector has 1.9 times more modules than the original one, resulting in a 4 times higher power needed to cope with the ohmic losses of the 50 m long power cables. Therefore, a new powering system based on DC-DC converters was established. The schematic sketch of a buck converter is shown in Fig. 4.5. A DC-DC converter converts the high input voltage to lower output voltage by using two transistors, T_1 and T_2, as a switch. Either of them is open, which leads to a periodically

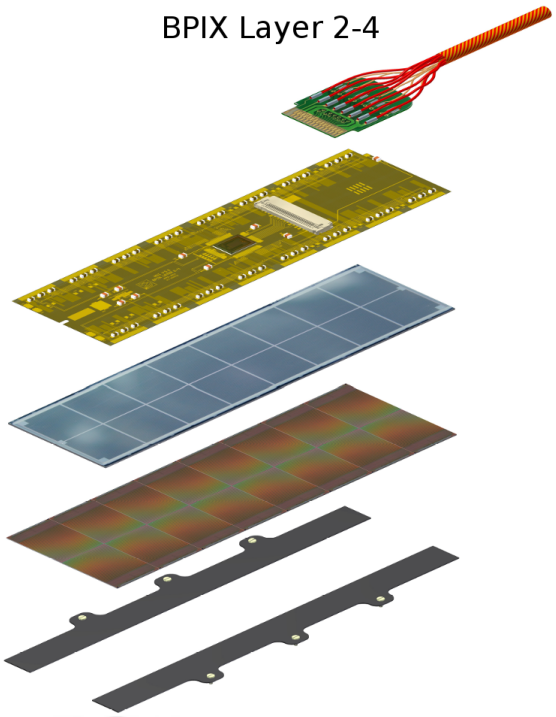


Figure 4.4: Schematic sketch of a module from BPix layers 2-4 from the phase 1 detector. From top to bottom: the cable, the High Density Interconnect with TBM, the silicon sensor, the readout chips and the base strips. Taken from Ref. [101]

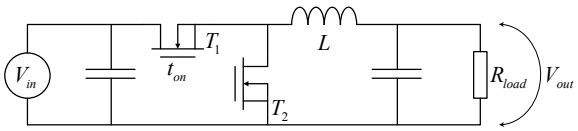


Figure 4.5: Schematic sketch of a DC-DC buck converter installed in the CMS phase-1 detector. The feedback control loop is not shown. Taken from Ref. [100]

connecting and disconnecting of the output from the power supply. The inductor stores energy while T_2 is open and the circuit is connected to the power supply and release energy when T_1 is open.

The DC-DC converters receive as an input 11 V and convert them down to 2.4 and 3.3 V, respectively, resulting in a reduction of the power losses by a factor of 10 [105]. Each pair of DC-DC converters (analog and digital) can feed up to 4 modules depending on the layer. All DC-DC converters are positioned outside the tracker volume at $|\eta| \sim 4$. Since they need to survive a high radiation dose a custom DC-DC converter [105–108] was developed at CERN.

4.4.2 Challenges and Consequences with the New System

Although the modules were tested extensively during the development and construction phase [106, 109, 110], serious problems were faced during the first operation in the unique environment of the LHC. A TBM stopped working because of a single event upset (SEU), which is an interaction of a single ionizing particle with a logical element causing a change of the state. As a result no data are recorded by this module. The rate of stuck TBMs decreases with the distance from the beam pipe, e.g. BPix layer 1 has an inefficiency of about 10% after a fill at the LHC (∼10h) whereas BPix layer 3 reaches an inefficiency of 10% only after a few days. Since reprogramming the modules did not help to recover the stuck TBMs the only way was to turn them off and on again. This power cycle is performed by disabling and enabling the corresponding DC-DC converter. Disabled means that the DC-DC converter does not deliver any output power while the input power is still on. In the enabled state, the DC-DC converter than passes the needed output power to the modules. Although performing a power cycle for one or more TBMs results in a downtime for CMS, it is needed to recover the hit efficiency.

It was later found that the *disabled state* of the DC-DC converter is not safe in the high radiation environment of the CMS pixel detector. A high total ionizing dose (TID) ($>$ 500 krad) leads to a source drain leakage current in the CMOS transistors. This has been known, which is why enclosed layout transistors are typically used in high radiation environments. The enclosed layout transistor can not stand the high input voltages (12 V) of the DC-DC converter and is therefore not used for all transistors implemented in a DC-DC converter. An other mechanisms has been used to protect the high-voltage transistors from leakage current. However, one of the transistors was not sufficiently protected, so the leakage current was integrated with a capacitor while the DC-DC converter was disabled. This lead to an over-voltage in the circuit, causing irreversible damage to the DC-DC converter [111].

A result of this is that at the end of the operation in 2017 around 5% of the DC-DC converters stopped working. If a DC-DC converter does not deliver the power needed any more, the connected modules do not record data. This has lead to significant holes in the efficiency maps, especially in BPix layer 1, where 8 modules out of 96 modules were affected. An ongoing trend of failure in 2018 would have led to a significant impact on physics analysis, e.g. a decrease in the b-tagging efficiency.

As a consequence of the DC-DC converter failure also the modules were damaged. If a module is not powered by the DC-DC converter but the bias voltage to operate the silicon semiconductor sensor is still on, the leakage current damages the readout

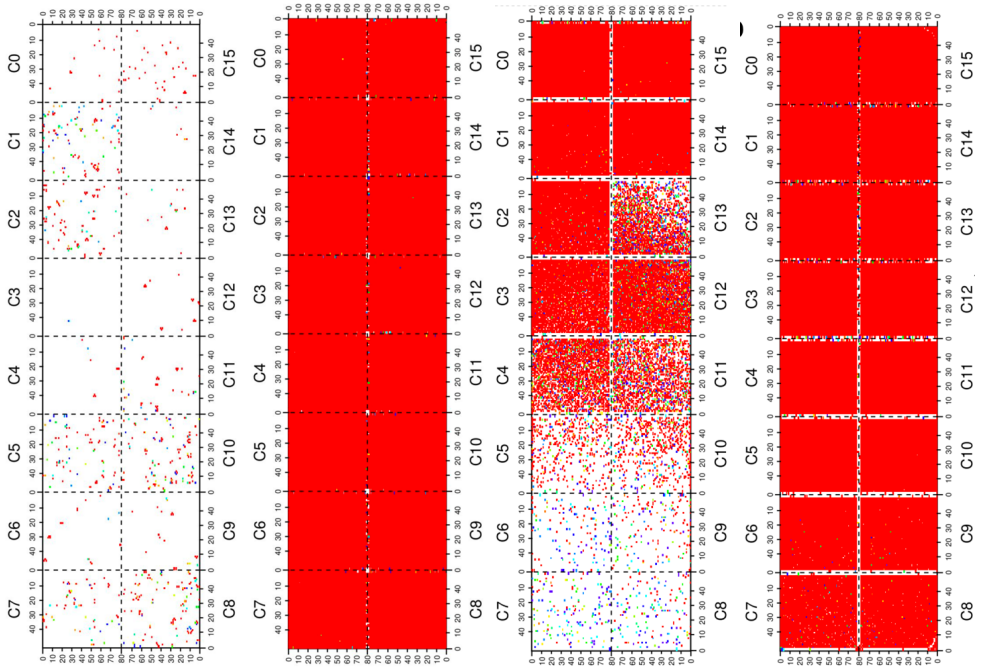


Figure 4.6: Damaged modules that were behind broken DC-DC converters at different distances to the beam pipe during data taking in 2017. From left to right: BPix layer 1, BPix layer 3, FPix ring 1 and FPix ring 2. Red areas are working pixels, white areas are not working pixels. Each module has eight squares indicating the readout chips. The first two modules are oriented parallel to the beam pipe, the last two perpendicular.

chip. This damage is proportional to the time and the leakage current which means that the modules closest to the beam pipe are affected the most, as can be seen in Fig. 4.6, which shows the occupancy of the modules after refurnished with working DC-DC converts. The first module in Fig. 4.6 is from BPix layer 1 behind a broken DC-DC converter. Nearly no pixel on this module is working (white areas). A module in BPix Layer 3 (Fig. 4.6 second plot) has nearly all pixel, working (red areas) and only a few broken ones. The two other modules on the right of Fig. 4.6 are from FPix. The first one is closer to the beam pipe than the other one. Also here the degradation decreases with increasing distance from the beam pipe. Note that these modules are not parallel to the beam pipe, but perpendicular. This is the reason why they show a nonuniform degradation compared to the two modules from BPix.

4.4.3 Investigation and Activities

Despite various stress tests of the CMS pixel detector, such as a warm up test or attempts in the laboratory to produce the same malfunction as in the running experiment, it was not clear until the end of operation in 2017 why the DC-DC converters

broke. Therefore it was decided to take out the CMS pixel detector, replace all DC-DC converters and study them. All 1200 DC-DC converters were extracted in early 2018 and categorised based on their current-voltage characteristic curve (IV curves). An IV curve can show if a device uses more power than expected or does not deliver enough output power. Therefore, it is a useful measurement to understand the behaviour of malfunctioning DC-DC converter.

In the setup used to measure the IV curve of the DC-DC converter it is possible to vary the input voltage and measure the input current (IV curve). In addition, a programmable load can be connected to the DC-DC converter to simulate the consumption of modules as it would be in the detector. It is also possible to disable or enable the DC-DC converter as it was done during operation. Two tests were performed: the *disabled* IV curve and the *enabled* IV curve.

The *disabled* IV curve is measured when the DC-DC converter is disabled and no load is connected. The input voltage (V_{in}) of the DC-DC converter is scanned from 0 to 5.5 V while the input current (I_{in}) is measured. An example for two different IV curves can be seen in the upper panel of Fig. 4.7. A DC-DC converter needs 4.6 V to operate its internal electric circuits. This feature is visible by the increase in I_{in} at around 4.6 V.

A normally working DC-DC converter has an input current between 2 and 4 mA when disabled. The upper panel of Fig. 4.7 shows two categories of DC-DC converters, extracted from the CMS pixel detector: the normal and the high-current DC-DC converters. High-current means that I_{in} is higher than 4 mA at $V_{in} = 5.5$ V when the DC-DC converter is disabled. A summary of all DC-DC converter in the disabled IV curve is depicted in Fig. 4.8 (left). While most of the DC-DC converter show a normal behavior when disabled, 31% showed an abnormal behavior. A high-current behavior in the first place is not crucial for the correct working in the detector.

The second IV curve (*enabled* IV curve) is measured when the DC-DC converter is enabled and a programmable load, which compares to a current of 1.5 A, is connected. The input voltage (V_{in}) of the DC-DC converter is scanned from 0 to 10 V while the input current (I_{in}) is measured. This test reflects how the DC-DC converter is functioning in the detector during operation. An example of this test is shown in the lower panel of Fig. 4.7. The turn on at around 4.6 V is still visible but due to the connected load I_{in} is by a factor of 3×10^2 higher than in the disabled IV curve. Both, the normal and the high-current DC-DC converter, show the same behavior meaning that both are working during operation. A summary of all DC-DC converter's IV curves in the enabled test are depicted in Fig. 4.8 (right). The difference between the above mentioned digital and analog converters is visible. All broken DC-DC converters

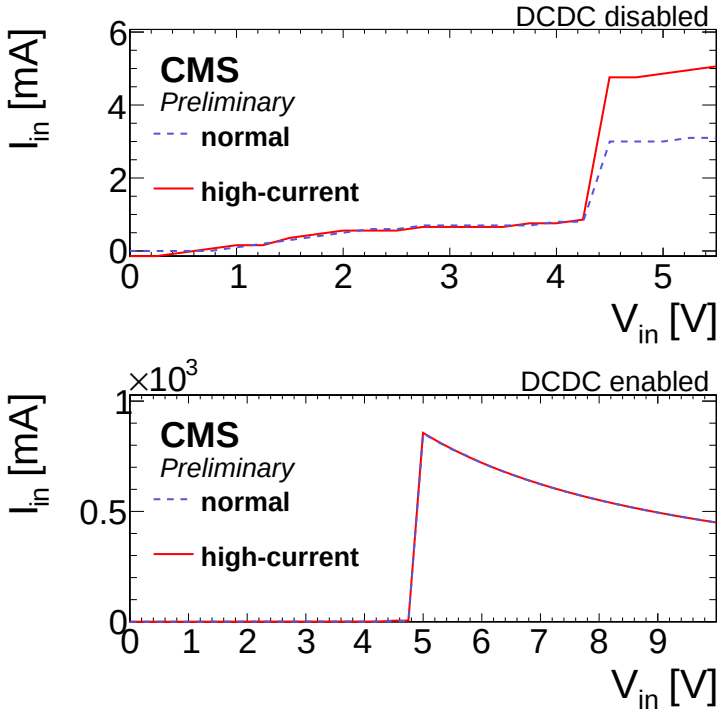


Figure 4.7: IV curves for extracted DC-DC converters from the phase 1 pixel detector. Upper panel shows the disabled IV curve where the DC-DC converter is disabled and no load is connected. Lower panel shows the enabled IV curve where the DC-DC converter is enabled and a programmable load, which compares to a current of 1.5 A, is connected. Published in Ref. [3].

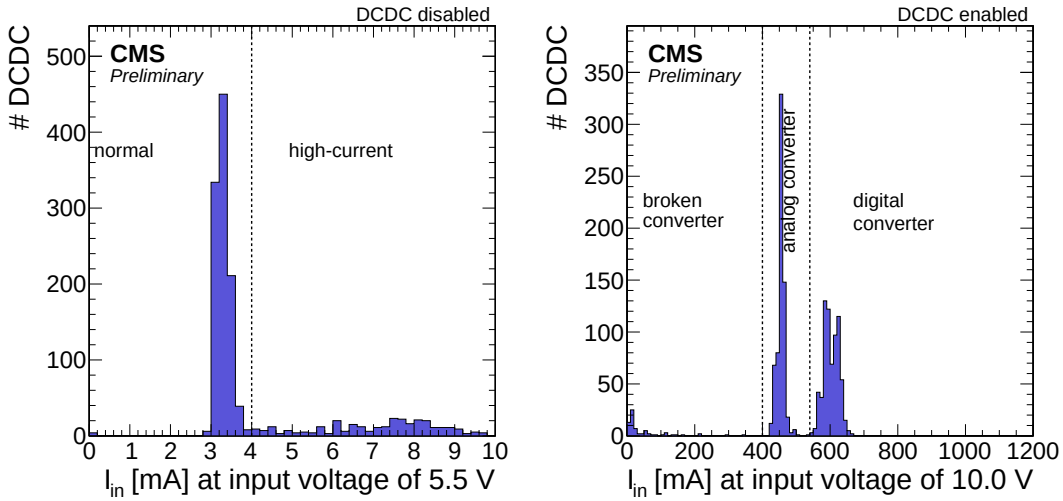


Figure 4.8: Summary of all DC-DC converters extracted from the phase 1 detector in the disabled test (left) and the enabled test (right). Published in Ref. [3].

show an input current far below 400 mA in this test.

The summary of all DC-DC converters in the different categories is given in Tab. 4.1,

Table 4.1: Summary of all extracted DC-DC converters categorised by the IV curves into normal, high-current and broken for BPix (left) and FPix (right). Published in Ref. [3].

Quadrant	normal behavior	high-current	broken	not measured
BPix BmI	155	41	9	3
BPix BmO	129	65	12	2
BPix BpI	159	40	6	3
BPix BpO	126	69	13	0
BPix All	569	215	40	8

Quadrant	normal behavior	high-current	broken	not measured
FPix BmI	56	32	8	0
FPix BmO	69	24	3	0
FPix BpI	52	39	5	0
FPix BpO	64	23	9	0
FPix All	241	118	25	0

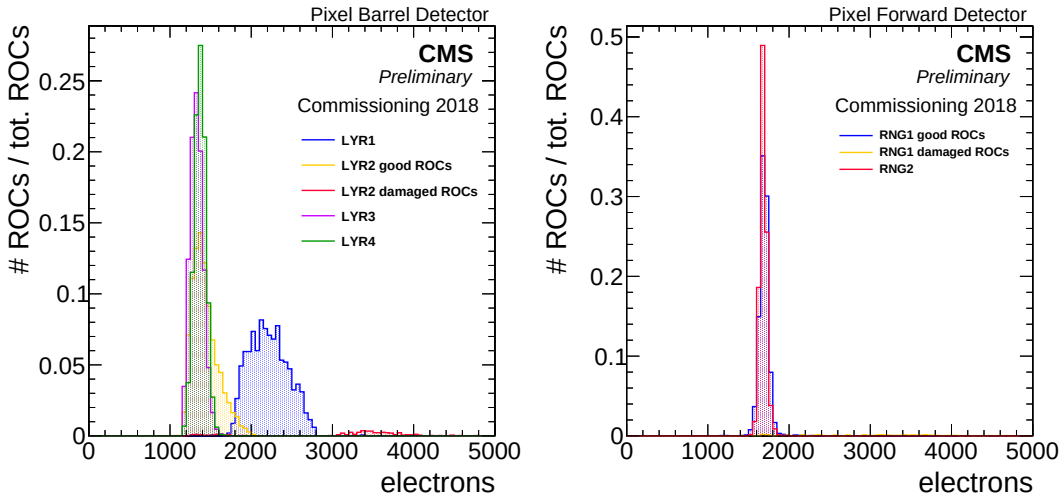


Figure 4.9: Thresholds during commissioning in 2018 for BPix (left) and FPix (right). Published in Ref. [3].

split in the quadrants of the detector geometry showing a uniformly spread of broken and high-current DC-DC converts across all sectors.

All these tests lead to the conclusion that a lower input voltage and avoiding the disabled state reduces the probability of failures. During data taking in 2018, the input voltage of the DC-DC converters was lowered to 9 V and the power cycling was performed by turning off the input voltage to the DC-DC converters. The result has been no broken DC-DC converters during the operation of 2018.

Due to the replacement of all DC-DC converters in the beginning of 2018 it was possible to read out every module again. However, as already mentioned above, some of the modules were not working correctly due to broken readout chips. While six out of eight modules in BPix layer 1 could be exchanged with new ones, two remain not working. In addition, the threshold for damaged modules in the other layers (BPix layer 2-4 and FPix) got increased, as shown in Fig. 4.9, with data from the commissioning in 2018. This means that the number of electrons needed to identify a hit with the sensor is higher for damaged modules.

By removing the CMS pixel detector between 2017 and 2018 also other, non-DC-

DC converter related components could be exchanged, the operation in 2018 started with more active channels than in 2017. New, radiation hard DC-DC converters are already in the production and will be installed before starting the operation in 2021.

4.5 The Electromagnetic Calorimeter

The purpose of the ECAL is to stop electrons and photons and measure the energy of the showers they produce. The high photon identification together with the good energy resolution of the ECAL allows for the precise measurement of another 25% of the particle spray from hadronization of quarks and gluons. The ECAL is build out of scintilators which uses the property of luminescence when excited by the ionizing particles. When the ECAL was designed the limited factor was the space inside the solenoid, which covers the tracking system, the ECAL and the HCAL. Therefore, a scintillator with a short radiation length and a small Molière radius is needed to stop photons and electrons up to high energies in the ECAL. Lead-tungstate ($PbWO_4$) crystals, which have a radiation length of $X_0 = 0.89\,\text{cm}$ and a Molière radius of 2.2 cm are used [98]. On crystal covers up to 25.8 radiation lengths resulting in around 98% of the energy of electrons and photons of up to 1 TeV that is measured in the ECAL. Due to the high granularity it is possible to separate and identify particles with a distance of around 5 cm. The energy resolution of the ECAL is measured with an ECAL submodul directly exposed to an electron beam. No (dead) material from the tracker system was simulated. The energy resolution relative to the electron energy [27] is given by

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\,\text{GeV}}} \oplus \frac{12\%}{E/\,\text{GeV}} \oplus 0.3\%. \tag{4.5}$$

The first term represents the stochastical effects from the shower development, the next term take into account the electronic noise and the last term the calibration errors and non-uniform energy collection, Also the ECAL crystals suffer from radiation damage especially in the endcaps ($|\eta| > 2.5$), which can result in noise if the thresholds are not adjusted properly or the calibration of the crystals is not properly. An increase in noise was seen in 2017 for $2.5 < |\eta| < 3$. This thesis present a possibility to mitigate the noise on analysis level after the data taking in Sec. 6.2.3.

4.6 The Hadronic Calorimeter

The HCAL is the last layer inside the magnet and measures the energy of hadrons. It has a coarser segmentation than the ECAL. Charged and neutral hadrons have a larger interaction length than electrons and photons, which means that they travel through the ECAL with only a few interaction. In order to stop these particles in the short amount of space left and alternating structure of absorber material and active detectors is implemented. Brass absorbers take care for the showering of the hadrons while plastic scintilators measure the energy. The energy resolution of the HCAL [98] is

$$\frac{\sigma_E}{E} = \frac{115.3\%}{\sqrt{E/\,\mathrm{GeV}}} \oplus 5.5\%. \tag{4.6}$$

Since 2018 the endcaps of the HCAL were upgraded to cope with the increasing radiation damage. An other advantage of this upgrade is the increased longitudinal segmentation [112]. This makes it possible to resolve the shower profile of particles better, which might help to identify pileup. The HCAL is in addition affected by out-of-time pileup which means that signals read out of the HCAL consist of different bunch crossings. In order to mitigate this effect a 4 pulse shapes including the previous and the next bunch crossing are fitted to the signal to extract the energy from the bunch crossing of interest [113].

4.7 The Solenoid

The central part of the CMS detector is the solenoid with an homogeneous magnetic field of 3.8 T on the inside. It is bending the trajectories of charged particles in the tracker and in the muon chambers which makes it possible to determine their momentum. It is placed outside the calorimeters to not harm the energy resolution with dead material in front the calorimeters. Outside the magnet iron return york are alternating with the muon chambers to return the magnetic field. In the muon chambers, which are outside the solenoid, the trajectory of a muon is bended in the opposite direction as inside, which gives an even better resolution of the momentum measurement.

4.8 The Muon System

The last layer is the muon system. Muons pass the whole detector without a significant loss of energy. The muon system consists out of three different types of detectors: drift tubes (DT), resistive plate chambers (RPC), cathode strip chambers (CSC). In the

barrel region, $|\eta| < 1.2$, DT and RPC are alternating with the return yorks for the magnetic field. In part of the endcaps, $1.2 < |\eta| < 1.6$, RPC and CSC are alternating. In the rest, $1.6 < |\eta| < 2.4$, CSC are used due to the high particle flux.

4.9 The Trigger System

About $\mathcal{O}(10^9)$ pp interactions per second take place in the CMS detector, which corresponds to a bunch crossing rate of 40 MHz. It is not possible nor needed to store each of these interactions, but only those of interest. A two-level trigger system [114] is used at the CMS experiment to filter out the collisions of interest. In order to be fast enough only a reduced amount of information, e.g. only some subsystems, are used and no full event reconstruction is done. The first trigger stage is hardware based and uses only information from the calorimeters and the muon system to decide weather a collision is of potential interest. The first stage needs 3.2 μs which reduces the event rate down to 100 kHz. The next stage, the high-level trigger, is software based and uses the information of all subsystems. Within several reconstruction and selection steps it is decided to keep or reject a collision. With this the rate is reduced to 100 Hz. At the high-level trigger stage several preliminary objects are reconstructed and based on their momentum it is decided to keep or reject an event. However, these objects are also affected by pileup, which makes it important to mitigate the effect also on trigger level. One possibility is a simplified, modified version of the PUPPI algorithm, which is currently under study and might be part of the trigger in future Runs of the LHC.

5 | Object Reconstruction

The CMS detector delivers raw electronic signals that need to be interpreted into particles from the pp collision. Particles that leave a distinct signature in the detector are e, μ, τ and photons. Quarks and gluons fragment and hadronize, resulting in collimated sprays of hadrons (jets). Each of them has a distinct signature, e.g. the electron leaves a track in the tracking system and showers in the ECAL, while charged hadrons leave a track in the tracking system and shower in ECAL and HCAL. The combination of subsystems makes it possible to identify each particle decay. In addition, it is also important to measure the momentum and energy of the particle, which is possible from the individual subsystems but combining them leads to a better energy resolution. This chapter will first describe the particle-flow (PF) algorithm that is used to combine subsystems of the detector in Sec. 5.1. After that the reconstruction of the interaction points (vertices) are described in Sec. 5.2 and the reconstruction of jets is discussed in Sec. 5.3. The identification of bottom-quark-initiated jets and of heavy objects is discussed in Sec. 5.4 and Sec. 5.5, respectively. A study comparing the latest machine learning algorithms and its decorrelation methods is presented in Sec. 5.5.1. The last Sec. 5.6 describes the reconstruction of the missing transverse momentum, $\vec{p}_T^{\,\text{miss}}$. All these objects are affected by particles originating from pp collisions, which happen in addition to the pp collision of interest in the same bunch crossing (pileup). Therefore a special emphasis is given to the pileup mitigation techniques available in the CMS Collaboration in Sec. 6, where the validation of a new algorithm called pileup per particle identification (PUPPI) is presented and the performance in terms of jet reconstruction is compared.

5.1 The CMS Particle Flow Algorithm

Each particle leaves their distinct signature in the detector as also shown in Fig. 4.1. While muons transverse the whole detector with only little interaction, hadrons are stopped in the HCAL and electrons and photons shower in the ECAL. Charged and neutral particles can be differentiated by their signature in the tracking system. The

energy and direction of stable hadrons, electron and photons can be measured with the energy deposit in a calorimeter cell and its neighbours (cluster). However, this measurement can be improved by using the signature of charged particles in the tracker. A charged particle leaves signals (hits) in the individual layers of the tracker system, which can be combined to the trajectory (track) of a particle. From the bending of the track the momentum can be determined and compared with the corresponding measurement of the calorimeter cluster. The linking between for example the tracking system and the calorimeter, but also between other subsystems is the main idea of the PF algorithm [27]. The three main steps of the PF algorithm are:

1. Reconstruct tracks and energy clusters in the individual detector subsystems.
2. Establish links between the detector subsystems (PF blocks).
3. Identify the PF blocks as particles (PF candidates).

The CMS detector is well suited to fit the needs of the PF algorithm which are

- a fine-grained tracker for a pure and efficient reconstruction of the trajectories,
- a large uniform magnetic field to measure the momentum of charged particles and separate the energy deposit of neutral and charged particles in the calorimeter,
- a highly-segmented calorimeter to measure the energy deposit from charged and neutral particles individually,
- and a muon system with a high efficiency of reconstructing the tracks of muons outside the calorimeters.

Each step of the PF algorithm will be summarized in the following. A more comprehensive description of each individual step can be found in Ref. [27].

In the first step of the PF algorithm tracks need to be reconstructed from a fit through the hits in the individual layers of the tracking systems (track fit). This is done in an iterative approach with a combinatorial track finder based on Kalman Filtering [115]. At the start strict requirements are placed on the quality of the track fit. In each iteration the hits used in the previous track fit are removed and the quality criteria on the next track fit are loosened to increase the efficiency while keeping the misidentification rate as low as possible.

The energy deposit in the ECAL and HCAL is collected to clusters which represent the energy and direction of photons, electrons, charged or neutral hadrons. The cell with the highest energy compared to its surrounding ones and an energy above a given

threshold is called the seed and is the starting point of the topological clustering. Cells that are surrounding the seed and have an energy twice as high as the noise threshold are added to the cluster. The different thresholds applied in the clustering step can result in a lower energy measured from the cluster than the particle has. This is especially relevant for low-p_T particles. A calibration of the the missing energy is determined from simulation.

Muons are leaving signals in the muon detectors, similar to charged particles in the tracking system. The hits in the different muon detectors are combined to the trajectory of the muon bended by the magnet. After all individual signals are reconstructed to tracks or energy clusters, they are combined (linked) to represent PF blocks.

A link can be established between

- the tracking system and the muon system,
- the ECAL and HCAL system,
- and the tracking system and the calorimeter system.

A link between the tracking system and the muon system is done by extrapolating the trajectory of the one system to other. If the tracks from the tracker system and the muon system agree within the uncertainties the link is established. A link between the calorimeter systems is done if the cluster of the ECAL is within the envelope of the HCAL cluster. A link between the tracker system and the calorimeter is done by extrapolating a track into the calorimeters. If the extrapolation ends up in an energy cluster both PF blocks are linked. The quality of the link is defined by the distance in the η-ϕ plane. Photons originating from Bremsstrahlung are clustered to an electron, while linking the track to the ECAL cluster, by adding energy clusters that lie within the tangent to the track at any tracking layer hit.

After linking the PF blocks, each combination is identified and removed from further consideration. The identification starts with the muon candidates, which consist of a track in the tracker system and a track in the muon system. Afterwards electron candidates, tracks that are connected with an energy cluster in the ECAL, are identified. Photon candidates are the remaining ECAL clusters without a track, if the cluster does not share more than half its energy with an HCAL cluster. Charged hadron candidates are called combinations that have a track linked to a HCAL cluster, while neutral hadron candidates are HCAL clusters without a track. Remaining HCAL and ECAL energy clusters might be linked to several tracks. The momenta of the tracks is compared to the energy in the calorimeter. If the difference between the momenta determined from the tracks and the energy in the calorimeter is larger

than the expected energy resolution of hadrons, this block might contain charged and neutral hadrons. If the difference is smaller than the energy in the ECAL cluster and larger than 500 MeV, this energy is considered to be a photon. Otherwise it is considered to be a neutral hadron. Each track is considered to be a charged hadron, where the energy and momentum is determined from the track.

Electron candidates from the PF algorithm have to fulfil a set of requirements to be considered in this analysis. These requirements consider criteria on the shape of the shower, the fraction of energy that is deposited in the HCAL, track based observables and difference in the momentum difference between the track and the calorimeter [116]. These requirements lead to an efficiency of around 70% in identifying electron with a misidentification rate of around 5% or less. The electron candidate has to have $p_T > 30\,\text{GeV}$ and be within the tracker acceptance $|\eta| < 2.5$.

Muon candidate from the PF algorithm needs to fulfil requirements on the number of hits in the tracker and/or the muon detectors and the quality of the link between the tracker and the muon chambers [117]. The efficiency is on average higher than 99% while the misidentification rate is less than 0.5% [117]. The muon candidate has to have $p_T > 30\,\text{GeV}$ and be within the tracker acceptance $|\eta| < 2.4$ to be considered in this thesis.

5.2 Reconstruction of the Primary Vertices

Reconstructing the position of the interactions of the pp collision is important to understand the origin of a particle. In each bunch crossing not only one pp collision happens but several, where only one interaction is of interest and the others are soft inelastic scatterings. In order to remove particles from the additional soft interactions, it is mandatory to define vertices which mark the origins of the pp interactions.

The vertices are reconstructed from tracks determined by the iterative approach with a combinatorial track finder [115] as described above, where strict criteria, like the number of hits in the tracker system or the transverse impact parameter with respect to the beam line, are placed on the tracks used [118]. The requirement on the impact parameter ensures that tracks from a secondary vertex do not interfere. After the selection of tracks that are used as an input, the tracks are clustered with a deterministic annealing algorithm [119]. The last step is to fit the position of the vertex based on all tracks that are clustered to originate from one vertex.

The primary vertex with the largest sum of physics objects p_T is called the leading vertex (LV), while the physics objects are jets clustered with the anti-k_t algorithm [120, 121] from the tracks assigned to the vertex and the negative vector sum of these tracks.

In simulation the number of reconstructed vertices is on average 30% smaller than the simulated number of interactions. The reconstruction efficiency is influenced by the pileup (PU) mitigation technique used and the selections on jets as discussed in more detail in Sec. 6.2.4.

5.3 Reconstruction of Jets

Coloured particles such as quarks can not be observed freely but they hadronize and appear as a spray of charged and neutral particles in the detector. These sprays are identified by jet clustering algorithms which are discussed in Sec. 5.3.1. After identification their energy needs to be calibrated as explained in Sec. 5.3.2. A special type of jets are the ones originating from b quarks which are discussed in Sec. 5.4. The number of jets in an event as well as the jet quantities such as $\mathrm{p_T}$ depends highly on the input to the jet clustering algorithm. Especially the mitigation of PU plays an important role and the influence on the jet reconstruction will be discussed in Sec. 6.

5.3.1 Jet Clustering Algorithm

In this section the general principle of jet clustering algorithms will be discussed while the exact input of jets used within the CMS Collaboration depends on the pileup mitigation used. The jet clustering algorithms presented here are infrared and collinear safe, which is important when comparing the simulation with theoretical calculations. Infrared safe means that the outcome of the algorithm is insensitive against the soft radiation of a gluon, while collinear safe means that the outcome is insensitive against the collinear splittings of particles. The two sequential clustering algorithms used in this thesis are the anti-k_T algorithm [120, 121] and the Cambridge-Aachen (CA) [122] algorithm, which both are based on comparing the distance between the input particles and the $\mathrm{p_T}$ of the particles. The distance between two particles i and j from the input list is defined as

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

where $p_{\mathrm{T},i/j}$ is the transverse momentum of particle i/j, $n \in -1, 0, 1$, ΔR_{ij} is the distance between the particles in the η-ϕ plane and R is the the distance parameter indicating the maximum radius of the jet. The second distance based on the $\mathrm{p_T}$ of the particle itself is defined as

$$d_{iB} = p_{\mathrm{T},i}^{2n}. \tag{5.2}$$

In an iterative procedure two particles are combined to a new particle with added

momentum if the lowest value of all distances d_{ij} is smaller than the lowest value of all d_{iB}. If, on the other hand, d_{iB} is smaller than all possible d_{ij}, the particle i is called a jet clustered from all particles contributed to the sum of particle i. The jet and inputs are removed from the list and the procedure repeats with the remaining particles until all particles are within a jet. The parameter n of Eq. 5.1 and 5.2 decides which particles are clustered first. While $n = -1$ (anti-k_t) means that the hardest particles are clustered first, $n = 1$ (k_t) means to start with the soft particles. Choosing the parameter $n = 0$ (CA) results in clustering that is purely geometrical. Two different jet collections based on the anti-k_t algorithm with a jet radius of 0.4 and 0.8 are used in this analysis.

5.3.2 Jet Energy Calibration and Identification

The energy of a measured jet needs to be corrected for effects originating from PU and detector effects to match the true jet energy. This is done in a factorized approach to tackle the effects from each source individually [123]. The first step is to remove residual PU contributions from the jet energy. This correction is only applied to jets which undergo the charged hadron subtraction (CHS) PU mitigation technique and is not needed for jets which undergo the PUPPI PU mitigation technique. The next step is to correct for detector effects, e.g. electronic noise bias, non-linearity in p_T, and non-uniformity in η, in the simulation. Therefore, the response of the jets, the ratio of reconstruction and particle-level jet p_T, is corrected to be on average unity. The last step is to correct residual effects in data to match the jet energy in simulation. This is done as a function of η in QCD dijet simulation and as a function of p_T in Z/γ+jets simulation. After this procedure the average jet energy scale should be equal in data and simulation. More details for each individual step can be found in Ref. [123].

In addition to the jet energy scale the jet energy resolution (JER) need to be corrected. The JER is typically lower in simulation as in data. The JER is measured for particle-level jets in simulation. Variables sensitive to JER are used to derive a correction factor for simulation to match the data as a function of η of the jet.

Jets used in this thesis have to pass certain quality criteria to reject noise origin from ECAL and HCAL electronics [4]. The efficiency of these criteria for genuine jets is greater than 98%. In the following two types of jets are distinguished: particle-level jets and reconstruction-level jets. While particle-level jets are clustered from stable (lifetime $c\tau > 1\,\text{cm}$) particles excluding neutrinos before any detector simulation, reconstruction-level jets are clustered from reconstructed PF particles.

5.4 Identification of Bottom-Quark-Initiated Jets

The analyses presented in this thesis consider for example the decay of t quarks and H bosons. The decay with the highest $\mathcal{B}$ of the t quarks and H boson are $t \to Wb$ and $H \to b\bar{b}$, respectively. Therefore, the identification of jets originating from b quarks is important to this thesis. In the process of identifying these jets the lifetime of the b hadrons is important. The b quark can, due to its lifetime, travel a few mm before it decays. This decay vertex of the b quark can be identified as a secondary vertex from PF particles and is used to distinguish between jets from b quarks and jets from e.g. light quarks.

Two algorithms are used to identify jets originating from b quarks in this thesis: CSVv2 [124] and DeepJet (also called DeepFlavour) [125]. The former is based on information from the secondary vertex and track-based lifetime information and is used for the first search which analyses 2016 data only. The latter is a deep-neural-network algorithm that was developed during Run 2. It includes 16 properties of charged constituents of a jet, 12 properties of neutral constituents of a jet and 12 properties related to the secondary vertex. This algorithm is used for the second search, which analyses the Full Run 2 data set. An improvement of the efficiency between DeepJet and CSVv2 of around 8% at a misidentification rate of 1% is expected [126].

5.5 Identification of Heavy Objects using Jet Substructure Techniques

This thesis considers a broad p_T range of H bosons that decay into $b\bar{b}$. While at low p_T ($p_T \lesssim 250\,\text{GeV}$) the distance between both b quarks in the η-ϕ plane is on average $\Delta R \sim 1$ at higher p_T the angular separation reduces to

$$\Delta R \sim \frac{2M}{p_T}, \tag{5.3}$$

where M is the mass of the particle that decays, e.g. the mass of the H boson. As described in Sec. 5.3 this thesis uses two jet collections with a radius parameter $R = 0.4$ and $R = 0.8$. The former one can be used to reconstruct two individual jets that originate from the decay products of the H decay at low p_T (resolved regime) the latter can be used to reconstruct both decay products in one jet at high p_T of the H boson (boosted regime).

This is relevant for all heavy objects used in this thesis like the Z and W boson, and

the t quark. In order to differentiate between the large-R jets ($R = 0.8$) the properties of the decay particle are considered. An intuitive choice is the jet mass, which is in the ideal case close to the mass of the heavy object. While the raw mass of the jet, i.e. the invariant mass of all constituents of the jet, is affected by soft and wide-angle radiation a grooming technique is applied to mitigate the effect of unwanted contributions from the underlying event.

The algorithm commonly used within the CMS Collaboration is the *soft drop* algorithm [127], while the ATLAS Collaboration prefers trimming [128]. Both algorithms attempt to remove soft and wide angle radiation by removing part of the jet energy. The soft drop algorithm reclusters the jet geometrically with the CA algorithm, where the input are all constituents of the jet under consideration. Afterwards the algorithm goes backwards in the clustering procedure and test for each pair of subjets if

$$\min(p_{\mathrm{T},i}, p_{\mathrm{T},j}) > z_{\mathrm{cut}}(p_{\mathrm{T},i} + p_{\mathrm{T},j}) \left(\frac{\Delta R_{ij}}{R}\right)^{\beta}, \tag{5.4}$$

with z_{cut} and β being tunable parameter, holds. In case it does, the two subjets are kept and the sum of them is the *soft drop mass*. In case it does not hold the subjet with the lower $\mathrm{p_T}$ is removed and the clustering history of the other jet is checked. The values used in this thesis are $z_{\mathrm{cut}} = 0.1$ and $\beta = 0$. It needs to be noted that with this β value the soft drop mass is infrared and collinear safe but the $\mathrm{p_T}$ of the soft drop jet is not.

The mass of the jet is already a good quantity to identify the heavy object. However, additional quantities are necessary to further reduce the amount of e.g. QCD multijet background. While large-R jets from a heavy object have either a two prong or three prong decay, large-R jets from QCD multijet production originate from a single particle and achieve large soft drop masses by large angle soft splitting. This decay structure, one-prong versus two/three-prong, is visible within the jet considering the energy distribution. There are different variables, e.g. N-subjettiness [129], energy correlation function [130] and D_2 [131], commonly used within the CMS and ATLAS Collaboration. Each of them trying to discriminate between the N-prong and $N+1$-prong decay inside the large-R jet.

The N-subjettiness variable, denoted τ_N, test the hypothesis that the energy within a jet is consistent with N energy axis

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T},k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}\}, \tag{5.5}$$

where k runs over all jet constituents, $\Delta R_{j,k}$ is distance between the constituent k

and the energy axis j and $d_0 = \sum_k p_{\mathrm{T},k} R$. If $\tau_N \approx 0$ the constituents of the jet are in alignment with the N energy axes, which means the jet has N or less energy axes. If $\tau_N \gg 0$ a large fraction of the constituents is not aligned with N energy axis, which means the jet has at least $N+1$ energy axes. This means that τ_2 should be small for two-prong decays like the W boson and large for one-prong decays like QCD multijet production. While τ_2 gives already some discrimination power, the best performance is achieved when calculating the ratio τ_N/τ_{N-1}. In this thesis $\tau_{21} = \tau_2/\tau_1$ is used to identify the W and Z decay and $\tau_{32} = \tau_3/\tau_2$ is used to identify the t quark.

While N-subjettiness is also useful to identify $\mathrm{H} \to q\bar{q}$ decays, the decay of the H boson offers an additional quantity that is helpful in the identification. With $\mathcal{B}(\mathrm{H} \to \mathrm{b\bar{b}}) = 58.4\%$ [29] jets originating from H boson decays can be classified with subjet b-tagging [124]. As discussed in Sec. 5.4, b hadrons have the unique feature of flying several mm before decaying. The displaced, secondary vertex from a b quark decay can also be identified in a large-R jet. In this case the CSVv2 algorithm [124] is applied to the subjets of the large-R jet. These subjets originate from the soft drop algorithm. A combination of requirements on those variables is called *tagger*, where each tagger has the goal to identify a specific heavy object.

Previously mentioned variables for the identification of heavy objects form the beginning of substructure techniques. Today several machine-learning based identification techniques, i.e. machine-learning taggers, are available in the CMS Collaboration. A description and the performance of each of them is presented in Ref. [132]. In this thesis a comparison of the identification techniques useful for a search of heavy resonances that decays into a pair of heavy bosons is performed, with a focus on the possibility to decorrelate them (Sec. 5.5.1). The machine-learning algorithm *DeepAK8* is of special interest for this study. This algorithm exploits the particle level information and is a multi-classifier with five main categories, H, Z, W, t and other. Each category can be further divided into the final state of the targeted particle, e.g. $\mathrm{Z} \to \mathrm{b\bar{b}}$ or $\mathrm{Z} \to q\bar{q}$, where $q = u,\ d,\ c$. Two input lists for each jet, one containing particle-level information and the other one containing information about the secondary vertices, are used as an input to the deep-neural-network. Information (42 variables) for up to 100 constituents of the jet considering for example the $\mathrm{p_T}$, the quality and origin of a track, the energy deposit, the charge, the angular separation between the particle and the jet axis, are used as an input (particle-level list). The information about angular separation and the jet axis are useful to understand the jet substructure needed for the decays considered. The secondary vertex list is used to further improve the information about the jet substructure for heavy flavor decays, e.g. $\mathrm{Z} \to \mathrm{b\bar{b}}$ or $\mathrm{H} \to \mathrm{b\bar{b}}$. It contains seven vertices with each 15 variables such as quality criteria and the displacement. In

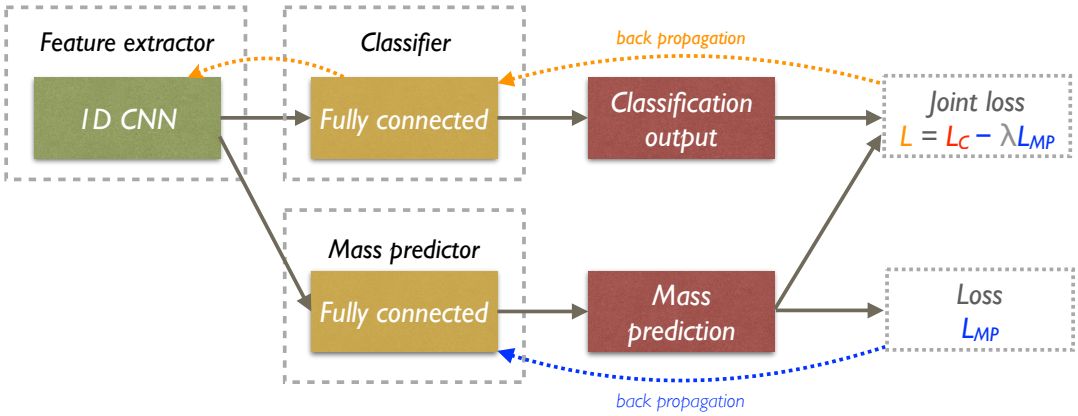


Figure 5.1: The architecture of the neural network “DeepAK8-MD".

order to take into account all correlations between the variables and cope with the high amount of input a convolutional neural network (CNN), one for each list, is used. The CNN, which has as an input the particle-level information, consist of 14 layer, the other one, which uses the secondary vertex information as an input, has 10 layers. The output layer of each CNN reaches from 32 to 128. The output is given to one fully connected layer with 512 units. Afterwards, a ReLU [133] activation function and a Dropout [134] layer with a 20% drop rate is used.

The output of the two CNNs might by highly correlated with the mass of the jet. This can lead to a shaping of the smoothly falling background as further discussed in Sec. 5.5.1. In order to overcome this issue a mass-decorrelated version of this classifier is obtained by adding a mass prediction network. The architecture is build in addition to the deep neural network mentioned above and influences the training of the CNNs by giving a penalty to the loss function. The structure of this network is shown in Fig. 5.1. The feature extractors (green box) are the two CNNs described above which have the particle-level list and the secondary vertex list as an input. The result of the CNNs is used to predict the mass of a jet with a resolution of down to 10 GeV in a range from 30 GeV to 250 GeV. The mass prediction network is build out of 3 fully-connected layers with 256 units and a SELU activation function [135]. If the precision of the mass predicted is to high, the CNN learned its discrimination power from variables highly correlated with mass. In order to suppress such a development, the precision of the mass prediction network is used as a penalty. However, this results in a reduced performance of the tagger itself, since the mass of a jet shows high discrimination power between background and signal events itself.

The output of the neural network is a score for each category. A combination of

scores is used to discriminated for example $\mathrm{Z} \to \mathrm{b\bar{b}}$ events from QCD $\to \mathrm{b\bar{b}}$ events

$$\mathrm{ZHbbvsQCD} = \frac{\mathrm{sc}(\mathrm{Z} \to \mathrm{b\bar{b}}) + \mathrm{sc}(\mathrm{H} \to \mathrm{b\bar{b}})}{\mathrm{sc}(\mathrm{Z} \to \mathrm{b\bar{b}}) + \mathrm{sc}(\mathrm{H} \to \mathrm{b\bar{b}}) + \mathrm{sc}(\mathrm{QCD\ all})} \tag{5.6}$$

$$\mathrm{WvsQCD} = \frac{\mathrm{sc}(\mathrm{W} \to q\bar{q}) + \mathrm{sc}(\mathrm{W} \to qc)}{\mathrm{sc}(\mathrm{W} \to q\bar{q}) + \mathrm{sc}(\mathrm{W} \to qc) + \mathrm{sc}(\mathrm{QCD\ all})} \tag{5.7}$$

$$\begin{aligned}\text{with } \mathrm{sc}(\mathrm{QCD\ all}) =& \mathrm{sc}(\mathrm{QCD} \to \mathrm{b\bar{b}}) + \mathrm{sc}(\mathrm{QCD} \to c\bar{c}) + \mathrm{sc}(\mathrm{QCD} \to \mathrm{b}) \\ &+ \mathrm{sc}(\mathrm{QCD} \to c) + \mathrm{sc}(\mathrm{QCD} \to \mathrm{other})\end{aligned} \tag{5.8}$$

5.5.1 Decorrelation of Taggers in the Context of a Search for Diboson Resonances

Searches that perform a bump hunt (resonant signal on smoothly falling background) in, for example, the jet mass distribution or that use certain regions of the jet mass as a control region to estimate the background, rely on the fact that the QCD multijet background is smoothly falling in this variable while the signal is resonant in the mass distribution. However, the identification of heavy objects might change this behaviour if the variables used are correlated with mass. In order to overcome this issue, which reduces the discrimination power between background and signal, hence the sensitivity of the search, the variables must be decorrelated. In this section a study comparing the most recent taggers commissioned by the CMS Collaboration [132] is presented. In addition, a two-dimensional decorrelation method is presented, which was first described in Ref. [136] for a energy correlation function (N_2^{DDT}). The effect on physics analysis is evaluated by performing this studies in the context of a search for a heavy resonance Z' that decays into a pair of heavy bosons.

The latest search of this decay done by the CMS Collaboration in the all-hadronic final state [66] analyzed the data set recorded in 2016 and 2017 corresponding to an integrated luminosity of $77.3\,\mathrm{fb}^{-1}$. Masses of the resonance above 1.2 TeV are studied, to ensure that the trigger used is fully efficient. The standard tagging technique, soft drop mass plus a requirement on a decorrelated N-subjettiness variable is used to identify the bosons.

Furthermore, the analysis makes use of the fact that the signal, $\mathrm{Z}' \to VV$ where V denotes a vector boson (W, Z), is resonant in three dimensions. A peak is expected in the distribution of the soft drop mass of the leading-$\mathrm{p_T}$ jet m_{j1}, the soft drop mass of the second leading-$\mathrm{p_T}$ jet m_{j2} and the invariant mass of both jets m_{jj}. The SM background

(mostly QCD multijet) is non-resonant in all three dimensions. A multidimensional fit in the three mass dimensions just explained results in an improvement of up to 30% [66] compared to the traditional one dimensional fit [72]. In order to reach the best sensitivity, the monotonically falling behaviour of the non-resonant background should not be influenced by the tagger. For example, the efficiency of the variable τ_{21} depends on the jet soft drop mass and the jet p_T. As a result the QCD multijet background is not monotonically falling but shaped, when a requirement on τ_{21} is applied. Figure 5.2 shows the mass spectrum of the leading-p_T jet in a QCD multijet sample. The blue rectangles show the nominal spectrum without any requirement on τ_{21}. The green crosses show the spectrum with $\tau_{21} < 0.5$, where a shaping of the QCD multijet background is seen. Therefore, it is important to decorrelate the variable (τ_{21}^{DDT}) from the jet p_T and mass. In Ref. [66] this is done by a designed decorrelated tagger (DDT) methodology presented in Ref. [137]. This means that the variable τ_{21} is transformed to τ_{21}^{DDT} by

$$\tau_{21}^{DDT} = \tau_{21} - M \log(M_{SD}^2/p_T/\,\text{GeV}), \tag{5.9}$$

where M is the slope measured in QCD multijet simulation from the distribution of τ_{21} as a function of $\log(M_{SD}^2/p_T)$ and is set to -0.08 in this thesis. This transformation reduces the p_T-dependence of the variable on the QCD multijet background significantly. The advantage of this decorrelation method is, that it results in a monotonically falling behaviour of the non-resonant background in m_{j1} and m_{j2}, but not in m_{jj} as can be seen in Fig. 5.2. In order to compare different decorrelation techniques the difference between the nominal spectrum and the spectrum with a requirement (χ^2/ndf) is given in the legend.

5.5.2 Efficiency and Misidentification Rate

The performance of the different taggers is evaluated with receiver operating characteristic (ROC) curves showing the misidentification rate as a function on the efficiency. The efficiency is calculated in a signal simulation of a heavy resonance $Z' \to ZH$, $G_{bulk} \to ZZ$ or $W' \to WH$ to identify the H, Z and W boson, respectively. The mass of each resonance is chosen to be 3 TeV. In each event the large-R jet that contains the two generator quarks from the decay of the boson ($\Delta R < 0.8$) is considered to be the H-, Z- or W-jet. In order to perform this study in the relevant phase space for the all-hadronic diboson search, only events with two large-R jets with $p_T > 200$ GeV, $|\eta| < 2.4$ and $55 < M_{SD} < 215$ GeV are taken into account. The pileup mitigation technique used in this analysis is PUPPI, which gives a more stable performance for large-R

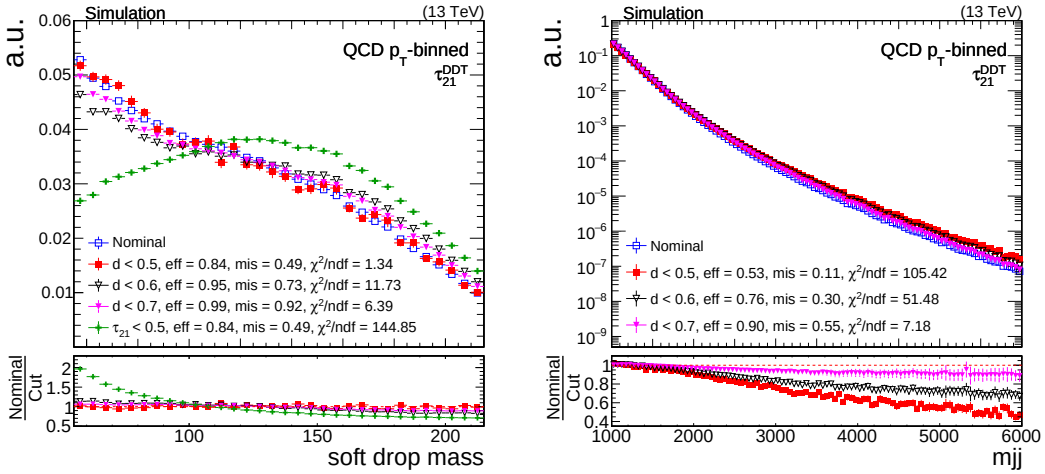


Figure 5.2: (left) Soft drop mass of the leading-p_T large-R jet in QCD multijet simulation for the τ_{21}^{DDT}tagger. (right) The invariant mass of the two leading-p_T large-R jets in QCD multijet simulation for the τ_{21}^{DDT}tagger. Different requirements on the discriminator corresponding to different misidentification rates (mis) and signal efficiencies (eff) are shown and compared to the nominal (without any requirement). The χ^2/ndf is calculated for each spectrum with respect to the nominal spectrum. The lower panel shows the ratio between the nominal spectrum and one of the spectra with a requirement on the discriminator.

jets as described in Sec. 6.2.5. The soft drop mass of large-R jets with PUPPI applied originating from a W decay shows a p_T dependent shift and does not peak at the PDG value of the W boson of 80.4 GeV [29]. Therefore, p_T dependent corrections are applied to remove the shift in p_T and correct the soft drop mass to the true W boson mass. In addition, requirements on the invariant mass of the two leading-p_T large-R jets of $m_{jj} > 1\,\text{TeV}$ is applied, needed due to trigger requirements. A separation of the two leading-p_T large-R jets of $|\Delta\eta_{jj}| > 1.3$ is used to reduce SM backgrounds. The last condition, $\rho = 2\log(M_{SD}/p_T) < -1.8$ must be fulfilled to reject events, where the soft drop mass of the two leading-p_T jets is large but the p_T of the jets is small. These are jets where higher order corrections are important, which we do not model with our parton showers.

The efficiency is then defined as the number of events after the tagger is applied on the H-, Z- or W-jet divided by all events in one of the signal simulations. The misidentification rate is defined as the number of events after the tagger is applied to the leading-p_T jet divided by all events in QCD multijet simulations. The resulting ROC curves can be seen in Fig. 5.3.

The DeepAK8 machine-learning tagger and the DeepFlavour tagger were already

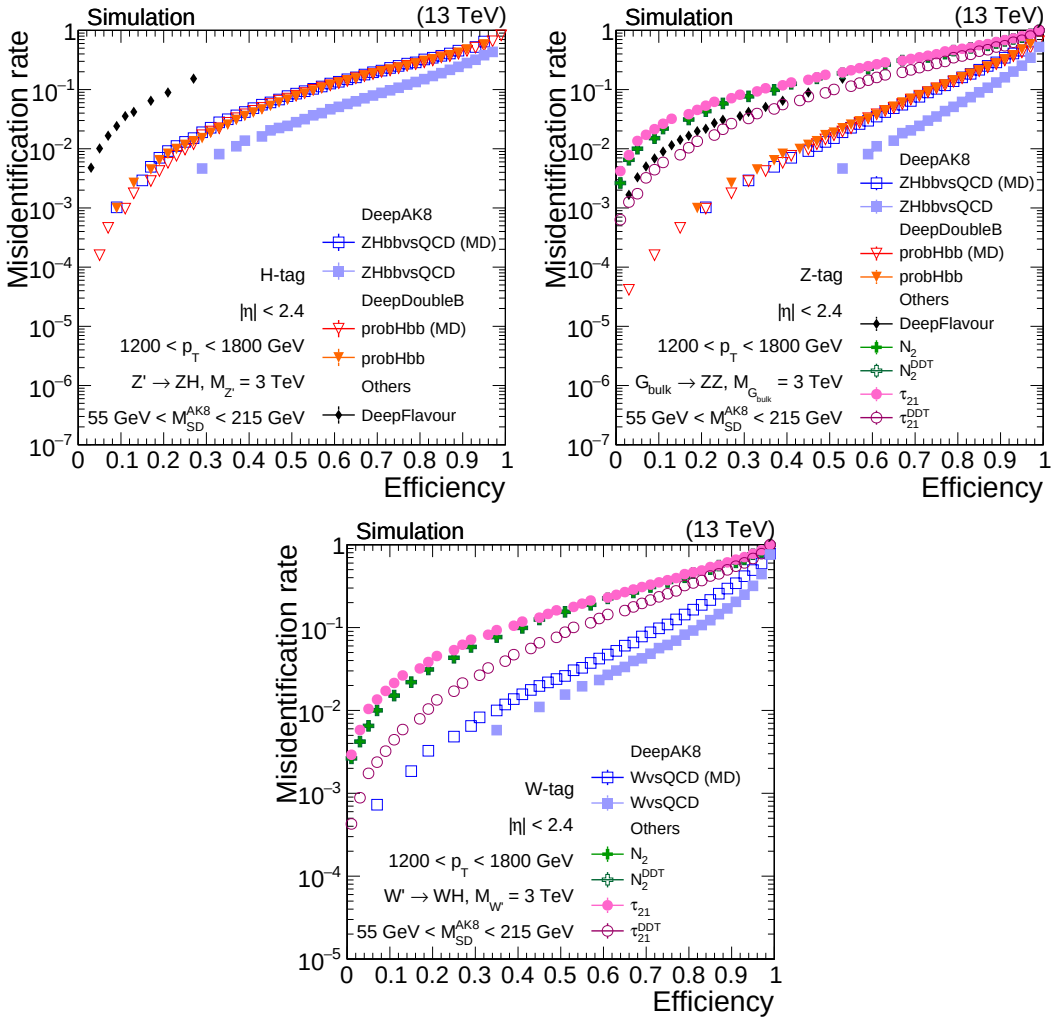


Figure 5.3: ROC curves for different taggers available in the CMS Collaboration to identify (upper left) a H boson, (upper right) a Z boson, and (lower) a W boson for $1200 < p_T < 1800\,\text{GeV}$. For the efficiency the large-R jet is called H, Z or W jet if it is matched to the decay products of generator level (two quarks) by $\Delta R < 0.8$. A signal sample with a resonance mass of 3 TeV is considered. The efficiency (misidentification rate) is defined as the number of events after the tagger is applied divided by the number of events before the tagger is applied in the signal sample (QCD multijet sample).

introduced in Sec. 5.5 and 5.4, respectively. The DeepDoubleB tagger is a deep-neural-network with 27 properties of the jet, involving the flight direction of the b quark, 8 properties of up to 50 charged constituent of the jet and information about the secondary vertices. The variable N_2 is a ratio of energy correlation function which in general shows a similar performance as τ_{21} but is more stable against p_T and jet mass.

The upper left of Fig. 5.3 shows the most recent taggers available to identify a H boson, the upper right row to identify a Z boson and the lower to identify a W

boson. Details about the individual taggers can be found in Ref. [126, 130, 132, 138]. Comparing the different H taggers (upper left in Fig. 5.3) shows that the DeepAK8 ZHbbvsQCD tagger results in the highest efficiency for a certain misidentification rate and shows therefore the best performance. The DeepAK8 ZHbbvsQCD tagger shows also the best performance for identifying Z bosons (upper right in Fig. 5.3). The DeepAK8 WvsQCD tagger shows the best performance in identifying W bosons.

5.5.3 2D-Decorrelation of the DeepAK8 Tagger

The sensitivity of the all-hadronic diboson search is improved if the taggers are decorrelated in p_T and m_j and thus do not introduce a non-monotonically falling behaviour in m_j and m_{jj}. In Fig. 5.3 the DeepAK8 taggers show the best performance, but these taggers are not decorrelated. A mass-decorrelated version (denoted with MD and explained in Sec. 5.5) is available with a slightly reduced performance. Therefore, in the following the DeepAK8 MD taggers are compared to standard taggers like τ_{21}^{DDT}. In addition, a new decorrelation method (2D-decorrelation) is introduced and applied to the nominal DeepAK8 tagger.

For the 2D-decorrelation method, first introduced in Ref. [136], a three dimensional histogram (ρ, p_T, discriminator) for the leading-p_T jet is filled. From that a 2D map is derived for each misidentification rate X% that is used in the analysis. The map is generated by finding the requirement on the discriminator of a certain tagger leading to a X% misidentification rate in each (ρ, p_T) bin. One example map for the DeepAK8 ZHbbvsQCD and WvsQCD tagger for a 5% misidentification rate can be seen in Fig. 5.4. The z-axis represents the requirement on the discriminator of the tagger to result in a 5% misidentification rate.

The performance of this new decorrelation method is evaluated in the difference of the soft drop mass distribution of the leading-p_T jet with and without a requirement of the tagger (mass sculpting).

The result can be seen in Fig. 5.5 for the H tagger and in Fig. 5.6 for the W tagger. Figure 5.5 (upper) shows the mass sculpting for the DoubleB tagger, which has a similar performance as the DeepAK8 ZHbbvsQCD MD tagger. The mass sculpting is less than 10% for misidentification rates above 3% over the whole mass range. The lower row shows the DeepAK8 ZHbbvsQCD tagger with two different decorrelation methods. The left plots shows the method implemented in the tagger training itself (mass-decorrelation, MD), whereas the right plot shows the decorrelation method with the maps explained above. The former method introduces a peak and a dip in the soft drop mass spectrum when requiring a certain value of the discriminator. The

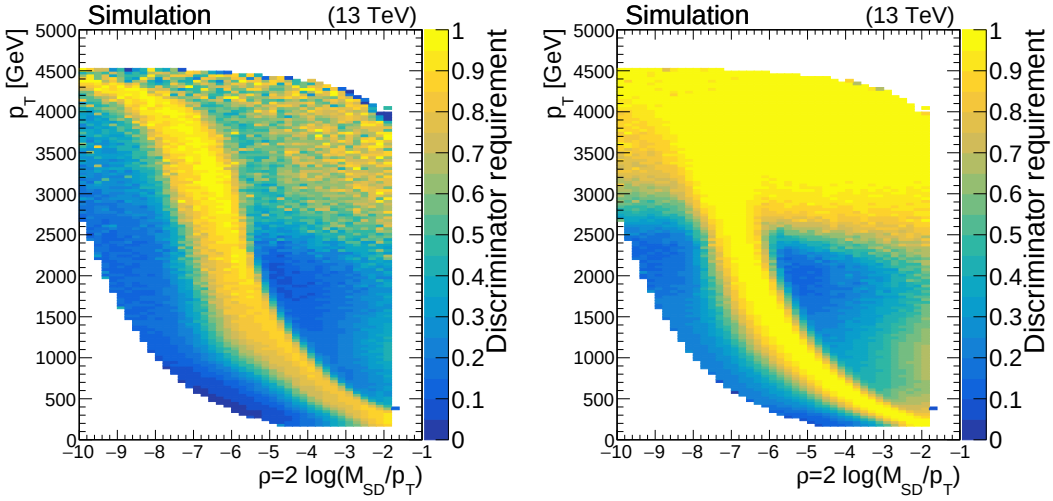


Figure 5.4: 2D map derived for (left) the DeepAK8 ZHbbvsQCD tagger and (right) the DeepAK8 WvsQCD tagger for a 5% misidentification rate for QCD multijet simulation.

2D-decorrelation does not introduce this behavior and results in a mass sculpting of less than 10% over the whole mass range, compatible with the results of the DoubleB tagger. However, the signal efficiency for the 2D-decorrelated DeepAK8 tagger is up to 10% higher compared to the DoubleB tagger.

Figure 5.6 (upper) shows the mass sculpting for the N_2^{DDT} tagger, which has a $\mathrm{p_T}$ dependent performance visible in Fig. 5.3 (middle row). The difference between the nominal spectrum and the different misidentification rates is less than 10%. Fig. 5.6 (lower row) shows the comparison between (left) DeepAK8 WvsQCD MD tagger and (right) DeepAK8 WvsQCD 2D-decorrelated tagger. DeepAK8 WvsQCD MD tagger shows a difference of up to 50% for high soft drop masses, whereas the difference for the DeepAK8 WvsQCD 2D-decorrelated tagger in the soft drop mass does not exceed 12%. Although the N_2^{DDT} tagger shows a better performance in the soft drop mass sculpting, the DeepAK8 WvsQCD 2D-decorrelated tagger has an efficiency that is around 10% higher than the one of the N_2^{DDT} tagger.

The last study concerns the invariant mass of the two leading-$\mathrm{p_T}$ jets. As described above the τ_{21}^{DDT} tagger shows a difference in the m_{jj} spectrum between the nominal spectrum and the spectrum obtained for different misidentification rates (see Fig. 5.2). The invariant mass spectrum of the two 2D-decorrelated machine learning (ML) taggers can be seen in Fig. 5.7. Both 2D-decorrelated taggers show small fluctuations in the invariant mass sculpting, but do not show a systematic shift as visible for τ_{21}^{DDT}.

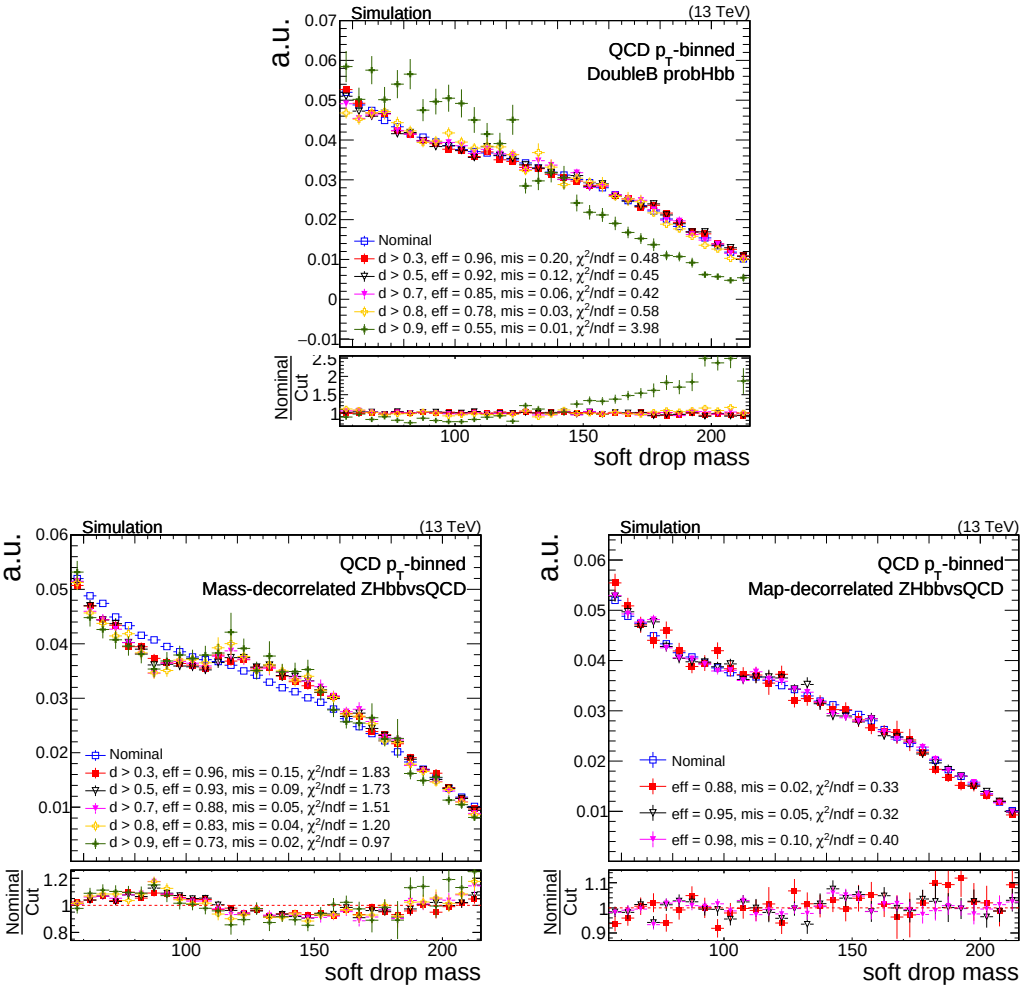


Figure 5.5: Soft drop mass of the leading-p_T large-R jet in QCD multijet simulation for (upper) DoubleB probHbb, (lower left) DeepAK8 ZHbbvsQCD MD, and (lower right) DeepAK8 ZHbbvsQCD 2D-decorrelated. Different requirements on the discriminator corresponding to different misidentification rates (mis) and signal efficiencies (eff) are shown and compared to the nominal (without any requirement). The χ^2/ndf is calculated for each spectrum with respect to the nominal spectrum. The lower panel shows the ratio between the nominal spectrum and one of the spectra with a requirement on the discriminator.

5.5.4 Summary

Three studies were performed to decide which tagger should be used for the diboson all-hadronic analysis. The ROC curves indicate that the DeepAK8 taggers without the mass-decorrelation give the highest efficiency at the same misidentification rate. A 2D-decorrelation method was presented to decorrelate the DeepAK8 tagger in m_j and m_{jj}. The mass sculpting in the soft drop mass of both 2D-decorrelated taggers

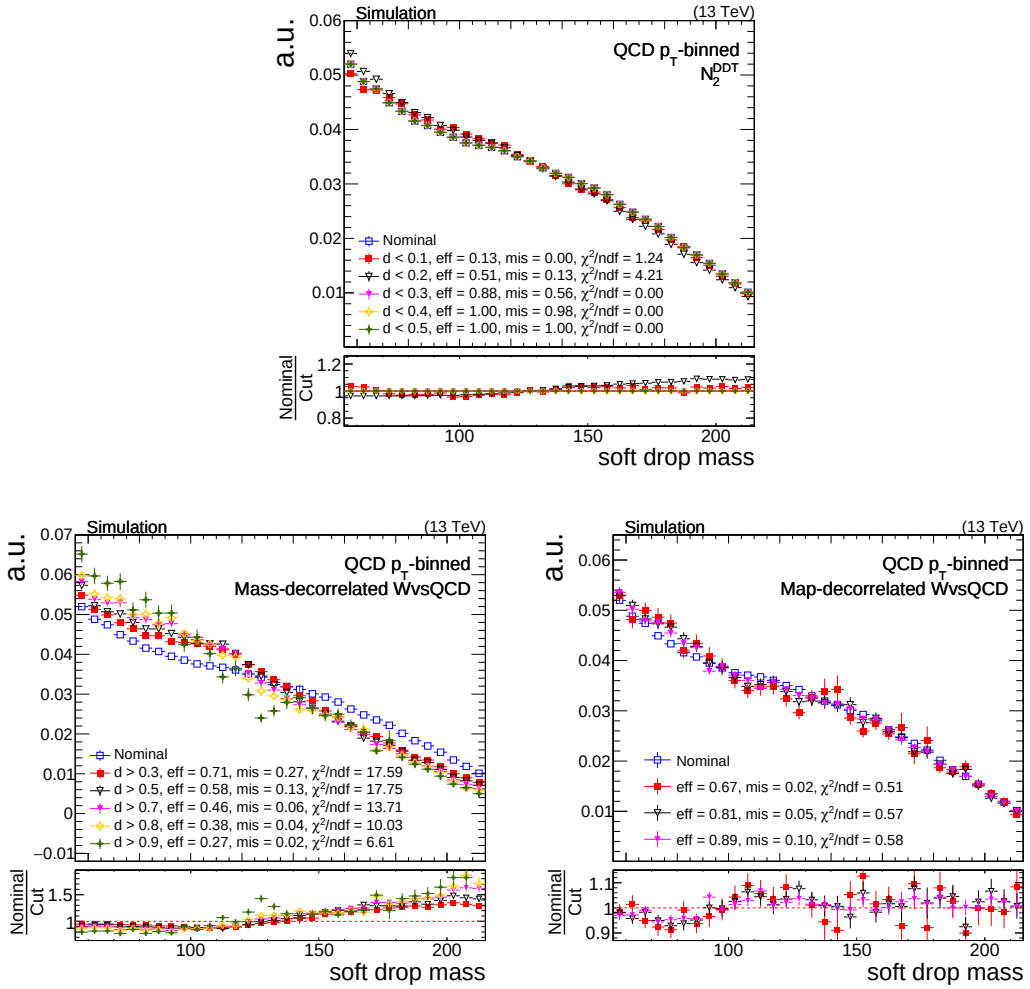


Figure 5.6: Soft drop mass of the leading-p_T large-R jet in QCD multijet simulation for (upper) N_2^{DDT}, (lower left) DeepAK8 WvsQCD MD, and (lower right) DeepAK8 WvsQCD 2D-decorrelated. Different requirements on the discriminator corresponding to different misidentification rates (mis) and signal efficiencies (eff) are shown and compared to the nominal (without any requirement). The χ^2/ndf is calculated for each spectrum with respect to the nominal spectrum. The lower panel shows the ratio between the nominal spectrum and one of the spectra with a requirement on the discriminator.

is compatible or smaller than for all other taggers tested. The last study, the mass sculpting in m_{jj}, showed that the 2D-decorrelation results in no systematic effect in all mass spectra as seen for τ_{21}^{DDT}. Based on this studies, the 2D-decorrelated DeepAK8 taggers are used in the diboson analysis. These lead to an approximate improvement of up to 30% in the expected sensitivity on the upper cross section limit [139, 140].

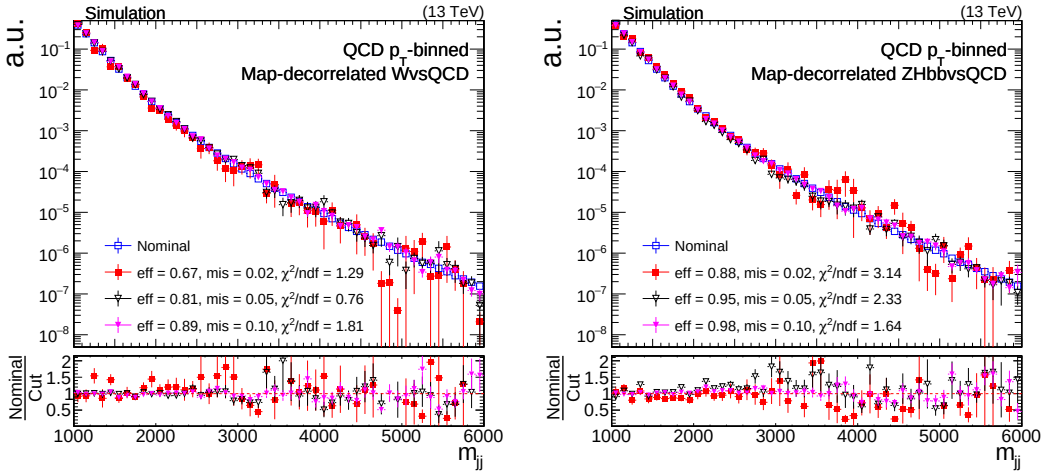


Figure 5.7: The invariant mass of the two leading-p_T large-R jets in QCD multijet simulation for (left) the DeepAK8 WvsQCD 2D-decorrelated tagger and (right) the DeepAK8 ZHbbvsQCD 2D-decorrelated tagger. Different requirements on the discriminator corresponding to different misidentification rates (mis) and signal efficiencies (eff) are shown and compared to the nominal (without any requirement). The χ^2/ndf is calculated for each spectrum with respect to the nominal spectrum. The lower panel shows the ratio between the nominal spectrum and one of the spectra with a requirement on the discriminator.

5.6 Missing Transverse Momentum

The neutrino is one SM particle that does not interact with the detector material and a sophisticated method is needed to evaluate its energy and momentum. The sum of the transverse momenta of the initial particles in each pp collision is zero. Due to momentum conservation the sum of the transverse momenta of all particles originating from this collision should be zero again. However, due to particles that leave no signature in the detector the sum is not zero. The imbalance is called the missing transverse momentum

$$p_T^{\text{miss, raw}} = |\vec{p}_T^{\,\text{miss, raw}}| = |-\sum_i \vec{p}_{T,i}|, \tag{5.10}$$

where i runs over each PF particle in the event. The calculation of p_T^{miss}does not only contain the neutrinos momentum, but also depends on the (mis-)measurement of the jets momenta. Therefore, a correction is applied to account for possible miscalibration

of the jets by propagating the changes due to jet calibration to $\mathrm{p_T^{miss}}$

$$\vec{p}_T^{\,\mathrm{miss}} = \vec{p}_T^{\,\mathrm{miss,\ raw}} - \sum_{\mathrm{jets}} (\vec{p}_{\mathrm{T,\ jet}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,\ jet}}), \tag{5.11}$$

where $\vec{p}_{\mathrm{T,\ jet}}^{\,\mathrm{corr}}$ is the jet energy calibrated jet and $\vec{p}_{\mathrm{T,\ jet}}$ is the uncalibrated jet.

6 | Pileup Mitigation Techniques

The LHC is continuously increasing its instantaneous luminosity, in order to collect more data to search for rare physics phenomena. Improved beam focusing results in more collisions per bunch crossing, hence more instantaneous luminosity. All collisions that occur in addition to the main collision are called PU. The particles coming from PU collisions add energy to all detector components. This leads to additional PF candidates in the event affecting several variables of the event, most notably the reconstruction of jets. For example, the peak of the mass spectrum of a large-R jet in a $Z' \to t\bar{t}$ simulation sample with $M_{Z'} = 2\,\text{TeV}$ shifts from 180 to 380 GeV when comparing no PU with an average PU of 60 interactions. In addition, the jet mass resolution becomes significantly worse [141]. However, not only the jet mass is influenced by PU, but also other jet substructure variables like N-subjettiness, the reconstruction of missing transverse momentum, jet multiplicity, jet energy resolution and the lepton reconstruction through isolation variables.

The mean number of pp collisions per bunch crossing can be found in Fig. 6.1 for Run 2 (2016–2018) data with an inelastic pp cross section ($\sigma_{\text{in}}^{\text{pp}}$) of 69.2 mb. This value was measured by the CMS [142] and ATLAS [143] Collaborations. Before the start of Run 2 all LHC experiments agreed on $\sigma_{\text{in}}^{\text{pp}} = 80$ mb as a baseline, which would result in an average PU of 34 rather than 29 (Fig. 6.1) for Run 2. However, analyses performed by CMS use $\sigma_{\text{in}}^{\text{pp}} = 69.2$ mb, which is therefore used in this thesis. During data taking in 2016–2018 the average number of collisions was around 30, during data taking in 2021–2024 a PU between 50 and 70 is expected and during the operation of the HL-LHC it is foreseen to increased to up to 200.

The average amount μ of PU per bunch crossing is given by

$$\mu = \frac{\mathcal{L}\sigma_{\text{in}}^{\text{pp}}}{f_{\text{rev}}}, \tag{6.1}$$

where $\mathcal{L}$ is the instantaneous luminosity, f_{rev} is the frequency at which collision happen. This estimation of the number of vertices per bunch crossing can be compared to reconstructed number of vertices, which is affected by the vertex reconstruction ef-

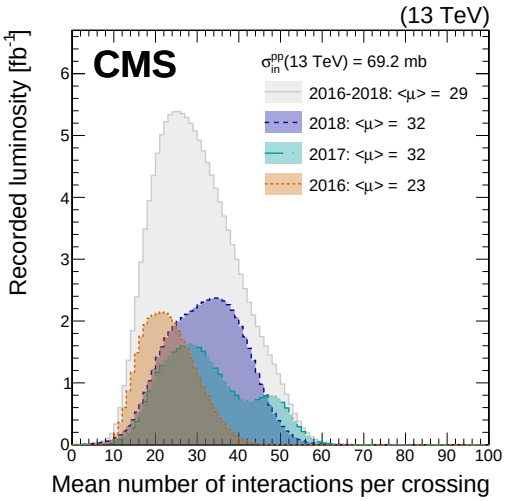


Figure 6.1: Distribution of the mean number of inelastic interactions per crossing (pileup) in data for pp collisions in 2016 (dotted orange line), 2017 (dotted dashed light blue line), 2018 (dashed navy blue line), and integrated over 2016–2018 (solid grey line). A total inelastic pp collision cross section of 69.2 mb is chosen. The mean number of inelastic interactions per bunch crossing is provided in the legend for each year.. Published in Ref. [4].

ficiency. In order to simulate the amount of PU accordingly, μ is taken as the mean of a Poisson distribution from which the actual simulated amount of PU is constructed. Therefore, the inelastic cross section directly enters the simulation in each analysis.

The easiest way to mitigate the effect of PU is to remove an average amount of pileup energy from the p_T of a jet based on the PU activity in the event. This jet-area based approach was used during the data taking in 2010–2012 by the CMS and ATLAS Collaborations [144, 145]. This technique shows a good performance for kinematic jet variables [28], however, it does not mitigate the effect of PU in the jet mass and substructure quantities of a jet. Therefore, a new technique based on track-vertex association was studied for the data taking [28], called CHS. As the name suggests, charged particles that are associated to a PU vertex are removed from the jet clustering. In this way, the effect of charged PU can be mitigated. To reduce the effect of residual neutral PU a jet-area-based correction [28], similar to the average PU subtraction mentioned above, is applied to the jets. CHS shows an improvement in the jet p_T resolution as well as in the jet mass resolution [123]. A remaining PU dependence in several variables such as the jet mass is visible mainly due to not properly removing local fluctuations in the neutral PU contributions.

Another problem are jets that are reconstructed nearly solely from PU particles, so-called PU jets. In a Z+jets $\rightarrow \mu\mu$+jets simulation sample, one real jet is expected which

balances against the Z boson. On average, three additional PU jets with $p_T > 25\,\text{GeV}$ are reconstructed at 40 collisions per bunch crossing [146]. The PU jets can consist out of particles from one or more PU vertices, which results into two distinct features of PU jets that are different from LV jets. PU jets have a large fraction of charged particles that do not point to the LV. In addition, LV jets have most of their energy close to the jet axis, while PU jets have a broader energy distribution within the jet. A boosted decision tree (BDT) is trained using 15 different input variables based on tracking and jet shapes to identify and reject PU jets, which is known as the PU jet ID [146].

Since both techniques, CHS and PU jet ID are most effective when tracking information is available, a third technique, called PUPPI, was developed and commissioned based on 2016 data [4] in order to mitigate the effect of neutral PU particles.

6.1 Pileup Per Particle Identification

The PUPPI algorithm [147] consist of three main steps:

1. Define a variable able to discriminate between charged particles from a PU vertex from charged particles from the LV.

2. Demonstrate that this variable is distributed similarly for neutral particles.

3. Scale the four-momentum of each particle in the event with its corresponding weight calculated from the discriminating variable.

Particles from PU interactions tend to be isolated[2], whereas LV particles are surrounded by many other high-energetic particles. Therefore, an intuitive choice to discriminate between LV and PU particles is the isolation variable α_i for a particle i:

$$\alpha_i = \log \sum_{j \neq i,\, \Delta R_{ij} < R_0} \left(\frac{p_{\mathrm{T}j}}{\Delta R_{ij}\,\text{GeV}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from LV,} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles,} \end{cases} \tag{6.2}$$

where j are the other particles in a cone around the particle in question and $p_{\mathrm{T}j}$ is the transverse momentum of particle j in units of GeV.

Two different samples are compared to test the data-to-simulation agreement for the main variables used within the PUPPI algorithm: the jet sample and the PU

[2] To measure the isolation of a PF particle, the transverse momenta of PF particles in a cone around the particle are summed.

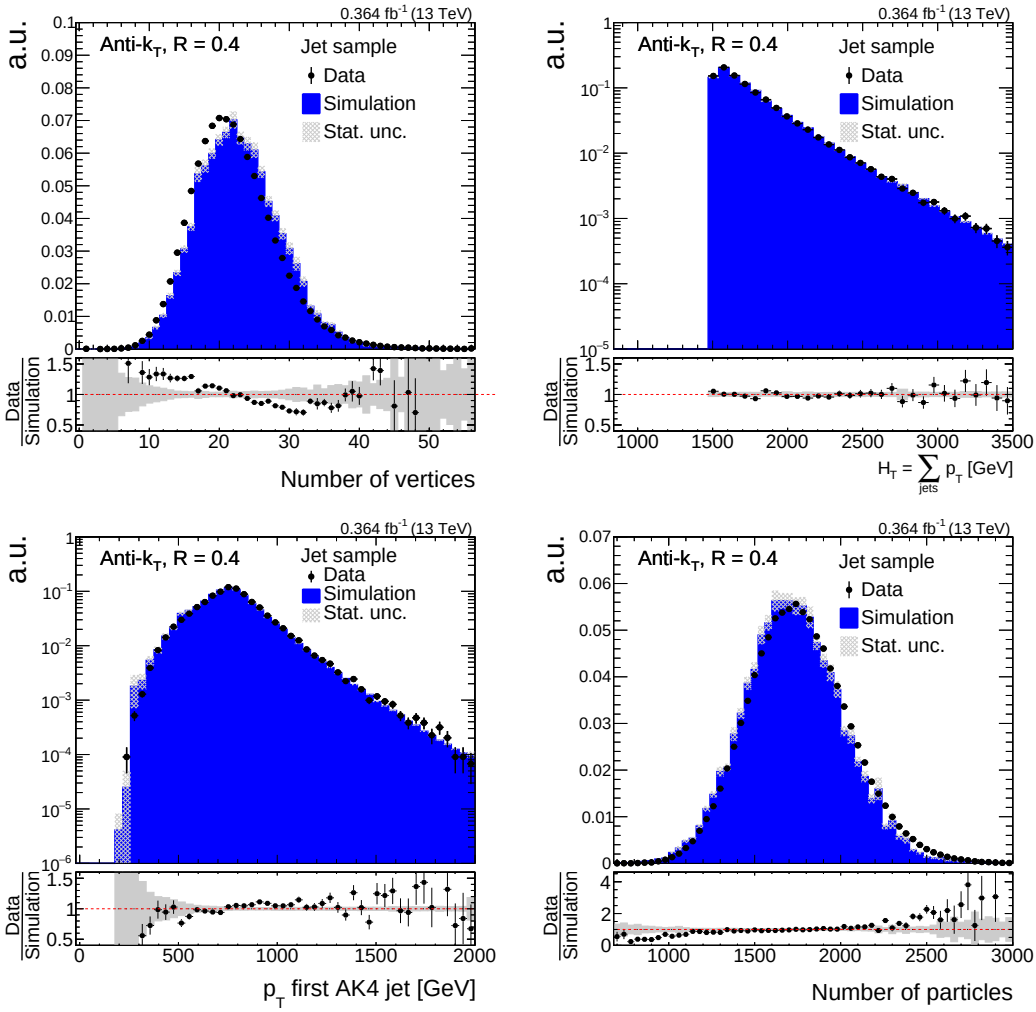


Figure 6.2: Data-to-simulation comparison of the (upper left) number of vertices, (upper right) scalar sum of all jets H_T, (lower left) p_T of the first small-R jet and (lower right) the number of all particles in an event. The black marker show the data of the jet sample, the blue area shows the QCD multijet simulation.

sample. The jet sample uses a subset of the data recorded in 2016 corresponding to an integrated luminosity of $0.364\,\mathrm{fb}^{-1}$ and requires a the scalar sum (H_T) of the p_T of jets with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 3$ to be $H_T > 1000\,\mathrm{GeV}$ and is compared to QCD multijet simulation. Control distribution of the jet sample are shown in Fig. 6.2. The distribution of the number of vertices (upper left) shows the know discrepancy between data and simulation, while event and object variables, as the variable H_T (upper right) and the p_T of the leading jet (lower left), show a good agreement between data and simulation. The number of particles (lower right) shows a reasonable agreement in the bulk of the distribution, while especially at low numbers of particles the simulation

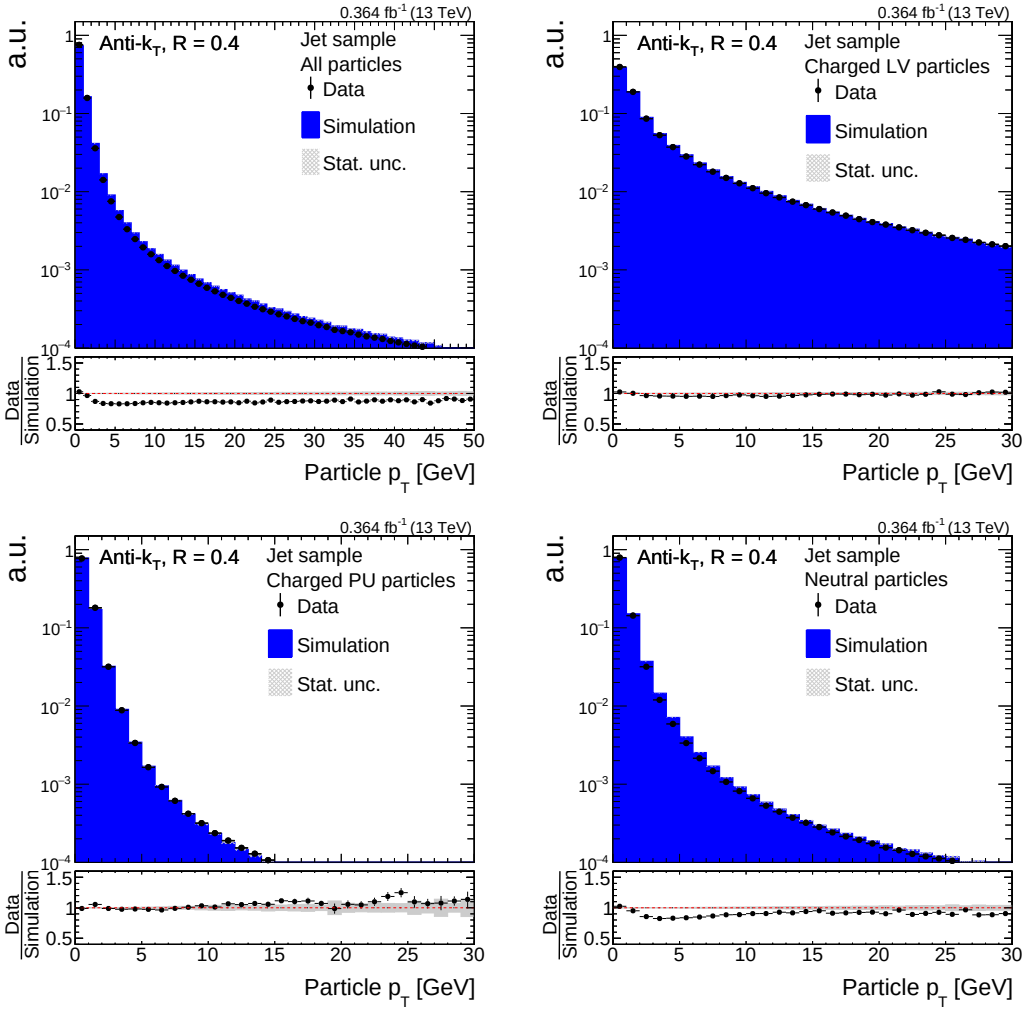


Figure 6.3: Data-to-simulation comparison of the p_T distribution of (upper left) all particles, (upper right) charged particles associated to the LV, (lower left) charged particles associated to the PU vertex and (lower right) neutral particles in an event. The black marker show the data of the jet sample, the blue area shows the QCD multijet simulation.

start to diverge from data. In general less particles are simulated in simulation as reconstructed in data. Especially particles at $p_T < 2\,\text{GeV}$ are not simulated well as can be seen in Fig. 6.14. While the charged particles associated to the LV are simulated well all other particles show a disagreement for $p_T < 2\,\text{GeV}$. At $p_T > 2\,\text{GeV}$ the ratio between data and simulation shows a flat behaviour, which shows that both agree in shape, while the normalisation is not corrected for the low p_T contribution.

As a reference, PU data recorded with a zero bias trigger that randomly selects a fraction of the collision events, corresponding to an integrated luminosity of 3.18nb^{-1}, are compared to PU-only simulation and are referred to as the PU sample. The PU-

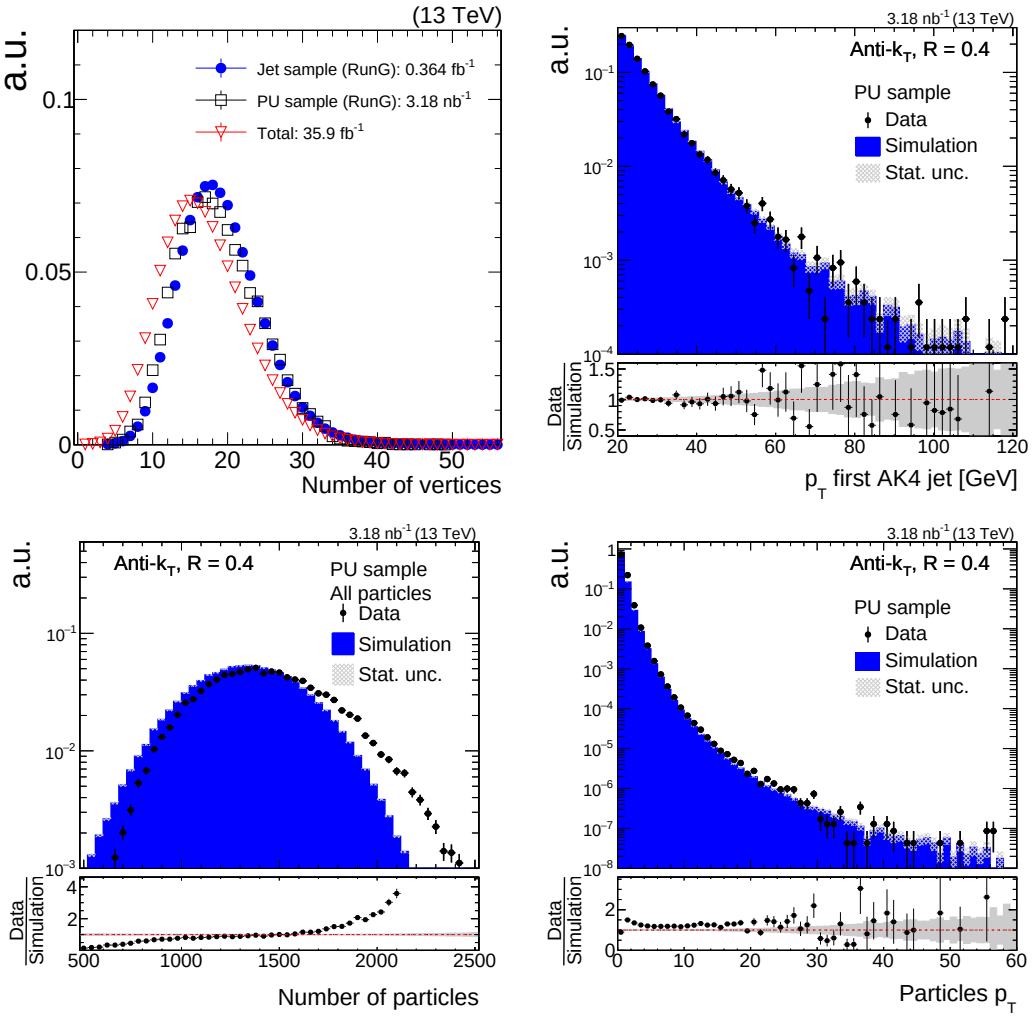


Figure 6.4: (upper left) Comparison of the number of vertices in the full 2016 data set, the jet sample subset and the subset of the PU sample. Data-to-simulation comparison of (upper right) the p_T distribution of leading-p_T jet, (lower left) the number of all particles in the event and (lower right) the p_T distribution of all particles in the event.

only simulation is generated by a single-neutrino production process. These simulation contains only pileup and detector noise since the neutrino is not interacting with the detector material. Both subsets show a similar PU profile as shown in Fig. 6.4 (upper left). The p_T distribution of the leading jet is shown in Fig. 6.4 (upper right). A good agreement between data and simulation is visible. The number of particles and the p_T distribution of particles (Fig. 6.4 (lower)) shows the same behaviour as the jet sample.

A data-to-simulation comparison of the α variable is displayed in Fig. 6.5 (upper row). While charged particles in Fig. 6.5 (upper left) are divided into LV and PU

particles based on tracking information[3], neutral particles are shown inclusively.

A good separation power between LV ($\alpha > 8$) and PU ($\alpha < 8$) is visible. However, charged particles from the LV show a double peak structure. While large values of α represent the particles within a jet originating from the LV, lower α values indicate isolated particles not relevant for physics object reconstruction. The median $\bar{\alpha}_{\mathrm{PU}}$ and RMS $\alpha_{\mathrm{PU}}^{\mathrm{RMS}}$ of the charged PU particle distribution (blue curve) is used to calculate a signed χ_i^2 for each particle in the event,

$$\text{signed } \chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{\mathrm{PU}})|\alpha_i - \bar{\alpha}_{\mathrm{PU}}|}{\left(\alpha_{\mathrm{PU}}^{\mathrm{RMS}}\right)^2}. \tag{6.3}$$

The median as well as the RMS of the charged PU particles are qualitatively similar to the ones from neutral PU particles as shown in Fig. 6.5 (upper right). The resulting χ^2 for neutral particles is shown in Fig. 6.5 (lower left).

A weight $w_i = F_{\chi^2,\mathrm{NDF}=1}(\text{signed } \chi_i^2)$ is calculated, where $F_{\chi^2,\mathrm{NDF}=1}$ is the cumulative distribution function of the χ^2 distribution. The weight for each neutral particle in the event is depicted in Fig. 6.5 (lower right). The four-momentum of each particle is scaled with the corresponding weight, whereas a weight close to zero indicates a PU origin. The disagreement of the weight between data and simulation at low values is due to the mismodeling of low-$\mathrm{p_T}$ PU particles in simulation. However, these particles receive a small weight and have therefore a negligible impact on physics analyses.

Particles with a weighted transverse momentum $w_i \cdot p_{\mathrm{T}i} < (A + B \cdot N_{\mathrm{vertices}})\,\mathrm{GeV}$, where N_{vertices} is the number of reconstructed vertices, are rejected, since these are most likely PU particles or originating from noise. The parameters A and B are η-dependent tunable parameters and are given in Tab. 6.1. While in the regions $|\eta| < 2.5$ and $2.5 < |\eta| < 3$ the parameters are chosen such that the ratio between the reconstructed jet energy and the true jet energy is near unity as a function of the number of interactions, in the region $3 < |\eta| < 5$ the parameters are chosen to optimize the $\mathrm{p_T^{miss}}$ resolution.

Challenges that occur when applying PU mitigation techniques are either removing not enough PU particles or too many LV particles. Considering the energy of a jet, removing too many LV particles results in a degraded energy resolution as well as in a decreased energy response. In the context of tagging, removing not enough PU particles results in PU dependent efficiencies and misidentification rates. Furthermore, the lepton isolation is affected by either case. While removing not enough PU leads to

[3] All charged particles can be either used in the fit of the LV, used in the fit of a PU vertex or not used in any vertex fit. The last category can be further split by distance in z-direction d_{z} to the LV. PUPPI LV particles are defined as particles used in the LV fit or are not used in any fit and $d_{\mathrm{z}} < 0.3\mathrm{cm}$, whereas CHS uses all charged particles that are not used in a PU vertex fit.

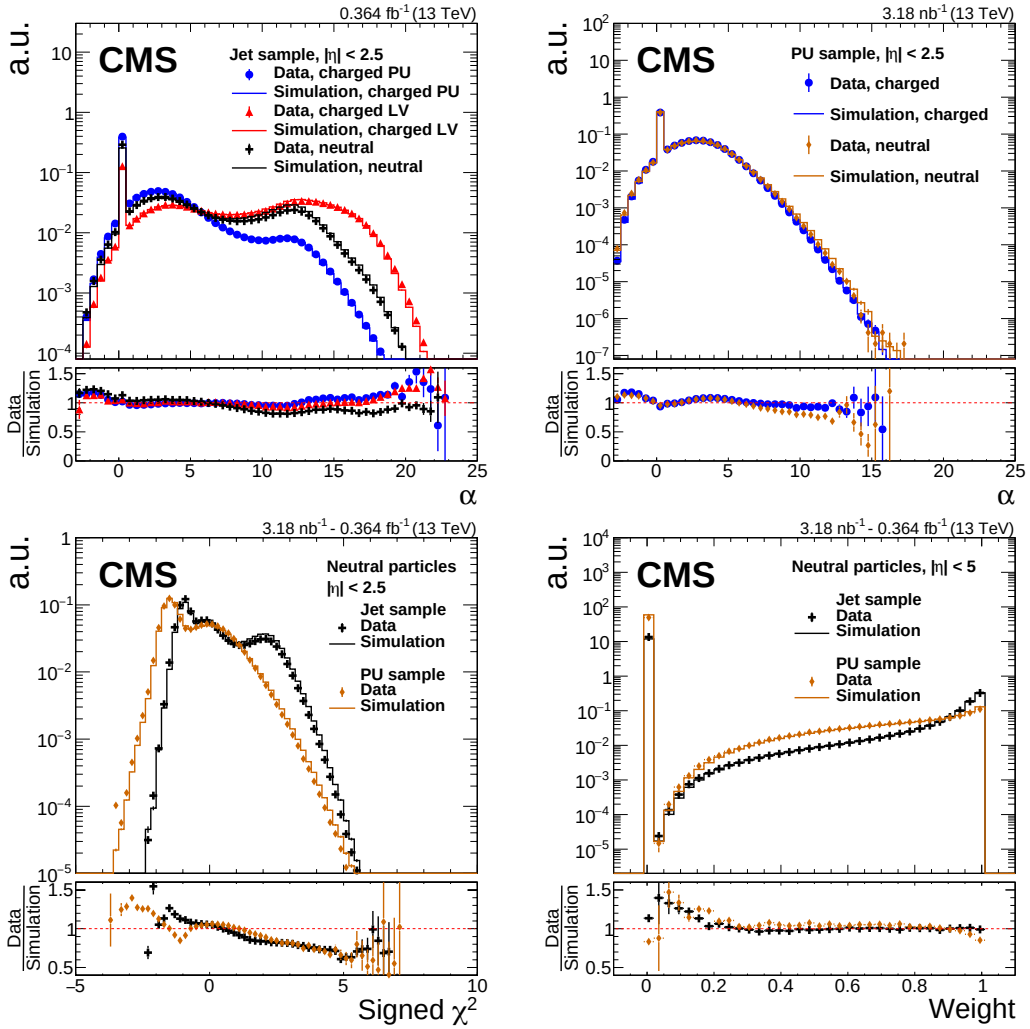


Figure 6.5: Data-to-simulation comparison for three different variables of the PUPPI algorithm. The markers show a subset of the data taken in 2016 of the jet sample and the PU sample, whereas the solid lines are QCD multijet simulations or PU-only simulation. The lower panel of each plot shows the ratio of data to simulation. Only statistical uncertainties are displayed. The upper left plot shows the α distribution in the jet sample for charged particles associated with the LV (red triangles), charged particles associated with PU vertices (blue circles), and neutral particles (black crosses) for $|\eta| < 2.5$. The upper right plot shows the α distribution in the PU sample for charged (blue circles) and neutral (orange diamond) particles. The lower left plot shows the signed $\chi^2 = (\alpha - \overline{\alpha}_{\mathrm{PU}})|\alpha - \overline{\alpha}_{\mathrm{PU}}|/(\alpha_{\mathrm{PU}}^{\mathrm{RMS}})^2$ for neutral particles with $|\eta| < 2.5$ in the jet sample (black crosses) and in the PU sample (orange diamonds). The lower right plot shows the PUPPI weight distribution for neutral particles in the jet sample (black crosses) and the PU sample (orange diamonds). The error bars correspond to the statistical uncertainty. Published in Ref. [4].

Table 6.1: The tunable parameters of PUPPI optimized for application in 2016 data. Published in Ref. [4].

$\lvert\eta\rvert$ of particle	A [GeV]	B [GeV]
$[0, 2.5]$	0.2	0.015
$[2.5, 3]$	2.0	0.13
$[3, 5]$	2.0	0.13

a PU dependent isolation, hence a reduced identification efficiency, rejecting too many LV particles results in a not trustworthy isolation, hence an increased misidentification rate.

6.2 Comparison of Pileup Mitigation Techniques for Jets

In the following chapter, the focus will be on small-R and large-R jets, as most analyses use them. As mentioned above the challenge when applying PU mitigation techniques is to find the right balance between removing and keeping too many particles. In the following, "number of interactions" (denoted μ), and "number of vertices" (denoted N_{vertices}) are used to quantify the amount of PU. While the number of interactions represent the simulated PU and is only available in simulation, the number of vertices can be determined in both data and simulation. More details on the effect from PU mitigation techniques on the relationship between the number of interactions and N_{vertices} are presented in Sec. 6.2.4.

6.2.1 Jet Energy Resolution

A good measure for the performance of the pileup mitigation techniques is the JER which is defined as the spread of the response distribution, that is to a very good approximation Gaussian. Jet energy corrections are applied to the reconstruction-level jets such that the ratio of reconstruction and particle-level jet $\mathrm{p_T}$ (the response) is on average one. The JER is shown in Fig. 6.6 for jets clustered with the anti-k_{T} algorithm and with one of the PU mitigation techniques, none (referred to as PF), CHS or PUPPI, applied. While jets without any pileup mitigation applied have an increased JER especially at low $\mathrm{p_T}$, PUPPI shows the best resolution at low $\mathrm{p_T}$ and $|\eta| < 2.5$ for both, small-R and large-R jets. PUPPI outperforms CHS at low $\mathrm{p_T}$ and $|\eta| < 2.5$ since it acts directly on neutral particles and keeps charged particles not associated to any

vertex only, other than CHS, if $d_z < 0.3$ cm to the LV.

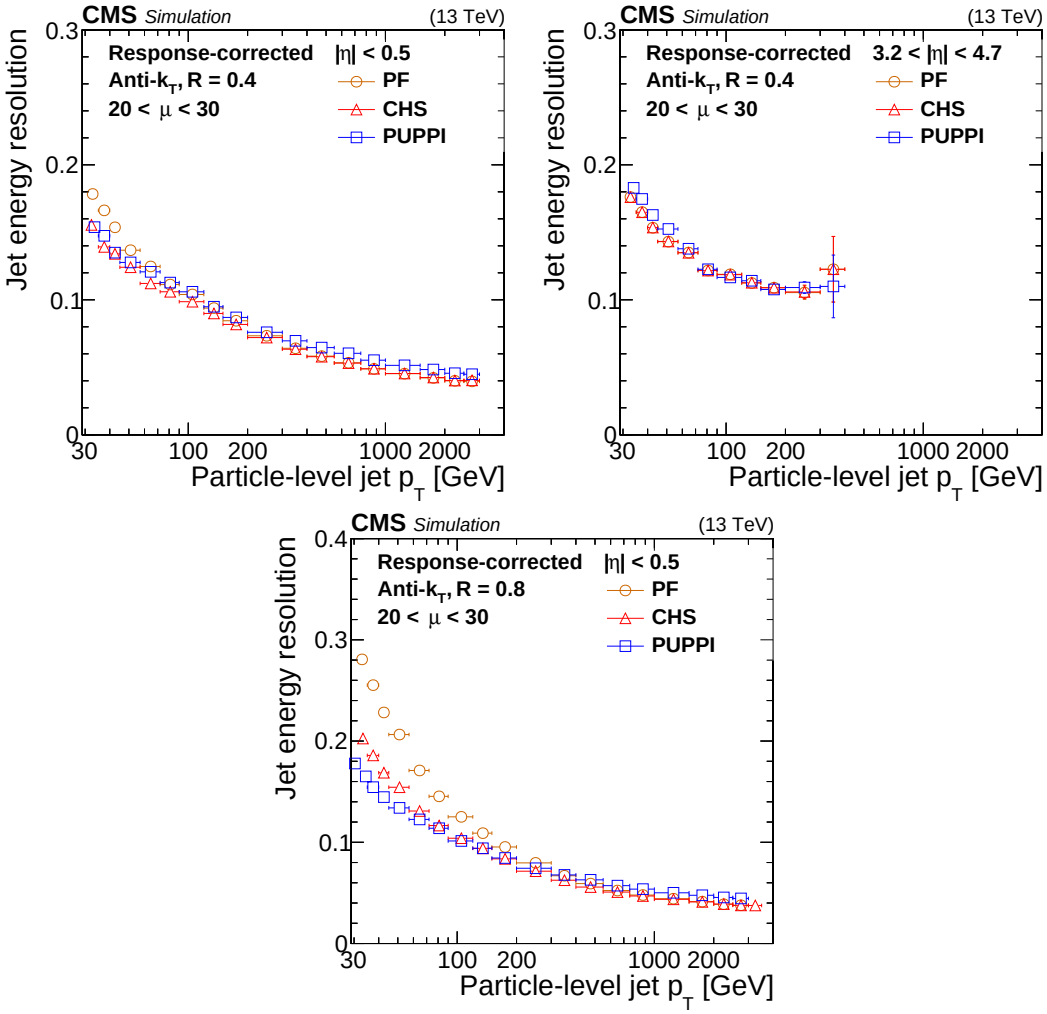


Figure 6.6: Jet energy resolution as a function of the particle-level jet p_T for PF jets (orange circles), PF jets with CHS applied (red triangles), and PF jets with PUPPI applied (blue squares) in QCD multijet simulation. The number of interactions is required to be between 20 and 30. The resolution is shown for AK4 jets with $|\eta| < 0.5$ (upper left) and $3.2 < |\eta| < 4.7$ (upper right), as well as for AK8 jets with $|\eta| < 0.5$ (lower). The error bars correspond to the statistical uncertainty in the simulation. Published in Ref. [4].

However, at high p_T, CHS outperforms PUPPI. The reason is, that at high p_T the track-vertex assignment is degraded such that PUPPI rejects too many charged particles. As a consequence α_i calculated for neutral particles is too low, hence neutral particles get a too small weight. To overcome this the PUPPI algorithm has been improved (PUPPI Chihuahua), leading to a resolution comparable to the one of CHS as shown in Fig.6.7 for different η regions. While PUPPI normally keeps charged particles not associated to any vertex only if $d_z < 0.3$ cm, with the new tune all charged

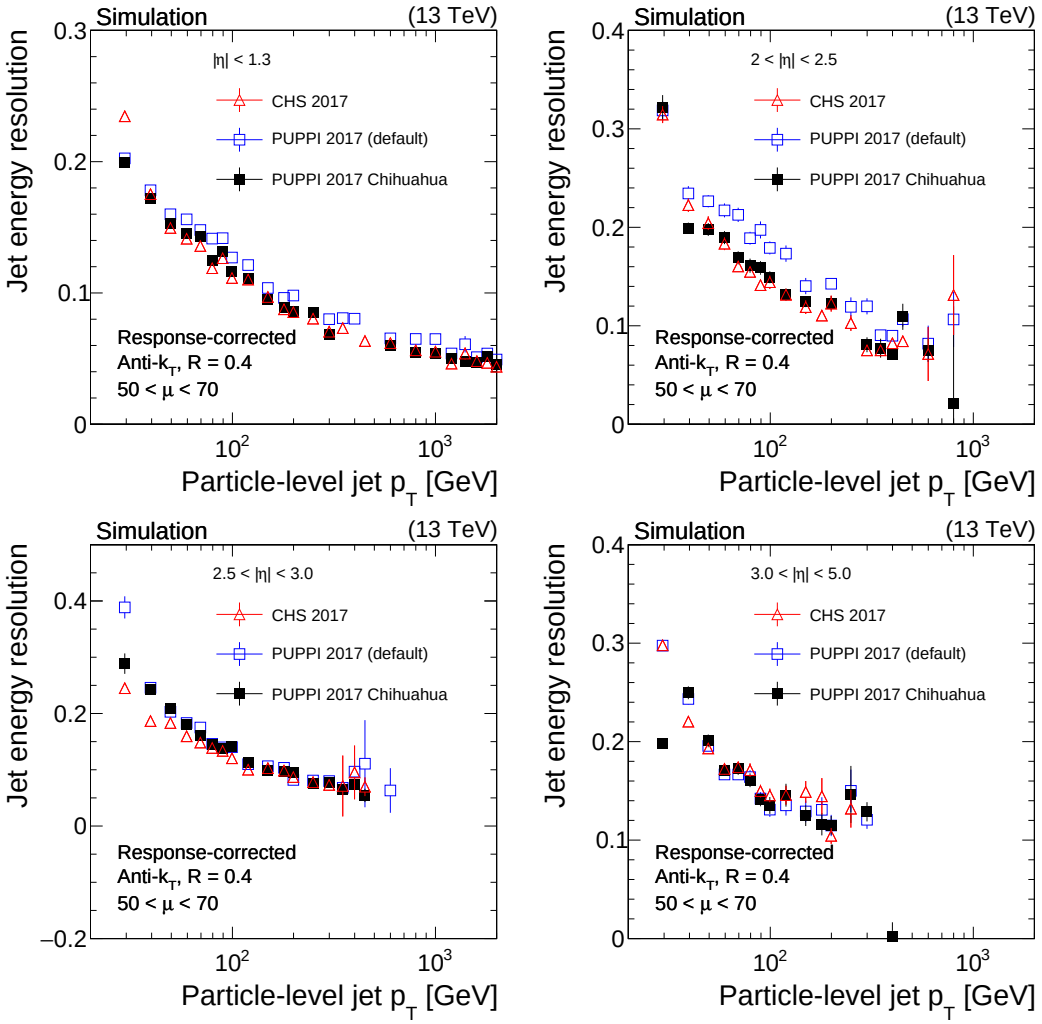


Figure 6.7: Jet energy resolution as a function of the particle-level jet p_T for PF jets with CHS applied (red triangles), and PF jets with PUPPI applied (blue open squares) and PF jets with the new tune of PUPPI (black closed squares) in QCD multijet simulation in different η regions.

particles not associated to any vertex and with $p_T > 20\,\text{GeV}$ are kept. Particles with $p_T > 20\,\text{GeV}$ are most likely LV particles due to their high p_T. This requirement avoids the misidentification of LV particles as PU particles due to the limited vertex association resolution.

The requirement on d_z is not reliable for particles with $p_T > 20\,\text{GeV}$ and only works for low-p_T particles. For charged particles with $2.5 < |\eta| < 3$ and $p_T < 20\,\text{GeV}$ the d_z requirement is kept. Charged particles, with $|\eta| < 2.5$, not associated to any vertex and with $p_T < 20\,\text{GeV}$ are excluded from the α_i calculation for neutral particles. A dedicated weight is calculated for these charged particles in the same way as it is done for neutral particles. Neutral particles receive a special treatment for $20 < p_T <$

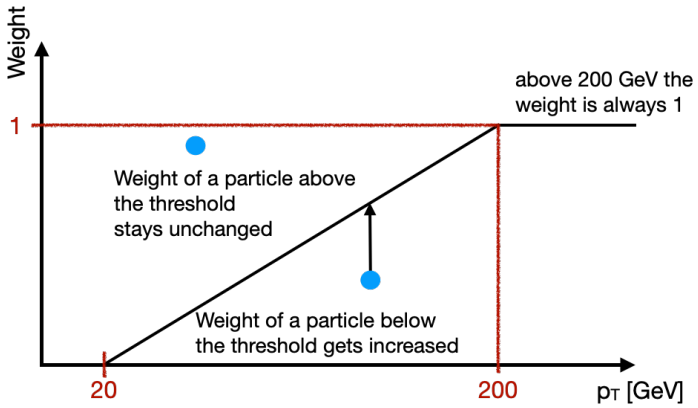


Figure 6.8: Protection applied to neutral particles with a linear function in the range $20 < p_T < 200\,\text{GeV}$ to improve the JER of jets clustered from PF particles after applying PUPPI.

200 GeV. The weights of neutral particles is increased to $w_i = \frac{p_T - 20}{200 - 20}$ as depicted in Fig. 6.8. Both changes lead to a significant improvement of the JER, however a small difference between CHS and PUPPI remains. If charged particles in PUPPI are treated the same as in CHS, a PU dependence of the jet mass reconstruction is observed as seen for CHS and discussed in Sec. 6.2.5.

6.2.2 Pileup Jet Rejection

PU particles do not only influence the jet energy but also create additional jets composed from mainly PU particles (PU jets). These PU jets affect especially measurements and searches relying on jets with $|\eta| > 2.5$, where PU jets represent a significant fraction. In the region $|\eta| < 2.5$ tracking information is available, which simplifies the identification of PU, whereas in $|\eta| > 2.5$ the identification relies on jet shape information. In order to evaluate the performance of PU jet rejection techniques, the efficiency and purity of such techniques is determined. In simulation a PU jet is defined as a reconstruction-level jet that does not match a particle-level jet from the hard interaction within $\Delta R < 0.4$. Reconstruction-level jets that have a match with a particle-level jet are called LV jet. The efficiency is then defined as the number of reconstructed LV jets divided by the number of all particle-level jets. The efficiency can be reduced if a technique removes too many LV particles or too many LV jets. The purity is defined as the number of LV jets divided by the number of reconstruction-level jets. The purity gets reduced if a technique does not remove enough PU particles or PU jets. In order to exclude possible inefficiencies of the matching procedure due to jet energy resolution when calculating the efficiency and purity, different p_T requirements between particle-

level jets and reconstruction-level jets are applied. For the efficiency the particle-level jets have $p_T > 30\,\text{GeV}$ while the reconstructed-level jets have $p_T > 20\,\text{GeV}$. In this way the p_T reconstruction of the jets can be off by around 30% without affecting the measurement. For the purity reconstructed-level jets have $p_T > 30\,\text{GeV}$ while the particle-level jets have $p_T > 20\,\text{GeV}$, for the same reason.

The comparison of all three techniques used in the CMS Collaboration, CHS, PU Jet ID and PUPPI, is shown in Fig. 6.9. CHS acting only on charged particles from PU shows the best efficiency in both $|\eta| < 2.5$ and $3 < |\eta| < 5$, because it is the technique that removes the least particles and therefore removes the least jets. However, due to this fact the purity for CHS is the smallest compared with other techniques in both regions. When adding the PU jet ID to CHS the efficiency decreases since jets (more PU than LV jets) are removed, but at the same time the purity increases. PUPPI shows a decreased efficiency with respect to CHS since it also acts on neutral PU, but at the same time it shows an improved purity without the need of the PU jet ID. While in the region $|\eta| < 2.5$ PUPPI has the highest purity compared with the other techniques, in the region $3 < |\eta| < 5$ PUPPI is comparable with the loose PU jet ID, which corresponds to a 95% efficiency on reconstruction-level jets originating from quarks. The purity of PUPPI especially in the region $|\eta| > 3$ can be improved by increasing the weighted p_T requirement $w_i\,p_T$. However, that would result in a reduced efficiency. Another approach is to develop a PU jet ID for PUPPI, which has not been studied yet.

The ATLAS Collaboration has developed a different technique, the forward Jet Vertex Tagging, to reject PU jets in the region $|\eta| > 2.5$ [148]. The ATLAS Collaboration is also using tracking-based variables as used in the BDT for the PU jet ID by the CMS Collaboration. In addition, the ATLAS Collaboration is using the fact that the total transverse momentum of each vertex from pp collisions is expected to be conserved. By calculating the missing transverse momentum for a specific PU vertex when associating a specific jet to that vertex a probability of originating from that PU vertex is determined. In addition, a requirement on the energy-weighted average of the timing of the constituent calorimeter cells is placed. The *Pile-up Jet Efficiency* used in Ref. [148] can be calculated from the purity used in Ref. [4] by:

$$\epsilon_{\text{PU}} = \frac{(\epsilon_{\text{PUPPI}} - p_{\text{PUPPI}} \cdot \epsilon_{\text{PUPPI}}) \cdot p_{\text{CHS}}}{(\epsilon_{\text{CHS}} - p_{\text{CHS}} \cdot \epsilon_{\text{CHS}}) \cdot p_{\text{PUPPI}}}, \tag{6.4}$$

where ϵ is the efficiency and p is the purity for the corresponding method. It is assumed that CHS does not reject any jets in $|\eta| > 2.5$, which is reasonable since CHS does not remove any particles in this region. Note, this calculation is not exact but an

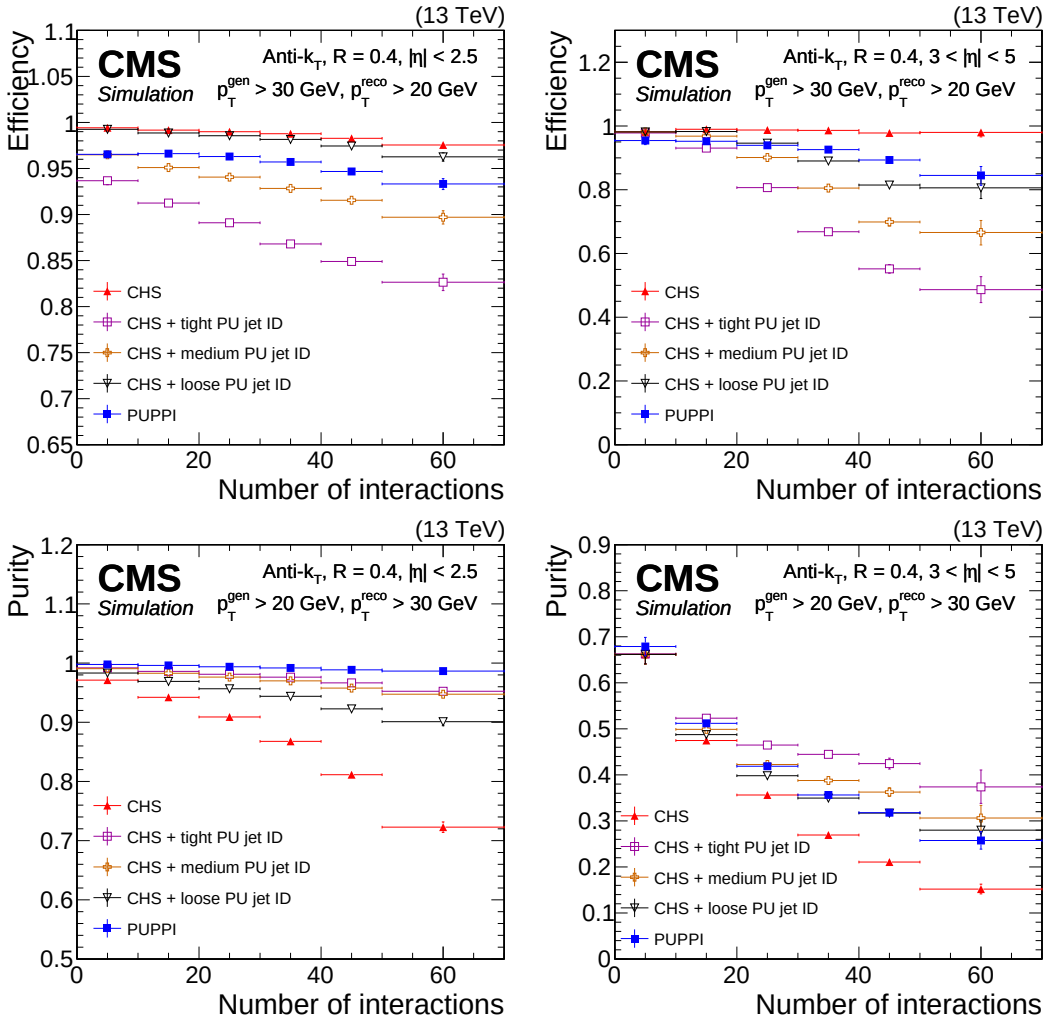


Figure 6.9: The LV jet efficiency (upper) and purity (lower) in Z+jets simulation as a function of the number of interactions for PUPPI (blue closed squares), CHS (red closed triangles), CHS+tight PU jet ID (magenta open squares), CHS+medium PU jet ID (orange crosses), and CHS+loose PU jet ID (black triangles). Plots are shown for AK4 jets $p_T > 20\,GeV$, and (left) $|\eta| < 2.5$ and (right) $|\eta| > 3$. The LV jet efficiency is defined as the number of matched reconstruction-level jets with $p_T > 20\,GeV$ divided by the number of particle-level jets with $p_T > 30\,GeV$ that originate from the main interaction. For the lower plots, the purity is defined as the number of matched particle-level jets with $p_T > 20\,GeV$ divided by the number of reconstructed jets that have $p_T > 30\,GeV$. The error bars correspond to the statistical uncertainty in the simulation. Published in Ref. [4].

approximation to compare the results between CMS and ATLAS. In order to do an exact comparison, both Collaborations would need to define an η region, a p_T range and a specific PU scenario, the same definition on PU jets and the same simulation

would be needed. However, most of these effects have an effect of $\mathcal{O}(1\%)$. The method used in ATLAS shows a compatible rejection power at the same efficiency compared with PUPPI. At an average of 22 interactions, the method from ATLAS has a PU jet efficiency of $\sim 75\%$ at a hard-scattered jet efficiency of $\sim 97\%$, while PUPPI has at the same PU jet efficiency a hard-scattered jet efficiency of $\sim 95\%$. The tight PU jet ID has an PU jet efficiency of $\sim 49\%$ at an efficiency of $\sim 80\%$, while the method used in the ATLAS Collaboration has an efficiency of $\sim 85\%$. Similar results are obtained for $|\eta| < 2.5$.

To evaluate the performance of PU jet rejection techniques in data the PU jet rate is measured. Z+jets events are split into a PU enriched region and a LV enriched region based on the azimuthal distance between the jet and the reconstructed Z boson $\Delta\phi$(jet, Z boson). The $\Delta\phi_{jZ}$ distribution for different $|\eta|$ bins can be seen in Fig. 6.10. Simulation is split into reconstruction-level jets matched to particle-level jets within $\Delta R < 0.4$ (quark and gluon jets based on the jet flavor) and not matched jets (PU or unassigned if the jet contains a generated particle from the hard scattering). The jet flavor is determined by associating generated particles to reconstruction-level jets.

While PU jets are uniformly spread in ϕ, quark and gluon jets are more likely in $\Delta\phi_{jZ} > 2.5$. Therefore, the PU jet enriched region is defined as $\Delta\phi_{jZ} < 1.5$ and the LV jet enriched region is given by $\Delta\phi_{jZ} > 2.5$. The amount of PU jets is at least two times higher in the PU jet enriched region as in the LV jet enriched region. The PU jet rate is defined as the rate of events in the PU jet enriched region divided by the events in the LV enriched region.

The PU jet rate for two η-regions is depicted in Fig. 6.11. The data-to-simulation disagreement for the three techniques shown is within the systematic uncertainties of using a different generator when showering. CHS shows a strong dependency of the PU jet rate in both regions on the number of vertices. Applying the PU jet ID reduces the dependency significantly. While PUPPI has a stable PU jet rate in $|\eta| < 2.5$, it shows a dependency on the number of vertices in $|\eta| > 2.5$.

6.2.3 Noise Reduction with PUPPI

PUPPI is not only successful in mitigating the effects of PU, but also reduces the effects of noise originating from malfunctioning detector parts. In 2017 part of the ECAL in $2.65 < |\eta| < 3.139$ was, due to radiation damage, miscalibration, and too low thresholds, affected by noise. Fig. 6.12 (upper left) shows the effect in 2017 data for PUPPI and CHS. An increase of jets in $2.5 < |\eta| < 3$ is observed which is not modelled in simulation. PUPPI already reduces the effect by a factor of two.

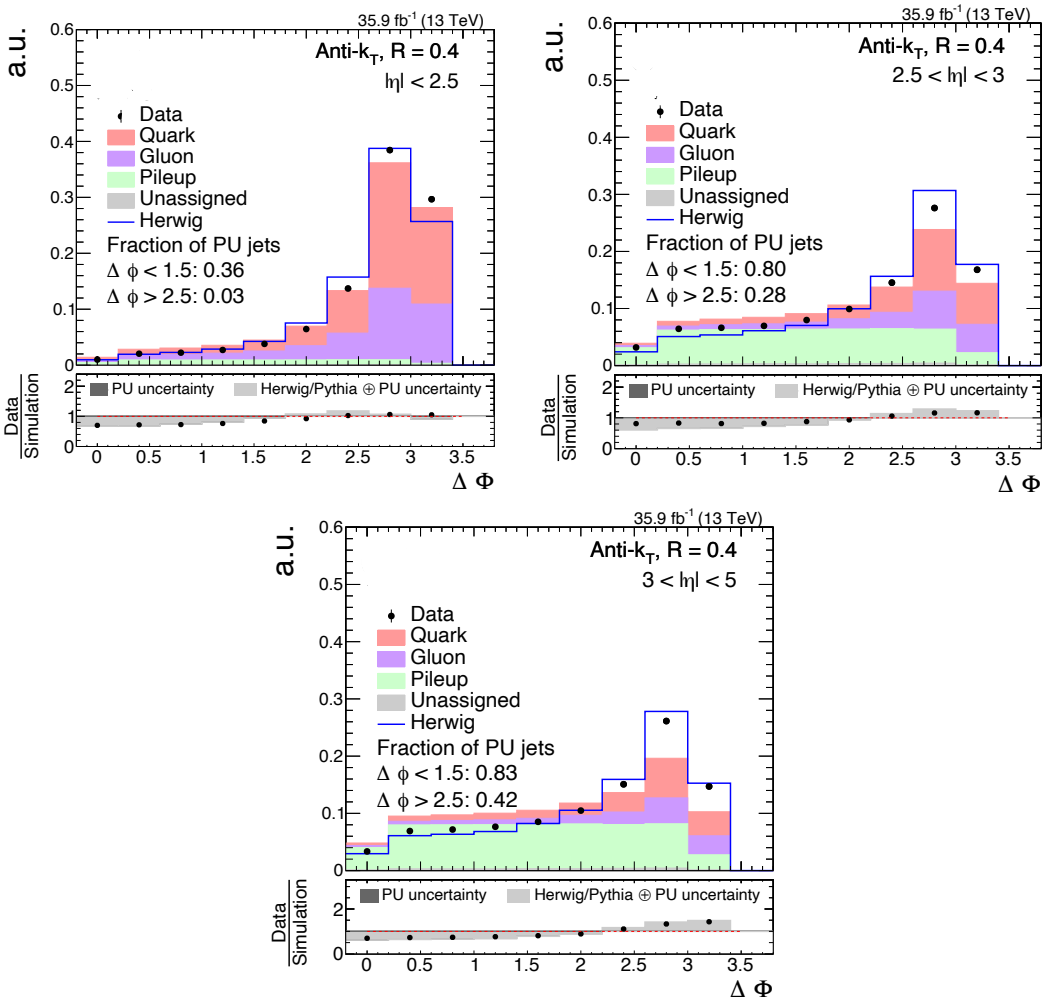


Figure 6.10: Data-to-simulation comparison of the $\Delta\phi$ distribution. Simulation is split into jets originating from quarks, gluons, pileup and, if the origin is not clear, unassigned. A gluon jet (solid purple) is defined as a jet having ΔR(reconstruction-level jet, particle-level jet) < 0.4 and the pdgID of the closest generator particle of 21. A quark jet (solid red) is defined as a jet having ΔR(reconstruction-level jet, particle-level jet) < 0.4 and the pdgID of the closest generator particle smaller than 7. A pileup jet (solid green) is defined as a jet having ΔR(reconstruction-level jet, particle-level jet) > 0.4 and and no generator particle within the jet. An unassigned jet (solid gray) is defined as a jet having ΔR(reconstruction-level jet, particle-level jet) > 0.4 and a generator particle is within the jet. Herwig simulation is shown in a solid blue line. The ratio plot shows data to MC simulation comparison between Pythia as points and the shaded gray band for Herwig.

In order to reduce the noise, the slope of the weighted p_T requirement listed in Tab. 6.1 was increased from 0.13 to 0.2 in the corresponding η-region, resulting in

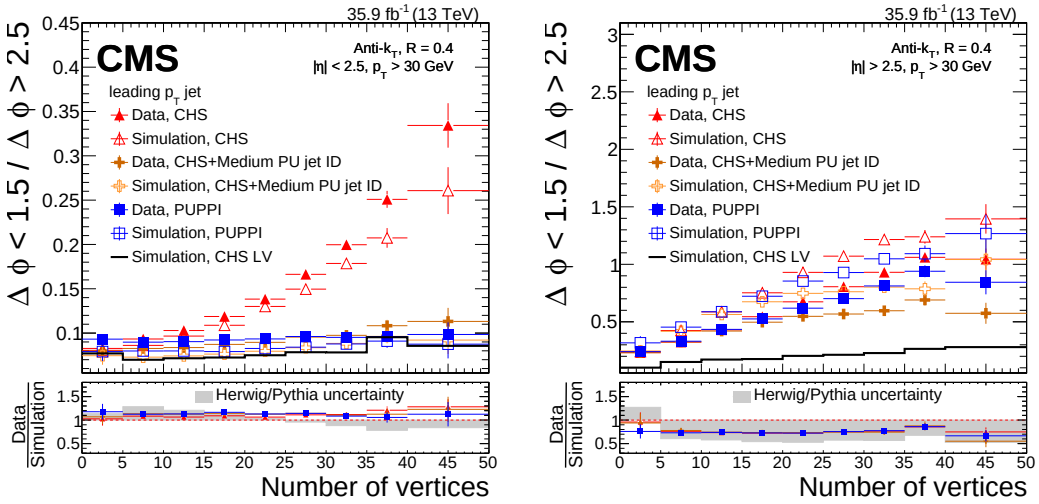


Figure 6.11: Rate of jets in the PU-enriched region divided by the rate of jets in the LV-enriched region as a function of the number of vertices for CHS jets (red triangles), CHS jets with medium PU jet ID applied (orange crosses) and PUPPI jets (blue squares) in Z+jets simulation (open markers), and data (full markers). The plots show the ratio for events with $|\eta| < 2.5$ (left) and $|\eta| > 2.5$ (right). The lower panels show the data-to-simulation ratio along with a gray band corresponding to the one-sided uncertainty that is the difference between simulated Z+jets events showered with the PYTHIA parton shower to those showered with the HERWIG++ parton shower. Published in Ref. [4].

a significant reduction of the jets originating from noise. The new tune was tested on a subset of the 2017 data, corresponding to an integrated luminosity of $1.5\,\text{fb}^{-1}$. The number of vertices of this subset are compared with the whole 2017 data and is shown in Fig. 6.12 (upper right). The subset has a sufficient mixture between high and low PU scenarios and is representative for the PU profile in 2017. Figure 6.12 (lower) shows the effect of the new tune compared to the original version of PUPPI on this subset. Increasing the requirements reduces the LV jet efficiency but at the same time increases the purity. This effect is of the order of a few percent in simulation as shown in Fig. 6.13. In data, where the noise jets are present, the loss in efficiency is outweighted by the large gain in purity in $2.5 < |\eta| < 3$.

6.2.4 Effects on the Number of Vertices

The relationship between the number of vertices and the number of interactions is affected by the jet reconstruction and requirements on the properties of a jet. Without a selection on jets, the number of vertices is on average 30% smaller [118] than the mean

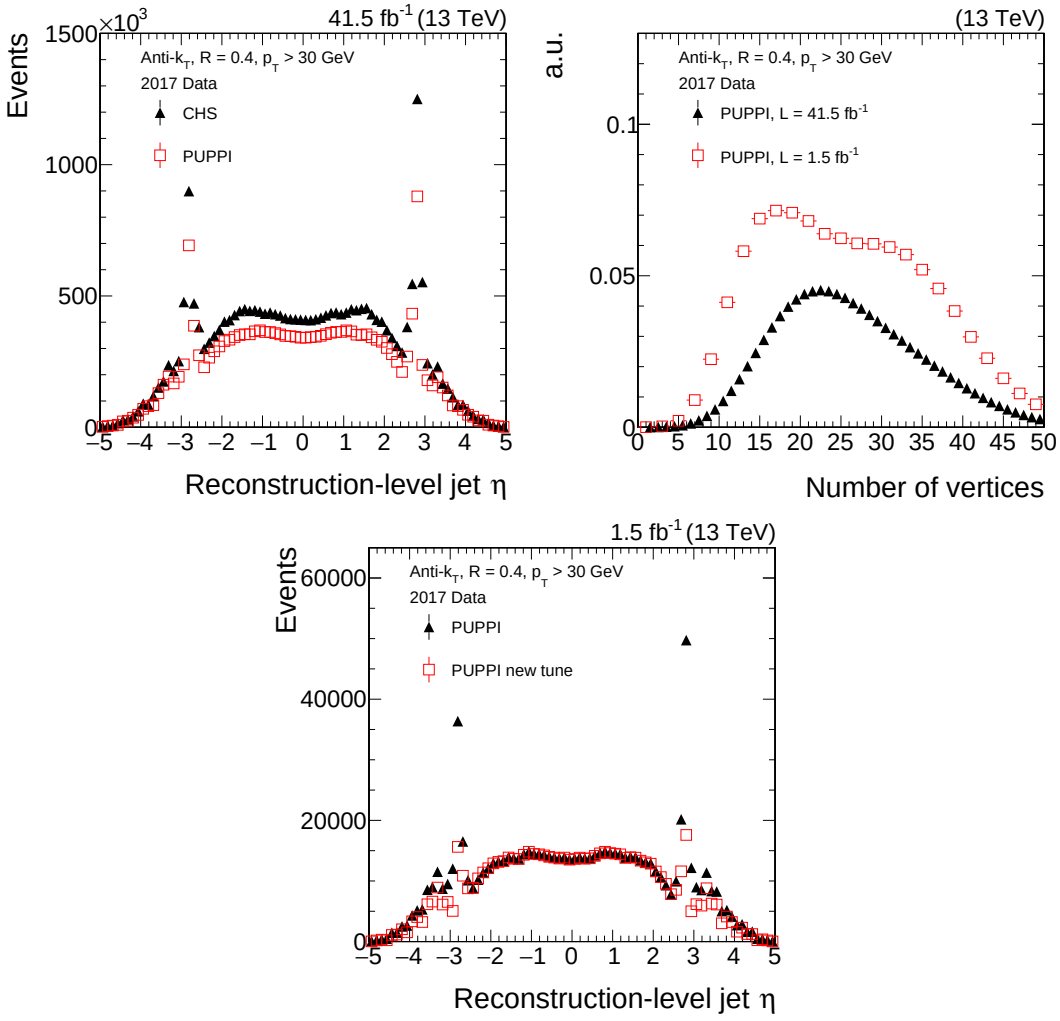


Figure 6.12: η-distribution of 2017 data with the ECAL noise problem between $2.5 < |\eta| < 3$.

number of interactions. This is because the vertex reconstruction and identification efficiency is around 70% for all vertices, while it is nearly 100% for the LV. With a requirement on the jet p_T, the number of vertices is even smaller. This effect can be seen in Fig. 6.14 (left). Event selections on CHS jets show the largest effect, since no treatment on PU jets is done. When applying a PU jet rejection technique, like PUPPI or PU jet ID, the bias gets reduced. This bias results from PU vertices close to or overlapping with the LV. The mean number of interactions as a function of the number of vertices is shown in Fig. 6.14 (right) and depends on the underlying PU profile. Here, the number of interactions is adjusted to match the PU profile during 2016 data taking, also shown in Fig. 6.1. For high numbers of vertices a difference between events with and without a p_T requirement is observed. The bias due to

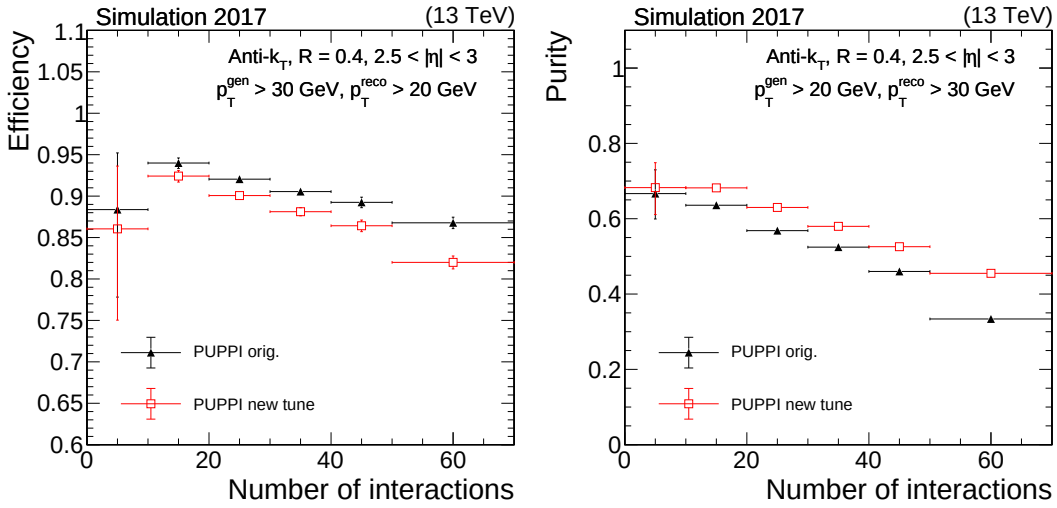


Figure 6.13: The LV jet efficiency (left) and purity (right) in Z+jets simulation as a function of the number of interactions for the default PUPPI (red closed triangles) and PUPPI with noise reduction (purple open squares). Plots are shown for AK4 jets $p_T > 20\,GeV$, and (left) $|\eta| < 2.5$ and (right) $|\eta| > 3$. The LV jet efficiency is defined as the number of matched reconstruction-level jets with $p_T > 20\,GeV$ divided by the number of particle-level jets with $p_T > 30\,GeV$ that originate from the main interaction. For the lower plots, the purity is defined as the number of matched particle-level jets with $p_T > 20\,GeV$ divided by the number of reconstructed jets that have $p_T > 30\,GeV$. The error bars correspond to the statistical uncertainty in the simulation.

different PU mitigation techniques is negligible.

6.2.5 Heavy Object Tagging

When searching or measuring in the boosted regime, as described in Sec. 7, the mass and other properties of the large-R jets are crucial observables. Large-R jets tend to collect more PU, hence PU mitigation techniques are of particular interest. Figure 6.15 shows the effect of PU on the soft drop jet mass scale, soft drop jet mass resolution and the N-subjettiness variable. The soft drop jet mass scale and resolution is calculated in simulation for large-R jets with $400 < p_T < 600\,GeV$. The jets originate from a W boson from the decay of a heavy bulk graviton. The PUPPI algorithm removes the PU dependence from the soft drop jet mass, whereas CHS shows a PU dependence. Similar to the jet energy resolution, the soft drop jet mass resolution is defined as the spread of the ratio of reconstruction- and particle-level jet mass (the response) divided by the mean of the response. CHS shows a dependency on the resolution as function

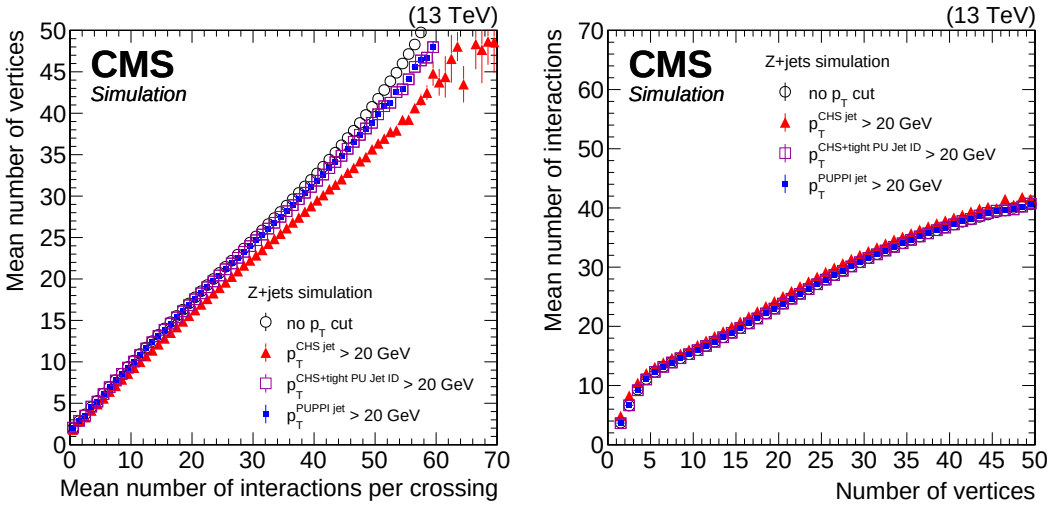


Figure 6.14: Left: distribution of mean number of reconstructed vertices as a function of the mean number of interactions in Z+jets simulation. Right: distribution of the mean number of interactions as a function of the number of vertices in Z+jets simulation. The black open circles show the behavior without applying any event selection, whereas for the other markers a selection on jets of $p_T > 20\,\text{GeV}$ is applied using the CHS (full red triangles), CHS+tight PU jet ID (violet open squares), and PUPPI (full blue squares) algorithms. The error bars correspond to the statistical uncertainty in the simulation. Published in Ref. [4].

of the number of vertices, whereas PUPPI shows a stable performance.

Fig. 6.15 (lower) shows the median of the τ_{21}-variable needed to identify boosted W bosons. It is shown for W jets and for jets originating from quarks and gluons (mistags). PUPPI shows a stable value over the different PU scenarios, whereas CHS shows a PU dependency. However, the shift for CHS is visible for both, W jets and q/g jets, hence the discrimination power of this variable is stable also for CHS.

In order to identify a W boson both variables – the soft drop jet mass and N-subjettniss – are combined. The performance of this combination in terms of efficiency and misidentification rate is shown in Fig. 6.16. PUPPI shows a stable efficiency and misidentification rate whereas CHS shows a decrease in both as a function of the number of vertices.

The ATLAS Collaboration tested a variety of possible inputs for jets and jet mass to identify boosted objects [149]. They tested different jet clustering methods, such as topocluster, PF or track-caloclusters, and different methods to reduce the effect of soft and wide-angular radiation, such as soft drop, recursive soft drop, bottom-up soft drop, pruning and trimming. The performance of each combination is evaluated

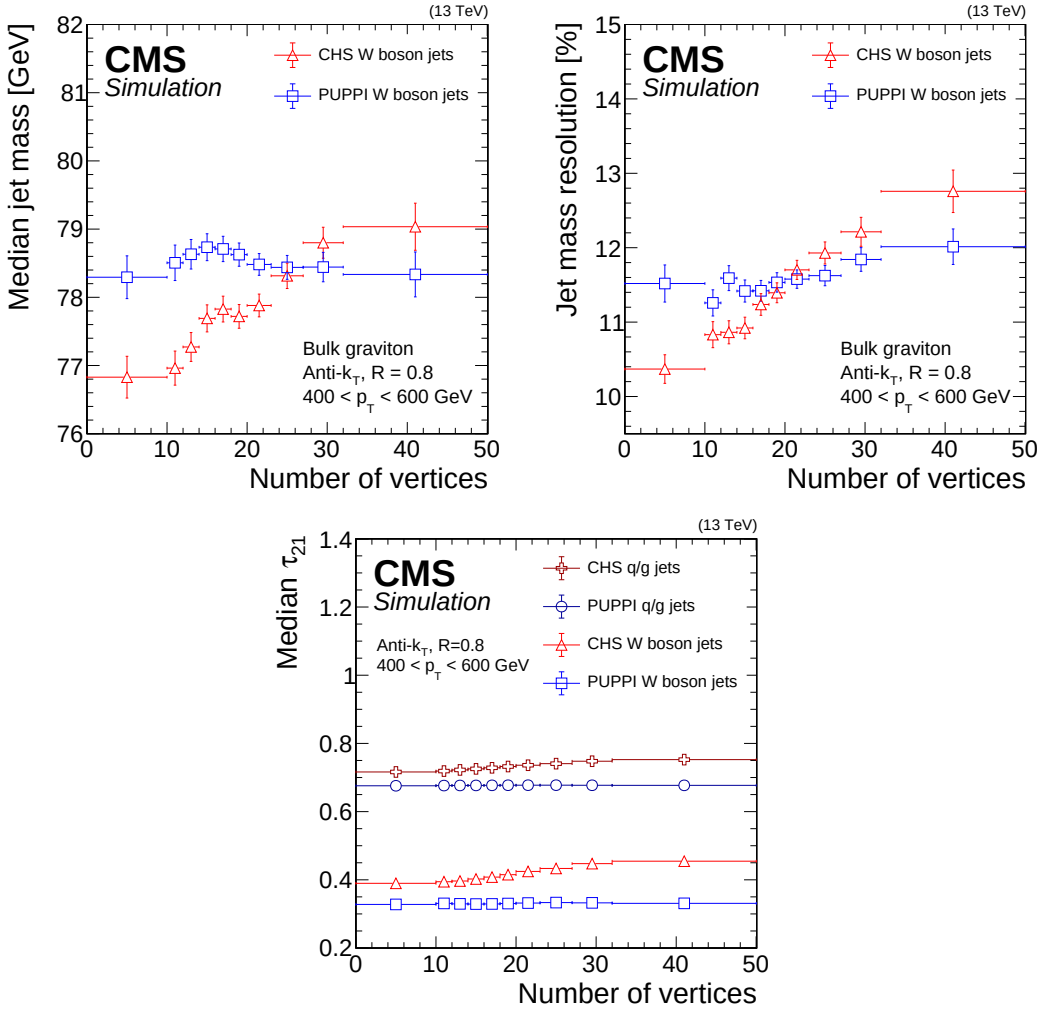


Figure 6.15: Median soft drop jet mass (upper left), soft drop jet mass resolution (upper right) and median τ_{21} (lower) for AK8 jets from boosted W bosons with $400 < p_T < 600\,\text{GeV}$ for CHS (red triangles) and PUPPI (blue squares) jets, as a function of the number of vertices. The error bars correspond to the statistical uncertainty in the simulation. Published in Ref. [4].

for different variables based on the slope of this variable as a function of the number of vertices. The ATLAS Collaboration also test a combination of PUPPI and PF as input. However, the used PF and PUPPI algorithms are not optimised and therefore not directly comparable with the version presented in this thesis. For example, with the PUPPI and PF version from the ATLAS Collaboration a slope of 11% per vertex for the soft drop jet mass of large-R W jets with a $300 < p_T < 500\,\text{GeV}$ is observed, whereas the CMS Collaboration sees a stable behavior for $400 < p_T < 600\,\text{GeV}$ as shown in Fig. 6.15. The ATLAS Collaboration also observers a 3% per vertex effect in the jet mass resolution. However, their standard method, topocluster with SoftKiller and constituent subtraction plus the use of trimming with $f_{\text{cut}} = 5\%$, results in a slope

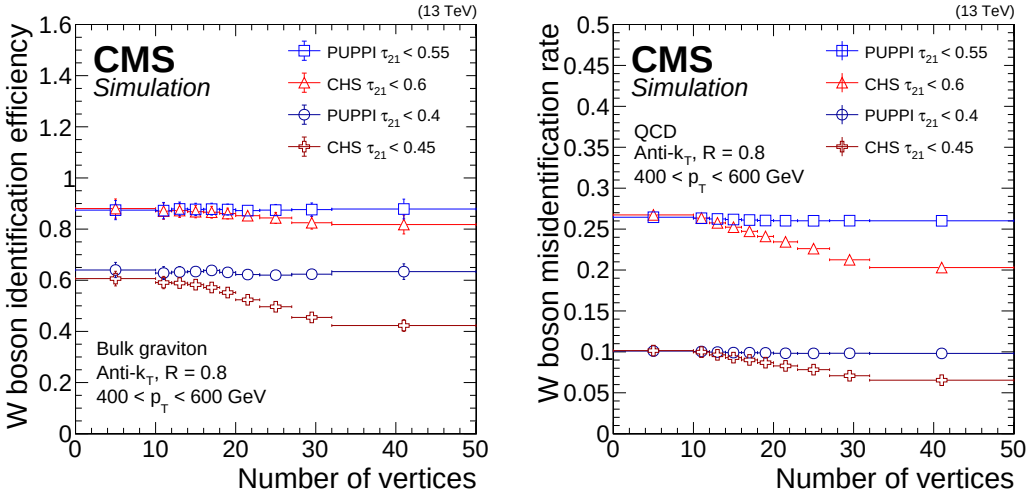


Figure 6.16: W boson identification performance using a selection on τ_{21} for CHS (red triangles and dark red crosses) and PUPPI (blue squares and circles) AK8 jets as a function of the number of vertices for loose and tight selections, respectively. Shown on the left is the W boson identification efficiency evaluated in simulation for a bulk graviton decaying to a WW boson pair and on the right the misidentification rate evaluated with QCD multijet simulation. The error bars correspond to the statistical uncertainty in the simulation. Published in Ref. [4].

of a 3% and 2% for the jet mass scale and resolution, respectively. In summary, PUPPI, especially when tuned to the corresponding detector configuration, shows a stable performance for variables needed to identify boosted heavy objects.

6.3 Summary and Outlook

Overall PUPPI shows a stable performance for PU mitigation on both, small-R jets and large-R jets, especially at high PU scenarios. The tagging performance for large-R jets is improved with PUPPI and stable against PU, which is particular important for measurements and searches in the boosted regime. Furthermore, PUPPI shows a compatible performance in terms of purity and efficiency in $|\eta| > 2.5$ compared with CHS + loose PU Jet ID and PUPPI is intrinsically able to remove detector noise. Both are needed when performing searches or measurements using jets with $|\eta| > 2.5$, as for example needed in the search for diboson resonances discussed in the following chapter. Another advantage of PUPPI is the consistent event representation unlike CHS +PU jet ID or the forward jet vertex tagging from the ATLAS Collaboration, where the effect of removing PU jets is not propagated to the missing transverse momentum or other global variables.

The decreased performance of PUPPI in the JER at high p_T could be solved with a p_T protection for charged and neutral particles. PUPPI with this new tune will be used as default algorithm by the CMS Collaboration for small-R and large-R jets during the operation in 2021-2023 and is crucial during the operation of the HL-LHC.

Further improvements are possible by studying

- the track-vertex association which is done during the PF reconstruction,
- the default value of α_i for particles without neighbouring particles in their isolation cone,
- the functional form used for the protection of LV neutral particles,
- to add information on the shower depth in the HCAL detector,
- the calculation of the signed χ^2, where a likelihood could yield in a better description of the weight,
- the use of missing transverse momentum for each PU vertex, and
- a multi-variant analysis instead of a simple single variable α_i.

Several simulation studies are done in preparation for the operation during the HL-LHC, showing that PUPPI is also working in harsher PU conditions [150] than studied here. It should be noted that for the HL-LHC detector configuration a timing detector has been added, ensuring a better vertex reconstruction, necessary for events with $< \mu > \ 140$.

7 | Search for a Heavy Spin-1 Resonance Z' decaying into a Vector-Like Quark T and a top Quark

Theories of compositness or extra dimension solve the hierarchy problem by extending the SM with a new heavy boson Z' and VLQs. Searches for a new heavy spin-1 resonance Z' decaying into a pair of SM particles have been performed in various decay channels at the LHC [62–82, 88–96]. The decay into $t\bar{t}$, for example, sets a stringent mass limit of 3.8 TeV for a leptophobic Z' with a 1% decay width [62, 64, 65], assuming $\mathcal{B}(Z' \to t\bar{t}) = 1$. However, if the mass of the new resonance is large enough ($M_{Z'} \sim$ TeV), the decay of $Z' \to Tt$, depicted in Fig. 7.1, becomes kinematically possible.

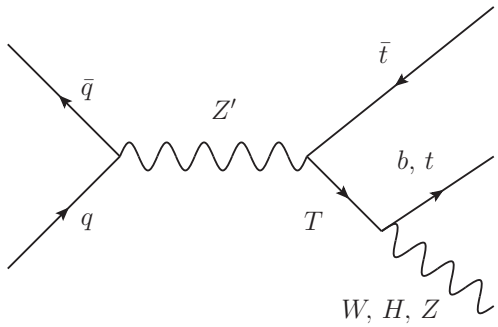


Figure 7.1: Leading order Feynman diagram for the production of a spin-1 resonance Z' and its decay, along with the possible decays of the vector-like quark T. Published in Ref. [1].

A search for $Z' \to Tt$ is possible in three different decay channels, tHt, tZt and tWb, according to the decay modes of the vector-like quark T. Two searches for a heavy resonance $Z' \to Tt$, focusing on the decay modes $T \to Wb$ [87] and $T \to Zt$ [21], already exist. The latter is a search for singly produced T, where the results are reinterpreted in the context of $Z' \to Tt$. When this work has been performed no dedicated search in the decay mode $T \to Ht$ existed. The search presented in this thesis and published in Ref. [1] is optimized in the $T \to Ht$ and $T \to Zt$ decay modes and is performed

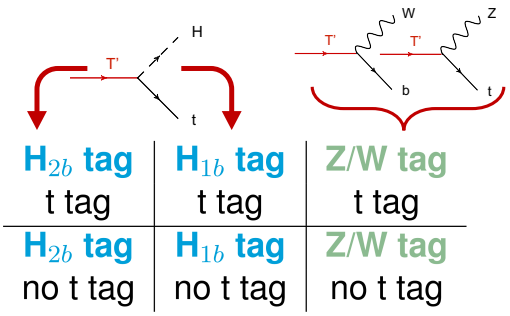


Figure 7.2: Sketch of the event categorisation. Each decay of the T is reconstructed in a separate category by the tag of the heavy boson. In addition, the events are split into the boosted regime, where a t tag is present and the resolved regime, where no t tag is present.

in the lepton+jets (ℓ+jets) channel, where l denotes an electron or a muon. The analysis is performed with pp collision data, corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$, recorded by the CMS experiment [98] at the LHC [97] at a center of mass energy of 13 TeV. This analysis uses jet substructure techniques to identify the boosted heavy bosons of the T decay as also proposed in Ref. [151].

7.1 Signature and Strategy of the Search

Although optimized in $\mathrm{T} \to \mathrm{Ht}$ and $\mathrm{T} \to \mathrm{Zt}$, this search covers all three final states of the T, resulting in a signature with two t quarks, one decaying leptonically ($\mathrm{t} \to \mathrm{Wb} \to l\nu\mathrm{b}$) and one decaying hadronically ($\mathrm{t} \to \mathrm{Wb} \to q\mathrm{q}\mathrm{b}$), and a heavy boson. Each of these decay modes is reconstructed in a separate category as shown in Fig. 7.2.

The mass range ($M_{\mathrm{Z}'}$, M_{T}) probed with this analysis, is limited by $M_{\mathrm{Z}'} > M_{\mathrm{t}} + M_{\mathrm{T}}$, when the decay becomes kinematically possible and $M_{\mathrm{Z}'} < 2M_{\mathrm{T}}$. Above this value, the $\mathcal{B}$ ($\mathrm{Z}' \to \mathrm{tT}$) reduces significantly because of the possibility of a $\mathrm{Z}' \to \mathrm{TT}$ decay. In addition, searches for VLQ T produced by gluon-gluon fusion or electroweak production place a lower bound on the mass of the T of around 1.3 TeV [11–25]. However, all searches presented in this chapter consider a different production of the T including a new boson, $\mathrm{Z}' \to \mathrm{tT}$, instead of electro-weak production (e.g. $q\mathrm{g} \to q'\mathrm{W}\mathrm{b}\overline{\mathrm{b}} \to \mathrm{T}q'\overline{\mathrm{b}}$) or the strong interaction ($\mathrm{gg} \to \mathrm{g} \to \mathrm{TT}$), lowering the constraints on the T mass. In this search, the mass of the T is at least 700 GeV.

The mass range probed with this search is presented in Fig. 7.3. The hashed areas show configurations which are excluded by the $M_{\mathrm{Z}'}$ mass requirements mentioned above. The white areas indicate mass configurations, where no signal sample was

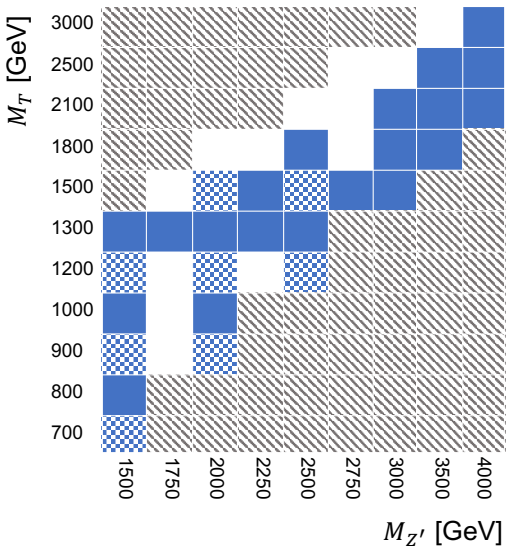


Figure 7.3: Possible configurations of $M_{Z'}$ and M_T. The checker board squares show the mass configurations probed by the two existing searches [21, 87]. Together with the filled squares it shows all configurations probed in the search presented in this thesis [1]. The hashed areas in the upper left indicates the region where the $Z' \to tT$ decay is kinematically forbidden, while in the lower right $Z' \to tT$ is suppressed by the preferred $Z' \to TT$ mode.

generated. The solid blue and checker board areas are the probed mass configurations. Compared to the two existing searches [21, 87], which only probe the checker board areas, this search extends the mass range probed.

The signature of this decay depends on the mass of the T and the mass difference between the Z' and the T, where the differences range from 200 GeV to 1900 GeV. At high masses of the T ($M_T \gtrsim 1000\,\text{GeV}$) the heavy boson receives a large Lorentz-boost resulting in a reduced angular separation ΔR of its decay products. The heavy boson is therefore identified with a large-R jet and substructure techniques. A large mass difference (e.g. $M_{Z'} - M_T \gtrsim 800$ GeV) causes a large Lorentz-boost of t quark that decays hadronically. While for a t quark that decays hadronically with low p_T ($p_T \lesssim 400$ GeV), each decay product can be reconstructed in a small-R jet with a distance parameter $R = 0.4$ (resolved regime), with $p_T \gtrsim 400$ GeV the angular separation of the decay products is reduced, such that a reconstruction with three small-R jets is not possible. Instead, a large-R jet with $R = 0.8$ can be used (boosted regime). Both scenarios are covered by a separate category in this search as indicated in Fig. 7.2.

The leptonic t quark is reconstructed from a small-R jet, p_T^{miss} and a lepton, where the lepton is not isolated because of the proximity of the b jet and the large Lorentz-boost. All three objects, the two t quark and the heavy boson, are used to reconstruct

$M_{Z'}$ which is the sensitive variable used to set upper cross section limits. The main backgrounds of this search are $t\bar{t}$ and W+jets processes, which can be predicted well by simulation. Two control region are used to constrain the cross section rates of the $t\bar{t}$ and W+jets processes and to validate the data-to-simulation agreement.

7.2 Data and Simulated Events[4]

The analysis is based on the data set of pp collisions recorded by the CMS detector during the year 2016. Events targeting the decay of a top quark to a final state including a muon are selected with a high-level single-muon trigger that requires the presence of at least one muon candidate with $p_T > 50\,GeV$ and $|\eta| < 2.4$. For events targeting a final state with an electron, the high-level trigger requires the presence of at least one electron candidate with $p_T > 115\,GeV$ and $|\eta| < 2.5$, or at least one photon with $p_T > 175\,GeV$ and $|\eta| < 2.5$. The latter requirement ensures events containing electrons with a high p_T are efficiently selected, as the requirements on ECAL shower shapes are less stringent for photons than for electrons. Given the highly boosted topology of the final-state objects, no isolation requirements are applied to the lepton candidates at the trigger level. The electron trigger threshold is significantly larger than the muon trigger threshold, since the non-isolated electron trigger selects a large number of hadrons incorrectly identified as electrons. Both recorded data sets correspond to an integrated luminosity of $35.9\,fb^{-1}$ [152] .

The spin-1 resonance signal samples are generated with the leading-order (LO) mode of MadGraph5_aMC@NLO 2.2.2 [44] as a high mass resonance with SM-like couplings using the G^* model [5]. The pythia 8.1 8.212 [46] event generator with the CUETP8M1 underlying event tune [49, 50] is used to model the parton showering and underlying event. Separate samples for the different decay channels of the T are produced, so that each sample has a branching fraction of 100% to the chosen decay channel. Throughout this paper, a generic spin-1 heavy resonance will be referred to as Z', whilst interpretations within a given model will refer to their specific resonance names. Two benchmark models are used to interpret the results: a heavy gluon G^* model described in Ref. [5] and a ρ^0 model described in Ref. [6].

In the G^* model [5], ten new VLQs (T, B, $\tilde{T}$, $\tilde{B}$, $T_{5/3}$, $T_{2/3}$, T', B', $B_{-1/3}$, $B_{-4/3}$) are predicted with well-defined relationships between their masses. In this analysis, the T mass is varied, whilst other masses are related by $M_{T_{5/3}} = M_{T_{2/3}} = M_T \cos\phi_L$. The mixing angle $\cos\phi_L$ governs the degree of compositness of the left-handed quark doublet

[4]This section is taken from Refs. [1] and was adjusted to fit the needs of this paragraph.

(t_L, b_L), and hence the relative coupling of the lightest spin-1 Kaluza–Klein excitation of the gluon, G^*, to third-generation quarks compared to the other two generations of quarks. A benchmark scenario with parameters $\tan\theta = 0.44$, $\sin\phi_{tR} = 0.6$, and $Y_* = 3$ is used in this analysis, leading to $\cos\phi_L = 0.84$. A description of the benchmark and its parameters can be found in Ref. [5]. In this model the branching fractions ($\mathcal{B}$) of the T decay to Wb, Ht, and Zt are chosen to be 0.5, 0.25, and 0.25, respectively.

The ρ^0 model [6] predicts a heavy spin-1 resonance, ρ, along with a multiplet of four new vector-like quarks, with two of the vector-like quarks (T, B) representing the heavy partners of the top and b quarks, respectively. Other exotic vector-like quarks are also predicted: $X_{2/3}$ with a charge of $2e/3$, and $X_{5/3}$ with a charge of $5e/3$, where e is the magnitude of the charge of the electron. A benchmark scenario with parameters $y_L = c_2 = c_3 = 1$ and $g_{\rho_L} = 3$ is used in this analysis, where a description of the benchmark and its parameters can be found in Ref. [**?**]. In this model the branching fractions of the T decay to Wb, Ht, and Zt are chosen to be 0, 0.5, and 0.5, respectively.

Two values of the Z' width are considered, corresponding to 1% or 30% of its mass. The T width is set to 1% of its mass. For the Z' and T mass parameter space considered in this analysis the total Z' decay width in the two considered theoretical models is always less than 20% of its mass. Since the experimental resolution is approximately 15%, the samples with the Z' width set to 1% are dominated by the experimental resolution, and are thus used in the interpretation of the results. The samples generated with the width of 30% are used as cross-checks and help to confirm that the conclusions do not change for scenarios with Z' widths somewhat larger than the experimental resolution for high masses of the Z'. Furthermore, it was checked that scenarios with T widths of up to 30%, with a Z' width equal to or larger than that of the T, do not significantly affect the resolution of the Z' mass, and therefore the experimental limits obtained with the T width set to 1% are also valid for larger T width scenarios.

The G^* model considers only left-handed T quarks. The ρ^0 model also allows for a right-handed ρ_R coupling to T quarks. For the $T \to Ht$ decay mode the kinematic distributions in the G^* model and ρ^0 model are the same. While for the $T \to Zt$ and $T \to Wb$ decay modes the Z/W boson p_T spectra are similar for the left-handed ρ_L and the G^*, the ratio of the distributions for left- and right-handed scenarios in the ρ^0 model deviates from unity by up to 30%. In this analysis only the decays of the left-handed ρ_L are considered.

Simulated event samples for the SM background processes Drell–Yan (DY)+jets, also referred to as Z+jets, and W+jets are computed at next-to-leading-order (NLO) precision in QCD with MadGraph5_aMC@NLO. The parton showering is calculated

using PYTHIA 8.1 8 following the FxFx merging scheme [47]. Background events from QCD multijet processes are simulated using PYTHIA 8.1 8. For the simulation of the underlying event, the tune CUETP8M1 is used in PYTHIA 8.1 8 for the W+jets, Z+jets, and QCD multijets samples.

The simulation of SM $t\bar{t}$ and single top quark (ST) background events is performed with the POWHEG event generator [41–43, 153–158], using POWHEG v1.0 for the simulation of tW events, whilst POWHEG v2.0 was used for the simulation of $t\bar{t}$ and all other single top quark processes. The PYTHIA 8.1 8 generator was used for the showering in both versions of POWHEG. An observed discrepancy between simulation and data in the top quark p_T spectrum is corrected with a reweighting procedure based on measurements of the top quark p_T spectrum [159, 160]. The underlying event tune CUETP8M2T4 [161] is used in PYTHIA 8.1 8 for the $t\bar{t}$ and single top quark samples.

All events are generated with the NNPDF 3.0 parton distribution functions (PDFs) [38]. The detector response is simulated with the GEANT4 package [52]. Simulated events are processed through the same software chain as used for collision data. All simulated event samples include the simulation of pileup, and are reweighted to match the observed distribution of the number of pileup interactions in data.

7.3 Event Selection

The search for a heavy spin-1 resonance Z′ with the decay $Z' \to tT$ and $T \to Ht/Zt$, results in two t quarks and a H/Z boson. The event selection is optimized for cases in which one of the t quark decays leptonically, while the other decays hadronically. Requiring exactly one lepton in the final state reduces the amount of QCD multijet background compared to no leptons in the final state. The QCD multijet background is complicated to estimate, whereas this search is dominated by $t\bar{t}$ and W+jets backgrounds, which are well known and can be estimated from simulations. Compared to a final state with multiple leptons, this final states offers a higher $\mathcal{B}$, hence higher signal efficiency.

The heavy bosons (H, Z and W) from the T decay as well as the boosted hadronically decaying t quark are reconstructed as large-R jets with $p_T > 250$ GeV and $|\eta| < 2.5$. In order to reconstruct the t quark, that decays leptonically, a non-isolated lepton, $p_T^{miss}$and a small-R jet with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$ is required. The lepton, either a muon or electron, has to have $p_T > 55\,\text{GeV}$ and $|\eta| < 2.4$ in case of the muon and $p_T > 125$ and $|\eta| < 2.5$ in case of the electron. In order to reduce the amount of QCD multijet background, the lepton is typically considered to be isolated, because leptons originating from QCD multijet processes (e.g. B mesons decays) are surrounded by

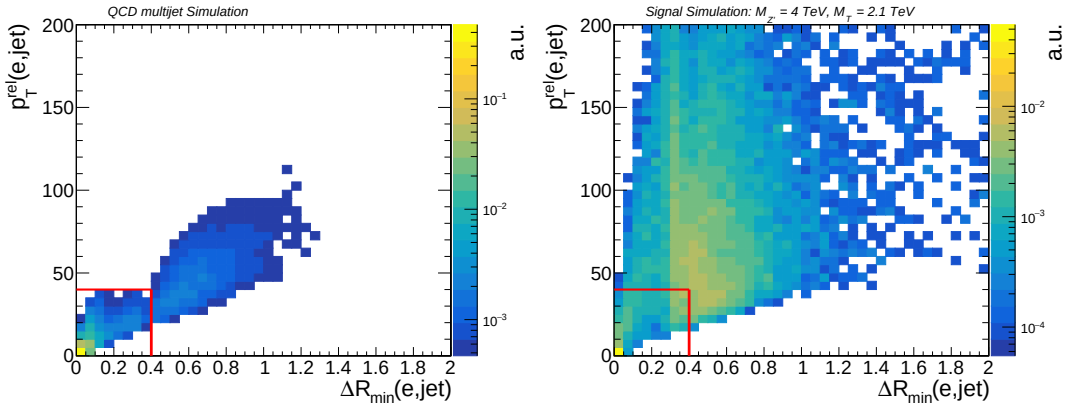


Figure 7.4: The isolation variables used in the 2D isolation for QCD multijet background (left) and a signal sample with $M_{Z'} = 4\,\text{TeV}$ and $M_T = 2.1\,\text{TeV}$ (right). The distributions are normalized to unity. The red lines indicate the region removed by the 2D isolation. In QCD multijet background this area is 76.1% of the total area and in the signal simulation it is 14.9% of the total area.

many other particles. An intuitive choice of a isolation is to sum up the p_T of particles surrounding the lepton within a cone in ΔR around it. Since for high p_T t quarks the lepton might overlap with the small-R jet originating from the b quark of the t quark decay, a standard cone-based isolation for the lepton is not appropriate to distinguish it from QCD multijet background. A two-dimensional isolation based on the relative momentum of the lepton and the closest jet, as well as the angular separation between the lepton and the closest jet is used. The isolation variables are shown for the QCD multijet background and a signal sample with $M_{Z'} = 4\,\text{TeV}$ and $M_T = 2.1\,\text{TeV}$ in the e channel in Fig. 7.4. For the 2D isolation used here it is required that either $p_T^{\text{rel}}(l,\text{jet}) > 40\,\text{GeV}$ or $\Delta R_{\text{min}} > 0.4$. The area excluded with this requirement is indicated by the red lines in Fig. 7.4. This requirement removes 76.1% of the QCD multijet background, while it only rejects 14.9% of the signal simulation events with $M_{Z'} = 4\,\text{TeV}$ and $M_T = 2.1\,\text{TeV}$.

In summary the event selection requires:

- exactly one lepton with $p_T > 50(125)\,\text{GeV}$ and $|\eta| < 2.4(2.5)$ for muon (electrons) with a 2D isolation of $p_T^{\text{rel}}(l,\text{jet}) > 40\,\text{GeV}$ or $\Delta R_{\text{min}} > 0.4$,
- at least two small-R jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$,
- at least one large-R jet with $p_T > 250\,\text{GeV}$ and $|\eta| < 2.5$ and
- (in the e channel only) $p_T^{\text{miss}} > 90\,\text{GeV}$.

The distributions of the main variables used in this events selection are shown in Fig. 7.5 in data, simulated SM background and three different signal samples with various $M_{Z'}$ masses. A good data-to-simulation agreement is seen in all distributions, except for the p_T of the electron (upper right). However only statistical uncertainties are shown and the disagreement seen is covered by systematic uncertainties described in Sec. 7.7.

7.3.1 Trigger Efficiency Measurement

This analysis uses single lepton trigger paths to select the events as described in Sec. 7.2. Events targeting a final state including a muon are selected with a A high-level single-muon trigger that requires the presence of at least one muon candidate with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.4$. For events targeting a final state with an electron, the high-level trigger requires the presence of at least one electron candidate with $p_T > 115\,\text{GeV}$ and $|\eta| < 2.5$, or at least one photon with $p_T > 175\,\text{GeV}$ and $|\eta| < 2.5$. The triggers are emulated in simulation but the efficiency is not the same between data and simulation. Therefore, correction factors are needed to adjust the efficiency in simulation to the one in data. Correction factors for the muon trigger combination are derived by the CMS Collaboration in Ref. [117] and are used in this analysis. No dedicated correction factors for the electron trigger combination are available. In order to measure these a control region orthogonal to the signal region is used. The control region is enriched in $t\bar{t}$ processes and the trigger efficiency is measured with a tag-and-probe method. The control region is defined by one muon and one electron, while the signal region requires exactly one lepton.

Event Selection

The trigger efficiency is measured with respect to the muon object (p_T, η and trigger requirements) selected, which ensures no bias in the event selection of the measurement from electron quantities. The tag-and-probe method requires a well-defined *tag* which is in this measurement a muon with $p_T > 55\,\text{GeV}$ and $|\eta| < 2.4$, which fired the trigger combination used in the signal region. In addition, the 2D isolation is applied to the muon as it is also done in the signal region. The *probe* is an electron with $p_T > 40\,\text{GeV}$ and $|\eta| < 2.5$. At least two small-R jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ are required to ensure a control region enriched in $t\bar{t}$ and single top production events. The electron p_T and η with these selections are shown in Fig. 7.6. As shown, $t\bar{t}$ and single top processes are the main backgrounds in this phase space.

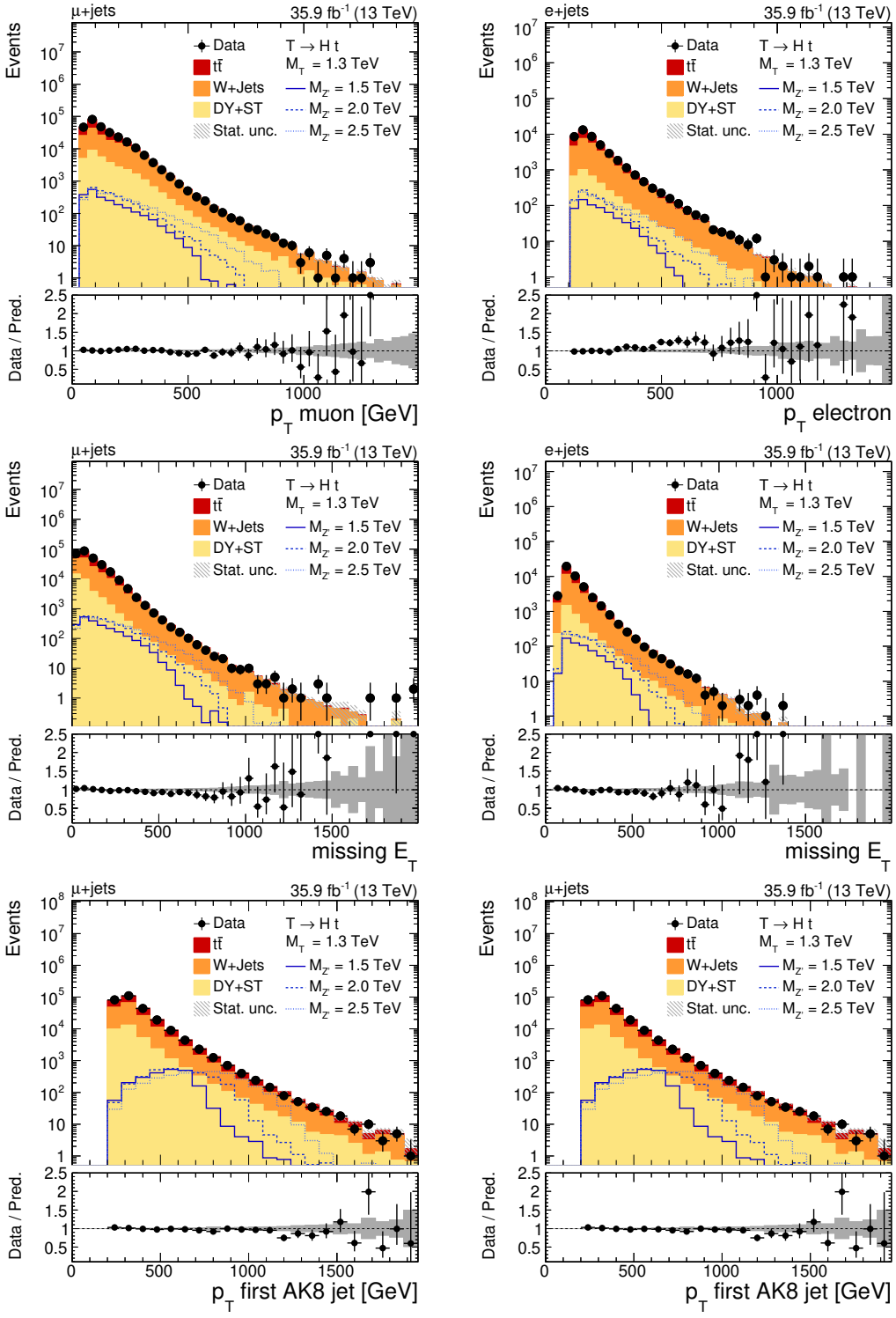


Figure 7.5: The distribution after the basic selection of the muon p_T (upper left) in the muon channel, the electron p_T in the e channel (upper right), p_T^{miss} in the μ channel (middle left) and in the e channel (middle right), the p_T of leading-p_T large-R jet in the μ channel (lower left) and the p_T of the small-R jets in the μ channel (lower right). Only statistical uncertainties are shown.

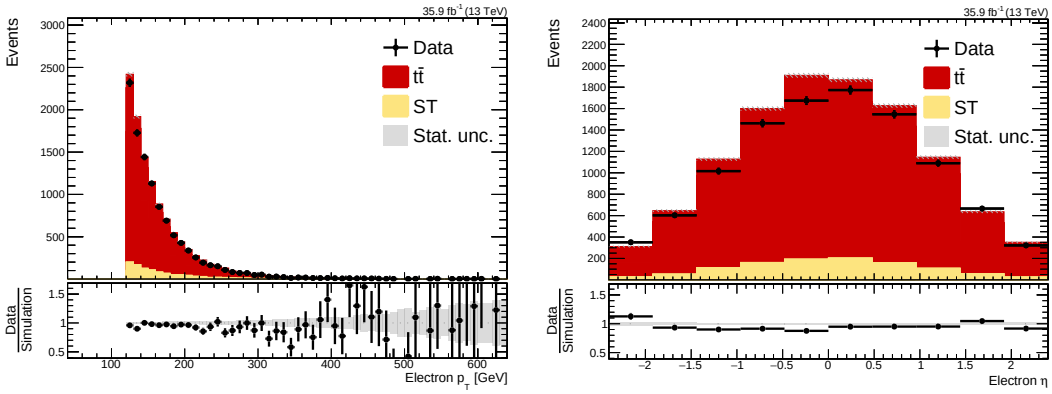


Figure 7.6: The p_T and η distribution of the probe electron. Only statistical uncertainties are shown.

Efficiency and Correction Factor Measurement

In the ideal case, the trigger efficiency is 100% above a certain p_T threshold. However, detector and trigger level effects result in a reduced efficiency. In order to model the detector effects, the correction factor is measured as function of p_T and $|\eta|$.

The trigger efficiency is defined as

$$\epsilon = \frac{\#\text{events passed (trigger + selection)}}{\#\text{events passed (selection)}}, \tag{7.1}$$

where the denominator is the number of events that pass the selection mentioned above and the nominator is the number of events that pass the selection and the trigger combination under test.

The resulting efficiency for data and simulation is shown in Fig. 7.7 (upper) as a function of the electron p_T. In data a turn-on behaviour at around 200 GeV is visible where the second trigger becomes efficient. The resulting correction factor, which is the ratio between the data efficiency and the simulation efficiency, as a function of p_T can be seen in Fig. 7.7 (lower left). Since the efficiency as function of $|\eta|$ is rather flat, a closure test with the p_T-dependent scale factors was performed. The result can be seen in Fig. 7.7 (lower right). Data and simulation mostly agree within the statistical uncertainties. However, a few outlier are visible. In order to cope with the differences in η an systematic uncertainty of 1% is added to the p_T-dependent correction factor.

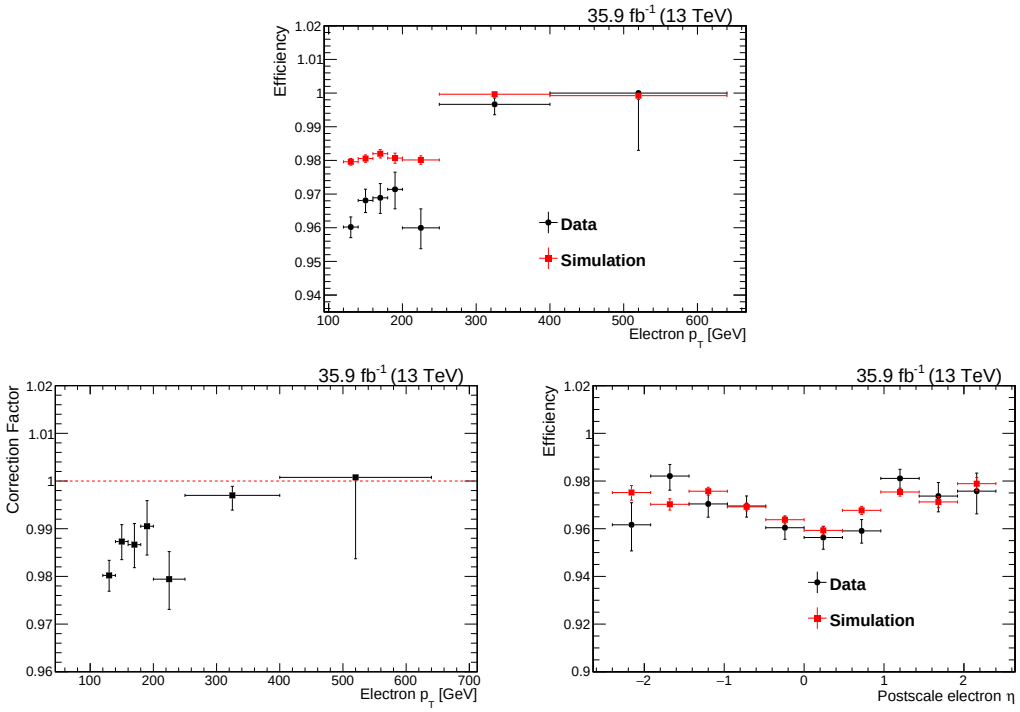


Figure 7.7: (upper) The trigger efficiency in data and simulation as function of the electron p_T. (lower left) The measured correction factor as function of the probe electron p_T. (lower right) The trigger efficiency in data and simulation as function of the electron η with the p_T-dependent correction factor applied.

7.4 Jet Substructure Selection

The challenging part of a search in the boosted regime is the discrimination of large-R jets originating from a boson or a t quark decay against SM multijet background. Several jet properties are combined to *tag* a certain particle decay. The most natural property to distinguish the heavy SM bosons and the t quark is the mass of the large-R jet. In order to mitigate the effect of initial state radiation and the underlying event[5], soft and wide angle radiation is removed from the jet. The most commonly used algorithm used in CMS is the *soft drop algorithm* [127] and the mass calculated from the remaining constituents of the jet is called the *soft drop mass*.

The soft drop mass (M_{SD}) of the signal and the SM background in this search is depicted in Fig. 7.8. Three different signal samples representing the three different decay modes of the T are shown. Each of the signal samples peaks at a different boson mass, allowing to use a mass requirement to select a specific signal. However, only

[5] Any hadronic activity that cannot be attributed to particles from the hard scattering.

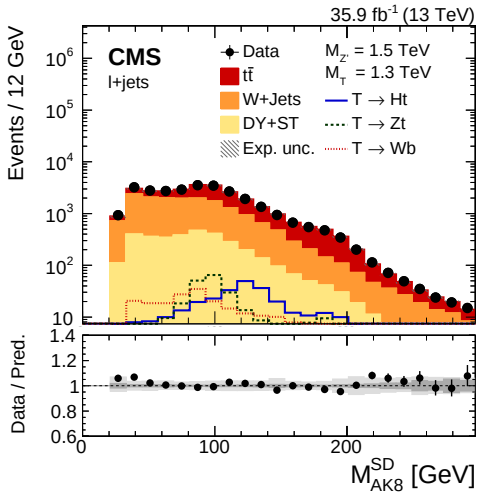


Figure 7.8: Distribution of the soft drop mass of jets as reconstructed with the anti-k_T jet algorithm with R = 0.8 after the event selection. Events are shown in the combined lepton+jets channel, with contributions from data, simulated signal samples, and the simulated SM backgrounds. The expected signal distribution from various T decay modes is shown for the example mass configuration $M_{Z'} = 1.5\,\text{TeV}$ and $M_T = 1.3\,\text{TeV}$ with a nominal cross section $\sigma(Z' \to tT)$ of 1 pb. The lower panel shows the ratio of data to predicted background. Here the darker grey band indicates the statistical uncertainty, whilst the lighter grey band shows the combined statistical and systematic uncertainty. Published in Ref. [1].

the simple mass requirement results in an overwhelming amount of events from SM processes. Therefore, additional jet substructure variables are used to suppress SM backgrounds that do not contain the particle of interest.

A commonly used variable is the *N-subjettiness* τ_N [129], which is a measure for the degree of compatibility with the jet having N energy axes. Typically, the N-subjettiness ratios $\tau_{21} = \tau_2/\tau_1$ or $\tau_{32} = \tau_3/\tau_2$ are used in a tagger, where small values of τ_2/τ_1 indicate a two prong decay (e.g. hadronic W/Z decay) and small values of τ_3/τ_2 indicate a three prong decay (e.g. hadronic t decay).

While N-subjettiness shows a good performance to distinguish Z/W bosons and t quarks from SM multijet background, for H boson tagging an additional property is exploited. With $\mathcal{B}(H \to b\bar{b}) = 58.4\%$ [29] jets originating from H boson decays can be classified with subjet b-tagging [124]. Subjet b-taggers use the unique property of the relatively long lifetime of B mesons that allow them to travel a certain distance before decaying to distinguish them from jets originating from light quarks. In order to increase the signal efficiency two H taggers are used: H_{2b} tagger, which requires two subjet b-tags, and H_{1b}, which requires one subjet b-tag. Identifying jets with subjet

b-tagging does not only result in jets originating from H decays but also from $Z \to b\bar{b}$ decays.

In summary, four taggers are used in this analysis:

- Z/W tagger: large-R jets are denoted Z/W-tagged if their soft drop jet mass is in the range $60 < M_{SD} < 115\,GeV$ and their N-subjettiness ratio fulfils $\tau_{21} < 0.5$.
- H tagger: two different H taggers are used:
  - H_{2b} tagger: large-R jets are denoted H_{2b}-tagged if their soft drop jet mass is in the range $100 < M_{SD} < 150\,GeV$ and two subjet b tags are found. This more stringent selection is used to reduce backgrounds in regions with significant background contributions.
  - H_{1b} tagger: large-R jets are denoted H_{1b}-tagged if their soft drop jet mass is in the range $100 < M_{SD} < 150\,GeV$ and exactly one subjet b tag is found. This less stringent selection is used in regions with low background contributions.
- t tagger: large-R jets are denoted t-tagged if their soft drop jet mass is in the range $150 < M_{SD} < 220\,GeV$ and their N-subjettiness ratio fulfils $\tau_{32} < 0.57$.

At least one Z/W or H tag is required, while the t tagger is only used to categorize the events into the boosted and resolved final state as shown in Fig. 7.2. Together with the event selection described in Sec. 7.3 this forms the signal region.

7.4.1 Measurement of the Correction Factors for the Misidentification Rate

All taggers used in this analysis do not only tag the particles which they are designed for, i.e. the H tagger tags large-R jets originating from the H boson (identification) but also large-R jets from W bosons and from quarks and gluons (misidentification). In order to account for the different misidentification rates of the taggers in data and simulation correction factors are measured. The misidentification rates of large-R jets originating from quarks and gluons (called quark and gluon jets in the following) for all taggers is measured as a function of the large-R jet p_T in a QCD multijet enriched region, where a negligible amount of W, Z and H bosons are present. A jet sample is used for this measurement, which is selected using trigger paths based on the scalar sum (H_T) of the p_T of small-R jets with $p_T > 30\,GeV$ and $|\eta| < 3$, requiring $H_T > 900\,GeV$. Only large-R jets with $p_T > 200\,GeV$ and $|\eta| < 2.5$ are considered. The

soft drop mass of large-R jets used has to be at least 30 GeV. A veto on leptons with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$ is placed such that this region is orthogonal to the signal region. Distributions of the large-R quantities for data and simulation are shown in Fig. 7.9. The simulation can not describe the data in the QCD multijet enriched region to a perfect agreement when considering only statistical uncertainties. However, when measuring the correction factors as a function of p_T the disagreement cancels out and is not influencing the measurement.

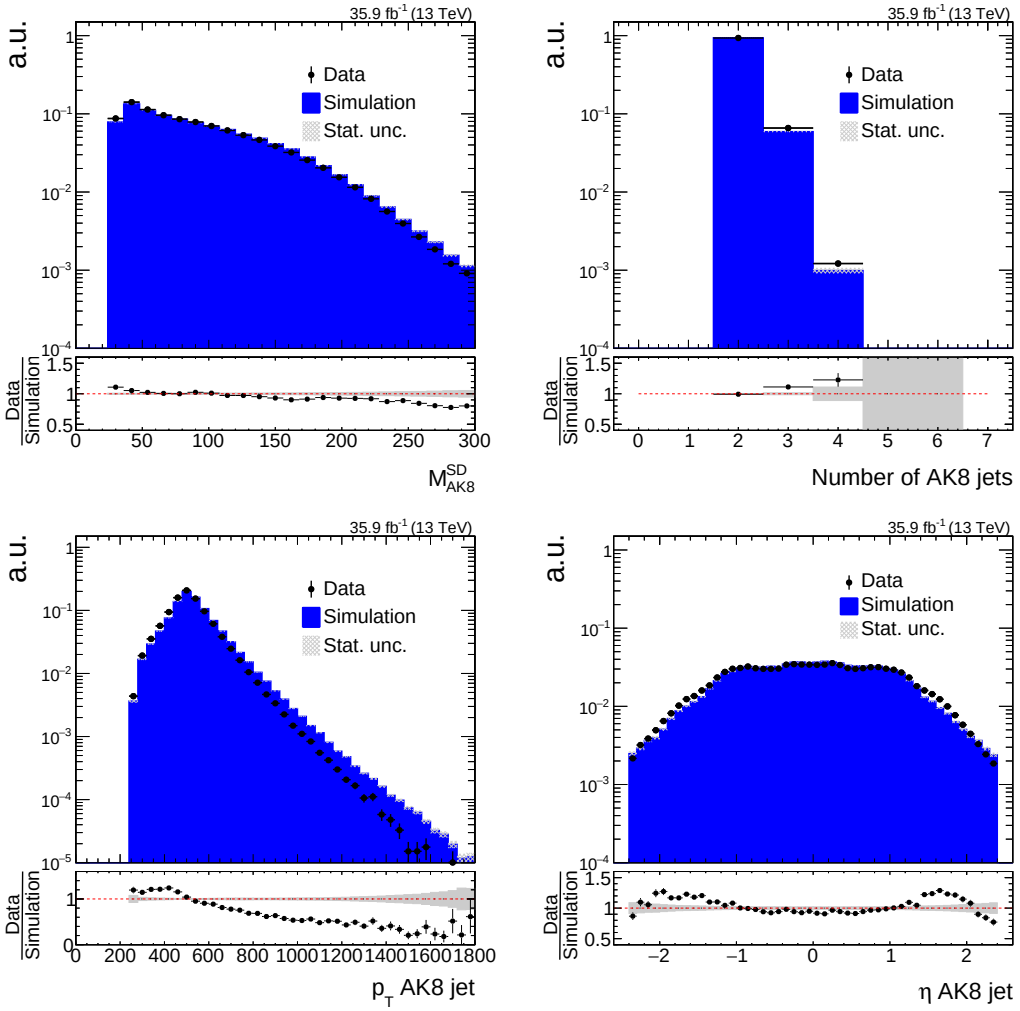


Figure 7.9: Control distribution in the jet sample to determine the misidentification rate. Shown are the mass of all large-R jets in the event (upper left), the number of large-R jets in the event (upper right), p_T distribution of all large-R jets (lower left) and η distribution of all large-R jets (lower right).

The misidentification rate is defined as the number of quark and gluon jets before the tagger is applied divided by the number of quark and gluon jets after the tagger is applied. The misidentification rate for the Z/W tagger in simulation and data is

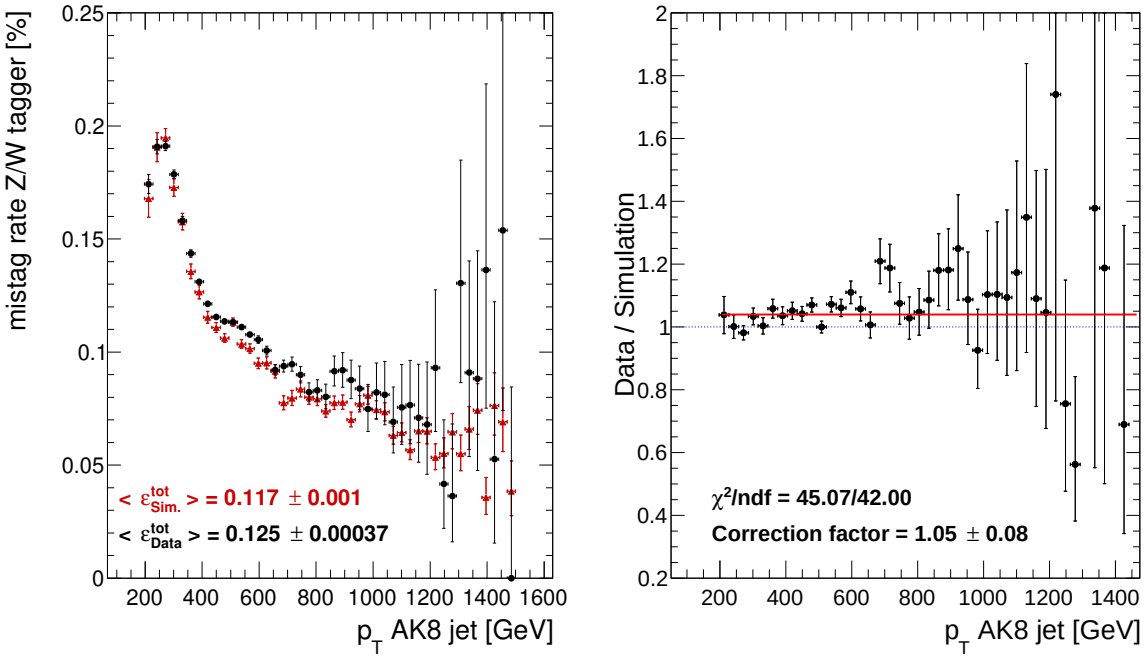


Figure 7.10: (left) The misidentification rate of the Z/W tagger in data and simulation. (right) The ratio of the two misidentification rates with a constant line fitted (red line).

shown in Fig. 7.10 (left) as an example. While the misidentification reduces with increasing p_T, due to the correlation between p_T and mass, the difference between the two misidentification rates (right) is to first order flat as a function of p_T within the uncertainties. Therefore, a constant function (red line) is fitted to the ratio to determine the correction factor. A systematic uncertainty is derived based on the choice of the fit function by calculating the differences between a constant fit, a linear fit and a quadratic fit.

The correction factors derived are 1.05 ± 0.08, 1.15 ± 0.18, 1.22 ± 0.05 and 0.95 ± 0.02 for the Z/W tagger, the H_{2b} tagger, the H_{1b} tagger and the t tagger, respectively.

The two H taggers, which require subjet b-tags (H_{2b}, H_{1b}), have a broad mass window and can also misidentify jets originating from the t quark decay. The agreement between data and simulation in the H tagger misidentification rates for tquark decays can not be tested in the jets sample, since a negligible amount of jets results from tdecays. Therefore, an additional measurement of the misidentification rate in a $t\bar{t}$ enriched region is carried out. This measurement is done in a tag-and-probe approach. This method is commonly used to measure the efficiency or misidentification rate of any object by exploiting a di-particle topology like $t\bar{t}$. The principle idea is to

1. consider processes where from the presence of one particle the presence of another particle can be deduced, e.g. $Z \to ll$ or $t\bar{t}$ production.

2. One of the particle decays, e.g. the leptonic t quark, has to pass strict identification criteria (*the tag*) to ensure a high purity of the required events, i.e. reduce the amount of other processes like W+jets or QCD multijet production.

3. The other particle decay, e.g. the hadronic t quark, has to pass loose identification requirements (*the probe*).

4. The efficiency or misidentification rate is then defined as the number of probes that pass the loose requirements and the tagger under test, divided by the number of probes that pass the loose requirements

$$\epsilon_{\text{mis}} = \frac{\#\text{probes(loose requirements + tagger requirement)}}{\#\text{probes(loose requirements)}}. \tag{7.2}$$

In this measurement the *tag* is a t quark that decays leptonically. The *probe* are the decay products of the t quark that decays hadronically. While the former has the same strict requirements as the leptonic t quark decay in this search, the latter requires a large-R jet and a small-R jet with the same p_T and η requirements as in this search. In this way the leading-p_T large-R jet originates from the t quark decay (the probe). The resulting correction factor is 1.019 ± 0.18 and 0.9966 ± 0.03 for the H_{2b} tagger and the H_{1b} tagger, respectively. This correction factor is only applied to $t\bar{t}$ events that have a H tag. While most of the $t\bar{t}$ background end up in the Z/W category, 2% of the $t\bar{t}$ background have a H_{2b} tag and 33% have a H_{1b} tag. Therefore, the large uncertainty, especially on the H_{2b} misidentification rate has a small impact on the analysis result.

7.4.2 Measurement of Correction Factors for the Efficiency

The principle idea of measuring the efficiency of a tagger is to measure the number of particles that are targeted by this tagger (e.g. W bosons) before and after the tagger is applied. In order to measure the efficiency of the Z/W tagger the amount of boosted W bosons (reconstructed by a large-R jet) before and after the tagger needs to be determined. However, it is nearly impossible to find a selection that results in events only containing a boosted W bosons with the p_T requirements set in the basic selection explained in Sec. 7.3. The region with boosted W bosons will also contain W bosons that are not boosted, hence resolved decays of the W boson. A $t\bar{t}$ enriched region gives a good starting point since it is enriched in processes containing a real W boson. But the problem remains that events from a $t\bar{t}$ decay that contain a resolved

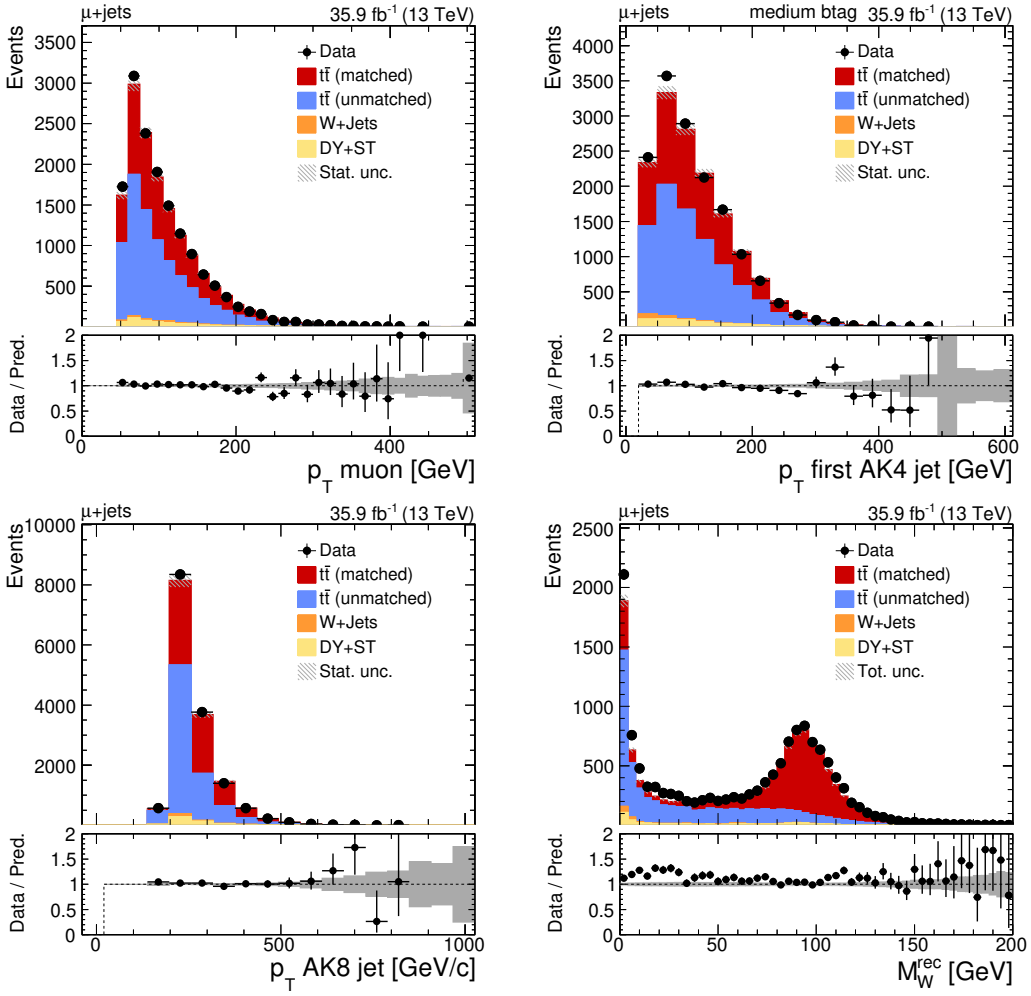


Figure 7.11: Control distributions in the $t\bar{t}$ enriched region to measure the efficiency of the Z/W tagger with a tag-and-probe method. (upper left) the distribution of the muon p_T, (upper left) the distribution of the p_T of the leading-p_T small-R jets, (lower left) the distribution of the p_T of the large-R jet (probe jet) and (lower right) distribution of the soft drop mass of the leading-p_T large-R jet. Events are shown in the μ+jets channel, with contributions from data and the simulated SM backgrounds. $t\bar{t}$ events are split into matched (ΔR(large-R jet, W) < 0.4) and unmatched (ΔR(large-R jet, W) > 0.4) events. The lower panel shows the ratio of data to predicted background. Here the lighter grey band shows the statistical uncertainty.

W boson decay enter the phase space. However, it is possible to differentiate between both contributions in the distribution of the mass of the leading-p_T large-R jets. A peak around the W boson mass is expected from events with a boosted W decay, while a smoothly falling spectrum is expected from resolved W boson decays and events not containing a W boson. In order to differentiate between both contributions a background and signal parametrisation will be used. *Background events* are events

where no real W boson is present (e.g. Drell-Yan) or where the W boson is not boosted. *Signal events* are only events where the hadronically decaying W boson is boosted.

In order to ensure a $t\bar{t}$ enriched region with boosted W bosons the efficiency is measured in the same tag-and-probe method as described in Sec. 7.4.1. The *tag* is a t quark that decays leptonically. The *probe* is leading-p_T large-R jet in the event. In this way the large-R jet is supposed to be the boosted W boson, where the purity of this assumption can be verified in simulation.

The p_T distributions of the different objects used, the muon, the leading-p_T small-R jet and the leading-p_T large-R jet can be found in Fig. 7.11. All variables show a good agreement between data and simulation. The amount of $t\bar{t}$ processes is enhanced compared to other processes like W+jets or QCD multijet background. The $t\bar{t}$ events are split based on the origin of the leading-p_T large-R jet into matched and unmatched events. Events where ΔR between the large-R jet and the W boson based on generator-level information is less than 0.8, are called *matched* $t\bar{t}$ events and should represent the signal events. Events where ΔR between the large-R jet and the W boson is greater than 0.8, are called *unmatched* $t\bar{t}$ events and should represent the background events. The soft drop mass of the leading-p_T large-R jet can be seen in Fig. 7.11 (lower right). The matched $t\bar{t}$ events in red peak nicely around the W boson mass while the unmatched $t\bar{t}$ events show a smoothly falling distribution as expected.

The soft drop mass before and after the tagger is applied is used to determined the efficiency. The efficiency is defined as the number of large-R jets that originated from a boosted W boson after the tagger is applied divided by the number before the tagger is applied. The number of W bosons before the tagger is applied N_{W} will be determined with a fit to simulation and data, while the number of W bosons after the tagger is applied $N_{\text{sig, after tagger}}$ will be determined by a counting experiment.

The Fitting Procedure: Determination of N_{W}

The fit to determine N_{W} is carried out in four steps. :

1. Find a variable that is suited to differentiate between background and signal events, e.g. the soft drop mass of the p_T-leading large-R jet.

2. Find a parametrisation of the signal and background events and validate it on simulation.

3. The validation is performed by fits to signal-only and background-only events. The matched $t\bar{t}$ and unmatched $t\bar{t}$ events are a good starting point for this. Since

the separation into signal and background events is ambiguous and depends on the matching parameter, its influence on the result needs to be checked.

4. Fit the signal+background function to simulation and data. The number of W bosons is the integral of the resulting signal fit.

The soft drop mass of the large-R jet includes matched and unmatched events. Two functions are needed to separate background events and signal events. As a signal function a Voigt profile, which is a convolution of a Breit-Wigner function and a Gaussian is used, while as a background function an error function is used. The combined, signal+background function is defined by

$$\mathrm{f}(x) = p_3 \mathrm{Voigt}(x - m', \sigma', p_1) + p_6(1 - \frac{1}{1 + \exp(-(x - p_7)/p_8)}), \tag{7.3}$$

where $m' = (p_2 + p_4 x)$ and $\sigma' = (p_0 + p_5 x)$. The first six parameters (p_0, p_1, p_2, p_3, p_4, p_5,) belong to the signal function, whereas the last three parameters (p_6, p_7, p_8) describe the background function. The parameter p_1 is fixed to 2.09, since this represents the width of the W boson [29]. Both functions as well as the combination are tested on the matched and unmatched simulation to validate the parametrisation.

The signal function fit to matched $\mathrm{t\bar{t}}$ events can be seen in Fig. 7.12 (upper left). The matched $\mathrm{t\bar{t}}$ events should represent only boosted W bosons that are reconstructed in a large-R jet. However, it is visible that the distribution has a tail towards low masses resulting from large-R jets that do not originate from boosted W bosons. These events are considered to be background events. The amount of background events in the matched $\mathrm{t\bar{t}}$ events depends highly on the matching parameter used. If the matching parameter is small ($\Delta R < 0.4$) the amount of background events is negligible as shown in Fig. 7.12 (upper right). Vice versa the problem is visible in the unmatched events, where a contamination of signal events (large-R jets originates from a boosted W boson) is present, when using a matching parameter of 0.8 as shown in Fig. 7.12 (lower left). If the matching parameter increases the amount of signal events is reduced Fig. 7.12 (lower right). However, both functions can describe the background or signal events, respectively.

Since data can not be split into matched $\mathrm{t\bar{t}}$ and unmatched $\mathrm{t\bar{t}}$ a signal+background fit is performed for both simulation and data to determine N_{W}. The number of signal events is calculated by integrating the signal function with the parameters determined by the fit. The resulting fit for both simulation and data can be seen in Fig. 7.13. The solid red curve shows the background function plus the signal function, whereas the orange line shows the contribution from background events only.

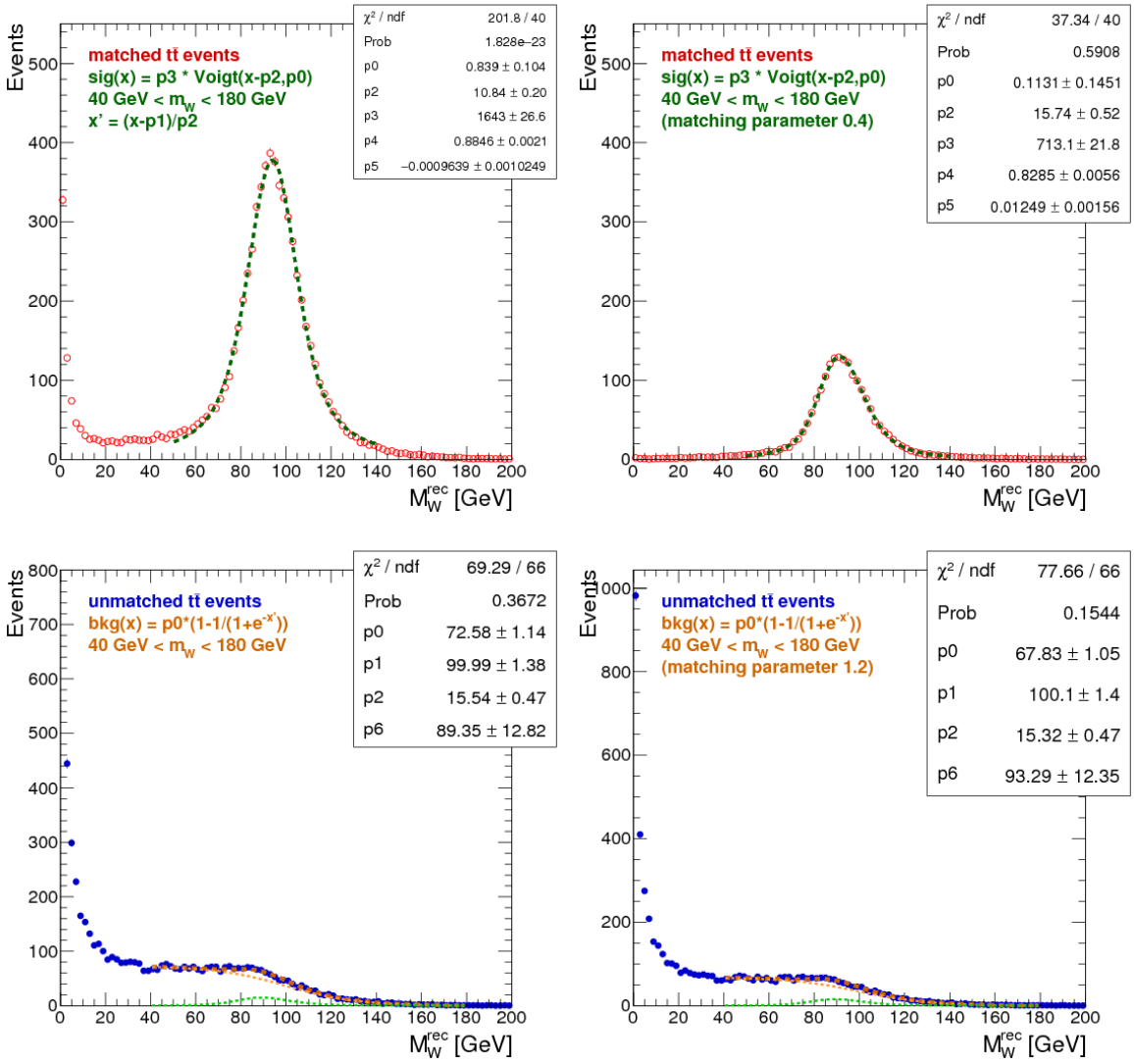


Figure 7.12: The fit function to determine the number of events before the Z/W tagger is applied. (upper left) The signal fit only with a matching parameter of 0.8, (upper right) the signal fit only with a matching parameter of 0.4, (lower left) the background+signal fit for a matching parameter of 0.8 and (lower right) the background+signal fit for a matching parameter of 1.2.

Counting Experiment: Determination of $N_{\text{sig, after tagger}}$

The number of W bosons after the tagger is applied ($N_{\text{sig, after tagger}}$) is needed to calculate the efficiency. It is not possible to perform the background+signal fit after the tagger is applied since the background function can not be constrained due to the missing low mass tails as seen in Fig. 7.14.

Instead a counting experiment is done. All events ($N_{\text{after tag}}$), signal+background events, are counted after the tagger is applied and the number of expected background

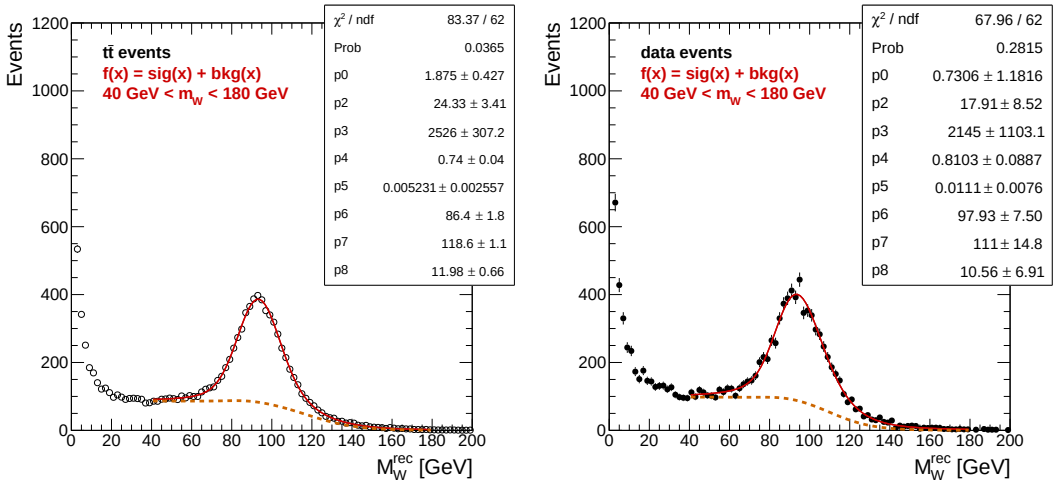


Figure 7.13: The fit of the soft drop mass of the leading-p_T large-R jet with the background plus signal function to simulation (left) and data (right). The circles show the simulation or data in the $t\bar{t}$ enriched region. The pink line shows the signal plus background fit, whereas the orange line shows the contribution of the background events.

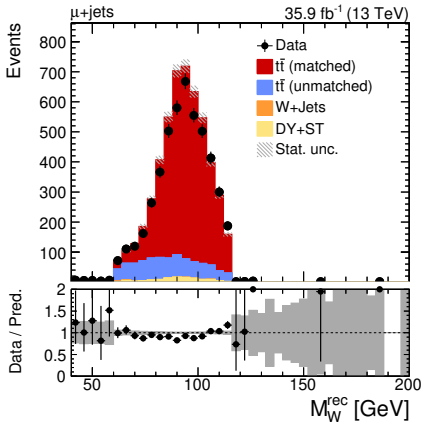


Figure 7.14: The reconstructed W Boson mass after the Z/W tagger is applied. Simulated events are split in matched $t\bar{t}$ events and unmatched $t\bar{t}$ events based on the ΔR between the large-R jet and the generator-level W boson.

events after the tagger is applied ($N_{\text{bkg, after tagger}}$) is subtracted

$$N_{\text{sig, after tagger}} = N_{\text{after tag}} - N_{\text{bkg, after tagger}} \tag{7.4}$$

$$= N_{\text{after tag}} - N_{\text{bkg, before tag}} \cdot \epsilon_{\text{miss}}, \tag{7.5}$$

where $N_{\text{sig, after tagger}}$ is the number of signal events after the tagger is applied. This number needs to be derived in order to calculate the efficiency. $N_{\text{bkg, before tag}}$ is the

number of background events before the tagger is applied and ϵ_{miss} is the misidentification rate of the background events. While $N_{\text{after tag}}$ is known by counting the events after the tagger, $N_{\text{bkg, before tag}}$ is known from the fit procedure and ϵ_{miss} needs to be calculated for simulation and data. The misidentification rate in simulation $\epsilon_{\text{miss}}^{\text{sim}}$ is defined as:

$$\epsilon_{\text{miss}}^{\text{sim}} = \frac{N_{\text{after tag}}^{'\text{bkg}}}{N_{\text{before tag}}^{\text{bkg}}}, \tag{7.6}$$

where $N_{\text{after tag}}^{'\text{bkg}}$ is the true number of background events after the tagger and $N_{\text{before tag}}^{\text{bkg}}$ is the number of background events before the tagger is applied. The former is unknown due to the contamination with signal events as described above. The latter is known from the fit procedure. The misidentification rate for data is calculated by using the correction factor derived in a QCD multijet enriched region as described in Sec. 7.4.1:

$$\epsilon_{\text{miss}}^{\text{data}} = \epsilon_{\text{miss}}^{\text{sim}} \times SF_{\text{mistag rate}}(1.053) \tag{7.7}$$

Since the background events are, due to the dependency on the matching parameter, always contaminated with signal events, the number of background events after the tagger is applied $N_{\text{after tag}}^{'\text{bkg}}$ needs to be corrected for the signal contamination:

$$N_{\text{after tag}}^{'\text{bkg}} = N_{\text{after tag}}^{\text{bkg}} - N_{\text{before tag}}^{\text{signal in bkg}} \times \epsilon_{\text{approx.}}^{\text{sim}}(\Delta R < 0.4), \tag{7.8}$$

where $N_{\text{after tag}}^{\text{bkg}}$ are the number of unmatched $\mathrm{t\bar{t}}$ events and other processes, $N_{\text{before tag}}^{\text{signal in bkg}}$ is the number of signal events in unmatched $\mathrm{t\bar{t}}$ before the tagger is applied, obtained from the fit procedure (Fig. 7.13) and $\epsilon_{\text{approx.}}^{\text{sim}}(\Delta R < 0.4)$ is the efficiency in simulation of the tagger derived from events with a matching of $\Delta R < 0.4$ between the W boson and the large-R jet

$$\epsilon_{\text{approx.}}^{\text{sim}}(\Delta R < 0.4) = \frac{N_{\text{after tag}}^{\text{signal}}}{N_{\text{before tag}}^{\text{signal}}} = 0.776, \tag{7.9}$$

both, $N_{\text{after tag}}^{\text{signal}}$ and $N_{\text{before tag}}^{\text{signal}}$ are known from the fit procedure. The efficiency is then defined as the number of W bosons after the tagger is applied (counting experiment) divided by the number of W bosons before the tagger is applied (fit procedure). The efficiency measured in simulation is $0.852^{+0.018}_{-0.017}$ and the efficiency in data correspond to $0.777^{+0.065}_{-0.061}$, leading to a scale factor of $0.91^{+0.078}_{-0.074}$, which includes an additional 1% systematic uncertainty due to the choice of the fit function.

7.5 Reconstruction of the Z′ Mass

The fundamental idea of reconstructing the Z′ mass is to reconstruct each decay product (tZt , tHt) and combine them to a Z′ candidate. Since each decay product can have multiple candidates when reconstructed, several Z′ candidates will be available while the best candidate is chosen based on a χ^2. The heavy boson (H/Z) is reconstructed with a tagged large-R jet, while the top quarks are reconstructed using the lepton, missing transverse energy and small-R jets in case of the leptonically decaying one. The hadronically decaying top quark is either reconstructed as one large-R jet (boosted regime) or with small-R jets (resolved regime). The following paragraphs will describe in more detail how the reconstruction of the neutrino and the reconstruction of the t quark candidates is performed.

7.5.1 Reconstruction of the Neutrino Momentum

The signature of the signal considered here, involves exactly one neutrino, which does not interact with the detector material but can be derived from $\vec{p}_T^{\,\text{miss}}$. Assuming the neutrino from the signal decay is the only source of $\vec{p}_T^{\,\text{miss}}$, it can be used to determine the p_T from the neutrino directly. The missing component, p_z, can be calculated from $\text{p}_\text{T}^\text{miss}$and the lepton using the fact that the neutrino originates from a W boson:

$$P_\text{W}^2 = M_\text{W}^2 = (P_l + P_\nu)^2\,, \tag{7.10}$$

where P_W is the four-momentum of the W boson, M_W is the mass of the W boson, P_l is the momentum of the lepton and P_ν is the momentum of the neutrino. The solution of this equation is

$$p_{z,\nu}^{\pm} = \frac{a p_{z,l}}{p_{T,l}^2} \pm \sqrt{\frac{a^2 p_{z,l}}{p_{T,l}^2} - \frac{E_l^2 p_{T,\nu} - a^2}{p_{T,l}^2}}, \tag{7.11}$$

where $a = M_\text{W}^2/2 + p_{T,l}p_{T,\nu}\cos\Delta\phi$, and $\Delta\phi$ the azimuthal angle between the $\vec{p}_T^{\,\text{miss}}$ and the lepton. This results in zero to two real solutions of the neutrino momentum. In case of zero real solution the real part of the complex solution is taken. In case of two real solutions both are considered as different neutrino candidates.

7.5.2 Reconstruction of the Top Quark Candidates

Two t quark candidates are reconstructed: a leptonic candidate and a hadronic candidate. The leptonic candidate is reconstructed as the vectorial sum of the lepton, the

Table 7.1: The expected mass and width values used in the χ^2 of the reconstruction procedure.

$M^{rec}_{\mathrm{t,lep}}$	$\overline{M}_{\mathrm{t,lep}}$	$M^{rec}_{\mathrm{t,had}}$	$\overline{M}_{\mathrm{t,had}}$
173 GeV	16 GeV	176 GeV	15 GeV

neutrino and at least one small-R jet. The hadronic candidate is reconstructed either with a top tagged large-R jet or with at least one small-R jet. Each small-R jet in the event can be assigned to either the leptonic candidate, the hadronic candidate or none of them. The t quark candidate is then the vectorial sum of the assigned objects. Each combination results in a new leptonic and hadronic t quark candidate and is stored.

7.5.3 Reconstruction of the Z′ Candidate

The Z′ candidate is then reconstructed by the vectorial sum of the H/Z tagged large-R jet, a leptonic t quark candidate and a hadronic t quark candidate. The invariant mass of the best candidate is called $M^{\mathrm{rec}}_{\mathrm{Z}'}$ and is used as sensitive variable. The best candidate is chosen by the lowest χ^2 value, which is defined as

$$\chi^2 = \frac{\left(\overline{M}_{\mathrm{t,lep}} - M^{rec}_{\mathrm{t,lep}}\right)^2}{\sigma^2_{\mathrm{t,lep}}} + \frac{\left(\overline{M}_{\mathrm{t,had}} - M^{rec}_{\mathrm{t,had}}\right)^2}{\sigma^2_{\mathrm{t,had}}}, \tag{7.12}$$

where $M^{rec}_{\mathrm{t,lep/had}}$ is the reconstructed mass of the leptonic/hadronic candidate, $\overline{M}_{\mathrm{t,lep/had}}$ is the expected mass from correctly matched events and $\sigma_{\mathrm{t,lep/had}}$ is the expected width of correctly matched events and is determined in simulation. Correctly matched means for the leptonic candidate that the ΔR between the generator-level lepton and the reconstruction-level lepton is less than 0.1, ΔR between the generator-level b quark and the reconstruction-level small-R jet is less than 0.4 and the $\Delta\phi$ between $\vec{p}_T^{\,\mathrm{miss}}$ and the generator-level neutrino is less than 0.3. In the case of the hadronic candidate the ΔR between the three generator-level quarks of the t quark decay and the small-R jets assigned to the candidate has to be less than 0.4. The values of the expected mass and width are listed in Tab. 7.1.

The χ^2 values for the best Z′ candidates are shown in Fig. 7.15 in the combined ℓ+jets channel. The values tend to zero for a correct reconstruction of the two t quarks and to higher values if one or both of the t quarks are misreconstructed. A correct reconstruction means that the jets, chosen in the reconstruction process, originate from the quarks of the t quark. The SM background feature a peak around $\chi^2 \approx 120$, which results from events where only one of the t quark is reconstructed correctly. This occur especially in single top production processes, where only one top is present and in $\mathrm{t\bar{t}}$

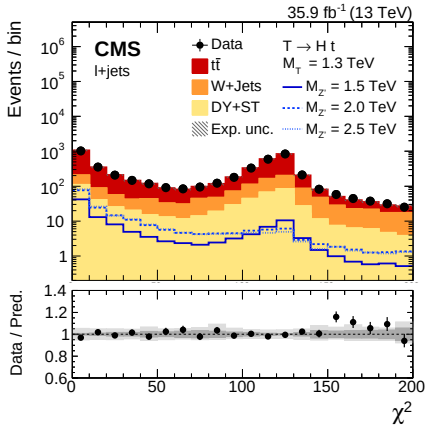


Figure 7.15: Distribution of the smallest χ^2 discriminator in each event for the combination of both top tag and no top tag categories, after the $\mathrm{t\bar{t}}$ reconstruction, combining both lepton channels. The expected signal distribution is shown for various $M_{\mathrm{Z'}}$ masses for a fixed mass $M_{\mathrm{T}} = 1.3\,\mathrm{TeV}$ in the $\mathrm{T} \to \mathrm{Ht}$ decay channel, each with a nominal cross section $\sigma(\mathrm{Z'} \to \mathrm{tT})$ of 1 pb. The lower panel shows the ratio of data to predicted background. Here the darker grey band indicates the statistical uncertainty, whilst the lighter grey band shows the combined statistical and systematic uncertainty. Published in Ref. [1].

processes if one t quarks is misreconstructed. It is also visible in signal simulation but only for the low $\mathrm{Z'}$ mass where the hadronically decaying t quark is not boosted.

7.6 Background Estimation

The main backgrounds of this analysis are $\mathrm{t\bar{t}}$ and W+jets. Both backgrounds are well understood and well modelled in simulation. Two effects of the simulations used here are known. The first effect is that the t quark $\mathrm{p_T}$ spectrum is significantly softer then the one predicted by simulation [159, 160] which results in a trend in the reconstructed $M_{\mathrm{Z'}}$ spectrum as can be seen in Fig. 7.16 (right). A dedicated corrections based on the ratio between data and $\mathrm{t\bar{t}}$ simulations [159, 160] was applied to overcome this issue. The effect of the correction can be seen in Fig. 7.16 (left). The data-to-simulation agreement improves with applying the correction.

The other known effect is that the overall normalisation of $\mathrm{t\bar{t}}$ simulation is about 10% lower, then measured in data. In order to validate this a dedicated control region is implemented.

In addition to the event selection that is described in Sec. 7.3, the control region is defined by one large-R jet with $M_{\mathrm{SD}} < 60\,\mathrm{GeV}$ or $M_{\mathrm{SD}} > 150\,\mathrm{GeV}$ to be orthogonal with

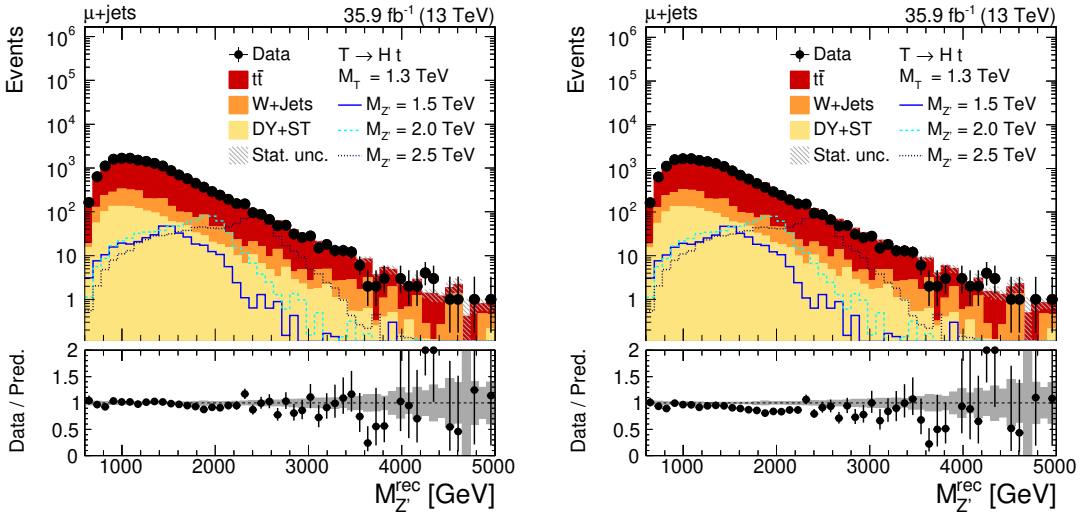


Figure 7.16: The distribution of the reconstructed $M_{Z'}$ in the $t\bar{t}$ enriched control region with (left) and without (right) the reweighting of the t quark p_T.

the signal region. The events in the control region are further split by the number of b-tagged small-R jets. Events that have zero b-tagged jets build the W+jets enriched control region, while events with at least one b-tagged jets build the $t\bar{t}$ enriched control region. Both control regions are used in the final fit to constraint the normalisation of $t\bar{t}$ and W+jets. A possible signal contamination of the control region is handled by the maximum-likelihood based fit for the upper cross section limit calculation.

Figure 7.17 shows both control regions for the electron and the muon channel with all systematic uncertainties after the maximum-likelihood based fit. A good data-to-simulation agreement is observed. Independently from the final fit, correction factors can be derived for the e and μ channel to understand the effect of the control region. The correction factor are not used for the final fit but show the shift suggested from the control region only. The correction factors are for the muon channel for $t\bar{t}$ 0.97 ± 0.015 and for W+jets 1.173 ± 0.048 and for the electron channel for $t\bar{t}$ 0.922 ± 0.03 and for W+jets 1.09 ± 0.105. A difference between the channels within the uncertainties is observed.

7.7 Systematic Uncertainties

The systematic uncertainties can affect either the normalisation only or the shape and the normalisation. The effect of all systematic uncertainties is summarized in Tab. 7.2.

Substructure Tagging: The efficiency and misidentification rate is varied by ± 1 standard deviation, where the variation is the systematic and statistical uncertainty

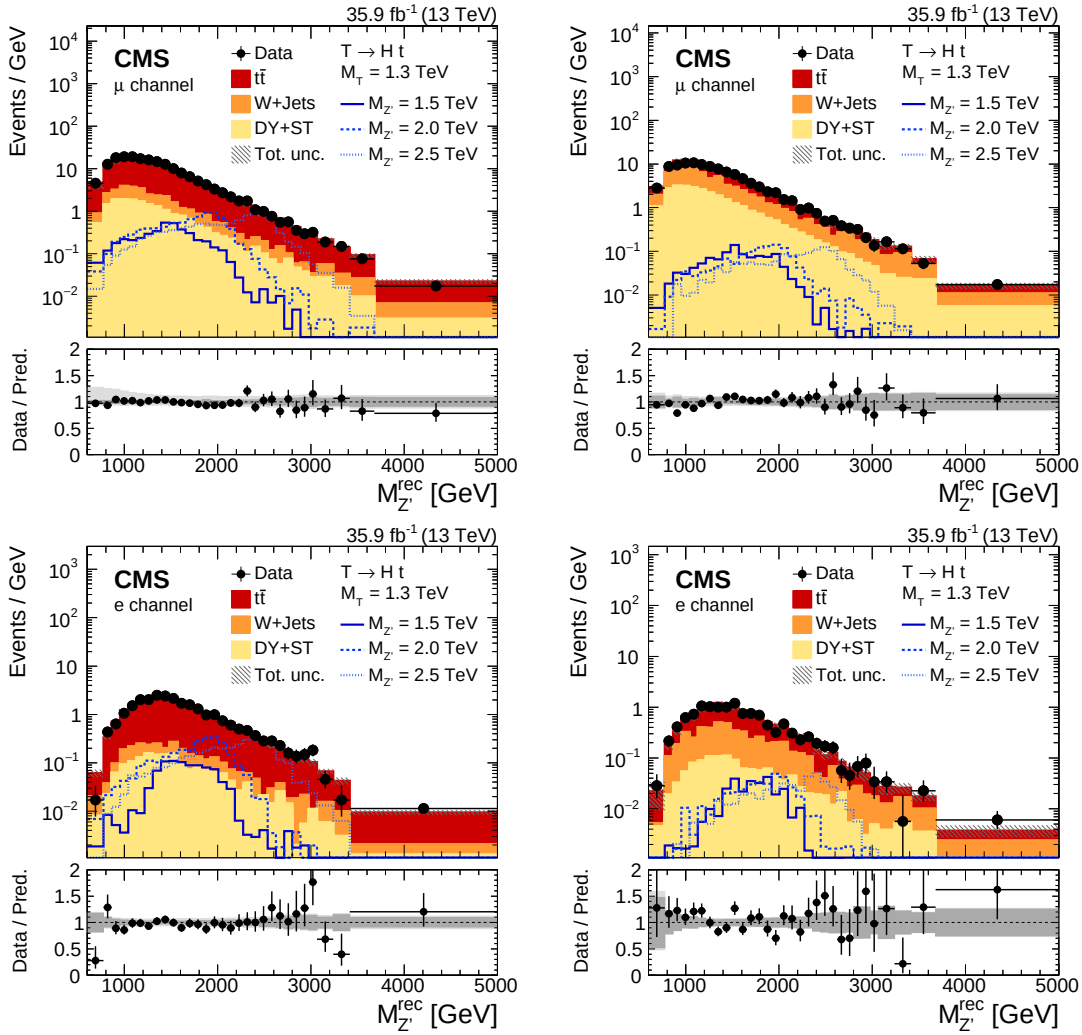


Figure 7.17: Distribution of the reconstructed Z′ boson mass in the μ channel (upper row) and e channel (lower row) for the $\mathrm{t\bar{t}}$-enriched control region (left) and for the W+jets-enriched region (right). The expected signal distribution is shown for various $M_{\mathrm{Z}'}$ masses for a fixed mass $M_{\mathrm{T}} = 1.3\,\mathrm{TeV}$ in the $\mathrm{T} \to \mathrm{Ht}$ decay channel, each with a nominal cross section $\sigma(\mathrm{Z}' \to \mathrm{tT})$ of 1 pb. The lower panel shows the ratio of data to predicted background. Here the darker grey band indicates the statistical uncertainty, whilst the lighter grey band shows the combined statistical and systematic uncertainty. Published in Ref. [1].

combined.

Jet Energy Scale and Resolution: The jet energy scale and resolution [123] is varied by ±1 standard deviation to estimate the effect on the signal and background templates and the uncertainty is propagated to $\mathrm{p_T^{miss}}$.

b-tagging Efficiency and Misidentification Rate: Data-to-simulation correction factors for the b quark identification and misidentification rate are applied as a function

of p_T and η. The uncertainty is estimated by varying the correction factor by ± 1 standard deviation [124]. The tagging efficiency of b- and c-jets are taken as fully correlated and therefore are varied simultaneously. The tagging efficiency of jets from light quarks and gluons is assumed fully uncorrelated and is varied independently.

Muon and Electron Identification and Trigger Efficiency: Data-to-simulation correction factors for the identification and trigger efficiencies are applied as a function of p_T and η. Each correction factor is varied by ± 1 standard deviation to estimate the uncertainties on the signal and background template.

Luminosity Uncertainty: The uncertainty on the integrated luminosity recorded by the CMS detector is 2.5% for 2016 data [152].

Pileup Uncertainty: The nominal minimum bias cross section for the pileup reweighting is 69 mb. The minimum bias cross section is varied by 5% and the samples are reweighted accordingly.

Renormalisation and Factorisation scales: The uncertainties in the factorisation and renormalization scales μ_F and μ_R are taken into account as proposed in Ref. [162, 163] by varying the default choice of scales by the following six combinations of factors, (μ_F, μ_R) $\times$ (1/2, 1/2), (1/2, 1), (1, 1/2), (2, 2), (2, 1), and (1, 2). The maximum and minimum of the six variations are computed for each bin of the reconstructed T mass distribution.

PDF: The PDF uncertainty is estimate from the choice of 100 different PDFs, provided with the NNPDF30_lo_as_0130 [38] set. All samples are re-weighted according to each of the varied sets. In each bin of the final distributions the symmetric standard deviation with respect to the nominal value is taken as the PDF uncertainty.

7.8 Results

A total of six categories are used in this search. The categories split the events in a boosted regime (t tag), relevant for a high mass difference $M_{Z'} - M_T$ and a resolved regime (no t tag). Furthermore, events are divided into separate categories corresponding to the tag of the boson jet. Two categories for H tagged jets are defined, based on the number of subjet b-tags. This leads to the best sensitivity in the low mass and in the high mass regime and in both decay channels $T \to Ht$ and $T \to Zt$. The signal efficiency for the different decay channels in each of the six categories is shown in Tab. 7.3. While the efficiency is similar for $T \to Ht$ and $T \to Zt$, the efficiency for $T \to Wb$ is reduced as the reconstruction might not be successful. Events, where $T \to Ht$, have the highest sensitivity in the two H categories and events, where $T \to Zt$, have the highest sensitivity in the Z/W categories.

Source	Uncertainty [%]	Shape	Categories
Z/W tagging efficiency	$8 \oplus 4.1 \times \ln(p_T/200\,\text{GeV})$		Z/W tag
Z/W mistag rate	±5.6–7.9	✓	Z/W tag
H_{2b}/H_{1b} tagging efficiency	9		H_{2b}/H_{1b} tag
H_{2b} mistag rate	±14–18	✓	H_{2b} tag
H_{1b} mistag rate	±3.2–4.6	✓	H_{1b} tag
H_{2b} mistag rate (only $t\bar{t}$)	18		H_{2b} tag
H_{1b} mistag rate (only $t\bar{t}$)	3		H_{1b} tag
t tagging efficiency	+7/−4		top tag
t mistag rate	1.8	✓	top tag
Jet energy scale	±0.1–5.5	✓	CR+SR
Jet energy resolution	<0.01	✓	CR+SR
b tagging AK4	±1.8–3.0	✓	CR
b tagging AK8	±2.7–7.3	✓	H_{2b}/H_{1b} tag
Muon ID	±0.1–2.6	✓	CR+SR
Muon trigger	±0.4–2.2	✓	CR+SR
Muon tracker	±0.5–1.8	✓	CR+SR
Electron ID	±0.3–3.1	✓	CR+SR
Electron trigger	±0.4–0.5	✓	CR+SR
Electron reconstruction	±0.1–3.0	✓	CR+SR
Luminosity	2.5		CR+SR
Pileup reweighting	±0.1–3.3	✓	CR+SR
μ_f and μ_r scales	6 variations	✓	CR+SR
PDF	100 samples	✓	CR+SR

The reconstructed Z' mass in the six categories are shown in Fig. 7.18 and Fig. 7.19 in the μ channel and the e channel, respectively. Events in the t tag category are shown on the left, events in the no t tag category are shown on the right. The first two rows show the two H tag categories, where the signal simulations with $T \to Ht$ are shown. The lower row shows the Z/W category, where the signal simulations with $T \to Zt$ are shown.

In each category three different signal samples with fixed T mass of 1.3 TeV are shown. Each signal sample peaks around the generated Z' mass. The SM backgrounds are taken from simulation and control regions are used in the maximum likelihood based fit to constrain the yield of the backgrounds. The binning in each category is chosen such that the statistical error on simulation is less than 30%.

No significant deviation from the SM expectation is observed in the reconstructed $M_{Z'}$ distribution in Fig. 7.18 and Fig. 7.19 . Upper limits at 95% CL are set on the production cross section times $\mathcal{B}$. In order to derive the 95% CL upper limits a Bayesian calculation with priors known to yield good frequentest properties [29, 164, 165] is used which is implemented in the THETA software package [166]. Limits for each

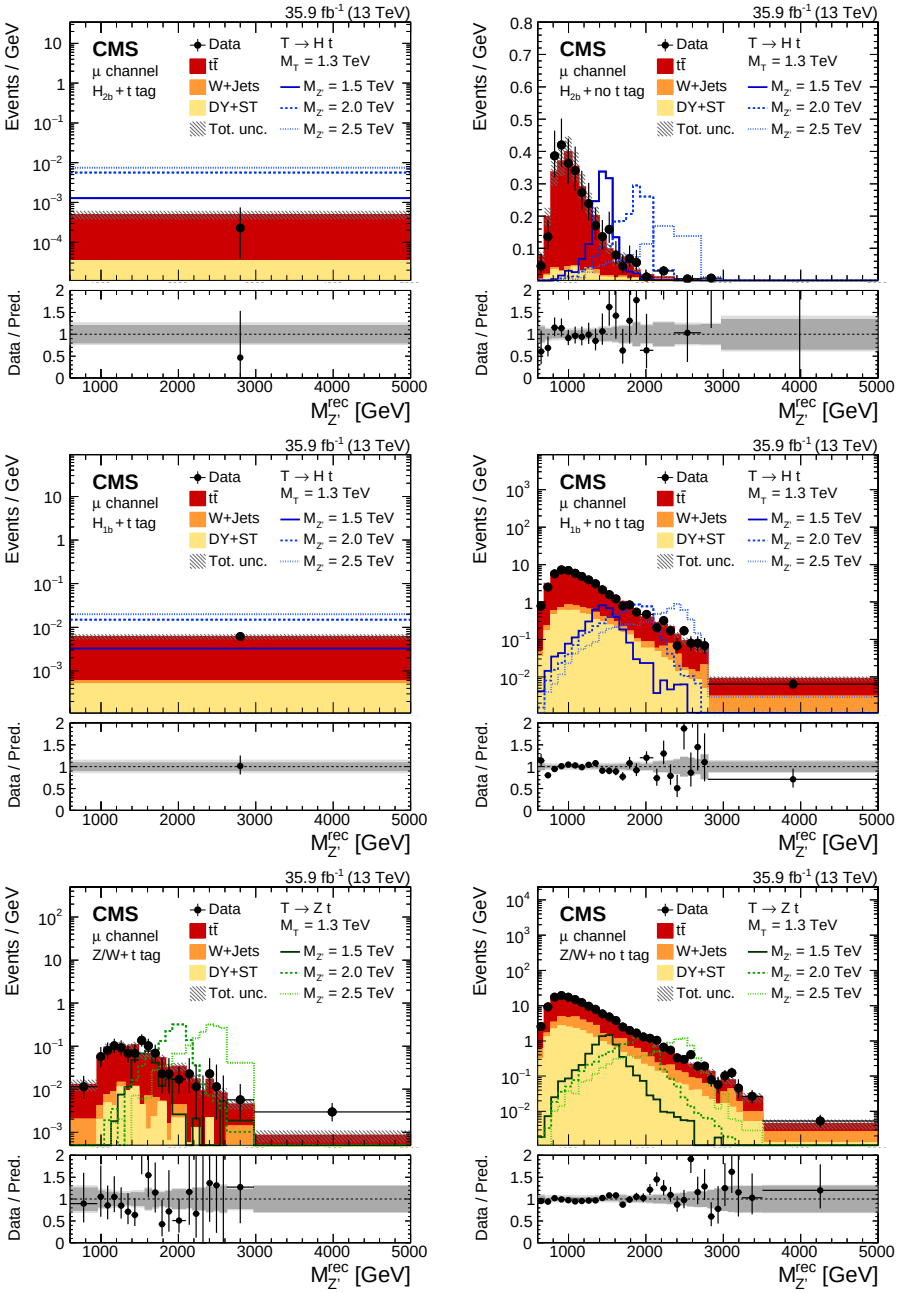


Figure 7.18: Distribution of the reconstructed Z′ resonance mass after the full selection in the μ+jets channel for the data, the expected SM background, and for the signal with different Z′ masses for a fixed T mass of 1.3 TeV. In the left (right) column the results in the top tag (no top tag) category are shown. Different rows display the distributions of events accepted by different taggers as well as the signal for the respective T decays: H_{2b} tagger and $T \to Ht$ decay (upper), H_{1b} tagger and $T \to Ht$ decay (middle), and Z/W tagger and $T \to Zt$ decay (lower). The signal histograms correspond to a nominal cross section $\sigma(Z' \to tT)$ of 1 pb. The lower panel shows the ratio of data to predicted background. Here the darker grey band indicates the statistical uncertainty, whilst the lighter grey band shows the combined statistical and systematic uncertainty. Published in Ref. [1].

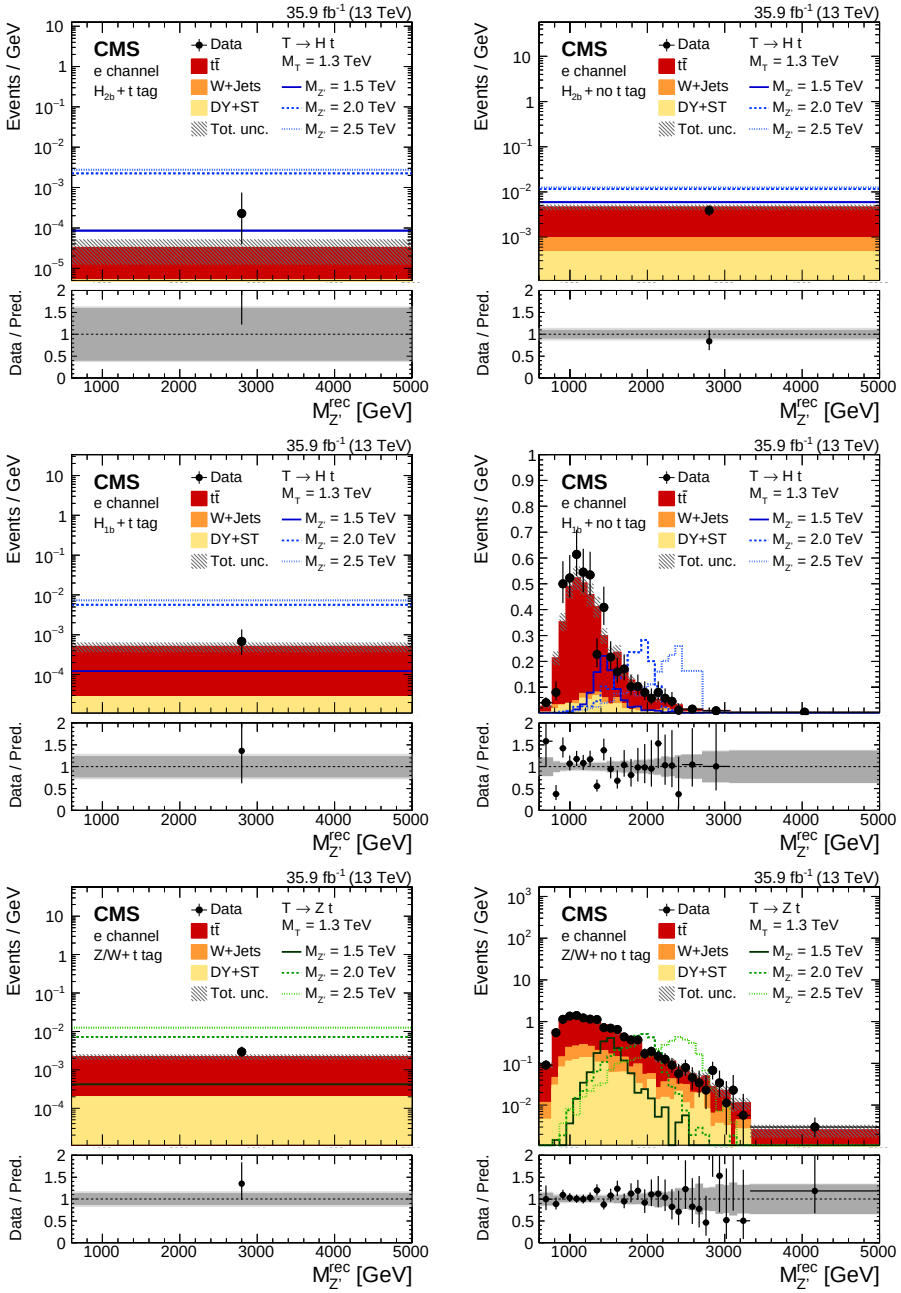


Figure 7.19: Distribution of the reconstructed Z′ resonance mass after the full selection in the e+jets channel for the data, the expected SM background, and for the signal with different Z′ masses for a fixed T mass of 1.3 TeV. In the left (right) column the results in the top tag (no top tag) category are shown. Different rows display the distributions of events accepted by different taggers as well as the signal for the respective T decays: H_{2b} tagger and $T \to Ht$ decay (upper), H_{1b} tagger and $T \to Ht$ decay (middle), and Z/W tagger and $T \to Zt$ decay (lower). The signal histograms correspond to a nominal cross section $\sigma(Z' \to tT)$ of 1 pb. The lower panel shows the ratio of data to predicted background. Here the darker grey band indicates the statistical uncertainty, whilst the lighter grey band shows the combined statistical and systematic uncertainty. Published in Ref. [1].

Table 7.3: Signal selection efficiency for the three T decay modes in each category for a signal with $M_{Z'} = 2.5\,\text{TeV}$ and $M_T = 1.3\,\text{TeV}$, taking into account branching fractions $\mathcal{B}(\text{tHt} \to \ell\text{+jets}) = 0.294$, $\mathcal{B}(\text{tZt} \to \ell\text{+jets}) = 0.317$, and $\mathcal{B}(\text{tWb} \to \ell\text{+jets}) = 0.255$ [29], where ℓ+jets is a final state with exactly one electron or muon originating from the decay of one of the top quarks, including electrons and muons from tau lepton decays. The last row of the table shows the total selection efficiency summed over all six categories. The efficiencies are shown after all selection requirements, including those on the reconstructed $t\bar{t}$ system as detailed in Ref. [1].

Category	$T \to Ht$ [%]	$T \to Zt$ [%]	$T \to Wb$ [%]
H_{2b} tag + t tag	0.35	<0.1	<0.1
H_{2b} tag + no t tag	1.7	0.15	<0.1
H_{1b} tag + t tag	0.93	0.12	<0.1
H_{1b} tag + no t tag	5.5	1.9	0.93
Z/W tag + t tag	0.33	0.15	<0.1
Z/W tag + no t tag	2.8	7.5	5.4
Sum	11.5	11.2	6.6

decay channel of the T are presented in Fig. 7.20 and specific, model-dependent limits assuming a certain $\mathcal{B}$ combination in Fig. 7.21.

Fig. 7.20 shows the upper limits on the cross section for various $(M_{Z'}, M_T)$ combinations for the decay channels $T \to Ht$ (upper left), $T \to Zt$ (upper right), and $T \to Wb$ (lower). These upper limits are set assuming a unit branching fraction for one of the decay modes of the T. Due to the tagging efficiency of the large-R jets used in the analysis, the sensitivity is better when the mass difference between the Z' and the T is large. Since this analysis is optimized for the decay modes $T \to Ht$ and $T \to Zt$, the sensitivity of these two channels is better than of $T \to Wb$. While $T \to Ht$ and $T \to Zt$ have two t quarks in the final state, $T \to Wb$ has one, resulting in the reconstruction not being successful.

The upper row of Fig. 7.21 shows upper limits on the production cross section of $Z' \to tT$ compared to the prediction of a heavy gluon G^* model [5] and a heavy vector resonance ρ^0 of the Composite Higgs model [6] at $M_T = 1.2\,\text{TeV}$ as a function of the resonance mass. The lower row shows the upper limits on the production cross section of $Z' \to tT$ compared to the prediction of the G^* model for two different values of M_T as a function of the resonance mass. The G^* can be excluded between a mass of 1.5 and 2.3 TeV if $M_T = 1.2\,\text{TeV}$, and between 2.0 and 2.5 TeV if $M_T = 1.5\,\text{TeV}$.

The G^* model in Ref. [5] predicts ten VLQs (T, B, $\tilde{T}$, $\tilde{B}$, $T_{5/3}$, $T_{2/3}$, T', B', $B_{-1/3}$, $B_{-4/3}$). While this search considers only the T in the decay chain, the other VLQs

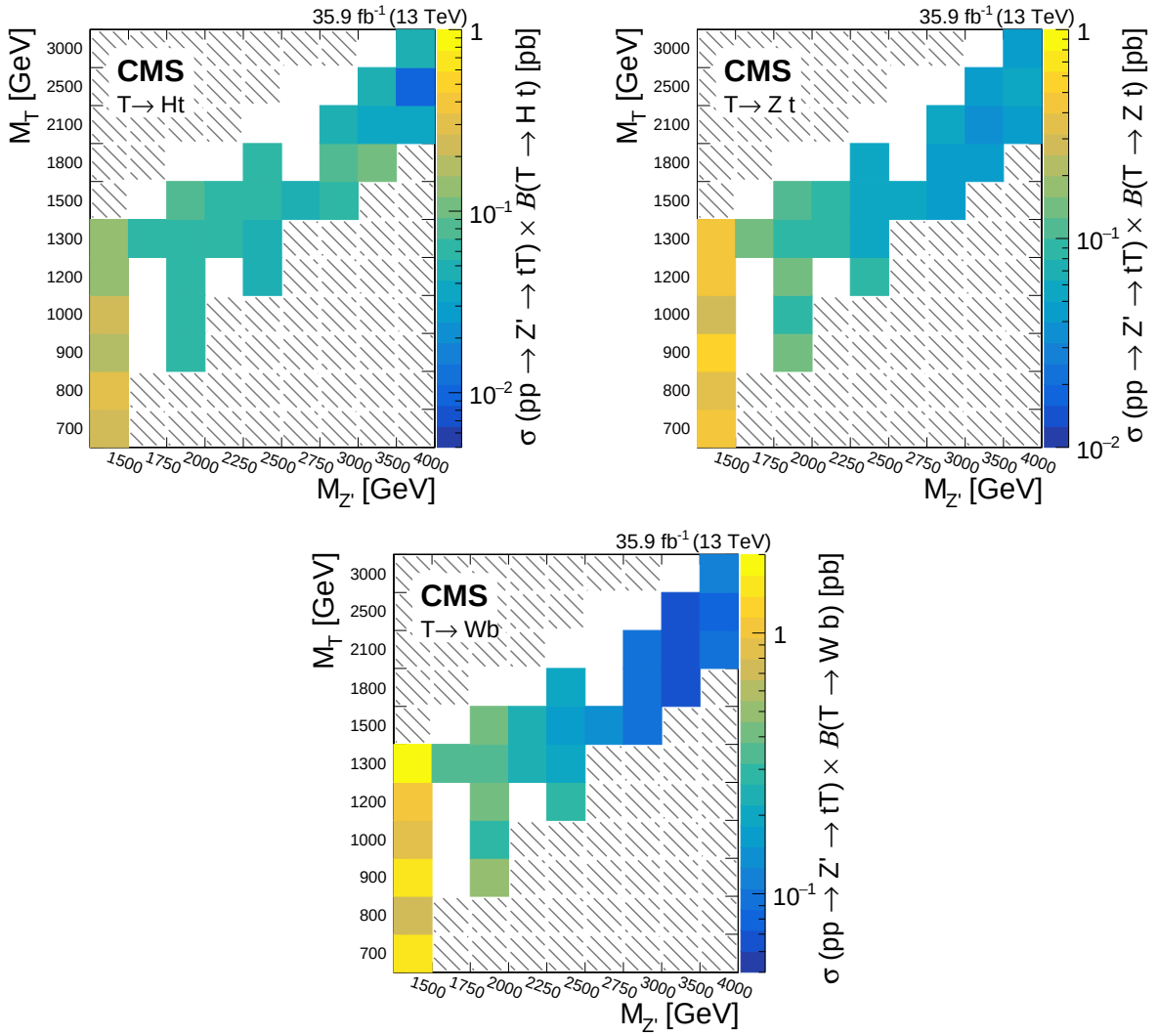


Figure 7.20: Observed exclusion limits at 95% CL on the production cross section for various $(M_{Z'}, M_T)$ combinations for the decay channels $T \to Ht$ (upper left), $T \to Zt$ (upper right), and $T \to Wb$ (lower). The hatched area in the upper left indicates the region where the $Z' \to tT$ decay is kinematically forbidden, while in the lower right $Z' \to tT$ is suppressed by the preferred $Z' \to TT$ mode. White areas indicate regions where signal samples have not been generated. Published in Ref. [1].

have an influence on the predicted signal cross section of the $Z' \to tT$. The VLQ $T_{2/3}$, with the mass $M_{T_{2/3}}$, has a strict mass correlation to the VLQ T considered in this search:

$$M_{T_{2/3}} = M_T \cdot \cos(\phi_L), \tag{7.13}$$

where ϕ_L is the mixing angle responsible for the degree of compositness of the left-handed quark doublet and is set to $\cos(\phi_L) = 0.84$ in the benchmark model considered

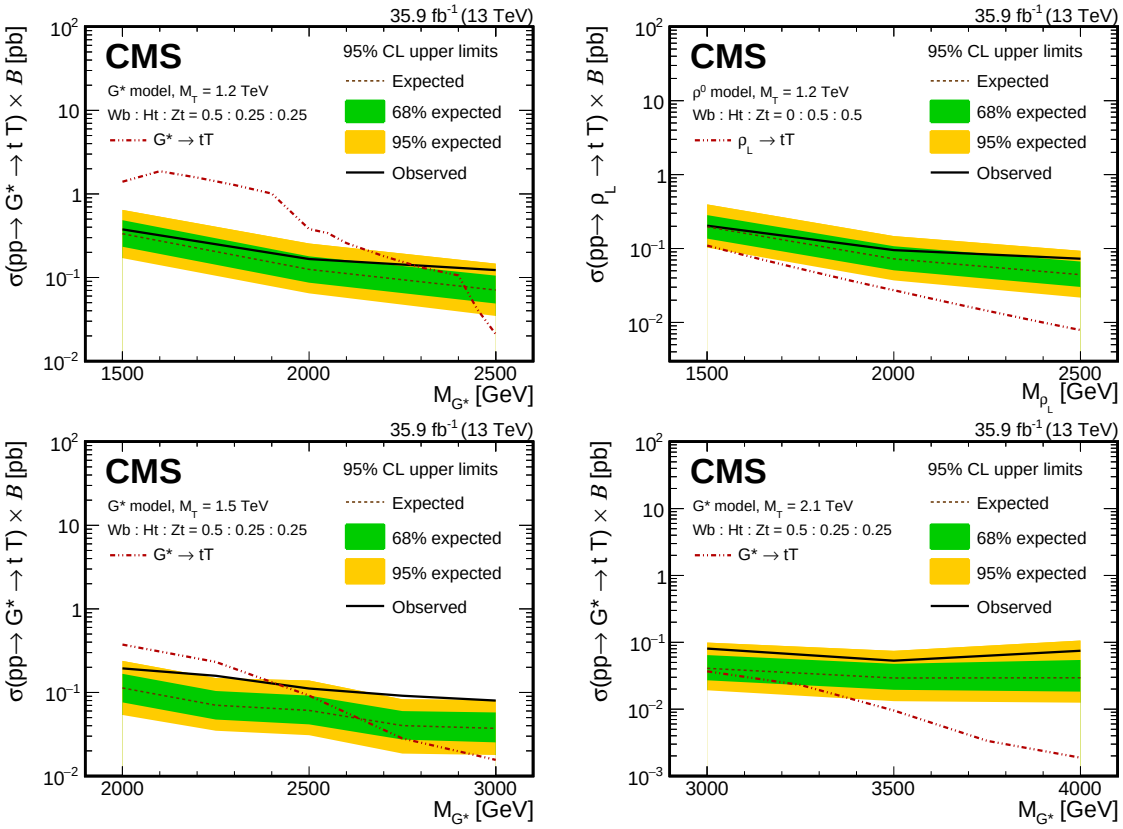


Figure 7.21: Exclusion limits at 95% CL on the product of the cross section and branching fraction for three T masses of 1.2 TeV (upper row), 1.5 TeV (lower left), and 2.1 TeV (lower right), as a function of the resonance mass. The branching fraction is defined as $\mathcal{B} = \mathcal{B}(\mathrm{T} \to \mathrm{Wb}) + \mathcal{B}(\mathrm{T} \to \mathrm{Ht}) + \mathcal{B}(\mathrm{T} \to \mathrm{Zt})$. Observed and expected limits are compared to the predictions from two different theory benchmark models: the G^* model (upper left and lower row), and the left-handed ρ_L in the ρ^0 model (upper right). Published in Ref. [1].

in this search [5]. Therefore, the two drops in the predicted signal cross section of the G^* are visible at $2 \cdot M_{T_{2/3}}$ and $2 \cdot M_\mathrm{T}$, where the new decay channels $\mathrm{Z}' \to T_{2/3}T_{2/3}$ and $\mathrm{Z}' \to \mathrm{TT}$ open up, resulting in a lower branching fraction for the decay channel under study. The most stringent cross section limits are set in the context of the ρ^0 model reaching a sensitivity already close to the theory prediction.

The exact branching fractions of the T decay dependent on the model specific parameter. In order to compare different branching fraction scenarios, Fig. 7.22 show the observed upper cross section limits for $M_{\mathrm{Z}'} = 1.5\,\mathrm{TeV}$ and $M_\mathrm{T} = 1.3\,\mathrm{TeV}$ as a function of the branching fractions $\mathcal{B}(\mathrm{T} \to \mathrm{Ht})$ and $\mathcal{B}(\mathrm{T} \to \mathrm{Zt})$. The left lower corner represents the decay $\mathrm{T} \to \mathrm{Wb}$, which has the lowest sensitivity. The right lower corner represents $\mathrm{T} \to \mathrm{Ht}$ and has the best sensitivity. In general, the higher the branching fraction into

Ht and Zt is, the higher the sensitivity.

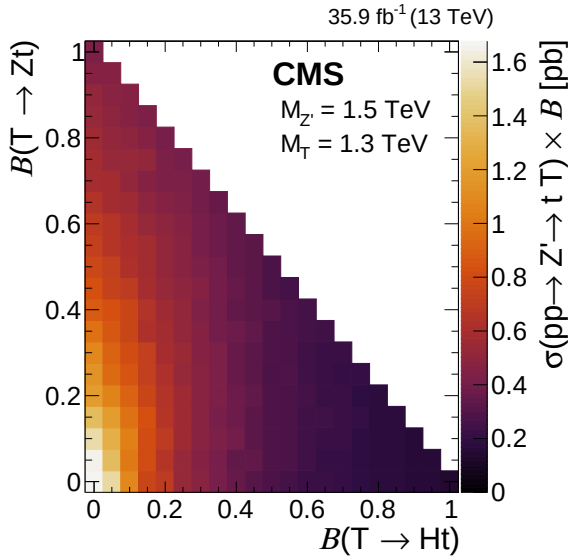


Figure 7.22: Model-independent observed exclusion limits at 95% CL on the product of the cross section and branching fraction $\mathcal{B} = \mathcal{B}(\mathrm{T} \to \mathrm{Wb}) + \mathcal{B}(\mathrm{T} \to \mathrm{Ht}) + \mathcal{B}(\mathrm{T} \to \mathrm{Zt})$ for an example mass configuration of $M_{\mathrm{Z}'} = 1.5\,\mathrm{TeV}$ and $M_{\mathrm{T}} = 1.3\,\mathrm{TeV}$ as a function of the branching fractions $\mathcal{B}(\mathrm{T} \to \mathrm{Ht})$ and $\mathcal{B}(\mathrm{T} \to \mathrm{Zt})$. Published in Ref. [1].

The sensitivity of this search is of the same order as the sensitivity of recent $\mathrm{Z}' \to \mathrm{t\bar{t}}$ searches [62–65]. In the next Sec. 7.9 the sensitivity is compared to other searches for $\mathrm{Z}' \to tT$. No dedicated search for $\mathrm{Z}' \to \mathrm{TT}$ exists. A search for $\mathrm{Z}' \to \mathrm{TT}$ would also have the disadvantage that the width of the resonance Z' is predicted to be rather large (40% of $M_{\mathrm{Z}'}$ [5]), while most searches are designed to probe a narrow resonance of 1% of $M_{\mathrm{Z}'}$ (experimental resolution of around 15%).

7.9 Comparison with Existing Results

Two other searches setting upper limits on the cross section of $\mathrm{Z}' \to \mathrm{tT}$ are available to date. The sensitivity of all three searches is compared in Fig. 7.23 for the mass range analysed in Ref. [21, 87]. The first search [87] uses a data set corresponding to an integrated luminosity of $2.6\,\mathrm{fb}^{-1}$ and is optimized for $\mathrm{T} \to \mathrm{Wb}$ in the *all-hadronic* final state. It uses a three-jet topology with one t-tag and one W tag leading to a reduced sensitivity for small $M_{\mathrm{Z}'} - M_{\mathrm{T}}$, which can be seen in Fig. 7.23 (upper, dotted lines). In addition, a small-R b-tagged jet is required. While the requirements on the W tag are similar to the ones presented before, the t tag used has an extended mass window range of $110 < M_{\mathrm{SD}} < 210\,\mathrm{GeV}$ with a looser selection on $\tau_{3/2}$.

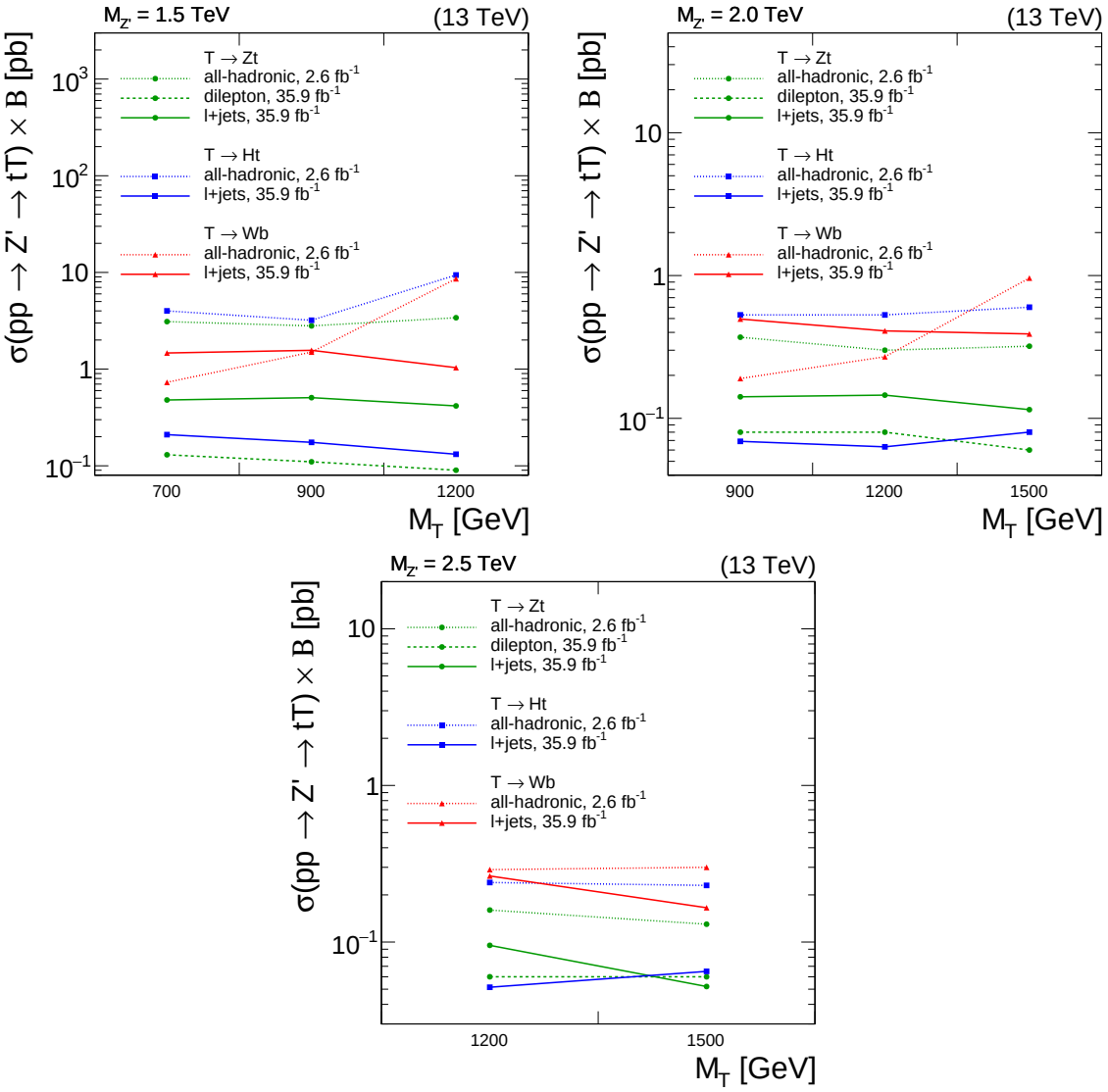


Figure 7.23: Exclusion limits at 95% CL on the product of the cross section and branching fraction for three Z′ masses of 1.5 TeV (upper left), 2.0 TeV (upper right), and 2.5 TeV (lower left), as a function of the T mass. Results are shown for the three different decay modes of the T: (green circles) T → Zt, (blue rectangles) T → Ht and (red triangles) T → Wb. Three different final states are compared: (dotted line) the all-hadronic analysis of Ref. [87], (dashed line) the dilepton analysis of Ref. [21] and (solid line) the ℓ+jets analysis of this thesis [1].

The all-hadronic search also uses the reconstructed $M_{Z'}$ as a sensitive variable and sets limits at 95% CL on all three decay modes of the T in the absence of a signal. The mass combinations studied cover a smaller range ($1.5 < M_{Z'} < 2.5$ TeV, $0.7 < M_T < 1.5$ TeV). The sensitivity is significantly reduced for T → Ht (blue dashed) and T → Zt (green dashed line) compared to T → Wb (red dashed line) as visible in Fig. 7.23.

Although not optimized, both decay modes ($Z' \to tT \to tHt$ and $Z' \to tT \to tZt$) can be reconstructed if the t tagger picks up one of the t quarks, the W tagger identifies the W from the other t quark while the b jet is used for the H/Z boson. However, this means that part of the decay is not reconstructed leading to a reduced sensitivity. It is not possible to exclude the ρ^0 model with this search but it is shown that such an analysis reaches good sensitivities. An improvement is expected when using the full data set collected from 2016 – 2018 by the CMS detector corresponding to an integrated luminosity of $137.2\,\text{fb}^{-1}$. With this integrated luminosity the statistical uncertainty as well as the upper cross section limits are expected to be reduced by a factor of $\sqrt{137.2/2.6}$, according to Poisson statistics.

The second search [21] is using an integrated luminosity of $36\,\text{fb}^{-1}$ and is optimized for $T \to Zt$ in the *dilepton* final state. Although the search primarily targets singly produced VLQs, the results are reinterpreted in the context of $Z' \to Tt$, resulting in a compatible sensitivity to the results presented in this thesis. The signature of Ref. [21] makes use of the fact that muons can be identified with high efficiency by the CMS detector and dileptonic final states in general result in a clean signature with reduced SM backgrounds. The two isolated leptons in the final state have to have an invariant mass close to the Z boson mass, while the t quark is reconstructed in three categories: resolved, merged and partially merged. The latter means that only the W of the hadronic t quark decay is identified in a large-R jet. Unlike the two previously described searches, this analysis uses the reconstructed mass of T as sensitive variable which would make it difficult to distinguish between different values of $M_{Z'}$ in the presence of a signal. However, some difference in the signal efficiency is visible between the boosted and the resolved categories.

No deviation from the SM expectation is found and upper limits on the production cross section of $Z' \to tT$ are set for $T \to Zt$ in the mass range $1.5 < M_{Z'} < 2.5\,\text{TeV}$ and $0.7 < M_T < 1.5\,\text{TeV}$. The limits set are slightly better as shown in Fig. 7.23 compared to the ℓ+jets analysis presented in this thesis. This is due to the category of partially merged t quarks. However, the models considered in this thesis require a mixing between the different decay channels of the T which makes it impossible to set mass limits on the two considered theoretical predictions, G^* and ρ^0 from the dilepton analysis only. The ℓ+jets search presented in this thesis would benefit from a combination with this result, leading to a better sensitivity in the Zt channel, improving the mass limits on the G^* and ρ^0 model.

In summary, the analysis presented here covers a larger mass range and sets the strongest upper limits on the cross section in $T \to Ht$. Furthermore, the upper limits in $T \to Wb$ were improved partially mainly by more data compared to the all-hadronic

analysis and the upper limits of the dilepton analysis [21] of $T \to Zt$ are slightly better to those of the analysis presented here. All three results where used in theory papers [8, 167] as lower bounds on new models and to make projections for future analysis.

Besides the heavy Z boson partner Z', a new heavy resonance W', the heavy partner of the W boson, appears also in theories beyond the SM. Like the W, the W' can also decay into the usual SM pairs of leptons and quarks. In theories predicting a VLQ, the decay $W' \to Tb$ is possible, which is similar to the one presented in this thesis. One search for a W' decaying into a VLQ and a quark was performed by the CMS Collaboration in the all-hadronic final state [168]. This search requires the VLQ to decay into Ht. The H and t quark are identified by large-R jets using jet substructure techniques similar to the ones used in this thesis. The b quark is identified by a small-R jet. No significant deviations from the SM prediction is found. Exclusion limits on the cross section times $\mathcal{B}$ are set, ranging from 0.01 to 0.43 pb in the W' mass rage between 1.5 to 4 TeV. It is not possible to exclude the W' model [169] studied, but it is shown that such an analysis reaches good sensitivities. An improvement is expected when using the full data set collected from 2016 – 2018 due to the reduced statistical uncertainty as explained above.

7.10 Outlook

The analysis presented in this thesis is optimised for $T \to Ht$ and $T \to Zt$ and sets the most stringent upper limits on the cross section in $Z' \to tT \to tHt$. The cross section limits on $Z' \to tT \to tWb$ were improved compared with existing one [87]. But, $T \to Wb$ shows the lowest sensitivity compared with the other decay channels in this search, because the analysis requires two reconstructed t quarks. The signature of $T \to Wb$ is two b quarks and two W bosons. The easiest way to adapt the selection is to remove the t tag and add the requirement of two b jets. With this change the W from the t quark can decay leptonically and the W from the T can decay hadronically or vice versa. Studies showed an improvement in the sensitivity by a factor of two over the whole mass range by such an extension [2].

However, the selection proposed in Ref. [2] considers the hadronically decaying t quark as partially merged. This means that the t quark that decays hadronically is reconstructed with a large-R jet and a small-R jet. This can lead to a reduced sensitivity at at high masses of the resonances where the t quark is fully merged. Two categories, one with a t tag and one with a W tag, can be used to resolve this short come.

Other improvements are possible by using more advanced taggers, e.g. based on

machine learning presented earlier in this thesis (Sec. 5.5.1). When redoing this search with the full data collected from 2016 – 2018 with the CMS experiment or during the operation of HL-LHC, the effect of pileup mitigation on e.g. the resolution of $\tau_{2/1}$ or the soft drop mass is important. Recent developments on pileup mitigation techniques are presented in Sec. 6, showing an improved performance in searches using jet substructure. The upgrade of the CMS pixel detector in 2017 (Sec. 4.4) results in an improved tracking performance especially important for b-tagging.

8 | Search for a Singly-Produced Vector-like T Quark decaying into Ht

Existing searches by the CMS and ATLAS Collaborations are excluding a VLQ T with $M_{\mathrm{T}} < 1.3\,\mathrm{TeV}$ [11–25]. However, these searches for pair and single production of VLQs take only the three decay channels ($\mathrm{T} \to \mathrm{Ht}$, $\mathrm{T} \to \mathrm{Zt}$, $\mathrm{T} \to \mathrm{Wb}$) into account. In models considering partial compositness of the new particle, a decay in a new scalar particle a is possible ($\mathrm{T} \to a\mathrm{t}$) [26]. The new scalar, light particle a can decay into a pair of SM particles. While studies of the decay $a \to \mathrm{t\bar{t}}$ are constrained by searches for supersymmetry and multitop final states [170], the decays into gg or bb are less constrained by current searches [26]. Depending on the branching ratio of the new decay and the mass of a, a vector-like T down to $M_{\mathrm{T}} \gtrsim 400\,\mathrm{GeV}$ is not excluded by current searches.

One recent search for a singly-produced T performed by the CMS Collaboration combines the resolved and boosted regime to be sensitive to masses down to 600 GeV [20]. This search was performed in the all-hadronic final state, analysing data collected by the CMS experiment corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. The analysis considers decays of the T into Ht and Zt and an excess of data compared to simulation is observed between $M_{\mathrm{T}} = 600$ and 700 GeV. No other dedicated search in the decay channel $\mathrm{T} \to \mathrm{Ht}$ for masses down to $M_{\mathrm{T}} = 600$ exist to date.

8.1 Strategy of the Search

In this thesis a search for a singly-produced vector-like T, as depicted in Fig. 8.1, decaying into Ht in the ℓ+jets final state is presented. The full Run 2 (2016–2018) data set collected by the CMS experiment, corresponding to an integrated luminosity of $137.2\,\mathrm{fb}^{-1}$, is analyzed. The search requires the t quark to decay leptonically ($\mathrm{t} \to \mathrm{Wb} \to \ell\nu\mathrm{b}$) and the H hadronically ($\mathrm{H} \to \mathrm{b\bar{b}}$). The main SM backgrounds are $\mathrm{t\bar{t}}$ and W+jets. The background is parametrized with a smoothly falling function and determined with a background plus signal fit in data to the measured spectrum. The

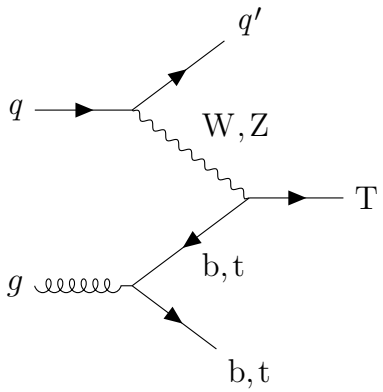


Figure 8.1: Leading order Feynman diagram for the single production of a vector-like T quark.

signal is parametrized by a Gaussian distribution. The sensitivity of this search is presented in the form of expected upper limits on the cross section times branching fraction of $T \to Ht$. This search is still blinded such that no data will be shown in the signal region. The expected sensitivity will be compared to the results from the all-hadronic analysis [20], which sees an excess.

8.2 Data and Simulated Events

This search analyses pp collision data recorded with the CMS experiment in 2016–2018, corresponding to an integrated luminosity of $137.2\,\text{fb}^{-1}$. Events targeting a ℓ+jets final state are selected with a single isolated lepton trigger. Events with a muon in the final state require the presence of an isolated muon with $p_T > 24\,\text{GeV}$ in 2016 and 2018 and $p_T > 27\,\text{GeV}$ in 2017 and $|\eta| < 2.4$. Events with an electron in the final state require the presence of an isolated electron with $p_T > 27\,\text{GeV}$ in 2016, $p_T > 35\,\text{GeV}$ in 2017 and $p_T > 32\,\text{GeV}$ in 2018 and $|\eta| < 2.5$. In order to increase the efficiency of the electron selection at high-p_T, events recorded by a single photon trigger requiring $p_T > 175\,\text{GeV}$ in 2016 and $p_T > 200\,\text{GeV}$ in 2017 and 2018 with $|\eta| < 2.5$ are used, since the trigger requirements on ECAL shower shapes are less stringent for photons than for electrons.

The signal samples are simulated with the event generator MadGraph5_aMC@NLO [47, 171, 172] at leading order. The parton shower is simulated with Pythia 8.1 [173] with the underlying event tune CUETP8M [50] in 2016 and CP5 [51] in 2017 and 2018. All signal samples represent an inclusive decay of $T \to Ht$ assuming $\mathcal{B}(T \to Ht) = 1$. Masses of the T between 600 GeV and 1.2 TeV are generated.

Simulated events for SM background processes like $t\bar{t}$, $t\bar{t}+V$, and single top (ST) are generated with the event generator Powheg [41–43, 153–158]. The showering is

done with PYTHIA 8.1 [173] with the underlying event tune CUETP8M [50] in 2016 and CP5 [51] in 2017 and 2018. Simulated events from Drell-Yan (DY) processes (W+jets, Z+jets) are generated at NLO precision with MADGRAPH5_aMC@NLO and the showering is done with PYTHIA 8.1 with the FxFx merging scheme [47]. Simulated events of diboson processes are generated at NLO precision with MADGRAPH5_aMC@NLO and POWHEG.

The detector response of the simulated samples is done with the GEANT4 package [52]. Pileup is simulated and reweighted to match the profile observed in data.

8.3 Event Selection

At least one vertex within a volume 24 cm in length and 2 cm in radius, centered on the mean pp collision point [118] is required for each event taken into account. Exactly one lepton, either a muon or an electron, is required with different offline p_T-requirements summarized in Tab. 8.1. The isolation [117, 174] of the leptons is defined as the sum of all PF particles' p_T in a cone around the lepton. The cone is defined by ΔR with a distance of 0.3 for electrons and 0.4 for muons. The isolation is corrected for the neutral contribution of pileup.

The three b quarks of the decay are reconstructed by small-R jets. The pileup mitigation technique used for these jets is CHS. No significant improvement is expected when using PUPPI since no boosted tagging is done. Each jet considered has to fulfill a set of requirements on the PF particles, reducing jets originating from noise and reconstruction failure, while retaining $98-99\%$ of the genuine jets [175]. The jet response is corrected to be unity on average. Jets can be also clustered from leptons, which are not used in this analysis. If $\Delta R(\text{lepton, jet}) < 0.4$ the jet is discarded.

The identification of jets originating from b quarks is done with the DeepJet algorithm [125]. At least three jets are required with a discriminator value > 0.309 in 2016 and 2017 and > 0.277 in 2018, which results in a misidentification rate of 1% and an efficiency around 85% [125]. The number of small-R jets that fulfil this requirement are shown in Fig. 8.2 for each year in the μ channel. While at low numbers of b-tagged jets QCD multijet simulation dominates, at high values $\bar{\text{tt}}$ is more relevant. The signal

Table 8.1: Offline p_T requirements on the lepton in both channels for each year.

Year	μ channel	e channel
2016	30 GeV	30 GeV
2017	30 GeV	40 GeV
2018	30 GeV	35 GeV

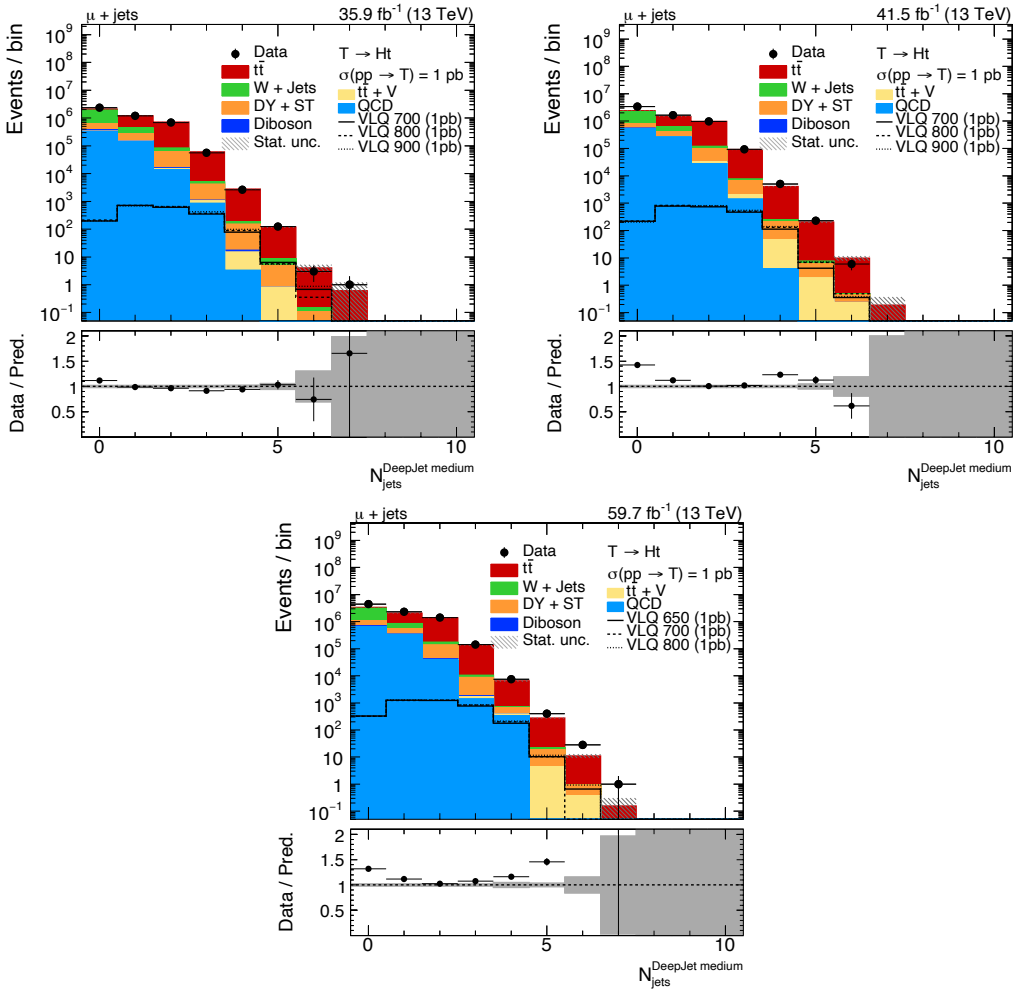


Figure 8.2: The distribution of the number of small-R jets with a DeepJet discriminator value > 0.309 in 2016 and 2017 and > 0.277 in 2018 in data and simulation for 2016 (upper left), 2017 (upper right) and 2018 (lower) in the μ channel.

simulation shows a significant fraction of events with three or more b-tags, while there is nearly no difference between the different mass points shown. After the requirement of at least three b-tags, QCD multijet simulation is around 2% of the total SM background and further reduced when reconstructing the signal. Therefore QCD multijet simulation will not be shown in the following plots.

During 2016 and 2017 data taking the timing of the ECAL in $2 < |\eta| < 3$ was not properly adjusted to the trigger. The timing ensures a correct matching of the energy recorded with the associated bunch crossing. In this case the energy is associated to the previous bunch crossing. If the energy deposit in the ECAL in $2 < |\eta| < 3$ is sufficiently large that it would fire the trigger, the actual event gets vetoed since the recording of two events in a row are forbidden by trigger rules. Since this effect is not modelled

in simulation, weights are applied to simulation to account for the difference in data. The weight is calculated out of the probability for each photon or jet in the event to cause the issue. The effect results in a loss in signal efficiency of less than 5% in 2016 and 2017, independent of M_{T}.

In 2018 two HCAL modules stopped working resulting in a loss of the energy measurement in $-3.2 < \eta < -1.3$ and $-1.57 < \phi < -0.87$. This effect is not modelled in simulation. Two scenarios, that have an impact on this analysis, can happen. The first case is the reduced energy measured of a jet in this region. The second case are misidentified electrons, since these are identified by an ECAL energy deposit without an HCAL energy deposit. In order to account for these two effects, events, where one of the b-tagged jets or the electron is in the affected region, are rejected. The effect on signal efficiency is about 3%, independent of M_{T}.

The distribution of the $\mathrm{p_T}$ of the electron and the muon for all three years after the selection described can be found in Fig. 8.3. The simulation can describe the data in general. A $\sim$ 10% difference in 2016 is seen, coming from the known effects of the $\mathrm{t\bar{t}}$ simulation as described in Sec. 7.6. In 2017 and 2018, where the $\mathrm{t\bar{t}}$ simulation has a different underlying event tune, this effect is not visible. The signal shows a higher $\mathrm{p_T}$ for higher M_{T} masses.

Figure 8.4 shows the $\mathrm{p_T}$ distribution of the small-R jets in the event in the μ channel (left) and the e channel (right) for all three years. The same effect in 2016 is visible, while in 2017 and 2018 the data are well described by simulation. The signal simulation shows in general a higher $\mathrm{p_T}$ than the SM background, where also here a dependency on the M_{T} is visible.

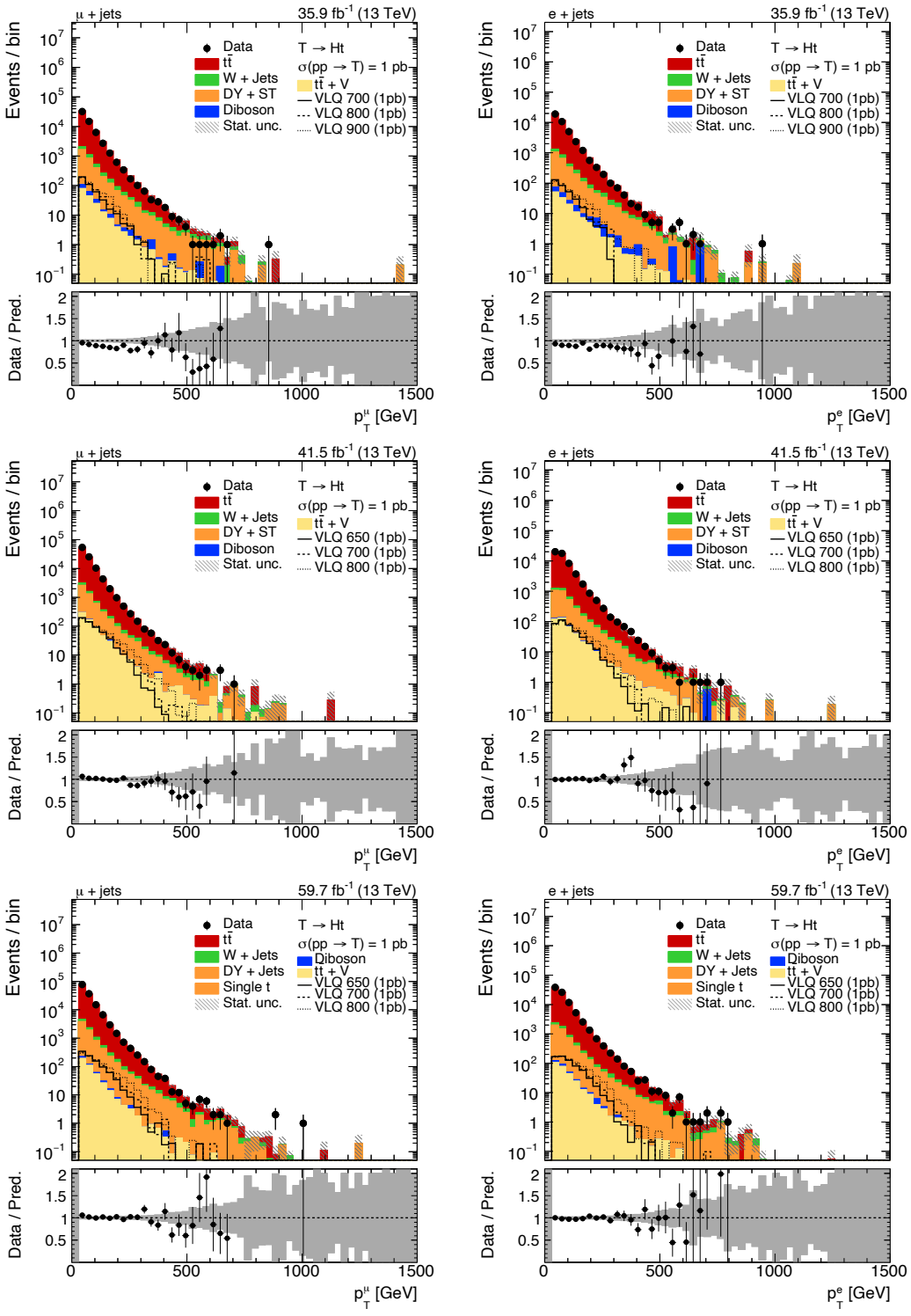


Figure 8.3: The distribution of the p_T of the muon in the μ channel (left) and the p_T of the electron in the e channel (right) for the three different years 2016 (upper), 2017 (middle) and 2018 (lower).

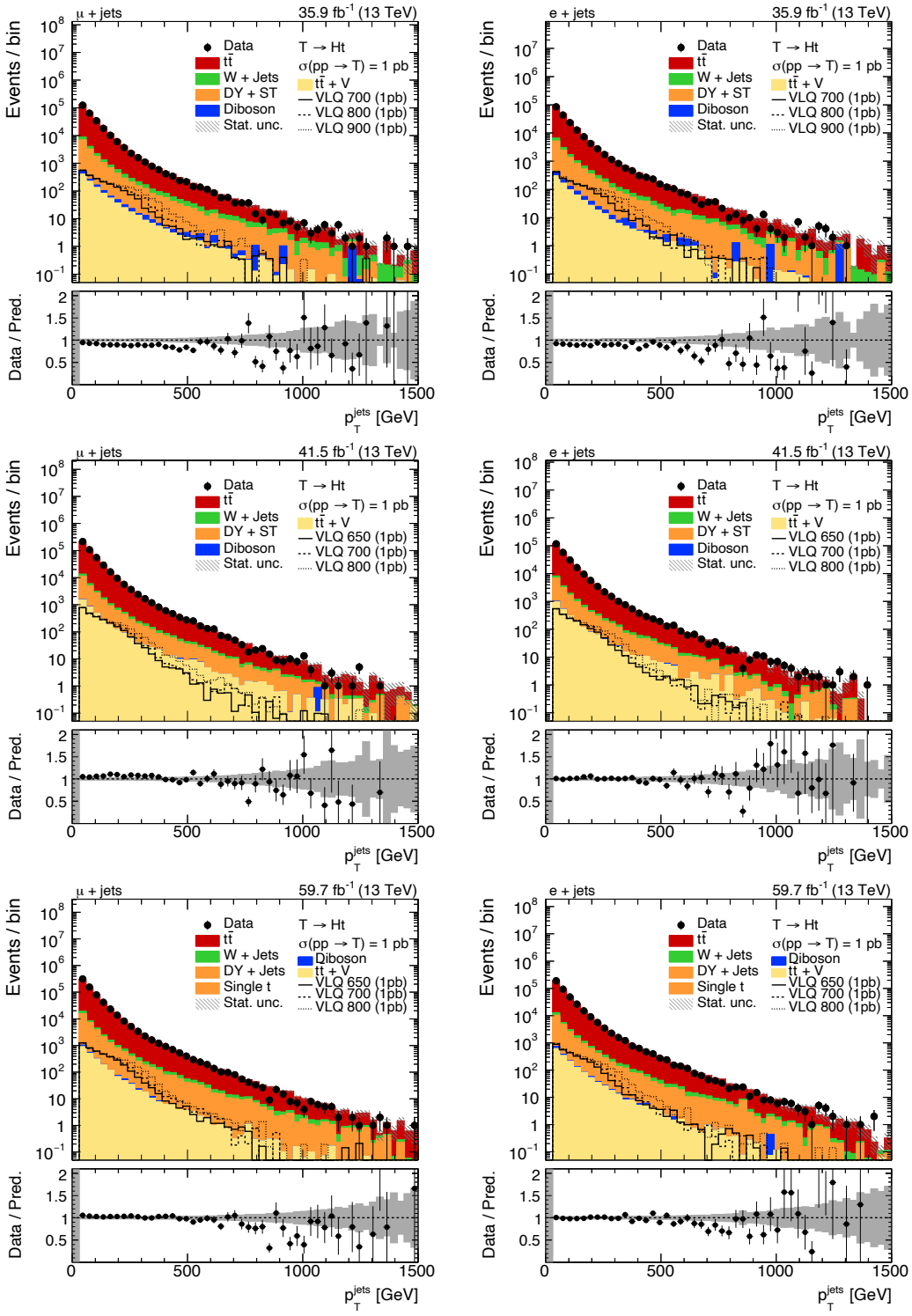


Figure 8.4: The distribution of the p_T of the small-R jets in the μ channel (left) and in the e channel (right) for the three different years 2016 (upper), 2017 (middle) and 2018 (lower).

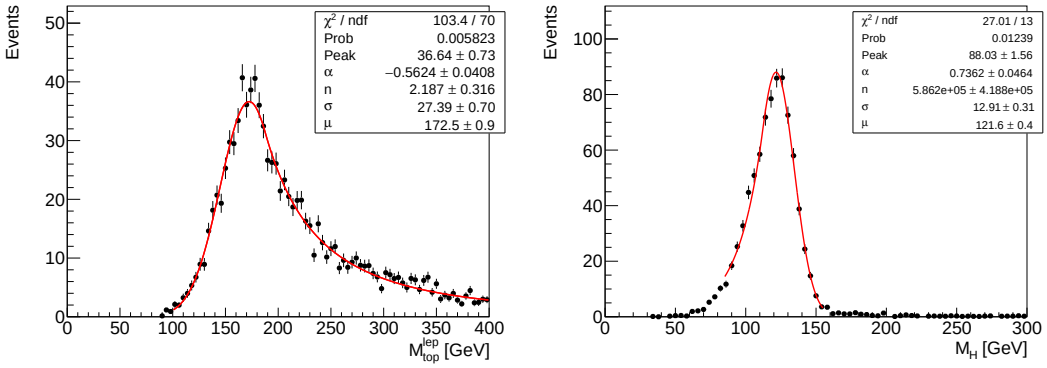


Figure 8.5: The reconstructed top mass (left) and H mass (right) in 2017. The fit is performed with a Crystal Ball function.

8.4 Reconstruction of the T Mass

The general idea of reconstructing the T mass is the same as already described in Sec. 7.5. The decay products of the T are reconstructed and a T candidate for each combination of the hadronic H candidates and the leptonic t candidates is stored. The best T candidate is chosen by a χ^2. The neutrino reconstruction of the leptonic t candidate is the same as in Sec. 7.5.

8.4.1 Reconstruction of the Top Quark and the Higgs Candidate

The t quark candidate consist of the lepton, the reconstructed neutrino and one b jet. The H boson candidate consists out of two b jets. The b-tagged jets are assigned either to the leptonic t candidate, the hadronic H candidate or none. All possible combinations are stored.

A χ^2 is calculated for each combination:

$$\chi^2 = \frac{(M_{\mathrm{t}}^{\mathrm{gen}} - M_{\mathrm{t}}^{\mathrm{rec}})^2}{\sigma_{\mathrm{t}}^2} + \frac{(M_{\mathrm{H}}^{\mathrm{gen}} - M_{\mathrm{H}}^{\mathrm{rec}})^2}{\sigma_{\mathrm{H}}^2}. \tag{8.1}$$

The reconstructed mass $M_{\mathrm{t/H}}^{\mathrm{rec}}$ of the t/H is obtained by the invariant mass of all objects of the candidate. The expected mass $M_{\mathrm{t/H}}^{\mathrm{gen}}$ and the expected width $\sigma_{\mathrm{t/H}}^2$ are calculated from the mass distribution of events, where each object of the candidate matches the generator particles with $\Delta R < 0.4$. Examples of these fits for 2017 can be seen in Fig. 8.5. The reconstructed mass distributions are fitted with a Crystal Ball function. The mean and width obtained from the fits in each year are shown in Tab. 8.2.

year	$\overline{M_{top}}$ [GeV]	σ_{top} [GeV]	$\overline{M_H}$ [GeV]	σ_H [GeV]
2016	177	30	122.7	13
2017	172.5	27	121.6	13
2018	171.4	28	120.4	13

Table 8.2: Values of the expected t mass and width as well as the expected H mass and width for all three years. The values are obtained from a fit to the reconstruct t mass and H mass, where all assigned jets match within $\Delta R <$ 0.4 the generator-level quarks of the decays.

For each event the combination with the lowest χ^2 value is chosen. All three χ^2 terms, the total, χ^2_{t} and χ^2_{H}, are shown in Fig. 8.6 for all three years combined in the μ channel. The signal simulation tend to have low values of χ^2, which means that the reconstruction is assigning the correct objects to the candidates. The χ^2 distribution is used to split the events into signal events (SR) and control region events (CR). Events with a $\chi^2 < 10$ and a $\chi^2_{\mathrm{H}} < 2$ are called SR events. All other events are called CR events.

8.4.2 Reconstruction of the T Candidate

The resulting distributions of the reconstructed T mass can be seen in Fig. 8.7 for the SR (upper row) and for the CR (lower row). The distribution has a turn on at low M_{T} and a smoothly falling background, which is important for the background estimation. Further requirements on the ΔR of the different objects are tested, leading to a negligible improvement in the sensitivity but a shift of the background turn on to higher masses. Therefore, no requirements on ΔR are used.

8.5 Background Estimation

The background is estimated from data by a background parametrisation to be as independent from simulation as possible. However, the simulation can describe the data very well as visible in Fig. 8.7 in the CR. It is planned to use the background estimation from simulation as a cross check on the final results, once unblinded.

As shown in Fig. 8.7 the SM background is smoothly falling, which makes it possible to parametrize it with an exponential function with two free parameters:

$$e_2(x) = e^{-p_0 x + p_1 x^2}, \tag{8.2}$$

where $x = M_{\mathrm{T}}/(1\,\mathrm{TeV})$ and p_i are the free parameters. This background function is

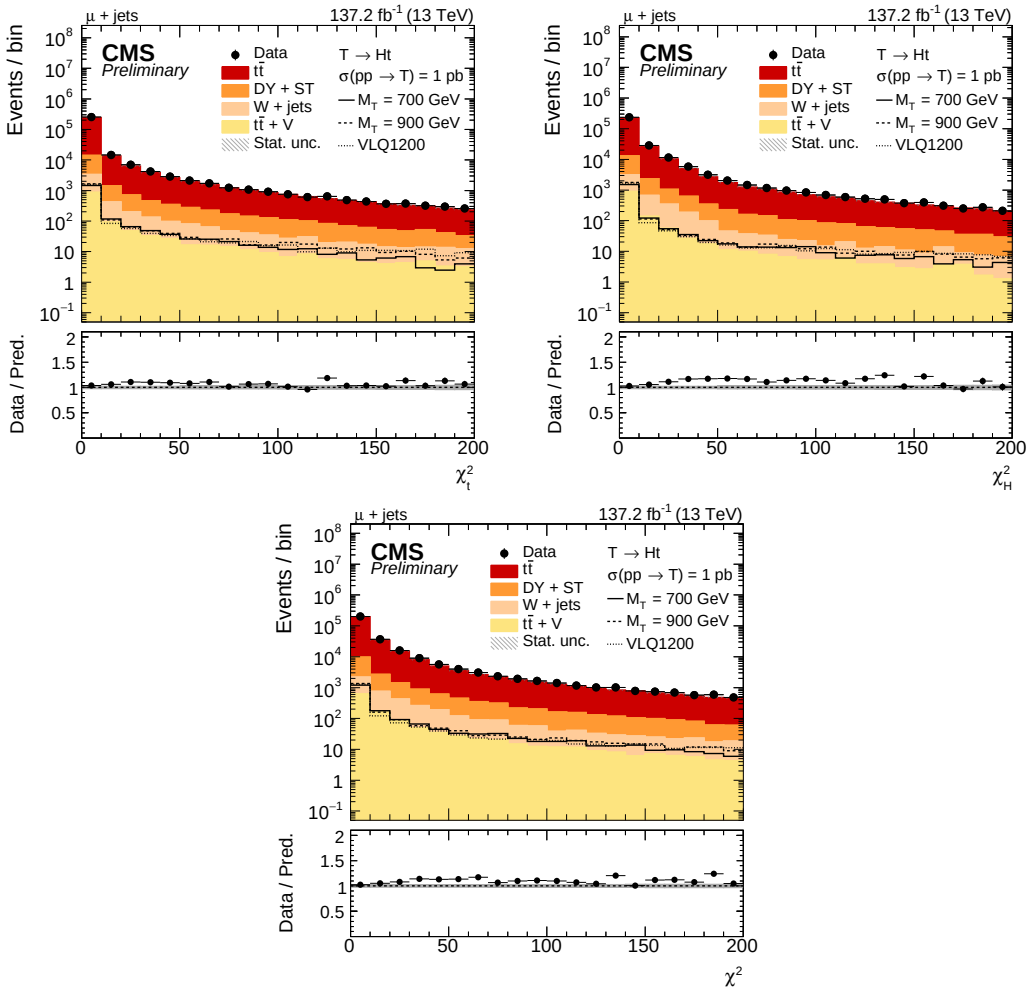


Figure 8.6: The distribution of the χ^2_{t} (upper left) and χ^2_{H} (upper right) and the total χ^2 (lower) in data and simulation in the μ channel for the combination of all three years.

validated on data in the control region. As a reference functions that have been previously used in dijet searches [72, 96, 176–180] with two, three or four free parameters are tested:

$$
\begin{aligned}
d_2(x) &= \frac{(5-x)^{p_0}}{x^{p_1}} \\
d_3(x) &= \frac{(5-x)^{p_0}}{x^{p_1+p_2\log(x)}} \\
d_4(x) &= \frac{(5-x)^{p_0}}{x^{p_1+p_2\log(x)+p_3\log(x)^2}},
\end{aligned}
\tag{8.3}
$$

where $x = M_{\mathrm{T}}/(1\,\mathrm{TeV})$ and p_i are the free parameters. The functions start 2.5 signal resonance widths below the lowest mass point in the signal region.

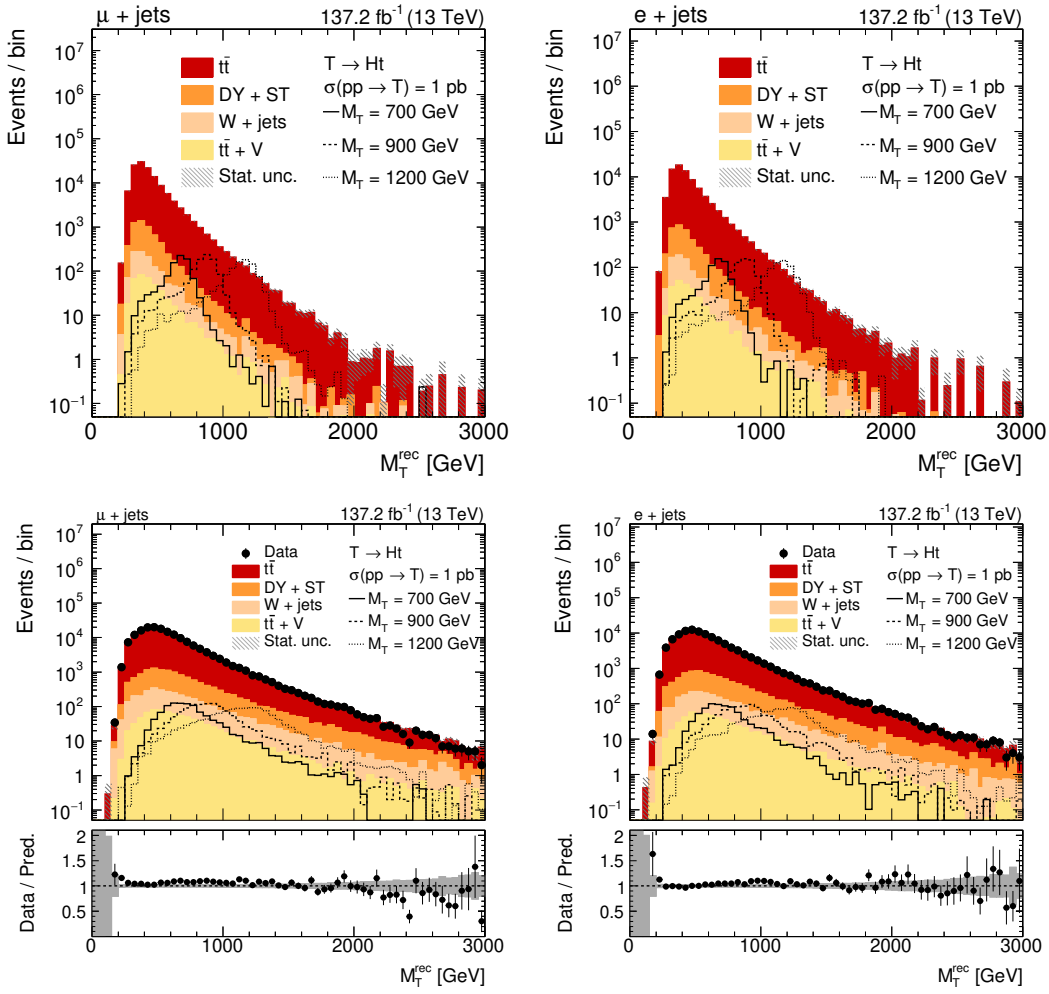


Figure 8.7: Distribution of the reconstructed T mass for the signal region (upper row) and the control region (lower row). Events are shown in the the μ channel (left) and the e channel (right) with contributions from data, simulated signal samples, and the simulated SM backgrounds. The expected signal distribution from various T masses is shown with a nominal cross section $\sigma(\mathrm{pp} \to \mathrm{T} \to \mathrm{Ht})$ of 1 pb. The lower panel shows the ratio of data to predicted background. Here the darker grey band indicates the statistical uncertainty.

8.5.1 Validation in the Control Region

The reconstructed T mass distribution in the CR has similar properties as the distribution in the SR, as shown in Fig. 8.7 and is therefore suited to validate the background parametrisation. All distributions are dominated by $\mathrm{t\bar{t}}$ and have the turn-on at low values of M_T. In addition, all backgrounds are smoothly falling necessary for the background parametrisation. The fit of three functions to the data in the CR can be seen in Fig. 8.8. All three parametrisations show a reasonable agreement to the data

Table 8.3: The results of the F-Test in the control region on data for the $e_2(x)$, the $d_3(x)$ and the $d_4(x)$ function. The function with more parameters is excluded if the CL (third column) is above 0.05.

Channel	Functions under test	CL to exclude model with more parameters
e+jets	$e_2(x)$ vs $d_3(x)$	0.085
e+jets	$d_3(x)$ vs $d_4(x)$	0.079
e+jets	$e_2(x)$ vs $d_4(x)$	0.049
μ+jets	$e_2(x)$ vs $d_3(x)$	0.295
μ+jets	$d_3(x)$ vs $d_4(x)$	0.567
μ+jets	$e_2(x)$ vs $d_4(x)$	0.493

in the CR. The $d_3(x)$ and $d_4(x)$ functions show a better flexibility to model the low mass region in the muon channel and the tail of the distribution. However, due to the flexibility these are more likely to absorb signal contributions in the background parametrisation.

8.5.2 F-Test

An F-test [181] is performed on data in the CR, to determine the necessary number of free parameters. The results of this test are presented in Tab. 8.3. The $e_2(x)$ and the $d_3(x)$ function are preferred against the $d_4(x)$ function. The $e_2(x)$ function is also preferred against the $d_3(x)$ function in both channels. Therefore, in the following the $e_2(x)$ function is used to parametrize the background, while the $d_3(x)$ function will be used as an alternative shape.

8.5.3 Goodness of Fit

In order to have a measure that indicates how well the function can describe the data, a Goodness-of-Fit test is performed in the CR on data. It shows the observed discrepancy in data in the context of the expected discrepancy measured in toys. The test is performed in each year in the μ and the e channel. The results are shown in Fig. 8.9. The red arrow represents the observed discrepancy from the data in the CR. The blue line shows the expected discrepancy from toys thrown around the background parametrisation. All observed discrepancies are within the expected ones, while the distribution in 2016 in the μ channel shows the lowest probability. The background parametrisation can describe the data in the control region. This test will be redone once the SR is unblinded.

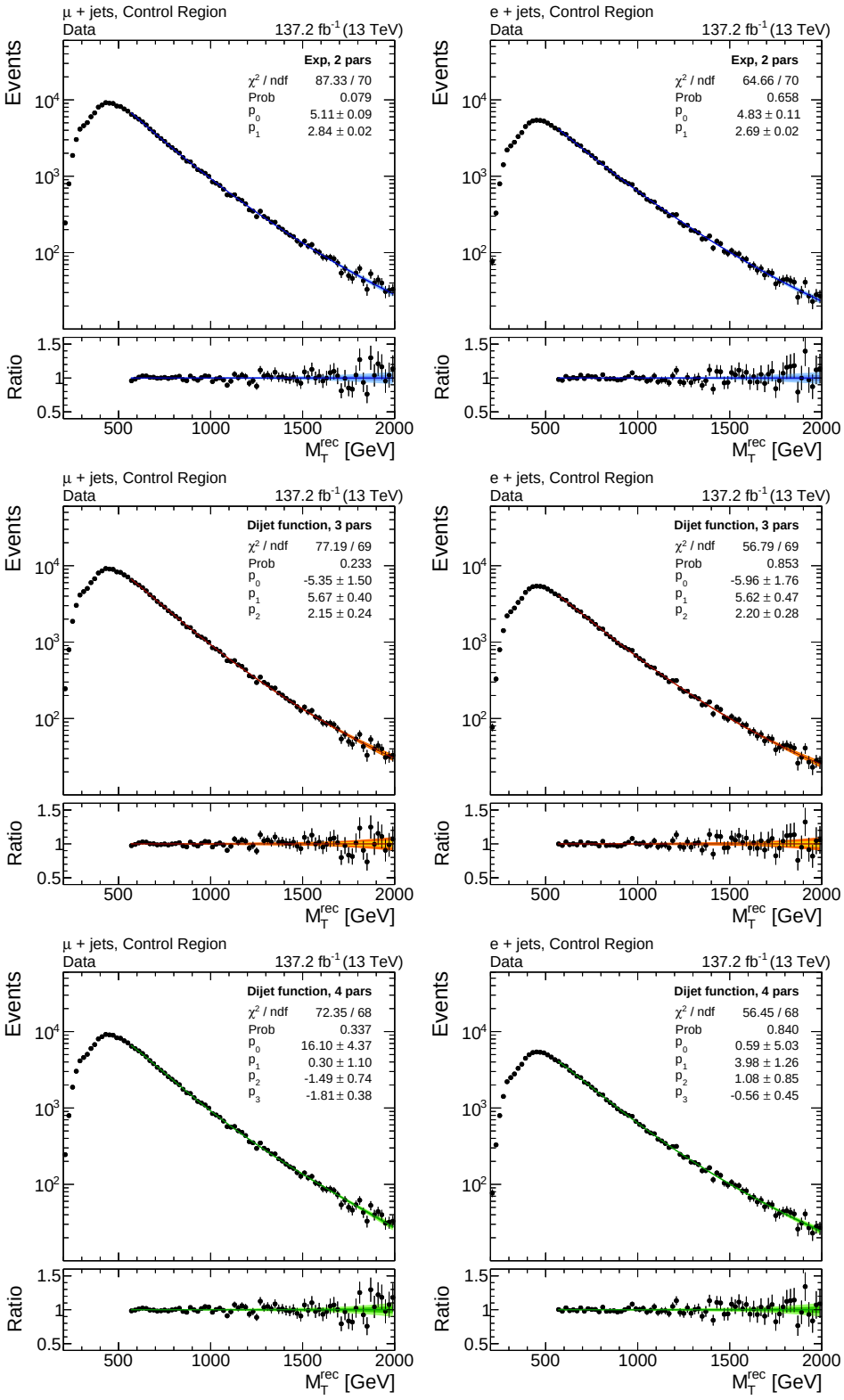


Figure 8.8: Distribution of the reconstructed T mass in the control region in the μ channel (left) and the e channel (right). The black dots represent the Run 2 data, whereas the blue/orange/green line shows the fit of the $e_2(x)$ (upper), the $d_3(x)$ (middle) and the $d_4(x)$ function (lower). The lower panel shows ratio between the data points and the fit with the one σ (dark blue/yellow/dark green) and two σ (light blue/orange/light green) band of the fit.

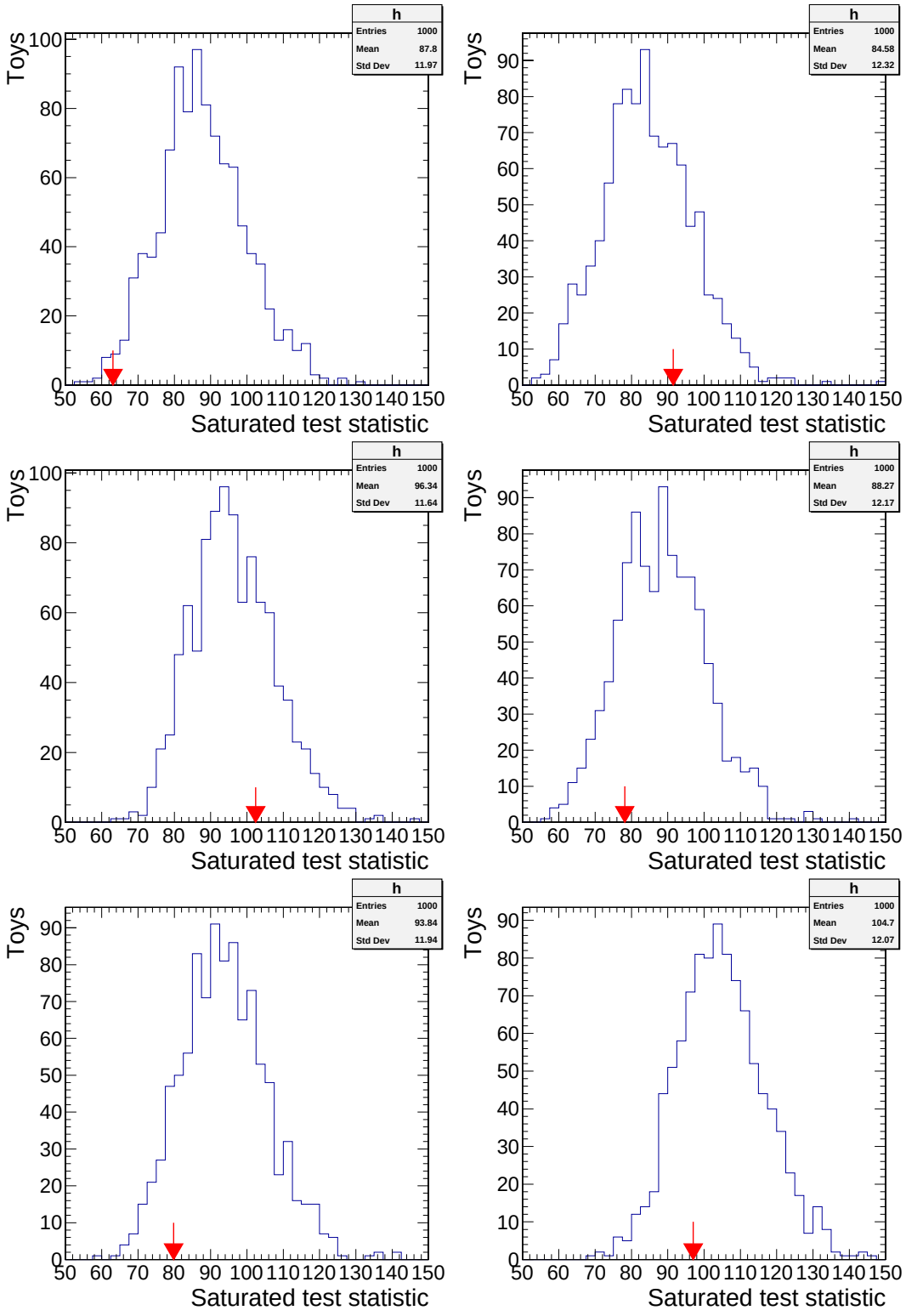


Figure 8.9: Goodness-of-fit test for data in the CR: (upper row) 2016, (middle row) 2017, (lower row) 2018. The left column shows the muon channel, the right column shows the electron channel. The red arrow shows the observed discrepancy while the blue line shows the expected discrepancy.

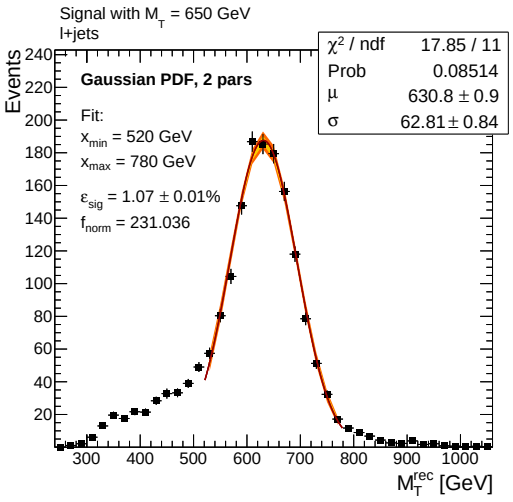


Figure 8.10: Distribution of the reconstructed T mass in the signal region in the combined ℓ+jets channel for signal simulation. The black dots represent the Run 2 simulation, whereas the red line shows the fit of a Gaussian function. The yellow band shows the one σ band of the fit and the orange band the two σ band of the fit.

8.6 Signal Parametrisation

The signal is parametrized with a Gaussian function as shown in Fig. 8.10 for an example mass of 650 GeV. The reconstructed signal mass peaks at the expected value with a width of around 10% of the mass. The low mass tail of the distribution originates from wrongly assigned b jets. With the Gaussian parametrisation a loss of around 15% in the signal events is expected due to the low mass tails which are not considered.

The mean and the width is extracted for each simulated mass point and a linear function is used to predict the signal shape for any mass. The resulting fits for the mean and the width in the ℓ+jets channel for the Run 2 simulation can be seen in Fig. 8.11. The reconstructed T mass is around 5% lower than the generator-level T mass and the width is around 10-15% of the generator-level T mass over the whole mass range.

The signal efficiency is shown in Fig. 8.12 for the three years for the combined ℓ+jets channel. The signal efficiency is defined as the number of events after full selection divided by the number of events before any selection. While the efficiency first rises with increasing M_{T}, due to the efficiency of the $\mathrm{p_T}$ requirements, at around $M_{\mathrm{T}} \approx 1000\,\mathrm{GeV}$ the efficiency drops again. This is due to the fact, that the H and t are getting boosted and their decay products start to get a lower angular separation, which is not covered in this search. The efficiency is similar between the years, which

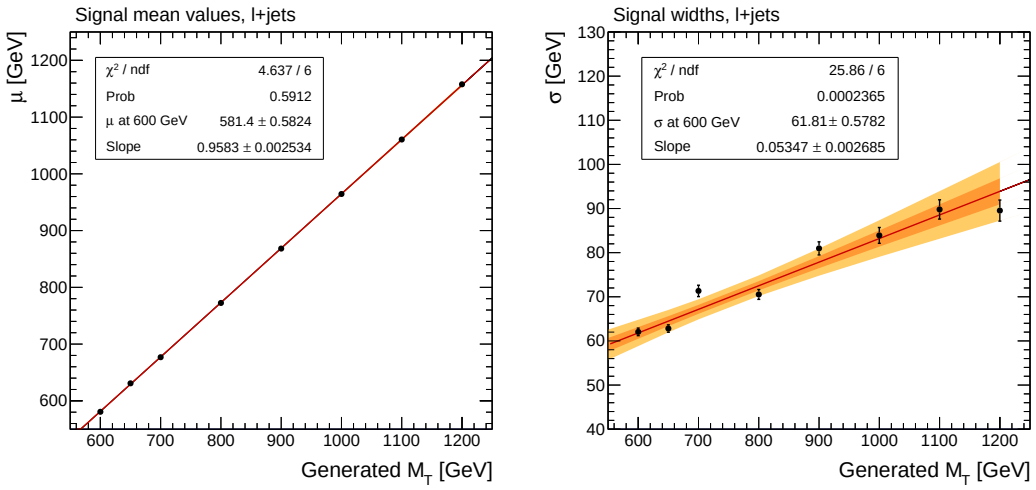


Figure 8.11: The mean (left) and width (right) of reconstructed T mass in the signal region in the combined ℓ+jets channel as a function of the generator-level T mass. The black dots represent the Run 2 simulation, whereas the red line shows the fit of a linear function. The yellow band shows the one σ band of the fit and the orange band the two σ band of the fit.

is expected due to similar trigger and detector conditions for the different years. The large systematic uncertainty in 2017 (Fig. 8.12, upper right, yellow band) results from a large b-tagging uncertainty in this year.

8.6.1 Bias test

A bias test is performed to understand the effect of the choice of the parametrisation. Pseudo data are generated and afterwards fitted with the $e_2(x)$ or the $d_3(x)$ function. One bias test with no signal injected in the toy generation ($r = 0$, where r is the signal strength) and one bias test with a signal of 1 pb injected in the generation ($r = 1$) is performed.

In Fig. 8.13 (left) the resulting signal strength divided by its error for $r = 0$ is shown. Since no signal is injected in the generation, the distribution is expected to peak around zero. The $e_2(x)$ function results in an average $\langle r/\sigma_r \rangle = 0.25$ for the full Run 2 data. While the error of r is around 150 fb, a small bias of 0.25 means that the fitted signal cross section is ~ 30 fb higher than injected. However the expected sensitivity is around 300 pb for this signal mass point. When using the $d_3(x)$ function the bias is reduced to 0.16. The bias is similar small when injecting a signal in the toy generation (Fig. 8.13 (right)).

In order to understand the origin of this bias, the M_{T} value at which our background parametrisation starts is reduced. In Fig. 8.13 the fit started at 450 GeV. When testing

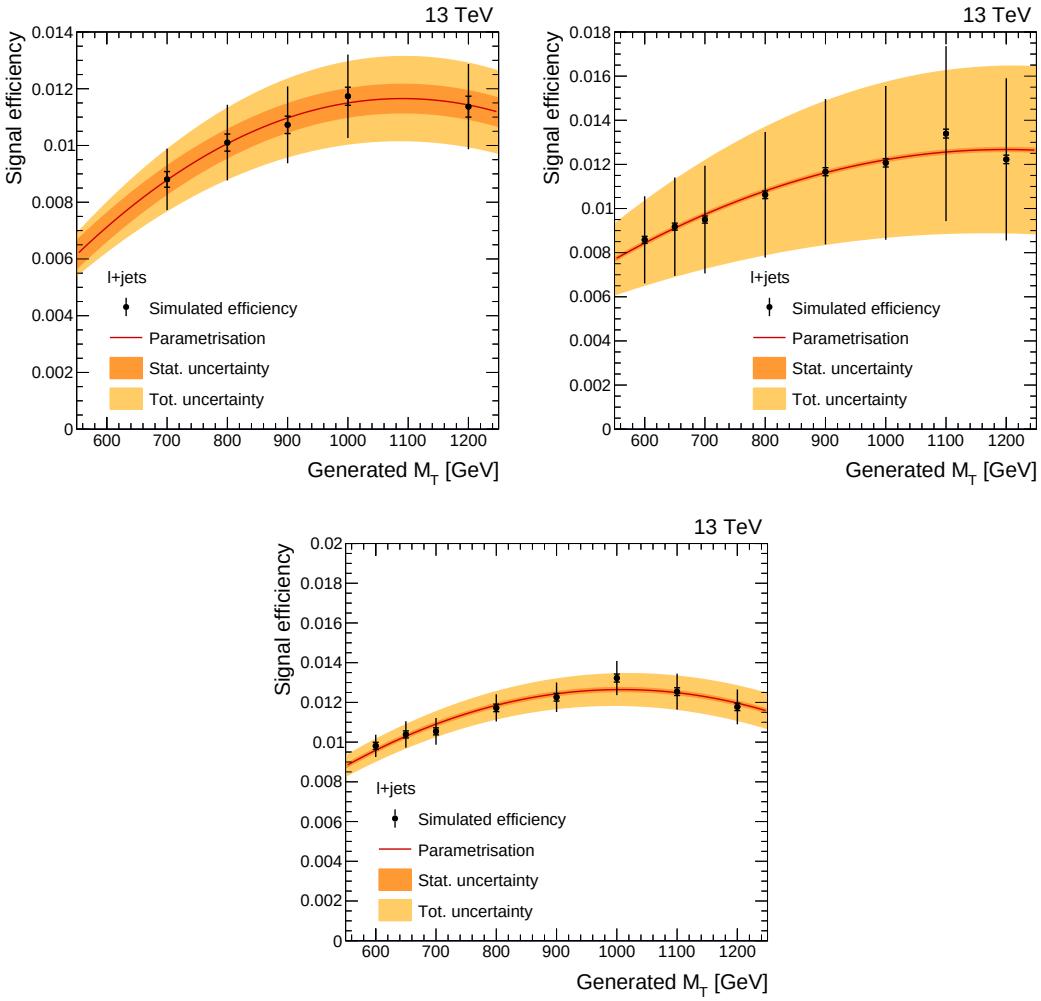


Figure 8.12: Efficiency of the final selection for 2016 (left), 2017 (right) and 2018 (bottom). The inner error bars and dark bands show statistical uncertainties. The outer error bars and light bands show the total uncertainties.

the mass point of $M_T = 600\,\text{GeV}$ assuming a reconstructed mean of 580 GeV and a width of the signal distribution of around 60 GeV (values from Fig. 8.11) starting the fit at $M_T = 450\,\text{GeV}$ is only ~ 2 standard deviations away from the peak position of the Gaussian. It might be possible that the bias results from a start value of the fit too close to the peak of the first signal mass point. Therefore, a bias test with a fit starting value of $M_T = 400\,\text{GeV}$ was performed. The comparison between the bias of the $e_2(x)$ for the two starting values of the fit is shown in Fig. 8.14. Reducing the starting value of the fit results in negligible amount of bias for both $r = 0$ (left) and $r = 1$ (right). However, from Fig. 8.7 it is known that the peak position of the background is around $M_T \sim 400\,\text{GeV}$, making it impossible to start the fit at $M_T = 400\,\text{GeV}$ without modelling the turn on.

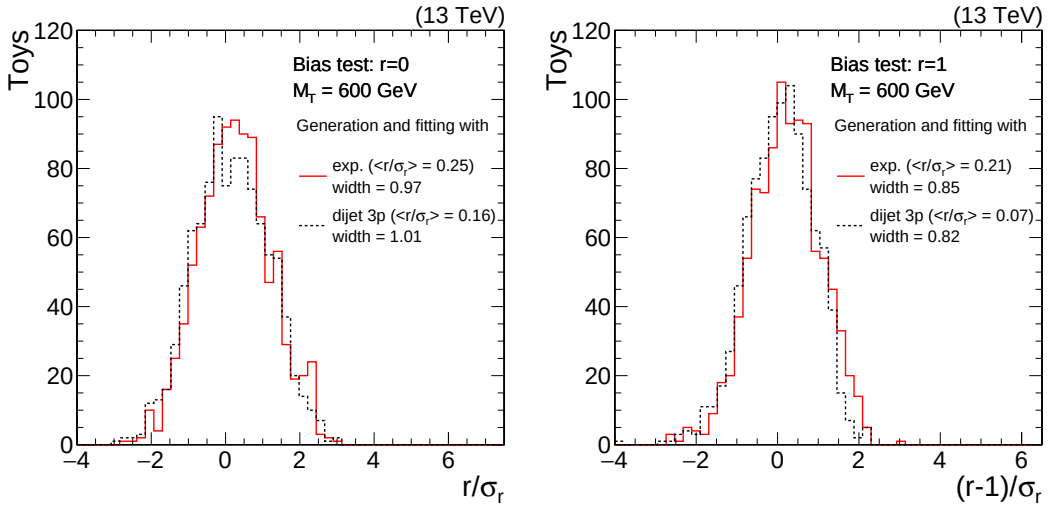


Figure 8.13: The bias test for three different background representations. Pseudo data without signal are generated according to the legend with the $e_2(x)$, the $d_2(x)$ or the $d_3(x)$ function. The fit is performed with the $e_2(x)$ function. The mean of the resulting signal strength (r) divided by its error (σ_r) is depicted in the legend.

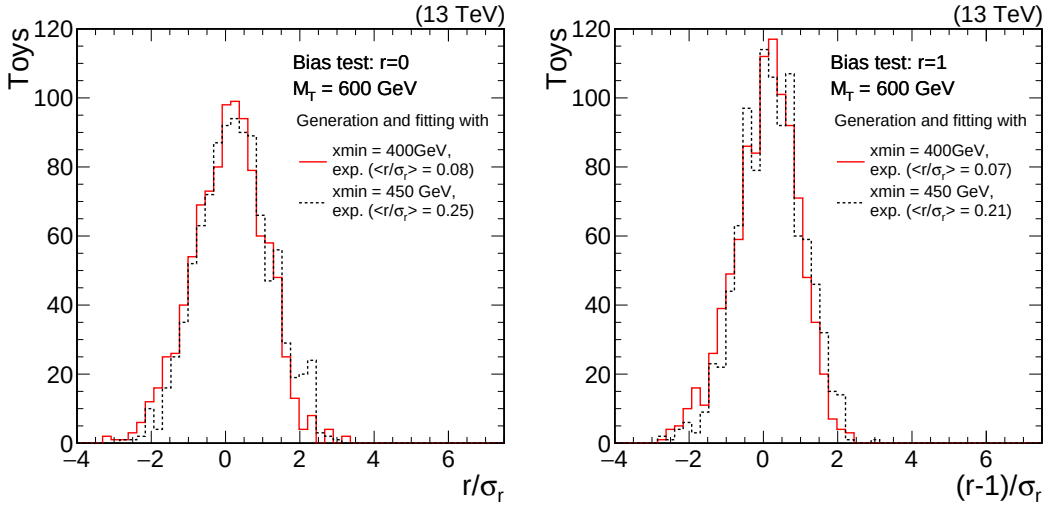


Figure 8.14: The bias test for the two different background representations at $M_{\mathrm{T}} = 600\,\mathrm{GeV}$. Pseudo data without signal are generated and fitted with the $e_2(x)$ function. Two different starting values of the fit ($M_{\mathrm{T}} = 400\,\mathrm{GeV}$ and $450\,\mathrm{GeV}$) are used. The mean of the resulting signal strength (r) divided by its error (σ_r) is depicted in the legend.

In order to confirm that the bias results from a starting value of the fit to close to the peak position of the signal simulation, another signal mass point with $M_{\mathrm{T}} = 700\,\mathrm{GeV}$ was tested. The bias for this signal mass point is shown in Fig. 8.15 for a starting value of the fit of $M_{\mathrm{T}} = 450\,\mathrm{GeV}$. It is seen that the $e_2(x)$ function results in a negligible

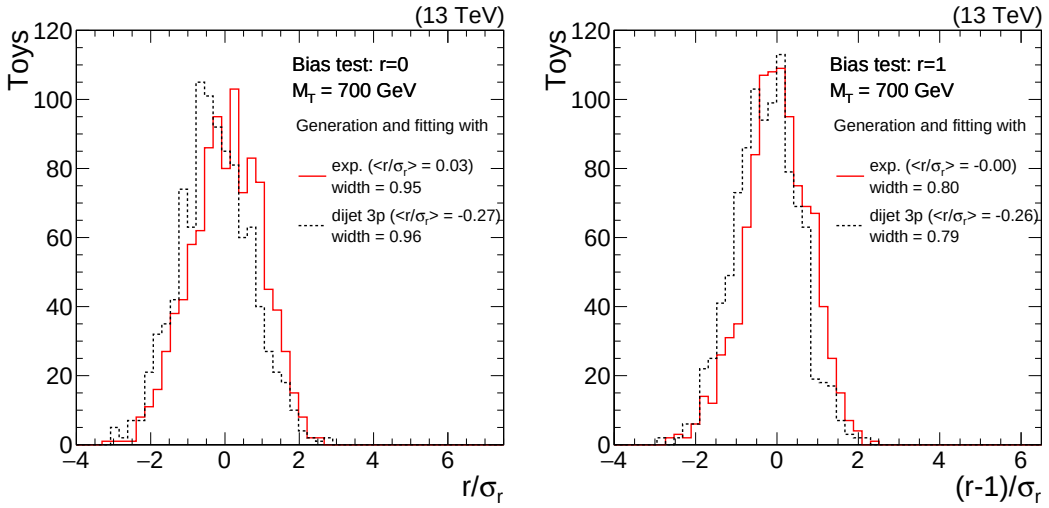


Figure 8.15: The bias test for the two different background representations at $M_{\mathrm{T}} = 700\,\mathrm{GeV}$. Pseudo data without signal are generated and fitted according to the legend with the $e_2(x)$ or the $d_3(x)$ function. The mean of the resulting signal strength (r) divided by its error (σ_r) is depicted in the legend.

amount of bias for both $r = 0$ (left) and $r = 1$ (right). However, the $d_3(x)$ function results in a bias of $\langle r/\sigma_r \rangle \sim -0.26$. In conclusion, the bias seen at $M_{\mathrm{T}} = 600\,\mathrm{GeV}$ most likely comes from the starting value of the background fit. In the following, the smallest possible value of 450 GeV is chosen as starting point. In addition, any possible bias is accounted for by using the $e_2(x)$ and $d_3(x)$ functions both for the final result, where the difference is used as a systematic uncertainty.

8.6.2 Signal Injection Test

A signal injection test is performed to verify the statistical model consisting of signal and background functions. The result of this test is shown in Fig. 8.16, which shows the mean of the fitted signal cross section as a function of the injected signal cross section for two mass points, $M_{\mathrm{T}} = 600$ and $1000\,\mathrm{GeV}$. The closed symbols show a fitted signal cross section close to the one injected. For the closed symbols, the low mass tail of the reconstructed T mass is not included in the signal parametrisation. However, it is observed that the fitted signal cross section of $M_{\mathrm{T}} = 600\,\mathrm{GeV}$ is around 5% higher than injected, which is consistent with the bias seen in Fig. 8.13.

In order to understand the effect of the approximation used for the signal parametrisation, namely excluding the low mass tail, the signal injection test is repeated by injecting a signal with a realistic line shape obtained by a parametrisation of the low mass tail with a broad Gaussian function.

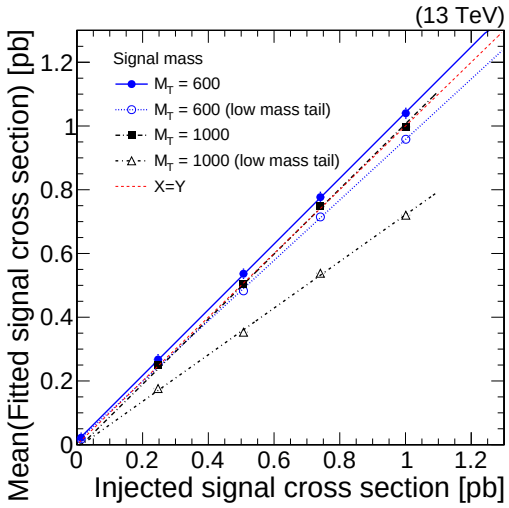


Figure 8.16: The mean of fitted signal cross section as a function of the injected signal cross section for two signal mass points: 600 (blue circles) and 1000 GeV (black rectangles). The open symbols represent the result where the signal was generated with a double Gaussian function including the low mass tail, whereas the fit was performed with one Gaussian function.

The resulting toys are fitted with the single Gaussian approximation. The results of this test are shown with the open markers in Fig. 8.16. At a signal mass of 600 GeV, a loss of 5% in the fitted signal cross section is observed, which originates from the fraction of events neglected by the Gaussian fit. At M_{T} of 1 TeV, a loss of 30% is observed, because the background parametrisation absorbs a part of the signal.

8.6.3 Parametrisation including the Low Mass Tail

In the bias test and the signal injection test it was seen that the approximation of the signal parametrisation, a single Gauss that excludes the modeling of the low mass tail resulting from false combinations in the reconstruction, leads to a reduced sensitivity. Therefore, a parametrisation of the signal including the low mass tail was tested defined as

$$f(x) = N\left[(1-f_{\mathrm{norm}})\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) + f_{\mathrm{norm}}\exp\left(\frac{(x-\mu_2)}{2\sigma_2^2}\right)\right], \tag{8.4}$$

where, N is the normalisation of the parametrisation, f_{norm} is the fraction of the normalisation of the second Gauss, μ and σ are the parameters of the main Gauss (similar as for the single Gauss) and μ_2 and σ_2 are the parameters of the second Gaussian, which parametrize the low mass tail. An example of this parametrisation can be seen in Fig. 8.17 for a signal mass point of $M_{\mathrm{T}} = 700\,\mathrm{GeV}$ in 2017 in the ℓ+jets channel.

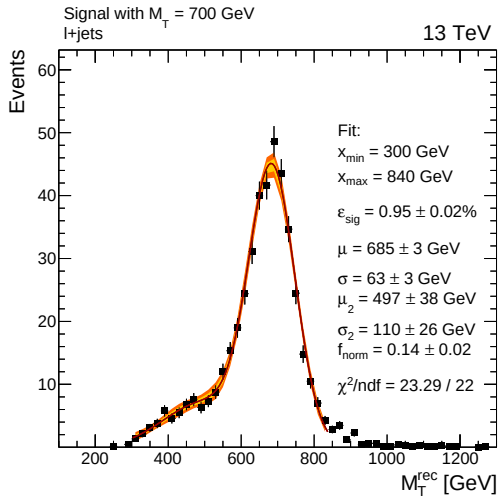


Figure 8.17: Signal parametrisation with two Gauss to model the left sided tail for $M_{\mathrm{T}} = 700\,\mathrm{GeV}$ in 2017 in the ℓ+jets channel.

Each parameter of $f(x)$ as a function M_{T} is shown in Fig. 8.18. The parameters μ and σ of the main Gauss are very similar to the values obtained in Fig. 8.11. The bias test for $M_{\mathrm{T}} = 600\,\mathrm{GeV}$ is performed again to understand the effect of the low mass tail on the choice of the parametrisation. The result is shown in Fig. 8.19. The bias is slightly reduced for both $r = 0$ (left) and $r = 1$ (right) and is consistent within the errors with the results obtained in Fig. 8.13.

This parametrisation is found to improve the expected and give a similar bias than the simple Gaussian approximation. Therefore this parametrisation is used to calculate the expected sensitivity in the following. However, the simple Gaussian parametrisation is useful to understand the bulk of signal shape and is a cross check for the mean and width of the distribution.

8.7 Systematic Uncertainties

Two uncertainty sources are evaluated for the estimate of the background: the choice of functional form and the uncertainties in the fit parameter values. The $e_2(x)$ function is compared to the $d_3(x)$ function to evaluate the effect of the choice of the functional form. The uncertainties in the fit parameter values result directly from the fit to data and result in the largest uncertainties in this analysis.

All other systematic uncertainties affect the signal and are discussed in the following. The jet energy scale and resolution have an effect on the signal shape, all other uncertainties influence only the normalisation.

Jet Energy Scale and Resolution: The jet energy scale and resolution [123]

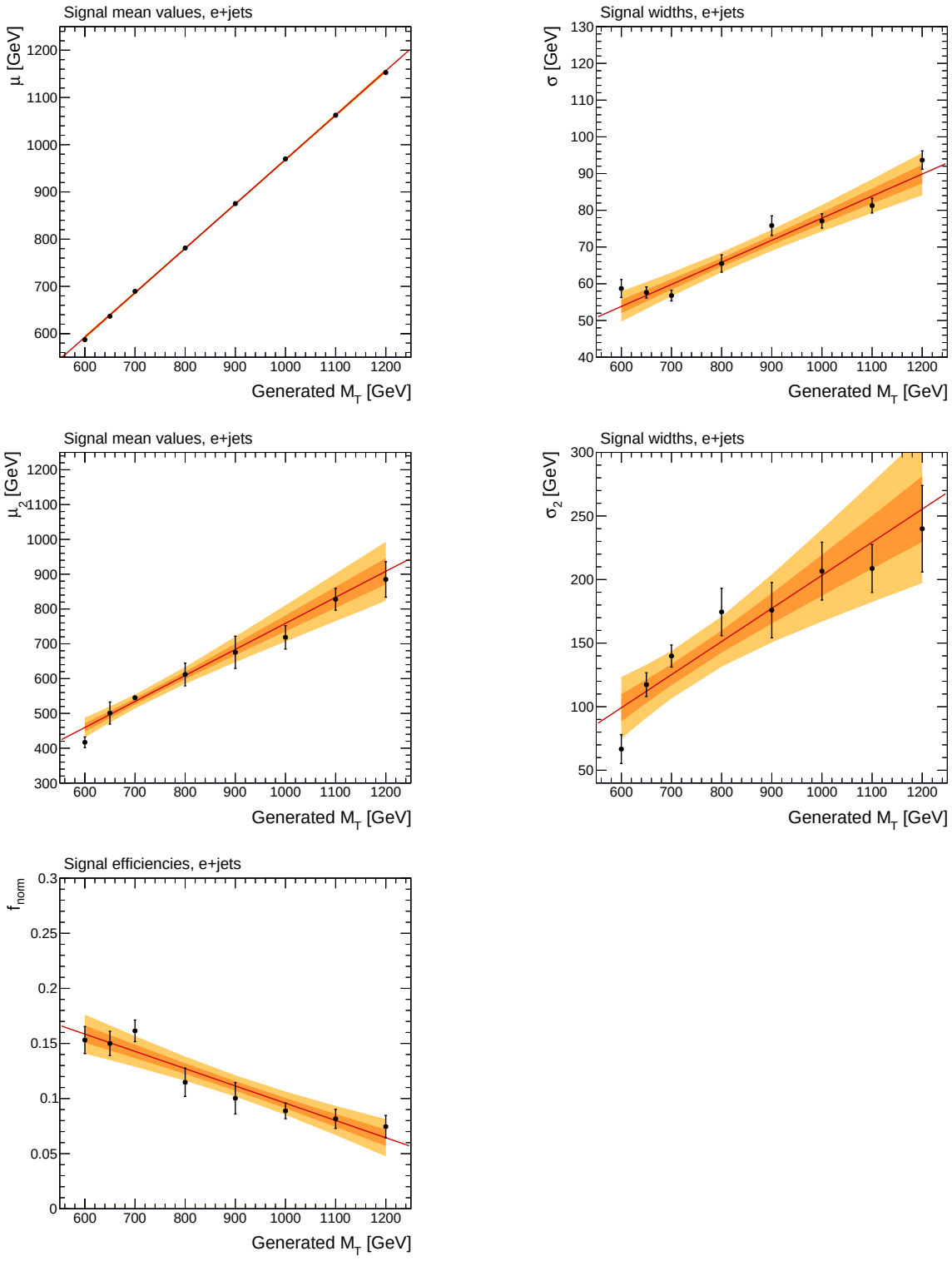


Figure 8.18: The mean (upper left) and width (upper right) of the main Gaussian parametrisation and the mean (middle left) and width (middle right) of the second broader Gauss used to parametrize the low mass tail, (lower) shows the fraction of normalisation for the second Gauss as function of M_{T} in the electron channel. The black dots represent the Run 2 simulation, whereas the red line shows the fit of a linear function. The yellow band shows the one σ band of the fit and the orange band the two σ band of the fit.

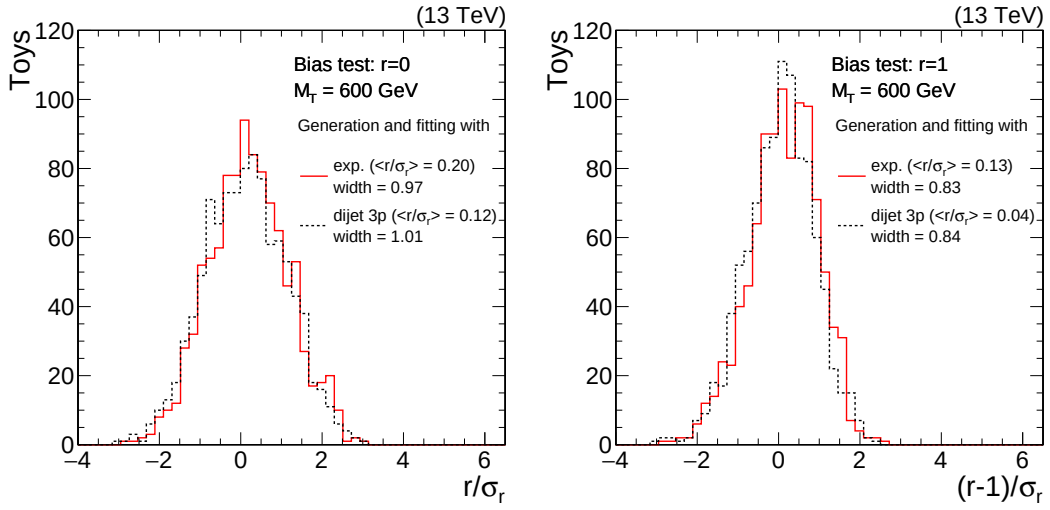


Figure 8.19: The bias test for three different background representations. Pseudo data without signal are generated according to the legend with the $e_2(x)$, the $d_2(x)$ or the $d_3(x)$ function. The fit is performed with the $e_2(x)$ function. The mean of the resulting signal strength (r) divided by its error (σ_r) is depicted in the legend.

is varied by ± 1 standard deviation to estimate the effect on the signal templates and the uncertainty is propagated to p_T^{miss}. The jet energy scale increases (decreases) the mean of the signal templates by more than the 1σ band (statistical uncertainty only) of the signal fits. The jet energy resolution increases (decreases) the width of the signal templates by more than the 1σ band of the signal fits. Therefore, alternative signal shapes for the jet energy resolution and for the jet energy scale variation are implemented.

b-tagging Efficiency and Misidentification Rate: Data-to-simulation scale factors for the b quark identification and misidentification rate are applied as a function of p_T and η. The uncertainty is estimated by varying the scale factor by ± 1 standard deviation [124]. The tagging efficiency of b- and c-jets are taken as fully correlated and therefore are varied simultaneously. The tagging efficiency of jets from light quarks and gluons is assumed fully uncorrelated and is varied independently. The b-tag efficiency uncertainty leads to an up to 30% effect of the efficiency in 2017 (see Fig. 8.20) and corresponds to the largest single uncertainty on the signal. This uncertainty is correlated between the years as a conservative choice.

Muon and Electron Identification and Trigger Efficiency: Data-to-simulation scale factors for the identification and trigger efficiencies are applied as a function of p_T and η. Each scale factor is varied by ± 1 standard deviation to estimate the uncertainties on the signal template.

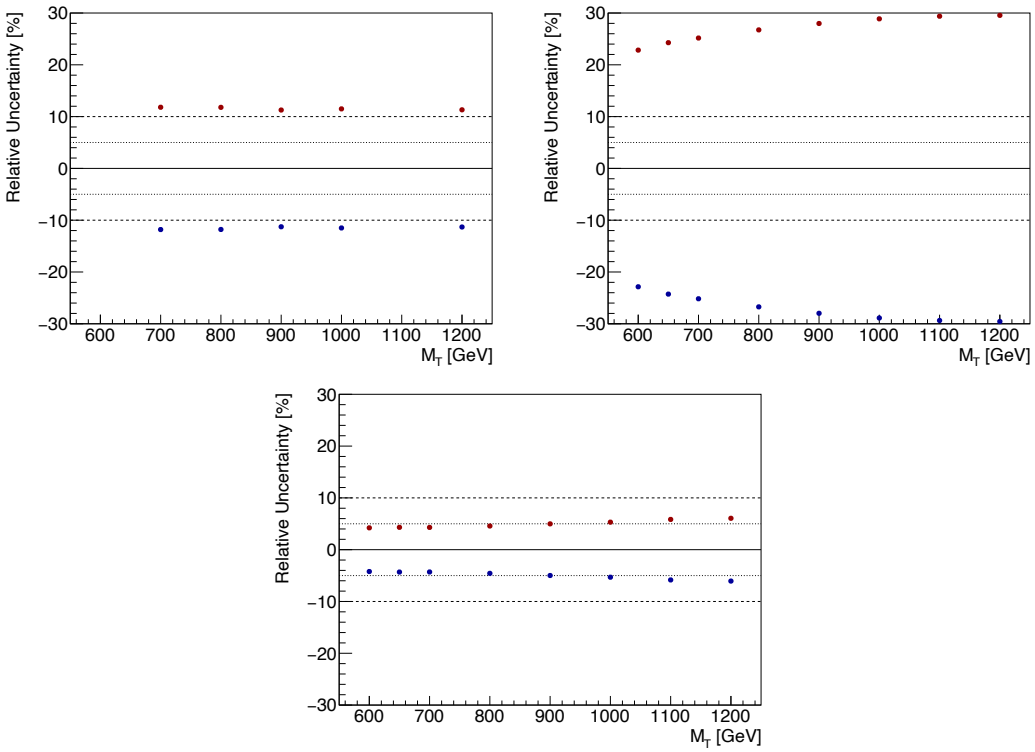


Figure 8.20: The effect of the b-tagging efficiency on the signal efficiency for 2016 (upper left), 2017 (upper right) and 2018 (lower).

Luminosity Uncertainty: The uncertainty on the integrated luminosity recorded by the CMS detector is 2.5% for 2016 and 2018 data and 2.3% for 2017 data [152, 182, 183].

Pileup Uncertainty: The nominal minimum bias cross section for the pileup reweighting is 69 mb. The minimum bias cross section is varied by 5% and the samples are reweighted again.

Renormalisation and Factorisation scales: The uncertainties in the factorisation and renormalization scales μ_F and μ_R are taken into account as proposed in Ref. [162, 163] by varying the default choice of scales by the following six combinations of factors, (μ_F, μ_R) × (1/2, 1/2), (1/2, 1), (1, 1/2), (2, 2), (2, 1), and (1, 2). The maximum and minimum of the six variations are computed for each bin of the reconstructed T mass distribution.

PDF: The PDF uncertainty is estimate from the choice of 100 different PDFs, provided with the NNPDF30_lo_as_0130 [38] set. All samples are re-weighted according to each of the varied sets. In each bin of the final distributions the symmetric standard deviation with respect to the nominal value is taken as the PDF uncertainty, where only the uncertainty on the acceptance is considered.

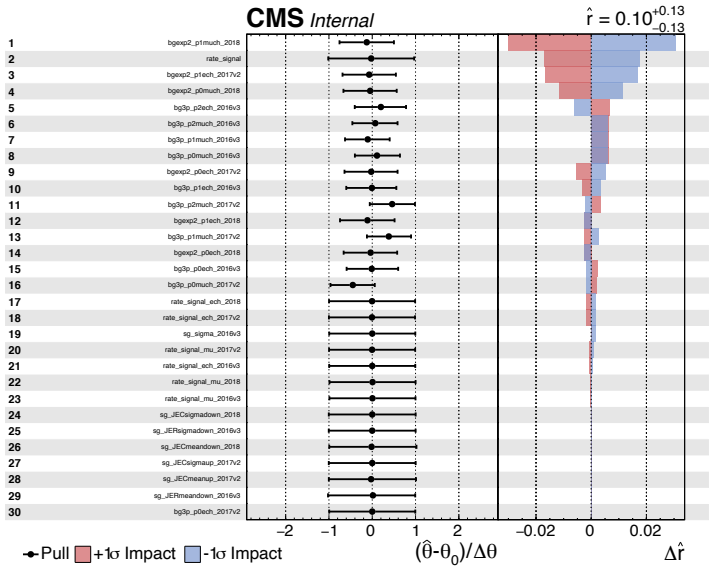


Figure 8.21: The impact of the nuisance parameter on pseudo data generated in the signal region for a signal mass of $M_{\mathrm{T}} = 600\,\mathrm{GeV}$. The left column shows the name of the nuisance parameter, the middle column shows the pull of the nuisance parameter and the right column shows the impact of each nuisance parameter. The nuisance parameters for the function of the background have the following form bgexp2 or bg3p for the $e_2(x)$ and $d_3(x)$ function respectively. The next term describes the parameter as defined in Eq. 8.2 and Eq. 8.3 and the lepton channel used. The last term describes the year for which the parameter is assigned. The nuisance parameter rate_signal covers all uncertainties on the signal efficiency, except the μ and e identification and trigger systematics. These are incorporate into rate_signal_(channel)_(year).

8.8 Results

Systematic uncertainties enter the statistical interpretation as nuisance parameters, which are marginalized during the statistical analysis. The impact of each nuisance parameter on the signal strength is studied. Fig. 8.21 shows the importance of the various sources of systematic uncertainties, obtained by changing a given nuisance parameter by $\pm 1\sigma$ and repeating the fit of the signal strength. The change of the extracted signal strength with respect to the nominal is denoted as $\Delta\hat{r}$.

The background parametrisation together with the uncertainty on the signal efficiency (rate_signal) has the largest impact. Changing, for example, the parameter $p1$ of the $e_2(x)$ function in the muon channel by $\pm 1\sigma$ results in a change of around 0.03 in Δr. The uncertainties in the signal efficiency show a comparable effect between 0.015 in Δr. All other nuisance parameters have an influence less than 0.02 on ΔR.

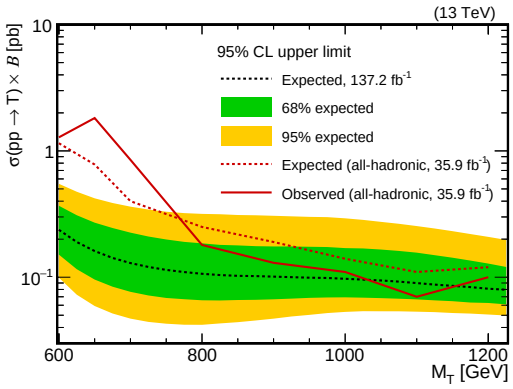


Figure 8.22: Expected upper limits on the T production cross section times the $\mathcal{B}$ at the 95% CL as a function of the T mass. The red line shows the expected (dotted) and observed (solid) upper cross section limit of the all-hadronic analysis [20].

A background-only fit is performed to the data in the signal region. The resulting background parametrisation is used to generate toy data, which are used to derive expected upper cross section limits. The asymptotic frequentest method is used to derive expected 95% CL upper limits on the production cross section of the T. The systematic uncertainties are included through nuisance parameters as explained above. The resulting expected upper limits are shown in Fig. 8.22 for the full Run 2 data set. Upper limits on the cross section times $\mathcal{B}$ reach from 0.24 pb at $M_{\mathrm{T}} = 600\,\mathrm{GeV}$ to about 0.08 pb at $M_{\mathrm{T}} = 1200\,\mathrm{GeV}$.

The red line in Fig. 8.22 shows the expected (dotted) and observed (solid) upper cross section limit of the all-hadronic analysis. With the sensitivity reached it is possible to confirm or exclude the excess observed by the orthogonal all-hadronic analysis [20]. The expected sensitivity of this analysis with the full Run 2 data set corresponds to a significance of five standard deviations for a possible signal with mass of 650 GeV, visible in the all-hadronic final state.

9 | Summary

In this thesis various different aspects have been reported in the search for new physics at the LHC. These range from detector developments, to the improvement of reconstruction algorithms, to developments for the identification of particle decays, and to full analyses of pp collider data.

In order to improve the sensitivity of searches for new physics and measurements of standard model quantities, three critical developments have been conducted during Run 2 (2016 – 2018): the upgrade of the CMS pixel detector, the validation and tuning of a new pileup mitigation technique and a study on the decorrelation of the machine-learning techniques to identify highly boosted particle decays. The developments are crucial for the data taking or improve the data analysis in the CMS Collaboration during Run 2 and beyond.

The new CMS pixel detector was installed between the end of 2016 and beginning of 2017. Since only a short time was planned for the installation, a large part of the off-detector services, such as power cables, were kept. The new pixel detector had an additional layer with sensors, leading to an improved b-tagging and tracking. In order to reduce the power losses, DC-DC converters are used for powering the sensors, which were never tested before in the unique and extreme conditions of the LHC. These DC-DC converters started to fail at the end of 2017 data taking. Projections of the failure showed that without a solution b-tagging and tracking efficiencies would decrease, resulting in a significant degradation of the data quality by summer 2018. A systematic analysis of the failure and the DC-DC converters extracted from the detector was presented in this thesis. This study led to a change in the operation of the CMS pixel detector in 2018, so that no DC-DC converter failed during data taking.

Once the data has been reconstructed, pileup (PU) mitigation plays an important role in nearly all physics analyses. A very efficient technique to reduce the effects of PU is the pileup-per-particle identification (PUPPI) algorithm. Already in 2017 high PU scenarios of up to 50 vertices were reached and it is planned to reach up to 200 vertices during the operation of the high-luminosity LHC. With standard PU mitigation techniques, such as charged hadron subtraction or a PU jet ID, the suppression of jets

mainly made from PU particles, the jet energy resolution or the discrimination power of substructure variables degrades. In this thesis the first analysis of the key variables entering the PUPPI algorithm and a comparison of pileup mitigation algorithms in terms of efficiency, purity and pileup jet rate have been presented. In addition, an improvement of the PUPPI algorithm to regain the jet energy resolution at high p_T has been developed. The improved version of the PUPPI algorithm has become the default pileup mitigation technique in the CMS Collaboration and will be used for future Run 2 data analyses and during the operation of Run 3 (planned: 2021-2023). Studies are ongoing to use the PUPPI algorithm at trigger level during the operation of Run 3.

Not only improvements in data quality and reconstruction lead to a better sensitivity in searches and measurements, but also substructure taggers can contribute through improved discrimination power between background and signal. In the past years new machine learning taggers were developed to identify boosted bosons and top quarks. Substructure taggers are crucial in the boosted regime to discriminate between jets originating from heavy boson and top quarks and jets originating from QCD multijet processes, e.g. when searching for a new heavy resonance that decays into a pair of standard model heavy bosons. In this case it can be important to decorrelate the taggers in order to search for a bump on a smoothly falling standard model background without p_T-dependent shapes introduced by requirements on the tagger. In this thesis a study of the performance of new tagging techniques and its decorrelation was presented. This has led to an improvement of about 30% in the expected upper cross section limits in a search for a resonance Z' decaying into a pair of bosons in the all-hadronic final state.

If the mass of the resonance Z' is of the order of a few TeV, the decay into a mixture of a top quark and a new heavy quark T is possible in a number of beyond the standard model theories. The T has three standard decay modes: $T \to Ht$, $T \to Zt$ and $T \to Wb$. Only two searches, focusing on $T \to Wb$ [87] and $T \to Zt$ [21], existed when this thesis was performed. It was not possible to exclude any of the considered models with this searches.

The search presented in this thesis analysed $35.9\,fb^{-1}$ of pp collider data at $\sqrt{s} = 13\,TeV$ in the lepton+jets final state including all three decay channels of the T with a focus on $T \to Ht$ and $T \to Zt$. This search led to the most stringent limits to date and excludes a heavy resonance of the G^* model between a mass of 1.5 and 2.3 TeV if $M_T = 1.2\,TeV$, and between 2.0 and 2.5 TeV if $M_T = 1.5\,TeV$.

A large effort has also been made to exclude vector-like quarks like the T in single and pair production [11–25]. Searches performed for single and pair production of

vector-like quarks have set lower bounds of around 1.3 TeV on the mass of the T. However, it is not mandatory that the T has only three decay channels. There are extensions of the standard model that predict a further decay channel: $\mathrm{T} \to a\mathrm{t}$, where a is a new scalar particle. As a result the mass limit on vector-like quarks from existing searches is reduced.

A search for a singly-produced vector-like T in the lepton+jets final state was presented in this thesis. The full Run 2 (2016-2018) data set ($137.2\,\mathrm{fb}^{-1}$) is analysed. The search is orthogonal to an existing search in the all-hadronic final state [20], which observes an excess between 600 and 700 GeV. The all-hadronic analysis selects events with at least five jets, three of which are required to be b tagged. The background is dominated by QCD multijet production and estimated by a complicated extrapolation method. In order to validate this result, the analysis in the lepton+jets channel has been constructed to search for a resonance structure on a smoothly falling background, which can be parametrized by an exponential or a power-law function. Extensive statistical tests have proven this method to work well and the sensitivity achieved is more than sufficient to validate or falsify the observed excess. The expected upper cross section limits are about an order of magnitude better at $M_{\mathrm{T}} = 600\,\mathrm{GeV}$ than the ones obtained by the all-hadronic analysis [20].

Bibliography

[1] CMS Collaboration, "Search for a heavy resonance decaying to a top quark and a vector-like top quark in the lepton+jets final state in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.* **C79** (2019), no. 3, 208, doi:10.1140/epjc/s10052-019-6688-5, arXiv:1812.06489.

[2] M. Czimmeck, "Suche nach einer Z' Resonanz mit dem CMS Experiment", (2019). Bachelor thesis.

[3] CMS Collaboration, "Phase-I Pixel ROC Thresholds and DC-DC Converter Characterization in 2018", CMS-DP-2020-005, CERN, Geneva, (Feb, 2020).

[4] 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.

[5] C. Bini, R. Contino, and N. Vignaroli, "Heavy-light decay topologies as a new strategy to discover a heavy gluon", *JHEP* **01** (2012) 157, doi:10.1007/JHEP01(2012)157, arXiv:1110.6058.

[6] D. Greco and D. Liu, "Hunting composite vector resonances at the LHC: naturalness facing data", *JHEP* **12** (2014) 126, doi:10.1007/JHEP12(2014)126, arXiv:1410.2883.

[7] D. Barducci and C. Delaunay, "Bounding wide composite vector resonances at the LHC", *JHEP* **02** (2016) 055, doi:10.1007/JHEP02(2016)055, arXiv:1511.01101.

[8] L. Da Rold and A. N. Rossia, "The Minimal Simple Composite Higgs Model", *JHEP* **12** (2019) 023, doi:10.1007/JHEP12(2019)023, arXiv:1904.02560.

[9] 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.

[10] L. Randall and R. Sundrum, "An Alternative to compactification", *Phys. Rev. Lett.* **83** (1999) 4690–4693, doi:10.1103/PhysRevLett.83.4690, arXiv:hep-th/9906064.

[11] ATLAS Collaboration, "Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector", *JHEP* **08** (2017) 052, doi:10.1007/JHEP08(2017)052, arXiv:1705.10751.

[12] ATLAS Collaboration, "Search for pair production of heavy vector-like quarks decaying into high-p_T W bosons and top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **08** (2018) 048, doi:10.1007/JHEP08(2018)048, arXiv:1806.01762.

[13] ATLAS Collaboration, "Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b-jets with the ATLAS detector", *JHEP* **07** (2018) 089, doi:10.1007/JHEP07(2018)089, arXiv:1803.09678.

[14] ATLAS Collaboration, "Search for pair- and single-production of vector-like quarks in final states with at least one Z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV", *Phys. Rev.* **D98** (2018), no. 11, 112010, doi:10.1103/PhysRevD.98.112010, arXiv:1806.10555.

[15] ATLAS Collaboration, "Search for pair production of heavy vector-like quarks decaying into hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev.* **D98** (2018), no. 9, 092005, doi:10.1103/PhysRevD.98.092005, arXiv:1808.01771.

[16] ATLAS Collaboration, "Combination of the searches for pair-produced vector-like partners of the third-generation quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. Lett.* **121** (2018), no. 21, 211801, doi:10.1103/PhysRevLett.121.211801, arXiv:1808.02343.

[17] ATLAS Collaboration, "Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **05** (2019) 164, doi:10.1007/JHEP05(2019)164, arXiv:1812.07343.

[18] CMS Collaboration, "Search for pair production of vector-like quarks in the bW$\bar{\text{b}}$W channel from proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett.*

B779 (2018) 82–106, doi:10.1016/j.physletb.2018.01.077, arXiv:1710.01539.

[19] CMS Collaboration, "Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV", *JHEP* **08** (2018) 177, doi:10.1007/JHEP08(2018)177, arXiv:1805.04758.

[20] CMS Collaboration, "Search for electroweak production of a vector-like T quark using fully hadronic final states", *JHEP* **01** (2020) 036, doi:10.1007/JHEP01(2020)036, arXiv:1909.04721.

[21] CMS Collaboration, "Search for single production of a vector-like T quark decaying to a Z boson and a top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett.* **B781** (2018) 574–600, doi:10.1016/j.physletb.2018.04.036, arXiv:1708.01062.

[22] CMS Collaboration, "Search for electroweak production of a vector-like quark decaying to a top quark and a Higgs boson using boosted topologies in fully hadronic final states", *JHEP* **04** (2017) 136, doi:10.1007/JHEP04(2017)136, arXiv:1612.05336.

[23] CMS Collaboration, "Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state", *Phys. Lett.* **B771** (2017) 80–105, doi:10.1016/j.physletb.2017.05.019, arXiv:1612.00999.

[24] CMS Collaboration, "Search for pair production of vectorlike quarks in the fully hadronic final state", *Phys. Rev.* **D100** (2019), no. 7, 072001, doi:10.1103/PhysRevD.100.072001, arXiv:1906.11903.

[25] CMS Collaboration, "Search for vector-like quarks in events with two oppositely charged leptons and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.* **C79** (2019), no. 4, 364, doi:10.1140/epjc/s10052-019-6855-8, arXiv:1812.09768.

[26] G. Cacciapaglia, T. Flacke, M. Park et al., "Exotic decays of top partners: mind the search gap", *Phys. Lett. B* **798** (2019) 135015, doi:10.1016/j.physletb.2019.135015, arXiv:1908.07524.

[27] 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.

[28] CMS Collaboration, "Pileup Removal Algorithms", CMS-PAS-JME-14-001, (2014).

[29] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (Aug, 2018) 030001, doi:10.1103/PhysRevD.98.030001.

[30] D. Griffiths, "Introduction to Elementary Particles". Physics textbook. Wiley, 2008.

[31] M. Peskin and D. Schroeder, "An Introduction to Quantum Field Theory". Advanced book classics. Addison-Wesley Publishing Company, 1995.

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

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

[34] 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. (1964).

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

[36] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[37] 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.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

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

[39] CMS Collaboration, "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **03** (2020) 056, doi:10.1007/JHEP03(2020)056, arXiv:1907.11270.

[40] CMS Collaboration, "Measurement of the jet mass distribution and top quark mass in hadronic decays of boosted top quarks in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.* **124** (2020), no. 20, 202001, doi:10.1103/PhysRevLett.124.202001, arXiv:1911.03800.

[41] 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.

[42] 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.

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

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

[45] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[46] T. Sjöstrand, S. Ask, J. R. Christiansen et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

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

[48] 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.

[49] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[50] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[51] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", arXiv:1903.12179.

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

[53] Gfitter Group Collaboration, "The global electroweak fit at NNLO and prospects for the LHC and ILC", *Eur. Phys. J.* **C74** (2014) 3046, doi:10.1140/epjc/s10052-014-3046-5, arXiv:1407.3792.

[54] J. Haller, A. Hoecker, R. Kogler et al., "Update of the global electroweak fit and constraints on two-Higgs-doublet models", *Eur. Phys. J.* **C78** (2018), no. 8, 675, doi:10.1140/epjc/s10052-018-6131-3, arXiv:1803.01853.

[55] D. Clowe, M. Bradac, A. H. Gonzalez et al., "A direct empirical proof of the existence of dark matter", *Astrophys. J.* **648** (2006) L109–L113, doi:10.1086/508162, arXiv:astro-ph/0608407.

[56] KATRIN Collaboration, "Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN", *Phys. Rev. Lett.* **123** (2019), no. 22, 221802, doi:10.1103/PhysRevLett.123.221802, arXiv:1909.06048.

[57] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, "The Hierarchy problem and new dimensions at a millimeter", *Phys. Lett.* **B429** (1998) 263–272, doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.

[58] M. E. Peskin, "On the Trail of the Higgs Boson", *Annalen Phys.* **528** (2016), no. 1-2, 20–34, doi:10.1002/andp.201500225, arXiv:1506.08185.

[59] D. Buttazzo, G. Degrassi, P. P. Giardino et al., "Investigating the near-criticality of the Higgs boson", *JHEP* **12** (2013) 089, doi:10.1007/JHEP12(2013)089, arXiv:1307.3536.

[60] R. M. Harris and S. Jain, "Cross Sections for Leptophobic Topcolor Z' Decaying to Top-Antitop", *Eur. Phys. J.* **C72** (2012) 2072, doi:10.1140/epjc/s10052-012-2072-4, arXiv:1112.4928.

[61] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, "Phenomenology of the Randall-Sundrum Gauge Hierarchy Model", *Phys. Rev. Lett.* **84** (2000) 2080, doi:10.1103/PhysRevLett.84.2080, arXiv:hep-ph/9909255.

[62] CMS Collaboration, "Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **04** (2019) 031, doi:10.1007/JHEP04(2019)031, arXiv:1810.05905.

[63] CMS Collaboration, "Search for $t\bar{t}$ resonances in highly boosted lepton+jets and fully hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2017) 001, doi:10.1007/JHEP07(2017)001, arXiv:1704.03366.

[64] ATLAS Collaboration, "Search for heavy particles decaying into a top-quark pair in the fully hadronic final state in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev.* **D99** (2019), no. 9, 092004, doi:10.1103/PhysRevD.99.092004, arXiv:1902.10077.

[65] ATLAS Collaboration, "Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J.* **C78** (2018), no. 7, 565, doi:10.1140/epjc/s10052-018-5995-6, arXiv:1804.10823.

[66] CMS Collaboration, "A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at 13 TeV", *Eur. Phys. J. C* **80** (2020), no. 3, 237, doi:10.1140/epjc/s10052-020-7773-5, arXiv:1906.05977.

[67] CMS Collaboration, "Combination of CMS searches for heavy resonances decaying to pairs of bosons or leptons", *Phys. Lett.* **B798** (2019) 134952, doi:10.1016/j.physletb.2019.134952, arXiv:1906.00057.

[68] 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.

[69] CMS Collaboration, "Search for a heavy resonance decaying to a pair of vector bosons in the lepton plus merged jet final state at $\sqrt{s} = 13$ TeV", *JHEP* **05** (2018) 088, doi:10.1007/JHEP05(2018)088, arXiv:1802.09407.

[70] CMS Collaboration, "Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV", *Eur. Phys.*

J. C **77** (2017), no. 9, 636, doi:10.1140/epjc/s10052-017-5192-z, arXiv:1707.01303.

[71] CMS Collaboration, "Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos, and b quarks", *Phys. Lett. B* **768** (2017) 137–162, doi:10.1016/j.physletb.2017.02.040, arXiv:1610.08066.

[72] CMS Collaboration, "Search for massive resonances decaying into WW, WZ, ZZ, qW, and qZ with dijet final states at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **97** (2018), no. 7, 072006, doi:10.1103/PhysRevD.97.072006, arXiv:1708.05379.

[73] ATLAS Collaboration, "Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **09** (2019) 091, doi:10.1007/JHEP09(2019)091, arXiv:1906.08589.

[74] ATLAS Collaboration, "Search for WW/WZ resonance production in $\ell\nu qq$ final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **03** (2018) 042, doi:10.1007/JHEP03(2018)042, arXiv:1710.07235.

[75] ATLAS Collaboration, "Searches for heavy ZZ and ZW resonances in the $\ell\ell qq$ and $\nu\nu qq$ final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **03** (2018) 009, doi:10.1007/JHEP03(2018)009, arXiv:1708.09638.

[76] ATLAS Collaboration, "Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett.* **B774** (2017) 494–515, doi:10.1016/j.physletb.2017.09.066, arXiv:1707.06958.

[77] ATLAS Collaboration, "Search for heavy resonances decaying into WW in the $e\nu\mu\nu$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J.* **C78** (2018), no. 1, 24, doi:10.1140/epjc/s10052-017-5491-4, arXiv:1710.01123.

[78] ATLAS Collaboration, "Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev.* **D98** (2018), no. 5, 052008, doi:10.1103/PhysRevD.98.052008, arXiv:1808.02380.

[79] ATLAS Collaboration, "Search for heavy resonances decaying into a W or Z boson and a Higgs boson in final states with leptons and b-jets in 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector", *JHEP* **03** (2018) 174, doi:10.1007/JHEP11(2018)051, arXiv:1712.06518.

[80] ATLAS Collaboration, "Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* **777** (2018) 91–113, doi:10.1016/j.physletb.2017.12.011, arXiv:1708.04445.

[81] ATLAS Collaboration, "Searches for heavy diboson resonances in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **09** (2016) 173, doi:10.1007/JHEP09(2016)173, arXiv:1606.04833.

[82] ATLAS Collaboration, "Search for new resonances decaying to a W or Z boson and a Higgs boson in the $\ell^+\ell^- b\bar{b}$, $\ell\nu b\bar{b}$, and $\nu\bar{\nu}b\bar{b}$ channels with pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* **765** (2017) 32–52, doi:10.1016/j.physletb.2016.11.045, arXiv:1607.05621.

[83] D. Pappadopulo, A. Thamm, R. Torre et al., "Heavy Vector Triplets: Bridging Theory and Data", *JHEP* **09** (2014) 060, doi:10.1007/JHEP09(2014)060, arXiv:1402.4431.

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

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

[86] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer et al., "Handbook of vectorlike quarks: Mixing and single production", *Phys. Rev.* **D88** (2013), no. 9, 094010, doi:10.1103/PhysRevD.88.094010, arXiv:1306.0572.

[87] CMS Collaboration, "Search for a heavy resonance decaying to a top quark and a vector-like top quark at $\sqrt{s} = 13$ TeV", *JHEP* **09** (2017) 053, doi:10.1007/JHEP09(2017)053, arXiv:1703.06352.

[88] CMS Collaboration, "Search for narrow resonances in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 13$ TeV and combination with 8 TeV data",

Phys. Lett. B **768** (2017) 57–80, doi:10.1016/j.physletb.2017.02.010, arXiv:1609.05391.

[89] CMS Collaboration, "Search for a high-mass resonance decaying into a dilepton final state in 13 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV", CMS-PAS-EXO-16-031, (8, 2016).

[90] CMS Collaboration, "Search for high mass resonances in dielectron final state", CMS-PAS-EXO-18-006, (3, 2018).

[91] ATLAS Collaboration, "Search for high-mass dilepton resonances using 139 fb^{-1} of *pp* collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector", *Phys. Lett. B* **796** (2019) 68–87, doi:10.1016/j.physletb.2019.07.016, arXiv:1903.06248.

[92] ATLAS Collaboration, "Search for high-mass new phenomena in the dilepton final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* **761** (2016) 372–392, doi:10.1016/j.physletb.2016.08.055, arXiv:1607.03669.

[93] CMS Collaboration, "Searches for dijet resonances in pp collisions at $\sqrt{s} = 13$ TeV using the 2016 and 2017 datasets", CMS-PAS-EXO-17-026, (9, 2018).

[94] CMS Collaboration, "Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:1911.03947.

[95] ATLAS Collaboration, "Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **03** (2020) 145, doi:10.1007/JHEP03(2020)145, arXiv:1910.08447.

[96] ATLAS Collaboration, "Search for new phenomena in dijet events using 37 fb^{-1} of *pp* collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector", *Phys. Rev. D* **96** (2017), no. 5, 052004, doi:10.1103/PhysRevD.96.052004, arXiv:1703.09127.

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

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

[99] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software",.

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

[101] CMS Collaboration, "Commissioning and first results from the CMS phase 1 upgrade pixel detector", *PoS* **Vertex 2017** (2018) 018, doi:10.22323/1.309.0018, arXiv:1807.08987.

[102] CMS Collaboration, "2017 tracking performance plots", , (Apr, 2017).

[103] H. C. Kastli, M. Barbero, W. Erdmann et al., "Design and performance of the CMS pixel detector readout chip", *Nucl. Instrum. Meth.* **A565** (2006) 188–194, doi:10.1016/j.nima.2006.05.038, arXiv:physics/0511166.

[104] H. C. Kästli, "Frontend electronics development for the CMS pixel detector upgrade", *Nucl. Instrum. Meth.* **A731** (2013) 88–91, doi:10.1016/j.nima.2013.05.056.

[105] L. Feld, W. Karpinski, K. Klein et al., "Experience from design, prototyping and production of a DC-DC conversion powering scheme for the CMS Phase-1 Pixel Upgrade", *JINST* **11** (2016), no. 02, C02033, doi:10.1088/1748-0221/11/02/C02033.

[106] L. Feld, C. Fimmers, W. Karpinski et al., "Development of a DC-DC conversion powering scheme for the CMS Phase-1 pixel upgrade", *JINST* **9** (2014), no. 01, C01048, doi:10.1088/1748-0221/9/01/C01048.

[107] F. Faccio, G. Blanchot, C. Fuentes et al., "FEAST2: A Radiation and Magnetic Field Tolerant Point-of-Load Buck DC/DC Converter", in *Proceedings, Radiation Effects Data Workshop (REDW): Paris, France, July 16, 2014*, p. 7004569. 2014. doi:10.1109/REDW.2014.7004569.

[108] F. Faccio, S. Michelis, S. Orlandi et al., "Development of custom radiation-tolerant DCDC converter ASICs", *JINST* **5** (2010), no. 11, C11016, doi:10.1088/1748-0221/5/11/C11016.

[109] CMS Collaboration, "Test Beam Campaigns for the CMS Phase I Upgrade Pixel Readout Chip", *JINST* **9** (2014), no. 12, C12001, doi:10.1088/1748-0221/9/12/C12001, arXiv:1410.1399.

[110] CMS Collaboration, "Beam test characterization of CMS silicon pixel detectors for the phase-1 upgrade", *Nucl. Instrum. Meth.* **A796** (2015) 64–67, doi:10.1016/j.nima.2015.03.074.

[111] CMS Collaboration, "The CMS Phase-1 pixel detector – experience and lessons learned from two years of operation", CMS-CR-2019-019. 07, CERN, Geneva, (Feb, 2019).

[112] CMS Collaboration, "Results related to the Phase1 HE upgrade", CMS-DP-2018-019, (May, 2018).

[113] CMS Collaboration, "HCAL Out Of Time Pileup Subtraction and Energy Reconstruction", CMS-DP-2018-018, (May, 2018).

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

[115] W. Adam, B. Mangano, T. Speer et al., "Track Reconstruction in the CMS tracker", CMS-NOTE-2006-041, CERN, Geneva, (Dec, 2006).

[116] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{\text{s}} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[117] 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.

[118] 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.

[119] 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.

[120] 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.

[121] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[122] Y. L. Dokshitzer, G. Leder, S. Moretti et al., "Better jet clustering algorithms", *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

[123] 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.

[124] 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.

[125] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector", CMS-DP-2018-058, (Nov, 2018).

[126] CMS Collaboration, "CMS Phase 1 heavy flavour identification performance and developments", CMS-DP-2017-013, (May, 2017).

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

[128] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming", *JHEP* **02** (2010) 084, doi:10.1007/JHEP02(2010)084, arXiv:0912.1342.

[129] 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.

[130] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure", *JHEP* **06** (2013) 108, doi:10.1007/JHEP06(2013)108, arXiv:1305.0007.

[131] A. J. Larkoski, I. Moult, and D. Neill, "Power Counting to Better Jet Observables", *JHEP* **12** (2014) 009, doi:10.1007/JHEP12(2014)009, arXiv:1409.6298.

[132] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", arXiv:2004.08262.

[133] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines", in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, p. 807. Omnipress, USA, 2010.

[134] N. Srivastava, G. Hinton, A. Krizhevsky et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research* **15** (2014), no. 56, 1929–1958.

[135] G. Klambauer, T. Unterthiner, A. Mayr et al., "Self-normalizing Neural Networks", in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, p. 972. Curran Associates Inc., USA, 2017.

[136] CMS Collaboration, "Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **01** (2018) 097, doi:10.1007/JHEP01(2018)097, arXiv:1710.00159.

[137] J. Dolen, P. Harris, S. Marzani et al., "Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure", *JHEP* **05** (2016) 156, doi:10.1007/JHEP05(2016)156, arXiv:1603.00027.

[138] CMS Collaboration, "Performance of Deep Tagging Algorithms for Boosted Double Quark Jet Topology in Proton-Proton Collisions at 13 TeV with the Phase-0 CMS Detector", CMS-DP-2018-046, (Jul, 2018).

[139] I. Zoi, "Search for new physics in diboson signatures with CMS and upgrade of the CMS pixel detector". PhD thesis, Universität Hamburg, 2021. (Expected).

[140] CMS Collaboration, "A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, ZZ, WH or ZH boson pairs in the dijet final state at 13 TeV", (2020). CMS internal analysis note AN-19-131.

[141] G. Soyez, "Pileup mitigation at the LHC: A theorist's view", *Phys. Rept.* **803** (2019) 1–158, doi:10.1016/j.physrep.2019.01.007, arXiv:1801.09721.

[142] 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.

[143] ATLAS Collaboration, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC", *Phys. Rev.*

Lett. **117** (2016), no. 18, 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.

[144] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[145] ATLAS Collaboration, "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector", *Eur. Phys. J.* **C76** (2016), no. 11, 581, doi:10.1140/epjc/s10052-016-4395-z, arXiv:1510.03823.

[146] CMS Collaboration, "Pileup Jet Identification", CMS-PAS-JME-13-005, (2013).

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

[148] ATLAS Collaboration, "Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector", *Eur. Phys. J.* **C77** (2017), no. 9, 580, doi:10.1140/epjc/s10052-017-5081-5,10.1140/epjc/s10052-017-5245-3, arXiv:1705.02211. [Erratum: Eur. Phys. J.C77,no.10,712(2017)].

[149] ATLAS Collaboration, "Impact of Alternative Inputs and Jet Grooming on Large-R Jet Performance", ATL-PHYS-PUB-2019-027, CERN, Geneva, (Jul, 2019).

[150] ATLAS, CMS Collaboration, "Addendum to the report on the physics at the HL-LHC, and perspectives for the HE-LHC: Collection of notes from ATLAS and CMS", *CERN Yellow Rep. Monogr.* **7** (2019) Addendum, doi:10.23731/CYRM-2019-007.Addendum, arXiv:1902.10229.

[151] A. Deandrea and A. M. Iyer, "Vectorlike quarks and heavy colored bosons at the LHC", *Phys. Rev.* **D97** (2018), no. 5, 055002, doi:10.1103/PhysRevD.97.055002, arXiv:1710.01515.

[152] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", CMS-PAS-LUM-17-001, (3, 2017).

[153] S. Alioli, P. Nason, C. Oleari et al., "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111,

doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: JHEP 02, 011 (2010)].

[154] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[155] 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.

[156] J. M. Campbell, R. K. Ellis, P. Nason et al., "Top-Pair Production and Decay at NLO Matched with Parton Showers", *JHEP* **04** (2015) 114, doi:10.1007/JHEP04(2015)114, arXiv:1412.1828.

[157] S. Alioli, S.-O. Moch, and P. Uwer, "Hadronic top-quark pair-production with one jet and parton showering", *JHEP* **01** (2012) 137, doi:10.1007/JHEP01(2012)137, arXiv:1110.5251.

[158] R. Frederix, E. Re, and P. Torrielli, "Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO", *JHEP* **09** (2012) 130, doi:10.1007/JHEP09(2012)130, arXiv:1207.5391.

[159] CMS Collaboration, "Measurement of normalized differential $\mathrm{t\bar{t}}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **04** (2018) 060, doi:10.1007/JHEP04(2018)060, arXiv:1708.07638.

[160] 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) 092001, doi:10.1103/PhysRevD.95.092001, arXiv:1610.04191.

[161] CMS Collaboration, "Investigations of the impact of the parton shower tuning in PYTHIA 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, (2016).

[162] M. Cacciari, S. Frixione, M. Mangano et al., "The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence", *JHEP* **04** (2004) 068, doi:10.1088/1126-6708/2004/04/068, arXiv:hep-ph/0303085.

[163] S. Catani, D. de Florian, M. Grazzini et al., "Soft gluon resummation for Higgs boson production at hadron colliders", *JHEP* **07** (2003) 028, doi:10.1088/1126-6708/2003/07/028, arXiv:hep-ph/0306211.

[164] D. Lindley, "Kendall's Advanced Theory of Statistics, volume 2B, Bayesian Inference, 2nd edn", volume 168. Blackwell Publishing Ltd, 2005.

[165] L. Demortier, S. Jain, and H. B. Prosper, "Reference priors for high energy physics", *Phys. Rev. D* **82** (2010) 034002, doi:10.1103/PhysRevD.82.034002, arXiv:1002.1111.

[166] T. Müller, J. Ott, and J. Wagner-Kuhr, "Theta—A framework for template-based modeling and inference", 2010.

[167] D. Liu, L.-T. Wang, and K.-P. Xie, "Prospects of searching for composite resonances at the LHC and beyond", *JHEP* **01** (2019) 157, doi:10.1007/JHEP01(2019)157, arXiv:1810.08954.

[168] CMS Collaboration, "Search for a W' boson decaying to a vector-like quark and a top or bottom quark in the all-jets final state", *JHEP* **03** (2019) 127, doi:10.1007/JHEP03(2019)127, arXiv:1811.07010.

[169] N. Vignaroli, "New W' signals at the LHC", *Phys. Rev.* **D89** (2014), no. 9, 095027, doi:10.1103/PhysRevD.89.095027, arXiv:1404.5558.

[170] H. Han, L. Huang, T. Ma et al., "Six Top Messages of New Physics at the LHC", *JHEP* **10** (2019) 008, doi:10.1007/JHEP10(2019)008, arXiv:1812.11286.

[171] 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.

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

[173] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[174] 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.

[175] CMS Collaboration, "Jet algorithms performance in 13 TeV data", CMS-PAS-JME-16-003, (3, 2017).

[176] ATLAS Collaboration, "Search for low-mass resonances decaying into two jets and produced in association with a photon using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* **795** (2019) 56–75, doi:10.1016/j.physletb.2019.03.067, arXiv:1901.10917.

[177] ATLAS Collaboration, "Search for low-mass dijet resonances using trigger-level jets with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.* **121** (2018), no. 8, 081801, doi:10.1103/PhysRevLett.121.081801, arXiv:1804.03496.

[178] CMS Collaboration, "Search for narrow resonances in the b-tagged dijet mass spectrum in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. Lett.* **120** (2018), no. 20, 201801, doi:10.1103/PhysRevLett.120.201801, arXiv:1802.06149.

[179] CMS Collaboration, "Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles", *JHEP* **08** (2018) 130, doi:10.1007/JHEP08(2018)130, arXiv:1806.00843.

[180] CMS Collaboration, "Search for dijet resonances using events with three jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **805** (2020) 135448, doi:10.1016/j.physletb.2020.135448, arXiv:1911.03761.

[181] R. A. Fisher, "On the interpretation of χ^2 from contingency tables, and the calculation of p", *J. Roy. Statis.* **85** (1922) 87, doi:10.2307/2340521.

[182] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", CMS-PAS-LUM-17-004, (6, 2018).

[183] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", CMS-PAS-LUM-18-002, (5, 2019).

Acronyms

SM standard model

H Higgs

t top

$\mathcal{B}$ branching ratio

TBM token bit manager

SEU single event upset

TID total ionizing dose

LHC Large Hadron Collider

CMS Compact Muon Solenoid

ATLAS A Toroidal LHC ApparatuS

ALICE A Large Ion Collider Experiment

PU pileup

LV leading vertex

PF particle-flow

CHS charged hadron subtraction

PUPPI pileup per particle identification

BDT boosted decision tree

JER jet energy resolution

BEH Brout-Englert-Higgs

VLQ vector-like quarks

ROC receiver operating characteristic

ST single top

DY Drell-Yan

NLO next-to-leading order

LO leading order

PDF Parton Density Function

ML machine learning

CERN Conseil Européen pour la Recherche Nucléaire

ECAL electromagnetic calorimeter

HCAL hadronic calorimeter

DT drift tubes

RPC resistive plate chambers

CSC cathode strip chambers

CNN convolutional neural network

Danksagung

An dieser Stelle möchte ich mich bei allen ganz herzlich bedanken, die mich während der letzten Jahre bei meiner Promotion unterstützt haben. Allen voran möchte ich mich bei Andreas Hinzmann bedanken, der es mir nicht nur ermöglicht hat, in seiner Arbeitsgruppe zu promovieren, sondern mich auch bei allen meinen Vorhaben unterstützt, gefordert und gefördert hat. Nicht nur hat er mich bei den verschiedenen Datenanalysen unterstützt, sondern mir auch die Möglichkeit gegeben, während eines halbjährigen Aufenthalts am CERN lehrreiche Einblicke in die dortige Forschungsarbeit zu erlangen. Außerdem hat Andreas mich in den unterschiedlichsten Projekten zum Thema Gleichstellung unterstützt, wodurch unter Anderem die Physik-Projekt-Tage zu einem tollen Erfolg wurden.

Desweiteren möchte ich mich bei Johannes Haller für die Übernahme der Rolle des Zweitgutachters bedanken. Seine hilfreichen Kommentare und seine konstruktivie Kritik haben mich immer dazu motiviert das Beste zu geben. Elisabetta Gallo, Dieter Horns und Kai Schmidt-Hoberg möchte ich für ihre Mitgliedschaft in meiner Prüfungskommission danken.

Bei Roman Kogler möchte ich mich für seine umfassenden Ratschläge und Hilfe bedanken. Ich konnte mich immer auf ihn verlassen und möchte mich für die Unterstützung beim Aufbau der Physik-Projekt-Tage, bei der Suche nach dem single VLQ und vielem mehr bedanken! Die Zusammenarbeit hat immer Spaß gemacht und war eine Erfahrung, die meinen Arbeitsalltag bereichert hat.

Bei Melanie Eich, Dennis Schwarz, Anna Albrecht, Mareike Meyer, Arne Reimers, Irene Zoi und Andrea Malara möchte ich mich für die Unterstützung im Beruflichen sowie die gemeinsame Zeit im Privaten bedanken. Mit euch hat das Arbeiten sich nicht wie Arbeit angefühlt. Insbesondere das Backen mit dir Melanie werde ich nicht vergessen. Danke, dass du mich bei allem unterstützt hast und immer dabei warst! Und natürlich nicht zu vergessen mein Dank an Steffen Albrecht, der sich mit mir gerne durch technische Probleme gequält hat.

Auch bei Paul Schütze, Simon Kurz und Joscha Knolle möchte ich mich für unvergessliche Sommerschulen, Tanzstunden und Uno-Abende bedanken, zumindest sobald

Paul seine Karten fertig sortiert hat. ;)

Meiner Familie, ganz besonders meinen Eltern und Alexander Schmitzer, danke ich für ihre unermütliche Unterstützung in jeder Situation. Ich konnte mich immer auf Euch verlassen und auf euch zählen!

Zum Schluss möchte ich mich noch beim Pixel-Team (Klaas Padeken, Benedikt Vormwald, Jory Sonnefeld, Danek Kotlinski, Stefanos Leontsinis und vielen mehr) sowie bei Franko Greiner, Dietmar Block und Jochen Wilms bedanken. Bei euch allen habe ich so viel gelernt! Danke für die tollen Erfahrungen und Möglichkeiten, die ihr mir gegeben habt.