Search for Leptoquarks and Dark Matter in final states with top quarks at the CMS experiment

Dissertation zur Erlangung des Doktorgrades

an der Fakultät für

Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

Mareike Meyer

Hamburg

2017

Gutachter/innen der Dissertation:	Prof. Dr. Johannes Haller Prof. Dr. Elisabetta Gallo
Zusammensetzung der Prüfungskommission:	Prof. Dr. Johannes Haller Prof. Dr. Elisabetta Gallo Prof. Dr. Caren Hagner Prof. Dr. Gudrid Moortgat-Pick Dr. Christian Sander
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Caren Hagner
Datum der Disputation:	12. Juli 2017
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Michael Potthoff
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Abstract

This thesis presents two searches for physics beyond the Standard Model (SM). A search for pair production of third-generation leptoquarks decaying into a top quark and a tau lepton using pp collision data recorded at a center-of-mass energy of 8 TeV with the CMS experiment is presented. The search is based on an event selection requiring an isolated muon or electron candidate, one hadronically decaying tau lepton candidate and at least three jets. The main reducible background originates from processes, where a jet is misidentified as a tau lepton candidate. A measurement of the tau lepton misidentification rate is performed in a sideband enriched in $W+\text{jets}$ events. The transverse momentum distribution of the leading tau lepton candidate is used for the statistical interpretation of the result. No excess over the SM expectation is observed. Upper cross section limits on the pair production cross section of leptoquarks decaying into a top quark and a tau lepton are set. By combining the presented search with an analysis requiring same-sign muon-tau lepton-pairs, leptoquarks with masses below 685 GeV (695 GeV expected) are excluded at 95% C.L., assuming a branching ratio of 100% into a top quark and a tau lepton.

The second analysis presented in this thesis is a search for Dark Matter (DM) produced in association with a top quark pair using data collected at a center-of-mass energy of 13 TeV by the CMS experiment. The analysis targets final states in which the top quarks receive large transverse momenta by recoiling against the DM particles. Therefore, the applied event selection allows for non-isolated leptons and uses top tagging techniques to identify merged top quark decays. The normalizations of the main background processes, $t\bar{t}+\text{jets}$ and $W+\text{jets}$, are determined in data using control regions enriched in the respective process. For the final statistical interpretation the spectra of the missing transverse momentum in two signal regions are studied. Data and the SM background expectation agree within the uncertainties. Upper cross section limits are set on the production cross section of DM produced with a top quark pair assuming different DM particle and mediator masses, as well as scalar and pseudoscalar couplings of the mediators. Scalar mediators below 36 GeV are excluded at 95% C.L. if DM particle masses of 1 GeV and coupling strengths of one are assumed. In addition, a novel reconstruction of the neutrino four-momentum based on a likelihood is presented, which aims at discriminating between $t\bar{t}+\text{jets}$ in the lepton+jets decay channel and signal events.

Furthermore, this thesis presents studies on the impact of different pileup mitigation techniques on top tagging algorithms using simulated events at a center-of-mass energy of 13 TeV. Optimal parameter choices for top tagging algorithms using these techniques are derived.

Zusammenfassung

Diese Arbeit präsentiert zwei Suchen nach Physik jenseits des Standardmodells. Zunächst wird eine Suche nach Paarproduktion von Leptoquarks der dritten Generation, die in ein Top-Quark und ein Tau-Lepton zerfallen, vorgestellt. Dazu werden vom CMS Experiment bei einer Schwerpunktsenergie von 8 TeV erfasste pp-Kollisionsdaten analysiert. Die Suche basiert auf der Selektion von Ereignissen mit einem isolierten Myon- oder Elektronkandidaten, einem hadronisch zerfallenden Tau-Leptonkandidaten und mindestens drei Jets. Der größte reduzible Untergrund besteht aus Jets, die fälschlicherweise als Tau-Leptonen identifiziert werden. Die Tau-Lepton-Misidentifizierungsrate wird in einem Seitenband dominiert von $W+$ Jets-Ereignissen bestimmt. Die Transversalimpulsverteilung des führenden Tau-Leptons wird für die statistische Interpretation der Resultate verwendet. Kein Ereignisüberschuss über die Standardmodellerwartung wird beobachtet. Obere Ausschlussgrenzen auf den Wirkungsquerschnitt von Paarproduktion von Leptoquarks der dritten Generation, die in ein Top-Quark und ein Tau-Lepton zerfallen, werden bestimmt. Durch eine Kombination der Ergebnisse dieser Suche mit einer Suche im Kanal mit Myon-Tau-Paaren gleicher elektrischer Ladung, wird die Existenz von Leptoquarks mit Massen unter 685 GeV (695 GeV erwartet) mit einem Konfidenzniveau von 95% ausgeschlossen. Dabei wird ein Verzweigungsverhältnis von 100% für den Zerfall in ein Top-Quark und ein Tau-Lepton angenommen.

In der zweiten Analyse wird eine Suche nach Dunkler Materie (DM), die zusammen mit einem Top-Quark-Paar produziert wird, vorgestellt. Dafür werden Daten untersucht, die bei einer Schwerpunktsenergie von 13 TeV vom CMS Experiment aufgezeichnet wurden. Die Analyse konzentriert sich auf die Untersuchung von Ereignissen in denen die Top-Quarks einen hohen Transversalimpuls durch den Rückstoß zu den DM-Teilchen erhalten. Daher erlaubt die Ereignisselektion nicht-isolierte Leptonen und verwendet Top-Tagging-Techniken um kollimierte Top-Quark-Zerfälle zu identifizieren. Die Normierung der Hauptuntergründe, $t\bar{t}+$ Jets- und $W+$ Jets-Produktion, wird in Daten unter Verwendung von Kontrollregionen bestimmt, die von dem jeweiligen Prozess dominiert werden. Für die statistische Interpretation der Suchresultate wird das Spektrum des fehlenden Transversalimpulses in zwei verschiedenen Signalregionen verwendet. Die Daten und die Standardmodellerwartung für Untergrundereignisse stimmen innerhalb der Unsicherheiten überein. Obere Ausschlussgrenzen auf den Produktionswirkungsquerschnitt von DM, die mit einem Top-Quark-Paar erzeugt wird, werden unter der Annahme verschiedener Massen der DM-Teilchen und des Mediators als auch unter der Annahme von skalaren und pseudoskalaren Kopplungen des Mediators gesetzt. Skalare Mediatoren mit einer Masse kleiner als 36 GeV werden mit einem Konfidenzniveau von 95% ausgeschlossen, wenn für die Masse des DM-Teilchens 1 GeV und Kopplungsstärken von eins angenommen werden. Zusätzlich wird eine neue Methode zur Rekonstruktion des Neutrinoviererimpulses basierend auf einem Likelihood vorgestellt. Ziel dieser Methode ist es, $t\bar{t}+$ Jets-Ereignisse im Lepton+Jets-Zerfallskanal von Signalereignissen zu unterscheiden.

Außerdem werden in dieser Arbeit Studien zum Einfluss zwei verschiedener Methoden zur Verringerung der Pileup-Aktivität auf Top-Tagging-Algorithmen vorgestellt, basierend auf simulierten Ereignissen bei einer Schwerpunktsenergie von 13 TeV. Die Parameterwahl in Top-Tagging-Algorithmen wurde optimiert für die Verwendung dieser zwei Methoden zur Verringerung der Pileup-Aktivität.

List of publications

The CMS Collaboration, ”Jet algorithms performance in 13 TeV data” CMS Physics Analysis Summary CMS-PAS-JME-16-003, March 2017.

The CMS Collaboration, ”Search for third-generation scalar leptoquarks in the $\mathrm{t}\tau$ channel in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$” Journal of High Energy Physics 1507 (2015) 042, arXiv:1503.09049, [Erratum: JHEP11,056(2016)], doi:10.1007/JHEP11(2016)056, 10.1007/JHEP07(2015)042, July 2015.

The CMS Collaboration, ”Search for Third Generation Scalar Leptoquarks Decaying to Top Quark - Tau Lepton Pairs in pp Collisions,” CMS Physics Analysis Summary CMS-PAS-EXO-13-010, July 2014.

Contents

1 Introduction

The Standard Model (SM) of particle physics successfully describes known elementary particles and their interactions. Its validity has been confirmed in numerous measurements and searches performed until today. However, the SM leaves important questions unanswered. These unanswered questions indicate that physics beyond the SM (BSM physics) exists. Hence, many theories extending the SM have been proposed. One of the main goals of the *Large Hadron Collider (LHC)* and its experiments is to test the proposed theories and search for new phenomena not yet considered. This thesis presents two searches for BSM signatures that aim at addressing two of the main limitations of the SM. The analyzed data are pp collision data recorded by the CMS experiment at the LHC.

In the SM, quarks and leptons appear completely decoupled from each other, despite their similarities. This is remarkable since quarks carry an electric charge of exactly one or two thirds of the electric charge of charged leptons, which leads to electrically neutral atoms. This relation hints at a connection between leptons and quarks, which could be established by the presence of leptoquarks. Leptoquarks are bosons that couple to a quark and a lepton and thus mediate lepton-quark transitions. They are postulated by many theories beyond the SM like Grand Unified Theories or compositeness models. In certain models, leptoquarks can be realized at the TeV scale and are therefore interesting particles to search for at the LHC. Furthermore, leptoquark searches can be interpreted in the context of R-parity violating supersymmetric models, which are among the most common SM extensions. Searches for leptoquarks decaying into third-generation SM fermions are of particular interest since many BSM theories, especially models trying to explain tensions with the SM in B meson decays, predict leptoquark decays with enhanced couplings to third-generation SM fermions. This thesis presents a search for third-generation leptoquarks decaying into a top quark and a tau lepton using data corresponding to $19.7\,\mathrm{fb}^{-1}$ collected by the CMS experiment at the LHC. The data were recorded at a center-of-mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$ in the year 2012. The presented analysis is published in [1] and [2], and provides the most stringent exclusion limits of third-generation leptoquarks decaying into a top quark and a tau lepton to date.

The SM only describes roughly 5% of the energy content in our universe, while approximately 27% is attributed to an unknown kind of matter, referred to as Dark Matter (DM). Based on the assumption that interactions of DM particles χ with SM particles exist, DM should also be produced at the LHC. This thesis presents a search for DM produced in association with a top quark pair using data corresponding to $12.9\,\mathrm{fb}^{-1}$ collected with the CMS experiment at $\sqrt{s} = 13\,\mathrm{TeV}$ in the year 2016. The associated production of DM with a top quark pair is particularly interesting because of the frequent assumption of Yukawa-type couplings between the SM fermions and the particle mediating the interaction between SM and DM particles. Furthermore, in the $t\bar{t}+\chi\overline{\chi}$ channel also pseudoscalar mediators can be probed, which are not accessible by direct DM searches. Most importantly, direct searches have found no evidence for DM. However, these can only probe couplings between DM particles and light quarks, which possibly indicates dominating couplings to top quarks. In contrast to already existing results in the $t\bar{t}+\chi\overline{\chi}$ channel by the LHC experiments (e.g. [3–5]), which study resolved decays of the $t\bar{t}$ system, the analysis presented here focuses on boosted $t\bar{t}$ decays. In these decays, the final-state leptons are not isolated or the fully-hadronic decay of the top quark is not resolvable, but instead reconstructed within a large-cone jet. With this approach, the sensitivity at low mediator masses is enhanced compared to other approaches. Furthermore, a likelihood reconstruction of the neutrino in boosted lepton+jets $t\bar{t}$ events is discussed, which can potentially further improve the sensitivity of searches in the future.

The importance of identifying highly-boosted fully-hadronic top quark decays has been pointed out in many phenomenological studies. These techniques, referred to as top tagging, have been successfully used in a number of experimental analyses at the LHC, and play a major role in the $t\bar{t}+\chi\overline{\chi}$ search. Therefore, a stable performance of these substructure techniques for a large number of simultaneous pp interactions in one bunch crossing is of particular importance. Top tagging efficiencies and misidentification rates obtained after the application of different pileup mitigation techniques are compared within this thesis. Furthermore, existing top tagging algorithms are reviewed and updated working points are derived. The derived working points are now used by analyses using top tagging algorithms within the CMS experiment and are published in [6].

The thesis is organized as follows. In Chapter 2 phenomenological aspects of SM and BSM physics are introduced, with the main focus on leptoquarks and Dark Matter. Chapter 3 describes the LHC and the CMS experiment, followed by an introduction to event simulation with MC methods in Chapter 4. Afterwards, the reconstruction and identification of all particle candidates used in the experimental analyses is explained (see Chapter 5). Chapter 6 presents the search for third-generation leptoquarks decaying into a top quark and a tau lepton. Chapter 7 summarizes the studies on the impact of different pileup mitigation techniques on top tagging and the determination of top tagging working points. This is followed by a presentation of the search for Dark Matter produced with a top quark pair in Chapter 8. The thesis ends with a summary of its main results in Chapter 9.

Natural units ($\hbar = c = e = 1$) are used in this thesis.

2 Phenomenological aspects of Standard Model and Beyond Standard Model physics

The Standard Model (SM) of particle physics aims at a complete description of all known elementary particles and their interactions. So far, it has been extremely successful in describing the available experimental particle physics data and was validated to high precision. Additionally, predictions made by the SM could be verified experimentally. With the discovery of the Higgs boson predicted by the SM, the SM is regarded to be complete. However, the SM leaves some important questions unanswered. These open questions suggest that physics beyond the SM exists.

This chapter gives a short phenomenological introduction to the SM (Section 2.1) and lists its main limitations (Section 2.2). Section 2.3 shortly outlines some of the most common extensions of the SM. The main focus of the discussion of Beyond Standard Model (BSM) physics is put on leptoquarks (LQs) (Section 2.4) and Dark Matter (DM) (Section 2.5) since searches for these particles have been performed within this thesis.

2.1 The Standard Model of particle physics

The Standard Model [7] is mathematically described by a gauge invariant relativistic quantum field theory, based on the gauge group

$$U(1)_Y \otimes SU(2)_L \otimes SU(3)_C, \tag{2.1}$$

where the electromagnetic interaction is described by a $U(1)$ symmetry with associated quantum number Y, which is called hypercharge. The group $SU(2)_L$ describes the weak force acting on left-handed (L) fermions only. The strong force, whose associated charge is denoted as color C, is described by the group $SU(3)_C$. A sketch of the particle content of the SM and the interactions comprised in it is presented in Fig. 2.1. In the following, all these particles and interactions are characterized briefly.

The SM contains spin-$\frac{1}{2}$ fermions, spin-1 gauge bosons and the spin-0 Higgs boson. The fermions are the matter particles. They are divided into leptons and quarks, which are arranged in three generations each. In principle, each generation is a heavier copy of its preceding generation. While each of the lepton generations includes a negatively charged lepton and an associated electrically neutral neutrino, each quark generation comprises an up-type quark with electric charge $+\frac{2}{3}$ and a down-type quark with electric charge $-\frac{1}{3}$. The lepton generations are given by the electron e and the electron neutrino ν_e, the muon μ and the muon neutrino ν_μ as well as the tau lepton τ and the tau-lepton neutrino ν_τ. The quark generations are made up of the up (u) quark and the down (d) quark, the charm (c) quark and the strange (s) quark, as well as the bottom (b)

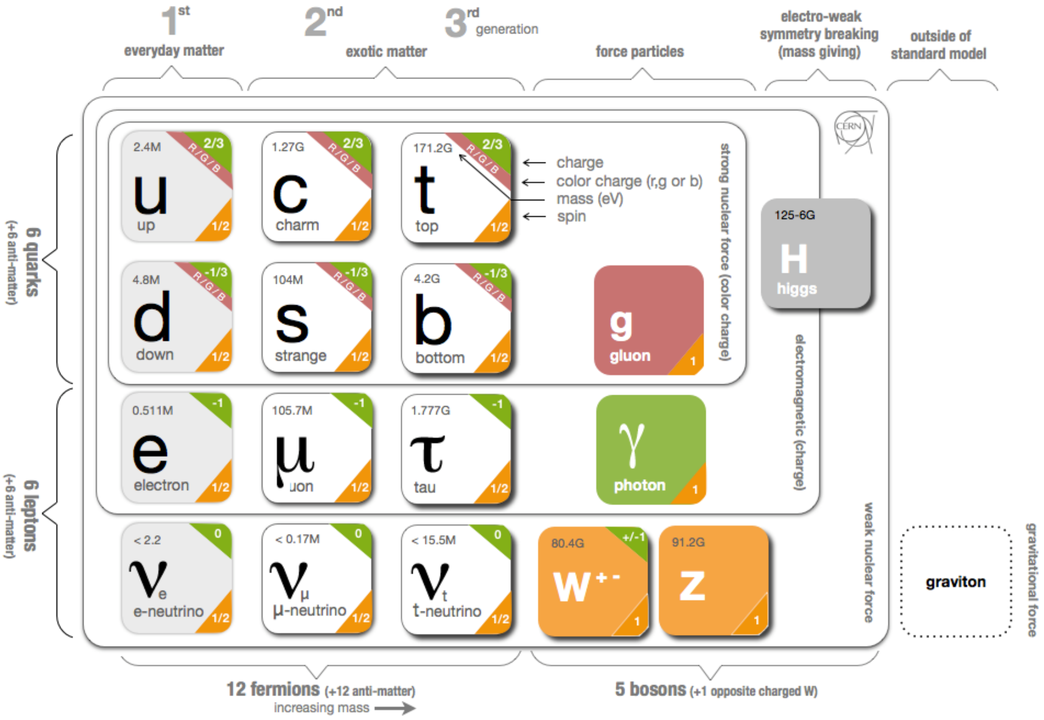


Figure 2.1: Sketch of the SM particles and their interactions [8]. Additionally, the mediator of the gravitational force, the graviton, is shown.

quark and the top (t) quark. In contrast to leptons, which are color-neutral, quarks carry color charge. All fermions are assigned a quantum number called weak isospin T_3. While left-handed fermions have a weak isospin of $\pm\frac{1}{2}$ and are arranged as weak isospin doublets within one fermion generation, right-handed fermions are weak isospin singlets with $T_3 = 0$. For each SM fermion there is an anti-particle, which has exactly the same mass but opposite charge than the associated particle.

Interactions between particles included in the SM are mediated by the exchange of spin-1 gauge bosons. The gauge bosons and the allowed interactions arise from gauge symmetries associated to the corresponding gauge groups. All colored objects contained in the SM are subject to the *strong force*, which is described by a quantum field theory called *Quantum Chromodynamics (QCD)*. The quantum number color has three representations, often referred to as *red, blue* and *green*. The gauge bosons associated to the strong force are eight massless gluons, which themselves carry color. Thus, gluons interact with other gluons. This has important consequences for the behavior of the coupling strength of the strong force as a function of the distance of two colored objects. Instead of decreasing, as it is the case for the coupling strength of the electromagnetic force, the coupling strength of the strong force increases with the distance. This means that the potential energy stored in the gluon field between two colored objects increases proportional to the distances between the two objects. Thus, colored objects cannot be separated from each other and appear only in color-neutral bound states, so called *hadrons*. The formation process of hadrons is called *hadronization*. One distinguishes between hadrons containing a quark-antiquark pair, denoted as *mesons*, and hadrons containing three quarks, denoted as *baryons*. Recently, evidence for the existence of a bound-state consisting of two quarks and two antiquarks (*tetraquarks*) and four quarks and an antiquark (*pentaquarks*), has been found by the LHCb experiment at the LHC [9–11]. However, the quantum numbers and the exact structure of these states are still under investigation. The described behavior of the strong coupling constant at large distances, and thus low momentum transfer, is referred to as *confinement*. Related to hadronization, the top quark differs from other quarks contained in the SM since its lifetime is so short that it decays before the hadronization process starts. At low distances, the strong coupling constant is small and the colored objects behave nearly as free particles, which is called *asymptotic freedom*.

The long-ranged *electromagnetic force* is mediated by the electrically neutral and massless photon, which interacts with all electrically charged particles contained in the SM. The underlying quantum field theory is the theory of *Quantum Electrodynamics (QED)*.

The *weak force* occurs as a neutral current arising by the exchange of a neutral Z boson and as a charged current arising from the exchange of a W^+ or W^- boson. While W bosons only couple to left-handed particles and right-handed antiparticles, the Z boson couples to particles of both chiralities, but with different strengths. Thus, the weak force violates parity (P) and charge-conjugation (C) symmetry. The ability of the Z boson to couple to both chiralities is explained within the unification of the electromagnetic and the weak force to the electroweak force, first described by Salam, Glashow, and Weinberg [12, 13]. In the unified theory, the Z bosons and photons are produced by rotating the representations of the gauge bosons W^0 and B^0 by the Weinberg angle θ_W. The W^0 boson and the $W^\pm$ bosons arise from the $SU(2)_L$ gauge group, while the B^0 is associated to the $U(1)_Y$ gauge group. From this unification, relations between the electromagnetic charge Q, the weak isospin T_3 and the hypercharge Y of a particle as well as a relation between the coupling strengths of the electromagnetic force e and the weak force g can be derived. These are given by

$$Q = T_3 + \frac{Y}{2} \tag{2.2}$$

and

$$e = g \sin\theta_W . \tag{2.3}$$

Charged leptons can be transformed into their corresponding neutrinos by the exchange of W bosons via the weak force. However, this is only possible within the same lepton generation, transitions between different lepton generations are forbidden. In the SM, this is implemented by introducing a conserved quantity for each lepton generation. These quantities are referred to as *lepton numbers* L_e, L_μ and L_τ, where e.g. $L_e = 1$ for electrons and electron-neutrinos, $L_e = -1$ for their anti-particles, and $L_e = 0$ for leptons from the other two generations. Each interaction allowed within the SM has to preserve all three lepton numbers. However, neutrino oscillations violating lepton number conservation have been observed. Neutrino oscillations occur since neutrinos are not massless and their mass eigenstates are not equal to their weak eigenstates, which results in a mixing of the different states. This is in contradiction with the SM, where neutrinos are massless. Also for quarks the mass eigenstates differ from the weak eigenstates. Within the SM, this is described in a basis where the weak eigenstates of the up-type quarks are equal to their mass eigenstates, while the weak eigenstates of the down-type quarks are obtained from the mass eigenstates by a rotation with the *CKM matrix*. The unitary CKM matrix can be parametrized by three mixing angles and a complex phase. According to the CKM matrix, transitions between quark generations are possible. The transition probability is given by the square of the entries of the CKM matrix. However, transitions between different quark generations are suppressed compared to transitions within one generation, since the diagonal elements of the CKM matrix have values close to unity, while off-diagonal elements are small. The phase contained in the CKM matrix is responsible for a small violation of CP conservation observed in weak interactions.

While the photons and the gluons are massless gauge-bosons, the W and the Z bosons have larges masses of $80.385 \pm 0.015\,\mathrm{GeV}$ and $91.1876 \pm 0.0021\,\mathrm{GeV}$ [7]. However, in the described formalism, mass terms from which the bosons could acquire masses are forbidden as these would not be invariant under local gauge symmetries. This problem is solved by introducing electroweak symmetry breaking via the *Brout-Englert-Higgs (BEH) mechanism*. The mechanism postulates a new scalar complex field Φ_H, which is a doublet in $SU(2)_L$ and contains a charged component Φ_H^+ and a neutral component Φ_H^0. Furthermore, a corresponding potential $V(\Phi_H)$ is postulated, which is given by

$$V(\Phi_H) = \mu^2|\Phi_H|^2 + \lambda|\Phi_H|^4. \tag{2.4}$$

It includes two free parameters μ and λ. If the parameters are chosen such that $\mu^2 < 0$ and $\lambda > 0$, a potential is created that features minima different from $\Phi_H = 0$. Thus, the symmetry of the potential and the formalism itself is spontaneously broken by the choice of a specific ground state, which corresponds to the *vacuum expectation value* υ. By choosing a ground state such that the vacuum is electrically neutral and by expanding the potential around the minimum υ, the quantized Higgs field is generated. Its excitations are the scalar SM Higgs bosons. Through couplings to Φ_H the Z and W bosons receive masses, given by

$$m_W = \frac{1}{2} g\upsilon, \tag{2.5}$$

and

$$m_Z = \frac{m_W}{\cos\theta_W}. \tag{2.6}$$

Since the Higgs boson is neutral, it does not couple to the photon such that photons remain massless. The Higgs boson couples to itself with a coupling strength proportional to λ. Its mass is given by

$$m_H = \sqrt{-2\mu^2}. \tag{2.7}$$

As the BEH mechanism does not yield a prediction for μ, the Higgs boson mass is a free parameter within the SM. The generation of fermion masses m_f is incorporated in the BEH mechanism, which yields a stronger coupling c_f for heavier fermions. Thus, fermion masses m_f are expressed by

$$m_f = c_f \frac{v}{\sqrt{2}}. \tag{2.8}$$

However, since the coupling constants c_f, called Yukawa couplings, are not predicted by the BEH mechanism, the mechanism only provides the means to include fermion masses into the SM, but does not predict their value.

2.2 Open questions within the Standard Model

Despite the great success of the SM in describing experimental particle physics data, important questions remain unanswered. In the following, some limitations of the SM are briefly summarized [7].

Gravity: The fourth known fundamental force, gravity, is not included in the SM. While the description of gravity is based on general relativity, the SM is described by an gauge invariant quantum field theory. So far, a quantized description of the gravitational force that is renormalizable could not be formulated.

Matter-antimatter asymmetry: In the big bang, matter and antimatter were produced in equal portions. Nonetheless, in today's universe, much more matter than antimatter is observed. According to [14], C and CP violation, baryon number violation[1] and a phase in that the universe expanded so fast that the interaction between matter and antimatter got out of equilibrium is needed to explain the observed asymmetry. While with inflation the last condition is satisfied, no baryon number violation has been observed so far. As stated before, CP conservation is violated in the weak interaction, which is described by the complex phase contained in the CKM matrix. However, the effect of the phase in the CKM matrix is too small to explain the observed matter-antimatter symmetry.

Charge quantization: The SM features a remarkable symmetry between the three generations of leptons and quarks, which manifests mainly in the electric charge. The electric charge of quarks is exactly one third or two thirds of the lepton charge making atoms electrically neutral. Despite this symmetry, quarks and leptons occur completely decoupled from each other in the SM.

Free parameters, mass hierarchy: As stated earlier, the BEH mechanism only provides the means to generate the masses of the fermions, but does not predict their magnitude. The same is true for other parameters e.g. the mixing angles of the quarks, the vacuum expectation value, the Higgs boson mass, the coupling strengths of the forces and other parameters, which are free within the SM. Additionally, the SM does not provide an explanation for the large differences in the SM fermion masses, which range from less than 2 eV for the neutrinos to 173 GeV for the top quark [7].

---

[1] In the SM, a baryon number of $B = 1$ is assigned to all baryons, $B = -1$ to all antibaryons and $B = 0$ to the mesons. Only baryon number conserving interactions are allowed.

Unification of forces: Although the electromagnetic and the weak force are described in a common model within the SM, the electroweak and the strong force cannot be unified. For unification, the running SM couplings have to meet at a high energy scale. At an energy scale of approximately 10^{15} GeV, the three coupling strengths of the SM are of the same order, but differences in the strength of the forces remain.

Hierarchy problem, fine-tuning: Since the Higgs boson is a scalar particle, the tree-level Higgs boson mass receives higher-order corrections that are, if studied at one-loop level, proportional to the energy scale Λ^2. The parameter Λ is the energy scale to which the SM is considered to be valid. Thus, under the assumptions that the SM can be extrapolated up to the Planck Scale ($M_{\text{Planck}} \approx 10^{19}$ GeV), where gravity becomes important and thus the validity of the SM is expected to break, the loop corrections get very large. This is known as the *hierarchy problem* since it occurs because of the different energy regimes of the known forces. Hence, to explain the measured Higgs boson mass of approximately 125 GeV within the SM, very specific fine-tuned loop level corrections are necessary. This is considered to be unnatural and referred to as *fine-tuning*.

Dark Matter (DM): Only approximately 5% of the energy content of our universe is made up of the particles contained in the SM, while roughly 27% is attributed to an unknown kind of matter, referred to as Dark Matter (DM) [15].

The existence of DM is mainly inferred from observations of stars, galaxies, galaxy clusters and other astrophysical objects, whose gravitational interactions with other objects cannot be explained solely with the visible matter [16]. Examples are measurements of galaxy rotation curves. Under the assumption that only the visible matter in the disk of the galaxy exists, the circular velocities of gas and stars around the center of the galaxy is expected to decrease proportional to $\frac{1}{\sqrt{r}}$ at large distances r to the center of the galaxy. However, the measured velocities stay roughly constant at large distances (see Fig. 2.2, where the rotation curve of NGC 6503 is presented). This behavior is only explainable if an halo of gravitational interacting not visible particles exits, whose enclosed mass increases proportional with the distance. Additionally, measurements of the temperature of galaxy clusters in relation to their mass, which is e.g. determined by gravitational lensing effects, provide evidence for DM. From these and other observations (see [16]), only the existence of DM can be inferred, its total amount cannot be determined. Measurements of the total amount of DM are performed by studying the *cosmic microwave background (CMB)* [7, 15, 16]. The CMB is the redshifted black body radiation originating from the very early universe, when the temperature became too low to keep matter ionized such that the universe got transparent for photons. In general the CMB is isotropic, but tiny temperature fluctuations of the order of 10^{-5} are observed, which are associated with density fluctuations that lead to structure formation. By performing a multipole expansion in spherical harmonics of the measured CMB, the total amount of DM can be determined. This measured relic DM density cannot be explained by any SM particle. Despite the fact that

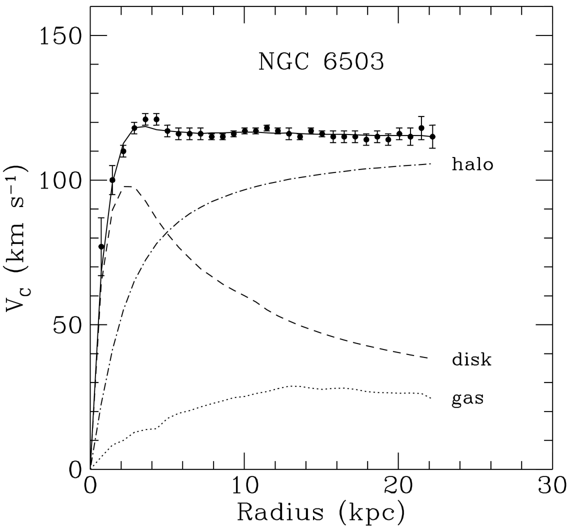


Figure 2.2: Rotation curve of NGC 6503 [17]. Taken from [16].

neutrinos only interact very weakly, are neutral, and have masses, their mass is too small to make up the measured relic DM density. Additionally, if neutrinos were DM particles, the universe would have formed differently than observed [16]. Attempts to explain the described gravitational interactions without the existence of DM, e.g. by modifying Newton's laws, only success in explaining phenomena inside galaxies, but not at larger scales [18].
In conclusion, the existence of DM can be called well established. However, the SM does not provide a DM candidate that can account for the whole measured density and thus astrophysical observations that hint at the existence of DM provide strong evidence for physics beyond the SM.

2.3 Beyond Standard Model theories

To address the limitations and unanswered questions of the Standard Model many extensions have been proposed. Some of the most common theories going beyond the Standard Model are shortly introduced in this section.

Grand Unified Theories (GUTs): The similarity between quarks and leptons in the SM hints at the existence of a description of these particles within common representations of larger gauge groups. GUTs [7, 19] aim at embedding the SM particles and their interactions into a larger symmetry group to unify the SM forces at an energy scale called *GUT scale*. At this scale the GUT symmetry is spontaneously broken to the SM gauge groups by a GUT Higgs field. The simplest possibilities to arrange all SM fermions into common multiplets arise in gauge groups based on $SU(4)$, firstly proposed by J.C. Pati and A. Salam [20], and in gauge groups based on $SU(5)$, established by H. Georgi and S.L. Glashow [21]. In the Pati-Salam $SU(4)$ model, a common representation of quarks and leptons is achieved by expressing the lepton number as the fourth color. However, in this model the three gauge couplings still appear independently from each other and cannot be unified. Contrary, $SU(5)$ models predict one common gauge coupling at the GUT scale, where the forces are unified. Still, in the minimal $SU(5)$, the SM gauge couplings do not unify. This limitation together with the fact that $SU(5)$ models predict proton lifetimes smaller than the experimentally determined lower limit on the proton lifetime exclude the minimal $SU(5)$ model. Nonetheless, in conjunction with e.g. supersymmetry an $SU(5)$ model is still realizable and provides unification of the forces. More complex GUTs e.g. propose models based on $SO(10)$ and $E(6)$.

GUTs can answer many of the questions not answered within the SM. Besides describing quarks and leptons in a common representation and unifying the forces, GUTs provide explanations for the fractional electric charges of the quarks by establishing relationships between lepton and quarks charges that arise from their common representation. Moreover, GUTs make, amongst others, predictions for free parameters in the SM and incorporate C, CP and baryon number violation.

Compositeness models: Another idea to explain the similarity of quarks and leptons as well as between the fermion generations in the SM is established by compositeness models [22, 23]. These state that the SM fermions, and in some models also the massive SM gauge bosons as well as the Higgs boson, are not fundamental, but comprised out of new fundamental particles. These are often referred to as *preons*. Preons are described within larger gauge groups than the SM and get assigned a new quantum number, often denoted as *hypercolor*. This hypercolor has the same properties as the color associated to the SM $SU(3)$ gauge group and thus binds preons to SM fermions through confinement. In order to construct colored quarks, preons have to carry $SU(3)$ color, or color has to be generated by certain combinations of preons. Since in compositeness models quarks and leptons are made of the same particles and the second and third SM fermion generation are described as excited states of the first generation, the number of free parameters the SM exhibits is drastically reduced in compositeness theories. Matter-antimatter symmetry can be constructed if expressed in terms of preons. Most notably, the common substructure of quarks and leptons explains the relation of the quark and lepton charges.

Supersymmetry (SUSY): One of the most discussed theories beyond the SM is supersymmetry (SUSY) [24, 25]. In supersymmetric theories, a symmetry between the bosons and fermions contained in the SM is introduced by adding a bosonic partner for each SM fermion, called *sfermion*, and a fermionic partner for each boson, called *bosonino*. Thus, each SM particle receives a partner with exactly the same quantum numbers except for the spin, which differs by $\frac{1}{2}$. If SUSY were an exact symmetry, the masses of the superpartners were exactly the same as those of their SM counterparts. However, these particles have not been detected experimentally. Thus, SUSY must be a broken symmetry. In order to prohibit lepton number L and baryon number B violating processes, which e.g. would allow, in contradiction to experiments, proton decays, a new multiplicative quantum number R, referred to as *R parity*, is introduced in SUSY. It is defined by

$$R = (-1)^{3(B-L)+2S}, \tag{2.9}$$

where S describes the spin of the particle. Hence, SM particles receive $R = +1$, while SUSY particles, called *sparticles*, receive $R = -1$. Most SUSY models impose R-parity conservation. This has important consequences, which are that SUSY particles can only be produced in pairs and that the lightest supersymmetric particle (LSP) is stable, which makes the LSP a very suitable DM candidate. However, there are also R-parity violating models being thought of. In these models, proton decay is prevented by requiring one of the involved couplings to be very small.

Since loop processes involving bosons are signed oppositely than loop processes involving fermions, contributions to the Higgs boson mass arising from higher-order corrections would cancel exactly if SUSY were an exact symmetry. Nonetheless, the corrections cancel to some degree if SUSY particles have masses around the TeV scale. Sparticles with TeV-scale masses also modify the evolution of the coupling constants of the SM interactions such that they unify at a scale of 10^{16} GeV. Additionally, since the existence of SUSY is a condition for the validity of many string theories, which in turn allow a unification of the SM interactions with the gravitational force, and SUSY also allows for explanation of the observed matter-antimatter asymmetry (see e.g. [26]), SUSY is a very attractive theory extending the SM.

2.4 Leptoquarks

All discussed extensions of the SM, amongst others, postulate the existence of *leptoquarks (LQs)* [27, 28], which are introduced as bosons that couple to a quark and a lepton. Leptoquarks are especially interesting since they would establish a connection between the quark and lepton sector of the SM. As stated before, quarks and leptons exhibit a remarkable symmetry concerning their charges and occurrence in three generations in the SM, while appearing completely decoupled from each other. If a leptoquark was found, it would hint to a more fundamental symmetry between the fermions in the SM and could help to establish unified interactions between SM particles.

2.4.1 Properties of leptoquarks

Leptoquarks can occur as scalar or vector bosons[2] that carry color and electric charges of $\pm\frac{1}{3}$, $\pm\frac{2}{3}$,$\pm\frac{4}{3}$ of $\pm\frac{5}{3}$, depending on the representation. Leptoquarks are most commonly described in a model first introduced by Buchmüller, Rückl and Wyler [27], referred to as *BRW model* in the following. The BRW model introduces renormalizable and under the SM gauge group invariant leptoquark-lepton-quark interactions. Each leptoquark gets assigned a lepton number L and a baryon number B. Only interactions between SM

---

[2] The focus here is on scalar leptoquarks.

S	F	$SU(3)_C$	$SU(2)_W$	$U(1)_Y$	allowed coupling
0	-2	$\bar{3}$	1	1/3	$q_L \ell_L$, $u_R e_R$
0	-2	$\bar{3}$	1	4/3	$d_R e_R$
0	-2	$\bar{3}$	3	1/3	$q_L \ell_L$
0	0	3	2	7/6	$\bar{q}_L e_R$,$\bar{u}_R \ell_L$
0	0	3	2	1/6	$\bar{d}_R \ell_L$
1	-2	$\bar{3}$	2	5/6	$q_L \gamma^\mu e_R$,$d_R \gamma^\mu \ell_L$
1	-2	$\bar{3}$	2	$-1/6$	$u_R \gamma^\mu \ell_L$
1	0	3	1	2/3	$\bar{q}_L \gamma^\mu \ell_L$,$\bar{d}_R \gamma^\mu e_R$
1	0	3	1	5/3	$\bar{u}_R \gamma^\mu e_R$
1	0	3	3	2/3	$\bar{q}_L \gamma^\mu \ell_L$

Table 2.1: Possible leptoquark states according to the BRW model [27]. The number F is defined as $F = 3B + L$. The absolute value of the electric charge $|Q|$ can be derived from the given numbers with $|Q| = |T_3 + Y|$. The numbers in the third and fourth row give the dimension of the LQ representation under the given symmetry group. Left-handed fermion doublets are described by q_L and ℓ_L, while e_R, u_R and d_R are right-handed electron, up- and down-type quark singlets. An adapted version of the table in [7] is presented.

fermions and leptoquarks that conserve both L and B are allowed, which prohibits proton decay. As a result, theories based on the BRW model can comprise leptoquarks with masses around the TeV scale without violating constraints on the proton lifetime [27]. A summary of the states allowed within the BRW model, the assigned quantum numbers and the couplings is presented in Table 2.1. If right-handed neutrinos were added to the SM, additional leptoquark states could be constructed [28]. In order to avoid stringent bounds on lepton-quark transitions arising e.g. from searches for FCNC and precision measurements of rare decays (see [29, 30] for summaries of these constraints), one type of leptoquark is allowed to couple to exactly one lepton and one quark generation. Thus, three generations of leptoquarks are constructed, which are called first-, second- or third-generation leptoquark depending on the generation of the lepton the leptoquark decays in. Since additionally tight constraints have been set on particles coupling to left- and right-handed fermions at the same time [30], only *chiral leptoquarks* that couple exclusively to one handedness at a time are in general studied.

Models that deviate from the assumptions made within the BRW are considered as well, e.g. models allowing couplings of one leptoquark to different fermion generations [31]. However, in these models mechanisms that prevent contradictions with the above mentioned stringent bounds have to be found. Thus, models allowing off-diagonal flavor decays often include couplings to top quarks since these are heavy and do not build bound states, such that constraints from rare processes and FCNC are often not applicable. Most recently, leptoquarks decaying into a top quark and different charged leptons gained interest due to anomalies observed in rare B meson decays, which are discussed below (see Section 2.4.2).

In this thesis, a search for third-generation leptoquarks decaying into a top quark and a tau lepton is presented. The studied leptoquarks have a charge of $-\frac{1}{3}$ and are described by the BRW model. They correspond to the LQ state listed in the first row in Table 2.1.

Thus, they can decay via $LQ \rightarrow t\tau^-$ or $LQ \rightarrow \nu_\tau b$. However, since for the main result of this thesis, only the decay into a top quark and a tau lepton is studied, the obtained result can also be re-interpreted in models assuming mixed decay modes by scaling the branching ratio into a top quark and a tau lepton accordingly.

2.4.2 Realization in BSM theories

Except for supersymmetric models, the mentioned extensions of the SM are formulated as theories based on gauge groups $SU(n)$. In a theory based on $SU(n)$, $n^2 - 1$ gauge bosons exits. Thus, new bosons are postulated since the gauge groups are larger than the SM gauge group in all cases. In these theories quarks and leptons are arranged in common multiplets, interactions that mediate transitions between these fermions occur. Hence, all these models postulate leptoquarks.

Because most GUTs are broken at the GUT scale to the SM gauge group, the predicted leptoquarks have GUT scale masses [29]. However, in certain cases, leptoquarks can receive masses of O(TeV) also in GUT theories, where the simplest ones are based on $SU(4)$ [20] and $SU(5)$ [21]. Leptoquarks arising in Pati-Salam GUTs are generally very heavy [20]. The same is true for the Georgi-Glashow $SU(5)$ model, which predicts vector leptoquarks with masses around the GUT scale [21]. In [32] a model based on $SU(5)$ is proposed, where only third-generation SM fermions violate B and L conservation, such that the model is not in contradiction with results from FCNC, measurements of rare decays and lower limits on the proton lifetime. In this model, leptoquarks decaying into third-generation SM fermions are predicted with masses within the reach of the LHC. Leptoquarks decaying predominantly into $t\tau$ or $b\nu$ with masses detectable at the LHC are also predicted by [33], which describes a supersymmetric GUT theory based on $SU(5)$.

In compositeness models, leptons and quarks are comprised of a common set of preons. Therefore, transitions between them and thus leptoquarks are predicted. In [22] a compositeness model based on the Pati-Salam $SU(4)$ symmetry group is discussed, where the $SU(4)$ group is spontaneously broken to the SM at the hypercolor confinement scale, which lies at a scale of 250 GeV. Thus, low-scale leptoquark masses are predicted. In [34], models with fundamental and composite fermions are studied. These postulate leptoquarks with masses of several hundred GeV, which decay predominantly into third-generation SM fermions.

As already mentioned, recently, leptoquarks gained interest due to anomalies observed in decays of B mesons. Examples are presented in [31, 35]. In [35] the rate of lepton flavor universality violation measured in the interactions $\bar{B} \rightarrow \bar{K}\ell^+\ell^-$, the muon anomalous magnetic moment and the increased decay rate of the interaction $\bar{B} \rightarrow D^{(*)}\tau\bar{\nu}$ compared to the SM prediction are explained with a leptoquark with mass of around one TeV that couples to second- and third-generation fermions. Similar leptoquark characteristics are predicted in [31], where a composite Higgs model is introduced that aims at explaining the tension with the SM prediction measured in the ratio of $B^+ \rightarrow K^+\mu^+\mu^-$ to $B^+ \rightarrow K^+e^+e^-$ decays.

R-parity violating supersymmetric models predict squarks decaying into a quark and a

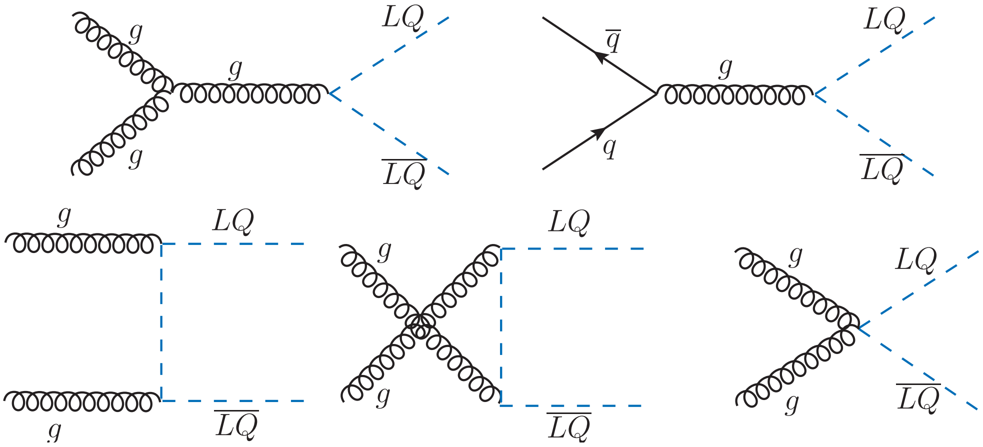


Figure 2.3: Dominant leading-order Feynman diagrams of third-generation leptoquark pair production at the LHC.

lepton of the same generation [36]. Thus, searches for leptoquarks can also be interpreted in this subgroup of supersymmetric models.

2.4.3 Leptoquark production at the LHC

The LHC provides an excellent opportunity to search for leptoquarks due to its large center-of-mass energy. At the LHC, leptoquarks are produced predominantly in pairs. Single production of leptoquarks at the LHC is possible as well, but highly suppressed for third-generation leptoquarks since it requires third-generation quarks in the initial state. Therefore, single production is not considered in the presented analysis. The dominant pair production modes at the LHC are leptoquark production through gluon-gluon fusion processes or quark-antiquark annihilation [37]. While gluon-gluon fusion is the dominant production mode for small leptoquark masses, quark-antiquark annihilation processes become more important with higher leptoquark masses. The leading-order Feynman diagrams of these processes are depicted in Fig. 2.3. In principle, lepton-mediated t(u)-channel pair production of leptoquarks is also possible. However, this process is again unlikely to happen at the LHC due to third-generation quarks needed in the initial state. Apart from the lepton-mediated t(u)-channel production, which also depends on the lepton-quark-leptoquark coupling, the pair production cross section is determined by SM gauge interactions and solely dependent on the mass of the leptoquark, the LQ spin and the center-of-mass energy [37]. Since the t(u)-channel production is suppressed at the LHC, the dependence of the cross section on the unknown coupling can be neglected. At leading order the parton pair production cross sections of scalar leptoquarks are given by [37]

$$\hat{\sigma}_{\text{LO}}^{q\bar{q}} = \frac{2\alpha_s^2\pi}{27\hat{s}}\beta^3 \tag{2.10}$$

and

$$\hat{\sigma}_{\text{LO}}^{gg} = \frac{\alpha_s^2\pi}{96\hat{s}}\left[\beta(41-31\beta^2) + (18\beta^2-\beta^4-17)\log\frac{1+\beta}{1-\beta}\right], \tag{2.11}$$

generation	channel	experiment	$\sqrt{s}$ [TeV]	L_{int} [fb^{-1}]	mass limit [GeV]	
1st	$eejj$	CMS	13	2.6	1130	[38]
1st	$eejj$	ATLAS	13	3.2	1100	[39]
1st	eej	CMS	8	19.6	1755	[40]
2nd	$\mu\mu jj$	CMS	13	2.7	1165	[41]
2nd	$\mu\mu jj$	ATLAS	13	3.2	1050	[39]
2nd	$\mu\mu j$	CMS	8	19.6	660	[40]
3rd	$bb\tau\tau$	CMS	13	12.9	900	[42]
3rd	$bb\tau\tau$	ATLAS	7	4.7	534	[43]
3rd	$bb\nu\nu$	CMS	7	4.7	450	[44]
3rd	$bb\nu\nu$	ATLAS	8	20.1	625	[45]
3rd	$tt\nu\nu$	CMS	8	19.7	660	[46]
3rd	$tt\nu\nu$	ATLAS	8	20.3	$(200, 640)$	[45]

Table 2.2: Summary of searches for scalar leptoquarks performed at the LHC. Presented are the most sensitive available analyses in the respective channel to date. Shown are the LQ generation, the channel the analysis has been conducted in, the experiment at that the analysis has been performed, the center-of-mass energy $\sqrt{s}$, the integrated luminosity L_{int} the studied data correspond to, and the derived mass exclusion limit. All results are given at 95% confidence level (C.L.). In the two presented searches for single production of first- and second-generation leptoquarks (third row and sixth row) the lepton-quark-leptoquark coupling has been assumed to be 1.

with $\beta = \sqrt{1 - 4\frac{M_{\text{LQ}}^2}{\hat{s}}}$. The value $\hat{s}$ describes the squared center-of-mass energy of the parton process, α_s is the strong coupling constant, and M_{LQ} the mass of the leptoquark. The cross section has been calculated to next-to-leading order precision in [37]. At $\sqrt{s} = 8\,\text{TeV}$, the cross sections for pair produced scalar leptoquarks range between 17.4 pb for leptoquarks with masses of 200 GeV and 2.69 fb for leptoquarks with masses of 800 GeV [37], and are well within the reach of the LHC.

2.4.4 Recent search results

Direct searches for pair or singly produced scalar leptoquarks have been performed at LEP, HERA, Tevatron and the LHC in a variety of different channels. However, none of these searches indicated the existence of leptoquarks. Exclusion limits on the cross sections of leptoquark production and the masses of leptoquarks were derived. The most stringent direct limits on the masses and production cross sections of scalar leptoquarks to date are placed by the ATLAS and CMS experiments at the LHC. These searches and their results are summarized in Table 2.2. A summary of the results of the analyses performed at LEP, HERA and Tevatron can be found e.g. in [47]. Searches for pair produced third-generation leptoquarks, like the one presented in Chapter 6, have been performed in the channels LQ $\rightarrow b\tau$ [42,43], LQ $\rightarrow b\nu$ [44,45] and LQ $\rightarrow t\nu$ [45,46] by the LHC experiments. The analysis presented in this thesis is a search for pair production of third-generation scalar leptoquarks decaying into a top quark and a tau lepton. It has been published in [2].

2.5 Dark Matter

From the astrophysical observations providing evidence for DM (see Section 2.2) properties that a DM candidate should have are inferred [7, 16, 18]. The most important properties are:

- DM particles should be stable and account for the observed relic density.
- Interactions of DM with SM particles, except for the gravitational force, have to be unlikely if present. In most theories, DM particles are electrically neutral.
- From the CMB measurements, it is known that the larger amount of DM should be non-baryonic.
- DM is supposed to be *cold*, which means that at the time galaxies formed it was non-relativistic. Hot (i.e. relativistic) DM would have lead to the formation of firstly big than small structures, which is in contradiction with observations in the universe.
- Since DM forms large halos, DM particles should, in great parts, not be able to cool by radiation of energy since otherwise DM would e.g. clump to disks in galaxies as baryonic matter does.
- From gravitational lensing observations of two galaxies permeating each other, it is known that self-interactions of DM should also be weak.

Various DM candidates have been proposed. Among those are sterile neutrinos, primordial black holes, axions and WIMPs. However, multiple particle types could account for the total DM content, e.g. it is believed that a small part of DM is made up by SM neutrinos [16]. Since in the analysis presented in Chapter 8 a search for WIMPs is performed, these particles are discussed in the following. A detailed review of other DM candidates can be found e.g. in [7, 16].

WIMPs (weakly interacting massive particles) are particles with masses around the electroweak scale and coupling strength to SM particles similar to the weak force [7, 16, 18]. They are strongly motivated by the so called *WIMP Miracle*. It is assumed that the DM particles were in thermodynamic equilibrium with SM particles in the early universe. As the universe expanded, the temperature dropped and the production rates became smaller than the annihilation rates (SM particles are assumed to be lighter than DM particles). At some point, the annihilation rate of DM particles also drops below the expansion rate. This way, the interaction of DM and SM particles got out of equilibrium and the DM particles decoupled from the thermal bath of particles. Since the temperature T_F at which decoupling of DM particles happened is given by $T_F \approx m_\chi/20$, where m_χ is the mass of the DM particles, WIMPs classify as cold DM [7]. Assuming an interaction cross section and a mass, the relic density that is observed today can be calculated in the standard cosmological model. It comes out that the measured relic density can be achieved by assuming DM particles with masses around 10 GeV and a few TeV and cross sections in the range of the weak force [7]. This is referred to as the WIMP Miracle and strongly motivates to search for WIMPs at the LHC. The most discussed particle classifying as a WIMP that occurs in a complete model extending the SM, is the LSP in R-parity conserving supersymmetric theories.

2.5.1 Direct, indirect and collider searches for WIMPs

Although only gravitational interactions of DM particles with the known matter are thought to be observed so far, it is considered as certain that interactions between DM and SM particles exist. These interactions are searched for by different complementary experiments, split up into three main groups: *collider searches, direct detection and indirect detection experiments* [16, 48, 49], depending on the type of interaction the experiment searches for. Direct detection experiments search for DM particles scattering with the detector material. The interaction is detected, depending on the detector material, by scintillation light, ionization processes, or phonons produced by the interaction. Indirect detection experiments search for highly-energetic SM particles, in most cases antiparticles, photons or neutrinos, originating from annihilation processes of DM particles. A few deviations from the expectation have been observed by some indirect detection experiments. However, most of these signals are also consistent with signals arising from other astrophysical sources than DM. Hence, both direct and indirect detection experiments have not found any evidence for DM so far.

If interactions of SM and DM particles exist, DM is also produced at the LHC. Especially, if DM particles occur as WIMPs, the production cross section of DM should be large enough such that these interactions are detectable. However, WIMPs are long-lived and weakly-interacting. They pass through the LHC detectors without interaction with the detector material. Thus, DM particles produced in pp collisions at the LHC can only be detected by measuring an energy imbalance in the final state, expressed by the missing transverse energy $\not{E}_T$, and, at the same time, by the presence of a SM particle indicating that an interaction occurred. These SM particles, which are used to tag the interaction, are either produced in initial state radiation or in association with the DM particles. Searches for these kind of interactions are performed by the CMS and ATLAS experiments at the LHC exploiting multiple signatures (like e.g. monojet searches). None of the collider searches performed so far has found any evidence for the existence of DM particles. Thus, lower bounds on their masses and the masses of the mediators have been derived. A summary of the searches performed at the CMS experiment and the mass exclusion limits they set are presented in Fig. 2.4. All these searches have been performed using pp collision data at $\sqrt{s} = 13\,\text{TeV}$. The ATLAS experiment has a very similar search program.

The three described detection approaches provide quite complementary results [16, 49]. However, it has to be noted that comparisons of the results between the different type of experiments can be complicated due to different assumptions made (e.g. couplings) and large uncertainties in the respective process (e.g. nuclear form factors, the local DM density and the DM velocity distribution). In general, direct detection experiments are very sensitive to spin-independent interactions (corresponding to exchanges

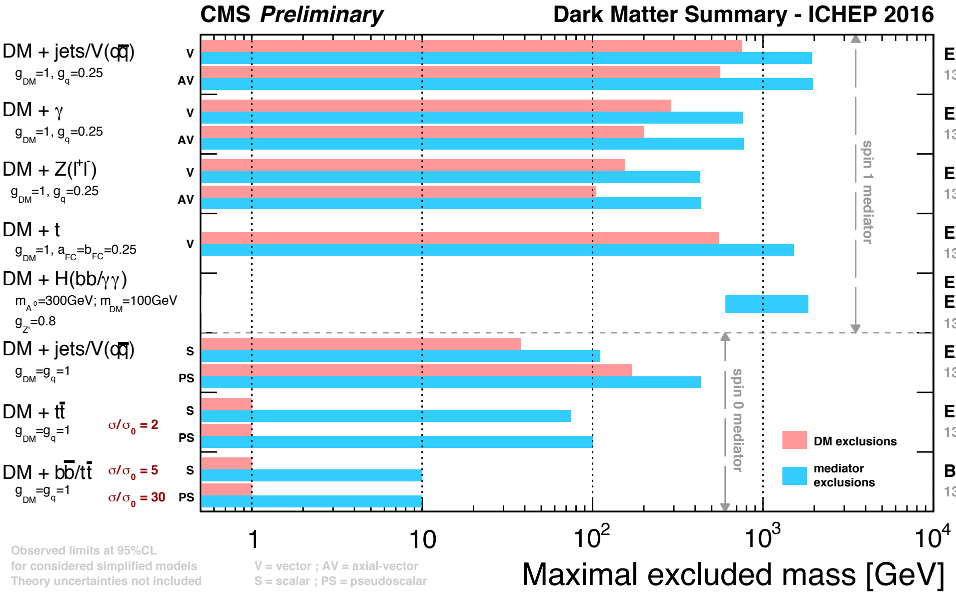


Figure 2.4: Summary of observed exclusion limits at 95% C.L. on the mediator and DM masses set by different searches performed at the CMS experiment. The figure is taken from [50].

of scalar or vector particles mediating the interaction between DM and SM particles), while LHC searches often outperform direct detection experiments for spin-dependent interactions (corresponding to exchanges of a axial-vector or pseudoscalar mediator). Since for low DM masses the momentum transfer in the scattering process is too small to be detectable, collider searches exhibit better sensitivity than direct searches at low DM masses. Contrary, for high DM masses direct detection experiments in general perform better. Most interestingly, pseudoscalar interactions cannot be probed by direct detection experiments and are therefore interesting interactions to search for at the LHC. However, indirect experiments do have sensitivity to this interaction, but are difficult to compare to LHC results [49]. Indirect experiments achieve comparable sensitivities for spin-dependent interactions by studying neutrinos produced by DM annihilation in the sun. Additionally, while direct and collider searches are less sensitive to DM exclusively coupling to leptons, these interactions are detectable by indirect searches.

2.5.2 Interpretation in simplified models

In principal, different approaches can be pursued for the theoretical interpretation of DM searches [51]. One the one hand, complete models, like supersymmetric theories, which are designed to solve the limitations of the SM and to provide a consistent extension of the SM, can be studied. However, these models are often quite complicated and specific assumptions are made. As so far only little is known about DM, a second approach is to use models that are as simple and general as possible such that only a few assumptions have to be made and the results can be re-interpreted in complex and complete models. The second approach is currently adopted in most searches performed for DM at the LHC and also pursued in the search presented in this thesis. The studied models are simple in the sense that they only assume exactly one type of DM particle χ and one interaction with SM particles, although the DM sector could contain multiple particles and interactions. It is assumed that other particles making up the DM sector are too heavy to be accessible by the performed search.

In initial searches performed at the LHC, it has been assumed that the particle mediating the interaction between SM and DM particles is very heavy compared to the momentum transfer in the interaction. Thus, interactions between DM and SM particles were interpreted as an contact interaction in an *Effective Field Theory (EFT)* with only two free parameters (the mass of the DM particle and the effective suppression scale). In several DM searches (e.g. [52]) performed at $\sqrt{s} = 8\,\mathrm{TeV}$, it turned out that LHC analyses exhibit sensitivity also in phase spaces were EFTs are not valid, meaning in regions were the momentum transfer is comparable to the mass of the particle mediating the interactions and in regions where the couplings are too large such that the theory is not longer perturbatively calculable. Additionally, for small coupling strengths of the SM particles to DM, EFTs do not correctly re-produce the kinematical distributions predicted by more complete theories.

Instead of using EFTs, searches performed at $\sqrt{s} = 13\,\mathrm{TeV}$ are now interpreted adopting *simplified models* [49, 51, 53], which explicitly predict a particle Φ that mediates the interaction between the SM and DM particles and thus avoid the limitations of the EFTs. In simplified models only interactions that have dimension four or less, thus that are renormalizable, are studied. Simplified models have the advantage that they are also valid for models with low mediator masses, which are favored by measurements of the relic density [54], and provide sensitivity also for models predicting softer $\not{E}_T$ spectra than EFTs. The simplified model used for the analysis presented in this thesis follows the recommendations of the LHC Dark Matter Forum given in [49, 53]. These simplified models contain four free parameters, the mass of the particle mediating the interaction M_Φ, the mass M_χ of the DM particles χ, the coupling strength g_{DM} of the DM particles to the mediator and the coupling strength g_{SM} of the SM particles to the mediator. In Fig. 2.5 a Feynman diagram of $t\bar{t} + \chi\overline{\chi}$ production in the simplified model framework, which is searched for in this thesis, is presented. Scalar mediators and pseudoscalar mediators are considered. As recommended in [55], mixing of the scalar mediators with the SM Higgs boson is not taken into account. For the mediators only couplings to SM quarks and one kind of DM particle are considered. The resulting widths of the mediators are determined with the calculation performed in [56]. The DM particles are assumed to be Dirac particles with spin $\frac{1}{2}$, which do neither decay nor interact within the detector. SM Yukawa couplings are assumed for the SM quark-mediator couplings, while the coupling of the mediator to the DM particles is fixed to one.

While simplified models provide a simple and general framework to search for DM, which enables comparisons to direct and indirect experiments as well as to the measured relic density, in specific cases limitations of the models have been identified (see [51] for a discussion of some of these). Concerning the $t\bar{t} + \chi\overline{\chi}$ channel, these include that the studied coupling of the $t\bar{t}$-pair to the scalar mediator is not gauge invariant. Studies aiming to improve this by adding mixing of the mediator with the SM Higgs boson to the simplified models are currently ongoing (e.g. [55]).

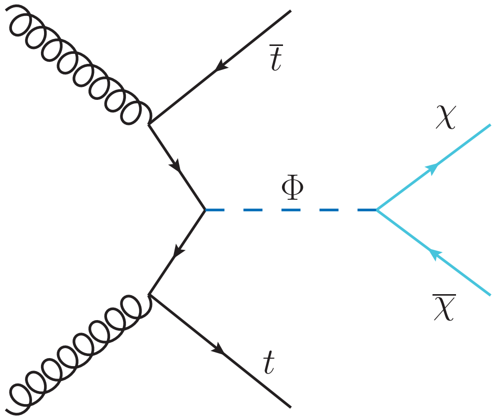


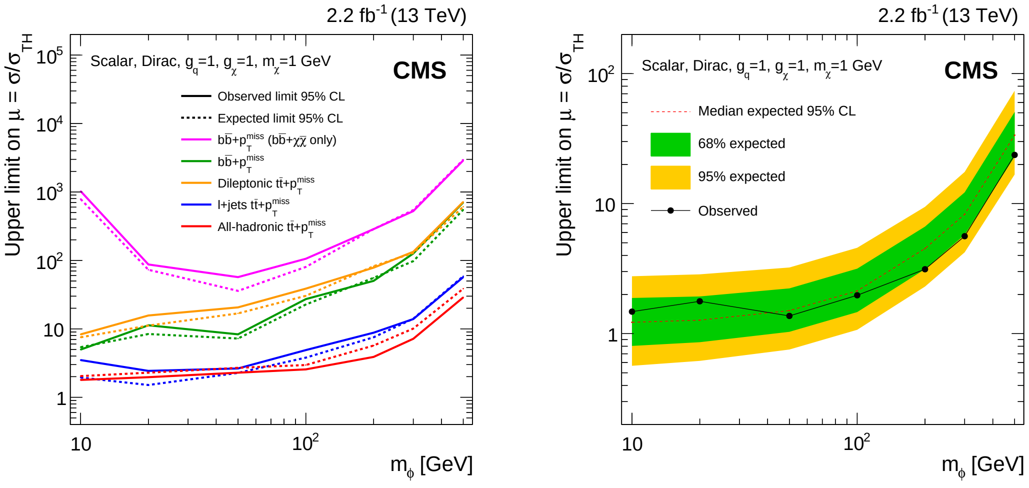
Figure 2.5: Feynman diagram for DM particles χ produced in association with a top quark pair in the simplified model framework. The interaction is mediated by the scalar or pseudoscalar boson Φ.

2.5.3 Motivation $t\bar{t} + \chi\overline{\chi}$ and current results

In this thesis, a search for DM produced in association with a top quark pair is presented (see Chapter 8). Many phenomenological studies have demonstrated enhanced sensitivity in the $t\bar{t}+\chi\overline{\chi}$ channel compared to other DM search channels if a scalar or pseudoscalar particle mediating the interaction between SM and DM particles is assumed (see e.g. [57–60]). In these studies the couplings between the mediator and the SM particles is of Yukawa-type, which is motivated by the principle of *minimal flavor violation* [61, 62]. Assuming Yukawa-type couplings, spin-0 mediators predominantly couple to SM top quarks. This assumption, that DM particles couple much stronger to top quarks than to light quarks, is particularly interesting since neither the direct nor the indirect detection experiments, which both only probe the interaction to light quarks, have found any evidence for the existence of DM so far. The study of pseudoscalar models is furthermore of particular interest since direct searches do not exhibit sensitivity to this interaction [16, 49].

For scalar mediators the strongest constraints on the masses of the DM particles as well as mediators are set by a search for DM produced in association with a jet, a W, or Z boson [63]. The sensitivity of the search is dominated by the signatures assuming associated W or Z boson production. These signatures only occur if, in addition to the coupling of DM to SM fermions, also DM couplings to SM bosons exist. If this is not case, the exclusion limits of [63] are significantly weaker. In this case, the mono-jet processes are suppressed at loop-level. This possible suppression also strongly motivates a search for $t\bar{t} + \chi\overline{\chi}$ production.

The ATLAS and the CMS experiments have already performed searches in the $t\bar{t} + \chi\overline{\chi}$ channel with data collected at $\sqrt{s} = 13\,\text{TeV}$ [3–5, 64, 65]. Both experiments exploited the fully-hadronic, di-leptonic and lepton+jets decay modes of the $t\bar{t}$-system. The observed exclusion limits in each of these channels obtained by the searches performed at the CMS experiment are summarized in Fig. 2.6(a) [3, 4]. The highest sensitivity is reached in the fully-hadronic channel. The exclusion limit of the combination of the three channels and a search for $b\bar{b} + \chi\bar{\chi}$ is presented in Fig. 2.6(b) [3, 4]. For $g_{\text{DM}} = g_{\text{SM}} = 1.5$, scalar



(a) (b)

Figure 2.6: (a) Comparison of the observed (*solid lines*) and expected (*dashed lines*) upper limits (95% C.L.) normalized to the simplified model cross section as a function of the mediator mass obtained for the individual channels (*different colors*) of [3,4]. (b) Combination of the channels presented in (a). Shown are the expected (*dashed red line*) and observed (*solid line*) upper limits (95% C.L.) normalized to the simplified model cross section as a function of the mediator mass. In both figures a DM particle mass of 1 GeV as well as a scalar mediator is assumed. The couplings g_{DM} and g_{SM} are set one. The figures are taken from [3, 4].

(pseudoscalar) mediators below masses of 124 GeV (128 GeV) are excluded at 95% C.L. if DM particles with masses of 1 GeV are assumed [3,4]. The ATLAS experiment derives the most stringent limit in the lepton+jets channel. At 95% C.L., exclusion limits are set on scalar (pseudoscalar) mediator masses below 320 GeV (350 GeV), where DM particles of masses of 1 GeV and couplings of $g_{\mathrm{DM}} = g_{\mathrm{SM}} = 3.5$ are assumed [5].

3 Experiment

The data analyzed in the search for third-generation leptoquarks decaying into a top quark and a tau lepton (Chapter 6) and for DM produced with a top quark pair (Chapter 8) were collected by the CMS experiment at the Large Hadron Collider (LHC). In this chapter the basic concepts of the LHC (Section 3.1) and the CMS experiment (Section 3.2) are presented.

3.1 The Large Hadron Collider

The *Large Hadron Collider (LHC)* [66–69] is an accelerator and storage ring based at the *European Organization of Nuclear Research (CERN)*, which is located near Geneva. The LHC provides proton-proton collisions at center-of-mass energies $\sqrt{s}$ of up to 14 TeV and heavy ion collisions up to $\sqrt{s} = 5.5\,\mathrm{TeV}$ per nucleon[1]. This chapter focuses on the operation with proton beams, since only *pp* collision data are relevant for the work presented in this thesis. A detailed description of heavy-ion operation at the LHC can be found elsewhere, e.g. in Chapter 21 of [69].

The LHC has been built in the former *LEP* tunnel [70]. Before being injected into the LHC, the protons pass through various pre-accelerators, illustrated in Fig. 3.1. The protons are obtained by ionizing hydrogen. While passing through the linear accelerator *Linac2* they reach an energy of 50 MeV. The energy is increased to 1.4 GeV by the *Proton Synchrotron Booster (PSB)*, before being elevated further to 25 GeV inside the *Proton Synchrotron (PS)*. The last step of the pre-acceleration is the *Super Proton Synchrotron (SPS)* from which the protons are transferred into the LHC via two tunnels, creating two opposing beams with an initial energy of 450 GeV. In the LHC the two beams are further accelerated to their final energy.

The LHC ring, which has a circumference of 27 km, is made up of eight arcs and eight straight sections. The arcs are equipped with superconducting NbTi dipole magnets, which can reach a magnetic field strength of up to 8.33 T. The strong field is needed to force the protons onto the circular trajectory defined by the LEP tunnel. The focusing of the proton beams is done with superconducting quadrupole and sextupole magnets installed in the straight sections of the LHC. Superconducting cavities are utilized for the acceleration of the protons.

Each proton beam consists of $n_b = 2808$ bunches, which each contain $N_b = 1.15 \cdot 10^{11}$ protons. The time interval between the bunches is 25 ns, which results in a revolution frequency f of 11 245.5 Hz. From these quantities the *instantaneous luminosity* L can be calculated with

---

[1] Numbers given in this chapter are design values. The actual operation parameters in the past data acquisition periods will be described in 3.3.

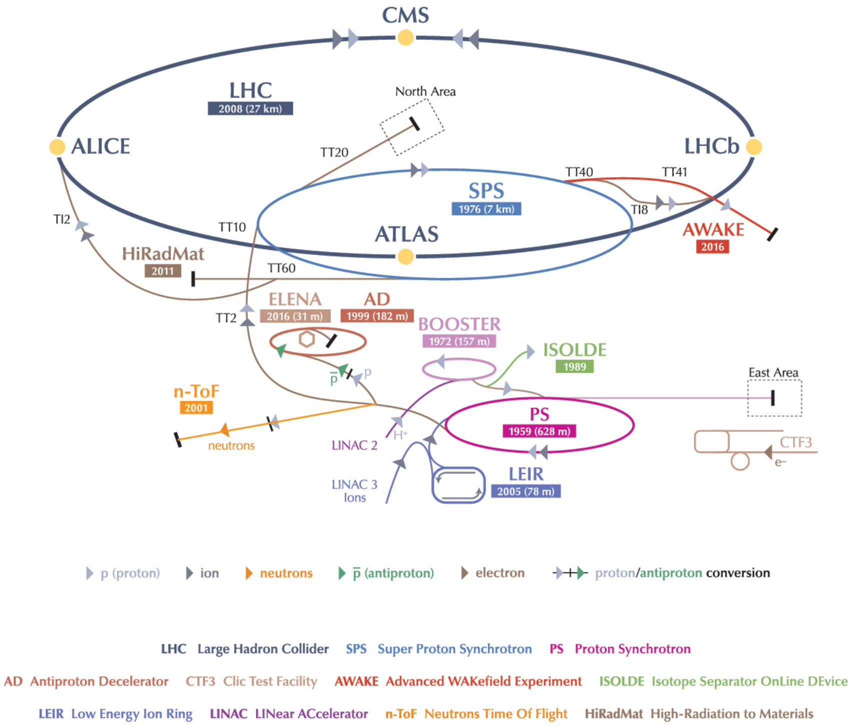


Figure 3.1: Illustration of the accelerators based at CERN. Additionally, the experiments at the LHC are shown [71].

$$L = f\frac{n_b N_b^2}{4\pi\sigma_x\sigma_y}F(\Theta_c, \sigma_x, \sigma_y). \tag{3.1}$$

The variables σ_x and σ_y represent the beam dimensions. The function F describes a reduction of the instantaneous luminosity due to the crossing angle Θ_c of the two beams at the collision point. With the given design parameters an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ can be reached for pp collisions at the LHC. The number of expected events N of a given process can be calculated with the cross section σ of the process and the *integrated luminosity* $L_{\mathrm{int}} = \int L\mathrm{d}t$:

$$N = \sigma L_{\mathrm{int}}. \tag{3.2}$$

Thus, large instantaneous luminosities are needed to enable the analysis of rare processes occurring at the LHC and therefore reach its physics goals.

Four main experiments were constructed at the four interaction regions of the LHC, where the two beams collide. These four experiments are CMS (***C**ompact **M**uon **S**olenoid*) [67], ATLAS (***A** **T**oroidal **L**HC **A**pparatu**S***) [72], LHCb [73] and ALICE (***A** **L**arge **I**on **C**ollider **E**xperiment*) [74]. CMS and ATLAS are multi-purpose experiments, which try to cover the complete physics program of the LHC. In contrast to that, ALICE and LHCb focus on very specific tasks. ALICE was built for the analysis of heavy ion col-

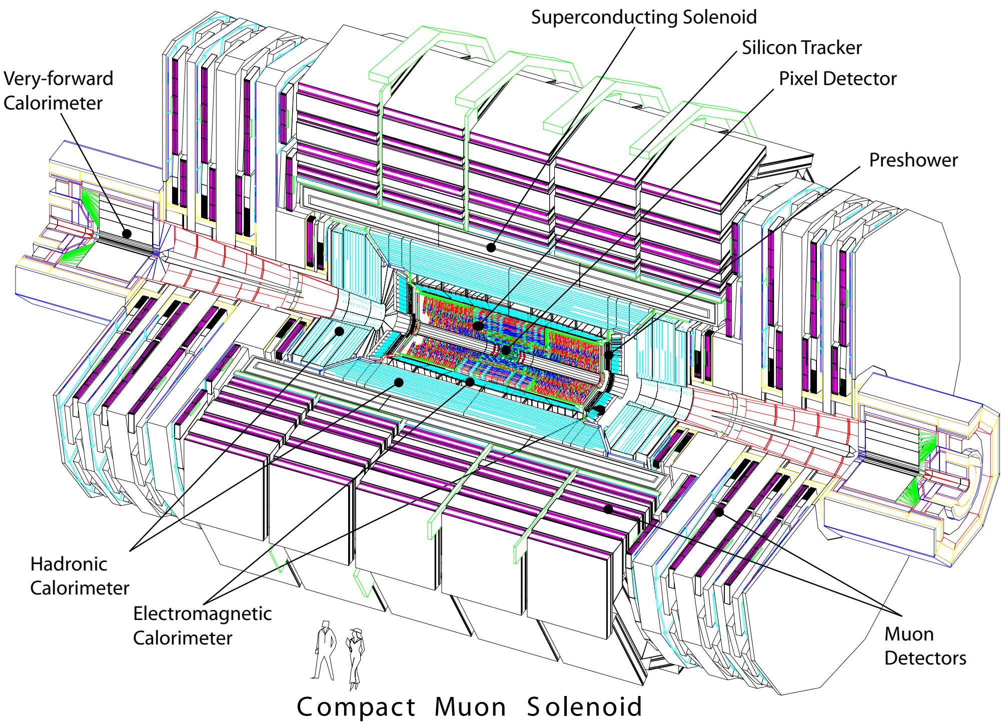


Figure 3.2: Sketch of the CMS experiment [67].

lisions. The aim is to study the strong interaction in the extreme conditions produced by high-energy heavy-ion collisions and to investigate the properties of the produced quark-gluon plasma. LHCb is specialized in b quark physics. The experiment studies CP-violation, rare decays of B and D hadrons and searches for indirect hints for physics beyond the SM in these processes.

3.2 The CMS experiment

The CMS experiment [67] is one of two multi-purpose experiments at the LHC. It surrounds an interaction region of the LHC in a cavern about 100 m under ground. The CMS experiment is 21.6 m long, has a diameter of 14.6 m and a weight of 12 500 t. A sketch of the detector is shown in Fig. 3.2. The experiment is divided into a cylindrical barrel part and two endcaps, which are orientated perpendicular to the beam axis, one at each side of the barrel detector. Different subsystems of the detector, which are needed to distinguish different particle types, are built in an onion-like structure around the beam pipe, starting with the tracking system as the innermost part of the experiment. The tracking system is surrounded by the electromagnetic and the hadronic calorimeters. Around these detector components a solenoidal superconducting magnet is placed. The outermost part of the detector is made up by the muon system.

The experiments at the LHC face many difficulties [67], which arise from the high instantaneous luminosities and the large center-of-mass energy provided by the collider. 10^9 inelastic events per second are expected for the LHC running at design conditions.

Due to limited storage capacities and computing power this high event rate has to be lowered to a few hundred Hz by an efficient trigger system. The trigger system and the read-out system have to be very fast to cope with the high collision frequency. The high instantaneous luminosities also lead to the production of a large number of charged particles per collision. This necessitates radiation-hard detectors and read-out electronics to keep the efficiency loss of the measurements due to radiation damage at a minimum. A further consequence of the high instantaneous luminosities reached by the LHC is that, besides the studied hard interaction, many additional interactions, called *pileup (PU) events*, occur and overlie with the event of interest [75]. *In-time pileup events* are distinguished from *out-of-time pileup events.* In-time pileup describes additional proton-proton interactions that take place within the same bunch crossing. Out-of-time pileup arises if a proton-proton collision from the previous or the subsequent bunch crossing is overlaid with the interaction of interest since the detector needs a certain amount of time to process the signals produced in a collision. Confusion of pileup interactions with the event of interest can lead to mismeasurements of various quantities. To be able to diminish the effect of pileup detector components with a high granularity and a good time resolution are needed. All subsystems of the CMS experiment have been designed with these requirements in mind.

In the following sections all detector components and the trigger system of the CMS experiment are presented shortly (see Sections 3.2.2 to 3.2.7). Prior to these, a description of the applied coordinate system and important variables is given in Section 3.2.1. The presentation of the CMS experiment is based on [67], which provides additional details.

3.2.1 Coordinate conventions and important variables

The coordinate system used to describe the pp collisions in the CMS experiment has its origin at the *nominal interaction point*, which is given by the center of the experiment. The x-axis points towards the center of the LHC ring, the y-axis is orientated upwards and the z-axis along the beam axis, creating a right-handed coordinate system. The angle measured with respect to the z-axis is denoted as θ. The angle in the x-y plane determined with respect to the x-axis is called ϕ. Radial distances in the x-y-plane are described by the variable r.

Given the angle θ, the *pseudorapidity* can be calculated with

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

In the ultra-relativistic limit the pseudorapidity is approximately equal to the *rapidity* y, defined by the energy E of a particle and its momentum component in z-direction, p_z, as

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right). \tag{3.4}$$

As a consequence, for high velocities, the Lorentz-invariance of rapidity differences under boosts in z-direction also applies to pseudorapidity differences. This makes the pseudorapidity a very helpful variable to describe distances of particles produced in pp interactions, where the boost of the partons in the initial state in z-direction is unknown.

Angular distances ΔR between two objects i and j can be described using the pseudorapidity η and the angle ϕ with

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}, \tag{3.5}$$

where $\Delta\phi = \phi_i - \phi_j$ and $\Delta\eta = \eta_i - \eta_j$.
Since the transverse motion of the partons in the initial state is negligible, the total transverse momentum in the final state must be equal to zero due to momentum conservation. Contrary, the total momentum p of the initial state is unknown since each of the colliding partons carries an unknown fraction of the total proton momentum in z-direction. Thus, only the conservation of the total transverse momentum in the particle interaction can be exploited for the description of pp collisions.
The *transverse momentum* p_T of a particle is defined by the x- and y-components of its total momentum, p_x and p_y, as

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{3.6}$$

3.2.2 The inner tracking detector

The innermost component of the CMS experiment is the inner tracking detector, which encloses the beam pipe, has a total length of 5.8 m and a diameter of 2.5 m. It covers the region up to $|\eta| = 2.5$.
Since the superconducting magnet of the CMS experiment (see Section 3.2.5) provides a strong and uniform magnetic field, the trajectories of all charged particles are bend while they traverse the tracking system. Hence, a measurement of the sign of their charge and their momentum per charge is possible. The inner tracking detector aims to precisely measure these for particles with transverse momenta above 1 GeV and to accurately reconstruct primary vertices as well as secondary vertices, which are produced in the decays of B and D mesons. This necessitates a highly-granular and fast tracking detector. Figure 3.3 shows an overview of the inner tracking detector in the r-z-plane. It consists exclusively of silicon-based detector components, which are relatively radiation-hard and enable the detector to operate in the expected large particle flux for a reasonable time-period before the tracking detector has to be exchanged due to serious radiation damage. In total, the silicon-covered area amounts to 200 m^2.
Around the beam pipe, where the particle flux is highest and a very good spatial resolution of the measurement is needed for a precise determination of primary and secondary vertices, 1440 *hybrid silicon pixel detectors* with a size of $100 \times 150\,\mu\text{m}^2$ in $(r\text{-}\phi) \times z$ are utilized. In the barrel region three layers of pixel detectors are positioned at radii between 4.4 cm and 10.2 cm. Each of the endcaps contains two disks of pixel detectors, which are placed at $z = 34.5$ cm and $z = 46.5$ cm and cover the region between 6 cm $< r <$ 15 cm. This layout leads to three position measurements per charged particle traversing the pixel detector with a single point resolution in the barrel of 9.4 µm in the r-ϕ-plane and 20 µm to 45 µm in z-direction [76].
At radii between 20 cm and 116 cm, where the expected particle flux is smaller and thus the usage of less-granular detector components is possible, silicon strip detectors are

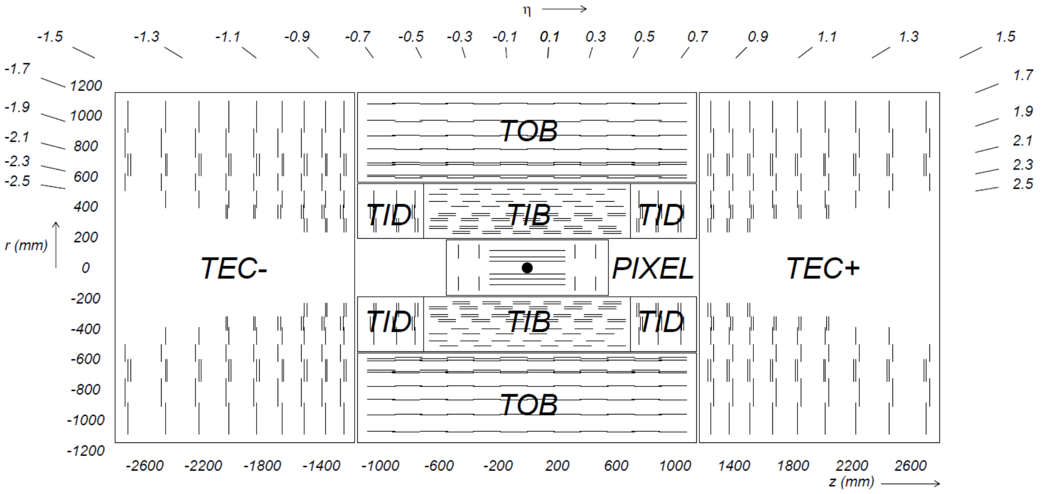


Figure 3.3: Sketch of the inner tracking detector of the CMS experiment in the r-z-plane [67]. Tracking detector modules are illustrated by a line.

installed. This part of the detector is divided into different subcomponents: the *Tracker Inner Barrel (TIB)*, the *Tracker Inner Disks (TIDs)*, the *Tracker Outer Barrel (TOB)* and the *Tracker Endcaps (TECs)*. In total, 15 148 strip modules are installed in the tracking system. In the range between $r = 20\,\mathrm{cm}$ and $r = 55\,\mathrm{cm}$ silicon strip detectors with a typical strip size of $10\,\mathrm{cm} \times 80\,\mu\mathrm{m}$ are installed. Going to higher radii, where the area that has to be covered is larger and the particle flux is much lower, larger strips with sizes up to $25\,\mathrm{cm} \times 180\,\mu\mathrm{m}$ are used. The TIB consists of four layers of silicon micro-strip detectors, whose strips are orientated parallel to the z-direction. Three disks in each of the endcaps with radially running strips make up the TIDs. Both subsystems provide up to four measurements of the trajectories of charged particles in the r-ϕ-plane. In the TIB a single point resolution of $23\,\mu\mathrm{m}$ or $35\,\mu\mathrm{m}$ is achieved depending on the size of the strip pitch, which differs between the inner and outer layers. The TOB surrounds the TIDs and TIB. It has a radius of $116\,\mathrm{m}$ and ranges up to $|z| = 118\,\mathrm{cm}$. Due to its six layers of silicon strip detectors, the TOB provides up to six measurements in the (r-ϕ)-plane. The single point resolution is $53\,\mu\mathrm{m}$ in the inner four layers and $35\,\mu\mathrm{m}$ in the outer layers. The different resolutions are again a result of different strip pitches. The TECs cover the region $124\,\mathrm{cm} < |z| < 282\,\mathrm{cm}$ and $22.5\,\mathrm{cm} < r < 113.5\,\mathrm{cm}$. They consist of nine disks of silicon microstrip detectors in each endcap providing up to nine measurements of the ϕ-coordinate of the trajectory of charged particles.

Some layers in the TIB, TIDs and TOB as well as some rings in the TECs are equipped with a second silicon strip detector module. These additional modules are tilted by a small angle with respect to the other detector modules, such that in the disks also the radial component and in the barrel also the z-coordinate of a trajectory is measured. In the TIB a single point resolution of $230\,\mu\mathrm{m}$ is achieved, in the TOB the resolution is $530\,\mu\mathrm{m}$.

The described tracking system enables a highly efficient vertex and track reconstruction [76], which contributes to the high reconstruction efficiency of charged particles and vertices with the CMS experiment (see Section 5).

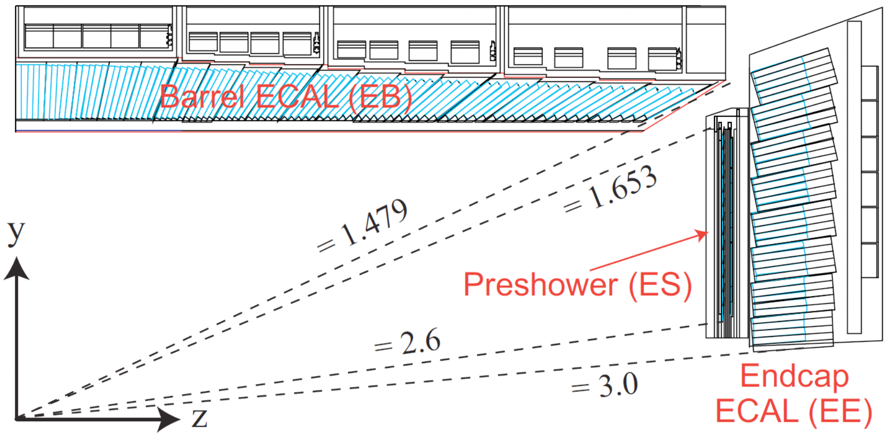


Figure 3.4: Sketch of one quarter of the ECAL in the y-z-plane [77].

3.2.3 The electromagnetic calorimeter

The *electromagnetic calorimeter (ECAL)* of the CMS experiment is, in its main parts, a homogeneous calorimeter. It surrounds the inner tracking system and consists of a barrel part, two endcaps and preshower detectors located in front of the endcaps. A sketch of one quarter of the ECAL in the y-z-plane is depicted in Fig. 3.4. The ECAL is made out of lead tungstate ($PbWO_4$) crystals. The purpose of the ECAL is to precisely measure the energy of photons and electrons. To accommodate these tasks a hermetic, fast, radiation-hard and highly-segmented calorimeter is needed, which is capable to completely contain the electromagnetic showers produced by photons and electrons. Since $PbWO_4$ crystals have a high density ($8.28\,\frac{\mathrm{g}}{\mathrm{cm}^3}$), a short Molière radius (2.2 cm) and a small radiation length X_0 (0.89 cm) they have the needed characteristics to build this compact and highly-segmented calorimeter. The blue-green scintillation light emitted by the crystals is detected and amplified by avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps. These kind of photodetectors were chosen because they can be used in a strong magnetic field, are radiation-hard, and able to amplify the small light yield emitted by the crystals. Additionally, approximately 80% of the scintillation light is emitted by the crystals within 25 ns, which makes the ECAL fast enough to cope with the high LHC collision rate.

In the *barrel part of the ECAL (EB)*, which covers the region up to $|\eta| < 1.479$, 61 200 crystals with a cross section of circa 0.0174×0.0174 in $\Delta\eta \times \Delta\phi$ are installed. Each of these crystals is 230 mm long, which is equivalent to $25.8\,X_0$. The crystals are installed with an angle of 3° between their axes and the vector to the center of the detector to avoid coincidences between two crystals and the trajectory of a particle.

Between $1.479 < |\eta| < 3.0$ the *endcaps of the electromagnetic calorimeter (EE)* are installed. In z-direction they are placed at $|z| = 315.4$ cm. Each endcap disk is divided in two D-shaped halves and contains 7324 crystals in total. The crystals in each halve are arranged in units of 5×5 crystals. With a length of 220 mm, which corresponds to a radiation length of $24.7\,X_0$, the crystals in the endcaps are slightly smaller than their counterparts in the barrel.

In order to improve the identification of neutral pions that decay into two photons,

which is an important background process for $H \to \gamma\gamma$ searches, a *preshower detector* can be found in front of each endcap disk. It covers the pseudorapidity region between $1.653 < |\eta| < 2.6$. The preshower detectors are sampling calorimeters, which consist of two layers of active and passive material. Lead is used as passive material, while silicon strip sensors, which are arranged orthogonal to each other, are used as active material. In total, the preshower detectors have a thickness of 20 cm, corresponding to $3\,X_0$. In addition to the identification of neutral pions, the preshower detectors also improve the capability to differentiate between minimum ionizing particles and electrons and to determine the position of photons and electrons.
The relative energy resolution of a calorimeter $\frac{\sigma}{E}$ is described by [67]

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{3.7}$$

where S, N and C are calorimeter specific constants, and the relative resolution improves with the energy of the measured particle. The first term in Eq. (3.7) is the *stochastic term*, which describes statistical fluctuations in the number of particles contained in the shower and, in case of the ECAL of the CMS experiment, also in the measurement of the scintillation light emitted by the crystals. The second term is the *noise term*, while the third is the so called *constant term*. The latter describes effects on the energy resolution due to dead material and non-uniformities of the calorimeter. For electron candidates, the parameters in Eq. (3.7) have been measured as $S = 2.8\%\sqrt{\mathrm{GeV}}$, $N = 12\%\,\mathrm{GeV}$ and $C = 0.3\%$ [78].

3.2.4 The hadronic calorimeter

The *hadronic calorimeter (HCAL)*, which completes the calorimetry of the CMS experiment, surrounds the ECAL. The sampling calorimeter aims at a measurement of the energy of all hadrons contained in jets with a good resolution and at a complete coverage in ϕ, which is especially important for a precise measurement of the missing transverse energy in an event.
The HCAL of the CMS experiment is divided into a *barrel part (HB)*, *endcaps (HE)*, an *outer calorimeter (HO)* and the *hadron forward calorimeter (HF)*. In total, the HCAL covers the pseudorapidity region up to $|\eta| = 5.2$. A sketch of one quarter of the HCAL of the CMS experiment is depicted in Fig. 3.5.
The HB covers the pseudorapidity region up to $|\eta| = 1.3$. It is split into two half-barrel regions containing 18 azimuthal wedges, which are arranged in rings. Hence, a complete coverage in ϕ is ensured. Each wedge is made up of several planes of passive and active material, which are orientated parallel to the beam axis. As passive material, brass is used. To ensure the needed structural strength the innermost and outermost layers of the absorber are made of stainless steel. The active material is a moderately radiation-hard and long-term stable plastic scintillator. The emitted light is processed by wavelength shifting fibres and detected with hybrid photodiodes. In addition to the subdivision into four different ϕ regions of the active and passive material, the active material is segmented in 16 η regions. This leads to a segmentation of 0.087×0.087

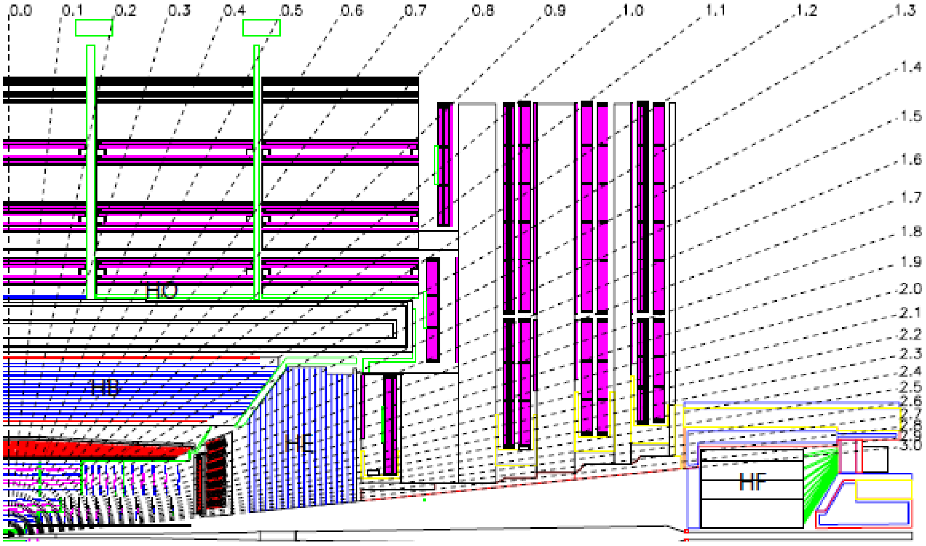


Figure 3.5: Sketch of one quarter of the HCAL of the CMS experiment [67]. Fixed η-coordinates are represented by dashed lines.

in $\Delta\eta \times \Delta\phi$. The active material in the η sections closest to the transition region between the barrel and the endcaps is also segmented in the longitudinal direction. This granularity corresponds to 70 000 assembled tiles. Since the brass absorber plates have thicknesses of 50.5 mm and 56.5 mm and the steel plates are 40 mm and 75 mm thick, the total thickness in interaction lengths at a polar angle of 90° is 5.82 λ, increasing to 10.6 λ at $|\eta| = 1.3$. Additional 1.1 λ are added by the material of the ECAL in front of the HB.

The η region between $1.3 < |\eta| < 3$ is covered by the HE. In this part of the HCAL 79- mm-thick brass plates are used as absorber material. In the 9 mm wide gaps between these, plastic scintillators are installed as active material. The active material features the same granularity as in the HB for $|\eta| < 1.6$, while for larger η the segmentation is coarser. This leads to 20 916 tiles in total. Similar to the HB, wavelength shifting fibres and multipixel hybrid photodiodes are used to detect the signal. In longitudinal direction the HE is segmented in two or three parts, depending on the distance to the beam line. Together with the ECAL, a thickness of approximately 10 λ is reached.

The HF is placed 11.2 m away from the nominal interaction point in the pseudorapidity region $3 < |\eta| < 5.2$. Since in this region very high particle fluxes are expected - on average 760 GeV per collision is deposited in this region in contrast to 100 GeV in the remaining part of the detector - the usage of a highly radiation-hard active material is particularly important. Therefore quartz fibres running parallel to the beam line and merged to towers with a size of 0.175×0.175 in $\Delta\eta \times \Delta\phi$ are chosen as active material in this part of the HCAL. Charged particles traversing this pseudorapidity region emit Cherenkov light, which is detected by photomultiplier tubes. The fibre bundles are arranged in 18 wedges on each side of the beam line, which each cover 20° in ϕ. In longitudinal direction two segments can be found. 5 mm-thick steel plates are used as absorber.

To ensure a good resolution of the energy measurements of the HCAL, it is essential that the complete hadron shower is contained in the hadronic calorimeter. Therefore, at $|\eta| < 1.3$, the HB is extended by an additional outer calorimeter, placed outside the solenoid within the return yoke of the magnet. Thus, the solenoid is used as additional absorber material. This extends the thickness of the HCAL in the barrel region to $11.8\,\lambda$. Plastic scintillators are the active material in the HO, which is segmented similar as the HB. The signal is likewise detected with wavelength shifting fibres and photo detectors. The structure of the HO closely follows the structure of the return yoke and the segmentation of the muon system.

The resolution of the HCAL has been determined in a pion test beam [78, 79] and is given by

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{110\%\sqrt{\mathrm{GeV}}}{\sqrt{E}}\right)^2 + (9\%)^2, \tag{3.8}$$

with a contribution of approximately 200 MeV for typical noise [78].

3.2.5 The magnet

To precisely measure the momentum and the sign of the electric charge of highly-energetic particles, a large bending power and therefore a strong and uniform magnetic field penetrating the tracking and muon system of the CMS experiment is required. The strong field is one of the most important features of the CMS experiment. It is provided by a superconducting solenoid placed between the calorimeters and the muon system. The chosen layout has the advantage that the amount of material before the calorimeters is kept low, such that the amount of energy lost by the particles before they reach the calorimeters is smaller compared to other layouts. Thus, a better energy resolution in the calorimeters is reached.

The solenoid of the CMS experiment is 13 m long, its bore has a diameter of 6 m and incorporates the tracking system and the calorimetry of the CMS experiment. The total weight of the magnet amounts to 220 t. Four layers of niobium-titanium windings build the coil. A current of approximately 19 kA in the coil produces a 3.8 T-magnetic field in z-direction, which enables the measurement of the transverse momentum of charged particles in the x-y-plane. To be able to return the magnetic flux and at the same time generate a strong magnetic field inside the muon system, an iron yoke consisting of five barrel wheels and a total of six endcap disks permeates the muon system.

3.2.6 The muon system

The outermost detector component of the CMS experiment is the muon system. Its task is to provide efficient muon identification, precise momentum measurements and fast and reliable triggering of muons (see Section 3.2.7). Since muons often occur in final states of interesting physics process, like e.g. $H \to ZZ \to \mu^+\mu^-\mu^+\mu^-$ and because of their clean signatures in the detector, a very good performing muon system is particularly important for the CMS experiment.

The muon system of the CMS experiment, which covers the region up to $|\eta| < 2.4$,

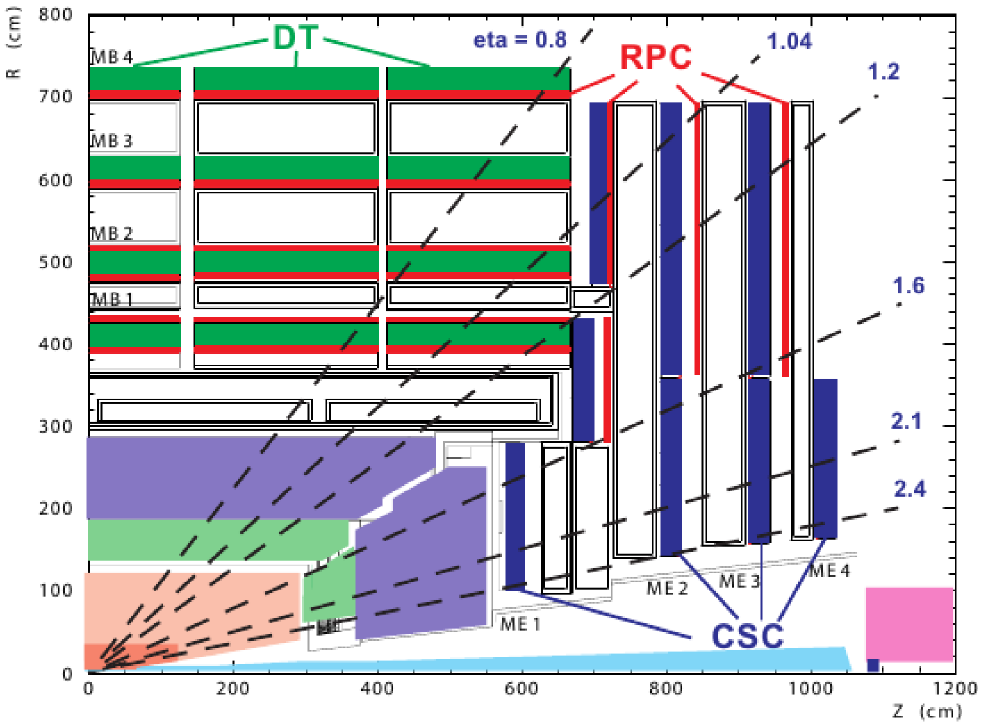


Figure 3.6: Schematic overview of one quarter of the muon system in the r-z-plane [77]

consists of a cylindrical barrel region and two endcap regions. One quarter of the muon system is illustrated in Fig. 3.6. Three different detector types, all operating with gas, are embedded into the magnet return yoke. The detector type to be used depends on the expected flux of muons and the homogeneity of the magnetic field.

In the barrel region of the muon system up to $|\eta| < 1.2$ four stations of rectangular *drift tube (DT) chambers* are installed. Due to the low expected muon rate, the small background generated by neutrons and the homogeneous magnetic field in this region, DTs provide a good choice for this part of the muon system. The drift length within one DT is 21 mm, which leads to an acceptable number of read-out channels on the one hand and to negligible occupancy on the other. All four stations measure the position of muon candidates in the r-ϕ-plane. Combining all measurements in r-ϕ, a resolution of 100 µm is achieved. Additionally, the innermost three chambers are able to determine the position of the muon candidates in the z-direction. Since the chosen arrangement of the drift tubes provides an excellent time resolution, which lies in the order of a few nanoseconds, DTs are also used for triggering.

Cathode strip chambers (CHSs) are installed in the endcap regions of the muon system ($0.9 < |\eta| < 2.4$). Each endcap comprises four stations of CSCs, which are multiwire proportional chambers. These fast radiation-hard detectors are particularly suitable for this part of the detector, where a high rate of muons and backgrounds as well as an intense and inhomogeneous magnetic field can be found. The cathode panels of the CSCs are segmented into radially running strips with a constant width in $\Delta\phi$. The anode wires are arranged almost perpendicular to the cathode planes. By determining the charges induced on the strips by a charged particle that ionizes the gas while it passes the chamber a coarse measurement of the position of muon candidates is provided. Since this position measurement can be performed very fast, it is used for triggering of muon candidates. More precise position measurements, with resolutions between 75 µm and 150 µm in the r-ϕ-plane, are achieved by analyzing the center-of-gravity of the charge distribution induced on the cathode strips. Additionally, a measurement of the η-coordinate is obtained by reading out the anode wires.

In the η region of $|\eta| < 2.1$ the muon system is additionally equipped with *resistive plate chambers (RPC)*. RPCs are double-gap chambers used in avalanche mode. In between the two gaps common read-out strips are installed. The strips are oriented along the z-direction and segmented into parts. Since these detectors are very fast and therefore able to assign each signal to the correct bunch crossing, RPCs are especially suitable for operation in areas with a high flux of muon candidates and also provide complementary triggering information to DTs and CSCs. The latter are slower in read-out but have a better position resolution than the RPCs. In each of the first two stations of the barrel region two RPC layers can be found. In this way, low-p_T muons that do not reach the outermost stations of the muon system can be triggered efficiently as well. In each of the last two stations only one layer of RPCs is installed, while each of the innermost four stations of the endcaps are equipped with one plane of RPCs. The segmented strips in the endcaps run in radial direction.

The described design of the muon system provides highly efficient muon reconstruction and identification capability (see Section 5.3).

3.2.7 The trigger system

At design luminosity of the LHC, bunch crossings occur with a frequency of 40 MHz. Due to limited computational processing power and storage capacities, not all of the resulting collision events can be recorded and analyzed. This necessitates an efficient trigger system, which identifies and stores the interesting events, while simultaneously reducing the rate to a few hundred Hz. The trigger decision, whether to keep or discard an event, has to be made very fast in order to keep track with the collisions occurring each 25 ns.

At the CMS experiment a two-stage trigger system, consisting of the *Level-1 (L1) trigger* and the *High-Level-Trigger (HLT)*, is implemented. The L1 trigger aims at a reduction of the initial event rate to 100 kHz. In contrast to the fully software-based HLT, the L1 trigger is based on custom-designed hardware completely, which is mainly located in a service cavern near the experiment. It analyzes collision events by evaluating information from the calorimeters and the muon system, which are detector components that enable a rapid read-out of signals. The trigger decision is based on information obtained from coarser detector resolution to speed up the analysis process. In the calorimeters signals collected by *trigger towers*, which in most cases summarize several ECAL crystals and multiple HCAL read-out units, are examined. Up to $|\eta| < 1.74$ these trigger towers have a size of 0.087×0.087 in $\eta \times \phi$, which corresponds to the granularity in the HB. In the remaining parts of the calorimeters larger trigger towers are used. By combining the information from an array of usually 4×4 trigger towers *trigger objects* are built, which in this case are candidates for electrons and photons. Trigger towers up to $|\eta| = 5$ are utilized to calculate event based variables, e.g. the total transverse energy deposited in the calorimeters and the total missing transverse energy. Furthermore, they are used for the reconstruction of jets, which is based on a special four-step clustering procedure. Muon candidates are reconstructed within $|\eta| < 2.4$ by combining information from track segments and hit patterns measured by the DTs, CSCs and RPCs. For a positive trigger decision, which means that the event is passed to the HLT for further investigation, the trigger objects have to fulfill certain pre-defined momentum, energy or multiplicity thresholds. The L1 trigger decision is made in approximately $3.2\,\mu$s, in which the signals are passed to the service cavern, analyzed and transferred back to the front-end electronics. In the meantime, the full event containing the high resolution measurements is held in pipeline buffers.

In contrast to the L1 trigger, the HLT trigger decision is based on quantities calculated with the full available granularity of the mentioned detector components as well as the inner tracking system. Rather detailed and complex calculations are performed during the decision process, which, in some cases, can already be quite similar to the offline reconstruction and identification algorithms. However, not the full event, but only interesting regions are analyzed, making the HLT faster than the complete event reconstruction performed afterwards. Again, all trigger objects have to fulfill certain thresholds to pass this triggering step, which takes 50 ms on average. With the HLT the final reduction to a rate of a few hundred Hz is achieved.

The combination of criteria that an event has to fulfill in order to be accepted by the trigger system is called a *trigger path*. Data accepted by trigger paths that require similar objects, e.g. a muon with different p_T thresholds, are collected into corresponding data streams, e.g. the SingleMuon data stream. Each trigger path is assigned a certain event rate that must not be exceeded in order to achieve the required total output rate. If certain trigger paths surpass their assigned trigger rates due to changes in the instantaneous luminosities, the trigger paths get a *prescale* n. This means that only the n^{th} event that passes the trigger conditions is actually stored for data analysis.

3.3 Operation periods and future plans

In September 2008, operations of the LHC had to be stopped only a few days after start-up due to a technical incident causing an explosion that destroyed several of the superconducting magnets inside the LHC. The necessary repairs were carried out in the following year and the LHC restarted in November 2009. The first pp collisions were performed at a center-of-mass energy of 900 GeV. The first longer operation periods of the LHC lasted from March to October in 2010 and 2011. During this time, the LHC ran at a center-of-mass energy of 7 TeV with a time spacing of 50 ns between subsequent proton bunches. Peak instantaneous luminosities of $2.04 \cdot 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$ ($4.02 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$) were reached in 2010 (2011). In total, data equivalent to $45\,\text{pb}^{-1}$ ($6.1\,\text{fb}^{-1}$) were recorded in 2010 (2011). These data were used for the commissioning of the detector operations and first data analyses. In the subsequent data acquisition period (from April to December

2012), the center-of-mass energy was increased to 8 TeV. In this period, up to 1374 bunches circulated in the LHC with a time spacing of 50 ns. Each bunch contained up to $1.6 \cdot 10^{11}$ protons, thus exceeding the design number for the bunch population [80]. With instantaneous luminosities of up to $7.7 \cdot 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, data corresponding to $23.3\,\mathrm{fb}^{-1}$ were delivered to the CMS experiment. On average, each event contained 21 pileup vertices [81]. These data are analyzed in the search presented in Chapter 6. The search is denoted as 8 TeV analysis in the following, while the data acquisition periods from 2010 to 2012 are referred to as *Run-I*.

Between the end of 2012 and the beginning of 2015 LHC operation was paused in order to perform maintenance and upgrade work on the LHC and its experiments. During this so called *Long Shutdown 1 (LS1)* [82,83] several magnets inside the LHC were repaired and new security systems were installed. Pixel and silicon strip modules of the tracking system as well as the beam pipe inside the CMS experiment were replaced. Additionally, the whole tracking system was prepared to enable operation at significantly lower temperatures than before to minimize radiation damage. Modules were repaired in the ECAL and in the muon system, while new photodetectors with improved signal-to-noise ratio were installed in the HO. Additionally, new RPCs and CSCs were added to the muon system to achieve a muon acceptance of $|\eta| < 2.4$ instead of $|\eta| < 2.1$ as in Run-I. Furthermore, improvements in the trigger system, in the read-out system of most of the detector components as well as in the *Data Acquisition (DAQ)* were implemented to enable faster event reconstruction. After LS1, the LHC successfully restarted operation at $\sqrt{s} = 13\,\mathrm{TeV}$ with a bunch spacing of 25 ns in May 2015, reaching instantaneous luminosities of up to $5.13 \cdot 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ due to, amongst others, an improved beam focusing capability at the interaction point. Unfortunately, the magnet of the CMS experiment faced problems with its cryogenic system. As a result the CMS experiment recorded a significant portion of the delivered data, equivalent to an integrated luminosity of $4.2\,\mathrm{fb}^{-1}$, without magnetic field. Nonetheless, the recorded data were used for recommissioning of the detector and for first data analyses at $\sqrt{s} = 13\,\mathrm{TeV}$. The 2015 data acquisition period was followed by a very successful year 2016. In this period up to 2040 bunches with up to $1.15 \cdot 10^{11}$ protons each were injected into the LHC [80]. Peak luminosities of $1.53 \cdot 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ were achieved, which exceeded the LHC design values. Hence, pp collision data equivalent to $41.1\,\mathrm{fb}^{-1}$ could be recorded by the CMS experiment. The data acquisition periods of 2015 and 2016 are summarized as *Run-II* in the following. Roughly one third of the data collected in 2016 is analyzed in the search for Dark Matter presented in Chapter 8. The search is referred to as *13 TeV analysis* in the following. The mean number of pileup interactions during this run period was 20. A summary of the recorded data as a function of time in the different data acquisition periods described in this section is presented in Fig. 3.7, the peak instantaneous luminosities as well as integrated luminosities of the collected data mentioned in this section are taken from [81].

After the end of the data acquisition in October 2016, the *Extended Year End Technical Stop (EYETS)* started, which lasts until May 2017. Besides maintenance work performed on the LHC and its experiments, the main task for the CMS experiment during

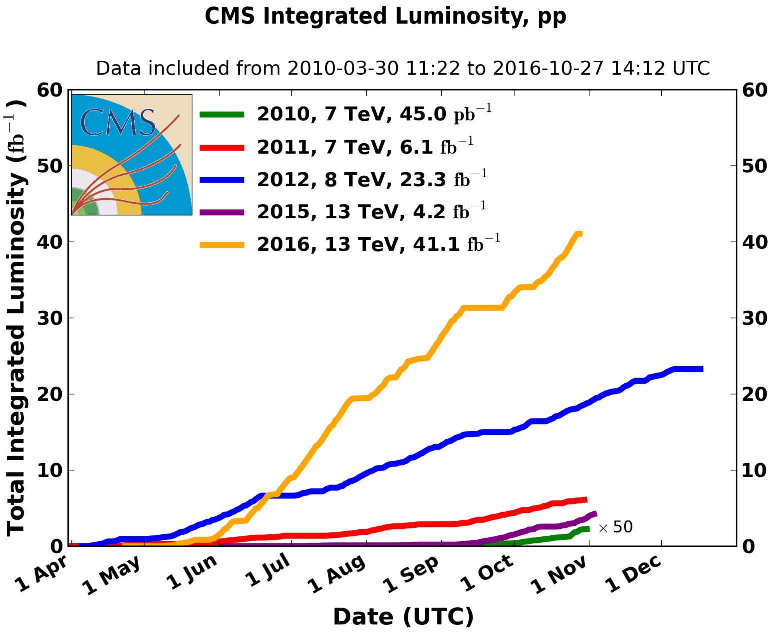


Figure 3.7: Total integrated luminosity recorded by the CMS experiment as a function of time in the different data acquisition periods (*colored lines*). The figure is taken from [81].

the EYETS is the replacement of its pixel detector [84]. The new pixel detector has a fourth barrel layer to extend the coverage of the pixel detector in r. Furthermore, the first pixel layer is located closer to the beam pipe than before. Additionally, a third endcap disc is added. The geometry of the modules inside the endcaps is changed slightly to ensure a better resolution in r and ϕ. In both the barrel and the endcaps, more radiation-hard modules with faster read-out are installed. This modification will result in a more efficient track reconstruction with lower misidentification rate. Having a layer closer to the beam pipe is thereby especially important for higher vertex reconstruction efficiencies and thus particularly helpful for b tagging. In May 2017, the LHC resumes operation at $\sqrt{s} = 13\,\mathrm{TeV}$ for the last data acquisition period of Run-II. The goal is to collect approximately $50\,\mathrm{fb}^{-1}$ of pp collision data. Afterwards, the *Long Shutdown 2 (LS2)* is planned, in which the LHC injectors will be upgraded. Moreover, the accelerator will be prepared to run at instantaneous luminosities corresponding to twice the design value ($2 \cdot 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$) and the experiment will undergo the *Phase 1 upgrade*. During the Phase 1 Upgrade [84] the CMS experiment will be prepared to cope with higher luminosities and thus with higher particle fluxes and enhanced radiation damage by replacing electronics and read-out systems in many subdetectors. Additionally, the photodetectors in the HCAL will be replaced with more radiation-hard detector modules segmented in longitudinal direction. The trigger system and the DAQ will be equipped with new advanced technologies enabling a faster operation and better capability to cope with higher rates and more complex events with high numbers of pileup vertices. After LS2, the LHC will operate at a center-of-mass energy of 14 TeV. During Run-III, lasting from 2021 to 2023, the recording of pp collision data equivalent to $300\,\mathrm{fb}^{-1}$ is foreseen. In *Long Shutdown 3 (LS3)* (from 2024 to mid of 2026), the LHC and the experiments will go through a second major upgrade in order to arrange for the running of the *High Luminosity Large Hadron Collider (HL-LHC)*. For the HL-LHC, instantaneous luminosities shall be increased to up to $7 \cdot 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. It is planned to take data until 2037, thereby recording $3\,\mathrm{ab}^{-1}$ of data. To handle these extremely high luminosities, where 140 pileup vertices are expected on average, the main tasks during the *Phase 2* [85] upgrade of the CMS experiment during LS3 will be to replace the whole tracking system as well as the endcap calorimeters with new detectors featuring higher granularities and more radiation-hard modules. Furthermore, the coverage of the tracking system will be extended to $|\eta| < 4$. Additionally, new electronics will be installed in most of the subsystems and the trigger system will be further improved to ensure faster operation and capability of handling higher output rates. The description of the time line of the LHC from 2017 on given here is based on [86].

4 Event Simulation

Simulated events are frequently used in particle physics to predict event yields and event kinematics for SM processes as well as signal models describing physics beyond the SM. In this chapter, the basic ideas of event simulation with Monte Carlo (MC) generators are shortly introduced. The descriptions are based on [87, 88].
Event generation performed in particle physics is based on *MC methods* [89], which simulate quantum-mechanical processes according to probability functions that describe the physics process under evaluation. The generation of a pp collision event proceeds in several steps, which can be studied to a good approximation independently from each other due to different characteristic energy scales involved. The event generation starts with the *hard process*, which describes the interaction characterized by the highest momentum transfer Q in the event. Here, the hard interaction is denoted as an interaction of two partons a and b contained in the colliding protons, such that n particles are produced. The cross section σ of this process is given by

$$\sigma_{pp} = \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a(x_a, \mu_F) f_b(x_b, \mu_F) \cdot \frac{1}{2\hat{s}} |\mathcal{M}_{ab\to n}(\Phi_n; \mu_F, \mu_R)|^2. \tag{4.1}$$

In this formula $f_a(x_a, \mu_F)$ and $f_b(x_b, \mu_F)$ describe the *parton distribution functions (pdfs)*. The pdf $f_a(x_a, \mu_F)$ describes the probability that a parton a that carries the longitudinal momentum fraction x_a of the total momentum of the proton is found inside the proton. They depend on the factorization scale μ_F. Pdfs cannot be calculated theoretically, but are determined by global fits to different kinds of data, e.g. deep-inelastic scattering data from HERA or jet data from the LHC. The pdf sets used in this thesis are the NNPDF3.0 [90], the CTEQ6L1 [91] and the CT10 [92] pdf sets. The formula given in Eq. (4.1) additionally contains the differential phase space element $d\Phi_n$ for the final state particles n and the squared center-of-mass energy of the subprocess $\hat{s}$, which is related to the squared center-of-mass energy of the pp collision s by $\hat{s} = x_a x_b s$. The matrix element $\mathcal{M}_{ab\to n}(\Phi_n; \mu_F, \mu_R)$, where μ_R is the renormalization scale, can in principle be calculated by summing the respective Feynman diagrams according to the Feynman rules. The order to which the matrix elements are calculated depends on the used matrix element generator. In this thesis, the tree-level generator MadGraph [93, 94], the next-to-leading order event generators MadGraph5_aMC@NLO [94] and powheg [95, 95, 96, 96, 97, 97, 98, 98–100] as well as the leading-order generator pythia [101, 102] are used to simulate the hard process. In some cases, MadSpin [103,104] is utilized in addition to model the angular distributions of top quark decay products correctly. Since there is no rule for setting the unphysical scales μ_R and μ_F, choices have to be made and systematic uncertainties, which are obtained by varying both scales, have to be applied. Often $\mu_R = \mu_F = Q^2$ is chosen since Q^2 is a measure for the energy scale of the interaction.

To simulate the cascade of gluons, quark-antiquark pairs and also photons emitted by the colored objects in the initial and final state described by the hard interaction, a *parton shower* model is applied. This model describes how the shower evolves going from the high scales of the hard process down to scales of approximately 1 GeV based on splitting functions [105–107]. At a scale of about 1 GeV, pertubation theory is not applicable anymore and hadronization processes start. The two most commonly used hadronization models are the *string model* and the *cluster model.* The string model describes the linearly with the distance between two colored objects increasing potential by the formation of a color-flux tube. If the energy contained in the tube is large enough, the color-flux tube breaks and quark-antiquark pairs are built. From these, hadrons are constructed. In the cluster model, intermediate clusters of objects with masses of a few GeV are built, which than further decay into the final-state hadrons. For all processes considered in this thesis, PYTHIA is used to model the parton shower and hadronization process, for which the string model is applied. MADGRAPH is also capable to simulate processes with additional partons, up to four are simulated for processes in this thesis. In these cases, the hard process generated with MADGRAPH has to be matched to the output of the parton shower obtained by PYTHIA to avoid double counting of objects. Rules for this matching are provided by the MLM [108] and FxFx merging scheme [109] for leading order and next-to-leading order generators, respectively.

The last step in the generation procedure is the simulation of the *decay* of hadrons that are not stable on detector timescales. To model these decays, many available MC generators determine the matrix element and also take spin correlations into account. The outcome of this step depends on the hadrons included in the MC generator and on the considered decay modes. TAUOLA [110], which is a generator specialized on the correct modelling of tau lepton decays, is used.

To achieve a complete description of a *pp* collision event the interactions of partons inside the protons that are not part of the hard interaction as well as interactions of the proton remnants have to be modeled. These interactions make up a main part of the so called *underlying event* and lead to a higher particle multiplicity and an increased scattering energy content in the event. While a small part of these interactions are hard and can therefore in principal be described perturbatively, most of these interactions are soft. The soft components have to be described by appropriate models, which depend on parameters that are summarized in so called *tunes.* PYTHIA is used to model the underlying event in all processes described in this thesis.

In order to model the interactions of the generated particles with the detector material, all simulated processes are subsequently processed by a detailed simulation of the CMS experiment based on GEANT4 [111]. Furthermore, detector signals from $2 \rightarrow 2$ processes simulated with PYTHIA are added to the simulated processes to model contributions to the event arising from pileup. Afterwards, the generated signals can be handled as if they would arise from data events and are passed to the same reconstruction and identification algorithms as the real *pp* collision data. These algorithms are described in the next chapter (Chapter 5).

5 Reconstruction and identification of particle candidates and jets

From the signals recorded in the different subsystems of the CMS experiment particle candidates are reconstructed and identified. By aiming at a reconstruction of all produced stable particles individually, the CMS experiment is capable of providing a complete event description. The high granularity of its tracking system and ECAL, the hermetic calorimeter, the strong magnetic field and the very efficient muon system make this approach feasible. The corresponding reconstruction method, called *Particle Flow (PF) algorithm*, is presented in Section 5.1. The output of this algorithm are muon, electron, photon, neutral and charged hadron candidates, which are passed on to specialized algorithms for further reconstruction and identification. The reconstruction of primary vertex candidates and the applied identification criteria are described in Section 5.2. For muon candidates this is explained in Section 5.3, which is followed by a section presenting the identification and reconstruction of electron candidates (Section 5.4). Particles built by the PF algorithm are used to cluster jets. The clustering procedure, pileup mitigation techniques, the applied jet identification criteria, and jet energy corrections are detailed together with b quark identification algorithms in Section 5.5. Clustered jets are used as input for the reconstruction of tau lepton candidates, which is described together with the applied identification criteria in Section 5.6. The chapter ends with the definition of the event based variables $\not{E}_T$ and S_T in Section 5.7.
If differences between the analyses performed with 8 TeV (Chapter 6) and 13 TeV (Chapter 8) data exist with regard to reconstruction algorithms or identification criteria, the respective section is divided into two parts describing the specific aspects applied in Run-I and Run-II.

5.1 Event reconstruction with the Particle Flow algorithm

The Particle Flow (PF) algorithm [78] targets the reconstruction of all stable particles present in the event by combining information from all available detector components. It starts with the reconstruction of tracks and calorimeter clusters. For the track reconstruction an iterative tracking algorithm is utilized [76]. The algorithm begins its first iterations with the reconstruction of tracks based on seeds that consist of at least three hits in the pixel detector, while tight criteria on a small required distance of the resulting tracks to the beam axis are applied. The next iterations proceed with successively looser distance and seed criteria by requiring reduced numbers of hits in the pixel detector or by allowing hits in the strip detectors to seed tracks. Thus, displaced tracks as well as short-lived particles are reconstructed and inefficiencies of the tracking system and interactions of charged particles with the detector material are taken into account. During the reconstruction process, clusters of pixel detector hits pointing to high energy deposits in the calorimeters are split into several individual hits to check for close-by tracks. Hits present in the muon system are used to seed tracks in the last steps of the algorithm. From the seeds tracks are reconstructed based on a *Kalman-filter (KF) technique*[1] [112]. The algorithm iteratively adds hits to the tracks taking energy losses and multiple scattering into account. The tracking approach is highly efficient in reconstructing tracks, while a low misidentification rate is achieved, since the most easily reconstructed tracks are studied first and the examined hits are removed afterwards. This drastically reduces the amount of hits in successive steps of the algorithm. Hence, the reconstruction of the more difficult tracks with missing hits or of displaced tracks is easier, while the probability to confuse hits of different tracks is kept low.

After the track reconstruction, energy clusters are formed from the energy deposits in the calorimeters, starting with the cells that measured the highest signals. If their neighboring cells also measured energies above a certain threshold, the central cell is chosen as a seed. From these seed cells *topological clusters* are built by combining all cells that measured energies above a certain threshold energy that lie next to a cell already part of the cluster. On these topological clusters a *Gaussian-mixture model* [78] is applied to disentangle signals arising from different particles. This algorithm iteratively determines the number of clusters and their positions by maximizing a likelihood that describes the energy distribution inside the topological cluster assuming Gaussian energy distributions for each particle. With this approach an efficient clustering, also for particles with low transverse momenta, is achieved, which is capable of resolving close-by energy deposits of two different particles.

In the next step of the PF algorithm the reconstructed tracks and calorimeter clusters are linked to form particle candidates, based on the assumption that different kind of particles produce different kind of signals in the detector. By extrapolating the tracks to the ECAL, the HCAL and to the preshower detectors, a link between a track and a calorimeter cluster can be established. The length of the extrapolated tracks is determined by the typical maximum size of an electromagnetic shower and the average interaction length of an hadronic shower. If the extended track coincides with a calorimeter cluster or with its neighbouring cells, the track and the cluster are considered to be linked. In order to reconstruct photons produced by electron bremsstrahlung, tangents to the tracks are projected to the ECAL. Tracks produced by electrons from photon conversions are reconstructed by a conversion finder [113] and linked to the tangents of the tracks. An ECAL cluster is linked to an HCAL or preshower cluster by checking if the cluster, or its neighbouring cells, in the detector part with the worse granularity coincide with the second studied cluster. Remaining tracks not yet linked to an object are projected to the muon system to search for matching hits. If none are found, the track in the tracking system is still kept as a tracker muon candidate (see Section 5.3). In cases where one object can be linked to several objects, the objects with the smallest distances in the η-ϕ-plane are chosen to arise from one particle. An exception is made for HCAL clusters,

---

[1] The Kalman-filter technique is a linear method for parameter estimation based on a χ^2-minimization that can handle quantities affected by noise and other fluctuations [112].

which can have multiple tracks linked to it.
From the linked objects muon and electron candidates (see Sections 5.3 and 5.4) are reconstructed. The remaining elements are employed to build photons as well as charged and neutral hadrons by comparing the energy of all tracks linked to an HCAL cluster to the total energy of the cluster. If the total energy of the tracks is significantly larger than the energy deposited in the HCAL, a search for muon candidates with looser criteria and a removal procedure for fake tracks is performed. Otherwise, all tracks give rise to a charged hadron and the corresponding energy is removed from the cluster. From the residual energy and calorimeter clusters not associated to tracks, a photon candidate is built if an associated ECAL cluster is present. Otherwise, a neutral hadron is created. In the region not covered by the tracking system ($|\eta| > 2.5$) no differentiation between charged and neutral particles can be achieved. Here, single ECAL clusters become photon candidates. ECAL clusters associated to HCAL clusters build hadron candidates [78].
The resulting reconstructed PF particles are subsequently passed to specialized algorithms, described in the remainder of this chapter.

5.2 Reconstruction and identification of primary vertex candidates

For the determination of primary vertex candidates [76, 114] all interaction vertex candidates present in the event are reconstructed first. Tracks found by the PF algorithm that pass certain quality criteria are used for the reconstruction with a *deterministic annealing* (DA) algorithm [115]. The algorithm builds vertex candidates based on the distance of the tracks to the beam spot in z-direction. Afterwards, candidates with at least two tracks are passed to the *adaptive vertex fitter* [116], which gives weights w_i to each track i [76, 114]. These weights are measures for the probability that the studied track is part of the vertex candidate. Tracks with a weight $w_i < 0.5$ are discarded. From the weighted tracks the positions of the vertex candidates are determined using the Kalman-filter technique [112]. All weights are re-evaluated afterwards. This procedure is repeated iteratively until the vertex position converges. The weights are also used to estimate the quality of the vertex fit by defining the number of degrees of freedom n_{dof} in this fit as

$$n_{\mathrm{dof}} = -3 + 2\sum_{i=1}^{N_{\mathrm{Tr}}} w_i. \tag{5.1}$$

Here, N_{Tr} denotes the number of tracks the vertex candidate consists of.
To qualify as a primary vertex candidate the fitted vertex candidates have to have $n_{\mathrm{dof}} > 4$ and distances to the nominal interaction point smaller than 2 cm in radial direction and less than 24 cm in the z-coordinate. Of the reconstructed vertex candidates passing these criteria, the vertex with the maximum sum of the squared transverse momenta of its assigned tracks is associated with the hard interaction and denoted as the primary vertex candidate of the event. All other vertices are regarded to arise from pileup or secondary interactions.

With the described method, which is also used in **Run-II**, the CMS experiment is able to reconstruct primary vertex candidates with a precision between 10 µm to 12 µm in each of the three spatial coordinates [76].

5.3 Reconstruction and identification of muon candidates

Both analyses presented in this thesis rely on an efficient reconstruction and identification of muons. Three different reconstruction algorithms are utilized by the CMS experiment: *the Standalone Muon reconstruction*, *the Tracker Muon reconstruction* and *the Global Muon reconstruction* [117, 118]. All approaches are based on tracks, either reconstructed exclusively in the tracking system or built from muon system hits only. The former are called *tracker tracks*, while the latter are referred to as *standalone-muon tracks*. The three different reconstruction approaches are summarized below.

> The **Standalone Muon reconstruction** exploits standalone-muon tracks. The reconstruction of the muon candidate's track is seeded by *muon segments*, which are short linear tracks consisting of hits in a chamber of the DTs or CSCs. Starting from these seeds, a fit of the signals found in the muon system is performed using the Kalman-filter technique [112] with the requirement that the track direction is compatible with the beam spot.
>
> The **Tracker Muon reconstruction** is based on tracker tracks with a total momentum of $p > 2.5\,\text{GeV}$ and $p_T > 0.5\,\text{GeV}$. All tracker tracks passing these criteria are extrapolated to the muon system, where the compatibility with at least one muon segment is tested. Since only few hits in the muon system are needed for the reconstruction of a Tracker Muon, better results are obtained with this reconstruction for low-p muons ($p \lesssim 5\,\text{GeV}$) than with the Global Muon reconstruction.
>
> In the **Global Muon reconstruction** a standalone-muon track is linked to a tracker track. The corresponding global-muon track is obtained by fitting all hits utilizing the Kalman-filter technique [112].

During the linking process of hits in the tracking and muon system the Tracker Muon and the Global Muon reconstruction consider the bending of the muon candidate's trajectory in the magnetic field, its expected average energy loss while traversing the detector, and Coulomb scattering.

The momentum of the muon candidate is determined through the bending of its trajectory in the magnetic field due to the Lorentz force. For muons with $p_T \gtrsim 200\,\text{GeV}$ the momentum obtained with the Global Muon reconstruction can have a better resolution than the one determined in the tracking system-only fit because of the larger available lever arm. Therefore, the momentum extracted by the global fit is utilized if both fits assign a momentum higher than 200 GeV to the track and if the ratio of the charge to the momentum q/p determined in the global fit agrees within two standard deviations with the one extracted by the tracking system-only approach. If one of these criteria fails, the momentum is taken from the tracking system-only fit [117].

For high-p_T muons additional reconstruction and momentum assignment techniques are available. Since these are not used in the presented analyses, these approaches are not described here. A description can be found in [117].

Following their reconstruction, different working points for identification can be applied on the muon candidates [119, 120]. Tighter working points yield lower identification efficiencies, but purer muon candidate samples.

Run-I identification criteria

In the presented 8 TeV analysis the **tight identification working point** [119], which is based on the **Particle Flow identification requirements** [78, 119], is applied.

> The **Particle Flow identification criteria** [78] require that the muon candidate is reconstructed and identified with the PF algorithm, described in Section 5.1. The muon identification in the PF algorithm distinguishes between isolated and non-isolated muon candidates. The isolation of a global muon candidate is checked by adding the transverse momenta of all tracker tracks and the energy deposits of all calorimeter clusters in a cone of $\Delta R \leq 0.3$ around the muon candidate. If this sum does not exceed 10% of the muon candidate's p_T, it passes the PF identification. Muon candidates that fail the isolation requirement either have to fulfill certain hard identification criteria (detailed in [121]) and requirements on the reconstructed track [78], or are required to have a high-quality standalone-muon track or tracker track consisting of a certain large amount of hits in the corresponding subdetectors of the CMS experiment. In the latter case the linked energy deposits in the calorimeters additionally have to fulfill compatibility criteria testing the minimal ionizing particle hypothesis to pass the PF identification requirements.
>
> To pass the **tight identification criteria** [119] the following requirements have to be met by the muon candidate:
>
> - The muon candidate passes the Particle Flow identification criteria and is reconstructed by the Global Muon reconstruction.
> - The global fit of the muon track yields $\chi^2/n_{\mathrm{dof}} < 10$.
> - The global fit of the muon candidate's track contains at least one hit in the muon chambers.
> - Muon segments in two or more muon stations are assigned to the tracker track of the muon candidate. This also means that it is reconstructed by the Tracker Muon reconstruction.
> - The muon candidate consists of a tracker track whose closest approach to the primary vertex in the transverse plane is within 2 mm.
> - The tracker track lies within a distance of 5 mm to the primary vertex in z-direction.

- The tracker track is reconstructed with at least six hits in the tracking system whereby at least one hit in the pixel detector is included.

Run-II identification criteria

For Run-II a new identification working point, called **medium** working point, has been developed by the CMS experiment [120]. It offers a higher identification efficiency than the tight working point, while the misidentification rate lies at a comparable level. Nevertheless, the identification efficiency of muons produced in decays of B hadrons is quite high with the medium working point as well, while a lower efficiency is obtained with the tight working point. The medium working point is used in the presented 13 TeV analysis and detailed below.

The **medium identification requirements** [120] include the following criteria:

- The muon candidate fulfills the **loose identification criteria**. These require that the muon candidate passes the Particle Flow identification working point and is additionally either reconstructed by the Global Muon or Tracker Muon reconstruction.
- More than 49% of the hits included in the tracker track are valid.

Furthermore, one of the selection sets mentioned hereafter have to be passed:

- The segment compatibility[2] is above 0.451.

or

- The muon candidate is reconstructed by the Global Muon reconstruction.
- The global fit of the track of the muon candidate has a $\chi^2/n_{\mathrm{dof}} < 3$.
- The fit performed to match the trajectories of the Tracker Muon and the Standalone Muon has to yield a $\chi^2 < 12$.
- The χ^2 of the fit performed by the kink finder[3] on the tracker track is smaller than 20.
- The segment compatibility is higher than 0.303.

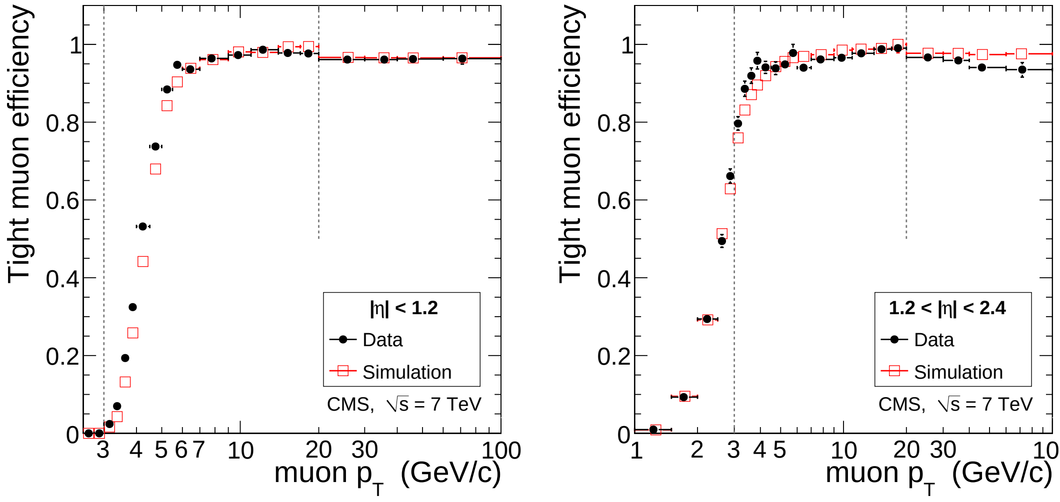
Isolation criteria in Run-I and Run-II

In addition to the identification requirements mentioned above, isolation criteria can be applied to further reduce the number of muons produced in decays of heavy quarks or heavy hadrons that are misidentified as prompt muons. In the presented analyses the **Particle Flow relative isolation** is used [117, 119, 120]. For this isolation criteria the transverse momenta of all PF charged hadrons assigned to the primary vertex, all PF neutral hadrons, and all PF photons within $\Delta R < 0.4$ to the muon candidate's track

---

[2] The segment compatibility, which can have values between 0 and 1, describes the quality of the spatial agreement between muon segments and the reconstructed track of the muon candidate.

[3] The kink finder algorithm searches for an interaction of the studied particle with the layers of the tracking system by performing a fit of the section of the track before the layer with the one after the layer. If an interaction happened, the fit will not work properly because of a change in curvature of the track.



(a) (b)

Figure 5.1: Comparison of the efficiencies of the tight muon identification working point in data (*black*) and simulation (*red*) as a function of p_T for muon candidates with (a) $|\eta| < 1.2$ and (b) $1.2 < |\eta| < 2.4$ [117].

are added and their ratio to the total transverse momentum of the track is studied. In order to reduce the effect of pileup on the isolation variable, the **$\Delta\beta$-correction** [122] is applied. 50% of the energy of all charged PF hadrons that lie in the isolation-cone but are not assigned to the primary vertex is subtracted from the energy sum. This corresponds to the ratio of neutral to charged hadrons in pileup interactions [122]. The recommended **tight requirements** on the isolation are 0.12 and 0.15 for Run-I and Run-II, respectively.

The performance of the muon reconstruction and identification has been measured in samples enriched in $J/\psi \to \mu^+\mu^-$ and $Z \to \mu^+\mu^-$ events using pp collision data recorded at $\sqrt{s} = 7\,\text{TeV}$ by the CMS experiment [117]. The efficiency, measured with the *Tag-and-Probe* method, is shown in Fig. 5.1 as a function of the p_T of the muon candidate in two regions of $|\eta|$ for the tight identification working point [117]. The plateau efficiency in data, reached for $p_T \approx 10\,\text{GeV}$, amounts to circa 96%. The misidentification rate of hadrons as muon candidates, which is caused by hadrons reaching the muon system, in-flight decays of hadrons and a mismatching of an hadron track in the tracking system to a muon track of a muon produced in a jet, is smaller than 1% (0.1%) for all working points (the tight working point). The muon transverse momentum scale has been calibrated using the same data samples and muon candidates from cosmic ray data. For muons with $p_T < 100\,\text{GeV}$ the transverse momentum resolution lies between 1% and 6% and below 10% for central muons with larger p_T [117].

5.4 Reconstruction and identification of electron candidates

While events with electron candidates are vetoed in the 13 TeV analysis in order to suppress Standard Model background processes, they constitute an important part of the signal selection in the presented 8 TeV analysis. The reconstruction of electron candidates and the applied identification requirements in the two analyses are discussed here. The description is based on [123], where more details are given.

The reconstruction of electron candidates is based on standalone algorithms in addition to the PF reconstruction. The former are needed because the high bremsstrahlung energy losses that occur when an electron traverses the tracking system of the CMS experiment alter the track curvature such that the standard track reconstruction and track-cluster matching algorithms yield low efficiencies.

The reconstruction algorithm starts with the clustering of energy deposits in the ECAL to *superclusters (SC)*. This step aims at collecting the energy deposited by the electron as well as photon candidates that are produced through bremsstrahlung. The clustering is performed using the so called *hybrid algorithm* in the central part of the detector and the *multi*-5×5 *algorithm* in the forward region. Starting points for both algorithms are cells that measure the highest energy deposits above a threshold energy. If energy requirements are passed, cells in certain windows around these seed cells are added to arrays and the resulting cluster arrays are combined with adjoining arrays to form superclusters. The main difference between the two algorithms lies in the window adopted for the clustering. In the hybrid algorithm these windows are enlarged in the ϕ-direction to take into account the spread between the bremsstrahlung photons and the electron trajectory due to the bending of the latter in the magnetic field. In the multi-5×5 algorithm a symmetric window in η and ϕ is employed. The built superclusters are linked to matching clusters in the preshower detectors if any exist.

PF tracks and energy clusters, built as described in Section 5.1, are used together with the superclusters to seed the reconstruction of electron tracks. The track seeding targets the identification of hits in the tracking system that belong to electron candidates, which in turn can be used to reconstruct their tracks. Two approaches, the *ECAL-based seeding* and the *tracker-based seeding*, are combined for this purpose. In the ECAL-based seeding, the supercluster is projected to the tracking system under the assumption that the energy-weighted position of all arrays contained in the supercluster corresponds to the coordinates the electron would have if it traversed the tracking system without radiation of bremsstrahlung. By applying requirements on the maximal distance between the predicted hits in the tracking system and the actual measured signals, track seeds are selected.

The tracker-based seeding operates on PF tracks that can be associated with a PF energy cluster in the ECAL. If the standard track fit yields a bad quality, or if only a few hits could be assigned to the track, the electron possibly lost large amounts of its energy due to bremsstrahlung. Thus, the track fit is repeated with the *Gaussian sum filter (GSF)* [124]. The GSF reconstructs the electron candidate trajectory by describing the energy loss due to bremsstrahlung in each layer of the tracking system. In order to identify electron seeds from the PF and GSF tracks a multivariate analysis (MVA) is used, based on the number of hits included in the tracks as well as the quality of the matching between the tracks and the calorimeter clusters and of the track fits.

Starting from the seeds identified by the ECAL- and tracker-based approaches, hits belonging to the electron tracks are selected using the PF track reconstruction, whereby energy losses are taken into account. Afterwards, a GSF fit is used again to reconstruct the tracks based on the selected hits. ECAL clusters produced by candidates for bremsstrahlung photons are associated with the tracks by applying the procedure described in Section 5.1.

The momentum of the resulting electron candidates is determined by combining measurements in the ECAL and the tracking system. The amount and the structure of the emitted bremsstrahlung, the position of the electron candidates and the resolution of the measurements in the ECAL and the tracking system define which of the measurements dominates the combined momentum. Details of this procedure can be found in [123].

Run-I identification criteria

All electron candidates considered in the presented 8 TeV analysis have to pass the identification criteria as recommended by [125]. This includes the requirements listed below:

- The supercluster of the electron candidate is not allowed to lie within $1.4442 < |\eta_{\mathrm{SC}}| < 1.5660$, which is the transition region between the barrel and the endcaps of the electromagnetic calorimeter.
- The distance of the GSF track to the primary vertex has to be smaller than 0.02 cm in the transverse plane at the point of their closest approach.
- A photon conversion veto is applied. This rejects electron candidates produced by the conversion of photons in the tracking system by checking if the electron candidates' tracks emerge from a common vertex and share a tangent at this vertex. The χ^2 of the corresponding fit as well as the impact parameter and its significance are used to discard the photon conversion tracks [123].
- The output of an MVA has to exceed 0.5. Input for the MVA are, among others, variables based on the track-cluster matching, the energies deposited in the calorimeters and the shape of an electromagnetic shower. Furthermore the amount of emitted bremsstrahlung as well as GSF and KF track information are studied. Before electron candidates are passed to the MVA, they have to pass loose preselection requirements, which are similar to selections performed at HLT level [123].
- In each of the modules of the tracking system a hit has to be present.

Run-I isolation criterion

In order to reduce the contamination with electron candidates that are misreconstructed jets, or electrons produced in decays of quarks or hadrons inside jets, the **Particle Flow relative isolation** requirement is applied. To pass this isolation requirement the sum of the transverse momenta of all charged PF hadrons assigned to the primary vertex as well as neutral PF hadrons and PF photons within $\Delta R < 0.3$ with respect to the electron candidate, corrected for pileup, is not allowed to exceed 10% of the transverse momentum of the electron candidate. To correct the sum for pileup, the average of the energy density ρ is determined for each event. Afterwards, ρA_{eff} is subtracted from the isolation sum. The density ρ is the median of the energy per jet area of the jets in the event clustered with the k_T algorithm (see Section 5.5) with $R = 0.6$, $|\eta| < 2.5$ and $p_T > 3\,\text{GeV}$. The observable A_{eff} is determined for each particle type in the isolation sum by weighting R^2 with the fraction in the energy density ρ associated to the particle type.

Run-II identification criteria

Electron candidates studied in the 13 TeV analysis are required to pass the subsequent criteria for $|\eta_{\text{SC}}| \leq 1.479$ ($|\eta_{\text{SC}}| > 1.479$) [123,126,127]:

- The lateral shower extension in η described by $\sigma_{\eta\eta} = \sqrt{\frac{\sum(\eta_i-\bar{\eta})^2 w_i}{\sum w_i}}$ has to be smaller than 0.011 (0.0314). For the calculation of $\sigma_{\eta\eta}$ the positions η_i of each crystal i in an array of 5×5 crystals around the crystal containing the highest transverse momentum in the superclusters, their average $\bar{\eta}$, and a weight w_i, which is a function of the logarithm of the measured energy in each crystal i, are used.

- Electron candidates reconstructed with the ECAL-based seeding have to pass $\Delta\eta < 0.00477$ (0.00868). The variable $\Delta\eta$ is defined as the difference between the pseudorapidity of the reconstructed track projected from the center of the tracking system to the ECAL without consideration of energy losses and the pseudorapidity of the energy-weighted position of all arrays contained in the supercluster at their point of closest approach.

- The variable $\Delta\phi$, defined analogously to $\Delta\eta$, has to fulfill $\Delta\phi < 0.222$ (0.213).

- The ratio of the energy deposited in the HCAL that can be matched to the supercluster and the energy of the latter has to be smaller than 0.298 (0.101).

- The energy of the supercluster E_{SC} and the momentum p of the track determined at the production vertex have to fulfill $|\frac{1}{E_{\text{SC}}} - \frac{1}{p}| < 0.241$ (0.14).

- At most one hit is allowed to be missing in the modules of the tracking system.

- The conversion veto is applied.

The efficiency of the applied electron identification and isolation criteria in the 8 TeV analysis has been determined with the Tag-and-Probe method in a sample dominated by $Z \to e^+e^-$ events using pp collision data recorded at $\sqrt{s} = 8\,\text{TeV}$ by the CMS experiment [128,129]. The efficiency for electron candidates with $p_T > 50\,\text{GeV}$ is slightly above 90%. An average efficiency of approximately 90% has also been measured for the applied identification criteria in the 13 TeV analysis [126].

The momentum scale and resolution of electron candidates have been determined by the CMS experiment using pp collision data enriched in $Z \to e^+e^-$, $J/\psi \to e^+e^-$ and $\Upsilon \to e^+e^-$ events. The data were collected at $\sqrt{s} = 8\,\text{TeV}$ and correspond to

$19.7\,\mathrm{fb}^{-1}$ [123]. The scale of electron candidates is known up to an uncertainty of less than 0.3%. The momentum resolution of electron candidates depends on the amount of emitted bremsstrahlung and the pseudorapidity of the electron candidate. It ranges between 1.7% and 4.5% [123].

5.5 Reconstruction of jets

The hadronization process of partons produced in the hard interaction results in collimated jets, which are measured in the experiment. In order to relate the measurements of the energy and positions of the hadrons inside the jets to the initial parton a proper jet definition is needed [130]. A jet definition includes a jet algorithm, which provides a clustering scheme of particles into a jet, a set of parameters, and a recombination scheme, which represents a rule how the momentum of the jet is calculated. A good jet definition should not only be easily implementable in experiments and theoretical calculations but especially be *infrared and collinear (IRC) safe* [130], which means that the jets found by the algorithm should not change if a soft emission or a collinear splitting is added to the event. IRC safety is important for the cancellation of real and virtual singularities in higher-order calculations. Therefore, the interpretation of the event has to be insensitive to these effects to be able to compare experimental data to theory. Moreover, using infrared and collinear unsafe algorithms in fixed-order perturbative QCD calculations can lead to infinite cross sections, which are unphysical. Due to finite resolutions and momentum thresholds, detector measurements can even out IRC unsafety to some degree. But since this is highly dependent on the details of the detector and the way the measurements are performed, a comparison to theoretical calculations is hardly possible. This section and the next section follow the description in [130], which provides additional details.

5.5.1 Jet clustering algorithms

In general, two different types of jet algorithms are distinguished: *cone algorithms* and *sequential recombination algorithms*. **Cone algorithms** are based on the assumption that the direction of the main energy flow in an event is not changed by hadronization and splitting processes. Most cone algorithms use seeds, which can be e.g. the particle with the highest p_T in the event, to define the initial direction of the jet. In this case, all particles in a cone of radius R are added, according to the utilized recombination scheme, to the initial seed particle. In *iterative cone algorithms* the resulting direction is used as a new seed direction. This procedure is repeated until a stable cone direction is achieved. In *fixed cone algorithms* no iteration over the cone orientation is performed. The usage of seeds makes cone algorithms infrared and collinear unsafe. Therefore, *seedless cone algorithms* were developed, which try to find stable cones by studying all possible combinations of particles. The drawback of these kind of algorithms is the computing time needed to study the large amount of different possibilities to build stable cones. Even though a faster implementation, which only studies a subset of all combinations determined by geometrical arguments is available, called *SISCone algorithm* [131], sequential recombination algorithms require less computation time.

Sequential recombination algorithms iteratively combine the closest particles using a metric that is specific to each algorithm. One of the advantages of sequential recombination algorithms is that they attach a clustering history to the event, which can be used to invert the clustering sequence to find substructure in large-cone jets (see Section 7 for more details). At the CMS experiment, the **anti-k_T algorithm** and the **Cambridge/Aachen algorithm** are commonly used. For both algorithms two distance measures d_{iB} and d_{ij} are introduced, where

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{5.2}$$

and

$$d_{iB} = p_{T,i}^{2p} \tag{5.3}$$

with $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, $p = -1$ for the anti-k_T algorithm, $p = 0$ for the Cambridge/Aachen algorithm and $p = 1$ for the inclusive k_T algorithm [130]. The parameter R is a dimensionless parameter of the clustering algorithm, which can be interpreted as a cone size. All three algorithms advance as follows. For all possible combinations of particles i and j from a list of input particles the distances d_{ij} are calculated. Moreover, for each particle i the value d_{iB} is determined. For all these values, it is evaluated which of all distances d_{ij} and d_{iB} is smallest. If the minimal value is d_{iB}, the particle i is considered to be a jet and removed from the input particle list. If the minimal distance is d_{ij}, the two particles i and j are combined according to the chosen recombination scheme. The default recombination scheme utilized at the CMS experiment is to add the four-momenta of the particles, which results in massive jets (*E-scheme*). Afterwards, all possible values of d_{ij} and d_{iB} are calculated and compared again. The procedure ends when all particles have been assigned to a jet.

The described approach of clustering particles into jets is infrared and collinear safe. In case of the k_T algorithm, soft and collinear particles are combined first and in later steps of the algorithm combined to form jets. The Cambridge/Aachen algorithm only uses angular information, particles close to each other will be recombined first. The anti-k_T algorithm tends to combine hard particles first.

Figure 5.2 illustrates the shapes in the ϕ-y-plane of jets clustered with different algorithms and $R = 1$. For this study, one event has been generated on parton-level using Herwig [132] and randomly distributed soft particles have been added to the event [130]. It can be clearly seen that only the jets clustered with the anti-k_T and SISCone algorithm have regular circular shapes. The jet areas produced by the Cambridge/Aachen algorithm and the k_T algorithm have very irregular shapes, which are in addition very dependent on the distribution of soft particles added to the event [130].

In the CMS experiment, different types of objects are used as input to the jet algorithms. PF particles are used to cluster the jets used in this thesis. These jets are referred to as **PF jets** in the following. Additionally, generator-level jets built out of all stable final-state particles on generator level except for neutrinos are considered. These jets are denoted as **particle-level jets**. Furthermore, jets with different cone sizes are

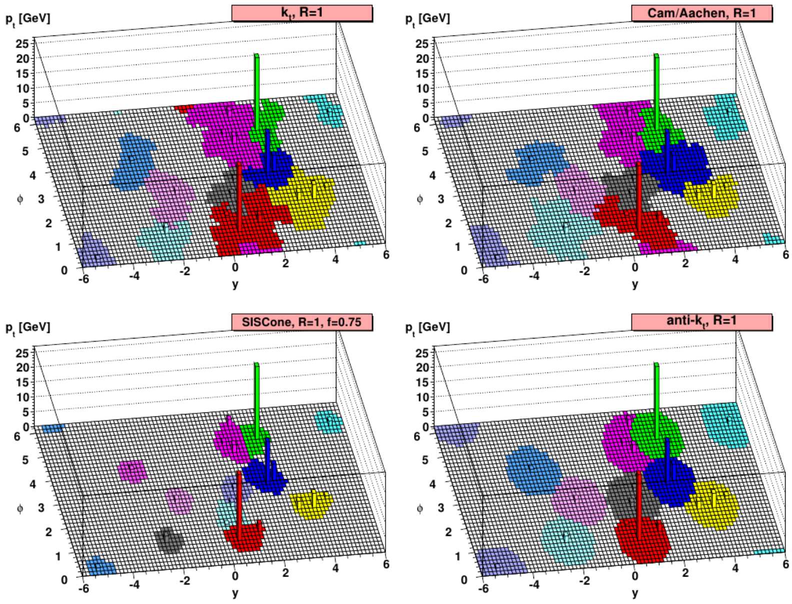


Figure 5.2: Illustration of the shapes of jets clustered with the k_T algorithm (upper left), the Cambridge/Aachen algorithm (upper right), the SISCone algorithm (lower left) and the anti-k_T algorithm in the ϕ-y-plane. Different jets are represented by different colors. The figure is taken from [130].

available. Small-cone jets clustered with a distance parameter of $R = 0.5$ ($R = 0.4$) are studied in Run-I (Run-II). Distance parameters of $R = 0.8$ and $R = 1.5$ are applied for the used large-cone jets. The implementation of the jet algorithms provided by the FastJet package [133] is used in the CMS experiment.

5.5.2 Pileup mitigation techniques

The large amount of pileup activity present in LHC pp collision data alters the reconstruction of jets and their properties, e.g. jet masses. In order to reduce the effect of pileup on jet observables, different pileup mitigation techniques are available in the CMS experiment. A summary of these can be found in [134]. Two different algorithms, summarized in the following, are used in studies presented in this thesis.

Charged hadron subtraction

Charge Hadron Subtraction (CHS) [134] is a method targeting the mitigation of charged in-time pileup in the detector region where tracking information is available ($|\eta| < 2.5$). It discards charged PF particles that are assigned to pileup vertex candidates and not to the primary vertex candidate, before the jet clustering is performed. A charged PF particle is assigned to a vertex candidate if the normalized χ^2 of the fit of the PF particle's track to the examined vertex candidate, reconstructed without the track in question, yields $\chi^2/n_{\mathrm{dof}} < 20$. The studied vertex in this procedure can either be the primary vertex candidate or additional vertices that pass requirements categorizing them as pileup vertices. Only charged particles assigned to the primary or no vertex at all are considered for the subsequent jet clustering step.

By applying CHS, the number of reconstructed jets that contain exclusively particles from pileup interactions is reduced. Additionally, an improved angular resolution in general and a better p_T resolution for low-p_T jets is observed [134].

Pileup Per Particle Information

Similar to CHS, Pileup Per Particle Information (PUPPI) [6,135] is a pileup mitigation technique operating on PF particles before these are clustered into jets. In contrast to CHS, PUPPI is defined over the whole η range and also for neutral PF particles. For each PF particle i the variable α_i given by

$$\alpha_i = \log \sum_{\substack{j\in \text{ PF particles} \\ i\neq j}} \left(\frac{p_{T,j}}{\Delta R_{ij}}\right)^2 \Theta(R_0 - \Delta R_{ij}) \tag{5.4}$$

is calculated, where Θ denotes the Heaviside step function. The variable α_i is a measure for the structure of the distribution of the particles in a cone with size $R_0 (= 0.4)$ around particle i. Particles i produced in hard processes tend to be surrounded by other high-p_T particles from the same process, resulting in a large value of α_i. Contrary, particles produced in pileup interactions are randomly distributed and lie only accidentally close to a particle emerging from the hard interaction. For these particles α_i is small.

In the part of the CMS experiment covered by the tracking system ($|\eta| < 2.5$) only charged particles that are assigned to the primary vertex candidate (PV) or that lie within 3 mm to the primary vertex candidate in z-direction are considered for the calculation of α_i. The corresponding quantity, called α_i^C in the following, is therefore given by

$$\alpha_i^C = \log \sum_{\substack{j\in \text{ Ch, PV} \\ i\neq j}} \left(\frac{p_{T,j}}{\Delta R_{ij}}\right)^2 \Theta(R_0 - \Delta R_{ij}). \tag{5.5}$$

It is assumed that considering only charged particles assigned to the primary vertex candidate approximates all particles produced in the hard interaction. In the forward region ($|\eta| > 2.5$) all particles j are considered for the calculation of α_i. This is given by Eq. (5.4) and referred to as α_i^F. Since particles produced in pileup interactions tend to have small transverse momenta $p_{T,j}$, the contribution of particles from pileup to α_i^F is small.

In the next step of the algorithm α_i^C and α_i^F are calculated for all charged particles i in the event assigned to a pileup vertex candidate to get values of α_i^C and α_i^F that can be compared to the corresponding values for particles with an unknown source. The median values $\bar{\alpha}_{\text{PU}}^{C/F}$ and the RMS $\sigma_{\text{PU}}^{C/F}$ of these distributions are compared to the $\alpha_i^{C/F}$ values of all other particles i to decide whether these can be associated with pileup interactions or the hard interaction. The comparison is achieved by calculating the variable χ^2, where

$$\chi^2 = \frac{(\alpha_i^{C/F} - \bar{\alpha}_{\text{PU}}^{C/F})^2}{(\sigma_{\text{PU}}^{C/F})^2}. \tag{5.6}$$

From this, a weight w_i given by the cumulative distribution F of the χ^2 function for pileup particles is determined,

$$w_i = F(\chi_i^2). \tag{5.7}$$

Particles with a small weight are classified as pileup, while particles with a weight near unity have an α_i that strongly deviates from $\bar{\alpha}_{\mathrm{PU}}$. Hence, they are classified as particles from the hard interaction. Considering only charged particles classified as pileup from the central part of the detector in the calculation of $\bar{\alpha}_{\mathrm{PU}}^{C/F}$ and $\sigma_{\mathrm{PU}}^{C/F}$ implies the assumption that neutral and charged particles from pileup vertices behave similarly. In order to account for differences in the distributions of the particles from pileup interactions in the central and in the forward part of the detector, η dependent correction factors determined from simulated events are applied on $\bar{\alpha}_{\mathrm{PU}}^{C/F}$ and $\sigma_{\mathrm{PU}}^{C/F}$.

The four-momentum of each particle is multiplied with the weight w_i. Only particles that receive a weight $w_i > 0.01$ and a rescaled p_T above a threshold, which linearly scales with the number of reconstructed vertices in the event and depends on $|\eta|$, are kept and used to cluster jets [6, 135].

Jets corrected with PUPPI have a better mass resolution than jets corrected with CHS. Furthermore, quantities derived from PUPPI corrected jets show a very good stability over a wide range of different pileup conditions [134].

In Fig. 5.3 a simulated dijet event overlaid with 80 pileup events is shown [135]. By applying CHS the amount of particles produced in pileup interactions is greatly diminished in the central region of the event display. For both the CHS and PUPPI corrected event the jets are clustered correctly. In the case of PUPPI most of the pileup particles inside and outside of the jets are discarded, leading to a reproduction of many characteristics of the event, also in areas not covered by jets.

5.5.3 Jet identification

On jets corrected for pileup the **loose identification requirements for PF jets** [136] are applied. These include the following criteria for jets with $|\eta| \leq 2.4$ [136]:

- The share of the energy of neutral hadrons in the total energy of the jet must not exceed 98%.
- The proportion of neutral electromagnetic energy deposits with respect to the total energy of the jet has to be less than 99%.
- The jet consists of at least two particles.
- Less than 80% of the total energy of the jet can be assigned to a muon candidate.
- The jet must contain at least one charged particle.
- Energy deposits of charged hadrons have to be present in the HCAL.
- Less than 99% of the total energy of the jet can be associated to charged electromagnetic energy deposits.

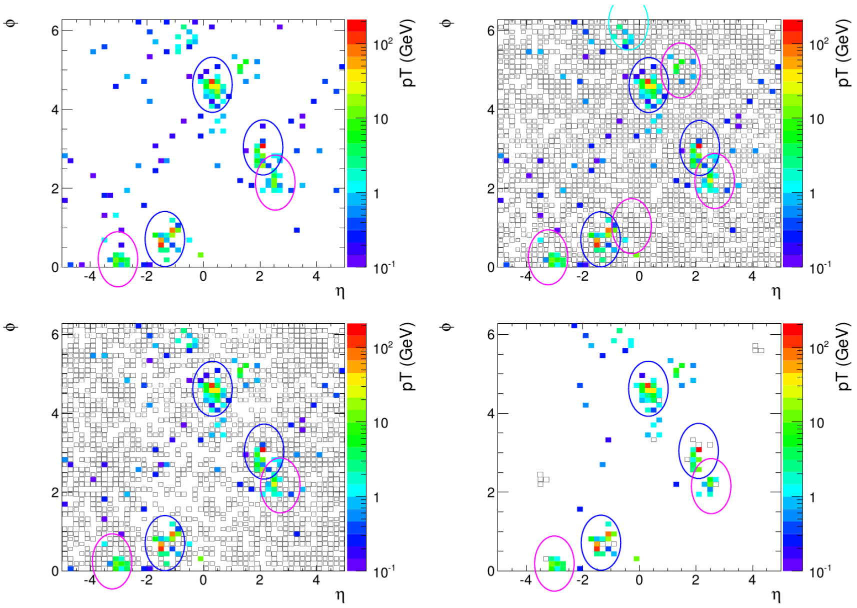


Figure 5.3: Simulated dijet event overlaid with 80 pileup events. Particles from the hard interaction are drawn as colored squares, the color-code is given next to the event display. The colored circles represent anti-k_T jets clustered with a cone size of R=0.7. Jets with 25 GeV $< p_T <$ 50 are shown in magenta, jets with 50 GeV $< p_T <$ 200 GeV in cyan. Higher-p_T jets are colored in blue. The uncolored boxes represent pileup particles, whereby their size scales logarithmically with the transverse momentum of the particles. Shown are the event displays for the hard interaction (*upper left*), the event with all particles including the particles produced in pileup interactions (*upper right*), the event with CHS applied (*lower left*) and the event display for the PUPPI corrected event (*lower right*). The L1 Offset Corrections described in 5.5.4 have been applied on the jets in the upper right and lower left event displays. The figure is taken from [135].

The application of the last three criteria is not possible for jets with $|\eta| > 2.4$ due to the absence of tracking information, which makes charged hadrons indistinguishable from neutral hadrons in this region.

In **Run-II** the same identification criteria are applied except for the requirement of the muon energy fraction, which is not recommended [136]. Moreover, jets with $|\eta| > 2.7$ have to fulfill additional criteria, which lead to higher identification and lower misidentification rates than before [136]:

– The amount of energy assigned to neutral electromagnetic energy deposits makes up less than 90% of the total energy of the jet.

– More than two (ten) neutral particles are included in the jet (for $|\eta| > 3.0$).

In simulation, the efficiencies of the identification requirements in Run-I and Run-II amount to 99% or more with an acceptable rate of other objects misreconstructed as jets [75, 136].

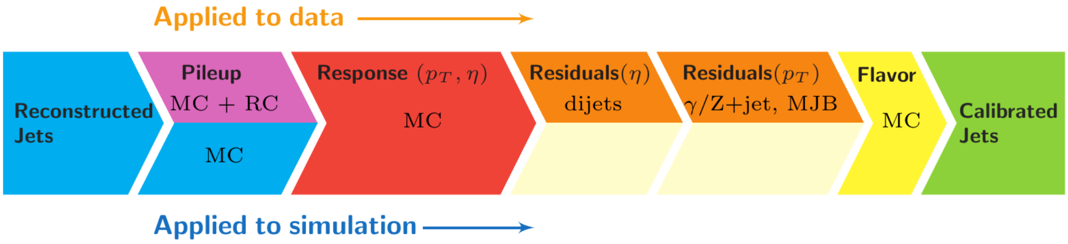


Figure 5.4: Sketch of the successively applied JEC [75]. Residual corrections are only applied on data events. The label MC indicates that the corrections have been determined using simulated events. The label dijets, γ/Z+jet and MJB (multijet background) show that data samples enriched in the indicated processes have been studied to derive the corrections. RC is the abbreviation of random cone method (see text for details).

5.5.4 Jet energy corrections

The measured transverse momenta of reconstructed jets do not agree with the transverse momenta of the jets formed by all stable visible particles in the final state after hadronization, denoted as particle-level jets. The size of the shift can be quantified by studying the *jet energy response*, defined as the ratio of the measured transverse momentum of the jet and the transverse momentum of the jet at the particle-level p_T/p_T^{ptcl}. The deviation of the measurement with respect to the particle level can be caused e.g. by the non-linear response of the detector to hadrons, nuclear interactions of the particles inside the jets with the material present in front of the calorimeters, inefficiencies of the track reconstruction, and the inclusion of particles belonging to pileup interactions in the jet p_T measurement. Therefore, *Jet energy corrections (JEC)* have to be applied. The JEC target a calibration of the jet momenta such that the average jet energy response, referred to as the *jet energy scale (JES)*, coincides with the jet energy on particle level. A sketch of the successively applied multiplicative corrections is shown in Fig. 5.4. The descriptions of the jet energy corrections given here is based on [75], which provides more detailed information.

The **L1 Offset corrections** aim at diminishing effects of pileup interactions on the jet energy response. Their application is not needed if PUPPI has already been applied. If CHS has been used instead, contributions from pileup interactions that arise from neutral particles and charged PF hadrons not assigned to any vertex as well as out-of-time pileup can still be present. The Pileup Offset correction is based on an η and p_T dependent estimate of the average energy offset ρ due to pileup and noise in simulation. The offset ρ is obtained by determining the median of the jet transverse momenta per jet area of all clustered jets in an event. The jet area is computed by adding randomly distributed soft particles covering the whole η-ϕ-plane to the event. Small residual differences between data and simulation are corrected for by using the random cone (RC) method [75].

The **L2 Relative Correction and L3 Absolute Correction** target the removal of the η and p_T dependencies of the jet energy response, respectively. The corrections are derived in bins of p_T^{ptcl} and jet η using simulated QCD multijet events

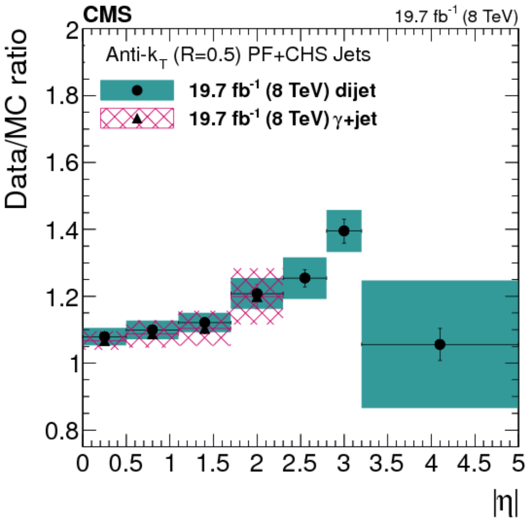


Figure 5.5: Run-I JER correction factors for simulation as a function of jet $|\eta|$ derived in samples enriched in dijet events (*colored area*) and γ+jet events (*hatched area*). Taken from [75].

and are applied to simulation and data.

To eliminate remaining differences between the jet energy response in data and simulation **L2L3 Residual corrections** are evaluated. They are applied on data only and are determined using data samples enriched in dijet events and $Z(\rightarrow \mu^+\mu^-/e^+e^-)$+jet, γ+jet and multijet events, respectively. The measurements are based on the fact that a deviation of the JES from unity causes a momentum imbalance between the studied jet and a well understood reference object recoiling against the jet.

Jet Flavor Corrections are not used in the presented analyses. The JES can be determined with very high precision. At $\sqrt{s} = 8\,\text{TeV}$, their total uncertainty is smaller than 3% for jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 5.0$ [75].

After the application of the JEC the *jet transverse-momentum resolution (JER)*, which describes the Gaussian width of the distribution p_T/p_T^{ptcl}, is estimated. For the determination of the JER in data and simulation the recoil of a jet against a reference object is used, where the reference object can either be a jet or a photon. The JER ranges between 5% and 20% for jets in the central part of the detector, larger values correspond to low-p_T jets. Correction factors for the JER in simulation dependent on η are calculated to compensate for differences to data. In Run-I the corrections, shown in Fig. 5.5, were found to be smaller than 20% in the central part of the detector [75]. Correction factors derived for the Run-II analyses have a similar magnitude.

5.5.5 *b* tagging

Jets produced through the hadronization of b quarks are characterized through their special properties. The technique for the identification of b quark jets is called *b tagging* [114, 137, 138].

In the hadronization process of jets originating from b quarks long-lived B hadrons are produced. Their decays can produce tracks that are displaced by a few millimeters from the primary vertex. Algorithms for b tagging exploit either the large impact parameter[4] of these tracks with respect to the primary vertex or the existence of secondary vertices, which are reconstructed from the displaced tracks. Furthermore, the large masses of B hadrons, the high transverse momenta of their decay products and the existence of light leptons, which are produced in 20% of all B hadron decays, are special features of b quark jets that are utilized by b tagging algorithms.

In the CMS experiment, many different b tagging algorithms are available. The **Combined Secondary Vertex (CSV) algorithm** and an improved version called **CSVv2** developed for Run-II are detailed in the following, a description of other algorithms can be found e.g. in [114].

Run-I b tagging: The Combined Secondary Vertex algorithm

Input for the CSV algorithm [114,137] are all tracks of the examined jet with $p_T > 1\,\mathrm{GeV}$ that fulfill quality criteria based on the number of hits in the tracking system, the quality of the track fit and on the distances of the tracks to the jet axis and to the primary vertex. These tracks are used to reconstruct the primary vertex candidate and secondary vertex candidates with the Adaptive Vertex Reconstruction (AVR) [139] algorithm (see Section 5.2). Afterwards, vertex candidates that have a large fraction of their tracks in common with the primary vertex, whose flight direction is too far away from the jet axis, or whose distance to the primary vertex and associated invariant mass is consistent with a decay of a K^0 meson are discarded.

The CSV algorithm then combines secondary vertex related observables like the track multiplicity, the invariant mass of the tracks associated with the vertex candidate, the position in η of the tracks measured relative to the jet axis, and the energy of the tracks compared to the total jet energy into an MVA discriminator. Furthermore, the total track multiplicity, the significance of the flight distance in the x-y-plane, and dedicated impact parameter significances are used.

An important feature of the CSV algorithm is that it is able to assign a discriminator value to jets in which no secondary vertex candidate could be reconstructed by adding the categories *no vertex* and *pseudo vertex* to the MVA. A pseudo vertex is defined as a vertex candidate that contains tracks with a large impact parameter significance. For the two additional categories a subset of the mentioned variables is used in the MVA.

The output of the CSV algorithm is a combined variable, which tends to zero for jets originating more likely from b quarks. It tends to unity if the probability that the jet is associated with a light quark is high. Three different working points, called **loose, medium, and tight**, are provided by the CMS experiment by defining cuts on this discriminator such that jets originating from light quarks with a p_T of circa 80 GeV have a probability to be identified as a b quark jet of roughly 10%, 1% and 0.1% on average [137].

---

[4] The impact parameter is the distance of two studied objects measured at their closest approach [137].

Improvements for Run-II: CSVv2

In Run-II secondary vertex candidates are reconstructed using the *Inclusive Vertex Finder (IVF)* algorithm [138], described in [140]. A main difference to the AVR is that the IVF operates on all tracks present in the event, not only on the tracks part of the studied jet. The considered tracks have to fulfill a transverse momentum requirement ($p_T > 0.8\,\mathrm{GeV}$) and additional quality criteria on the number of hits in the tracking system and the impact parameter. In the first step of the IVF, seed tracks are defined based on criteria on their impact parameter and its significance. On these seed tracks and nearby tacks the adaptive vertex fitter [116] is applied in order to reconstruct vertex candidates. If one of these vertex candidates has a large amount of tracks in common with another vertex candidate or if applied quality criteria fail, it is discarded. Additionally, tracks that fail quality criteria or can also be assigned to the primary vertex are removed from the reconstructed secondary vertex candidate. The adaptive vertex fit is repeated on the remaining vertex candidates that still contain more than one track after the cleaning procedure. The resulting secondary vertex candidates are assigned to reconstructed jets by performing a track matching between the two objects. The jets are subsequently passed to the CSVv2 algorithm.

Except for using vertices reconstructed with the IVF, the CSVv2 proceeds very similar to the CSV algorithm used in Run-I. The algorithms differ mainly in a more complicated multivariate technique used in the CSVv2, which is capable of operating on more input variables. From the MVA output loose, medium, and tight working points are defined with the same misidentification rates as for Run-I, but determined for jets with $p_T > 30\,\mathrm{GeV}$ [138].

In Fig. 5.6 a comparison of the performance of different b tagging algorithms is shown. Jets with $p_T > 30\,\mathrm{GeV}$ in simulated $t\bar{t}$ events at $\sqrt{s} = 13\,\mathrm{TeV}$ were used to derive these curves. At a given misidentification rate for tagging jets originating from light quarks, the CSVv2 algorithm outperforms the CSV algorithm in efficiency by up to 4%. Reconstructing vertices in the CSVv2 algorithm with the IVF instead of the AVR algorithm increases the efficiency by 1% to 2% [138].

5.6 Reconstruction and identification of tau lepton candidates

For the presented search for third-generation leptoquarks in the channel $\mathrm{LQ} \rightarrow t\tau^-$ an efficient reconstruction and identification of tau leptons is particularly important. Since tau leptons are, with a lifetime of $290.3 \times 10^{-15}\,\mathrm{s}$ [7], short-lived and decay after travelling only few millimeters, they can only be detected by the reconstruction of their decay products. Due to the large mass of tau leptons ($m_\tau = 1776.86\,\mathrm{MeV}$ [7]) hadronic decay modes are accessible and with a branching ratio of circa 64.8% [7] also the dominant ones. The most probable hadronic decay modes and the corresponding branching ratios are summarized in Table 5.1. Leptonic decay modes occur, with branching ratios of 17.41% [7] for the channel $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ and 17.83% [7] for $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$, less frequently.Distinguishing light leptons originating from tau lepton decays from prompt

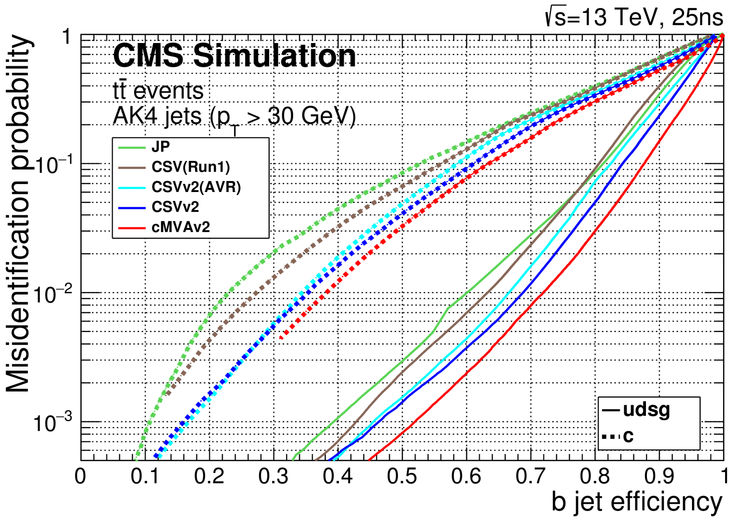


Figure 5.6: Misidentification rate of an u, d, s quark, or a gluon (*solid lines*) or a c quark (*dashed lines*) as a jet originating from a b quark in dependence of the efficiency to correctly assign a b quark jet for different b tagging algorithms (*colored lines*). JP (Jet Probability) and cMVAv2 are two further b tagging algorithms, described in [138]. The label CSVv2(AVR) indicates that vertices reconstructed with the AVR algorithm are used as input for the CSVv2, while the label CSVv2 shows that the IVF algorithm is used for the reconstruction of vertices. The figure is taken from [138].

light leptons is hardly possible. Hence, no distinct algorithm is available for the reconstruction of the leptonic decay modes. For the reconstruction of the hadronic decay modes the *hadron-plus-strips (HPS)* [141] algorithm is employed. The HPS algorithm aims at reconstructing the most important decays into hadrons, flagged with a check mark in Table 5.1, separately. The reconstruction of the decay $\tau^- \to h^-h^+h^-\pi^0\nu_\tau$ is currently not included in the HPS algorithm due to a large misidentification rate of jets as tau leptons in this channel.

5.6.1 The hadron-plus-strips algorithm

The HPS algorithm [141] operates on jets with $|\eta| < 2.5$ and an uncorrected $p_T > 14\,\text{GeV}$, which have been clustered using the anti-k_T algorithm with a distance parameter of $R = 0.5$. The first step of the algorithm aims at building neutral pion candidates, which occur in a large fraction of all targeted hadronic decays. Neutral pions decay with a probability of circa 98.8% [7] to two photons, which very likely convert into electron-positron pairs while traversing the tracking system. These decays are reconstructed by clustering the photon and electron candidates with $p_T > 0.5\,\text{GeV}$ found inside the jets into strips. The η- and ϕ- coordinates of the candidate with the highest p_T determine the starting position of the strip. Afterwards, the electron or photon candidate with the highest p_T found in a $(\Delta\eta \times \Delta\phi)$ region of (0.05×0.2) around the strip center is added to the strip. The larger strip size in ϕ is needed due to the curvature of the electrons' trajectories in the magnetic field. The new strip location is given by the energy-weighted average position of all photons and electrons in the strip. The clustering procedure is

hadronic decay mode	resonances	branching ratio [%]	reconstructed by HPS
$\tau^- \rightarrow h^- \nu_\tau$		11.5	✓
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0	✓
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5	✓
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8	✓
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$		4.8	
other hadronic decay modes		3.2	

Table 5.1: Summary of the hadronic decay modes of tau leptons [7, 141]. $h^\pm$ stands either for a charged pion or a charged kaon. If an intermediate resonance is produced during the decay of the tau lepton, the corresponding meson is listed in the column called *resonances*. Check marks indicate decay modes reconstructed by the hadron-plus-strips (HPS) algorithm [141].

repeated until all electron and photon candidates in the region around the strip center are merged into the strip. If the sum of the transverse momenta of all electron and photon candidates contained in the strip exceeds 2.5 GeV, the resulting strip is considered to be a π^0 candidate. Otherwise, the strip is discarded. The particle with the highest p_T of the remaining electron and photon candidates is afterwards utilized as starting point for a new strip. The strip building is performed until no unclustered photon or electron candidate is left.

Afterwards, up to two out of the six highest-p_T π^0 candidates are combined with one or three out of the six highest-p_T charged hadrons inside the jet. All possible combinations are tested. Combinations in which the sum of all charges is not $\pm e$ are discarded. Furthermore, combinations where one of the strips or charged hadrons has a large distance to the tau lepton candidate are not considered. Since the decay products of high-p_T tau leptons have a larger Lorentz-boost and are therefore closer to each other, the allowed distance shrinks with p_T.

For each of the remaining combinations the compatibility with one of the targeted hadronic decay modes is tested by comparing the invariant mass built by the examined strips and hadrons with the mass of the meson resonances involved in tau lepton decays. For the decay mode into three charged hadrons the compatibility with a common production vertex is imposed additionally. Charged particles not assigned to a strip are associated to the decay $\tau^- \rightarrow h^- \nu_\tau$. If one combination of strips and charged particles fulfills the requirement of more than one decay mode, the combination that results in the highest-p_T tau lepton candidate is kept.

On the tau lepton candidates reconstructed in this manner different identification criteria are applied, which are detailed in the following.

5.6.2 Isolation criteria and discriminators against muons and electrons misidentified as tau lepton candidates

Tau leptons and their decay products are in general well isolated from other particles in the event. Therefore, **isolation discriminators** [141] are very helpful for the reduction of the misidentification rate of the tau lepton identification. For the isolation discriminator I_τ the scalar sum of the transverse momenta of all charged PF particles and PF photons with $p_T > 0.5\,\mathrm{GeV}$ that were not used to build the tau lepton candidate and that lie in a cone of $\Delta R = 0.5$ around the direction of the tau lepton candidate is determined,

$$I_\tau = \left(\sum_{\text{charged},\Delta z<0.2\,\text{cm},\Delta r<0.03\,\text{cm}} p_T \right) + \max\left\{ 0, \sum_\gamma p_T - \Delta\beta \right\}. \tag{5.8}$$

As noted in the formula, the tracks of the considered charged particles additionally have to fulfill criteria on their distances to the production vertex of the tau lepton candidate in Δz and Δr. The $\Delta\beta$ in Eq. (5.8) term describes the effect of pileup on the reconstruction of the photon candidate. It is determined statistically and defined by

$$\Delta\beta = 0.46 \sum_{\text{charged},\Delta z>0.2\,\text{cm}} p_T. \tag{5.9}$$

For the $\Delta\beta$-correction, all charged particles that lie in a cone of $\Delta R = 0.8$ around the tau lepton candidate and whose tracks pass the distance requirements to the production vertex of the tau lepton candidate are considered. Depending on the selected working point, the isolation I_τ has to be smaller than 2.0 GeV (*loose*), 1.0 GeV (*medium*) and 0.8 GeV (*tight*).

Muons and electrons can be misidentified as tau lepton candidates decaying into charged hadrons or, in the case of electrons emitting bremsstrahlung, as tau lepton candidates decaying into an hadron and a neutral pion. Hence, the CMS experiment also provides discriminators for the reduction of these rates [141].

For the discriminator rejecting **muons misidentified as hadronically decaying tau leptons** two different working points, called *loose* and *tight*, are available.

For the **loose working point** the following criteria have to be met:

- Within $\Delta R \leq 0.3$ around the tau lepton candidate track segments are found in maximally one muon station.
- The energy deposited in the ECAL and HCAL exceeds 20% of the momentum of the tau lepton candidate's leading track.

In order to pass the **tight working point** the criteria are:

- The tau lepton candidate fulfills the criteria of the loose working point.
- No hits are found in the CSC, DT and RPC detectors of the two outermost muon stations that lie in a distance of $\Delta R \leq 0.3$ around the tau lepton candidate.

In order to be able to better discriminate **electrons falsely identified as tau leptons** from real tau leptons an MVA discriminator has been developed [141]. In this MVA approach different variables that take into account the different amount of energy deposited in the ECAL compared to the HCAL by electrons and hadrons, the share of the photon's energy in the total energy of the tau lepton candidate, the position of the photons found in a strip relative to the highest-p_T charged particle and the mass of the reconstructed tau lepton candidate are combined into a boosted decision tree (BDT) discriminator. Additionally, variables which exploit the larger amount of bremsstrahlung radiated by electrons than by hadrons and the quality and location of the reconstructed GSF track are included into the BDT. Tau lepton candidates found in the transition region between the barrel part and the endcaps of the ECAL ($1.45 < |\eta| < 1.56$), where no active detector material is present, are discarded. Different working points (called *very loose, loose, medium, tight*, and *very tight*) are defined by cutting on the BDT output variable.

Run-I identification criteria

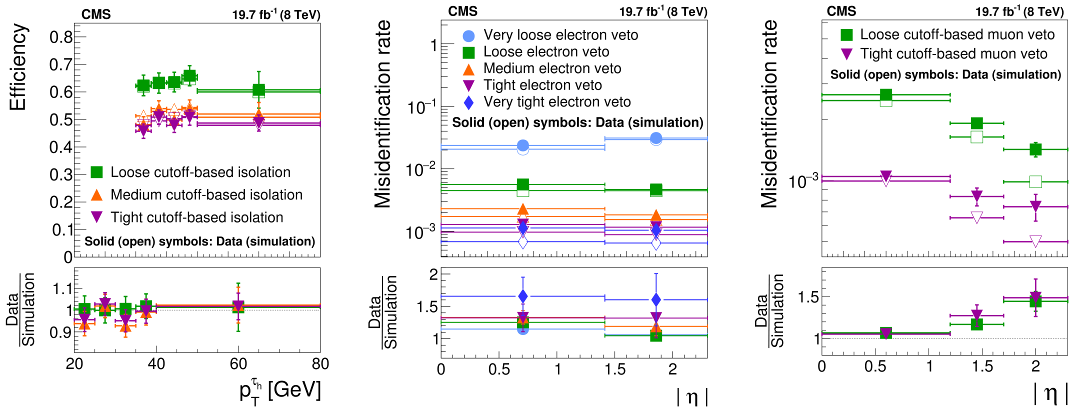
For the 8 TeV analysis two different combinations of the mentioned identification criteria are utilized, referred to as *decay mode finding* and *medium working point* [142].

> The **decay mode finding** criterion only requires that an hadronic decay mode has been reconstructed by the HPS algorithm, no further isolation or identification criteria are applied.
>
> For the **medium** working point the decay mode finding criterion has to be passed. Additionally, the medium isolation criterion is applied in combination with tight requirements on the discriminator against muons and the MVA discriminator against electrons.

Studies in simulated $Z/\gamma^* \to \tau^+\tau^-$ events [141] show that the efficiency of the HPS algorithm to reconstruct the hadronically decaying tau lepton in the correct decay mode lies between 83% and 97%. The efficiency of the applied medium isolation criterion is shown in Fig. 5.7 together with the misidentification rate of the used discriminators against muons or electrons reconstructed as tau lepton candidates [141]. They have been determined in data collected at 8 TeV with the CMS experiment, the former in a sample enriched in $Z/\gamma^* \to \tau^+\tau^-$, the latter two in $Z/\gamma^* \to e^+e^-$ and $Z/\gamma^* \to \mu^+\mu^-$ events, respectively. The efficiency of the medium isolation requirement amounts to circa 50%. The misidentification rates of electrons and muons range between less than one per mill and a few per cent. The efficiency of prompt tau leptons to fulfill the discriminators against muons and electrons varies between 95% to 99% and between 60% to 95% [141]. The tau lepton candidate energy scale, measured in a data sample enriched in $Z/\gamma^* \to \tau^+\tau^-$ events, is on average a bit lower than unity but is in agreement with simulation except for the $h^\pm\pi^0$ decay mode, where a difference of 1% is observed [141].

5.7 Measurement of event based variables

The variables $\not{E}_T$ and S_T are helpful observables to describe the characteristics of an event. They are defined in this section.



(a) (b) (c)

Figure 5.7: Comparison of (a) the efficiency of the isolation criterion as a function of the transverse momentum of the tau lepton candidate and the rates of (b) muons and (c) electrons misidentified as tau lepton candidates in dependence of $|\eta|$ of the tau lepton candidate between data (*solid symbols*) and simulation (*open symbols*) for the different working points (*colored*) [141].

5.7.1 Measurement of $\not{E}_T$

Stable particles that interact solely weakly escape the experiment without leaving signals in the detector. Thus, their properties are not directly measurable with algorithms like the ones described in the previous sections. Nevertheless, conclusions about their existence and properties can be drawn by measuring the transverse momentum imbalance that these particles cause in the visible event. The imbalance in the transverse momentum plane is described by the vector $\vec{\not{E}}_\mathrm{T}$ [143]. It is calculated, assuming momentum conservation in the transverse plane, as the negative vectorial sum of the transverse momenta of all detected final state particles. The absolute value of this quantity is called missing transverse energy and is referred to as $\not{E}_T$. For a precise determination of $\not{E}_T$ accurate measurements of the energies of all visible particles and a well working detector are essential since mismeasurements of energies and misinterpreted signals create spurious $\not{E}_T$.

Several types of $\not{E}_T$, differing in the applied corrections and in the particles studied to determine the momentum imbalance, are available in the CMS experiment. In this thesis **type-1-PF-$\not{E}_T$** [143] is utilized.

Type-1-PF-$\not{E}_T$

For the calculation of type-1-PF-$\not{E}_T$ [143] all PF particles present in the event are considered. Furthermore, type-1-PF-$\not{E}_T$ is corrected for the non-uniform jet energy response by subtracting the energy difference between calibrated ($\vec{p}_{T,\,\text{jet}}^{\,\text{corr}}$) and uncalibrated ($\vec{p}_{T,\,\text{jet}}^{\,\text{raw}}$) jets from the transverse momentum sum of all PF particles. Hence, type-1-PF-$\not{E}_T$ is given by

$$\not{E}_T = -\sum_{\text{PF particles}} \vec{p}_T - \sum_{\text{jets}} \left(\vec{p}_{T,\,\text{jet}}^{\,\text{corr}} - \vec{p}_{T,\,\text{jet}}^{\,\text{raw}}\right) \tag{5.10}$$

Only jets with a calibrated transverse momentum larger than 10 GeV (15 GeV) in Run-I (Run-II) and only jets whose share of ECAL energy deposits in the total energy is less than 90% are considered.

5.7.2 Measurement of S_T

The variable S_T is helpful to discriminate signal events that contain many high-p_T particles in the final state against background events. It is defined as the scalar sum of the transverse momenta of all jets and leptons in the event together with $\not{E}_T$, and is given by

$$S_T = \left(\sum_{\text{jets, leptons}} |\vec{p}_T| \right) + \not{E}_T. \tag{5.11}$$

In contrast to $\not{E}_T$, S_T depends on the applied event selection.

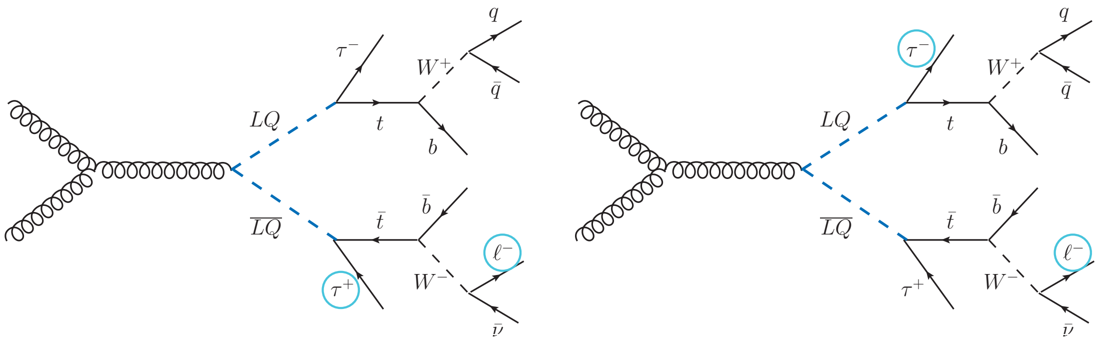
6 Search for third-generation scalar leptoquarks decaying into a top quark and a tau lepton at $\sqrt{s} = 8$ TeV

A search for pair production of third-generation leptoquarks decaying into a top quark and a tau lepton is presented. The chapter is organized as follows. The search strategy is described in Section 6.1. A summary of the studied data samples and the used triggers is presented in Section 6.2, while simulated background and signal processes are listed in Section 6.3. Section 6.4 describes the applied event filter and the quality criteria imposed on all objects considered in the search. The signal selection is detailed in Section 6.5. The estimation of the Standard Model backgrounds is presented in Section 6.6, its validation in Section 6.7. In Section 6.8 the evaluation of the systematic uncertainties is summarized. The results of this search are presented in Section 6.9, followed by their interpretation in Section 6.10. The chapter ends with a short outlook (Section 6.11).

6.1 Search strategy

Due to the rich final state produced by two leptoquarks decaying into a top quark and a tau lepton, many different search strategies can be pursued. Two possible decay chains with the studied event signature are depicted in Fig. 6.1. Performing the search by requiring an isolated light lepton candidate provides a good handle to suppress the QCD multijet background, which is the dominant background at the LHC. In the signal process, isolated light leptons (electrons or muons) are produced in leptonic decays of top quarks or tau leptons. On the other hand, tau leptons, which are produced either directly in leptoquark decays or in top quark decays, exhibit a high probability of approximately 64.8% to decay hadronically (see Section 5.6). Therefore, this search requires at least one isolated muon or electron candidate, one isolated hadronically decaying tau lepton candidate, denoted as τ_h in the following, and at least three jets. The final state of the signal is characterized by the presence of up to six jets since top quarks almost exclusively decay into a b quark and a W boson, and W bosons in turn decay with a probability of approximately 67.4% [7] into two quarks. Events containing a $\mu\tau_h$-candidate pair will be referred to as *muon channel* in the following, events with a $e\tau_h$-candidate pair will be called *electron channel.*

Standard Model processes, which are usually produced with much higher cross sections than leptoquarks, can mimic the signature of the signal in the detector. The Standard Model background of this search mainly consists of $t\bar{t}$ + jets, W + jets and Drell Yan (DY) +jets events, with importance decreasing in this order. Smaller contributions arise



(a) (b)

Figure 6.1: Sketch of an example decay chain that produces the studied event signature $pp \to \mathrm{LQLQ} \to \ell\tau_h + X$ with (a) an opposite-sign and (b) a same-sign $\ell\tau_h$-pair.

from single top quark production, diboson events, $t\bar{t}$ production in association with a Z boson and QCD multijet events. These background processes can be categorized into reducible and irreducible processes. In the context of this study, irreducible background processes are processes with prompt pairs of tau leptons and light leptons in the final state. These are produced if in $t\bar{t}$ decays both W bosons decay into leptons, or in case of $t\bar{t}$ $+Z$ production if the $t\bar{t}$-pair decays into a lepton and jets and the Z boson into leptons. DY + jets events become irreducible if the Z boson decays into two tau leptons, one of which decays leptonically and one hadronically. The same also applies to diboson events. All other processes, together with lepton+jets decays of $t\bar{t}$ events and all other possible decays of DY + jets and diboson events are reducible backgrounds and are selected if a jet is misidentified as an hadronically decaying tau lepton candidate. Therefore, the misidentification rate of jets as tau lepton candidates has been studied in detail in the course of this thesis. The fraction of events with a misidentified muon or electron candidate is negligible. This especially concerns the QCD multijet background, which only contributes if one jet is misreconstructed as a light lepton candidate and a second jet as an hadronically decaying tau lepton candidate.

For the statistical evaluation of the search result the transverse momentum distribution of the leading tau lepton candidate divided into four different p_T regions is utilized in the muon and electron channel, resulting in eight different search regions. The p_T regions are given by $20\,\mathrm{GeV} \leq p_T < 60\,\mathrm{GeV}$, $60\,\mathrm{GeV} \leq p_T < 120\,\mathrm{GeV}$, $120\,\mathrm{GeV} \leq p_T < 200\,\mathrm{GeV}$ and $p_T \geq 200\,\mathrm{GeV}$. The transverse momentum is very well suited for the final statistical interpretation of the result. The background processes tend to produce tau lepton candidates with lower transverse momenta, while the tau lepton candidates in signal events have higher transverse momenta since they are produced in the decay of a heavy leptoquark in most cases. Thus, the first one or two p_T regions, dependent on the leptoquark mass, can be used to constrain the normalization of the Standard Model background processes, while the remaining two regions provide sensitivity to the signal processes.

First studies of the presented $\mu\tau_h$-channel have been performed in the master thesis of the author of this work [144]. These studies have been extended and have been published in a first combination with a same-sign analysis in the muon channel in [1] in the course of this thesis. The studies for the electron channel have been started in the master thesis of [128]. These efforts have been continued by the author of this work, who also performed the combination of the $\mu\tau_h$- and $e\tau_h$-channels. This combination has been published together with the combination with the same-sign analysis in the muon channel in [2].

6.1.1 Same-sign analysis in the muon channel

The result of the analysis presented in this thesis is combined with a search [145, 146] for pair production of third-generation scalar leptoquarks decaying into a top quark and a tau lepton in the $\mu\tau_h$-channel, in which only events with same-sign charges of the muon and tau lepton candidate are studied. This same-sign analysis is denoted as EXO-12-030 in the following. The muon and the tau lepton can have same-sign charges if e.g. the muon candidate is produced through the top quark decay and the tau lepton produced in the decay of the respective second leptoquark is reconstructed as the hadronically decaying tau lepton candidate (see Fig. 6.1 for a sketch of an example decay chain). In total, this applies to roughly 50% of all signal events. In the SM, background processes with same-sign $\mu\tau_h$ pairs appear much less frequently, such that the SM background can be suppressed significantly by selecting only events with same-sign $\mu\tau_h$ pairs. This selection results in a much lower background yield and a very different background composition. By requiring a pair of same-sign muon and tau lepton candidates, backgrounds with prompt muon and tau lepton pairs are almost completely discarded. This leaves only backgrounds where a jet has been misidentified as a tau lepton candidate, or - but much less frequently - as a muon candidate. The different background composition can be exploited in a data-based estimation of Standard Model background processes as it has been done in EXO-12-030. This advantage balances the significantly smaller signal efficiency, which results in a better sensitivity for low leptoquark masses and a worse sensitivity for high masses if compared to the search presented in this thesis. The statistical interpretation of EXO-12-030 is performed in two different regions that exploit the centrality of signal events to enhance the search sensitivity. A more detailed description of EXO-12-030 can be found in [145, 146].

6.2 Data samples and trigger

The search for third-generation leptoquarks decaying into a top quark and a tau lepton is performed using pp collision data collected by the CMS experiment at a center-of-mass energy of 8 TeV. The analyzed data amount to an integrated luminosity of $19.7\,\text{fb}^{-1}$ [147, 148], which corresponds to the complete data set recorded in the year 2012. A detailed description of the machine parameters of the LHC in this time period is given in Section 3.3.

The analysis is based on the requirement of a muon or an electron candidate. These can be produced in decays of top quarks or tau leptons emerging from the leptoquark decays.

Since the studied leptoquarks are relatively lightweight (with masses between 200 GeV and 800 GeV), the top quarks are produced with rather small transverse momenta, and therefore with a moderate Lorentz boost. The leptons produced in top quark decays are thus expected to be isolated from the other final state objects. Leptons produced in the decays of tau leptons are also well isolated since the accompanying neutrinos are not detectable in the experiment. Therefore, isolated lepton triggers are very suitable for selecting the events of interest. Events considered in the muon channel are selected by a trigger requiring isolated muon candidates with $|\eta| < 2.1$ and $p_T > 24$ GeV at HLT level. The online isolation, calculated as described in Section 5.3 using trigger level PF objects, has to be less than 0.15. Within the CMS experiment this trigger is denoted as HLT_IsoMu24_eta2p1. For the electron channel a trigger selecting isolated electron candidates with $p_T > 27$ GeV that fulfill loose identification criteria at HLT level has been chosen (HLT_Ele27_WP80).

Detailed studies of the muon trigger efficiency have been performed by the group of the CMS experiment responsible for muon triggering, reconstruction and identification [149] using the Tag-and-Probe method in a data sample dominated by $Z \to \mu^+\mu^-$ events. The efficiencies of the electron trigger have been determined in the course of [128] by applying the Tag-and-Probe method in a data sample enriched in $Z \to e^+e^-$ events [129]. The muon trigger reaches its plateau efficiency between approximately 80% and 90%, depending on the pseudorapidity, for muon candidates with $p_T > 40$ GeV [149]. The electron trigger reaches a plateau efficiency of around 90% for electron candidates with $p_T > 50$ GeV [128, 129]. Performance checks have been made for both lepton triggers to ensure that they operated reliably and with roughly constant efficiencies over the complete duration of data acquisition. In both channels, a trigger simulation is used for simulated events and correction factors that diminish the differences to data are taken into account. A more detailed discussion of the trigger efficiencies and the correction factors is given in Section 6.6.2.

The data streams combining events collected by the single-muon triggers and single-electron triggers are called SingleMuon and SingleElectron data streams, respectively. The integrated luminosities corresponding to these data streams are summarized in Table 6.1 divided into the four runs of the 2012 data acquisition periods.

6.3 Simulated signal and background processes

Simulated events are used to describe all Standard Model processes that constitute a background for the search, as well as signal processes. Standard model $t\bar{t}$ and single top quark events have been generated using POWHEG (v1.0) [95–98]. PYTHIA (v6.426) [101] has been utilized to produce simulated QCD multijet and diboson events. W + jets events and DY + jets events have been simulated with up to four additional partons in the matrix element using the MadGraph (v5.1) [93] generator. For the generation of $t\bar{t}$ + Z events also the MadGraph (v5.1) generator has been chosen, but only up to one additional parton has been added to the matrix element.

Signal events with leptoquark masses between 200 GeV and 800 GeV have been produced

data set	run	L_{int} [fb^{-1}]
/SingleMu/Run2012A-22Jan2013-v1/AOD	2012 A	0.876
/SingleMu/Run2012B-22Jan2013-v1/AOD	2012 B	4.411
/SingleMu/Run2012C-22Jan2013-v1/AOD	2012 C	7.055
/SingleMu/Run2012D-22Jan2013-v1/AOD	2012 D	7.369
/SingleElectron/Run2012A-22Jan2013-v1/AOD	2012 A	0.876
/SingleElectron/Run2012B-22Jan2013-v1/AOD	2012 B	4.411
/SingleElectron/Run2012C-22Jan2013-v1/AOD	2012 C	7.055
/SingleElectron/Run2012D-22Jan2013-v1/AOD	2012 D	7.369

Table 6.1: Summary of the integrated luminosities of the single-muon and single-electron data streams collected by the CMS experiment at $\sqrt{s} = 8\,\mathrm{TeV}$ divided into the four runs of the 2012 data acquisition periods [147, 148].

using PYTHIA. Only pair production of leptoquarks that decay into a top quark and tau lepton is simulated. Other possible decay channels are not taken into account. The cross sections of pair production of scalar leptoquarks in *pp* collisions have been determined with NLO accuracy [37].

For all simulated events PYTHIA has been utilized to simulate the parton shower and the hadronization process. For samples produced with MADGRAPH the matching between particles associated with the matrix element process and particles produced in the shower and the hadronization processes has been performed using the MLM matching scheme [108]. For the proper simulation of tau lepton decays the MC generator TAUOLA [110] is interfaced with the generator. GEANT4 [111] has been used to simulate the passage of the particles through the detector.

To enable a comparison of simulation and data events, the simulated samples are weighted according to their cross section and scaled to the total integrated luminosity of the collected data. For the weighting the most accurate calculations for the cross sections that were available when the analysis was published [2] were used [101, 150–154]. Table 6.2 summarizes for simulated background and signal samples the event generator, the cross section used for the re-weighting, the available accuracy of the theory calculation of the cross section, and the pdf set the sample has been generated with.

6.4 Event filters and object quality criteria

For the analysis, only data events that have been certified by the CMS experiment are considered. Hence, it is ensured that only events are studied that were recorded with a completely functional detector. Sometimes problems in the data acquisition have been found only after the certification. These include anomalous signals reconstructed in the ECAL or HCAL, signals near ECAL cells that are marked as dead, events for that the track reconstruction failed, interactions of the beams with residual gas molecules still present in the LHC, or interactions of the beams with the optics used to focus and deflect

process	generator	σ [pb]	order theory calc.	pdf set
LQ ($M = 200\,\text{GeV}$)	PYTHIA	17.4 [37]	NLO	CTEQ6L1 [91]
LQ ($M = 250\,\text{GeV}$)	PYTHIA	5.26 [37]	NLO	CTEQ6L1
LQ ($M = 300\,\text{GeV}$)	PYTHIA	1.89 [37]	NLO	CTEQ6L1
LQ ($M = 350\,\text{GeV}$)	PYTHIA	$7.69 \cdot 10^{-1}$ [37]	NLO	CTEQ6L1
LQ ($M = 400\,\text{GeV}$)	PYTHIA	$3.42 \cdot 10^{-1}$ [37]	NLO	CTEQ6L1
LQ ($M = 450\,\text{GeV}$)	PYTHIA	$1.63 \cdot 10^{-1}$ [37]	NLO	CTEQ6L1
LQ ($M = 500\,\text{GeV}$)	PYTHIA	$8.20 \cdot 10^{-2}$ [37]	NLO	CTEQ6L1
LQ ($M = 550\,\text{GeV}$)	PYTHIA	$4.31 \cdot 10^{-2}$ [37]	NLO	CTEQ6L1
LQ ($M = 600\,\text{GeV}$)	PYTHIA	$2.35 \cdot 10^{-2}$ [37]	NLO	CTEQ6L1
LQ ($M = 650\,\text{GeV}$)	PYTHIA	$1.32 \cdot 10^{-2}$ [37]	NLO	CTEQ6L1
LQ ($M = 700\,\text{GeV}$)	PYTHIA	$7.61 \cdot 10^{-3}$ [37]	NLO	CTEQ6L1
LQ ($M = 750\,\text{GeV}$)	PYTHIA	$4.48 \cdot 10^{-3}$ [37]	NLO	CTEQ6L1
LQ ($M = 800\,\text{GeV}$)	PYTHIA	$2.69 \cdot 10^{-3}$ [37]	NLO	CTEQ6L1
$t\bar{t}$ + jets	POWHEG	245.8 [151]	NNLO	CT10 [92]
W + jets	MadGraph	37509 [150]	NNLO	CTEQ6L1
$Z(\to \ell\ell)$+ jets ($10\,\text{GeV} < m(\ell\ell) < 50\,\text{GeV}$)	MadGraph	850 [150]	NNLO	CTEQ6L1
$Z(\to \ell\ell)$+ jets ($m(\ell\ell) > 50\,\text{GeV}$)	MadGraph	3504 [150]	NNLO	CTEQ6L1
Single t, s channel	POWHEG	3.79 [152]	approx. NNLO	CT10
Single $\bar{t}$, s channel	POWHEG	1.76 [152]	approx. NNLO	CT10
Single t, t channel	POWHEG	56.4 [152]	approx. NNLO	CT10
Single $\bar{t}$, t channel	POWHEG	30.7 [152]	approx. NNLO	CT10
Single t, tW production	POWHEG	11.1 [152]	approx. NNLO	CT10
Single $\bar{t}$, tW production	POWHEG	11.1 [152]	approx. NNLO	CT10
WW	PYTHIA	54.8 [153]	NLO	CTEQ6L1
ZZ	PYTHIA	8.059 [153]	NLO	CTEQ6L1
WZ	PYTHIA	33.2 [153]	NLO	CTEQ6L1
$t\bar{t}Z$	MadGraph	0.2057 [154]	NLO	CTEQ6L1
QCD (electron and photon enriched)	PYTHIA	$3.6 \cdot 10^{8}$ [101]	LO	CTEQ6L1
QCD (B and C meson enriched)	PYTHIA	$3.6 \cdot 10^{8}$ [101]	LO	CTEQ6L1
QCD (muon enriched)	PYTHIA	$3.6 \cdot 10^{8}$ [101]	LO	CTEQ6L1

Table 6.2: Summary of the studied simulated samples including the MC generator used to simulate the respective process, the theoretical cross section utilized to weight the process with respect to all other background samples, the theoretical accuracy of that cross section and the pdf set the sample has been generated with.

the beams. To remove these events, the CMS experiment recommends the application of certain filters [143,155,156]. All recommended filters are applied.

In order to ensure that only well understood objects are studied, certain quality criteria are imposed on all considered particle candidates and jets. The applied criteria are based on the reconstruction algorithms and identification criteria detailed in Chapter 5.

- All considered primary vertex candidates are reconstructed as described in Section 5.2 and have to fulfill the identification criteria detailed there. All candidates passing these criteria are denoted as *good primary vertex candidates.* In each event considered in the following at least one good primary vertex candidate has to be present.
- The tight identification and isolation requirements are imposed on all muon candidates. Only muon candidates with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$ are taken into account.
- Electron candidates not fulfilling the identification and isolation criteria described in Section 5.4 are discarded. Additionally, a p_T-criterion of $p_T > 35\,\text{GeV}$ and a requirement on the pseudorapidity ($|\eta| < 2.1$) is enforced on all studied electron candidates. Electron candidates within $\Delta R < 0.5$ of a jet are not considered to avoid double counting of objects due to misidentification.
- All studied tau lepton candidates have to satisfy the decay mode finding criterion, $p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$.
- Charge Hadron Subtracted jets clustered out of PF particles using the anti-k_T jet algorithm with a distance parameter of $R = 0.5$ are utilized. All jets have to pass the loose identification requirements for PF jets. The JEC as well as the JER correction factors are applied. Only jets with a corrected transverse momentum of $p_T > 30\,\text{GeV}$ and within $|\eta| < 2.5$ are considered.
- The Type-1-PF description of $\not{E}_T$ is used.

All tracks of studied objects have to be assigned to the primary vertex candidate of the event to make sure that all particle candidates emerge from the same interaction. The p_T thresholds and $|\eta|$ restrictions of the selected light lepton candidates have been chosen such that the correction factors for the trigger efficiencies mentioned in Section 6.6.2 are applicable.

6.5 Signal selection

After the quality criteria have been applied, an event selection is constructed that efficiently selects signal events and strongly suppresses Standard Model background processes. The design and the optimization of the event selection have mainly been performed in the course of [144] and [128]. This section summarizes the approach of the selection and the results of its optimization.

To select signal events and efficiently suppress the background from QCD multijet events either at least one muon candidate or at least one electron candidate has to be present. Events that contain a muon candidate are vetoed in the electron channel. In this way the orthogonality of the two channels is ensured and they can be statistically combined for the final interpretation of this search.

In addition to the light lepton candidate, only events with one hadronically decaying tau lepton candidate passing the decay mode finding criterion are selected. Together with the requirement of $S_T > 350\,\mathrm{GeV}$, which describes the scalar sum of the transverse momenta of all jets and particle candidates in the event together with $\not{E}_T$ (see Section 5.7), and at least two jets with $p_T > 50\,\mathrm{GeV}$ and $|\eta| < 2.5$, this set of selection criteria defines the *pre-selection*. The main backgrounds after this selection step are $W+$jets and $t\bar{t}+$jets production. The requirements of the pre-selection are relatively loose such that Standard Model background processes are suppressed, while most of the signal events are kept. Based on this sample different selections with more stringent criteria can be studied. However, the pre-selection is also constructed such that the number of expected signal events is still negligible compared to the number of expected background events. Thus, the pre-selection defines a control region that can be used to verify the description of data by simulated events. This is discussed in detail in Section 6.7.

Besides the basic selection requirements of the pre-selection, many further selection steps are possible, due to the very rich final state of the studied signal process. Different selections consisting of various combinations of requirements on variables with high separation power between background and signal events have been studied [128, 144]. The selection with the best sensitivity has been found by comparing, as a function of the leptoquark mass, the expected upper limit at 95% confidence level (C.L.) on the cross section of pair produced scalar leptoquarks decaying with a branching ratio of 100% into a top quark and a tau lepton. All tested selections include the requirement of at least one tau lepton candidate passing the medium working point. The expected limits are determined by a binned likelihood fit in the four regions of the p_T distribution of the leading tau lepton candidate. Only statistical uncertainties of the simulated samples are considered in this study.

The two top quarks produced by the signal process decay almost exclusively into a b quark and a W boson [7]. Since the W bosons decay with a probability of approximately 67.4% [7] hadronically, signal events are characterized by a final state with two or four jets in addition to the two b quark jets. Therefore, the studied selections have been based on a requirement of at least three jets with different p_T thresholds. Due to the two b quarks, the number of b-tagged jets can be an interesting variable to study as well. However, b tagging has been found to deteriorate the expected sensitivity since the main background of this search, $t\bar{t}+$jets, also contains two b quarks and thus cannot be suppressed with this selection.

Since the final state particles of the signal process are produced through decays of heavy objects, all these particles tend to have higher transverse momenta than final state particles from Standard Model processes. Hence, tight cuts on transverse momenta and $\not{E}_T$, which is caused by neutrinos produced in the decays of the top quarks or tau leptons, have been tested. Jet p_T thresholds of 50 GeV, 100 GeV and 150 GeV imposed on the jet with the highest transverse momentum have been studied. Additionally, two different cuts on the transverse momentum of the jet with the third-highest p_T, $p_T > 30\,\mathrm{GeV}$ and $p_T > 50\,\mathrm{GeV}$, have been investigated. The best results have been obtained by selecting events with $p_T > 100\,\mathrm{GeV}$ for the leading jet, $p_T > 50\,\mathrm{GeV}$ for the sub-leading jet and $p_T > 30\,\mathrm{GeV}$ for the jet with the third-highest transverse momentum. Cuts on $\not{E}_T$ of $\not{E}_T > 50\,\mathrm{GeV}$ and $\not{E}_T > 100\,\mathrm{GeV}$ have been studied. The lower $\not{E}_T$ requirement has been chosen for the analysis. The light lepton p_T thresholds, set by the requirements of the trigger, have been found to be already optimal and are not changed for the analysis. The loose p_T requirement of $20\,\mathrm{GeV}$ on the hadronically decaying tau lepton candidate is also kept since in this way the different p_T spectra of the background and the signal processes can be exploited in the statistical interpretation of the result.

A variable very well suited to characterize events with many final state objects that have high transverse momenta is S_T. Selections with cuts on S_T between $400\,\mathrm{GeV}$ and $1\,\mathrm{TeV}$ in steps of $100\,\mathrm{GeV}$ have been tested. Only very little variations in the search sensitivity have been observed for S_T thresholds between $400\,\mathrm{GeV}$ and $700\,\mathrm{GeV}$. Tighter criteria have resulted in improved expected limits for leptoquark masses above approximately $450\,\mathrm{GeV}$. The best expected limit for high leptoquark masses has been obtained for a selection with $S_T > 1\,\mathrm{TeV}$, while the sensitivity for low leptoquark masses has been only slightly deteriorated.

Selections with a same-sign charge requirement of the $\mu\tau_h$-candidate pair, as utilized in EXO-12-030, have been found to give comparable results as the above presented selections for low leptoquark masses. At high masses (above circa $500\,\mathrm{GeV}$), the expected limit was deteriorated significantly. This is expected since for high leptoquark masses the produced tau leptons have a very high p_T. In this p_T region hardly any background processes are present. Therefore, a high signal selection efficiency rather than a good background suppression is needed to improve the sensitivity. Thus, the relatively low signal selection efficiency of the same-sign requirement worsens the sensitivity in that phase space.

In order to be able to combine the analysis described in this thesis with EXO-12-030, events considered in EXO-12-030 are discarded in this analysis. This concerns events with a $\mu\tau_h$-candidate pair, where the muon candidate and the tau lepton candidate are charged equally while the tau lepton candidate passes the tight tau lepton identification criteria, described in Section 5.6. Afterwards, the muon channel of this analysis consists of events with same-sign $\mu\tau_h$-candidate pairs where the tau lepton candidate passes the medium identification criteria but fails the tight identification criteria, and of events with opposite-sign $\mu\tau_h$-candidate pairs. Since the $e\tau_h$-channel is not studied in EXO-12-030, the electron channel is not affected by this event veto.

Finally, the event selection applied in this search can be summarized. The pre-selection is defined by the following criteria:

- The event contains at least one muon candidate with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 2.1$, fulfilling the tight identification and isolation requirements or at least one electron candidate passing the identification and isolation criteria described in Section 5.4

with $p_T > 35\,\text{GeV}$ and $|\eta| < 2.1$.

- Events that contain a muon candidate with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$ fulfilling the tight identification and isolation requirements are discarded in the electron channel.
- At least one hadronically tau lepton candidate with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$ passing the decay mode finding criterion is present.
- Each event has at least two jets with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.5$.
- The event has to pass $S_T > 350\,\text{GeV}$.

Events pass the final selection if they fulfill the additional criteria hereafter:

- Each event has at least three jets, whereby p_T thresholds of 100 GeV, 50 GeV and 30 GeV are imposed on the leading p_T jet, the sub-leading p_T jet and the jet with the third-highest momentum, respectively.
- The presence of at least one tau lepton candidate with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$ fulfilling the medium working point is required.
- The event has to pass $\not{E}_T > 50\,\text{GeV}$ and $S_T > 1\,\text{TeV}$.
- Events with a same-sign $\mu\tau_h$-candidate pair where the tau lepton candidates passes the tight identification criteria are discarded in the muon channel.

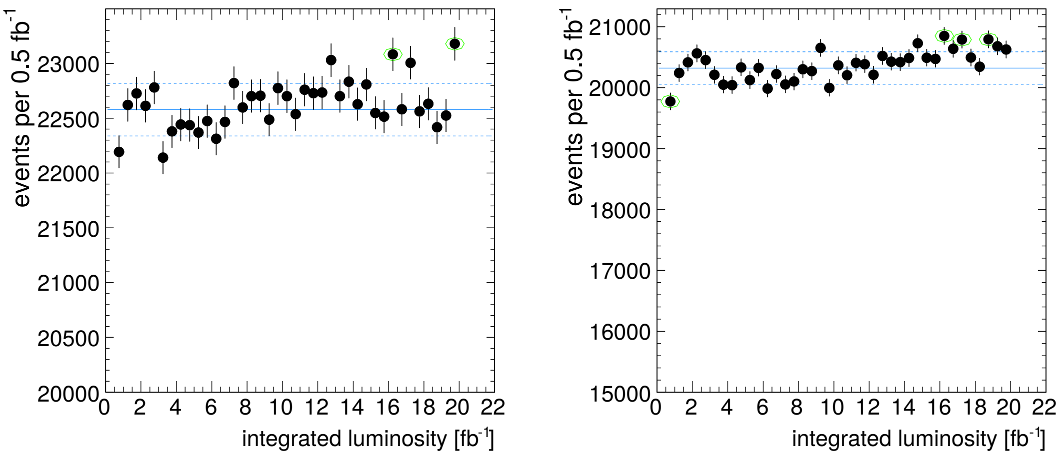
Furthermore, only tau lepton candidates passing the medium identification requirements are studied in the following.
A particle candidate might be listed twice as two different objects. To avoid double counting, certain distance requirements are imposed on the considered particle candidates:

- Tau lepton candidates have to have a distance $\Delta R > 0.5$ to all muon candidates.
- Jets with $\Delta R < 0.5$ to a tau lepton candidate are discarded.

Finally, also different options for the choice of the variable used for the final statistical interpretation have been studied in [144]. Comparisons of the expected limits based on the transverse momentum of the leading tau lepton candidate, S_T and the sum of the transverse momenta of the leading tau lepton candidate and the leading muon candidate have revealed that the best results can be obtained using the distribution of the p_T of the leading tau lepton candidate.

Time stability of data acquisition conditions

In order to check whether the data acquisition has worked smoothly over the considered time period or whether unwanted effects such as changes in the trigger efficiency or detector failures occurred, the event yield as a function of the integrated luminosity, which can be interpreted as a measure of time, is studied. In Fig. 6.2 the number of



(a) (b)

Figure 6.2: Event yield per $0.5\,\mathrm{fb}^{-1}$ after the pre-selection has been applied in (a) the muon channel and (b) the electron channel. A fit to data with a constant, represented by the blue line in the figure, has been performed. The dashed blue lines show the 1σ-band of the fitted average value. If a data point deviates with more than 3σ from the average, it is marked with a green circle.

selected events per $0.5\,\mathrm{fb}^{-1}$ is shown for the muon and the electron channel after the pre-selection. A constant number of events per integrated luminosity indicates accurate data acquisition and that the conditions were stable over the considered time period. The event yield is fitted by a constant to check whether the assumption that no trend is present is valid. Overall a very stable performance is observed. A more detailed look reveals a slight increase of the event yield with increasing integrated luminosity. This is explainable by an increase of the trigger efficiencies in both channels. In both cases, this effect is covered by the uncertainties in the trigger correction factors described in Section 6.2 and the uncertainty in the luminosity [157] (see Section 6.8 for a detailed description of the applied systematic uncertainties).

6.6 Estimation of Standard Model backgrounds

The estimation of the Standard Model background processes is performed using simulated events. In order to ensure a realistic detector simulation and an accurate description of SM background processes, multiple correction factors determined in control regions are applied on reconstruction and identification efficiencies and misidentification rates of particle candidates in simulation. These corrections are either measured by responsible groups within the CMS experiment or are determined within this analysis. All corrections will be described in this chapter. A detailed validation of the background estimation with simulated events is described in the next section (see Section 6.7).

6.6.1 Pileup re-weighting

Although the simulated samples are generated with the expected pileup conditions in the respective data acquisition period, the distribution of the number of vertex candidates differs between data and simulation. Therefore, all simulated events are re-weighted according to [158]. The number of pileup interactions in data is estimated per lumi-block[1] by the total inelastic pp cross section of 69.4 mb [159] and the integrated luminosity in the respective lumi-block. This number is compared to the known number of pileup interactions in simulation[2]. The correction factors are designed in a way that only the shape, not the scale of the simulated distribution, is changed.

6.6.2 Trigger and light lepton identification and isolation efficiencies

Re-weighting factors that account for different efficiencies of the trigger and the muon and electron identification as well as isolation criteria in data and in simulation are considered in the respective channel of this analysis. The correction factors are determined with the *Tag-and-Probe* method [160] applied on a sample enriched in $Z(\ell\ell)$+jets events. The correction factors needed for the muon channel are provided by a group within the CMS experiment responsible for muon triggering, reconstruction and identification [149]. The correction factors for the electron identification and isolation have been determined together with the correction factors for the single-electron trigger efficiencies in [128] with the Tag-and-Probe method [129].

The derived single-muon trigger correction factors range from approximately 0.95 to 1.002 with uncertainties at the per-mill level or below [149]. Although the plateau of the trigger efficiency is only reached for muon candidates with $p_T > 40\,\text{GeV}$, muon candidates with lower transverse momenta ($p_T > 25\,\text{GeV}$) can be used since the trigger turn-on curve is well described by simulation. For the tight muon candidate identification requirements the corrections are usually smaller than 1%, and range up to 2% for the applied isolation criteria [149].

The combined electron trigger, identification and isolation correction factors depend on the p_T and η of the electron candidate and range from a few per cent up to more than 50% at high $|\eta|$ [128]. Also in this case, the trigger turn-on curve is described by simulation after the application of the derived correction factors. Hence, electron candidates with $p_T > 35\,\text{GeV}$ can be included.

6.6.3 Misidentification rates of light leptons as tau lepton candidates

As described in Section 5.6.2, discriminants are applied to reduce the misidentification rate of electrons and muons as tau lepton candidates. The efficiency of the discriminant against muons in data is well modeled in simulation. Therefore, no correction is needed [142]. Contrary, disagreements between data and simulation have been observed after the application of the criteria that are ought to reduce the misidentification rate of electrons as tau lepton candidates [142]. Thus, a factor of 0.85 is applied on top of

---

[1] A lumi-block summarizes events that were recorded with very similar detector conditions. Several lumi-blocks make up a run.

[2] Simulated events are assigned event numbers, and therefore also to a certain lumi-block, randomly.

the event weight for all tau lepton candidates with $|\eta| \leq 2.1$ that can be matched to a generator-level electron. For reconstructed tau lepton candidates with $|\eta| > 2.1$ this correction factor is 0.65 [142]. The matching of a reconstructed tau lepton candidate to a generator-level electron is performed by checking if a generator-level electron exists within $\Delta R < 0.5$ with respect to the tau lepton candidate.

6.6.4 Re-weighting of generator properties in $W+$ jets and $t\bar{t}+$ jets events

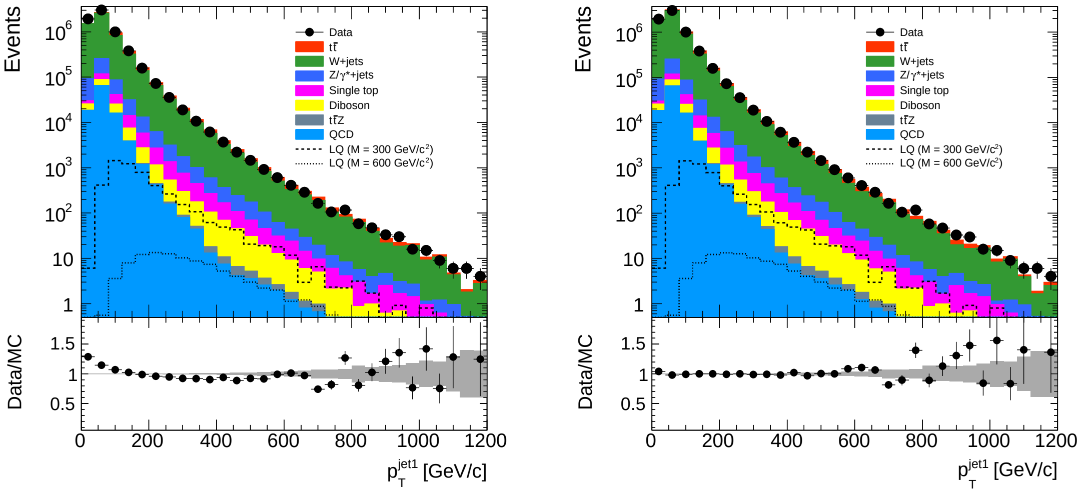
The leading jet p_T spectrum is found to be poorly modeled by simulation after the pre-selection requirements have been applied. The observed discrepancy is not only transferred to the description of the S_T distribution, but also to the modeling of the tau lepton candidate p_T spectrum, since a large portion of tau lepton candidates in the background is made up of misidentified jets. This mismodeling has to be corrected in simulation. Otherwise a bias is introduced through kinematic requirements on these observables to the analysis as well as to the measurement of the tau lepton misidentification rate, described later in this section. The observed discrepancy can be traced back to arise from MadGraph simulated W+jets events. This phenomenon has also been observed by other analyses, e.g. an analysis measuring the differential cross section of W+jets events [161,162]. Thus, re-weighting factors for simulated events are derived for this analysis in a sideband region dominated by $W(\rightarrow \mu\nu_\mu)$+jets events with a negligible contribution from signal events. All studied objects have to pass the object criteria detailed in Section 6.4. All event filters, also mentioned in Section 6.4, PU re-weighting as well as the correction factors that account for different muon reconstruction and identification efficiencies in data and simulation are applied. The sideband region is defined by the following selection requirements:

- Exactly one muon candidate with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$ passing the tight identification criteria is present in all considered events.
- Each event is required to have at least one jet with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$.
- $\not{E}_T > 40\,\text{GeV}$ has to be fulfilled by all considered events.

Figure 6.3(a) shows the distribution of the leading jet transverse momentum after having the aforementioned selection criteria applied. The discrepancy between data and simulation is most prominent for low transverse momenta, where the excess of data compared to simulation amounts to almost 30%. For higher transverse momenta the discrepancy is decreasing, until, for jets with $p_T > 200\,\text{GeV}$, simulation starts to slightly overpredict the amount of observed data events.

In order to determine the re-weighting factors for the leading jet transverse momentum in the W+jets simulation the small contribution from non-W+jets events is determined in simulation and subtracted from data accordingly. Afterwards, the ratio of observed events in data and expected events in simulation, shown in Fig. 6.4, is fitted as a function of the leading jet transverse momentum with an exponential function $w(p_T)$ defined by

$$w(p_T) = \exp(p_0 + p_1 \cdot p_T) + p_2, \tag{6.1}$$



(a) (b)

Figure 6.3: The distribution of the leading jet p_T in the $W+$ jets sideband region defined by the selection described in Section 6.6.4 before (a) and after (b) the correction factors defined by Eq. (6.1) are applied on the W+jets simulation.

with $p_0 = -0.690$, $p_1 = -0.009$ and $p_2 = 0.879$. According to this function a weight determined for the transverse momentum of the leading jet on generator level is applied to each simulated W+jets event.
The effect of applying the correction factors, which range up to approximately 20% for low-p_T jets, is shown in Fig. 6.3(b). Re-weighting the W+jets events according to the transverse momentum of the leading generator-level jet significantly improves the description of data by simulation, leading to good agreement.

Analyses measuring the normalized differential cross section of top quark pair production in data recorded with the CMS experiment at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV [163, 164] have observed that the $t\bar{t}$ + jets simulation predicts a p_T spectrum of the top quark candidates that is too hard compared to the measured spectrum in data. Therefore, re-weighting factors as a function of the generator-level p_T of the top and the antitop quark are derived in order to compensate for the shape difference between simulation and data [165]. These re-weighting factors yield corrections up to 30% and are applied in this analysis.

6.6.5 Misidentification rates of jets as tau lepton candidates

A large portion of selected background events does not contain prompt tau leptons. The events are selected because a jet is falsely reconstructed as a tau lepton candidate. These make up roughly 50% of all events after the final signal selection. The main contributions to these reducible background processes arise from $t\bar{t}$ + jets production ($tt \to WbWb \to qqbb\ \mu\nu_\mu/e\nu_e$) and W + jets events ($W \to \mu\nu_\mu$ or $W \to e\nu_e$), where the lepton originates from a W boson decay and the tau lepton candidate from a misreconstructed jet. Single top quark, diboson and QCD multijet production can also contribute

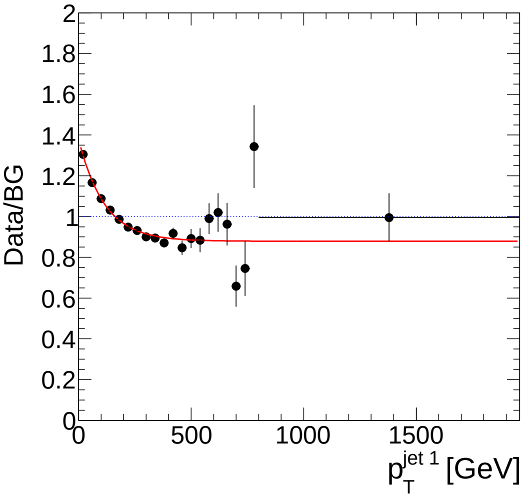


Figure 6.4: Ratio of the number of events in data and simulation as a function of the transverse momentum of the leading jet in the W+jets sideband region defined by the selection described in Section 6.6.4. The ratio is fitted with an exponential function defined by Eq. (6.1) (*red line*).

to the reducible background, but less frequently.

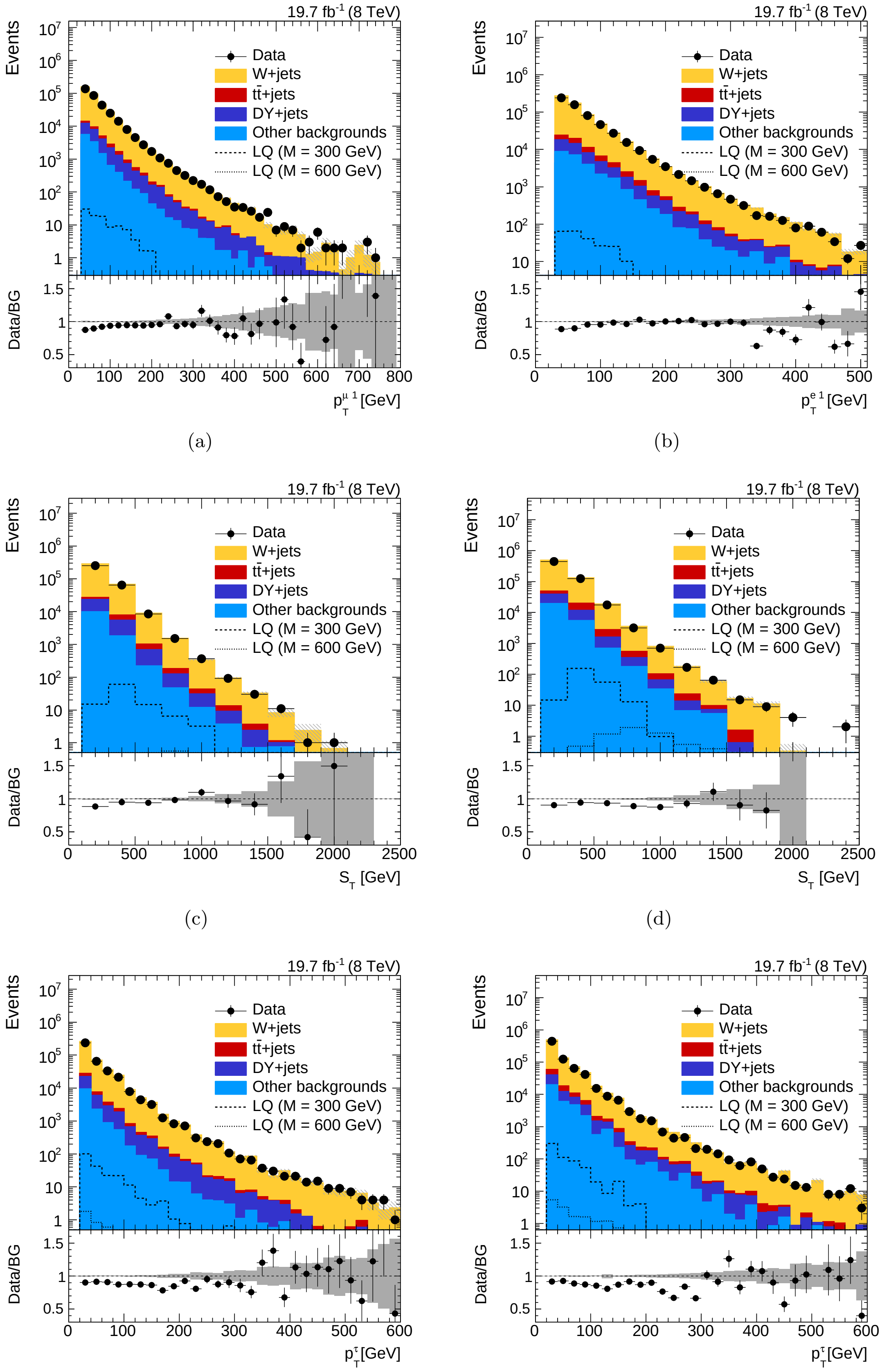
Since the p_T distribution of the leading tau lepton candidate is used for the statistical interpretation of this search, it is crucial for the performed analysis to understand the misidentification rate of jets as tau lepton candidates and the kinematics of these misidentified tau lepton candidates. Additionally, it has to be ensured that this rate is well modelled by simulated events.

The misidentification rate is separately measured in the four regions of the p_T distribution of the tau lepton candidates in data and simulation in both channels. Re-weighting factors that compensate for differences in the misidentification rate in data and simulation are derived. For the measurement, a sideband enriched in W+jets events is selected. By requiring exactly one light lepton candidate and $\not{E}_T > 40\,\text{GeV}$ it is ensured that the W boson candidate in these events decays leptonically via one of the decay chains $W \to \mu\nu_\mu$, $W \to e\nu_e$ or $W \to \tau\nu_\tau \to e/\mu\ \nu_\tau\nu_{e/\mu}$, such that all hadronically decaying tau lepton candidates reconstructed in these events are misidentified jets. The presence of at least one tau lepton candidate passing the decay mode finding criterion is required for each selected event, such that the misidentification rate of the medium identification working point can be determined with respect to this requirement. The misidentification rate of the decay mode finding criterion is determined separately. In order to suppress $t\bar{t}$ + jets events, a veto against b-tagged jets using the loose working point of the CSV algorithm (see Section 5.5.5) is applied. The contribution from DY+jets events decaying via $Z \to \mu\mu$ or $Z \to ee$, where a light lepton is misidentified as an hadronically decaying tau lepton candidate, is suppressed by discarding events where the invariant mass of the light lepton candidate and the hadronically decaying tau lepton candidate is close to the Z boson mass.

All objects considered for the selection of the sideband region have to pass the object quality criteria described in Section 6.4. The event filters mentioned in Section 6.4 and all corrections for simulation described in this section are applied. To avoid double counting of objects, all studied particle candidates have to pass the distance requirements described in Section 6.5. The selection of the sideband region is summarized in the following.

- In each event exactly one electron candidate fulfilling the identification criteria described in Section 5.4 with $p_T > 35\,\text{GeV}$ and $|\eta| < 2.1$ or exactly one muon candidate with $p_T > 35\,\text{GeV}$ and $|\eta| < 2.1$ passing the tight identification requirements is found. If an electron candidate passing the aforementioned selection criteria is present, a veto against muon candidates passing the mentioned muon candidate selection criteria is applied in order to avoid overlap in the events of the two channels. A higher cut on the muon candidate transverse momentum than in the signal region is chosen, such that the kinematics are the same in the electron and the muon channel. As in the signal region, a p_T requirement of $p_T > 35\,\text{GeV}$ is needed in the electron channel due to the high-p_T threshold of the trigger used in the electron channel.
- The measured $\not{E}_T$ has to be higher than 40 GeV.
- Each event has to have at least one jet with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$.
- To veto events that contain a b jet candidate, the event is discarded if a jet passing the loose working point of the CSV algorithm is found.
- At least one tau lepton candidate with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$ passing the decay mode finding criterion has to be present.
- If the invariant mass of the selected tau lepton candidate and the light lepton candidate $M(\tau\ell)$ fulfills $80\,\text{GeV} < M(\tau\ell) < 100\,\text{GeV}$, the event is discarded.

With the given selection a sideband that is dominated by W+jets events is defined. Remaining background contributions mainly arise from DY+jets processes and $t\bar{t}+$jets events. Also a few single top quark, diboson and multijet events are present. This is illustrated in Fig. 6.5, where the p_T distributions of the leading light lepton candidate, S_T and the p_T of all tau lepton candidates are shown for the muon channel and for the electron channel. The signal contribution in this sideband is very small compared to the large amount of expected events from Standard Model processes. Also the contribution from events that contain real tau leptons is negligible, and is estimated to be approximately 1%. The number of real tau leptons is thereby derived by performing a ΔR-matching of the selected tau lepton candidates to tau leptons on generator level. If a generator-level tau lepton is found within $\Delta R < 0.5$ to a reconstructed tau lepton candidate, this candidate is considered to be a real tau lepton. The sideband has been designed in a way that the kinematics are very similar to the signal region, justifying the assumption that the measured misidentification rates in the sideband region are also valid for the signal region. Despite the fact that there is no explicit requirement that makes the sideband region statistically independent from the signal region, the event overlap between the two regions is smaller than 0.5% and can be neglected.



(a) (b)

(c) (d)

(e) (f)

Figure 6.5: Comparison of data (*black dots*) and simulation (*colored areas*) in the muon channel (*left*) and in the electron channel (*right*) after the $W+$jets sideband selection used to derive the misidentification rate of jets as tau lepton candidates. Shown are the distributions of (a) and (b) the p_T spectra of the leading light lepton candidates, (c) and (d) S_T, (e) and (f) the p_T spectra of the tau lepton candidates passing the decay-mode-finding criterion.

In all distributions shown in Fig. 6.5 good agreement between data and simulated events is observed. The normalization discrepancies visible in all distributions can be traced back to the application of the b tag veto since correction factors that take into account different b tagging efficiencies in data and simulation are not applied. The application of these factors is not needed, since these cancel out in the determination of the tau lepton misidentification rate.

The misidentification rate of the medium working point can then be determined by counting the number of tau lepton candidates fulfilling the medium working point, denoted as N_{med} in the following, with respect to the total number of tau lepton candidates passing the decay mode finding criterion, N_{dmf}. It should be noted that this misidentification rate does not include the efficiency of the decay mode finding criterion. In contrast to the only 1% contribution of events that contain real tau leptons when the medium working point is not applied, this number is approximately 10% after application of the medium working point, meaning that this background cannot be neglected anymore. All tau lepton candidates that can be matched to a generator-level tau lepton are removed from the list of candidates in simulation. Since this procedure cannot be performed in data, the number derived in simulation is also subtracted from the number of reconstructed tau lepton candidates in data. This ansatz is justified by the good description of the tau lepton identification efficiency in data by simulation observed in other studies [141].

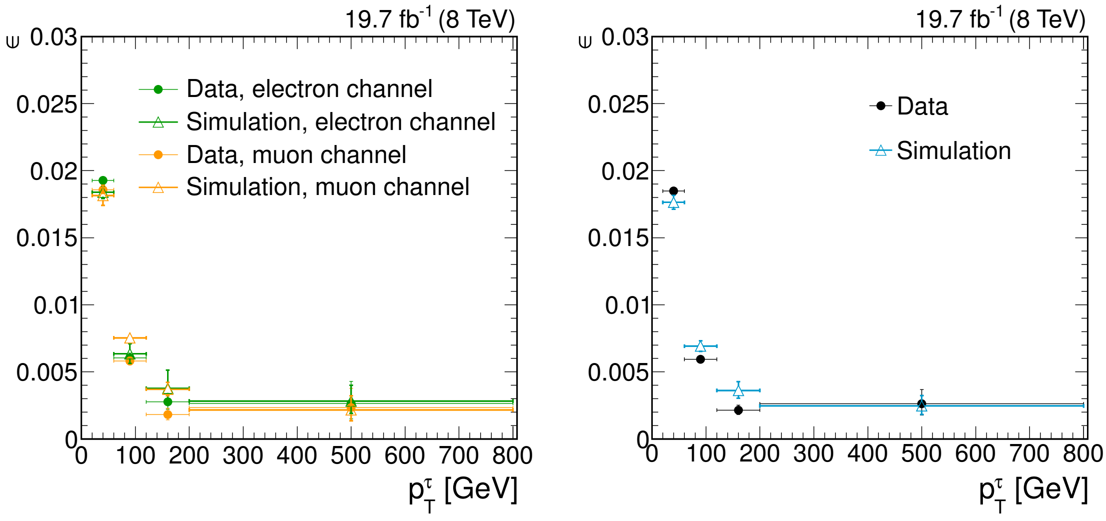
The same approach as for tau lepton candidates matchable to real tau leptons is applied for tau lepton candidates that can be matched to an electron or a muon candidate on generator level. These misidentification rates are treated separately, as described before.

The number of tau lepton candidates matched to a generator-level tau lepton, electron, or muon and passing the medium identification working point is referred to as $N_{\mathrm{med}}^{\mathrm{real,MC}}$ or $N_{\mathrm{med}}^{\mathrm{e}/\mu,\mathrm{MC}}$, respectively. The parameter $N_{\mathrm{dmf}}^{\mathrm{real,MC}}$ and $N_{\mathrm{dmf}}^{\mathrm{e}/\mu,\mathrm{MC}}$ denote the number of reconstructed tau lepton candidates that fulfill the decay mode finding criterion and can be matched to a generator-level tau lepton, or electron or muon, respectively. With this notation the misidentification rate ϵ in the four regions of the p_T distribution of the tau lepton candidate in data and simulation is given by

$$\epsilon(p_T) = \frac{N_{\mathrm{med}}(p_T) - N_{\mathrm{med}}^{\mathrm{real,MC}}(p_T) - N_{\mathrm{med}}^{\mathrm{e}/\mu,\mathrm{MC}}(p_T)}{N_{\mathrm{dmf}}(p_T) - N_{\mathrm{dmf}}^{\mathrm{real,MC}}(p_T) - N_{\mathrm{dmf}}^{\mathrm{e}/\mu,\mathrm{MC}}(p_T)}. \tag{6.2}$$

A comparison of the misidentification rate in the electron channel and in the muon channel in the four p_T regions of the distribution for the tau lepton candidates derived with Eq. (6.2) is presented in Fig. 6.6(a). The measured misidentification rates are of the order of a few per cent and decrease as a function of p_T of the tau lepton candidates. The measured rates in the electron channel are compatible, within the statistical uncertainties, to the ones derived in the muon channel. They can be combined into a single misidentification rate, shown in Fig. 6.6(b), yielding smaller statistical uncertainties.

The misidentification rates in simulation differ from the ones observed in data. Thus, correction factors for simulation in the four p_T regions of the tau lepton candidates are derived. These correction factors, defined as the ratio of the misidentification rate in data and in simulation, are shown in Fig. 6.7. Figure 6.7(a) compares the correction



(a) (b)

Figure 6.6: Misidentification rates of a jet as a tau lepton candidate as a function of p_T of the tau lepton candidate. The error bars show the statistical uncertainties only. (a) Comparison of the misidentification rates in the electron channel (*green*) and in the muon channel (*orange*) in data (*solid circles*) and in simulation (*open triangles*). (b) Combined results of the muon and the electron channel in data (*black dots*) and in simulation (*light blue triangles*).

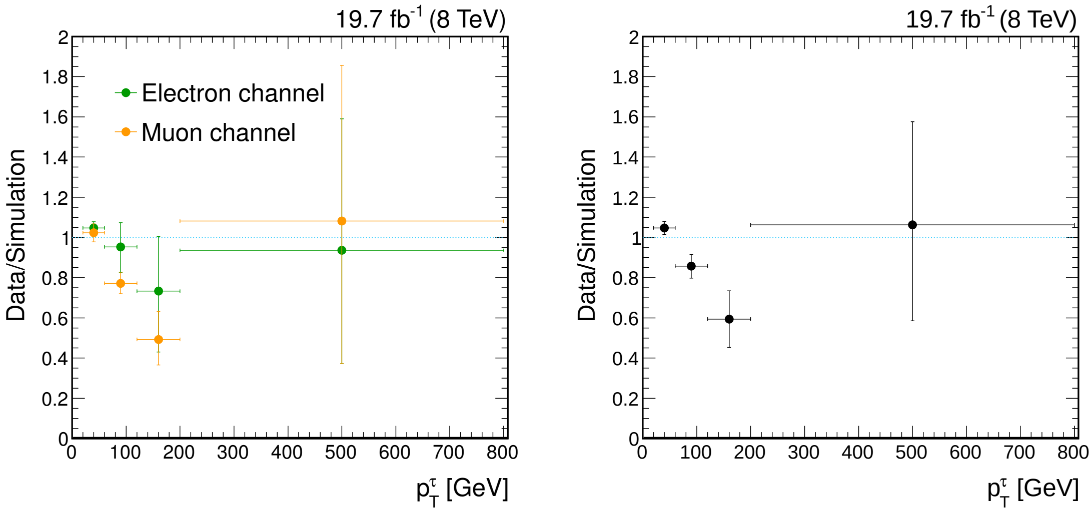
factors obtained from the muon channel to the ones from the electron channel. The ratios agree within their statistical uncertainties. The correction factors derived from the combined misidentification rate in the two channels are shown in Fig. 6.7(b). The combined corrections and their statistical uncertainties are

$$\begin{aligned}
s &= 1.05 \pm 0.03 \text{ (stat) for } 20\,\text{GeV} \leq p_T^\tau < 60\,\text{GeV}, \\
s &= 0.86 \pm 0.06 \text{ (stat) for } 60\,\text{GeV} \leq p_T^\tau < 120\,\text{GeV}, \\
s &= 0.59 \pm 0.14 \text{ (stat) for } 120\,\text{GeV} \leq p_T^\tau < 200\,\text{GeV}, \\
s &= 1.06 \pm 0.50 \text{ (stat) for } p_T^\tau \geq 200\,\text{GeV}.
\end{aligned} \tag{6.3}$$

These p_T dependent correction factors are applied in simulation as an event weight for each tau lepton candidate that cannot be matched to a generator-level tau lepton, electron or muon in order to compensate for the differences of the misidentification rates in data and in simulation. If more than one of those tau lepton candidates is found in the event, the corresponding correction factors are multiplied.

Additionally, the misidentification rate of the decay mode finding-criterion with respect to all jets is studied, analogously to the determination of the misidentification rate of the medium working point. The corresponding correction factor for simulated events is 0.92, independent of the transverse momentum of the tau lepton candidates. This factor is also applied as an event weight for each tau lepton candidate that cannot be matched to a generator-level tau lepton, electron or muon.

A systematic uncertainty is assigned to the misidentification rate measured in data due to the subtraction of tau lepton candidates matched to generator-level electrons, muons

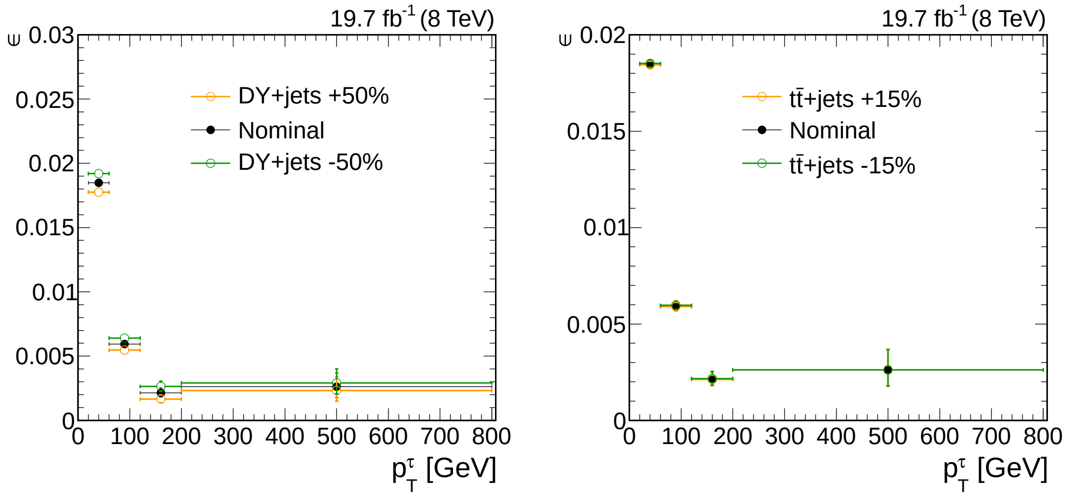


(a) (b)

Figure 6.7: Correction factors for the tau lepton misidentification rate as a function of p_T of the tau lepton candidate. The error bars represent statistical uncertainties only. (a) Comparison of the correction factors in the muon channel (*orange*) and in the electron channel (*green*). (b) Combined correction factors of the muon and the electron channel.

or tau leptons since these numbers are determined in simulation and can therefore deviate from the actual numbers in data. The main contributions arise from DY + jets and $t\bar{t}$ + jets events still present in the W+jets sideband. In order to determine the uncertainty due to the subtraction of the events, the cross section of DY + jets production is varied by ±50% and the $t\bar{t}$ cross section by 15% [166]. For each variation the misidentification rates in data are calculated again using Eq. (6.2). The results are shown in Fig. 6.8. The variation of the $t\bar{t}$ + jets production cross section has a negligible effect on the derived misidentification rate, while the change of the DY + jets cross section leads to variations of the misidentification rates of up to 24%. In order to determine the uncertainties, the two up variations and down variations are added quadratically, respectively, and the resulting systematic uncertainty in the misidentification rate in data is propagated to the determination of the correction factors. The resulting uncertainties in the correction factors range between 4% and 24%.

To check if p_T dependent correction factors are sufficient for the removal of the discrepancies between the misidentification rates in data and simulation, the misidentification rates and the corresponding correction factors are additionally calculated as a function of η of the tau lepton candidates, S_T, the number of jets with $p_T > 30\,\text{GeV}$ and in dependence of the distance between the tau lepton candidates and the respective closest jet in ΔR. Prior to the determination of these rates, the p_T dependent correction factors (see Eq. (6.3)) are applied. The resulting correction factors for the combination of the electron channel and the muon channel are presented in Figs. 6.9(a) to 6.9(d). Only in η a dependence of the correction factors is observed at high $|\eta|$. All other correction factors are flat and compatible with unity. In principle, p_T and η dependent correction factors are needed to diminish the difference in the misidentification rates in data and



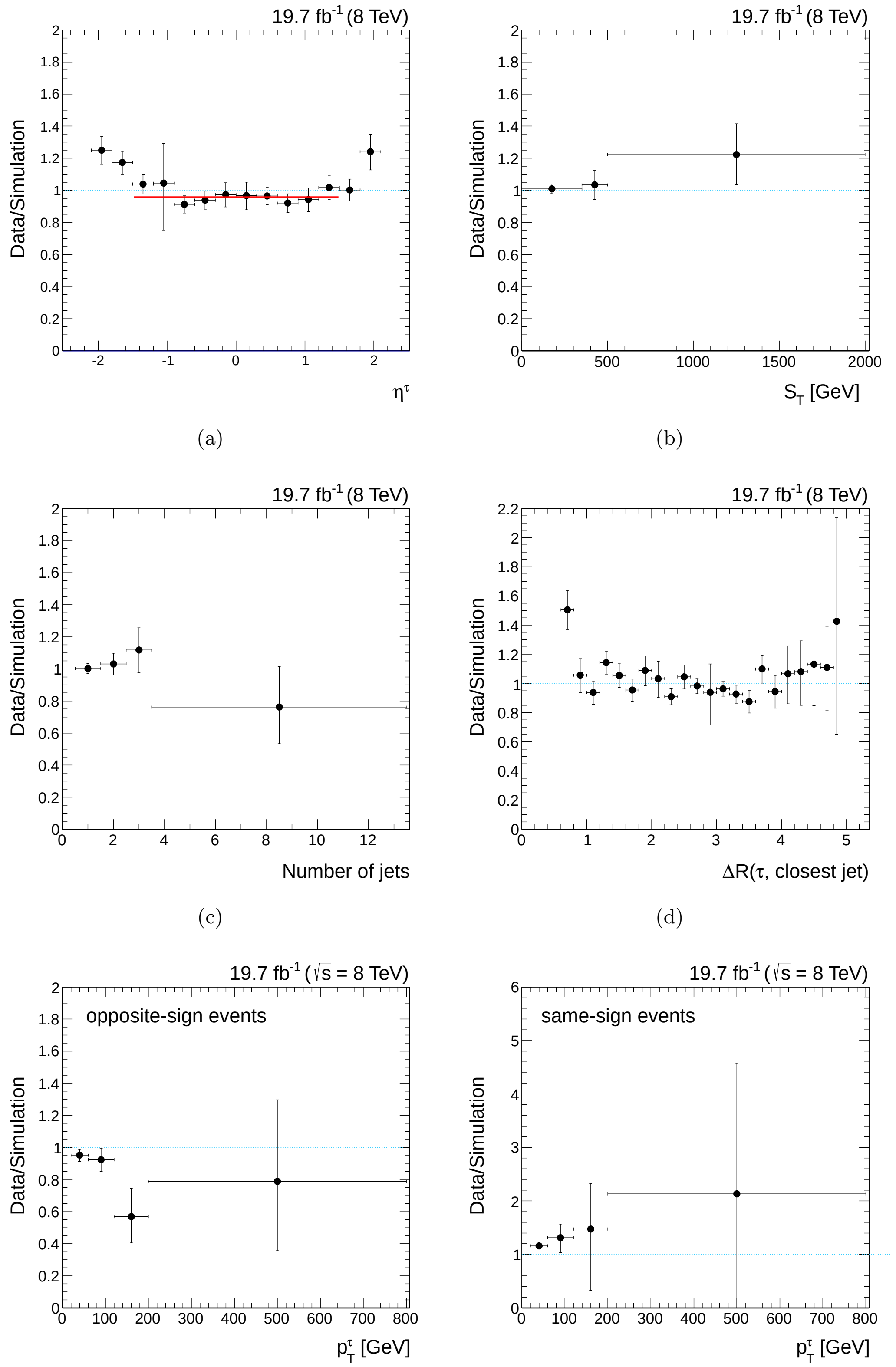
(a) (b)

Figure 6.8: Combined electron and muon channel misidentification rates in data as a function of p_T of the tau lepton candidate, with variations of the cross section of (a) DY + jets events by 50% and (b) $t\bar{t}$ + jets events by 15%. The up-variations are shown in orange, the down-variations in green.

simulation. Due to the limited statistical precision available for this measurement, this approach is not feasible. Instead, an additional systematic uncertainty is assigned to the correction factors for tau lepton candidates with $|\eta| > 1.5$ in order to take this dependence into account. The uncertainty is chosen according to the observed difference between data and simulation for tau lepton candidates with $|\eta| > 1.5$ and amounts to 20%. In order to check if the correction factors are flat for more central tau lepton candidates, a fit of the correction factors with a constant has been performed in this region, represented by the red line in Fig. 6.9(a). The fit yields $\chi^2/n_{\mathrm{dof}} \approx 0.4$ and justifies that no uncertainty is applied for tau lepton candidates with $|\eta| \leq 1.5$.
In Figs. 6.9(e) and 6.9(f) the correction factors for the misidentification rate are illustrated in the four regions of the p_T distribution of the tau lepton candidates for same-sign events and opposite-sign events. An event is classified as a same-sign event if the light lepton candidate and the studied tau lepton candidate have same-sign charges. If not, the event is assigned to the opposite-sign sample. In cases where more than one tau lepton candidate is reconstructed all possible combinations of the tau lepton candidates with the light lepton candidate are considered and each of these cases is assigned to the respective category of events. A large difference between the two channels in the data over simulation ratios is observed. The differences can be traced back to a dependence of the misidentification rate on the flavor of the jet that is misidentified as the tau lepton candidate and to the different flavor compositions in the opposite-sign and same-sign samples. This will be explained in detail in the following.

Jet flavor dependence

In order to understand this effect, the origins of the jets that are misidentified as hadronically decaying tau lepton candidates are studied. In general ten different possibilities,



(a) (b)

(c) (d)

(e) (f)

Figure 6.9: Correction factors for simulation as a function of (a) η of the tau lepton candidates, (b) S_T, (c) number of jets, (d) $\Delta R(\tau$, closest jet), (e) p_T of the tau lepton candidates in events with opposite-sign $\ell\tau_h$-pairs and (f) p_T of the tau lepton candidates in events with same-sign $\ell\tau_h$-pairs. Shown are the combined results of the muon and the electron channel. The red line in (a) shows the result of a fit to the ratio with a constant for tau lepton candidates with $|\eta| < 1.5$

depending on whether an up-type quark $q^{\pm\frac{2}{3}}$, a down-type quark $q^{\mp\frac{1}{3}}$ or a gluon g initiates the jet that is reconstructed as a tau lepton candidate, with electric charge $+1$ or -1 can be identified. The possibilities are

$$\begin{aligned}
&g \to \tau_h^-,\ g \to \tau_h^+,\\
&q^{+\frac{2}{3}} \to \tau_h^+,\ q^{-\frac{2}{3}} \to \tau_h^-,\\
&q^{+\frac{1}{3}} \to \tau_h^+,\ q^{-\frac{1}{3}} \to \tau_h^-,\\
&q^{+\frac{2}{3}} \to \tau_h^-,\ q^{-\frac{2}{3}} \to \tau_h^+,\\
&q^{+\frac{1}{3}} \to \tau_h^-,\ q^{-\frac{1}{3}} \to \tau_h^+.
\end{aligned}$$

The first two cases describe a gluon being reconstructed as a tau lepton candidate with positive or negative charge. The second and the third line refer to up-type quarks and down-type quarks, respectively, that are misidentified as a tau lepton candidate with the same charge sign as the initial quark. The last two lines specify cases in that the initial quark is reconstructed as a tau lepton candidate with opposite sign charge. These are referred to as *charge-flip categories* in the following.

The misidentification rate for each of these five different classes is studied inclusively in p_T in simulated W+jets events with the selection described above. Additionally, the sample is categorized in $\ell^+\tau_h^-$, $\ell^+\tau_h^+$, $\ell^-\tau_h^+$ and $\ell^-\tau_h^-$ events. In each of the resulting 20 categories the misidentification rate is determined according to Eq. (6.2). The result is shown in Fig. 6.10 for the muon channel[3]. The misidentification rate of jets originating from gluons is clearly smaller than the average rate from quark initiated jets. This dependence has also been observed by other analyses (e.g. by [141]). Gluon jets are generally wider and contain more particles than quark jets [141, 167]. Therefore, it is more unlikely that they are misidentified as an hadronic tau lepton decay, which usually decays in a much narrower cone. The misidentification rate additionally depends on the quark flavor of the original jet, and whether the tau lepton candidate gets assigned the same sign of the charge as the initial parton or not. The misidentification rates within the categories $\ell^+\tau_h^-$, $\ell^+\tau_h^+$, $\ell^-\tau_h^+$ and $\ell^-\tau_h^-$ are compatible, meaning that the rates do not depend on the charge combination of the light lepton candidate and the tau lepton candidate. The difference observed in the correction factors for the misidentification rates in same-sign events and opposite-sign events (see Figs. 6.9(e) and 6.9(f)) is therefore explainable with different compositions of the flavors of the partons initiating the jets that are reconstructed as tau lepton candidates in the two cases. The difference in the flavor composition can be explained by assuming that most of the studied events are $W+$ 1 jet events. Since in pp collisions the dominant Feynman diagrams for $W+$ 1 jet production are the ones with a gluon and a valence quark in the initial state, a W^+, which decays subsequently in a ℓ^+, is produced with a high probability in association with a d quark. A W^-, and thus a ℓ^-, is produced most likely together with an u quark. Since down-type quarks have a different probability to be reconstructed as a tau lepton

---

[3] Only distributions for the muon channel are presented hereafter. All conclusions drawn in this section are also valid for the electron channel.

candidate than up-type quarks and since in pp collisions W^+ events are more frequently produced than W^- events, different flavor compositions of opposite-sign and same-sign events are expected.

Despite the kinematic similarities, the flavor composition of the jets that are reconstructed as tau lepton candidates in the W+jets sideband differs from the composition in the signal region. The derived correction factors, depending on p_T of the tau lepton candidates, are therefore not directly applicable to the signal region. A parametrization of the misidentification rate in the flavor of the jet that is reconstructed as the tau lepton candidate in each of the four p_T regions is needed. Since the flavor of the jet cannot be determined in data, other measurable quantities have to be defined that depend on the flavor composition. One possibility is the categorization in $\ell^+\tau_h^-$, $\ell^+\tau_h^+$, $\ell^-\tau_h^+$ and $\ell^-\tau_h^-$ events in each p_T region. If we denote the misidentification rate of an up-type or down-type quark reconstructed as a tau lepton candidate with the same-sign charge as $\epsilon_{q,\frac{2}{3}}$ and $\epsilon_{q,\frac{1}{3}}$, respectively, the rate of a gluon misreconstructed as a tau lepton candidate as ϵ_g and the rate of the charge-flip categories as ϵ_{CH}, the parametrization for one p_T region is given by

$$\begin{aligned}
\epsilon^{\tau_h^+\ell^+} &= f_g^{\tau_h^+\ell^+} \cdot \epsilon_g + f_{q,\frac{2}{3}}^{\tau_h^+\ell^+} \cdot \epsilon_{q,\frac{2}{3}} + f_{q,\frac{1}{3}}^{\tau_h^+\ell^+} \cdot \epsilon_{q,\frac{1}{3}} + f_{\mathrm{CF}}^{\tau_h^+\ell^+} \cdot \epsilon_{\mathrm{CF}},\\
\epsilon^{\tau_h^+\ell^-} &= f_g^{\tau_h^+\ell^-} \cdot \epsilon_g + f_{q,\frac{2}{3}}^{\tau_h^+\ell^-} \cdot \epsilon_{q,\frac{2}{3}} + f_{q,\frac{1}{3}}^{\tau_h^+\ell^-} \cdot \epsilon_{q,\frac{1}{3}} + f_{\mathrm{CF}}^{\tau_h^+\ell^-} \cdot \epsilon_{\mathrm{CF}},\\
\epsilon^{\tau_h^-\ell^+} &= f_g^{\tau_h^-\ell^+} \cdot \epsilon_g + f_{q,\frac{2}{3}}^{\tau_h^-\ell^+} \cdot \epsilon_{q,\frac{2}{3}} + f_{q,\frac{1}{3}}^{\tau_h^-\ell^+} \cdot \epsilon_{q,\frac{1}{3}} + f_{\mathrm{CF}}^{\tau_h^-\ell^+} \cdot \epsilon_{\mathrm{CF}},\\
\epsilon^{\tau_h^-\ell^-} &= f_g^{\tau_h^-\ell^-} \cdot \epsilon_g + f_{q,\frac{2}{3}}^{\tau_h^-\ell^-} \cdot \epsilon_{q,\frac{2}{3}} + f_{q,\frac{1}{3}}^{\tau_h^-\ell^-} \cdot \epsilon_{q,\frac{1}{3}} + f_{\mathrm{CF}}^{\tau_h^-\ell^-} \cdot \epsilon_{\mathrm{CF}}.
\end{aligned} \tag{6.4}$$

The factors f_Y^X refer to the fraction of jets of the flavor category $Y (= g, (q, \frac{2}{3}), (q, \frac{1}{3}), \mathrm{CF})$ being reconstructed as a tau lepton candidate in the channel $X (= \tau_h^+\ell^+, \tau_h^+\ell^-, \tau_h^-\ell^+, \tau_h^-\ell^-)$. These numbers can be determined in simulation. For this approach, the misidentification rates in the two charge-flip categories are summarized to one misidentification rate. This is justified since the misidentification rates in these two categories, shown in Fig. 6.10, are statistically compatible. As illustrated for the W+jets sideband in the muon channel in Fig. 6.11, in all four p_T regions, the fraction f_Y^X, and thus the flavor compositions, are significantly different in the four different channels, meaning that this parametrization provides a very good sensitivity to the different flavor misidentification rates. Since the misidentification rates $\epsilon^{\tau_h\ell}$, on the left hand side of Eq. (6.4), can be measured in data, and the fraction f_Y^X can be taken from simulation, the different flavor misidentification rates ϵ_Y can in principle be determined for each of the four p_T regions by solving the set of equations given in Eq. (6.4). Practically this is not possible, since the limited statistical precision and numerical uncertainties lead to negative solutions for the rates. Also a likelihood fit of the four equations given in Eq. (6.4) is too unstable to yield reliable results. Moreover, the likelihood fit is not sensitive to all four channels at the same time due to limited statistical precision.

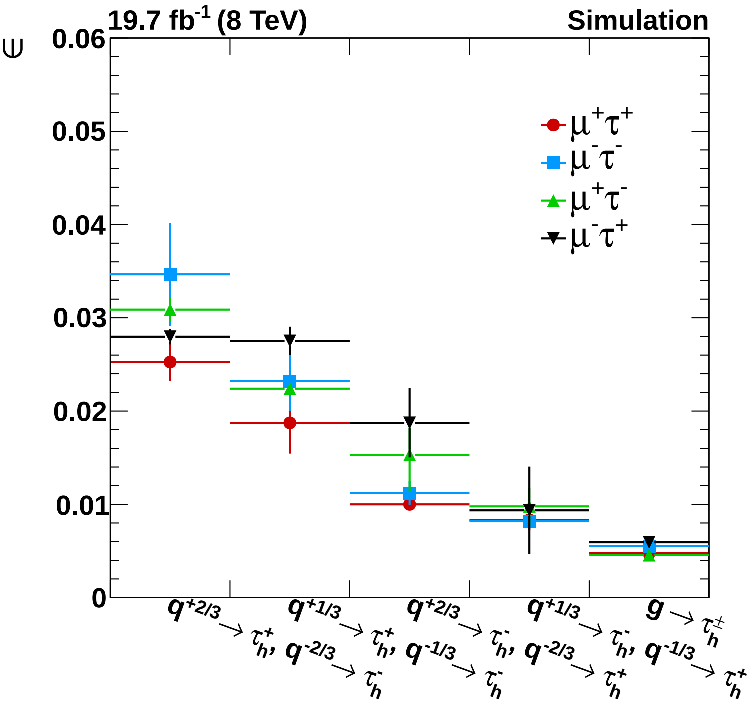
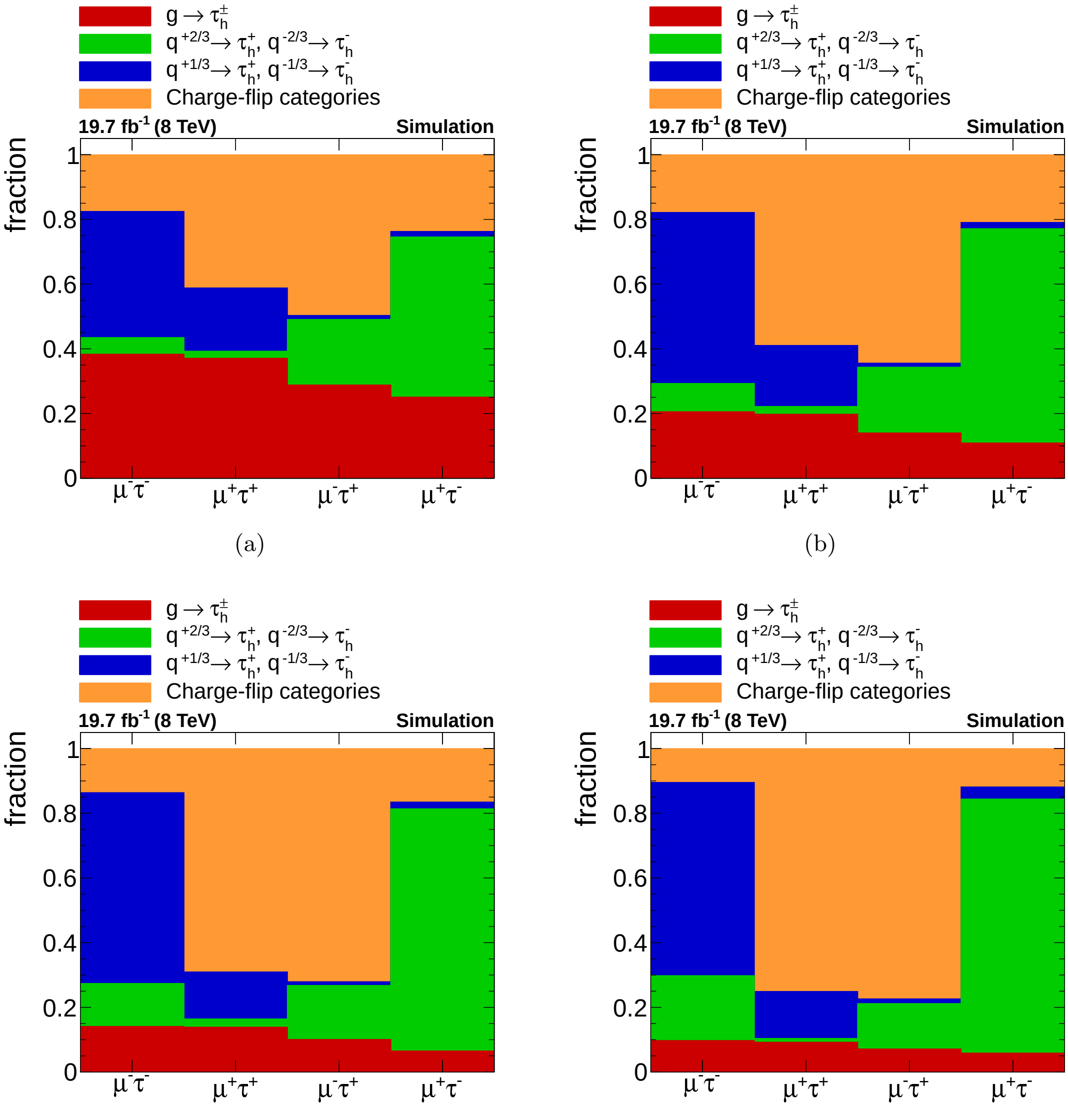


Figure 6.10: Misidentification rate ϵ of jets as tau lepton candidates as a function of the origin of the jet that is reconstructed as a tau lepton candidate and of the charge of the reconstructed tau lepton candidate. The misidentification rates are derived for the muon channel in simulated W+jets events after the application of the selection described in this section. The events are categorized in $\mu^+\tau_h^-$, $\mu^+\tau_h^+$, $\mu^-\tau_h^+$ and $\mu^-\tau_h^-$ events, but are studied inclusively in p_T.

Flavor uncertainties

Since a determination of the flavor misidentification rates is therefore impossible with the available data, an uncertainty is assigned to the flavor-independent correction factors given in Eq. (6.3) in order to take the dependence of the misidentification rate on the jet flavor into account. For the determination of this uncertainty the misidentification rates in data and simulation in the four channels $\ell^+\tau_h^-$, $\ell^+\tau_h^+$, $\ell^-\tau_h^+$ and $\ell^-\tau_h^-$ in each of the four p_T regions are derived. The inclusive correction factors (see Eq. (6.3)) are applied and the electron and muon channels are combined to obtain a higher statistical precision. The result is illustrated for all four p_T regions in Fig. 6.12. The remaining discrepancy observed between data and simulation in the rates studied in the four channels $\ell^+\tau_h^-$, $\ell^+\tau_h^+$, $\ell^-\tau_h^+$ or $\ell^-\tau_h^-$ is assumed to arise solely from different flavor misidentification rates and compositions in data and simulation. A likelihood fit is performed that adjusts one parameter ϵ_Y ($Y = g, (q, \frac{2}{3}), (q, \frac{1}{3}), \text{CF}$) in simulation individually such that the resulting misidentification rates ϵ^X ($X = \tau_h^+\ell^+, \tau_h^+\ell^-, \tau_h^-\ell^+, \tau_h^-\ell^-$) match the rates in data. The rates ϵ_Y are adjusted using the four equations in Eq. (6.4), where in each fit the other three parameters are set to their values in simulation. The uncertainty of ϵ_Y in the fit is a measure for the variation of ϵ_Y that is still possible, such that all ϵ^X in data remain compatible with the ϵ^X in simulation. The allowed shift in the misidentification rate



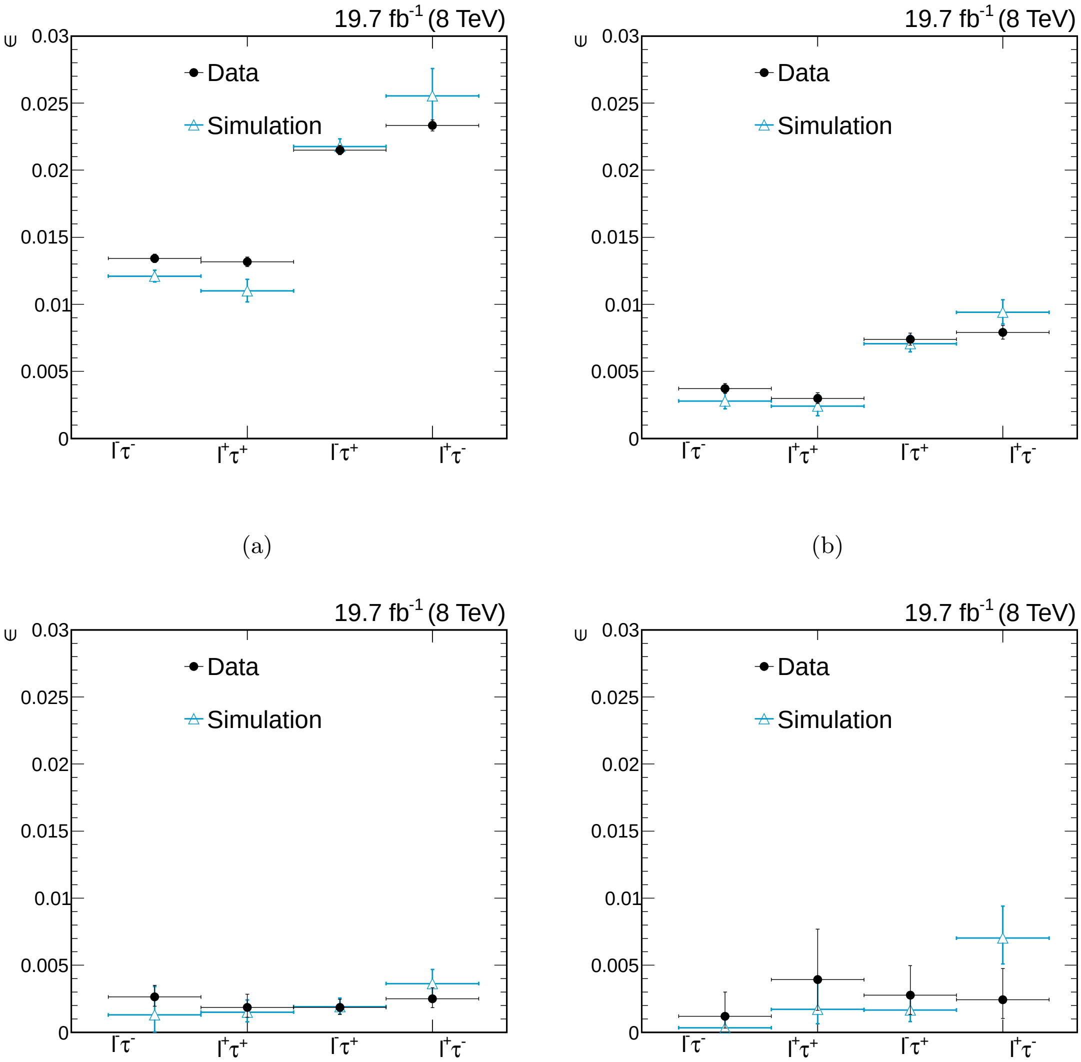
(a) (b)

(c) (d)

Figure 6.11: Composition of the flavor of the jets that are reconstructed as tau lepton candidates divided into events with $\mu^+\tau_h^-$, $\mu^+\tau_h^+$, $\mu^-\tau_h^+$ or $\mu^-\tau_h^-$-candidate pairs for tau lepton candidates with (a) $20\,\text{GeV} \leq p_T < 60\,\text{GeV}$, (b) $60\,\text{GeV} \leq p_T < 120\,\text{GeV}$, (c) $120\,\text{GeV} \leq p_T < 200\,\text{GeV}$ and (d) $p_T \geq 200\,\text{GeV}$. The flavor composition is derived in the W + jets sideband used to determine the tau lepton candidate misidentification rate of jets.

translates into a change of the correction factor. Thus, the uncertainty in each ϵ_Y yields an uncertainty in the correction factor due to the respective flavor as a function of p_T. In principle, this procedure has to be performed in each of the four considered p_T regions. However, since the fit gives stable results only in the first p_T region, the derived values from this p_T region are used for the final uncertainty.

For the transfer of these uncertainties from the sideband region to the signal region a comparison is done between the flavor composition of the jets reconstructed as tau lepton candidates in the signal region and in the W+jets sideband region. If these two regions had the same flavor composition, the application of this uncertainty would not



(a) (b)

(c) (d)

Figure 6.12: Comparison of the misidentification rates in data (*black dots*) and simulation (*light blue triangles*) in events with $\ell^+\tau_h^-$, $\ell^+\tau_h^+$, $\ell^-\tau_h^+$ or $\ell^-\tau_h^-$-candidate pairs for tau lepton candidates with (a) $20\,\text{GeV} \leq p_T < 60\,\text{GeV}$, (b) $60\,\text{GeV} \leq p_T < 120\,\text{GeV}$, (c) $120\,\text{GeV} \leq p_T < 200\,\text{GeV}$ and (d) $p_T \geq 200\,\text{GeV}$. The rates are derived in the $W+$jets sideband after the p_T-dependent corrections factors given in Eq. (6.3) have been applied. The combined results for the muon and the electron channel are shown.

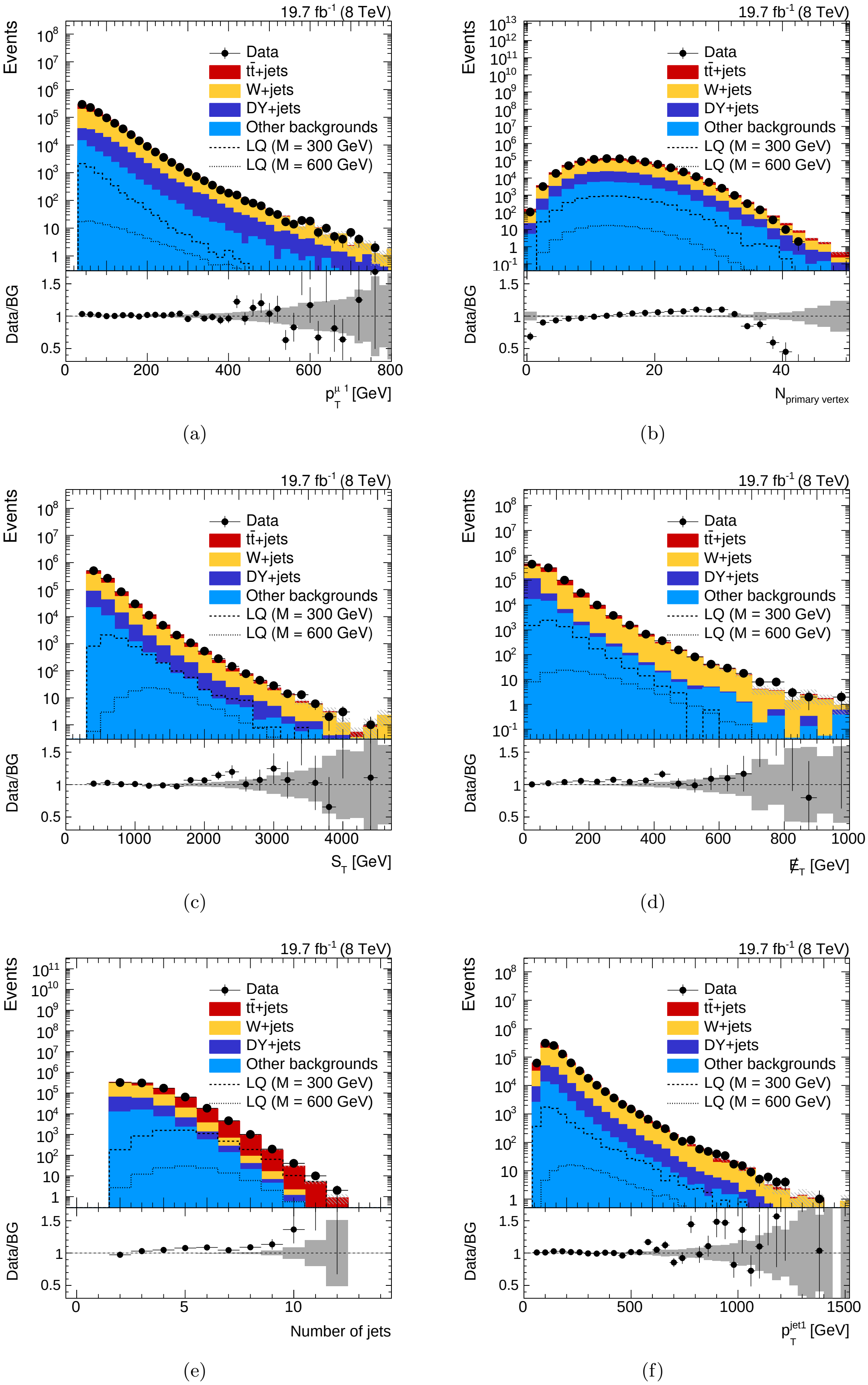
be necessary. This is not the case due to different background compositions in these two regions. Therefore, each of the four parts of the uncertainty is scaled according to the difference in the fraction of jets reconstructed as a tau lepton candidate of the respective flavor in the signal region and in the sideband region. The final systematic uncertainty in the correction factor of the misidentification rate then results from the scaled single statistical uncertainties of ϵ_Y determined in the fit by their addition in quadrature. The correction factors with all systematic uncertainties included are summarized in Section 6.8, where all sources of uncertainties relevant in this analysis are discussed.

6.7 Validation of the background estimation

In order to justify the background estimation by simulation with the corrections described in Section 6.6 applied, it has to be shown that the Standard Model processes measured in data are well described by the corrected simulation and that effects occurring in data are understood and can be estimated using simulation. For this purpose, two signal-depleted control regions are studied. The first control region is defined by the pre-selection, where the amount of predicted signal events is negligible compared to the huge amount of expected background events from Standard Model processes. This control region is mostly dominated by W+jets events and $t\bar{t}+$jets events, such that the modeling of these backgrounds can be studied in a phase space region with high statistical precision. This is helpful, since the two mentioned processes also constitute the main backgrounds for this search. The second control region is a region with stricter selection requirements applied and is thus closer to the signal region, but statistically orthogonal to both the signal region and the $W+$jets sideband used to derive the correction factors for the misidentification rate of jets as tau lepton candidates.

Control distributions for the muon channel after the pre-selection are depicted in Fig. 6.13. For the electron channel, control distributions are presented in Fig. 6.14. All corrections mentioned in the previous section have been applied to these distributions, except for the correction factors for the misidentification rate of jets as tau lepton candidates. These are not applicable here, since the pre-selection only includes a requirement of a tau lepton candidate passing the decay mode finding criterion, while the correction factors are designed to cure the discrepancy in the misidentification rate of the medium working point only. For both channels, the distributions of S_T, $\not{E}_T$, the number of jets and the p_T spectra of the leading jet are shown. For the muon channel, the distribution of the number of primary vertex candidates and the transverse momentum of the leading muon candidate are presented in addition. For the final signal selection tighter requirements are imposed on all these observables, except for the last two mentioned variables. Thus, a good description of these observables in simulation is essential, so that no bias is introduced through the imposed selection criteria.

In the muon channel in general a very good agreement between data and simulation is observed. Especially the variables that are further used in the final analysis selection show very good agreement. Large discrepancies are observed in the distribution of the number of primary vertex candidates shown in Fig. 6.13(b). However, these are ac-

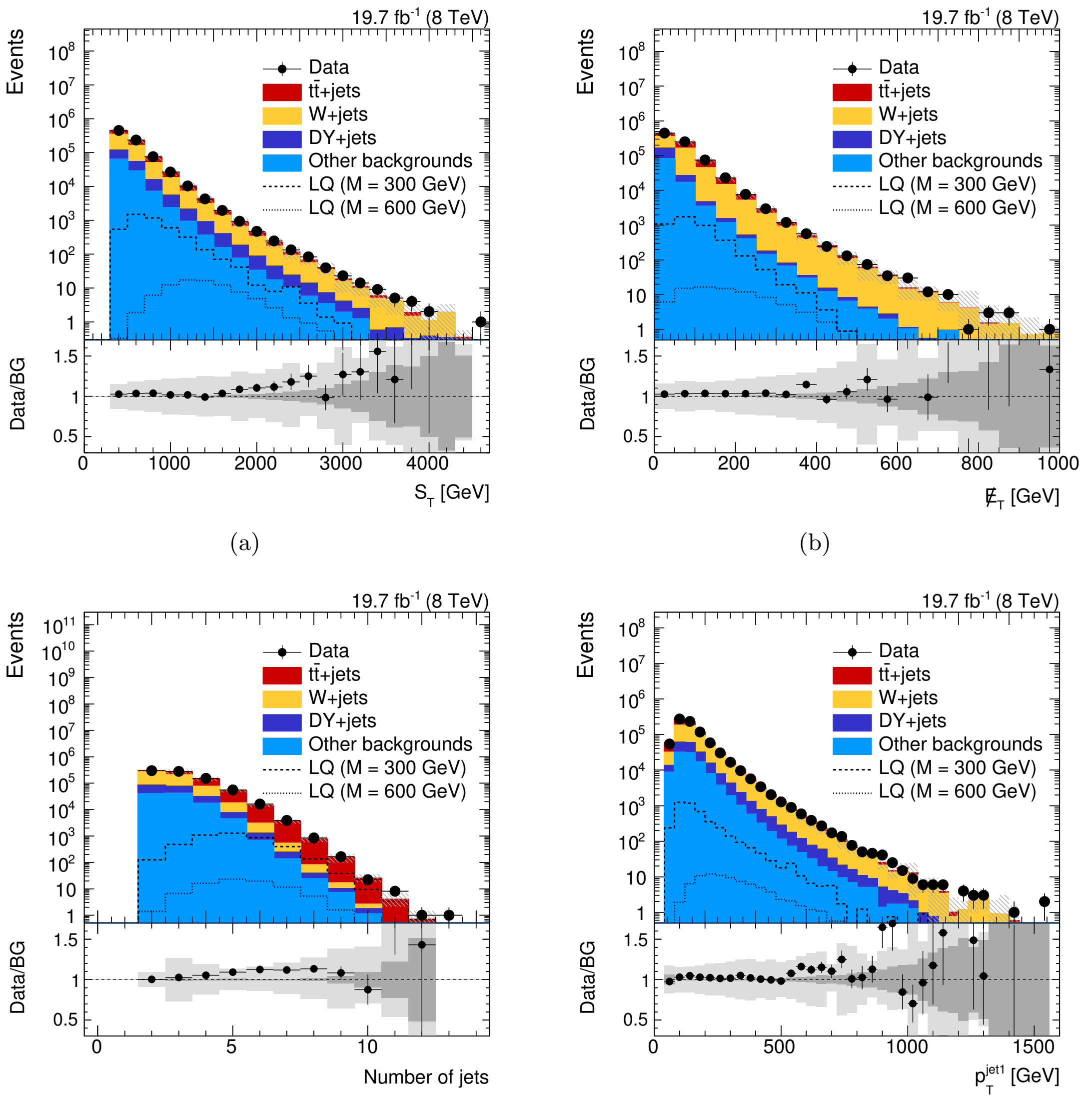


(a) (b)

(c) (d)

(e) (f)

Figure 6.13: Comparison of data (*black dots*) and simulation (*colored areas*) in the muon channel after the pre-selection for the distribution of (a) the p_T of the leading muon candidate, (b) the number or primary vertex candidates, (c) S_T, (d) $\not{E}_T$, (e) the number of jets and (f) the p_T of the leading jet. The dark gray areas in the ratio distributions illustrate the statistical uncertainty of simulated events only.



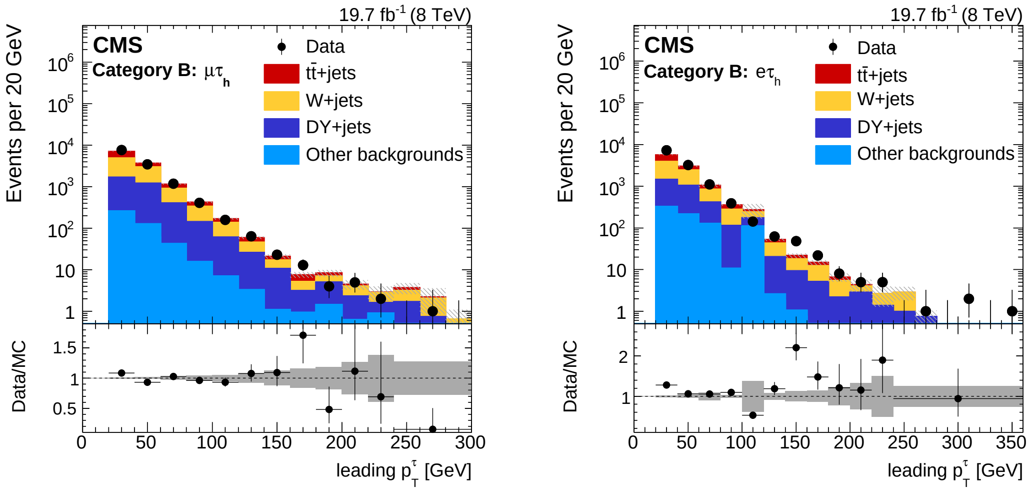
(a) (b)

(c) (d)

Figure 6.14: Comparison of data (*black dots*) and simulation (*colored areas*) in the electron channel after the pre-selection for the distribution of (a) S_T, (b) $\not{E}_T$, (c) the number of jets and (d) the p_T of the leading jet. The dark gray areas in the ratio distributions illustrate the statistical uncertainty of simulated events, the light gray bands the quadratically added systematic and statistical uncertainties.

counted for by the systematic uncertainty that is assigned due to the performed pileup re-weighting (see Section 6.8 for a detailed description of the systematic uncertainties). In the electron channel the agreement between data and simulation is slightly worse than in the muon channel. However, discrepancies observed in the electron channel are also well covered by the assigned uncertainties, leading to a satisfying description of data by simulated events in both channels.

The second control region is defined such that the p_T distribution of the leading tau lepton candidate is kept as similar as possible in terms of kinematics and background composition to the signal region. Since the presence of at least one medium tau lepton candidate is required, the correction factors for the tau lepton misidentification rates can



(a) (b)

Figure 6.15: Comparison of data (*black dots*) and simulation (*colored areas*) in the control region obtained by applying the signal selection with omitted S_T requirement and inverted $\not\!\!E_T$ requirement. Shown are the distributions of the p_T of the leading tau lepton candidate in (a) the muon channel and (b) the electron channel. The statistical uncertainties of simulated events are represented by the gray bands in the ratio distributions. The figures are published in [2].

be validated in this control region. Events are selected that pass the selection requirements as described in Section 6.5, except for the S_T requirement, which is completely omitted, and for the $\not\!\!E_T$ requirement, which is inverted. By inverting the $\not\!\!E_T$ requirement a control region is created that is statistically orthogonal to the signal region and that contains a negligible amount of signal events. Omitting the S_T requirement yields higher statistical precision in the control region. A stricter criterion is imposed on the electron MVA ID (MVA ID > 0.95) in order to suppress the QCD multijet background in the electron channel. QCD multijet events are present after the inversion of the $\not\!\!E_T$ requirement since jets are likely falsely identified as isolated electron candidates. An equivalent change in the muon channel is not necessary.

The p_T distribution of the leading tau lepton candidate is shown in Fig. 6.15(a) for the muon channel and in Fig. 6.15(b) for the electron channel. In the muon channel, a very good agreement is observed between data and simulation. In the electron channel an excess of data events over the Standard Model prediction is present for electron candidates with a transverse momentum of around 150 GeV. The excess corresponds to a local significance of 3.8σ if only uncertainties in the tau lepton misidentification rate and statistical uncertainties in simulation are considered. Since a similar excess is not observed in any other studied kinematic distribution, this discrepancy is considered to be a statistical fluctuation.

In summary, both studied control regions show good agreement between data and simulation. Small discrepancies are understood and covered by uncertainties assigned to simulated events (see Section 6.8 for a detailed discussion of these). These studies validate the usage of simulated samples for the prediction of Standard Model background processes in the signal region.

6.8 Evaluation of systematic uncertainties

Various systematic uncertainties affecting the normalization and the shape of the transverse momentum distribution of the leading tau lepton candidate are taken into account before the result of this search is evaluated. The considered uncertainties are applied to simulated background and signal events and will be listed in the following. If applicable, the recommendations given by expert groups within the CMS experiment have been followed. An overview of the systematic uncertainties can be found at the end of this section in Table 6.4.

6.8.1 Normalization uncertainties

The integrated luminosity used to scale the simulated samples to the actual number of collected data events has been determined by recording relative changes in the instantaneous luminosity with the cluster counting method in the silicon pixel detector. The absolute scale of the luminosity has been derived with Van der Meer scans [157]. The uncertainty in the integrated luminosity, determined with this procedure, amounts to 2.6% [157].

Additionally, uncertainties in the cross sections of the Standard Model background processes are taken into account. An uncertainty of 15% is applied to the cross section of $t\bar{t}$ + jets production [166]. Normalization uncertainties of 50% are considered for the production of W+jets events, DY + jets events and single top quark events. All other background processes get a 100% uncertainty assigned on their rate. All uncertainties in Standard Model cross sections are used as priors in the statistical evaluation and are adjusted during the limit setting procedure performed with the *theta package* [168] (see [169] for a detailed description of the procedure). The fit yields much smaller uncertainties in the normalization of the important background processes, meaning that background composition can be determined much more precisely during the limit-setting procedure. In order to test if the assumed normalization uncertainties in the cross sections have a large impact on the final results, all uncertainties have been decreased by 50% and the resulting change in the expected limit has been checked. Since the expected limit with the smaller uncertainties applied has varied only slightly compared to the expected limit obtained with the above mentioned uncertainties, the values chosen do not influence the final expected sensitivity.

6.8.2 Shape uncertainties

In order to evaluate the impact of the shape uncertainties listed in the following, the change in the number of events in each of the four considered p_T regions of the leading tau lepton candidate transverse momentum distribution is calculated after a variation of the studied observable by one standard deviation up and down. The difference to the nominal value is taken as uncertainty.

Tau lepton uncertainties

An uncertainty is assigned to the factors that correct the misidentification rate of jets as tau lepton candidates in simulation. The applied uncertainty is split into a statistical component and into uncertainties due to the subtraction of the non-W+jets backgrounds, the difference in flavor composition of the jets that are reconstructed as tau lepton candidates in the W+jets sideband and the signal region, and the non-uniform correction factors for tau lepton candidates with $|\eta| > 1.5$ (Section 6.6.5). Since the difference in flavor compositions between the sideband and the signal region is not the same in the electron and in the muon channel, these uncertainties differ in the two channels. In the electron channel the uncertainty is determined to 30.7%, while in the muon channel the uncertainty due to the flavor composition amounts to 24.2%. Both uncertainties are estimated with the likelihood fit described in Section 6.6.5 performed in the lowest of the four considered regions of the p_T distribution of the leading tau lepton candidate. The differences in the flavor composition between the muon channel and the electron channel are mainly caused by the veto of events selected by EXO-12-030 in the muon channel. Additionally, the veto on the presence of muon candidates in the electron channel and small differences in the kinematic cuts between the two channels can generate variations between the uncertainties in these channels.
The total uncertainties in the correction factors lie between 25% and 55% for the three low-p_T regions. In the highest p_T region uncertainties between 55% and 75% are assigned. A summary of the uncertainties in both channels is shown in Table 6.3.

An uncertainty is applied to the correction factors that account for the different misidentification rates of electrons as tau lepton candidates in data and in simulation. For tau lepton candidates with $|\eta| \leq 2.1$ that can be matched to a generator-level electron within $\Delta R(\tau, e) < 0.5$ this uncertainty amounts to 20%. For matched tau lepton candidates with $|\eta| > 2.1$ an uncertainty of 25% is applied [142].

For each tau lepton candidate in a simulated event that can be matched within $\Delta R < 0.5$ to a generator-level tau lepton an uncertainty of 6% is assigned due to uncertainties in its identification efficiency [142].

The energy of all tau lepton candidates is changed by ±3%, the energy resolution by ±10%, in order to account for uncertainties in the tau lepton energy measurement [142]. The value of $\not{E}_T$ and S_T are also recalculated using these shifts in the tau lepton candidate energy and resolution.

Uncertainties in the electron and muon identification, isolation, and trigger efficiencies

Uncertainties depending on η and p_T due to the trigger, identification, and isolation efficiencies of muon and electron candidates in simulation are assigned in the respective channel. For the muon channel these uncertainties are provided by a group within the CMS experiment responsible for muon triggering, reconstruction and identification.

channel	η	p_T^τ	s	stat. unc.	syst. unc.
muon channel	$\lvert\eta\rvert < 1.5$	$20\,\text{GeV} \leq p_T < 60\,\text{GeV}$	1.05	± 0.03	± 0.26
		$60\,\text{GeV} \leq p_T < 120\,\text{GeV}$	0.86	± 0.06	± 0.22
		$120\,\text{GeV} \leq p_T < 200\,\text{GeV}$	0.59	± 0.14	± 0.20
		$p_T \geq 200\,\text{GeV}$	1.06	± 0.50	± 0.29
	$\lvert\eta\rvert > 1.5$	$20\,\text{GeV} \leq p_T < 60\,\text{GeV}$	1.05	± 0.03	± 0.33
		$60\,\text{GeV} \leq p_T < 120\,\text{GeV}$	0.86	± 0.06	± 0.30
		$120\,\text{GeV} \leq p_T < 200\,\text{GeV}$	0.59	± 0.14	± 0.29
		$p_T \geq 200\,\text{GeV}$	1.06	± 0.50	± 0.35
electron channel	$\lvert\eta\rvert < 1.5$	$20\,\text{GeV} \leq p_T < 60\,\text{GeV}$	1.05	± 0.03	± 0.32
		$60\,\text{GeV} \leq p_T < 120\,\text{GeV}$	0.86	± 0.06	± 0.27
		$120\,\text{GeV} \leq p_T < 200\,\text{GeV}$	0.59	± 0.14	$^{+0.26}_{-0.23}$
		$p_T \geq 200\,\text{GeV}$	1.06	± 0.50	$^{+0.61}_{-0.34}$
	$\lvert\eta\rvert > 1.5$	$20\,\text{GeV} \leq p_T < 60\,\text{GeV}$	1.05	± 0.03	± 0.38
		$60\,\text{GeV} \leq p_T < 120\,\text{GeV}$	0.86	± 0.06	± 0.32
		$120\,\text{GeV} \leq p_T < 200\,\text{GeV}$	0.59	± 0.14	$^{+0.29}_{-0.25}$
		$p_T \geq 200\,\text{GeV}$	1.06	± 0.50	$^{+0.65}_{-0.40}$

Table 6.3: Summary of the statistical and systematic uncertainties in the derived correction factors that account for different misidentification rates in data and simulation for the muon and the electron channel as a function of $|\eta|$ and p_T of the tau lepton candidate. In the fourth column, the variable s refers to the derived correction factor (see Section 6.6.5).

They consist of a statistical uncertainty in the correction factors, an uncertainty of 0.5% due to the isolation efficiency, and uncertainties of 0.2% each due to the trigger and the identification efficiencies [149].

In the electron channel the uncertainties in the correction factors amounts to 0.3% for central electron candidates with $p_T > 50\,\text{GeV}$. For other p_T and η values, the uncertainty ranges up to 25% [128, 129].

Uncertainties in the energy measurement of jets

Uncertainties in the jet energy corrections and in the factors applied to compensate for differences in the jet energy resolutions in data and simulation are considered in the analysis. In both cases, also $\not{E}_T$ and S_T are recalculated using the jet p_T varied by one standard deviation. The η and p_T dependent uncertainties in the jet energy corrections are below 3%. The uncertainties in the η dependent jet energy resolution correction factors lie between 2% and 4% for jets with $|\eta| < 2.3$. For jets with higher pseudorapidities the uncertainty increases up to 10% [75].

Theory and generator uncertainties

In order to evaluate the effect of the choice of the parton density functions on the p_T distribution of the leading tau lepton candidate, a re-weighting of all simulated ($t\bar{t}$+jets) events with the CTEQ66 (CT10) pdf and its uncertainties is performed [91]. For the signal processes only uncertainties on the acceptance are considered, since changes in the normalization are covered by the theoretical uncertainties on the signal cross sections.

Uncertainties due to missing higher order terms are estimated through variations of the renormalization and factorization scales for the production of $t\bar{t}+\text{jets}$ and $W+\text{jets}$ events. The uncertainties are determined by studying the variation in the leading tau lepton candidate p_T spectrum obtained by using simulated samples produced with scales enlarged and reduced by a factor of 2.

The matching of the partons generated with MadGraph to the jets produced in the showering and hadronization process, which is simulated by pythia, includes a free parameter, the matching threshold, in the MLM matching scheme [108]. Therefore, the W+jets events are re-generated using a matching threshold changed by a factor of two up and down in order to assign a systematic uncertainty due to the choice of this parameter.

An extra uncertainty in the leading jet-p_T re-weighting procedure in $W+\text{jets}$ events is not applied since the described uncertainties in the matching threshold and scales in the $W+\text{jets}$ simulation are larger than the observed discrepancies and therefore cover the observed mismodeling completely.

For the determination of the uncertainty in the top p_T re-weighting, the correction factors that aim at diminishing the difference between the $t\bar{t}+\text{jets}$ simulation and data in the top p_T spectrum [165] are taken into account twice or are omitted completely to model the up and down variation.

Uncertainties in pileup re-weighting

An uncertainty of $\pm 3.5\,\text{mb}$ [159, 170] is assigned on the inelastic pp cross section of $69.4\,\text{mb}$ used in the pileup re-weighting procedure. This describes the uncertainty in the expected number of interactions due to the uncertainty in the total inelastic cross section and the integrated luminosity. The uncertainty is propagated to the analysis by studying the change of the p_T distribution of the leading tau lepton candidate when the total inelastic cross section used in the pileup re-weighting procedure is increased and decreased by one standard deviation.

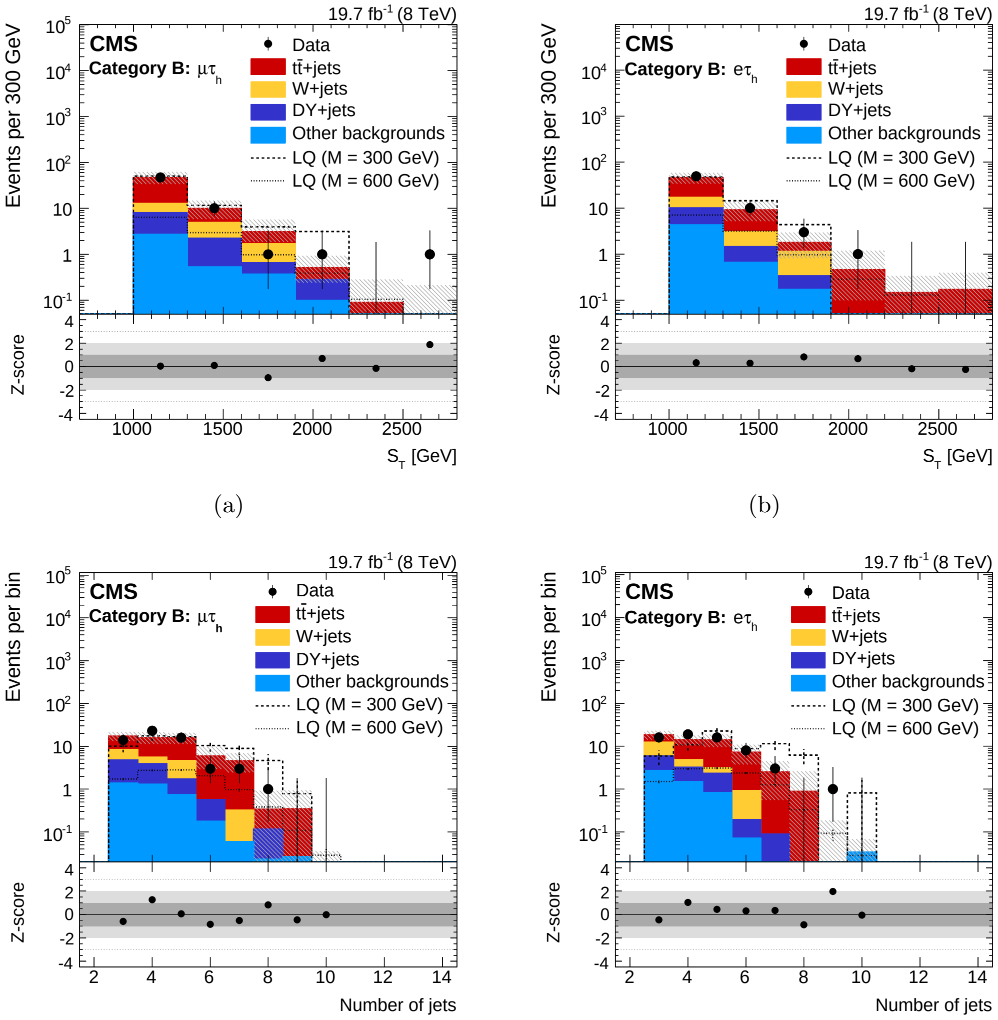
A summary of the applied uncertainties, their magnitude, and the respective size of these uncertainties after the application of the final signal selection (described in Section 6.5) can be found in Table 6.4 for both channels. The uncertainties, which are presented as the average over all four considered p_T regions of the distribution of the leading tau lepton candidate, do not exceed 11% in both channels. For the background processes the dominant uncertainties are the uncertainties in the correction factors for the misidentification rate of jets as tau lepton candidates and the pdf uncertainties. The uncertainty in the tau lepton candidate reconstruction, identification, and isolation con-

		muon channel		electron channel	
Systematic uncertainty	Magnitude (%)	B (%)	S (%)	B (%)	S (%)
Integrated luminosity	2.6	2.6	2.6	2.6	2.6
Electron reco/ID/iso & trigger	p_T, η dependent	—	—	1.4	2.2
Muon reco/ID/iso & trigger	1.1	0.9	0.9	—	—
τ lepton reco/ID/iso	6.0	1.5	3.0	0.6	3.1
τ lepton energy scale	3.0	2.3	2.7	0.6	1.5
τ lepton energy resolution	10.0	1.2	1.3	0.2	0.1
Jet energy scale	p_T, η dependent	4.2	1.9	5.6	2.7
Jet energy resolution	η dependent	0.8	0.3	1.6	0.8
Pileup	5.0	0.8	0.3	0.9	0.5
Pdf (on acceptance)	—	—	0.7	—	0.9
Pdf (on background)	—	8.7	—	8.3	—
Jet $\rightarrow \tau$ misidentification rate	p_T dependent	8.2	1.0	10.9	0.8
$e \rightarrow \tau$ misidentification rate	$\|\eta\|$ dependent	0.1	0.1	0.1	0.1
$t\bar{t}$ factorization/renormalization	$^{+100}_{-50}$	$^{+6.1}_{-5.9}$	—	$^{+2.9}_{-2.7}$	—
Top quark p_T re-weighting	p_T dependent	0.1	—	0.1	—
W+jets factorization/renormalization	$^{+100}_{-50}$	4.3	—	0.3	—
W+jets matching threshold	$^{+100}_{-50}$	1.3	—	2.5	—

Table 6.4: Summary of the sources of the considered systematic uncertainties, apart from the normalization uncertainties in the cross sections of the SM background processes. The magnitudes of the uncertainties and their effect on the total number of background events B as well as on signal events S for a leptoquark with a mass of 550 GeV are presented for the muon and the electron channel.

stitute the major uncertainty for the signal events, while the uncertainty in the correction factors for the misidentification rate of jets is relatively small for signal events. This can be explained by the different fractions of misidentified tau lepton candidates in signal events and the background processes. In the background a large portion of selected tau lepton candidates are misidentified jets. This results in a large uncertainty due to the misidentification rate, but a small uncertainty in the reconstruction, identification and isolation efficiency of real tau leptons. In the signal events most of the studied tau lepton candidates are real tau leptons. Therefore, the relative contributions of the systematic uncertainties are swapped.

The effect of the individual uncertainties in each of the four search bins of the leading p_T tau lepton candidate distribution is illustrated for the $t\bar{t}$ + jets and W + jets simulation as well as for the signal sample for a leptoquark with a mass of 500 GeV in Appendix A.



(a)

(b)

(c)

(d)

Figure 6.16: Comparison of the number of observed events (*black dots*) and the number of predicted events by simulation (*colored regions*) in the muon channel (*left*) and the electron channel (*right*) for the signal region after the full signal selection has been applied. The shaded areas in the figures represent the total uncertainties. Presented are the distributions of (a) and (b) S_T, (c) and (d) the number of jets. The figures are published in [2].

6.9 Results

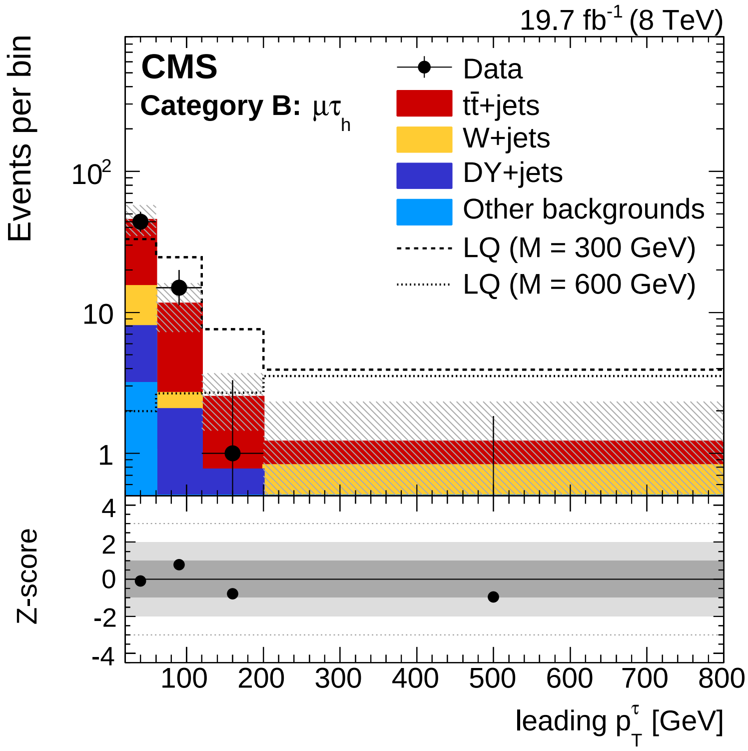
After the estimation of the systematic uncertainties, the final results of this search can be evaluated. Figure 6.16 presents the distributions of S_T and the number of jets for the muon channel and for the electron channel. In both channels, the final signal selection requirements described in Section 6.5 and all corrections detailed in Section 6.6 have been applied. The main backgrounds are $t\bar{t}$ + jets production and to a minor extent W+jets events. To check the agreement between data and simulation, the Z-score is presented. The Z-score expresses the deviation between simulation and data in standard deviations taking into account the statistical and systematic uncertainties in simulation. In all distributions data agree with the prediction from the Standard Model within the uncertainties.

For the final statistical interpretation of the results the p_T distributions of the leading tau lepton candidate in the muon and in the electron channel are utilized. The p_T spectra are presented in Fig. 6.17. Also in these distributions good agreement between data and prediction is observed within the uncertainties. In total, 60 data events have been selected in the muon channel. This agrees very well with 60.4 ± 10.9 predicted events from the Standard Model. The same is true for the electron channel, where 63 data events have been observed, while 57.4 ± 11.2 are expected from Standard Model processes. The expected event yields for the different signal processes in the various search regions are summarized in Tables 6.5 and 6.6, together with the observed events in data events and the SM background prediction. Additionally, the tables present the selection efficiencies for the considered signal processes. The efficiencies are calculated with respect to the expected number of signal events that contain a muon and an hadronically decaying tau lepton. The efficiencies range from 0.01% to 15%. For light leptoquarks the selection efficiencies are relatively small, but the efficiency increases with the mass of the leptoquarks. This is due to the high S_T requirement. The efficiency loss of this criterion decreases with the mass of the leptoquarks, since the decay products of heavier leptoquarks have higher transverse momenta, which results on average in higher S_T values of these events.

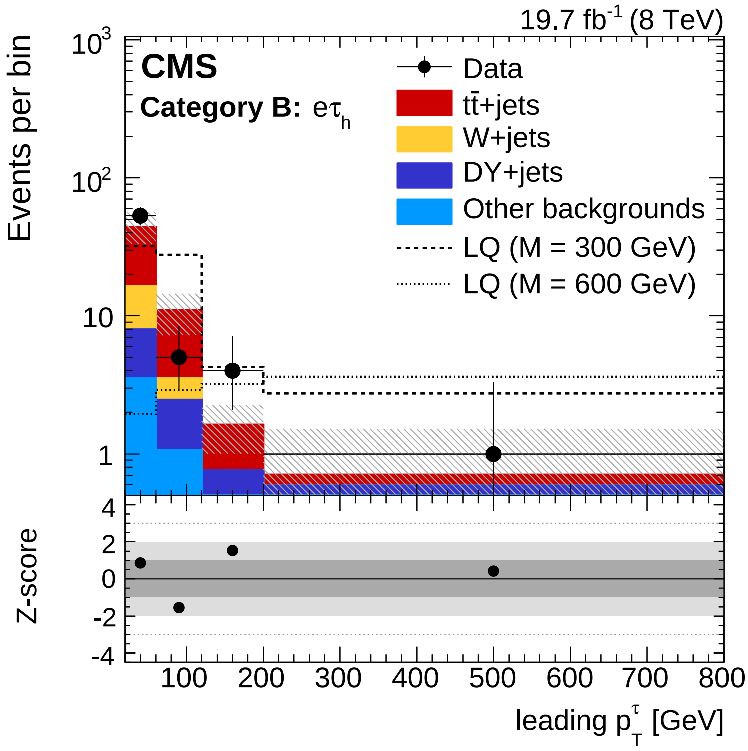
In summary, no significant deviation between data and simulation is observed. Therefore, no hint for new physics has been found in this search and the result will be used to constrain physics beyond the Standard Model.

6.10 Interpretation of the results

Based on the results described in the previous section, exclusion limits on the pair production cross section times branching ratio of third-generation leptoquarks decaying into a top quark and a tau lepton are set. For the determination of the exclusion limit a shape analysis based on the p_T distribution of the leading tau lepton candidates in the muon channel, shown in Fig. 6.17(a), and in the electron channel, presented in Fig. 6.17(b), is performed. The limits are determined using the Bayesian method of the *theta package* [168], described in detail in [169]. Nuisance parameters describe the effect of the systematic uncertainties, detailed in Section 6.8, and the statistical uncertainties of simulated events. Uncertainties present in both channels are handled as fully correlated. In the limit setting procedure, the sensitivity to the signal production arises mainly from the last two or three high-p_T regions, dependent on the leptoquark mass. The remaining p_T regions constrain the normalization of the SM background processes. The result of the limit calculation is shown in Fig. 6.18. The observed limit is contained in the 1σ band for all studied leptoquark masses. Assuming a branching ratio of 100% for the decay LQ → $t\tau$, pair production cross sections higher than approximately 340 pb (334 pb expected) for leptoquarks with masses of 200 GeV and higher than approximately 7 fb (8 fb expected) for masses of 800 GeV are excluded at 95% C.L. A comparison of the observed limit with the NLO theory cross section [37] yields an exclusion of third-generation leptoquarks with masses between 279 GeV and 678 GeV (269 GeV and 668 GeV expected) at



(a)



(b)

Figure 6.17: Distributions of the p_T of the leading tau lepton candidates in the event for (a) the muon channel and (b) the electron channel. These distributions are used for the final statistical interpretation of the search. The hatched areas in the figures represent the statistical and systematic uncertainties. The figures are published in [2].

Process	$p_T^\tau < 60\,\mathrm{GeV}$	$60 \leq p_T^\tau < 120\,\mathrm{GeV}$	$120 \leq p_T^\tau < 200\,\mathrm{GeV}$	$p_T^\tau \geq 200\,\mathrm{GeV}$	ϵ_{LQ} (%)
LQ ($M = 200\,\mathrm{GeV}$)	$21.0 \pm 12.0\,^{+7}_{-2}$	$0.0 \pm 0.1 \pm 0.0$	$0.0 \pm 0.1 \pm 0.1$	$0.0 \pm 0.1 \pm 0.1$	0.01
LQ ($M = 250\,\mathrm{GeV}$)	$31.0 \pm 8.2\,^{+6.6}_{-3.4}$	$13.1 \pm 5.5\,^{+1.1}_{-2.9}$	$0.0 \pm 0.1 \pm 0.1$	$0.0 \pm 0.1 \pm 0.1$	0.11
LQ ($M = 300\,\mathrm{GeV}$)	$33.1 \pm 5.3\,^{+2.8}_{-3.8}$	$24.6 \pm 4.6\,^{+2.8}_{-2.1}$	$7.6 \pm 2.6\,^{+1.1}_{-1.7}$	$3.9 \pm 1.8\,^{+0.9}_{-0.3}$	0.43
LQ ($M = 350\,\mathrm{GeV}$)	$18.1 \pm 2.6\,^{+1.8}_{-1.4}$	$13.3 \pm 2.2\,^{+1.0}_{-1.1}$	$7.2 \pm 1.6\,^{+0.8}_{-0.7}$	$2.9 \pm 0.9\,^{+0.5}_{-1.4}$	0.65
LQ ($M = 400\,\mathrm{GeV}$)	$13.9 \pm 1.4\,^{+1.1}_{-2.6}$	$13.4 \pm 1.4\,^{+1.0}_{-1.1}$	$7.8 \pm 1.1\,^{+0.8}_{-0.6}$	$4.1 \pm 0.8\,^{+0.6}_{-0.8}$	1.42
LQ ($M = 450\,\mathrm{GeV}$)	$10.1 \pm 0.9\,^{+0.8}_{-1.9}$	$8.6 \pm 0.8\,^{+0.8}_{-0.8}$	$7.1 \pm 0.7\,^{+0.5}_{-0.6}$	$5.8 \pm 0.6\,^{+0.7}_{-0.6}$	2.32
LQ ($M = 500\,\mathrm{GeV}$)	$5.2 \pm 0.4\,^{+0.5}_{-0.9}$	$6.0 \pm 0.5 \pm 0.5$	$5.3 \pm 0.4\,^{+0.4}_{-0.5}$	$4.4 \pm 0.4\,^{+0.7}_{-0.5}$	3.02
LQ ($M = 550\,\mathrm{GeV}$)	$3.2 \pm 0.3\,^{+0.3}_{-0.6}$	$4.4 \pm 0.3\,^{+0.4}_{-0.3}$	$4.3 \pm 0.3\,^{+0.5}_{-0.4}$	$4.0 \pm 0.3 \pm 0.4$	4.4
LQ ($M = 600\,\mathrm{GeV}$)	$2.0 \pm 0.1\,^{+0.2}_{-0.5}$	$2.7 \pm 0.2 \pm 0.2$	$2.7 \pm 0.2 \pm 0.2$	$3.5 \pm 0.2 \pm 0.4$	5.54
LQ ($M = 650\,\mathrm{GeV}$)	$1.3 \pm 0.1\,^{+0.1}_{-0.3}$	$1.8 \pm 0.1\,^{+0.1}_{-0.2}$	$2.0 \pm 0.1 \pm 0.2$	$2.5 \pm 0.1\,^{+0.3}_{-0.2}$	6.74
LQ ($M = 700\,\mathrm{GeV}$)	$0.7 \pm 0.1 \pm 0.1$	$1.1 \pm 0.1 \pm 0.1$	$1.1 \pm 0.1 \pm 0.1$	$1.6 \pm 0.1\,^{+0.2}_{-0.1}$	7.23
LQ ($M = 750\,\mathrm{GeV}$)	$0.4 \pm 0.1 \pm 0.1$	$0.5 \pm 0.1 \pm 0.1$	$0.7 \pm 0.1 \pm 0.1$	$1.1 \pm 0.1 \pm 0.1$	7.3
LQ ($M = 800\,\mathrm{GeV}$)	$0.2 \pm 0.1 \pm 0.1$	$0.4 \pm 0.1 \pm 0.1$	$0.5 \pm 0.1 \pm 0.1$	$0.8 \pm 0.1 \pm 0.1$	8.38
$t\bar{t}$ + jets	$29.9 \pm 2.9\,^{+7.3}_{-7.2}$	$8.8 \pm 1.3\,^{+3.2}_{-3.4}$	$1.7 \pm 0.6\,^{+0.6}_{-0.6}$	$0.4 \pm 0.3\,^{+0.9}_{-0.4}$	
W + jets	$7.4 \pm 1.7\,^{+5.1}_{-5.1}$	$0.6 \pm 0.5 \pm 0.6$	$0.0 \pm 0.1 \pm 0.1$	$0.4 \pm 0.4 \pm 0.4$	
DY + jets	$4.8 \pm 0.7 \pm 2.5$	$1.8 \pm 0.4\,^{+1.1}_{-0.9}$	$0.5 \pm 0.2 \pm 0.3$	$0.4 \pm 0.2 \pm 0.2$	
Others	$3.1 \pm 0.9\,^{+1.8}_{-1.9}$	$0.2 \pm 0.1\,^{+0.8}_{-0.3}$	$0.2 \pm 0.1 \pm 0.4$	$0.1 \pm 0.1\,^{+0.1}_{-0.2}$	
Total $N_{\mathrm{Bkg}}^{\mathrm{Exp}}$	$45.2 \pm 3.5\,^{+9.4}_{-9.3}$	$11.5 \pm 1.4\,^{+3.4}_{-3.6}$	$2.5 \pm 0.6 \pm 0.8$	$1.2 \pm 0.5\,^{+1.0}_{-0.6}$	
N^{Obs}	44	15	1	0	

Table 6.5: Event yields in the four regions of the p_T spectrum of the leading tau lepton candidates predicted for the signal and background processes in the muon channel after the application of the full signal selection. Additionally, the total number of expected background events ($N_{\mathrm{Bkg}}^{\mathrm{Exp}}$) and the number of observed data events (N^{Obs}) in each of the four p_T regions is presented. The errors on the predicted events are split into statistical and systematic uncertainties. The last column shows the selection efficiency of the signal events (ϵ_{LQ}). The efficiencies are calculated with respect to all expected events with a $\mu\tau_h$-pair determined on generator level.

Process	$p_T^{\tau} < 60\,\text{GeV}$	$60 \leq p_T^{\tau} < 120\,\text{GeV}$	$120 \leq p_T^{\tau} < 200\,\text{GeV}$	$p_T^{\tau} \geq 200\,\text{GeV}$	ϵ_{LQ}
LQ ($M = 200\,\text{GeV}$)	$32.0 \pm 19.0\,^{+6}_{-4}$	$0.1 \pm 0.1 \pm 0.1$	$0.0 \pm 0.1 \pm 0.1$	$0.0 \pm 0.1 \pm 0.1$	0.02
LQ ($M = 250\,\text{GeV}$)	$33.3 \pm 8.7\,^{+8.2}_{-5.8}$	$11.9 \pm 5.3\,^{+2.6}_{-2.4}$	$0.0 \pm 0.1 \pm 0.1$	$0.0 \pm 0.1 \pm 0.1$	0.16
LQ ($M = 300\,\text{GeV}$)	$31.9 \pm 5.2\,^{+3.2}_{-9.1}$	$27.7 \pm 4.6\,^{+3.9}_{-4.4}$	$4.2 \pm 1.9\,^{+1.1}_{-0.3}$	$2.7 \pm 1.6\,^{+0.2}_{-0.3}$	0.70
LQ ($M = 350\,\text{GeV}$)	$19.6 \pm 2.6\,^{+2.7}_{-3.6}$	$19.6 \pm 2.5\,^{+2.0}_{-2.1}$	$8.6 \pm 1.7\,^{+1.4}_{-0.8}$	$4.7 \pm 1.3 \pm 0.5$	1.35
LQ ($M = 400\,\text{GeV}$)	$12.7 \pm 1.4\,^{+1.7}_{-2.6}$	$14.6 \pm 1.5\,^{+2.0}_{-1.4}$	$8.1 \pm 1.1\,^{+1.2}_{-1.3}$	$4.8 \pm 0.9\,^{+0.6}_{-0.4}$	2.22
LQ ($M = 450\,\text{GeV}$)	$7.8 \pm 0.7\,^{+1.3}_{-1.4}$	$10.1 \pm 0.9\,^{+0.9}_{-1.1}$	$7.2 \pm 0.7\,^{+1.0}_{-0.7}$	$5.5 \pm 0.6\,^{+0.6}_{-0.8}$	3.65
LQ ($M = 500\,\text{GeV}$)	$4.8 \pm 0.4\,^{+0.5}_{-1.2}$	$7.3 \pm 0.5\,^{+0.8}_{-0.9}$	$5.5 \pm 0.4 \pm 0.6$	$5.2 \pm 0.4\,^{+0.7}_{-0.6}$	5.34
LQ ($M = 550\,\text{GeV}$)	$3.3 \pm 0.2\,^{+0.4}_{-1.0}$	$4.3 \pm 0.3 \pm 0.4$	$4.4 \pm 0.3 \pm 0.4$	$4.3 \pm 0.3 \pm 0.5$	7.28
LQ ($M = 600\,\text{GeV}$)	$1.9 \pm 0.1\,^{+0.2}_{-0.6}$	$2.9 \pm 0.2 \pm 0.3$	$3.2 \pm 0.2 \pm 0.3$	$3.6 \pm 0.2 \pm 0.4$	9.61
LQ ($M = 650\,\text{GeV}$)	$1.2 \pm 0.1\,^{+0.1}_{-0.4}$	$1.8 \pm 0.1 \pm 0.2$	$2.0 \pm 0.1 \pm 0.2$	$2.4 \pm 0.1\,^{+0.3}_{-0.3}$	10.89
LQ ($M = 700\,\text{GeV}$)	$0.7 \pm 0.1\,^{+0.1}_{-0.2}$	$1.1 \pm 0.1 \pm 0.1$	$1.5 \pm 0.1 \pm 0.1$	$1.9 \pm 0.1 \pm 0.2$	13.11
LQ ($M = 750\,\text{GeV}$)	$0.4 \pm 0.1 \pm 0.1$	$0.7 \pm 0.1 \pm 0.1$	$0.7 \pm 0.1 \pm 0.1$	$1.4 \pm 0.1\,^{+0.1}_{-0.2}$	13.84
LQ ($M = 800\,\text{GeV}$)	$0.2 \pm 0.1 \pm 0.1$	$0.4 \pm 0.1 \pm 0.1$	$0.5 \pm 0.1 \pm 0.1$	$0.9 \pm 0.1 \pm 0.1$	14.82
$t\bar{t}$ + jets	$27.7 \pm 2.4 \pm 7.7$	$7.5 \pm 1.2\,^{+2.1}_{-2.8}$	$0.9 \pm 0.4 \pm 0.3$	$0.1 \pm 0.1\,^{+0.6}_{-0.1}$	
W + jets	$8.5 \pm 1.8\,^{+5.3}_{-5.4}$	$1.1 \pm 0.6\,^{+0.6}_{-0.7}$	$0.0 \pm 0.1 \pm 0.1$	$0.0 \pm 0.1 \pm 0.1$	
DY + jets	$4.4 \pm 0.7\,^{+2.3}_{-2.4}$	$1.4 \pm 0.4\,^{+0.8}_{-0.7}$	$0.6 \pm 0.2 \pm 0.3$	$0.2 \pm 0.1 \pm 0.1$	
Others	$3.5 \pm 1.0\,^{+2.5}_{-3.0}$	$1.1 \pm 0.5\,^{+0.7}_{-0.8}$	$0.2 \pm 0.1 \pm 0.2$	$0.4 \pm 0.3 \pm 0.2$	
Total $N_{\text{Bkg}}^{\text{Exp}}$	$44.1 \pm 3.2\,^{+10.0}_{-10.1}$	$11.0 \pm 1.5\,^{+2.5}_{-3.1}$	$1.6 \pm 0.5 \pm 0.5$	$0.7 \pm 0.4\,^{+0.7}_{-0.3}$	
N^{Obs}	53	5	4	1	

Table 6.6: Event yields in the four regions of the p_T spectrum of the leading tau lepton candidates predicted for the signal and background processes in the electron channel after the application of the full signal selection. Additionally, the total number of expected background events ($N_{\text{Bkg}}^{\text{Exp}}$) and the number of observed data events (N^{Obs}) in each of the four p_T regions is presented. The errors on the predicted events are split into statistical and systematic uncertainties. The last column shows the selection efficiency of the signal events (ϵ_{LQ}). The efficiencies are calculated with respect to all expected events with a $e\tau_h$-pair determined on generator level.

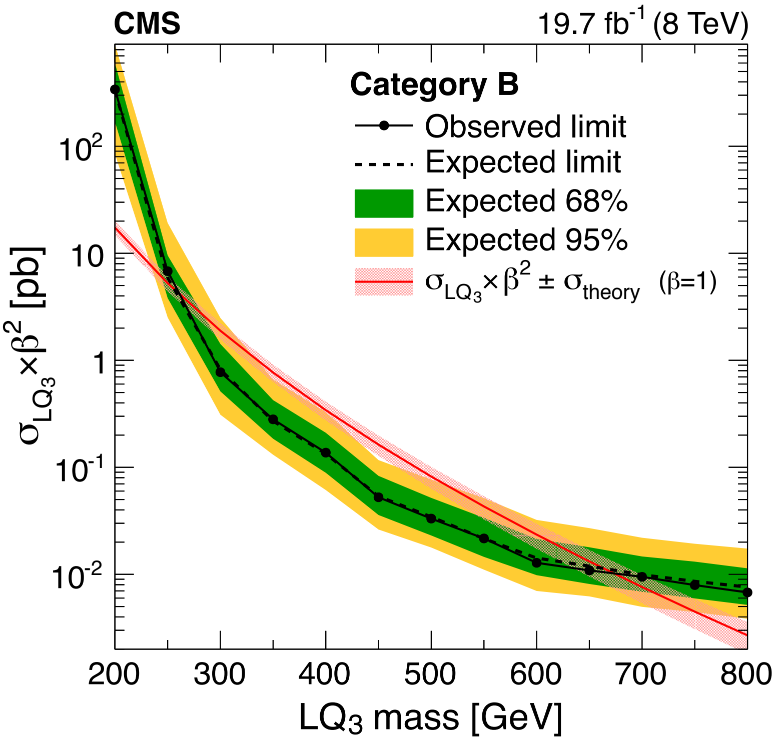


Figure 6.18: Expected (*dashed line*) and observed (*solid line*) upper limits on the cross section times branching ratio (called β here) of pair production of third-generation leptoquarks at 95% C.L. as a function of the leptoquark mass. For the expected limit also the one (*green*) and two sigma bands (*yellow*) are shown. The theory curve (*red line*) and its uncertainty (*hatched red area*), which consists of uncertainties due to the choice of the pdf and the factorization and renormalization scales, are published in [37]. The figure is published in [2].

95% C.L. under the assumption of $\mathrm{BR}(\mathrm{LQ} \to t\tau) = 1$. Due to the high S_T requirement included in the event selection, the sensitivity deteriorates towards lower leptoquark masses. Thus, leptoquarks with masses below 279 GeV cannot be excluded by this search.

6.10.1 Combination

Within [2] a combination of the analysis presented in this thesis with the analysis EXO-12-030 (*Category A*) [145] (see Section 6.1.1) has been performed. The analysis presented in this thesis will be called *inclusive analysis (Category B)* in the following. Also for the combination the Bayesian method of the *theta package* [168], described in [169], is utilized. The combination is a shape analysis in ten search regions, which are the four p_T regions of the distribution of the leading tau lepton candidate in the muon channel shown in Fig. 6.17(a), the equivalent four regions in the electron channel presented in Fig. 6.17(b) and the two search regions of EXO-12-030. Also here nuisance parameters are used to handle the statistical and systematic uncertainties. All common uncertainties between the inclusive analysis and EXO-12-030 are considered to be fully correlated. These are the uncertainty in the tau lepton identification, energy and energy resolution,

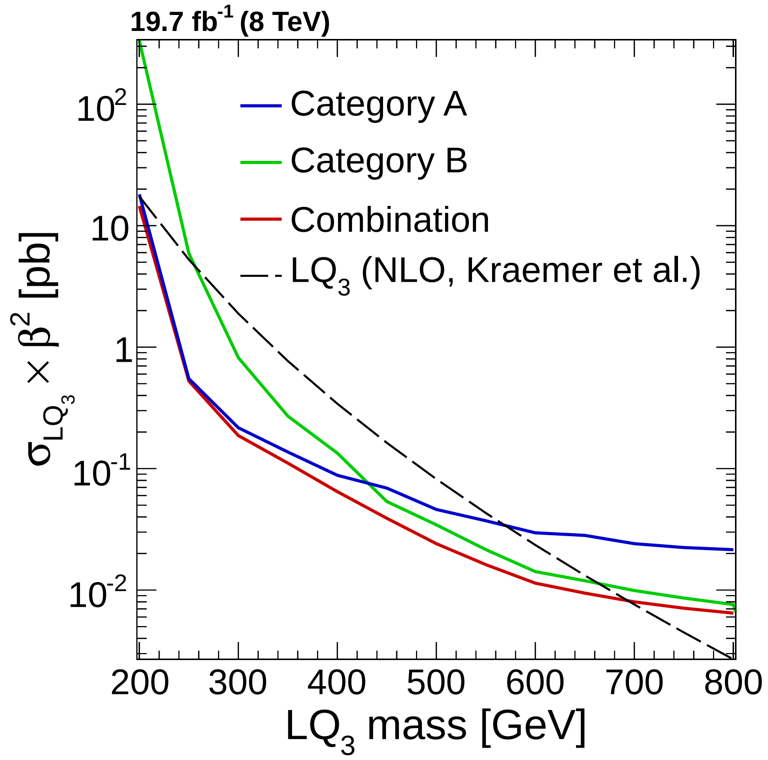


Figure 6.19: Expected limits at 95% C.L. on the cross section times branching ratio (called β here) as a function of the mass of the leptoquarks. Shown is a comparison of the results for EXO-12-030 (*Category A, blue*), the inclusive analysis presented in this thesis (*Category B, green*) and the combination of the two analyses (*red*).

the muon identification and isolation, the pileup re-weighting, in the pdfs, as well as the uncertainties in the jet energy resolution, jet energy corrections and the integrated luminosity. Other uncertainties can be treated as uncorrelated since these either are not considered in EXO-12-030, because of the data-driven background estimation approach, or since they are otherwise unique to one of the two analyses.

The expected limits at 95% C.L. on the cross section times branching ratio of pair produced third-generation leptoquarks obtained in the combination is illustrated together with the expected limit at 95% C.L. of the individual analyses in Fig. 6.19. Comparing the two analyses individually, it can be observed that the inclusive analysis has a significantly better sensitivity than EXO-12-030 at high leptoquark masses, while the latter yields better results at low leptoquark masses. This is due to the high S_T requirement imposed in the inclusive analysis, which becomes more efficient for higher leptoquark masses. The loss in efficiency of EXO-12-030 for high masses stems from the low signal selection efficiency of the requirement of a same-sign muon candidate and tau lepton candidate pair. Contrary, EXO-12-030 exhibits sensitivity for leptoquark masses below 270 GeV, where the inclusive analysis has not been able to exclude leptoquarks. In summary, the sensitivity of the combination is driven by EXO-12-030 at low leptoquark masses and by the inclusive analysis at high leptoquark masses, but is improved by the combination over the whole mass range.

Generally, the sensitivity of the combined result decreases for low leptoquark masses,

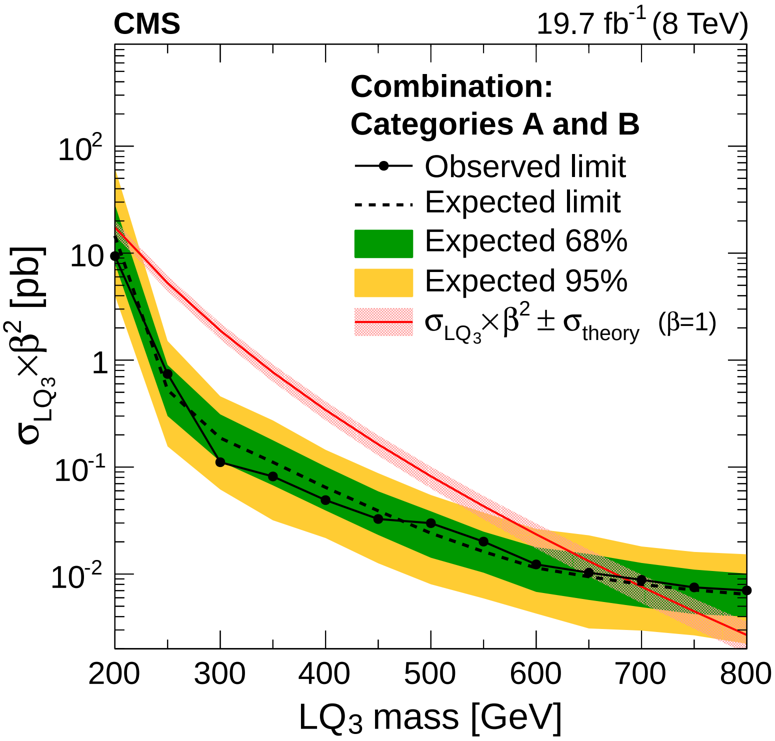


Figure 6.20: Expected (*dashed line*) and observed (*solid line*) upper limits on the cross section times branching ratio (called β here) of pair production of third-generation leptoquarks at 95% C.L. as a function of the leptoquark mass for the combination of the inclusive analysis and EXO-12-030 [145]. The one sigma bands (*green*) and two sigma bands (*yellow*) for the expected exclusion limit are presented. The theory curve (*red line*) and its uncertainty (*hatched red area*), which comprises uncertainties due to the choice of the pdf and the factorization and renormalization scales, have been determined in [37]. The figure is published in [2].

since the selection efficiencies, and also the discrimination power of the distribution of the p_T of the leading tau lepton candidate in case of the inclusive analysis, deteriorates due to smaller transverse momenta of the decay products of the leptoquarks. However, leptoquarks with masses of 200 GeV can still be excluded with the combined analysis, which means that the analysis can exclude leptoquarks down to masses close to the production threshold of leptoquarks decaying into a top quark and a tau lepton.

The expected and observed exclusion limits on the pair production cross section of third-generation leptoquarks times branching ratio of the combination are depicted in Fig. 6.20. The observed limit fluctuates within the 1σ band of the expected upper limit. The combined result excludes pair production cross sections above approximately 9 pb (15 pb expected) for leptoquarks with masses of 200 GeV and above approximately 7 fb (6 fb expected) for leptoquarks with masses of 800 GeV at 95% C.L.. Transferring this into a mass exclusion limit, exclusion limits at 95% C.L. of up to 685 GeV (695 GeV expected) are obtained. The results assume third-generation leptoquarks decaying with a branching ratio of 100% into a top quark and a tau lepton.

The presented exclusion limits can also be directly understood as an interpretation of a search of pair produced supersymmetric bottom squarks $\tilde{b}$ that decay via an R-parity violating interaction into a top quark and a tau lepton. Hence, these bottom squarks can be excluded up to masses of 685 GeV (695 GeV expected) at 95% C.L. under the assumption of pair production via the strong interaction and BR($\tilde{b} \rightarrow t\tau$)=1.

The third-generation leptoquark with charge $\frac{1}{3}$ considered in this analysis can also decay into a b quark and a neutrino. The exclusion limits derived by a search for pair produced bottom squarks decaying into a b quark and a neutralino performed by the CMS experiment [171] can be interpreted as exclusion limits for pair production of third-generation leptoquarks decaying into a b quark and a tau neutrino in case of a massless neutralino. Third-generation leptoquarks decaying with a branching ratio of 100% into a b quark and a neutrino can thus be excluded up to masses of 700 GeV (735 GeV expected) at 95% C.L.

The bottom squark search and the search presented in this thesis are used to constrain third-generation leptoquarks for arbitrary branching ratios to $t\tau$ and $b\nu$. The obtained upper exclusion limits on the pair production cross sections times branching ratio in both decay channels can be translated into upper limits on the branching ratios of third-generation leptoquarks decaying into a top quark and a tau lepton or a b quark and a neutrino with

$$\mathrm{BR(LQ)} = \sqrt{\frac{\sigma}{\sigma_{\mathrm{T}}}}, \tag{6.5}$$

where σ corresponds to the upper cross section limit derived before, and σ_T describes the theory cross section, taken from [37]. Hence, an exclusion limit at 95% C.L. on the branching ratio as a function of the leptoquark mass can be derived. This is shown for a combination of the two searches in the two different decay channels in Fig. 6.21. The combination is done under the assumptions of $\mathrm{BR(LQ} \rightarrow t\tau) + \mathrm{BR(LQ} \rightarrow b\nu) = 1$ and that signal processes decaying into $b\nu$ do not contribute to the channel $t\tau$ and vice versa. The excluded area has been obtained by scaling the signal selection efficiency for the $t\tau$-channel with the squared branching ratios BR^2, where the branching ratio runs between 0 and 1 in steps of 0.1, while the selection efficiency of the $b\nu$-channel is multiplied with $(1 - \mathrm{BR})^2$ at the same time. For each BR the expected and observed exclusion limits on the pair production cross section of third-generation leptoquarks are derived with the *theta package* and the upper limits on the BRs are calculated according to Eq. (6.5). Also here, common uncertainties between the channels are treated as fully correlated. Due to the combination of the two analyses, third-generation leptoquarks can be excluded over the full range of branching ratios up to 560 GeV (605 GeV expected) at 95% C.L.

6.11 Outlook

With the beginning of the LHC run period in the year 2015 the center-of-mass energy has been increased from 8 TeV to 13 TeV. At this energy, circa 40 fb^{-1} of pp collision data were collected by the CMS experiment in 2015 and 2016 [81]. In 2017, the LHC will continue to run at $\sqrt{s} = 13$ TeV. A data set at least as large as the one already collected is expected. This large amount of data, partly already available, will help to probe much higher mass scales and lower cross sections than possible with the 8 TeV

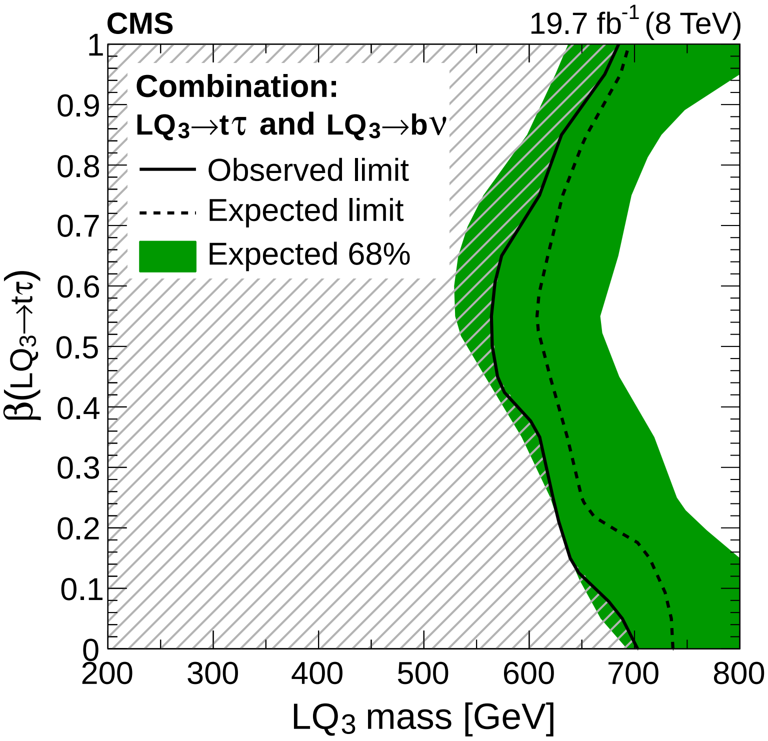


Figure 6.21: Expected (*dashed line*) and observed (*solid line*) upper limits at 95% C.L. on the branching ratio (called β here) of a third-generation leptoquark decaying into a top quark and a tau lepton candidate under the assumption of $\mathrm{BR}(\mathrm{LQ} \to t\tau) + \mathrm{BR}(\mathrm{LQ} \to b\nu) = 1$ in dependence of the leptoquark mass. The green band illustrates the 68% confidence band on the expected limit. The limit is a combination of the results obtained in [2] and [171]. The shaded region shows the excluded region at 95% C.L. The figure is published in [2].

data set. Besides the larger amount of available data, an improvement in the sensitivity of leptoquark searches is expected due to much higher pair production cross sections at $\sqrt{s} = 13$ TeV. A comparison of the pair production cross sections of scalar leptoquarks at $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV is presented in Fig. 6.22. While the cross sections of the main backgrounds of this search, $t\bar{t}$ + jets and W + jets, only increase by factors of approximately 3.4 [151] and 1.6 [150], respectively, the pair production cross sections are approximately nine times larger for leptoquarks with masses of 650 GeV and larger by a factor of approximately 15 for leptoquarks with masses of 950 GeV [37]. Studies suggest that with roughly 10 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV the expected exclusion limit will be better than the one in the 8 TeV-search presented in this thesis. The break-even point can also be much lower, if one manages e.g. to improve the reconstruction techniques or the selection efficiencies, or if data-driven estimations of the Standard Model backgrounds are used to reduce the systematic uncertainties in these.

Unfortunately, besides the mentioned advantages of the changed run conditions, also new challenges are imposed on the search due to the larger center-of-mass energy and higher instantaneous luminosities. As already said, the large center-of-mass energy helps to gain sensitivity for higher leptoquark masses. This means that their decay products receive a larger Lorentz boost due to the high mass of the leptoquark. Thus, the decay

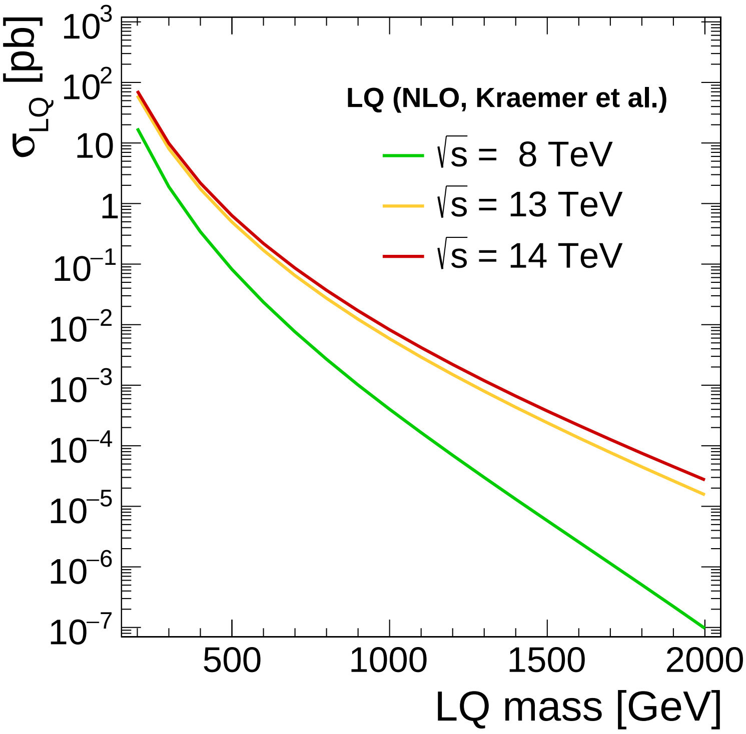


Figure 6.22: Pair production cross sections of scalar leptoquarks in dependence of the mass of the leptoquarks for different center-of-mass energies. The cross section have been calculated according to [37].

products of the top quarks and the tau leptons become more collimated, and therefore more difficult to reconstruct with conventional reconstruction techniques. Hence, techniques that operate on large-cone jets and aim at identifying the substructure within these will become more important for the analysis. Additionally, reconstruction and identification criteria like e.g. the isolation of particles have to be adopted in order to take the larger collimation of the final state particles into account. These new techniques possibly make also the fully hadronic channel accessible, since they are in principle able to significantly reduce the amount of background events from QCD multijet production to a manageable amount[4]. Challenges imposed on the analyses because of the higher instantaneous luminosity arise due to the higher amount of expected pileup interactions, that make the reconstruction and identification of final state objects more difficult[5].
Furthermore, searches targeting new final states will attract interest. Theories aiming at solving anomalies observed in decays of B mesons have been developed that e.g. predict leptoquarks decaying into a top quark and a muon with sizable cross section and branching ratio [31]. In certain parameter spaces, the branching ratio of leptoquarks decaying into $t\mu$ can exceed the branching ratio into $t\tau$ significantly [31], such that this channel provides an interesting opportunity to search for leptoquarks at the LHC. In searches for first- and second-generation leptoquarks, also single production of leptoquarks becomes interesting to study. For certain values of the lepton-quark coupling

---

[4] Some of these substructure techniques will be explained in Chapter 7 and applied in the analysis presented in Chapter 8.

[5] Chapter 7 presents studies of the influence of the application of new pileup mitigation techniques on substructure algorithms.

and leptoquark masses the single production cross section exceeds the pair production cross section already for $\sqrt{s} = 8\,\text{TeV}$ [40].

With the increased sensitivity for all leptoquark searches and the new accessible channels, LHC Run-II offers a promising opportunity to shed light on the question whether leptoquarks exist at the TeV scale or not.

7 Impact of pileup mitigation techniques on the top tagging performance in LHC Run-II

Many BSM signatures feature top quarks that can receive a large Lorentz boost since these are either produced in the decay of (very) heavy objects, like high-mass leptoquarks, or recoil against the latter as it is the case in the Dark Matter search presented in the next chapter (see Chapter 8). The decay products of these boosted top quarks are produced in narrow cones around the flight direction of the top quark and are therefore not well separated from each other. Since the standard reconstruction techniques of jets and leptons are optimized for the resolved case (see Chapter 5), these reconstruction techniques perform much worse or even fail for high-p_T top quarks. Thus, other techniques have been developed that aim at reconstructing non-isolated leptons close to the b quark jets in case of leptonically decaying top quarks[1] ($t \to W^+b \to \ell^+\nu_\ell b$) and at capturing all jets produced in hadronic decays of top quarks ($t \to W^+b \to q\bar{q}'b$), within one large-cone jet. In order to distinguish hadronically decaying top quarks from QCD multijet production, the large-cone jets subsequently have to pass additional identification criteria, including requirements on their mass, which should be compatible with the top quark mass, or requirements on their substructure, like the presence of three small-cone jets (so called subjets) within the large-cone jet.

The techniques used to identify hadronically decaying top quarks are referred to as *top tagging*. In both the ATLAS and the CMS experiments, many different top tagging algorithms are available and used in a variety of searches for BSM physics and in SM measurements. All of them rely on requirements on reconstructed jet quantities, like jet masses or the number of subjets. Since unintended clustering of pileup activity into large-cone jets can have a significant impact on these jet quantities, the performance of top tagging algorithms depends on an efficient mitigation of pileup effects. This is particularly important for LHC Run-II, where high instantaneous luminosities are achieved, which results in high pileup activity. The studies presented in this chapter aim at comparing the impact of the standard pileup mitigation technique used at the CMS experiment, Charge Hadron Subtraction (CHS) [134] (see Section 5.5.2), and the newer PUPPI algorithm [6,135] (see Section 5.5.2) on top tagging observables, as well as top tagging efficiencies and misidentification rates. The presented studies have been the first studies that used PUPPI-corrected jets for top tagging. Together with other studies performed at the CMS experiment, these studies led to the decision to make PUPPI the default algorithm recommended for pileup mitigation in top tagging algorithms. Moreover, top tagging working points have been derived for CHS- and PUPPI-corrected jets within the here presented studies, which entered the official recommendations at the CMS experiment for analyses using data collected at $\sqrt{s} = 13\,\text{TeV}$ in the year 2015 and

---

[1] These techniques are discussed and used in Chapter 8 and will not be detailed here.

2016. They are published with corresponding correction factors for the simulation in [6]. The chapter begins with a short introduction into the top tagging techniques used within the presented studies (Section 7.1). Afterwards, the simulated processes utilized for the investigation of the top tagging performance are detailed in Section 7.2, while the reconstruction of large-cone jets as well as the examined top tagging working points are presented in Section 7.3. In Section 7.4 the determination of the top tagging efficiencies and misidentification rates is described. Section 7.5 compares the discrimination power of the tagging observables as well as the top tagging efficiencies and misidentification rates of the studied algorithms for jets corrected with PUPPI- and CHS-corrected jets. The determination of new working points for the two pileup mitigation techniques is presented in Section 7.6. A short summary and outlook mark the end of the chapter (Section 7.7).

7.1 Top tagging techniques

In the studies presented here, top tagging algorithms are used that have been studied in [172]. The recommended algorithms represent a compromise of high top quark identification efficiency, low misidentification rates and simplicity of the algorithms. For high-p_T top tagging ($p_T \approx 500\,\mathrm{GeV}$) the recommendations are to select events based on the observable n-subjettiness and the large-cone jet mass corrected with the soft drop algorithm. Analyses using low-p_T top quarks ($p_T \approx 200\,\mathrm{GeV}$) are recommended to use the HTT V2 algorithm and n-subjettiness to identify hadronically decaying top quarks. The mentioned techniques are introduced shortly in this section, while the reconstruction of large-cone jets and the studied working points are detailed in Section 7.3.

7.1.1 The soft drop algorithm

The mass associated with large-cone jets is an important ingredient for most top tagging algorithms since it is expected to be small for jets emerging from light quarks or gluons, while it it is close to the top quark mass for jets comprising top quark decays. However, it can be distorted by pileup interactions, initial-state radiation or the underlying event if these are clustered into the jet. The larger the cone size parameter of the jet reconstruction algorithm is chosen, the larger the corrections can become. Therefore, so-called *grooming techniques* are applied to alleviate the contributions of such processes.

The top tagging algorithms used in the studies presented in this chapter exploit the soft drop algorithm [172,173], a generalization of the mass drop tagging algorithm [174], which removes low-p_T wide-angle radiation from the large-cone jet and thus enables a more efficient reconstruction of subjets and a reconstruction of the jet mass with better resolution. For the application of the soft drop algorithm, the Cambridge/Aachen clustering sequence of the large-cone jet is inverted (see Section 5.5.1 for a description of the jet clustering algorithms used at the CMS experiment). At each de-clustering step it is checked if the two examined clusters j_1 (with $p_{\mathrm{T},1}$) and j_2 (with $p_{\mathrm{T},2}$) fulfill the condition given by

$$\frac{\min(p_{\mathrm{T},1}, p_{\mathrm{T},2})}{p_{\mathrm{T},1} + p_{\mathrm{T},2}} > z \left(\frac{\Delta R_{12}}{R}\right)^{\beta}, \tag{7.1}$$

where ΔR_{12} describes the distance measured in η and ϕ of the two studied clusters and R denotes the distance parameter used to cluster the initial jet. The soft drop condition includes two free parameters, which are z and β. Throughout the studies in this thesis, values of 0.1 or 0.2 are used for z, while β is either 0 or 1. If the soft drop condition is met by the two clusters, j_1 and j_2 are classified as subjets and the jet built from j_1 and j_2 as the final groomed jet. If the condition given by Eq. (7.1) does not hold, the softer cluster is removed from the jet and the procedure is repeated with the remainder of the jet. While the fractional p_T requirement on the subjets can be tightened with a larger value of z, the selection criterion on the allowed radiation angle can be loosened by choosing β values above 0. For the latter case, the soft drop algorithm is infrared and collinear safe. By adding the four-momenta of the subjets identified by this method, which ideally correspond to the three decay products of the hadronically decaying top quark, the groomed mass of the large-cone jet, denoted as m_{SD}, is built.

7.1.2 The HEPTopTagger Version 2

In contrast to other top tagging techniques, the HEPTopTagger Version 2 (HTT V2) [172, 175–177] shows a high top quark identification capability already for low-p_T top quarks ($p_T \approx 200\,\mathrm{GeV}$) and is thus especially suited for low-p_T top tagging. However, stable performance of the HTT V2 algorithm has been observed over a wide range of top quark transverse momenta [172], which makes it an interesting approach for many BSM searches.

The HTT V2 algorithm is based on jets clustered with the Cambridge/Aachen algorithm using a distance parameter of $R = 1.5$. In order to identify the subjets associated to the three decay products of the hadronically decaying top quark, the clustering sequence of the large-cone jet is inverted. In each unclustering step, which breaks the jet j into two subjets j_1 and j_2, a so-called mass drop condition is imposed. The mass drop condition requires the heavier of the two subjets (with mass m_{sj}) to pass the condition $m_{\mathrm{sj}} < 0.8\,\mathrm{m}_j$, where m_j describes the mass of the jet before the unclustering step. This criterion implies that both subjets carry a certain minimal momentum fraction of the initial jet. If the mass drop condition does not hold, the subjet with the smaller mass is removed from the jet, and the de-clustering procedure is continued with the heavier subjet. If the subjets pass the mass drop condition, the decomposition of the jet is continued with both subjets. The procedure stops when the mass of all remaining subjets lies below 30 GeV. Only for jets that contain at least three of these subjets the HTT V2 algorithm continues. Subsequently, a *filtering* procedure [178] is performed, which seeks the removal of contributions from the underlying event and pileup to the large-cone jet. Filtering re-clusters the particles included in the three highest-p_T subjets with a smaller cone size. In the HTT V2 algorithm this cone size is set to the minimum of 0.3 and half of the distances between the three considered subjets. The constituents of the five subjets with the highest p_T remaining after the filtering procedure are passed to another clustering step, which builds again exactly three subjets out of the initial five. Only if the transverse momenta of these exceed 30 GeV the large-cone jet is kept. From the three subjets of the large-cone jet the mass m_{HTT}, which is used as discriminating variable in top tagging with the HTT V2 algorithm, is derived by adding the four-momenta of the subjets.

To improve the top tagging performance, the described procedure of building subjets and the large-cone jet mass m_{HTT} is repeated multiple times with successively smaller cone sizes of the initial large-cone jet ranging from 1.5 to 0.5. For the final reconstruction the minimal value of all tested cone sizes R is chosen, for which the final jet mass m_{HTT} is not reduced by more than 20% from the initial mass obtained for $R = 1.5$. This approach accommodates the fact that top quarks with larger transverse momenta are more collimated. All observables of the top tagging algorithm are evaluated at this minimal cone size R, denoted as R_{opt} in the following. Furthermore, the observable ΔR_{HTT} is defined by

$$\Delta R_{\mathrm{HTT}} = R_{\mathrm{opt}} - R_{\mathrm{opt}}^{\mathrm{calc}}, \tag{7.2}$$

where $R_{\mathrm{opt}}^{\mathrm{calc}}$ is the minimal cone size determined on simulated events. It depends on the transverse momentum of the jet.

A further discriminating variable used for top tagging with the HTT V2 algorithm is f_{REC}, defined by

$$f_{\mathrm{REC}} = \min_{ij} \left| \frac{\frac{m_{ij}}{m_{\mathrm{HTT}}}}{\frac{m_W}{m_t}} - 1 \right|, \tag{7.3}$$

where the sum runs over all pairs of subjets i and j. This variable serves as a measure of the compatibility of the ratio of the minimal invariant mass of two subjets and the large-cone jet mass with the ratio of the W boson mass m_W and the top quark mass m_t.

7.1.3 N-subjettiness

A further important quantity for top tagging is τ_i, referred to as n-subjettiness [172, 179,180], which is a measure for the compatibility with the hypothesis that a large-cone jet contains i subjets. It is infrared and collinear safe for $\tau_i > 0$. Values of τ_i close to 0 describe a high compatibility with the i-subjet hypothesis, while values close to 1 indicate that the large-cone jet is more likely composed of less than i subjets. The n-subjettiness τ_i for a jet clustered with the distance parameter R that consists of k particles is given by

$$\tau_i = \frac{1}{\sum_k p_{\mathrm{T},k} R} \sum_k p_{\mathrm{T},k} \min(\Delta R_{1k}, \Delta R_{2k}, ..., \Delta R_{ik}). \tag{7.4}$$

In this formula, $\Delta R_{jk} = \sqrt{(\phi_j - \phi_k)^2 + (\eta_j - \eta_k)^2}$ describes the distance between the axis of the candidate subjet j and the particle k. The reader is referred to [180] for a description of the axis determination. Since large-cone jets that comprise an hadronically decaying top quark are expected to contain three subjets, while jets arising from gluons or light quarks most likely are reconstructed with less subjets, the observable $\tau_{32} = \frac{\tau_3}{\tau_2}$

provides good discrimination power between signal jets and background processes and is used to enhance the sensitivity of top tagging algorithms.

7.1.4 Subjet b tagging

Since top quarks decay almost exclusively into a b quark and a W boson, the performance of top tagging algorithms can be significantly improved by including b tagging techniques in the top tagging definition. For the studies presented in this chapter, the CSVv2 algorithm (see Section 5.5.5) is applied on the reconstructed subjets of the large-cone jets. To decide whether the large-cone jet passes the b tagging requirement, the subjet with the highest CSVv2 discriminant is required to exceed the threshold value of the applied b tagging working point.

7.2 Signal and background processes

The studies presented in this chapter are based on simulated events only. Different processes are used to study top tagging efficiencies. On the one hand, signal processes of Z' bosons with masses of 1 TeV or 2 TeV and 10% width are studied. The modeled Z' bosons decay exclusively into a pair of top quarks and their couplings to the SM fermions are set to those of the SM Z boson. The samples are generated with up to three extra partons added to the matrix element process using MadGraph (v5.2) [94]. pythia (v8) [101,102] is used for the simulation of the shower and hadronization process, the matching of the shower particles to the hard process is done by applying the MLM matching [108] scheme. In addition to the Z' processes, SM $t\bar{t}+$jets production divided into two regions of the invariant mass of the top quark pair ($700\,\mathrm{GeV} < m_{t\bar{t}} < 1000\,\mathrm{GeV}$ and $m_{t\bar{t}} > 1000\,\mathrm{GeV}$) is used for the determination of top tagging efficiencies. The processes are generated at NLO using powheg [95, 98–100], while the parton shower and hadronization process is again simulated using pythia (v8).
To estimate the discrimination power of different top tagging observables between signal and background events as well as for the determination of the misidentification rates of the studied algorithms different QCD multijet samples are examined. A LO QCD multijet sample is generated using pythia (v6) [101] and uniformly distributed in $\hat{p}_{\mathrm{T}}$[2] ($15\,\mathrm{GeV} < \hat{p}_{\mathrm{T}} < 7000\,\mathrm{GeV}$) to ensure a constant statistical precision over the whole generated phase space. Additionally, QCD multijet processes simulated at LO with pythia (v8) in two different regions of $\hat{p}_{\mathrm{T}}$ ($300\,\mathrm{GeV} < \hat{p}_{\mathrm{T}} < 470\,\mathrm{GeV}$ and $800\,\mathrm{GeV} < \hat{p}_{\mathrm{T}} < 100\,\mathrm{GeV}$) are used to investigate the top tagging performance in low- and high-p_T regions with higher statistical precision.
For all simulated processes the NNPDF3.0 [90] pdf sets are used and the interactions of the particles with the detector material is simulated using Geant4 [111]. All samples are generated with additional pileup interactions included, assuming the pileup conditions expected for the LHC running with a time spacing of 25 ns between two subsequent

---

[2] The variable $\hat{p}_{\mathrm{T}}$ is defined as the momentum transfer in the simulated dijet processes.

proton bunches.

7.3 Reconstruction of large-cone jets and studied top tagging working points

The impact of CHS and PUPPI on the performance of the top tagging techniques described in Section 7.1 is tested by using different collections of large-cone jets. For the high-p_T top tagging studied in this chapter, jets clustered with the anti-k_T algorithm out of PF particles with a distance parameter of $R = 0.8$ are used as input. The particles included in the anti-k_T jets are afterwards re-clustered with the Cambridge/Aachen algorithm. The latter reconstruction step is necessary to make the clustering history invertible and thus the soft drop algorithm applicable. The parameter of the soft drop algorithm are chosen as $z = 0.1$ and $\beta = 0$. Furthermore, the observable τ_{32} is derived using the particles contained in the jet before the application of the soft drop algorithm. Jets reconstructed in the described way will be referred to as *Ak8-jets* in the following.

For low-p_T top tagging, jets with larger cone sizes are needed to be able to capture the hadronic top quark decay completely inside the large-cone jet. Hence, jets reconstructed with the Cambridge/Aachen algorithm adopting a distance parameter of $R = 1.5$ are studied. These are passed to the HTT V2 algorithm. The resulting jets are denoted as *HTT-jets*.

These two jet collections corrected with CHS and with PUPPI are studied. Jet energy corrections derived for jets using the respective pileup mitigation technique are applied. Since no jet energy corrections are available for jets clustered with the Cambridge/Aachen algorithm and a cone size of $R = 1.5$, the corrections derived for jets built with the anti-k_T algorithm and a cone size of $R = 0.8$ are applied on all jet collections. The corrected transverse momenta of the studied jets have to exceed 150 GeV and the jets have to lie within $|\eta| < 2.4$.

The presented studies are performed for medium top tagging working points pre-defined by the CMS experiment [181]. Other provided working points have been tested as well. Since the conclusions drawn for these do no differ from what is presented here, the results for other working points are not discussed.

The reconstructed Ak8-jets have to pass the following requirements to be top tagged:

- The soft drop mass m_{SD} lies within $110\,\mathrm{GeV} < m_{\mathrm{SD}} < 210\,\mathrm{GeV}$ and is thus compatible with the top quark mass.
- An n-subjettiness ratio requirement of $\tau_{32} = \frac{\tau_3}{\tau_2} < 0.69$ is imposed on all studied jets.

These requirements will be referred to as the *soft drop tagger* in the following.

To be identified as a top quark candidate jet by the HTT V2 algorithm the HTT-jets have to fulfill the criteria hereafter:

- The mass of the large-cone jet, which is derived as described in Section 7.1.2, passes $110\,\mathrm{GeV} < m_{\mathrm{HTT}} < 300\,\mathrm{GeV}$.

- The observable f_{REC} is smaller than 0.35.
- The difference ΔR_{HTT} between the optimal R_{opt} and the calculated $R_{\mathrm{opt}}^{\mathrm{calc}}$ lies within $-0.78 < \Delta R_{\mathrm{HTT}} < 0.26$.
- The n-subjettiness ratio τ_{32} is smaller than 0.97.

7.4 Determination of top tagging efficiencies and misidentification rates

In order to study top tagging efficiencies as well as misidentification rates, a matching between the large-cone jets and the generator-level particles is performed. For the matching of large-cone jets to generator-level top quarks only hadronically decaying generator-level top quarks with $p_T > 150\,\mathrm{GeV}$ are considered, while the association of a large-cone jet to quarks or gluons the two p_T leading light quarks or gluons on generator level that pass $p_T > 150\,\mathrm{GeV}$ are studied. The closest large-cone jet is matched with a maximal distance of $\Delta R < 1.0$. In all following studies, only matched Ak8-jets or HTT-jets are considered.

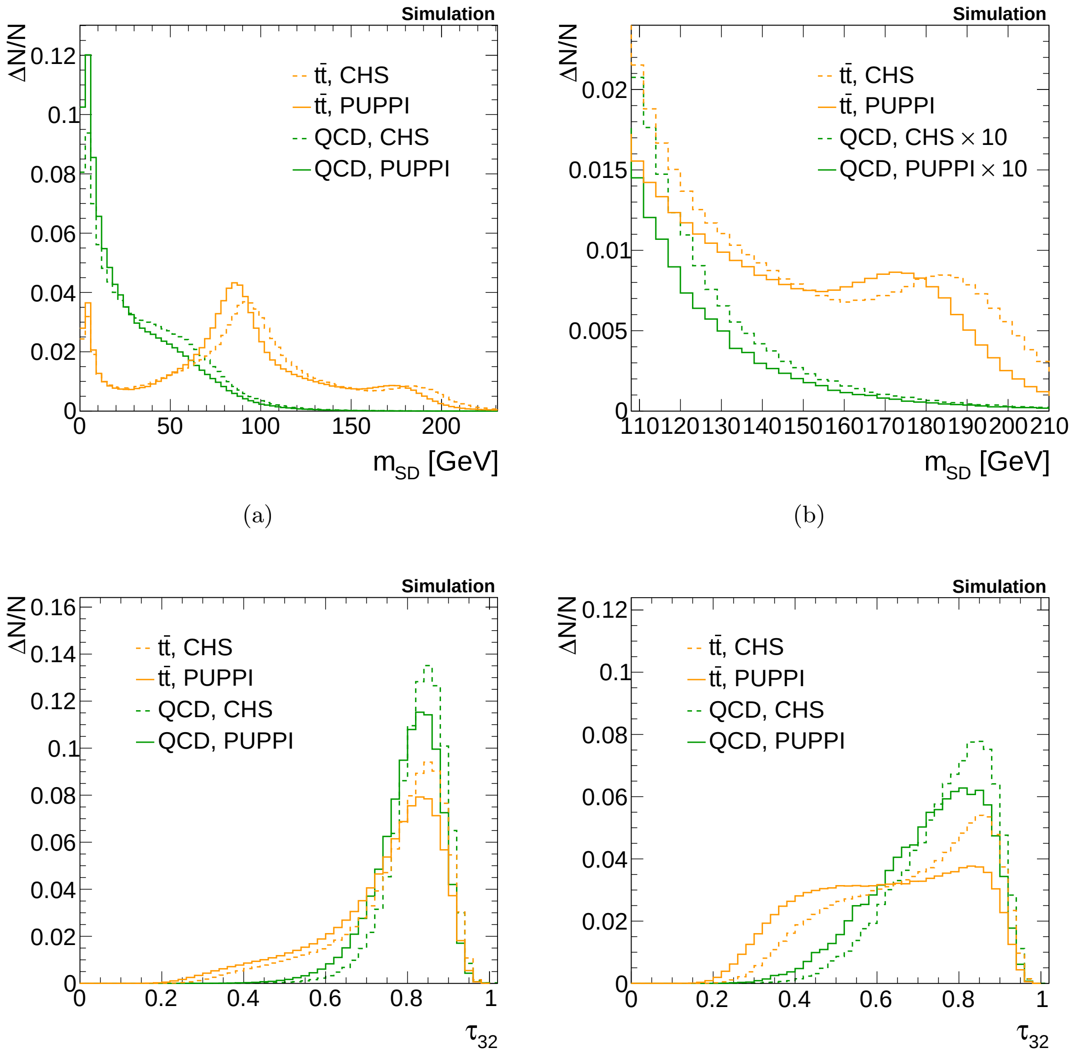
The efficiency ϵ and the misidentification rate f of the top tagging algorithm is derived by counting the number of the respective matched jets that pass the examined top tagging criteria with respect to all large-cone jets matched to a top quark or a light quark or gluon, respectively.

7.5 Performance studies

In this section, the shapes of top tagging observables corrected with CHS are compared to PUPPI-corrected ones in background and signal processes. The $t\bar{t}$+jets and the inclusive QCD samples described in Section 7.2 are analyzed. Furthermore, the top tagging efficiencies and misidentification rates of the working points detailed in Section 7.3 are compared after the application of the two mentioned pileup mitigation techniques.

7.5.1 High-p_T top tagging: soft drop tagger based on Ak8-jets

Figure 7.1 compares the normalized distributions of the soft drop mass m_{SD} and τ_{32} obtained from Ak8-jets corrected with PUPPI and corrected with CHS in signal and background processes. The suppression of the peak at the top quark mass compared to the peak at the W boson mass in signal processes (see Fig. 7.1(a)) is mainly caused by the dominant contributions of top quarks with transverse momenta close to 150 GeV, which have too little boost to be clustered in one jet of cone size $R = 0.8$. In the mass window chosen for the studied top tagging working point ($110\,\mathrm{GeV} < m_{\mathrm{SD}} < 210\,\mathrm{GeV}$) a falling spectrum of the soft drop mass is observed in the background processes (see Fig. 7.1(b)). In the mentioned top tagging mass window, PUPPI shifts the background and signal distributions to lower values of the masses. Figure 7.1(c) shows a comparison of τ_{32} in background and signal processes for Ak8-jets corrected with PUPPI and



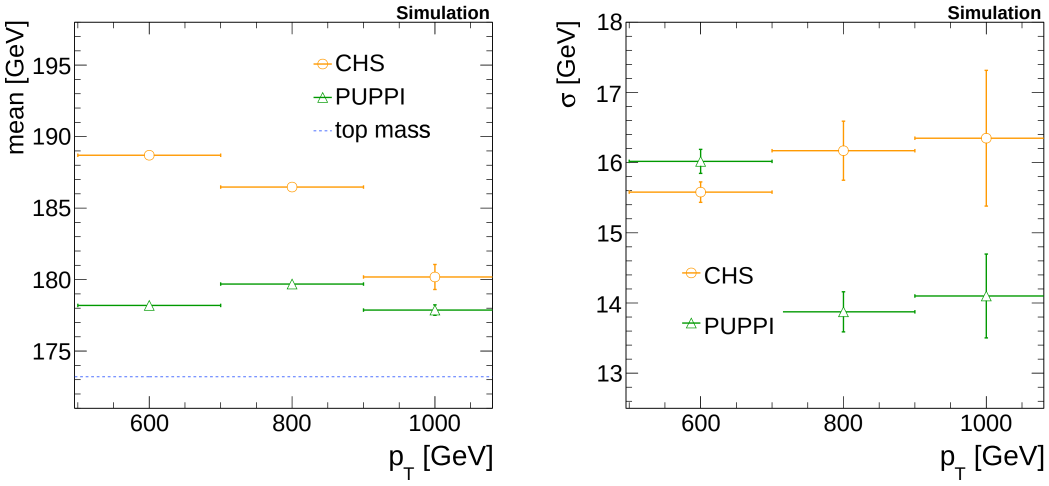
(a) (b)

(c) (d)

Figure 7.1: Comparison of the normalized distributions of (a) m_{SD} , (b) m_{SD} in the range $110\,\mathrm{GeV} < m_{\mathrm{SD}} < 210\,\mathrm{GeV}$ and (c) τ_{32} obtained from the Ak8-jet collection corrected with PUPPI (*solid lines*) and corrected with CHS (*dashed lines*) in signal (*colored in orange*) and background (*colored in green*) processes. In (b) the background processes are scaled with a factor of 10 and the normalization of (a) is taken. (d) shows the τ_{32} distribution after the application of the top quark mass constraint used for this study.

CHS. Both the signal and the background distributions peak at very similar positions. The only difference between the background and signal distributions lies in the longer tails of the signal at smaller values. This behavior of τ_{32} is expected since most of the top quark decays in the studied phase space are not yet fully merged into the Ak8-jet. Therefore, most of the studied Ak8-jets contain only two subjets and the corresponding n-subjettiness ratio value τ_{32} gets large. The discrimination power of τ_{32} is enhanced after imposing the top quark mass constraint used for this study (see Fig. 7.1(d)). The application of PUPPI shifts both the signal and the background distributions to smaller values of τ_{32}.

Figure 7.2 shows that a more stable mass peak position as a function of the p_T of the



(a) (b)

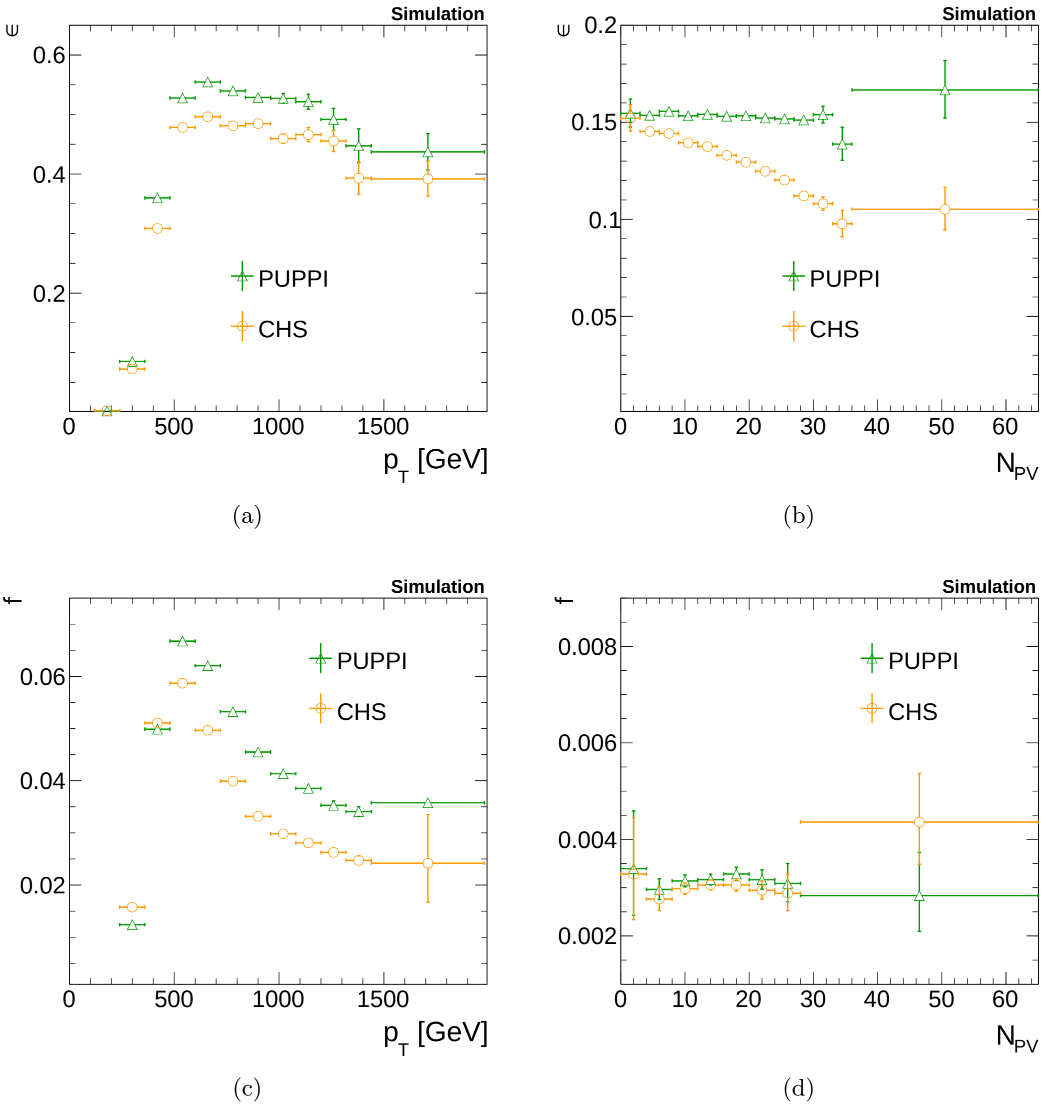
Figure 7.2: Comparison of the (a) mean and (b) standard deviation of a Gaussian fitted to the distribution of the soft drop mass obtained for different p_T ranges (starting with $p_T > 500\,\text{GeV}$) for PUPPI- (*green triangles*) and CHS- (*orange circles*) corrected Ak8-jets. In (a) the top quark mass (*dashed blue line*) is shown for illustration.

Ak8-jets as well as a smaller width of the mass distribution, is obtained by using PUPPI[3]. Additionally, the peak position is much closer to the mass of the top quark after running the PUPPI algorithm than the mass peak position obtained with CHS. A comparison of the top tagging efficiencies ϵ and misidentification rates f as a function of the p_T of the Ak8-jets and the number of primary vertices reconstructed in the events is presented in Fig. 7.3. For both PUPPI- and CHS-corrected jets, a turn-on behavior of the efficiency as a function of p_T is observed for Ak8-jets with transverse momenta between 150 GeV and 500 GeV due to not fully merged top quark decay products in the mentioned phase space. For both pileup mitigation techniques a decrease in the efficiency at high p_T is observed. The top tagging efficiency obtained with PUPPI is higher than the one obtained with CHS over the whole considered p_T range. This is also true over a wide range of transverse momenta for the misidentification rate f depicted in dependence of the p_T of the Ak8-jet in Fig. 7.3(c). Both misidentification rate distributions exhibit a peak at transverse momenta of approximately 500 GeV and decrease for larger p_T. For transverse momenta below 500 GeV the misidentification rate derived for PUPPI-corrected jets is smaller than for CHS-corrected jets.

Most importantly, Figs. 7.3(b) and 7.3(d) present the top tagging efficiency and the misidentification rate as a function of the number of primary vertices reconstructed in the events for PUPPI- and CHS-corrected events. For both pileup mitigation techniques the misidentification rate is relatively constant. Contrary, the top tagging efficiency determined for CHS-corrected events decreases rapidly with increasing number of primary vertices, while the efficiency derived for PUPPI-corrected events is stable. Due to the high level of pileup activity reached in LHC Run-II, a stable performance of the top

---

[3] The mass distribution is fitted for Ak8-jets in the p_T regions $500\,\text{GeV} < p_T < 700\,\text{GeV}$, $700\,\text{GeV} < p_T < 900\,\text{GeV}$ and $p_T > 900\,\text{GeV}$.



(a) (b)

(c) (d)

Figure 7.3: Comparison of (a) the soft drop tagger efficiency as a function of p_T of the Ak8-jets, (b) the soft drop tagger efficiency in dependence of the number of primary vertices in the event, (c) the misidentification rate of the soft drop tagger in dependence of p_T of the Ak8-jets and (d) the misidentification rate of the soft drop tagger as a function of the number of primary vertices between PUPPI (*green triangles*) and CHS (*orange circles*) corrected jets. The working point presented in Section 7.3 is studied.

tagging efficiencies and misidentification rates over a wide range of pileup conditions is of great importance. Thus, the more stable top tagging performance with increasing number of pileup vertices, resulted in the decision to use PUPPI-corrected jets instead of CHS-corrected jets for top tagging in the future.

7.5.2 Low-p_T top tagging: The HTT V2 algorithm

Figure 7.4 presents a comparison of the observables used to identify hadronically decaying top quarks with the HTT V2 algorithm obtained from HTT-jets corrected with

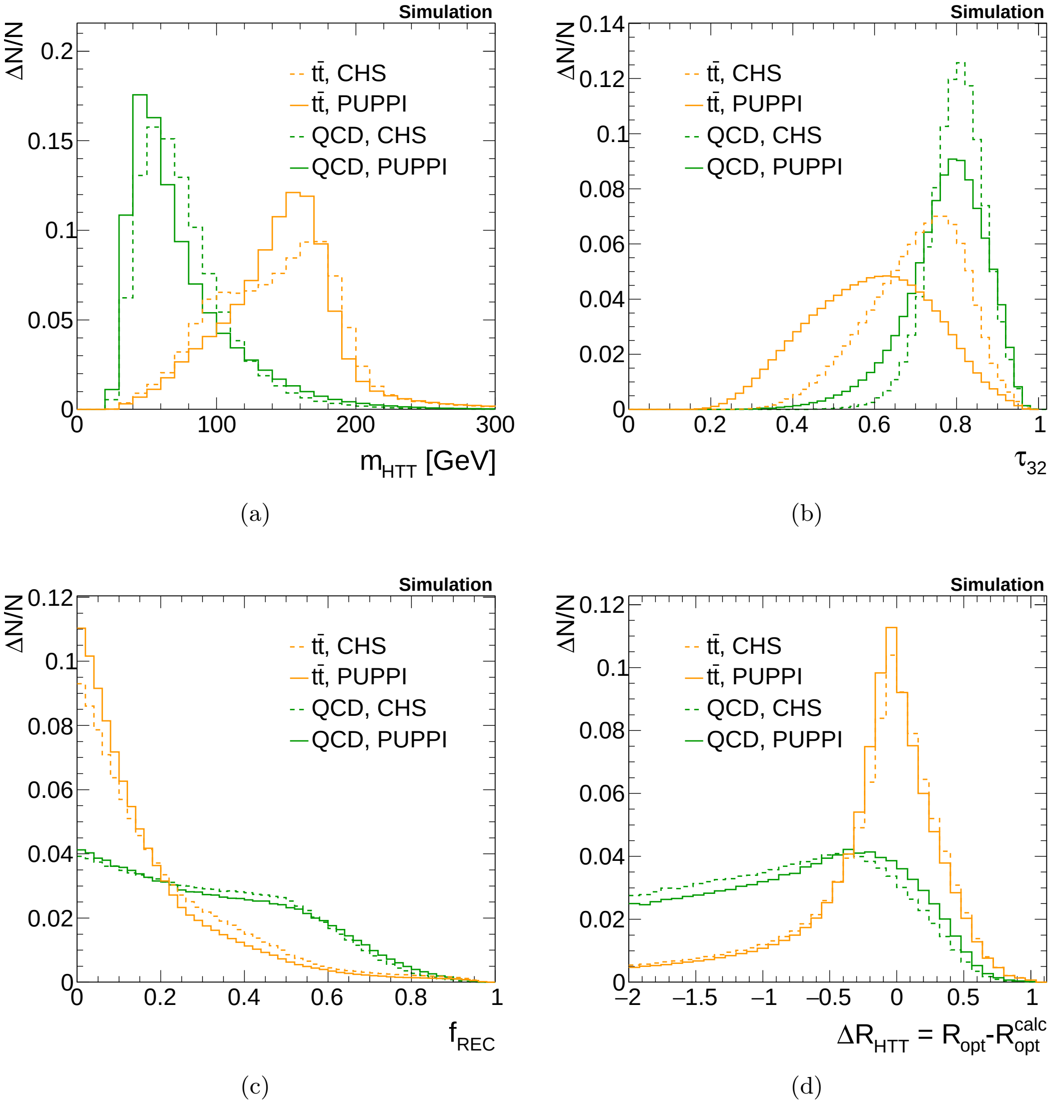
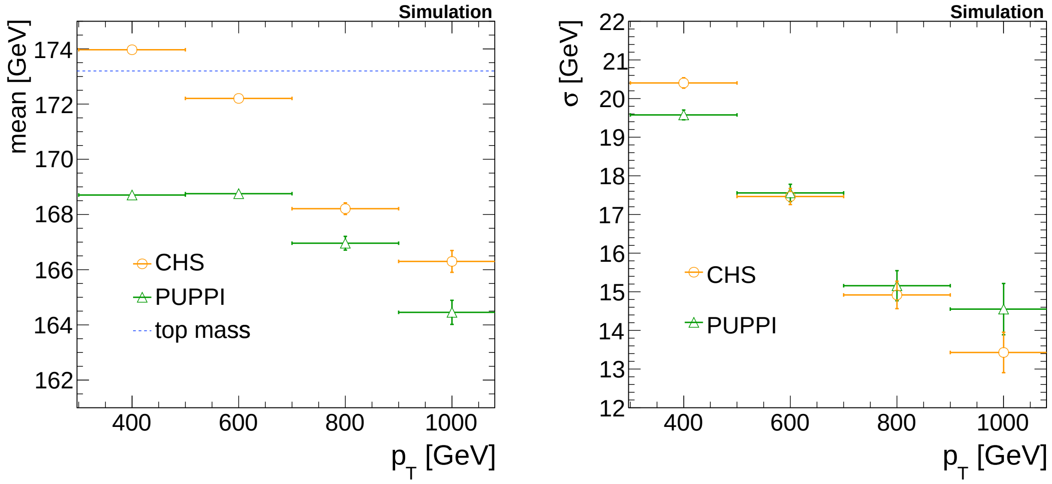


Figure 7.4: Comparison of the normalized distributions of (a) m_{HTT}, (b) τ_{32}, (c) f_{REC} and (d) ΔR_{HTT} determined from the HTT-jet collection corrected with PUPPI (*solid lines*) to the ones corrected with CHS (*dashed lines*) in the signal (*colored in orange*) and background (*colored in green*) processes.

PUPPI and CHS. The depicted observables are m_{HTT}, τ_{32}, f_{REC} and ΔR_{HTT}.
Since the HTT V2 algorithm rejects top quark candidates that are not completely contained inside the HTT-jet, no peak in m_{HTT} is observed around the mass of the W boson. Inside the mass window used for top tagging ($110\,\mathrm{GeV} < m_{\mathrm{HTT}} < 300\,\mathrm{GeV}$), a falling background distribution is observed. Within this mass window, applying the PUPPI algorithm instead of CHS shifts the background mass distribution to larger values. Contrary, the signal is moved to smaller values. While the CHS distribution peaks around the top quark mass, the maximal value of the PUPPI distribution lies at a value slightly below the top quark mass. This observation is confirmed by studies of the stability of the mass peak as a function of the transverse momenta of the HTT-jets illustrated



(a) (b)

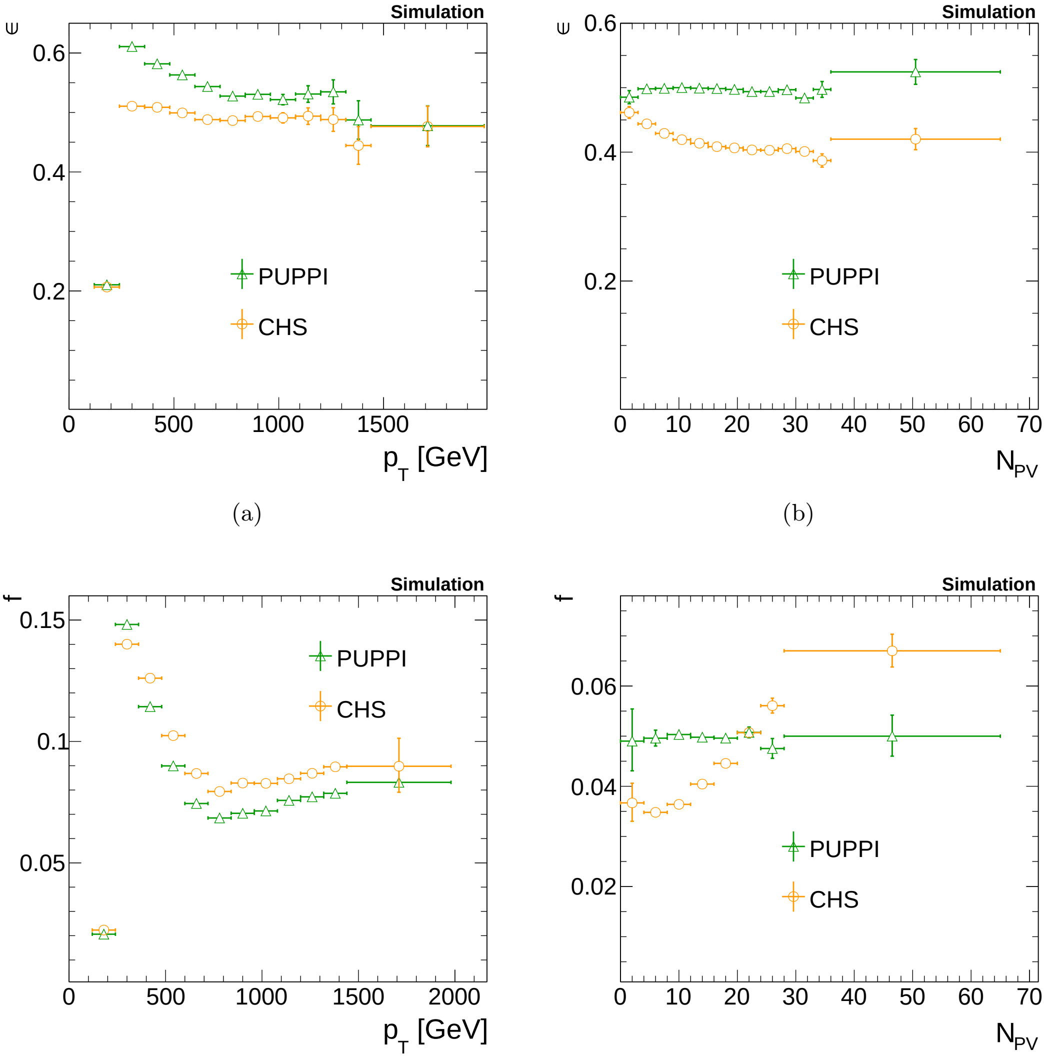
Figure 7.5: Comparison of the (a) mean and (b) standard deviation of a Gaussian fitted to the distribution of the mass obtained by applying the HTT V2 algorithm on HTT-jets split into different p_T ranges (starting with $p_T > 300\,\mathrm{GeV}$) for PUPPI (*green triangles*) and CHS (*orange circles*) corrected jets. In (a) the top quark mass (*dashed blue line*) is shown for illustration.

in Fig. 7.5[4]. Similar to the findings for Ak8-jets, also for HTT-jets the mass position is more stable after the application of PUPPI. However, the peak position is below the top quark mass over the whole p_T range, while with CHS a peak closer to the top quark mass is achieved. The widths of the peaks decrease with p_T and are comparable between the two approaches. Figure 7.4(b) depicts τ_{32} in background and signal processes after the application of CHS and PUPPI. A clear separation between signal and background events is achieved for HTT-jets. For PUPPI-corrected events, the separation power of τ_{32} is significantly higher than the one obtained for CHS-corrected jets. The pileup mitigation technique has only a small effect on the distribution of f_{REC} and ΔR_{HTT}.

The top tagging efficiencies and misidentification rates as a function of the transverse momenta of the HTT-jets and the number of primary vertices reconstructed in the events are illustrated in Fig. 7.6. In comparison to the soft drop tagger, the turn-on curve of the efficiency of the HTT V2 algorithm is much sharper in dependence of p_T. The HTT V2 algorithm reaches its plateau efficiency already for transverse momenta slightly above 200 GeV. If CHS is applied, the top tagging efficiency is also stable over a wide range of transverse momenta. With PUPPI a higher efficiency is reached, which is decreasing with increasing transverse momentum of the jet. Similar as for the soft drop tagger, the misidentification rate reaches a maximal rate for HTT-jets around transverse momenta of 300 GeV. In this p_T range applying PUPPI yields a larger misidentification rate than the application of CHS. In the remaining phase space, the misidentification rates behave opposite. Similar as for Ak8-jets, also here more stable results of the top tagging efficiency in dependence of the number of primary vertices are obtained for PUPPI than for CHS. For the HTT-jets, this is also true for the misidentification rate, while by apply-

---

[4] The mass distribution is fitted for HTT-jets in the p_T regions $300\,\mathrm{GeV} < p_T < 500\,\mathrm{GeV}$, $500\,\mathrm{GeV} < p_T < 700\,\mathrm{GeV}$, $700\,\mathrm{GeV} < p_T < 900\,\mathrm{GeV}$ and $p_T > 900\,\mathrm{GeV}$.



(a) (b)

(c) (d)

Figure 7.6: Comparison of (a) the HTT V2 algorithm efficiency as a function of p_T of the HTT-jets, (b) the HTT V2 algorithm efficiency in dependence of the number of primary vertices in the event, (c) the misidentification rate of the HTT V2 algorithm in dependence of p_T of the HTT-jets and (d) the misidentification rate of the HTT V2 algorithm as a function of the number of primary vertices between PUPPI- (*green triangles*) and CHS -(*orange circles*) corrected jets. The working point presented in Section 7.3 is studied.

ing CHS the efficiency decreases and the misidentification rate increases with increasing number of pileup vertices.

In conclusion, it could be shown that higher top tagging efficiencies can be achieved by applying PUPPI. Unfortunately at the same time, the misidentification rate increases in most of the phase space. Thus, the top tagging working points used for these studies, which were derived for CHS, are not optimal for PUPPI and the performances of the two pileup mitigation techniques are not directly comparable. A determination of new working points is needed for a fair comparison of the absolute values obtained with the two approaches. The determination of working points is presented in the next section.

Still, the trends observed by applying the two pileup mitigation techniques can be compared and, since a much more stable behavior has been observed for PUPPI-corrected events with increasing number of pileup vertices, PUPPI will become the default and recommended pileup mitigation technique for top tagging in the CMS experiment.

7.6 Determination of top tagging working points

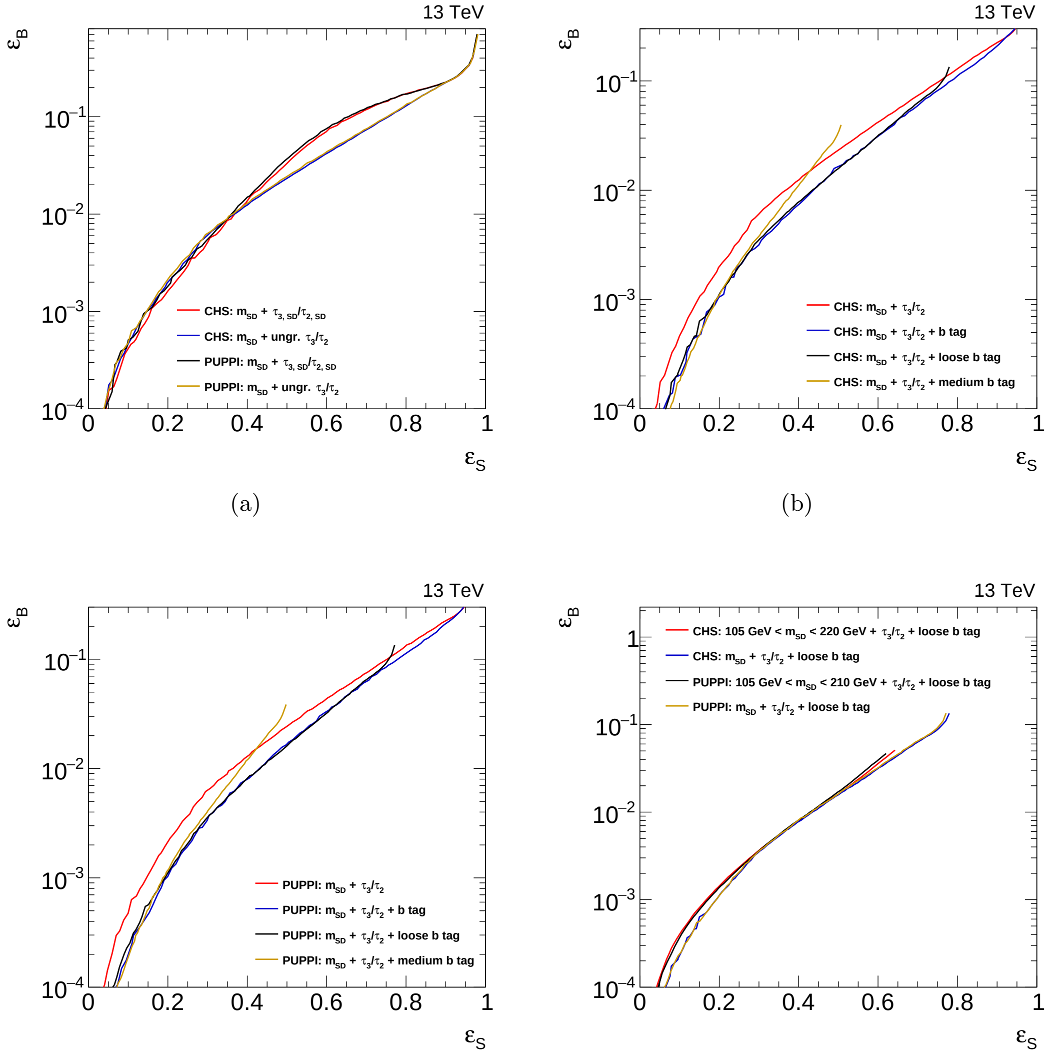
Working points for CHS- and PUPPI-corrected jets are derived from the evaluation of so-called *Receiver Operating Characteristics (ROC) curves* determined with the TMVA package [182]. A ROC curve displays for a given signal selection efficiency the minimal background selection efficiency that is reachable with requirements imposed on a set of input variables. The determination of the working points for high-pt and low-p_T top tagging will be discussed in the following.

7.6.1 High-p_T top tagging: soft drop tagger based on Ak8-jets

For the determination of the working points for high-p_T top tagging, Z' boson (with a mass of 2 TeV) and QCD multijet events (with $\hat{p}_{\mathrm{T}}$ between 800 GeV and 1 TeV) are used as signal and background processes, respectively. The transverse momenta of the studied Ak8-jets have to exceed 500 GeV. The Ak8-jets that lie closest within $\Delta R < 0.6$ to the generator-level top quark (generator-level gluon or light quark) and that passes $800\,\mathrm{GeV} < p_T < 1\,\mathrm{TeV}$ and $|\eta| < 1.5$ are studied for the determination of the signal selection efficiency (background selection efficiency). Based on the selected jets, ROC curves can be derived.

Figure 7.7 illustrates ROC curves for different set of input variables. In addition to using n-subjettiness calculated before the application of grooming techniques, n-subjettiness determined after the application of grooming can also be studied. The latter will be referred to as groomed n-subjettiness ratio τ_{32} or $\tau_{32,\mathrm{SD}}$ in the following. However, Fig. 7.7(a) shows that requirements on the groomed n-subjettiness ratio τ_{32} in conjunction with selections on the soft drop mass m_{SD} yield worse top tagging performances than applying the ungroomed n-subjettiness ratio instead, while no difference between applying PUPPI and CHS is observed. Figure 7.7(b) and Fig. 7.7(c) show that a significant improvement in the top tagging performance is achieved by applying subjet b tagging. Furthermore, the impact of using a predefined requirement on the subjet b tag discriminator and of letting TMVA determine the b tag requirement is presented. For high background suppression, adopting a loose subjet b tag requirement [138] in addition to criteria on m_{SD} and τ_{32} yields similar results as using the medium subjet b tag working point. However, high signal efficiencies can only be achieved by using the loose working point. Letting TMVA determine the selection requirement on the subjet b tag discriminator instead of using a predefined b tagging working point yields no gain in performance.

Based on these studies, the working points have been derived by applying requirements on m_{SD} and τ_{32} in addition to a loose subjet b tag requirement for PUPPI- and CHS-corrected jets. The working point is chosen such that it corresponds to a background



(a) (b)

(c) (d)

Figure 7.7: ROC curves for high-p_T top tagging with the following observables used as input: (a) m_{SD}, $\tau_{32,\mathrm{SD}}$, and ungroomed τ_{32} for CHS- and PUPPI-corrected jets, (b) m_{SD}, ungroomed τ_{32} and a floating subjet b tag discriminator for CHS-corrected jets. Additionally, the ROC curves obtained with the application of a fixed medium or loose b tag requirement instead of a floating one are shown. (c) the same as (b) for PUPPI-corrected jets. A comparison of the ROC curves obtained for ungroomed τ_{32} and a loose subjet b tag as well as a fixed mass window or a floating mass window for CHS- and PUPPI-corrected jets is shown in (d).

selection efficiency of 3%. The associated mass window of that working point overlaps with the window used for W boson tagging in the CMS experiment [6]. Therefore, the lower mass bound has been raised to exclude the overlapping masses. Afterwards, the mass selection has been fixed, and new ROC curves have been derived with τ_{32} as the only input variable. Finally, working points have been selected that correspond to background selection efficiencies of 0.1%, 0.3%, 1% and 3%. The resulting working points are summarized in Table 7.1. Additionally, it could be shown that removing the overlap with the W boson tagging mass window only slightly worsens the top tagging

	$\epsilon(B)$ [%]	$\epsilon(S)$ [%]	softdrop mass[GeV]	τ_{32}	subjet b tag
CHS	0.1	16.5	[105,220]	< 0.5	> 0.46
	0.3	27.4		< 0.57	
	1.0	43.2		< 0.67	
	3.0	58.1		< 0.81	
PUPPI	0.1	17.3	[105,210]	< 0.46	> 0.46
	0.3	28.1		< 0.54	
	1.0	42.8		< 0.65	
	3.0	56.6		< 0.8	

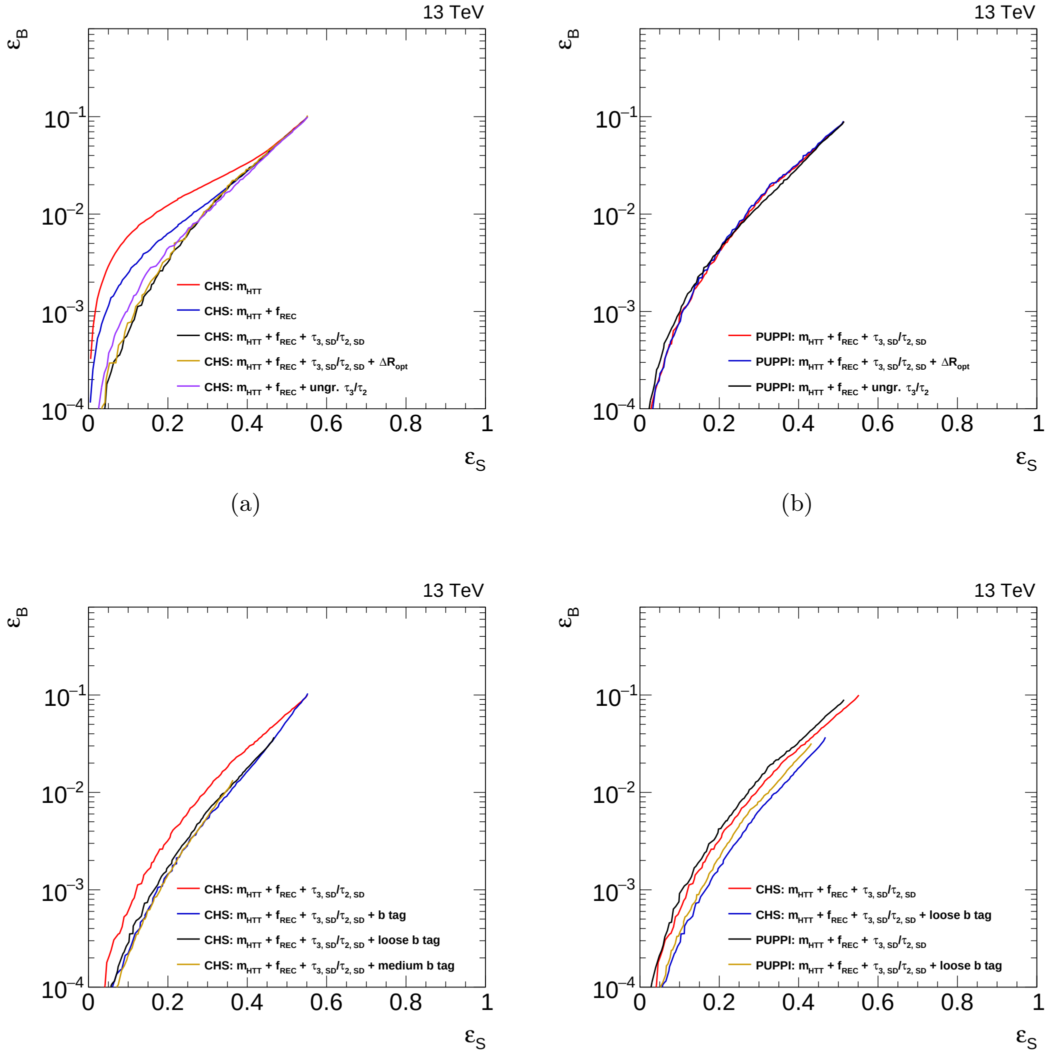
Table 7.1: Summary of the working points derived for the high-p_T top tagging based on Ak8-jets. The selection efficiencies of background and signal events are described by $\epsilon(B)$ and $\epsilon(S)$, respectively.

performance (see Fig. 7.7(d)). Since by the evaluation of the ROC curves selections optimized individually for PUPPI- and CHS-corrected jets are studied, a fair comparison of the performance of the two algorithms is now possible. The ROC curves displayed in Fig. 7.7(d) show that the application of PUPPI and CHS yield very similar performances if studied inclusively.

7.6.2 Low-p_T top tagging: The HTT V2 algorithm

For the determination of low-p_T top tagging working points the Z' boson (with a mass of 1 TeV) events represent the signal processes, while QCD multijet events with $\hat{p}_\mathrm{T}$ between 300 GeV and 470 GeV constitute the background events. All HTT-jets have to fulfill $p_T > 150$ GeV. The matching to a generator-level top quark, light quark or gluon is performed by requiring that the distance of the HTT-jet to the generator-level particle does not exceed $\Delta R = 1.2$. All generator-level particles considered for the matching have to have $|\eta| < 2.4$ as well as transverse momenta between 300 GeV and 470 GeV. In cases where more than one HTT-jets can be matched to a generator-level particle, the closest HTT-jets in ΔR to the generator-level particle is chosen.

ROC curves for different sets of input variables are presented in Fig. 7.8. Adding the groomed n-subjettiness ratio $\tau_{32,\mathrm{SD}}$ ($z = 0.2$ and $\beta = 1$) and f_REC in addition to m_HTT to the top tagging observables significantly improves the performance of the HTT V2 algorithm for CHS-corrected jets (see Section 7.8(a)). Contrary, the ungroomed n-subjettiness ratio deteriorates the top tagging performance, while no gain in performance is achieved by adding a requirement on ΔR_HTT, which is the least discriminating variable of all observables studied for the HTT V2 algorithm (see Fig. 7.4). For PUPPI-corrected jets no difference between the usage of groomed and ungroomed τ_{32} is observed (see Fig. 7.8(b)). Apart from that, the same conclusions as for the application of CHS are valid for PUPPI. Also for low-p_T top tagging the best results are achieved for a loose or a floating subjet b tag requirement (see Fig. 7.8(c)). Hence, for low-p_T top tagging a loose subjet b tag criterion is imposed on the HTT-jets in addition to requirements



(a) (b)

(c) (d)

Figure 7.8: ROC curves for low-p_T top tagging with the following observables used as input: (a) m_{HTT}, $\tau_{32,\mathrm{SD}}$, f_{REC}, ΔR_{HTT} and ungroomed τ_{32} for CHS-corrected jets, (b) m_{HTT}, $\tau_{32,\mathrm{SD}}$, f_{REC}, ΔR_{HTT} and ungroomed τ_{32} for PUPPI-corrected jets, (c) m_{HTT}, $\tau_{32,\mathrm{SD}}$ and a floating subjet b tag discriminator. Additionally, the ROC curves obtained for the application of a fixed medium or loose b tag requirement instead of a floating one are shown for CHS-corrected jets. A comparison of the ROC curves using m_{HTT}, f_{REC} and $\tau_{32,\mathrm{SD}}$ as input with and without additional requirement on a loose subjet b tag for PUPPI- and CHS-corrected jets is shown in (d).

on m_{HTT}, f_{REC} and $\tau_{32,\mathrm{SD}}$ to determine the working points. The working points are again derived by first extracting the mass window associated to a background selection efficiency of 3% from the corresponding ROC curves. Afterwards, selection requirements on f_{REC} and $\tau_{32,\mathrm{SD}}$ are derived by studying ROC curves obtained for the corresponding fixed mass window selection and a loose subjet b tag requirement. The resulting working points are listed in Table 7.2. A comparison of the top tagging performance obtained for PUPPI- and CHS-corrected jets is shown in Fig. 7.8(d). In this case, using PUPPI leads to a slightly worse performance than using CHS. However, since a much more sta-

	$\epsilon(B)$ [%]	$\epsilon(S)$ [%]	HTT mass [GeV]	f_{REC}	$\tau_{32,\mathrm{SD}}$	subjet b tag
CHS	0.1	16.6	[130,185]	< 0.17	< 0.55	> 0.46
	0.3	23.9	[115,180]	< 0.27	< 0.62	
	1.0	34.4	[110,185]	< 0.2	< 0.93	
	3.0	45.3	[85,280]	< 0.47	< 0.97	
PUPPI	0.1	15.3	[110,170]	< 0.16	< 0.5	> 0.46
	0.3	22.0	[120,175]	< 0.25	< 0.6	
	1.0	32.3	[105,180]	< 0.22	< 0.99	
	3.0	42.7	[70,320]	< 0.46	< 0.97	

Table 7.2: Summary of the working points derived for low-p_T top tagging with the HTT V2 algorithm. The selection efficiencies of background and signal events are described by $\epsilon(B)$ and $\epsilon(S)$, respectively.

ble top tagging performance for increasing number of pileup vertices has been observed (see Section 7.5.2), also for low-p_T top tagging PUPPI will become the recommended algorithm.

7.7 Summary and outlook

The studies presented in this chapter have shown that jets corrected with PUPPI lead to stable top tagging efficiencies and misidentification rates with increasing number of pileup vertices. Since a stable performance over a wide range of pileup conditions is especially important for LHC Run-II, where high instantaneous luminosities and therefore high levels of pileup vertices are reached, PUPPI will become the recommended pileup mitigation technique for top tagging. The mass peak position in signal events can be found at roughly the same values in dependence of the p_T of the large-cone jets if PUPPI is applied, while shifts in the mass peak position are observed for CHS. Furthermore, it has been confirmed that the working points optimized for CHS-corrected jets are not optimal for PUPPI-corrected jets. Therefore, specific working points for jets corrected with the PUPPI algorithm have been derived. At the same time, the performance of the to date recommended top tagging algorithms has been reviewed, which led to changes in the recommendations for top tagging. These include the usage of the groomed n-subjettiness ratio for low-p_T top tagging. Additionally, the provided working points are recommended to be used in analyses both with and without the subjet requirement. In terms of signal selection efficiencies and background rejection it has been observed that by the application of PUPPI instead of CHS for the mitigation of pileup contributions the performance stays stable.

The application of the working points has been validated in data and correction factors for simulation are available for all but the low-p_T PUPPI working points. These are widely used in searches analyzing data collected by the CMS experiment in the years 2015 and 2016. The working points and the corresponding correction factors are documented in [6].

8 Search for Dark Matter produced in association with a top quark pair at $\sqrt{s} = 13\,\mathrm{TeV}$

This chapter presents a search for Dark Matter, which assumes the production of DM particles in association with a top quark pair. An example Feynman diagram illustrating the studied interaction is presented in Fig. 8.1. Besides the DM particles χ, the signal process involves a new scalar or pseudoscalar boson that mediates the interaction between the top quarks and the DM particles.
All existing results by the CMS and ATLAS experiments that target the lepton+jets event signature [3–5] focus on final states, where the leptons and the jets are well separated from each other and can therefore be identified by applying isolated muon and small-cone jet reconstruction. These kind of searches are referred to as *resolved analyses* in the following. In contrast, the presented analysis aims to cover events where the top quarks receive a large Lorentz boost through the recoil against the DM particles, such that their decay products are strongly collimated and cannot be reconstructed with standard methods. This search strategy is denoted as *boosted approach.*

The chapter is divided into two parts. The first part is dedicated to the design of the search strategy and its optimization, which is based on simulated events only. In the second part the optimized selection requirements obtained in the first part are applied to data events recorded by the CMS experiment at $\sqrt{s} = 13\,\mathrm{TeV}$ in the year 2016.

Search channel and strategy

Given that DM particles do neither decay nor interact while they travel through the detector, their existence only manifests in an imbalance in the total measured transverse momentum of the visible final state particles. Thus, the studied event signature is characterized by a large amount of $\not{E}_T$, which arises from the kinematic energy and the mass of the DM particles. From Fig. 8.2 the dependence of the reconstructed $\not{E}_T$ on the mass of the boson Φ and on the mass of the DM particle χ can be deduced. The $\not{E}_T$ distribution for signal models with $M_\chi = 1\,\mathrm{GeV}$ and different mediator masses is presented for a scalar (Fig. 8.2(a)) and a pseudoscalar (Fig. 8.2(b)) mediator. Due to the larger amount of kinematic energy available for the DM particles, the reconstructed $\not{E}_T$ gets higher with higher masses of the mediator. Signal events with pseudoscalar mediators exhibit harder $\not{E}_T$ spectra than the corresponding processes with scalar mediators. For low mediator masses, the $\not{E}_T$ distributions for pseudoscalar mediators are also broader than those for scalar mediators. The differences between scalar and pseudoscalar mediator particles become less prominent with higher mediator masses. Figures 8.2(c) and 8.2(d) present

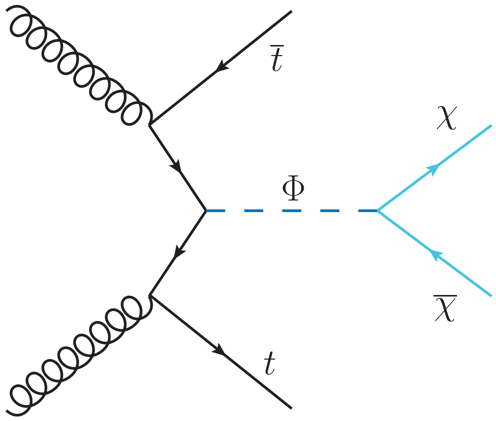
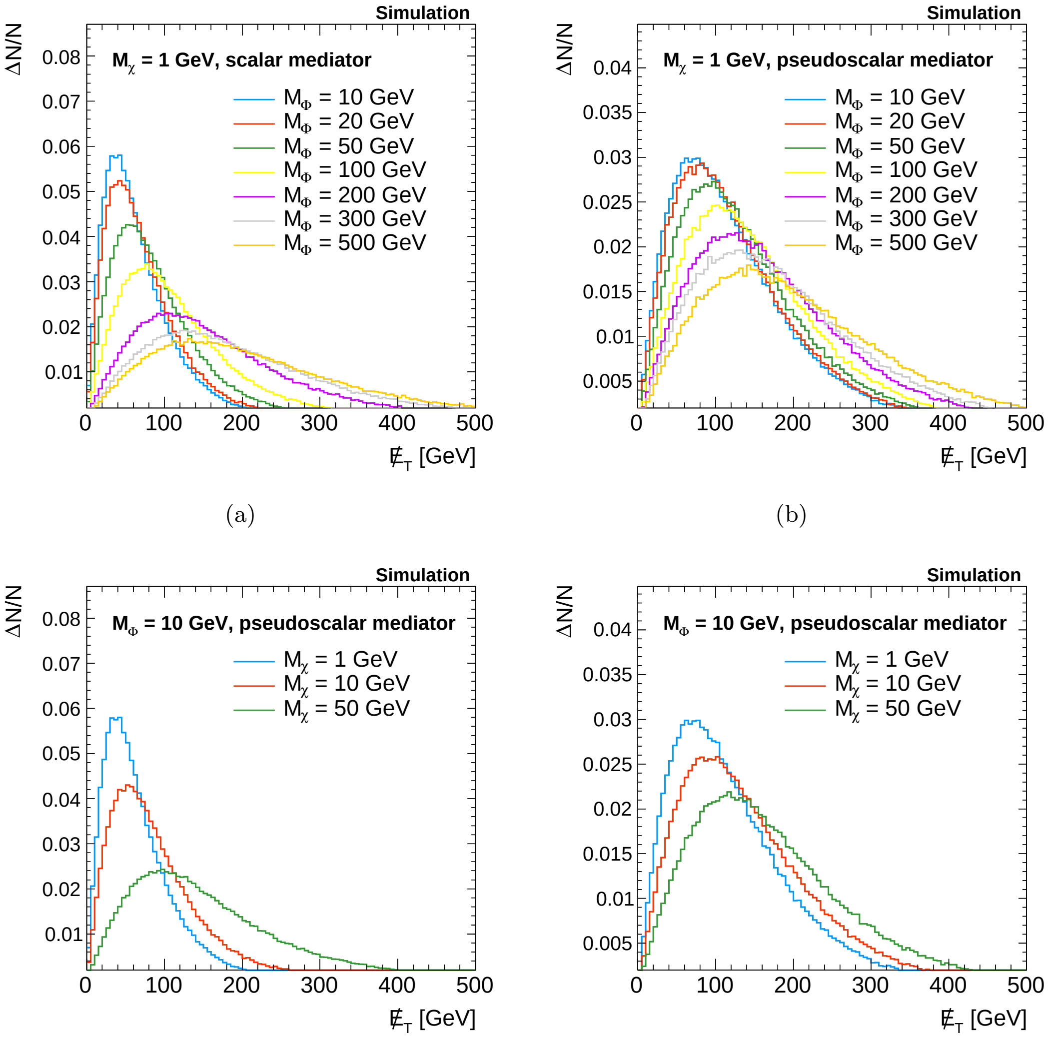


Figure 8.1: Example Feynman diagram for DM particles χ produced in association with a top quark pair. The interaction is mediated by the scalar or pseudoscalar boson Φ.

the $\not{E}_T$ distributions for signal models with $M_\Phi = 10\,\mathrm{GeV}$ and different masses of the DM particles. Figure 8.2(c) assumes a scalar mediator, in Fig. 8.2(d) the $\not{E}_T$ distributions for a pseudoscalar mediator are depicted. While the distributions for pseudoscalar mediators are again harder than those for scalar mediators, the $\not{E}_T$ spectra become harder with increasing mass of the DM particles in both cases. For scalar mediators, the $\not{E}_T$ distributions broaden significantly at the same time. The latter behavior is also linked to the fact that the signal models with the higher DM masses feature an off-shell mediator. Different kinematic behaviors of on-shell ($2M_\chi < M_\Phi$) and off-shell ($2M_\chi > M_\Phi$) mediators are expected.

Besides the large amount of $\not{E}_T$, the studied signal events contain two top quarks, which almost exclusively decay into a W boson and a b quark. Similar to the leptoquark search, one of the two produced W bosons is required to decay into a muon and the corresponding neutrino. This selection requirement significantly reduces the extremely large QCD background present at the LHC. To be able to achieve a high signal selection efficiency, the second W boson is required to decay into two quarks. Thus, in addition to the two b quark jets, signal events feature at least two additional jets.

The p_T distributions of generator-level top quarks in $t\bar{t}$ background events that decay in the lepton+jets channel and different signal models after some basic selection steps described later (see Section 8.1), including a tight $\not{E}_T$ requirement, are shown in Fig. 8.3. In both the signal and background processes the top quarks exhibit a moderate Lorentz boost with transverse momenta around 200 GeV and 300 GeV. While in the signal high $\not{E}_T$ can be achieved by the mass and kinematic energy of the DM particles, the $\not{E}_T$ in the background arises solely from the neutrinos in the top quark decays. Thus, in $t\bar{t}+\text{jets}$ events that pass the tight $\not{E}_T$ requirement the $t\bar{t}$ system has in general a larger boost than the $t\bar{t}$ system in signal events. Only very small kinematic differences between scalar and pseudoscalar mediators and different signal models are observed. The event selection applied in this search is based on the requirement that at least one of the two produced top quarks exhibit a relatively large p_T. By not requiring isolated muon candidates, boosted leptonically decaying top quarks are covered, in which the muon is spatially close to the b



(a)

(b)

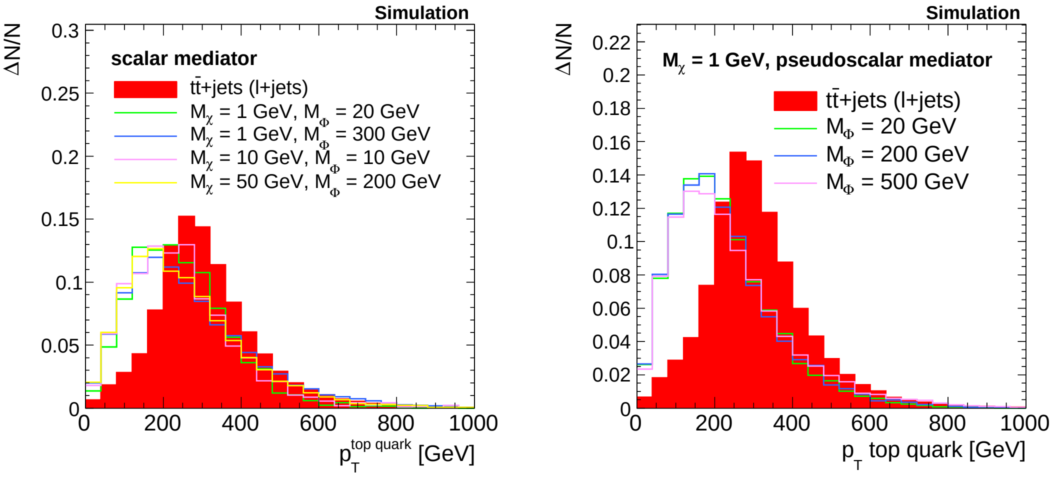
(c)

(d)

Figure 8.2: Comparison of the $\not{E}_T$ distribution for different signal models (*colored lines*) assuming (a) $M_\chi = 1\,\text{GeV}$ and scalar mediators with masses between 10 GeV and 500 GeV, (b) $M_\chi = 1\,\text{GeV}$ and pseudoscalar mediators with masses between 10 GeV and 500 GeV, (c) a scalar mediator with $M_\Phi = 10\,\text{GeV}$ and masses of the DM particles between 1 GeV and 50 GeV (d) a pseudoscalar mediator with $M_\Phi = 10\,\text{GeV}$ and masses of the DM particles between 1 GeV and 50 GeV.

quark jet. Boosted hadronically decaying top quarks are taken into account by including top tagging (see Chapter 7) in the event selection. Since the top quarks only posses a moderate Lorentz boost the HTT V2 algorithm is chosen for top tagging, which offers a very good performance in the required p_T range (see Sections 7.1.2 and 7.5.2). This search strategy is also designed with a possible combination of the boosted and resolved approaches (like [3, 4]) in mind. By requiring events that contain an isolated muon candidate to include also a top tagged jet, the two analyses could be easily combined. To ensure orthogonality between the two channels only a veto of events that pass the top tag selection would have to be added to the resolved analysis.

The main background contribution arises from $t\bar{t}$ + jets events decaying in the lep-



(a) (b)

Figure 8.3: Shape comparison of the p_T spectra of the generator-level top quarks in lepton+jets decays of $t\bar{t}+$jets events (*red area*) and different signal models with (a) scalar and (b) pseudoscalar mediators (*colored lines*).

ton+jets channel, meaning via $t\bar{t} \to W^+b\ W^-\bar{b} \to q\bar{q}'b\ \mu^-\bar{\nu}_\mu\bar{b}$, or fully leptonically ($t\bar{t} \to W^+b\ W^-\bar{b} \to \mu^+\nu_\mu b\ \ell^-\bar{\nu}_\ell\bar{b}$). In these events, $\not{E}_T$ is mainly produced by the neutrinos leaving the experiment unseen. Additionally, mismeasurements of the jet energies can create spurious $\not{E}_T$. Also $W+$jets events, single top quark and diboson events as well as $t\bar{t}$ production in association with a W or Z boson can produce the studied signature consisting of a muon candidate, $\not{E}_T$ and jets and thus constitute SM background processes for this search. A further background process is DY$+$jets events where the Z boson decays into a muon pair. These events contribute to the background processes if one of the produced muons is not reconstructed such that spurious $\not{E}_T$ is generated.

The main challenges of the analysis are two-fold. On the one hand, the assumed cross sections of $t\bar{t}+\chi\overline{\chi}$ production are much smaller than those of the SM background processes. In the simplified model framework [49,53], the signal model with a mediator mass of $M_\Phi = 10$ GeV and a DM mass of $M_\chi = 1$ GeV, which is the process with the largest cross section of the considered signal models, has a approximately 39 times smaller cross section than the $t\bar{t}+$jets production. Consequently, it is important to design a search that keeps the signal selection efficiency as high as possible, while at the same time the amount of SM background processes needs to be greatly reduced.

The second challenge for this search is that the main background process, which is $t\bar{t}$+jets production, features exactly the same visible final state as the signal events. This similarity makes the signal processes hardly distinguishable from SM $t\bar{t}+$jets production. The only difference between the two processes is the slightly harder $\not{E}_T$ spectra of the signal events in comparison to SM $t\bar{t}+$jets production. One of the main approaches applied to discover DM in the $t\bar{t}+\chi\overline{\chi}$ channel is to search for an excess of events over the SM background in the tails of the $\not{E}_T$ distribution. Unfortunately, due to the neutrinos produced in $t\bar{t}$ decays, the measured $\not{E}_T$ can also be quite large in the background processes, which leads to a quite similar shape of the $\not{E}_T$ distribution in the background and signal events. Hence, a main part of the performed optimization in this thesis has been dedicated to studies targeting at finding a variable with higher discrimination power between the $t\bar{t}$ background and the signal processes than $\not{E}_T$ (see Section 8.2).

8.1 Baseline selection

The mentioned optimization studies are based on the *baseline selection* summarized in the following. The baseline selection is designed with having a boosted search in mind. The $\not{E}_T$ requirement is adopted from the 8 TeV search presented in [52]. The baseline selection includes the following selection criteria, which have to be met by each studied event:

- At least one primary vertex candidate passing the quality criteria detailed in Section 5.2 is present.
- Exactly one muon candidate passing the medium working point, $p_T > 47\,\mathrm{GeV}$ and $|\eta| < 2.1$ is found. No isolation criterion is applied on the muon candidates.
- Events that contain electron candidates with $p_T > 50\,\mathrm{GeV}$ and $|\eta| < 2.5$ that fulfill loose identification requirements are discarded.
- At least two jets passing the loose PF identification requirements, $p_T > 50\,\mathrm{GeV}$ and $|\eta| < 2.4$ are present.
- $\not{E}_T$ exceeds 160 GeV.

The electron requirement helps to suppress SM background processes and facilitates a possible addition of the electron channel to this search. Due to the high $\not{E}_T$ requirement, the selected $t\bar{t}$ pairs in the SM $t\bar{t}$+jets background and signal processes already exhibit a moderate Lorentz boost (see Fig. 8.3). The fraction of boosted events in the signal region is additionally enhanced by applying the following event categorization and selection, which aims at selecting signal events with either a boosted leptonically or hadronically decaying top quark. Two categories are defined, denoted as *isolated lepton signal region* and *non-isolated lepton signal region*. The isolated lepton signal region contains events that fulfill the following criteria:

- The muon candidate passes the tight isolation requirements (see Section 5.3).
- Each event is top tagged. For the identification of the top quark jet the loose working point of the HTT V2 algorithm with subjet-b-tag requirement (see Table 7.2) is applied.

In this way, events that contain moderately boosted hadronically decaying top quark candidates are selected.

To be allocated to the non-isolated lepton signal region, events have to pass the criteria listed below:

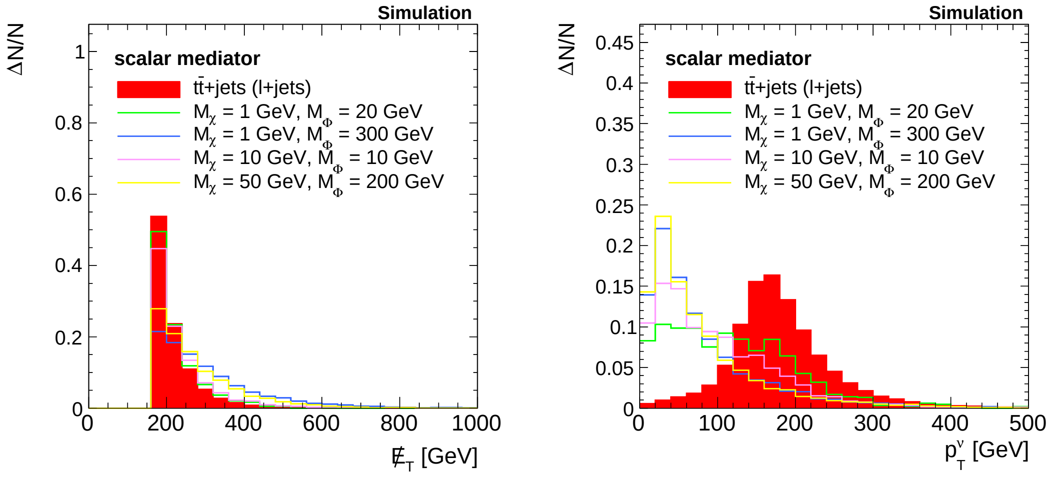
- The muon candidate fails the tight isolation requirements.

- The event passes the 2D-cut. The 2D-cut is designed to reduce the QCD multijet background to a negligible amount and in this manner replaces the isolation criterion in cases where non-isolated leptons are studied. To pass the 2D-cut the selected muon candidate in the event has to have a distance larger than $\Delta R > 0.4$ to all jets with $p_T > 15\,\mathrm{GeV}$ and $|\eta| > 3.0$ or a transverse momentum perpendicular to those jets ($p_{\mathrm{T}}^{\mathrm{rel}}$) larger than $40\,\mathrm{GeV}$. The p_T and $|\eta|$ restrictions as well as the thresholds on ΔR and $p_{\mathrm{T}}^{\mathrm{rel}}$ were optimized for the analysis presented in [183].

Thus, the non-isolated lepton signal region contains events in which the muon candidate and the b quark jet produced in the top quark decay are spatially close. These events are not covered by any of the analyses performed so far.
The main background is $t\bar{t}$ jets production decaying in the lepton+jets channel, with smaller contributions from di-leptonic $t\bar{t}$ decays. Also $W +$ jets production contributes significantly to the SM background processes. Other background processes arising from DY + jets, single top quark production, diboson and QCD multijet production are very small compared to $t\bar{t}$ and $W +$ jets. The number of events for background and signal processes is much higher in the isolated lepton signal region than in the non-isolated lepton signal region. Still, the latter contributes to the sensitivity of the presented search.

8.2 A new discriminating variable

As discussed before, in existing searches for DM produced in association with a top quark pair the higher $\not{E}_T$ values measured in $t\bar{t} + \chi\overline{\chi}$ events compared to background processes are exploited since it is search for an excess of signal events in the tails of the $\not{E}_T$ distribution. Unfortunately, the main background process, which is $t\bar{t}$ + jets production, also produces $\not{E}_T$ due to neutrinos emitted in lepton+jets and fully-leptonic decays of the $t\bar{t}$ pairs. Furthermore, mismeasurements of jet energies can create spurious $\not{E}_T$. In Fig. 8.4(a) a shape comparison between the $\not{E}_T$ distributions in $t\bar{t}$+jets events and different signal models is presented. The mentioned sources of $\not{E}_T$ in the $t\bar{t}$ background create a spectrum that is very similar to that produced in the signal processes. Since boosted $t\bar{t}$ events are selected by the high $\not{E}_T$ requirement of the baseline selection, which thus also contain a high-p_T neutrino, the shape similarity is further enhanced. Additionally, especially for signal models with low mediator and DM masses, great parts of the peaks of the signal distributions are rejected by the tight $\not{E}_T$ selection such that only events in the tails of the $\not{E}_T$ distributions are kept (see Fig. 8.2). Still, the high $\not{E}_T$ requirement is very efficient in suppressing SM background processes and constitutes an essential part of the selections applied in all existing searches.
The performed optimization studies target the design of a variable with improved discrimination ability between signal and background processes compared to $\not{E}_T$. The idea is to reconstruct the four-momentum of the neutrino produced in lepton+jets decays of the $t\bar{t}$ pair ($t\bar{t} \to W^+b\ W^-\bar{b} \to q\bar{q}'b\ \mu^-\bar{\nu}_\mu\bar{b}$), being the main background after the baseline selection, without using the measured $\not{E}_T$ value. Instead, other constraints of boosted lepton+jets SM $t\bar{t}$ decays are exploited. Afterwards, the reconstructed neutri-



(a) (b)

Figure 8.4: Shape comparison of (a) the distribution of the reconstructed $\not{E}_T$ and (b) the transverse momentum of the generator-level neutrino in lepton+jets decays of $t\bar{t}$ + jets events (*red area*) and different signal models assuming scalar mediators (*colored lines*).

nos are compared to the measured $\not{E}_T$. In $t\bar{t}$ + jets events, the reconstructed transverse momentum and the measured $\not{E}_T$ in the event are expected to be compatible, while in signal events the kinematics of the produced neutrinos in the top quark decays (see Fig. 8.4(b)) are different and large values of $\not{E}_T$ arises from the DM particles. Thus, the neutrino solution is not expected to match the measured $\not{E}_T$ value in direction and absolute value for signal processes. In other words, the reconstructed transverse momentum of the neutrino is used to design a variable that measures the compatibility of the measured $\not{E}_T$ with a neutrino from lepton+jets SM $t\bar{t}$ events.

Since the only difference between the signal and $t\bar{t}$ + jets processes is the presence of the DM particles and thus the higher amount of $\not{E}_T$ in the signal, only constraints targeting the $t\bar{t}$ system are used. This approach, together with not using $\not{E}_T$ in the reconstruction, prevents increasing the similarities between background and signal processes and the associated deterioration in discrimination power between the processes. The described effect has been observed by adopting the reconstruction of the $t\bar{t}$ system from a search for a heavy Z' boson decaying into a top quark pair [184, 185] to this search. In the method presented in [184, 185], the $\not{E}_T$ is assumed to arise solely from the neutrino produced in the decay of a top quark via $t \to W^+ b \to b\ell^+\nu_\ell$. Thus, the measured $\not{E}_T$ value translates directly into the transverse momentum of the neutrino such that the unknown p_z component of the neutrino four-momentum can be determined from the constraint that the invariant mass of the lepton candidate and the neutrino should yield the W boson mass. The $t\bar{t}$ system is subsequently obtained by assigning jets to the leptonically or hadronically decaying top quark such that the minimal χ^2 value is achieved. The studied χ^2 thereby mainly consists of constraints on the reconstructed top quark masses. In principal, the mentioned χ^2 distribution can be used to distinguish background and signal processes in the search for DM. In the DM signal processes the reconstruction of the $t\bar{t}$ system should not work properly since a large contribution of $\not{E}_T$ arises from the DM particles and the assumption that the neutrino four-momentum is described by the measured $\not{E}_T$ value in the event does not hold. Instead, it has been observed that the algorithm chooses the wrong jet combinations in the signal events to reconstruct the $t\bar{t}$-system in order to facilitate a small χ^2 value. This leads to quite similar χ^2 distributions in signal and background events such that no gain in sensitivity is achieved by applying this method. In a kinematic fit of the $t\bar{t}$ system, like the one adopted e.g. in [186, 187], this behavior would most likely be even more enhanced. Thus, in the presented analysis constraints that are applicable to $t\bar{t}+$ jets processes and to signal events are adopted, while, most importantly, the measured $\not{E}_T$ value is not used.

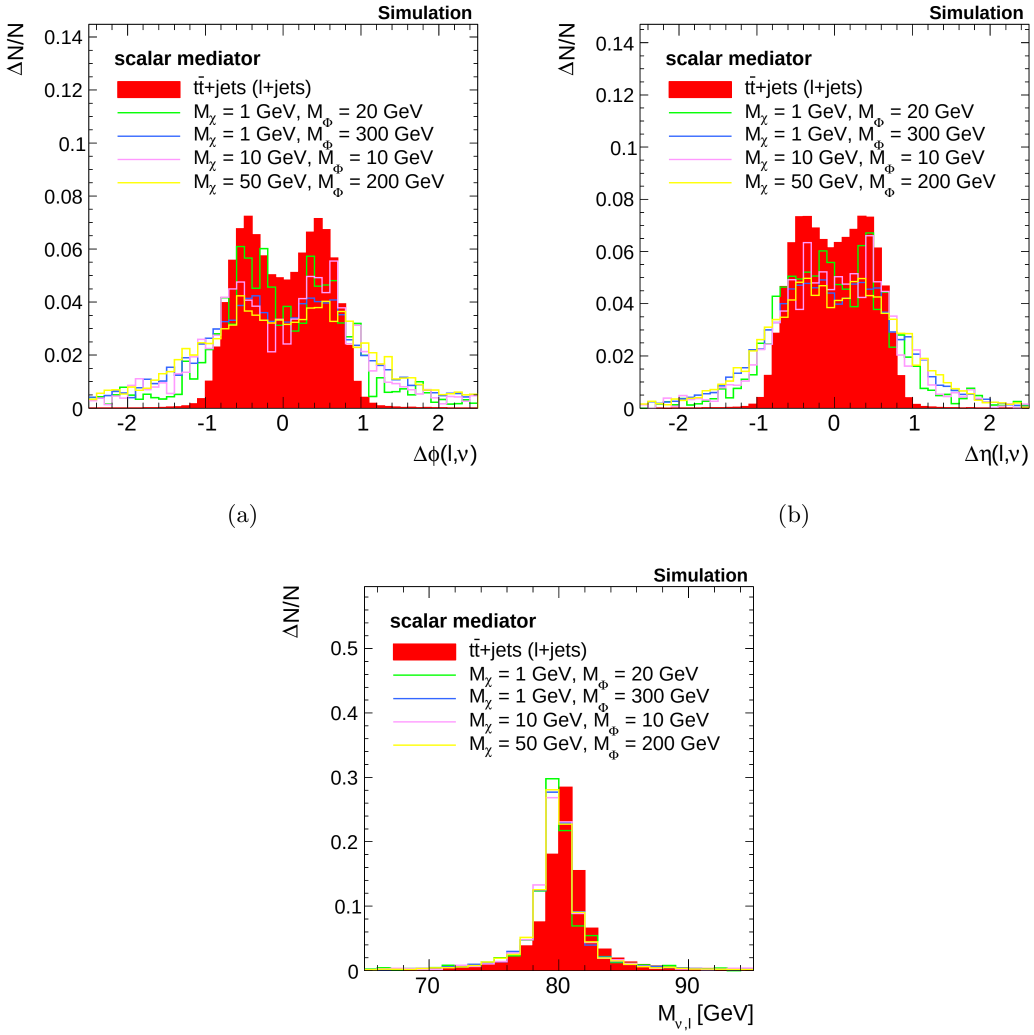
To enable a reconstruction of the neutrino four-momentum without the usage of the observable $\not{E}_T$, characteristics of boosted lepton+jets SM $t\bar{t}$ decays are examined. The studied variables are

$$\begin{aligned} \Delta\phi(\nu,\mu) &= \phi_\nu - \phi_\mu, \\ \Delta\eta(\nu,\mu) &= \eta_\nu - \eta_\mu, \\ M^2_{\nu\mu} &= (P_\nu + P_\mu)^2, \end{aligned} \tag{8.1}$$

where P_ν and P_μ denote the four-momentum of the neutrino and the muon candidate found in the $t\bar{t}$ system, respectively. All three distributions are depicted for lepton+jets decays of the $t\bar{t}$ system and different signal models in Fig. 8.5. Generator-level information is used for the neutrino. As expected in moderately boosted $t\bar{t}$ events, the differences in the angles between the lepton and the neutrinos are relatively small. In both distributions, two clear symmetric peaks are observed. This is also true for the signal processes. However, due to the smaller Lorentz boost of the top quarks in these samples (see Fig. 8.3), the signal distributions feature longer tails and are more smeared out. The invariant mass of the neutrino and the muon shows a sharp peak around the W boson mass in both the signal and the background processes. These three variables are suitable for a construction of a normalized three-dimensional likelihood function $\mathcal{L} = \mathcal{L}(\Delta\phi, \Delta\eta, M_{\nu\mu})$. Thus, for each measured lepton four-momentum the likelihood describes the probability to find a neutrino with a certain four-momentum. Since both studied angles are correlated with the invariant mass of the neutrino and the muon candidate, only two of the three free parameters of the neutrino four-momentum can be constrained with the likelihood function. This necessitates a fourth constraint, which is built by requiring that the b quark candidate jets, the muon candidate and the neutrino have an invariant mass $M_{\nu\mu b}$ compatible with the top quark mass. Since the available statistical precision is not sufficient for an inclusion of the top quark-mass constraint into the likelihood, a χ^2 variable is constructed out of the likelihood and the top quark-mass constraint. This χ^2 is given by

$$\chi^2 = -2\log(\mathcal{L}) + \frac{(M_{\nu\mu b} - M_t)^2}{\sigma_t^2}, \tag{8.2}$$

with $M_t = 164.7\,\mathrm{GeV}$ and $\sigma_t = 13.9\,\mathrm{GeV}$. These two values have been determined by a reconstruction of the top quark candidate using the generator-level neutrino, the reconstructed muon candidate and the jet closest to the muon candidate in ΔR. The



(a)

(b)

(c)

Figure 8.5: Shape comparison of the variables used as input for the likelihood function: (a) the difference in the angle ϕ, (b) the angle η between the muon candidate and the generator-level neutrino, and (c) the invariant mass $M_{\nu\mu}$ of the muon candidate and the generator-level neutrino in lepton+jets decays of $t\bar{t}+$ jets events (*red area*) and different signal models assuming scalar mediators (*colored lines*).

values M_t and σ_t are afterwards extracted by performing a fit of the distribution with a Gaussian function.

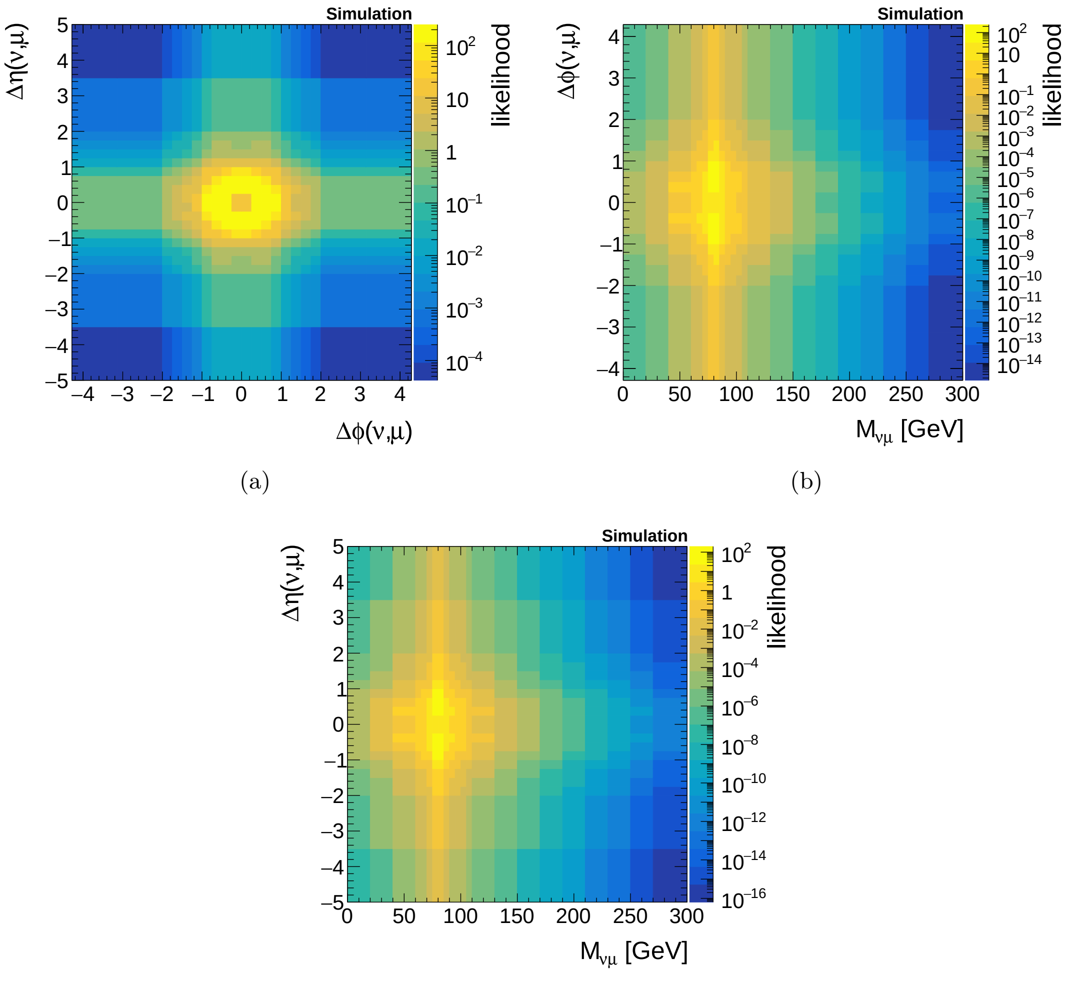
By minimizing the χ^2 function given in Eq. (8.2) the neutrino four-momentum can be determined. However, before the minimization is feasible, a few modifications to the likelihood have to be done. In principal, all three of the observables studied in the likelihood are continuous functions. However, due to the limited statistical precision of the simulated samples, the likelihood is discretized. To collect as much information of the structure of the likelihood as possible small bin sizes are used in the bulk of the distributions. In the tails of the distributions, where the statistical precision is limited, larger bin sizes are chosen. Despite the adapted binning of the likelihood histogram, not all bins are filled. This is a problem for the minimization process since information is needed in the whole phase space to be able to find the minimum of the distribution. Thus, bins without entries are filled by extrapolating the likelihood value from the filled entries, assuming factorization of the likelihood function in empty bins. For each of the three variables a projection into one dimension is derived by integrating over the other two variables. Based on the projection, a fit to the one dimensional function is performed in each variable to get the respective parametrization. The three parametrizations multiplied with each other are afterwards used to fill the empty entries in the likelihood histogram. Using a factorized likelihood function implies that the three studied observables are uncorrelated. This assumption has to be made since no information about the correlations of the variables in the affected bins is available.

In order to construct a rather smooth likelihood distribution without large fluctuations due to limited statistical precision, bins with only few entries and thus a large statistical uncertainty are discarded and refilled afterwards with the method explained before.

Figure 8.6 presents two-dimensional slices of the resulting likelihood histogram for all combinations of two of the three variables. In each figure the third observable, meaning the one not included in the two-dimensional histograms, is set to a fixed value. Since $\Delta\phi$ is periodic, one extra bin has been added beyond $|\Delta\phi| = \pi$, which is simply a copy of the corresponding value shifted by 2π. This is helpful for the minimization process described later. In all three distributions, a smooth likelihood without large fluctuations is found. The likelihood features clear maxima and decreases rapidly. One of the reasons why a likelihood function is chosen for the reconstruction, instead of a χ^2, are correlations between the observables $\Delta\phi$ and $M_{\nu\mu}$, and $\Delta\eta$ and $M_{\nu\mu}$ (see Fig. 8.6), which can be taken into account in the minimization process.

This likelihood histogram can afterwards be used for the construction of the χ^2 function as described in Eq. (8.1), which is minimized to reconstruct the neutrino four-momentum. The following aspects have to be taken into account to permit the minimization of the χ^2 function. The χ^2 function has to be continuously differentiable and has to exhibit well-defined minima. Otherwise, the minimization process, which exploits the gradient of the function, will not be able to find a solution. To check if the defined distribution fulfills these requirements, the χ^2 function is analyzed in single simulated lepton+jets SM $t\bar{t}$ events by scanning the p_x and p_y components in ranges between $-1050\,\text{GeV}$ and $1050\,\text{GeV}$. The p_z component is set to its value on generator level in these scans. Two examples are presented in Fig. 8.7. In both scans clear minima and no obvious discontinuities are observed. Moreover, the generator-level neutrino components lie in the minima of the χ^2 function. Hence, the χ^2 function seems to be well defined and in principle the minimization is possible. However, a small discontinuity is found for values of p_x and p_y close to 0. In these cases, the neutrino momentum is aligned almost completely in z-direction and the η coordinate gets large. Therefore, the corresponding $\Delta\eta$ is found in the tails of the likelihood function, where it becomes zero. Since $\log(0)$ is not defined, a discontinuity is created. To prevent the minimizer from running into this discontinuity, the likelihood is set to an arbitrary large value for transverse momenta of the neutrino below $5\,\text{GeV}$.

For the minimization of the χ^2 function MINUIT [188] is used. Before the final minimiza-



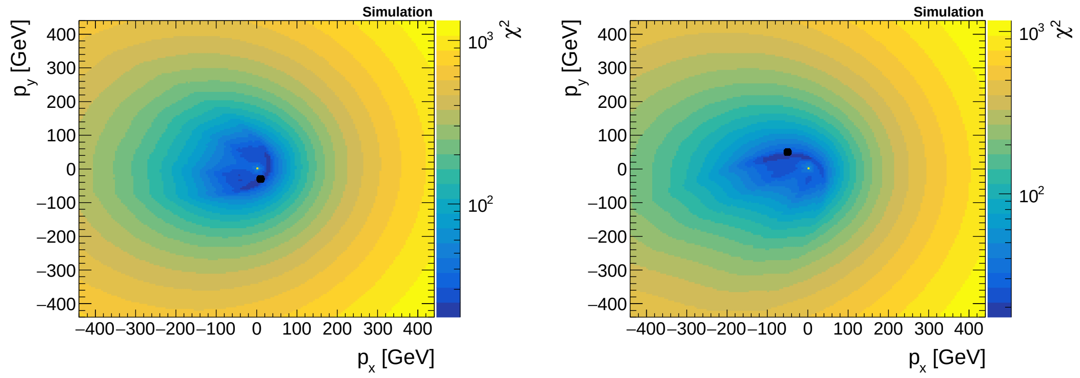
(a)

(b)

(c)

Figure 8.6: Slices of the binned likelihood including all selected $t\bar{t}$ + jets events as a function of (a) $\Delta\phi(\nu, \mu)$ (*x-axis*) and $\Delta\eta(\nu, \mu)$ (*y-axis*) for fixed $M_{\nu\mu}$, (b) $M_{\nu\mu}$ (*x-axis*) and $\Delta\phi(\nu, \mu)$ (*y-axis*) for fixed $\Delta\eta(\nu, \mu)$, (c) $M_{\nu\mu}$ (*x-axis*) and $\Delta\eta(\nu, \mu)$ (*y-axis*) for fixed $\Delta\phi(\nu, \mu)$.

tion is performed, the b jet associated with the leptonically decaying top quark candidate has to be identified and a suitable starting point for the neutrino four-momentum has to be found. To find the b quark jet associated with the leptonically decaying top quark candidate, several minimizations of the χ^2 function are performed. As starting point for the neutrino four-momentum in these fits the four-vector of the reconstructed lepton candidate shifted by the mean values of $\Delta\phi(\nu, \mu)$ and $\Delta\eta(\nu, \mu)$ is chosen. In each minimization a different jet is chosen as the b quark jet, whereby all jets with $p_T > 15\,\text{GeV}$ and $|\eta| > 2.4$ are tested. After each minimization procedure, the top mass constraint $\frac{(M_{\nu\mu b} - M_t)^2}{\sigma_t^2}$ is evaluated. The jet that yields the smallest top mass constraint is chosen as the final b quark jet. To identify the starting point for the neutrino four-momentum in the final minimization, the p_z component of the reconstructed neutrino found in the same minimization process as the chosen b quark jet and the b quark jet itself are used to perform a rough scan over the χ^2 function to find its minimum. This minimum in turn defines the starting points for p_x and p_y in the final minimization process performed



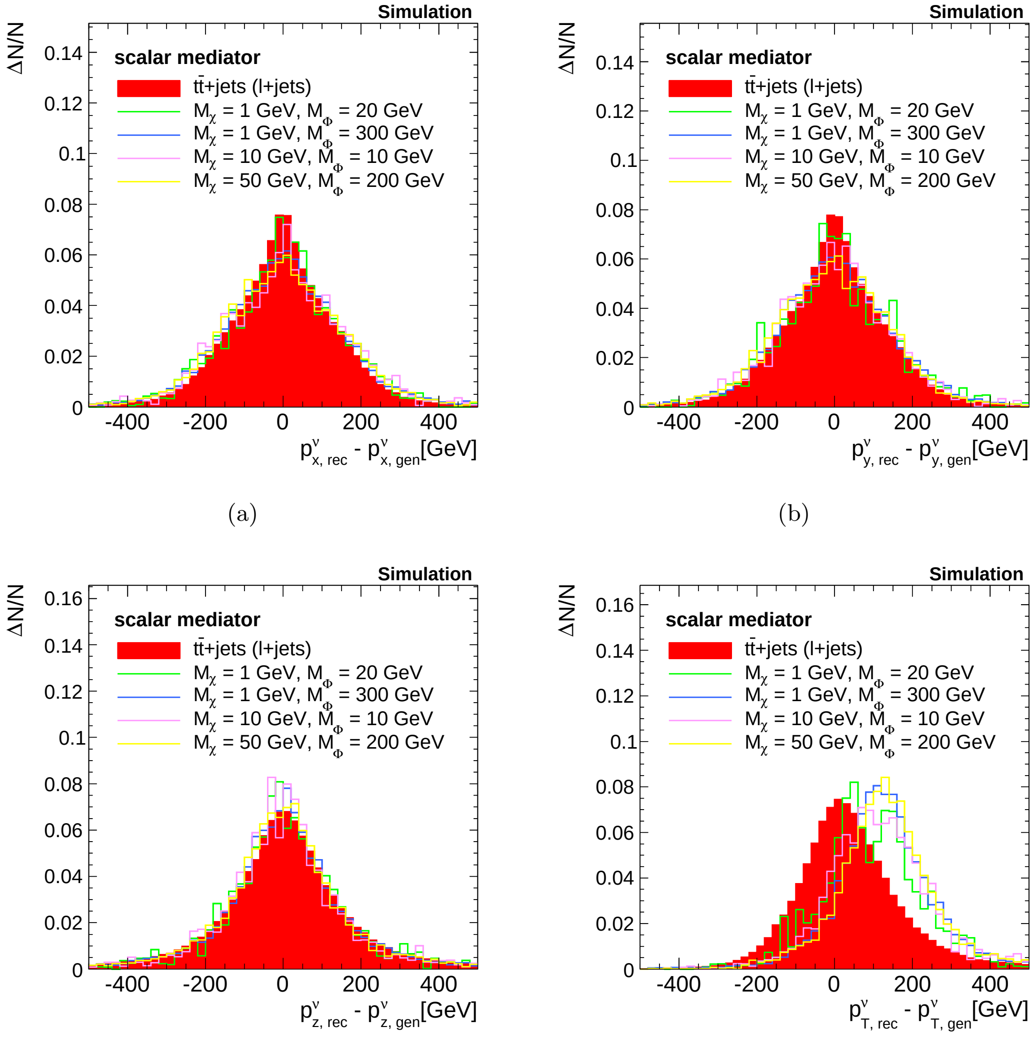
(a) (b)

Figure 8.7: Scan of χ^2 (*colors*) as a function of p_x and p_y of the neutrino. (a) and (b) each show the χ^2 for a single simulated $t\bar{t}+$jets event. The p_z component is set to its generator-level value. The generator-level p_x and p_y of the neutrino components are presented by the black dots.

thereafter.

In the minimization procedures several aspects have to be taken into account. As already said, MINUIT needs a continuously differentiable function to be able to find the minimum. Thus, to define a gradient within a bin of the likelihood histogram a linear interpolation in three dimensions using the centres of the eight neighbouring bins is performed. This also explains the purpose of the values added beyond π in $\Delta\phi$ in the likelihood histogram. Since $\Delta\phi$ is periodic with 2π, the values shifted by 2π can be used for the interpolation of the values around $|\Delta\phi| = \pi$. This does not work for the outermost bins of the likelihood histogram in the other two variables. For all values outside the likelihood histogram, the extrapolation used for filling empty spaces in the likelihood histogram is employed to determine the value of the likelihood in these cases.

The difference between the reconstructed x, y and z components of the momentum of the neutrino, meaning the result of the minimization, and the respective generator-level values for simulated $t\bar{t}+$jets events and different signal models are presented in Fig. 8.8. Additionally, the difference between the reconstructed transverse momentum of the neutrino and the generated transverse momentum is presented for the same processes. The latter distribution is particularly interesting since the reconstructed transverse momentum is the variable that will be compared to the measured $\not{E}_T$ in the following to design a new variable with improved discrimination power between background and signal processes than $\not{E}_T$. Despite the relatively long tails towards larger values, all four distributions peak around zero in the $t\bar{t}+$jets processes. This means that the reconstruction of the neutrino four-momentum works quite well in lepton+jets decays of $t\bar{t}+$jets events. For the signal the reconstructed p_x, p_y and p_z components also agree quite well with their generated values. This is not true for the reconstructed transverse momentum, which is, on average, too large compared to its generator-level value. This effect is not observed in the distributions of the p_x and p_y components since results with different signs can compensate each other in these plots such that a distribution symmetric around zero is observed nevertheless. Thus, the method does not work for signal processes. This is



(a) (b)

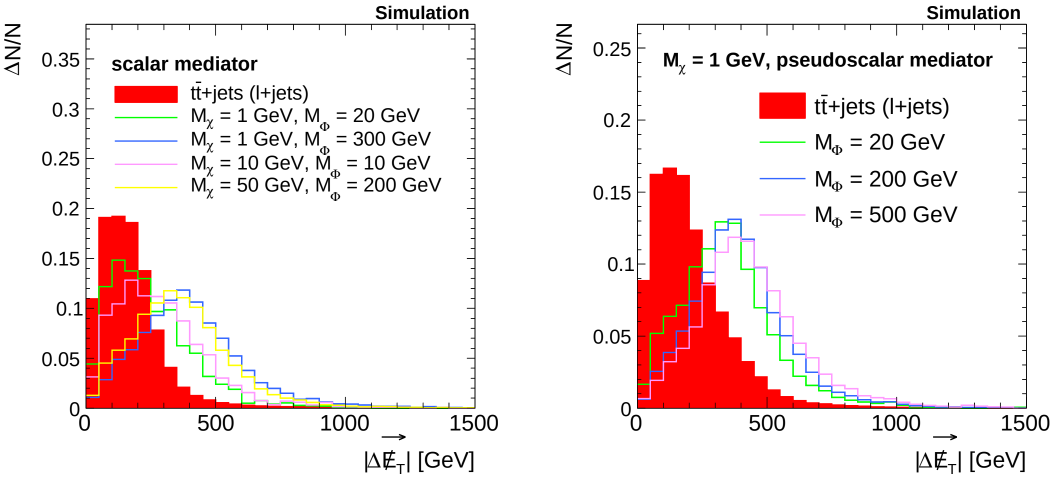
(c) (d)

Figure 8.8: Distributions of the differences between (a) the p_x component, (b) the p_y component, (c) p_z component and (d) the p_T of the four-momentum of the reconstructed neutrino and the generator-level neutrino. The results obtained for lepton+jets decays of the $t\bar{t}$+jets background (*red area*) are compared to those obtained for different signal models assuming scalar mediators (*colored lines*).

expected since the likelihood has been optimized on $t\bar{t}+$jets events and the kinematics are quite different between the signal and background processes. Still, the method is able to reconstruct the neutrino four-momentum in the main background process and can therefore be used for the construction of the desired variable.

With the reconstructed neutrino four-momentum, a variable can be defined that describes the compatibility of the measured $\not{E}_T$ value with a neutrino produced in lepton+jets SM $t\bar{t}$ events. In a lepton+jets $t\bar{t}$ event, the reconstructed transverse momentum of the neutrino should be compatible in direction and absolute value with the measured $\not{E}_T$. Hence, the observable given by

$$|\Delta\vec{\not{E}}_{\mathrm{T}}| = |\vec{\not{E}}_{\mathrm{T}} - \vec{p}_{\mathrm{T}}^{\,\mathrm{rec},\,\nu}| \tag{8.3}$$



(a) (b)

Figure 8.9: Shape comparison of $|\Delta\vec{\slashed{E}}_\mathrm{T}|$ in lepton+jets decays of $t\bar{t}$ + jets events (*red area*) and different signal models assuming (a) scalar mediators and (b) pseudoscalar mediators (*colored lines*).

is very well suited to test the mentioned compatibility since it compares the direction and absolute value of $\vec{\slashed{E}}_\mathrm{T}$ and $\vec{p}_\mathrm{T}^{\,\mathrm{rec},\,\nu}$ at the same time. The variable $\vec{p}_\mathrm{T}^{\,\mathrm{rec},\,\nu}$ denotes the solution from the minimization for the transverse momentum of the neutrino. Other possible ways to compare the measured $\slashed{E}_T$ with the reconstructed neutrino have been tested as well, but $|\Delta\vec{\slashed{E}}_\mathrm{T}|$ has been found to yield the best separation between lepton+jets $t\bar{t}$ background and signal events. A shape comparison of this variable between lepton+jets SM $t\bar{t}$ events and signal events is presented in Fig. 8.9. The $t\bar{t}$ background on average yields low values of $|\Delta\vec{\slashed{E}}_\mathrm{T}|$. Hence, in these events the reconstructed neutrino coincidences very well with the measured $\slashed{E}_T$. Small deviations from zero are expected, since on the one hand the reconstruction does not work for all events (see Fig. 8.8) and on the other hand $\slashed{E}_T$ is also created through mismeasurements of the energies of the jets found in $t\bar{t}+$ jets events. For the signal processes on average much larger values of $|\Delta\vec{\slashed{E}}_\mathrm{T}|$ are observed. In these events the measured $\slashed{E}_T$ does not match the reconstructed neutrino transverse momentum in direction and absolute value since most $\slashed{E}_T$ is produced by the DM particles in the event, which results in very different kinematics in the signal models than in the $t\bar{t}$ background. This effect is largest for signal samples where large values of $\slashed{E}_T$ arising from DM particles are expected, such as processes with high mediator masses or a pseudoscalar mediator. The shape comparisons show that in the variable $|\Delta\vec{\slashed{E}}_\mathrm{T}|$ a signal would appear as a peak over the falling $t\bar{t}$ background. It seems to be able to better discriminate between signal and background processes than the $\slashed{E}_T$ distribution, where similar shapes in background and signal events are observed (see Fig. 8.4(a)). Hence, a promising variable has been found. In order to test its performance, sensitivity studies are performed.

8.3 Optimization of the event selection

Starting from the baseline selection, a final event selection has to be constructed, which has to meet two main criteria. On the one hand, due to the small production cross sections of the signal processes in the simplified model framework, the signal selection efficiency has to be kept as high as possible to be able to design a sensitive analysis. On the other hand, the background processes have to be significantly suppressed at the same time. Meeting both demands simultaneously is very challenging due to the $t\bar{t} + \text{jets}$ background decaying in the lepton+jets and di-leptonic decay mode. To design an event selection fulfilling the mentioned requirements, variables with high separation power between background and signal processes are identified and different combinations and cut thresholds on those are tested. Furthermore, the cuts of the baseline selection are reviewed. The main results are summarized in this section. In parallel to the optimization of the event selection, also the sensitivity of the variable $|\Delta\vec{\not{E}}_{\mathrm{T}}|$, built with the reconstruction of the neutrino in lepton+jets $t\bar{t}$ events as described in Section 8.2, is compared to the results obtained from the $\not{E}_T$ spectrum only. The distributions of both variables after the baseline selection are presented in Figs. 8.10(a) and 8.10(b).

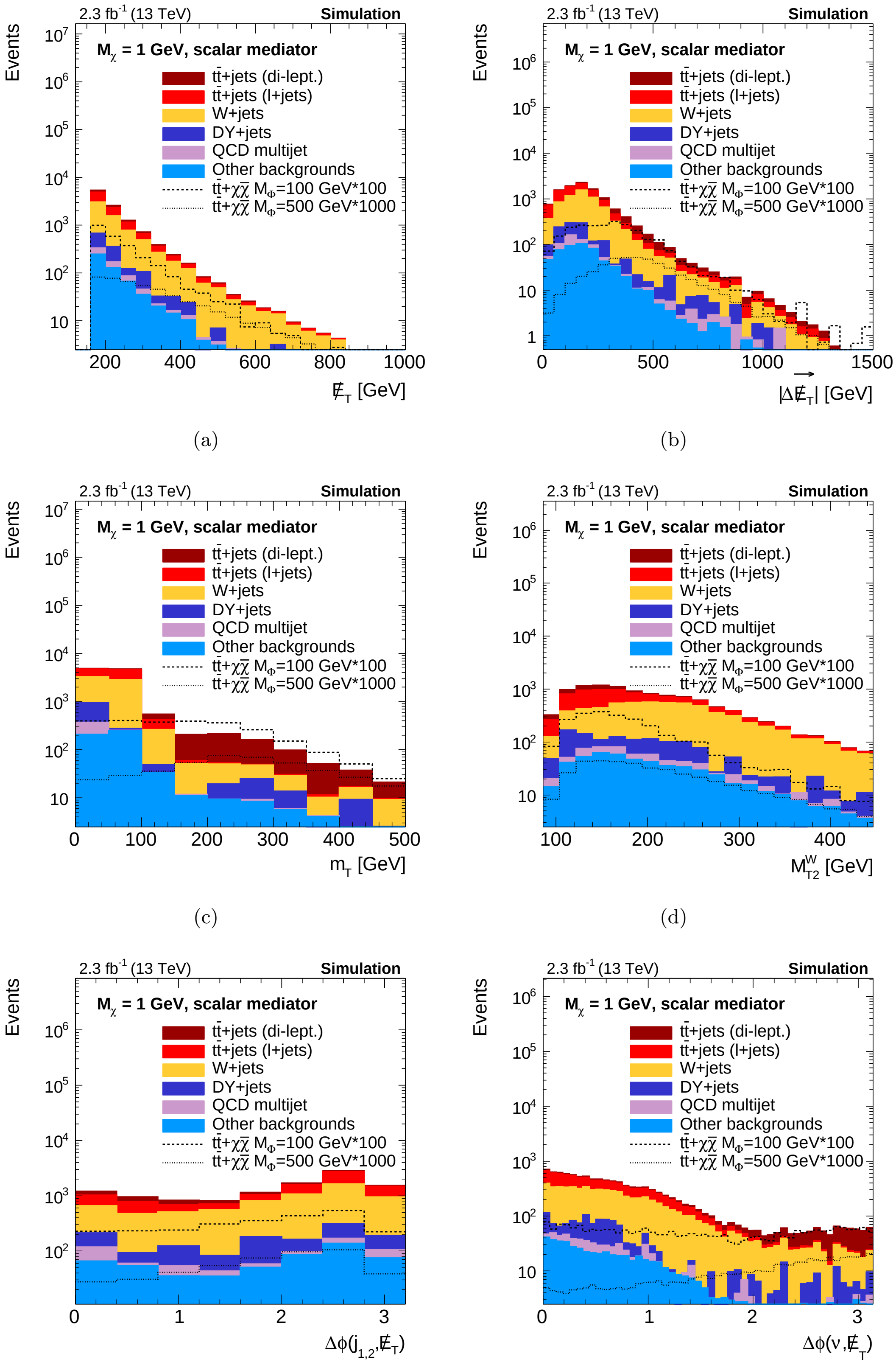
Interesting variables to study for the event selection are m_T and m_{T2}^W. The transverse mass m_T describes the invariant mass of the measured $\not{E}_T$ and the muon candidate in the transverse plane. It is given by

$$m_T = \sqrt{2\,\not{E}_T p_{\mathrm{T}}^{\mu}\,(1 - \cos{(\Delta\phi)})}, \tag{8.4}$$

where $\Delta\phi$ describes the angle between $\not{E}_T$ and the muon candidate in the ϕ-plane and p_{T}^{μ} the transverse momentum of the muon candidate. The m_T distribution after the application of the baseline selection is presented in Fig. 8.10(c). For events containing on-shell W bosons, like $t\bar{t} + \text{jets}$ and $W + \text{jets}$ production, m_T lies below the W boson mass. Larger values of m_T can be realized in events comprising off-shell W bosons or in processes where the neutrino produced in the W boson decay is not the only source of $\not{E}_T$, as e.g. in $t\bar{t} + \chi\overline{\chi}$ signal events and di-leptonically decaying $t\bar{t}$ events. By selecting only events above a certain m_T threshold great parts of the $W + \text{jets}$ and lepton+jets $t\bar{t}$ background are discarded.

Figure 8.10(d) presents the distribution of the observable m_{T2}^W [189] after the baseline selection. This variable is helpful for the reduction of di-leptonic $t\bar{t}$ events, in which only one of the leptons is reconstructed, while the second is not in the acceptance of the search and creates spurious $\not{E}_T$. It is constructed such that it ends at the top quark mass in $t\bar{t}$ events if a perfect detector resolution is assumed (see [189] for further details). In existing searches for $t\bar{t} + \chi\overline{\chi}$ production, like e.g. [5] and [3,4], $m_{T2}^W > 140\,\text{GeV}$, $m_{T2}^W > 170\,\text{GeV}$ or $m_{T2}^W > 200\,\text{GeV}$ is required.

To reduce the non-top quark background processes, the b tag multiplicity is an interesting variable to study. In existing searches for $t\bar{t} + \chi\overline{\chi}$ furthermore events are selected based on the minimum of the angle $\Delta\phi$ between the jet leading- or sub-leading in p_T and $\not{E}_T$, which is denoted as $\Delta\phi(\not{E}_T, \text{jet}_{1,2})$ and shown in Fig. 8.10(e). For signal processes the angle $\Delta\phi(\not{E}_T, \text{jet}_{1,2})$ is on average large, since the DM particles are expected



(a) (b)

(c) (d)

(e) (f)

Figure 8.10: Comparison of the background (*colored areas*) and different signal models (*black lines*) assuming a DM mass of 1 GeV and scalar mediators with masses of 100 GeV and 500 GeV. Shown are the distributions of (a) $\not{E}_T$, (b) $|\Delta\vec{\not{E}}_\mathrm{T}|$, (c) m_T, (d) m_{T2}^W, (e) $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2})$ and (f) $\Delta\phi(\not{E}_T, \nu)$ after the baseline selection.

to be spatially separated from the jets produced in the associated top quark decays. Additionally, also variables that involve the reconstructed neutrino can be studied. The most promising variable, despite $|\Delta\vec{\not{E}}_{\mathrm{T}}|$, is the difference in the angle ϕ between the reconstructed neutrino and $\not{E}_T$, presented in Fig. 8.10(f). For background processes the angle tends to smaller values, while for signal events the distribution is more uniformly distributed and has a small trend to large angles.

Selections involving requirements on these variables and others are tested. To identify the selection with the best sensitivity, the expected upper 95% C.L. cross section limits on the production of $t\bar{t} + \chi\overline{\chi}$ as a function of the mass of the mediator Φ are compared. Additionally, the sensitivity obtained by determining the exclusion limit based on $\not{E}_T$ is compared to the one derived with $|\Delta\vec{\not{E}}_{\mathrm{T}}|$. In these studies, signal models with DM masses of 1 GeV are examined. In a first step, only the selection for the isolated lepton signal region is optimized. On the non-isolated lepton signal region two additional selection steps are imposed, which are $\not{E}_T > 160\,\mathrm{GeV}$ and $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2}) > 1.4$. These requirements are optimized in a second step.

In addition to the baseline selection, all tested selections include a veto on events that contain additional light lepton candidates passing loose identification criteria, $p_T > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$ ($|\eta| < 2.4$) for electron (muon) candidates to reduce the di-leptonic $t\bar{t} +$ jets background processes.

For the optimization of the isolated lepton signal region the following selections are studied:

selection A: $\not{E}_T > 160\,\mathrm{GeV}$, a third jet with $p_T > 50\,\mathrm{GeV}$ and $|\eta| < 2.4$, $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2,3}) > 1.4$

selection B: $\not{E}_T > 320\,\mathrm{GeV}$

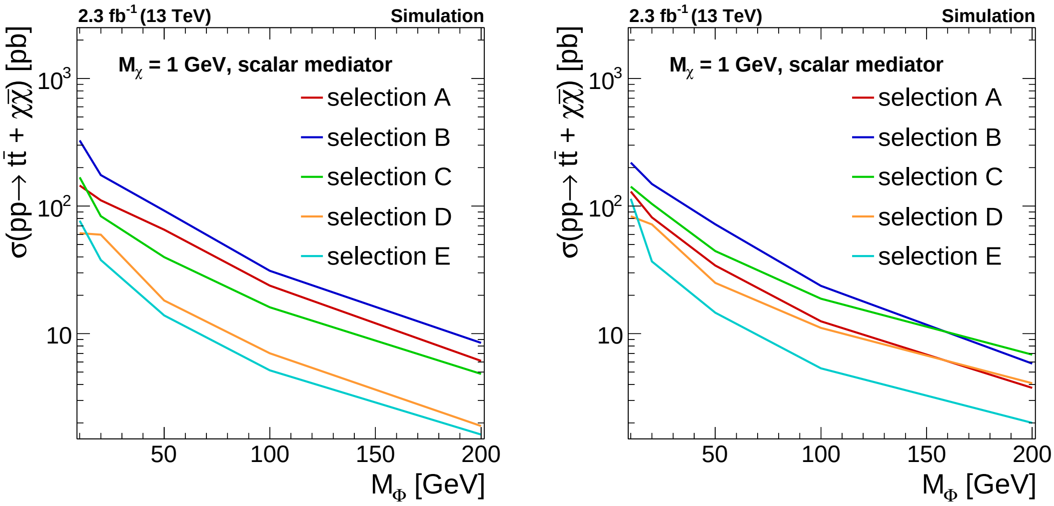
selection C: $m_{T2}^W > 200\,\mathrm{GeV}$, at least one b-tagged jet

selection D: $m_T > 200\,\mathrm{GeV}$, at least one b-tagged jet

selection E: $\not{E}_T > 160\,\mathrm{GeV}$, a third jet with $p_T > 50\,\mathrm{GeV}$ and $|\eta| < 2.4$, at least one b-tagged jet, $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2,3}) > 1.4$, $m_T > 100\,\mathrm{GeV}$, $\Delta\phi(\nu^{\mathrm{rec}}, \not{E}_T) > 1.4$

For the b tagging the medium working point of the CSVv2 algorithm (see Section 5.5.5) is used.

The expected limits obtained after applying the selections A to E are presented in Fig. 8.11. The statistical evaluation for the results presented in Fig. 8.11(a) has been performed using the respective $\not{E}_T$ spectra, while in Fig. 8.11(b) $|\Delta\vec{\not{E}}_{\mathrm{T}}|$ has been used. When studying the $\not{E}_T$ spectra, the best results are obtained with selection E, while selection B yields the worst results. The latter is expected, since the signal efficiency of the high $\not{E}_T$ requirement is relatively low for the studied signal processes and the high signal efficiency loss cannot be compensated by the strong background rejection due to this requirement. The same is true for selection A and selection C. Contrary, good sensitivities can be achieved by applying selection D and selection E, while the latter yields the best results. By applying b tagging a large fraction of background processes



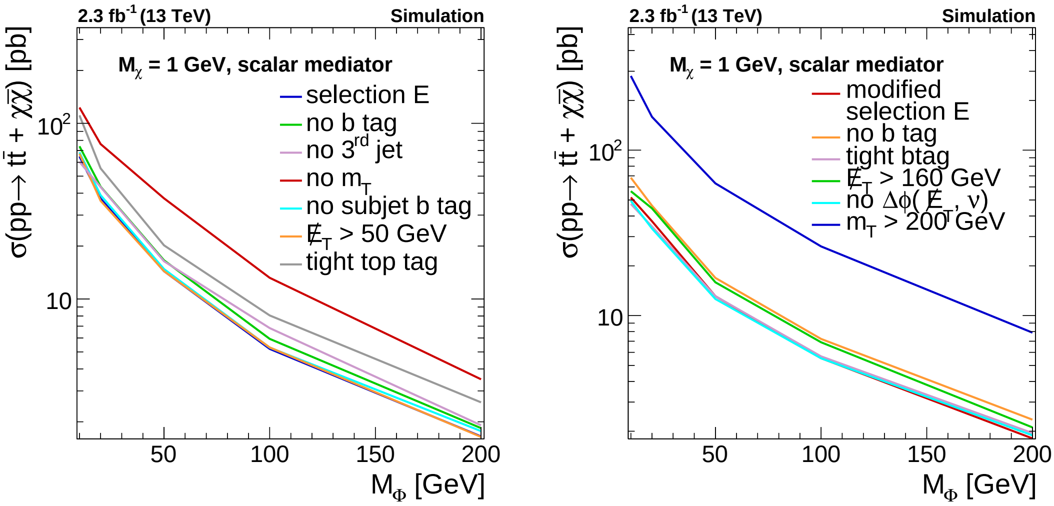
(a) (b)

Figure 8.11: Comparison of the expected upper 95% C.L. cross section limits on the production of $t\bar{t} + \chi\bar{\chi}$ as a function of the mass of the mediator Φ after the application of different selections (*colored lines*). The limits presented in (a) are based on the $\not{E}_T$ spectra after the respective selections have been applied, while the results of (b) are obtained from $|\Delta\vec{\not{E}}_{\mathrm{T}}|$.

that do not contain top quarks is rejected, while the $t\bar{t}$ background can be significantly reduced with the m_T requirement. Comparing the results obtained with $\not{E}_T$ to the ones obtained from $|\Delta\vec{\not{E}}_{\mathrm{T}}|$ for selection E, both variables yield very similar results, except for the lowest mass point where $\not{E}_T$ performs better. To explain the worse performance of selection D based on $|\Delta\vec{\not{E}}_{\mathrm{T}}|$ in comparison to the performance obtained with $\not{E}_T$, the background composition after the application of the m_T requirement is studied. By selecting only events with large values of m_T, most of the $t\bar{t}$+jets background decaying via the lepton+jets decay mode is suppressed, such that almost exclusively di-leptonically decaying $t\bar{t}$ events are left (see Fig. 8.10(c)). For the latter the reconstruction described in Section 8.2 does not work properly, such that large values of $|\Delta\vec{\not{E}}_{\mathrm{T}}|$ are obtained, which drastically reduces the discrimination power of $|\Delta\vec{\not{E}}_{\mathrm{T}}|$. If a high $\not{E}_T$ requirement is included in the selection, like in selection A and selection B, the shape of the $\not{E}_T$ distribution in $t\bar{t}$ background and the signal events get quite similar, such that the mentioned selections perform much better if $|\Delta\vec{\not{E}}_{\mathrm{T}}|$ is used for the statistical evaluation instead of $\not{E}_T$.

In the next optimization steps, selection E is investigated further. Since $|\Delta\vec{\not{E}}_{\mathrm{T}}|$ and $\not{E}_T$ perform similarly for selection E, the following exclusion limits are determined on $\not{E}_T$ only. A more detailed comparison of the characteristics and performance of the observable $|\Delta\vec{\not{E}}_{\mathrm{T}}|$ compared to $\not{E}_T$ is given in Section 8.4.

Since many selection steps are included in selection E, it is checked in the following if all of them are really necessary. Therefore, it is investigated how much the sensitivity changes without the requirement in question. Furthermore, selection steps are tightened and loosened to find the optimal selection. The respective expected upper limits are presented in Fig. 8.12. Due to the top tag requirement the background is already

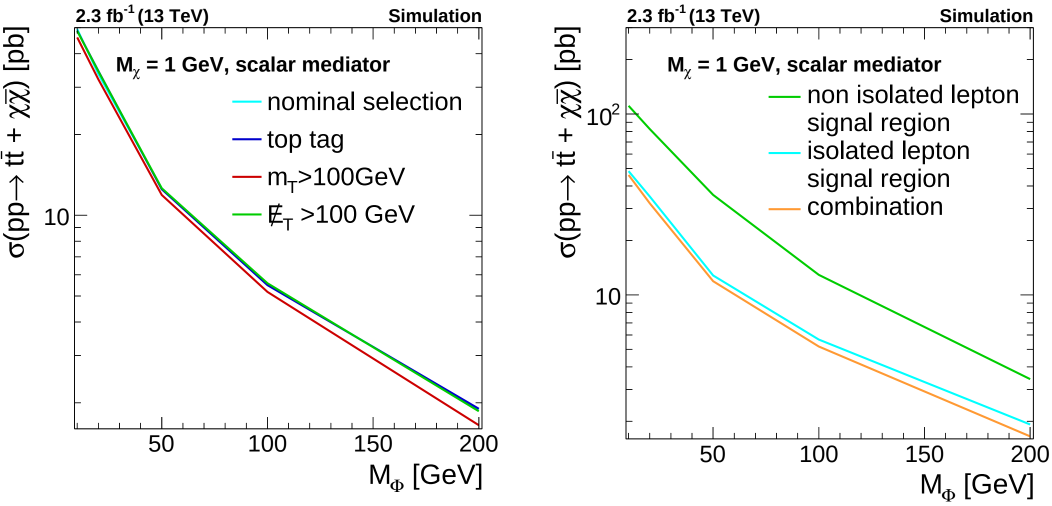


(a) (b)

Figure 8.12: Comparison of the expected upper 95% C.L. cross section limits on the production of $t\bar{t} + \chi\overline{\chi}$ as a function of the mass of the mediator Φ obtained for different selections (*colored lines*). Compared are the sensitivities of selections in which one of the cuts of (a) selection E and (b) the modified selection E is either omitted or tightened. For the statistical evaluation the $\not{E}_T$ spectrum after the respective selection is used.

greatly suppressed. Hence, the $\not{E}_T$ requirement can be lowered to $\not{E}_T > 50\,\mathrm{GeV}$. The third jet requirement does not improve the result significantly and is therefore not necessary. Leaving out the subjet b tag requirement on the top tagged jet or the b tag requirement on small-cone jets yields very similar results. Hence, only one of these cuts is needed. Contrary, the sensitivity deteriorates significantly by omitting the m_T selection or by tightening the top tag requirement. The latter selection leads to a large loss in the signal selection efficiency, while the former requirement is essential for the background suppression. Figure 8.12(b) shows a comparison of different selections to the expected upper limits obtained for selection E without the subjet b tag requirement, including a $\not{E}_T$ requirement of 50 GeV instead of 160 GeV, and without the third jet requirement and therefore also without the $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2,3}) > 1.4$ criteria. This selection is denoted as *modified selection E*. No gain is achieved by tightening the b tag requirement or by including the $\Delta\phi(\nu^{\mathrm{rec}}, \not{E}_T)$ selection. Imposing harder constraints on the m_T or $\not{E}_T$ selection leads to significant losses in the signal selection efficiency, which cannot be compensated by the gain obtained through the rejection of the background processes. Thus, the sensitivity worsens. Hence, requirements on $m_T > 100\,\mathrm{GeV}$ and $\not{E}_T > 50\,\mathrm{GeV}$ are utilized in the final selection. Since omitting the b tag requirement significantly degrades the sensitivity, at least one b-tagged jet is included in the final selection. The medium working point of the CSVv2 algorithm is applied.

In order to optimize the selection in the non-isolated lepton category, the most efficient selections identified above are also tested in this event category. The comparison of the resulting expected limits is presented in Fig. 8.13(a). The *nominal selection* includes a requirement on the presence of at least one medium b-tagged jet, $\not{E}_T > 160\,\mathrm{GeV}$ and $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2}) > 1.4$. Applying a top tag instead of the $\not{E}_T$ and $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2})$ selection



(a) (b)

Figure 8.13: Comparison of the expected upper 95% C.L. cross section limits on the production of $t\bar{t}+\chi\overline{\chi}$ as a function of the mass of the mediator Φ obtained for different selections (*colored lines*). In (a) different selections for the non-isolated signal region are tested, while (b) compares the sensitivity of the isolated lepton signal region, the non-isolated lepton signal region and their combination. The limits are evaluated from the $\not{E}_T$ spectrum after the respective selection.

does not yield any improvement in the sensitivity. Additionally, changing the $\not{E}_T$ requirement from 50 GeV to 100 GeV and leaving out the $\Delta\phi(\not{E}_T, \mathrm{jet}_{1,2})$ requirement also does not change significantly the sensitivity. The best performance is reached by applying a m_T selection of 100 GeV in addition to the b tag requirement. Also in the non-isolated category a great part of the $t\bar{t}$ background processes is rejected by this requirement, while the signal efficiency is still high enough to yield good results. Thus, the final selection identified for the non-isolated signal region is very similar to the one in the isolated lepton signal region.

The final event selection is summarized in the following. To pass the signal selection all events have to meet the following criteria:

- Exactly one muon candidate passing the medium identification requirements, $p_T > 47$ GeV and $|\eta| < 2.4$ is present. Isolation requirements are not applied on the muon candidates.
- Each event contains at least two jets clustered with the anti-k_T algorithm using $R = 0.4$. The jets have to pass $p_T > 50$ GeV and $|\eta| < 2.4$.
- The missing transverse energy has to fulfill $\not{E}_T > 50$ GeV.
- A veto on further light lepton candidates is applied, meaning that events with an electron candidate or further muon candidates that pass loose identification requirements, $p_T > 10$ GeV and $|\eta_{\mathrm{SC}}| < 2.5$ ($|\eta| < 2.4$) for electron (muon) candidates are discarded.

- Each event contains at least one b-tagged jet. For the b tagging the medium working of the CSVv2 algorithm (see Section 5.5.5) is utilized.

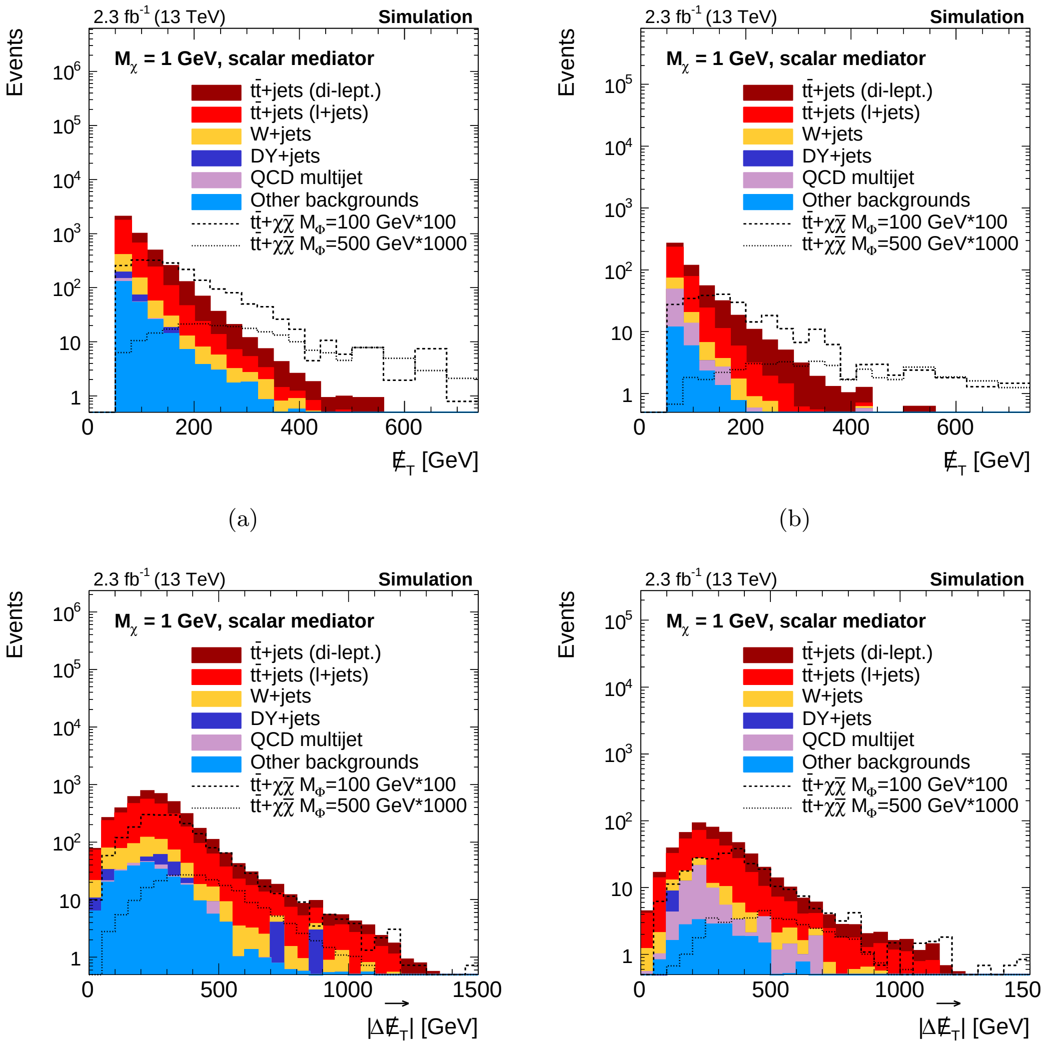
- The transverse mass has to pass $m_T > 100\,\mathrm{GeV}$.

Based on this selection the event categorization defined in Section 8.1 is performed. The $\not{E}_T$ and $|\Delta\vec{\not{E}}_\mathrm{T}|$ spectra after applying the full selection and the event categorization are depicted in Fig. 8.14. The main background processes are $t\bar{t}$ + jets events, which can be split up into lepton+jets and di-leptonic $t\bar{t}$ + jets events. The latter gets particularly important at higher $\not{E}_T$ and $|\Delta\vec{\not{E}}_\mathrm{T}|$ values. Also a small contribution from W+jets events is still present. In contrast to the isolated lepton category, the non-isolated lepton category still contains a very small contribution from QCD multijet processes. The sensitivity of the isolated lepton category and the non-isolated category is compared in Fig. 8.13(b). Additionally, the sensitivity of the combined analysis is presented. The exclusion limits have been obtained from the $\not{E}_T$ spectrum presented in Fig. 8.14. The sensitivity of the non-isolated lepton category is clearly worse that that of the isolated lepton category. This can be explained by the lower statistical precision available in the non-isolated lepton category. But still, a significant gain in sensitivity can be achieved by combining the two categories.

The presented event selection is applied on data events collected by the CMS experiment at 13 TeV with slightly adopted identification criteria as well as p_T and η restrictions, which have to be changed to match the conditions of the 2016 data acquisition period. The performed search is presented in the second part of this chapter (Sections 8.6 to 8.17).

8.4 Conclusions of the optimization studies

The shape comparison of the observable $|\Delta\vec{\not{E}}_\mathrm{T}|$ in the $t\bar{t}$+jets background decaying in the lepton+jets decay mode and different signal models, presented in Fig. 8.9, suggests that a variable with better separation power between signal and background processes than $\not{E}_T$ has been found by the kinematic reconstruction method explained in Section 8.2. Especially for signal models with high mediator or high DM masses, a clear separation between background and signal processes is achieved. However, the separation power of $|\Delta\vec{\not{E}}_\mathrm{T}|$ deteriorates when applying the optimized event selection (see Fig. 8.14). Hence, the expected sensitivity is significantly weaker when using $|\Delta\vec{\not{E}}_\mathrm{T}|$ in the statistical interpretation of the search than when using $\not{E}_T$. The same statement holds true for an extrapolation of the sensitivity to the data set collected in 2017. This is illustrated in Fig. 8.15, where the sensitivities of the two approaches are compared using integrated luminosities of $2.3\,\mathrm{fb}^{-1}$ and $40\,\mathrm{fb}^{-1}$. Three different effects are identified that lead to the deterioration of the search sensitivity. A small contribution stems from the background composition after the application of the optimized signal selection. The chosen selection criteria mainly reduce $t\bar{t}$ + jets events decaying in the lepton+jets decay mode, while di-leptonic $t\bar{t}$+jets events are mostly kept (see Fig. 8.10 for the background compositions of some variables used in the final event selection). Due to two neutrinos being present



(a)

(b)

(c)

(d)

Figure 8.14: Comparison of the background (*colored areas*) and different signal models (*black lines*) assuming a DM mass of 1 GeV and different mediator masses of (a) the $\not\!\!E_T$ spectra in the isolated lepton signal region and (b) the non-isolated lepton signal region, and of (c) $|\Delta\vec{\not\!\!E}_\mathrm{T}|$ in the isolated lepton signal region and (d) the non-isolated lepton signal region after applying the optimized event selection.

in di-leptonic $t\bar{t}$ + jets and thus larger values of $\not\!\!E_T$, this background is very difficult to distinguish from signal events and even harder to reduce than the lepton+jets $t\bar{t}$ background. Di-leptonic $t\bar{t}$ + jets events tend to populate the high $|\Delta\vec{\not\!\!E}_\mathrm{T}|$ region since the kinematic reconstruction method is not designed for this decay mode and fails in these cases. Despite of the applied lepton veto, no further selection requirements were found to reduce the di-leptonic $t\bar{t}$ background, while keeping the signal events.

Secondly, the changed $\not\!\!E_T$ requirement has a strong impact on the separation power and thus the sensitivity obtained using $|\Delta\vec{\not\!\!E}_\mathrm{T}|$. Due to the top tag requirement included in the final event selection, the background processes are notably reduced. Therefore, the $\not\!\!E_T$ requirement can be lowered from $\not\!\!E_T > 160$ GeV to $\not\!\!E_T > 50$ GeV, which significantly increases the signal selection efficiency. However, for events with $\not\!\!E_T > 50$ GeV the kine-

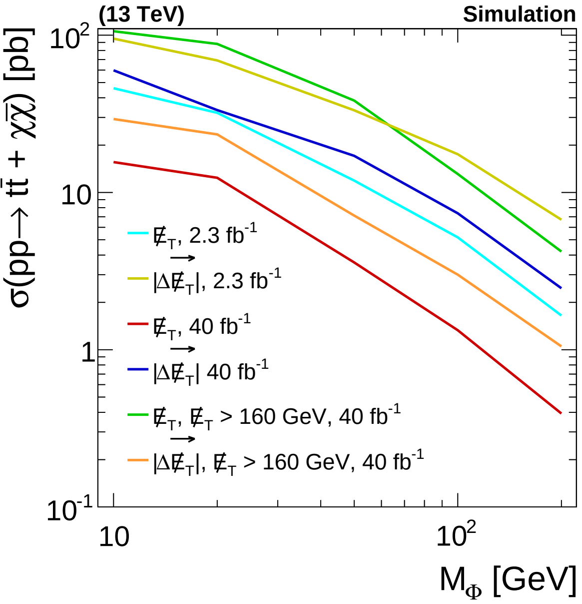


Figure 8.15: Comparison of the expected upper 95% C.L. cross section limits on the production of $t\bar{t}+\chi\bar{\chi}$ as a function of the mass of the mediator Φ determined based on $\not{E}_T$ or $|\Delta\vec{\not{E}}_\mathrm{T}|$. The sensitivities obtained by applying the optimized event selection and using 40 fb^{-1} of data are compared to the sensitivities obtained by applying the optimized event selection with a $\not{E}_T$ requirement of 160 GeV in addition and the sensitivities obtained for 2.3 fb^{-1}.

matic reconstruction does not work properly since these events are not boosted enough for the constraints used in the likelihood distribution to be applicable. Therefore, the sensitivity obtained with $|\Delta\vec{\not{E}}_\mathrm{T}|$ deteriorates, while the sensitivity obtained with $\not{E}_T$ improves due to the higher signal selection efficiency.

A third reason for the deterioration of the result is the applied m_T requirement. By imposing the requirement $m_T > 100$ GeV only misreconstructed $t\bar{t}$ events are selected, for which the reconstruction does not work. Thus, the peak of $|\Delta\vec{\not{E}}_\mathrm{T}|$ is shifted towards much higher values by applying the m_T selection. Still, the m_T requirement greatly reduces the background processes and thus improves the sensitivity obtained with $\not{E}_T$. Additionally, for a low $\not{E}_T$ requirement the $\not{E}_T$ spectra also exhibit a shape difference between signal and background processes (see Fig. 8.14). Due to the described effects, the idea to use $|\Delta\vec{\not{E}}_\mathrm{T}|$ for the final statistical interpretation of this search has to be abandoned. However, if a $\not{E}_T$ requirement of 160 GeV is applied in addition to the optimized event selection, which is the value the reconstruction has been designed for, $|\Delta\vec{\not{E}}_\mathrm{T}|$ yields much better sensitivity than $\not{E}_T$ (see Fig. 8.15). Thus, the likelihood reconstruction works very well in the phase space it was originally designed for and it seems to be a promising method for signal models with high mediator and high DM masses as well as pseudoscalar models (see Fig. 8.9). Nonetheless, to be applicable in a realistic analysis with background processes other than pure $t\bar{t}$+jets production present, improvements to the likelihood reconstruction and the signal selection are necessary. One way to identify a selection that yields high separation power between signal and background processes could be to use MVA techniques. This way, a selection can possibly be found without usage of m_T and other $\not{E}_T$ related variables, which should notably improve the sensitivity

obtained with $|\Delta\vec{\not{E}}_{\mathrm{T}}|$. Furthermore, the likelihood reconstruction can be improved. In some cases, events feature two minima in the χ^2 function. In the current implementation this is not accounted for, leading to the wrong minimum being selected in some events. Additionally, one could try to optimize the likelihood for slightly lower $\not{E}_T$ requirements. However, the $\not{E}_T$ criteria can only be decreased to a certain extent since the events need to be boosted for the method to work. Furthermore, including the p_T of the lepton into the likelihood function could help to improve the method.
To conclude, a promising method and observable have been found, which especially promise enhanced sensitivity for signal models with higher DM and mediator masses. Nonetheless, improvements to the likelihood reconstruction and event selection are needed in order to be able to potentially outperform the adopted strategy of the search presented in this thesis.

8.5 Comparison to the resolved analysis

To access how the presented search strategy compares to already existing results, a comparison of the sensitivity of this analysis to the resolved lepton+jets analysis of EXO-16-005 presented in [3, 4] is performed. Since only the combined results of the electron and muon channel after the application of systematic uncertainties are available from EXO-16-005, the main selection steps of EXO-16-005 (see [3, 4] and Fig. 2.6) are rebuilt to enable a comparison of the two analyses. The mentioned selection steps are listed hereafter:

- Each event contains exactly one muon candidate with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 2.1$ that passes the tight muon identification and isolation requirements (see Section 5.3, the tight identification requirements of Run-II match those used in Run-I).
- No further electron (muon) candidates fulfilling loose identification and isolation criteria, $p_T > 10\,\mathrm{GeV}$, and $|\eta| < 2.5$ ($|\eta| < 2.1$) are found.
- At least three jets passing the loose PF identification criteria with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$ are present. The jets are clustered with the anti-k_T algorithm and and a distance parameter of $R = 0.4$.
- At least one of the jets has to be identified as a b quark candidate jet. For the b tagging, the medium working point of the CSVv2 algorithm (see Section 5.5.5) is applied.
- Each event passes $\not{E}_T > 160\,\mathrm{GeV}$, $m_T > 160\,\mathrm{GeV}$ and $m_{T2}^W > 200\,\mathrm{GeV}$.
- The minimum angle $\Delta\phi$ between the leading/sub-leading p_T-jet and $\not{E}_T$ has to exceed 1.2.

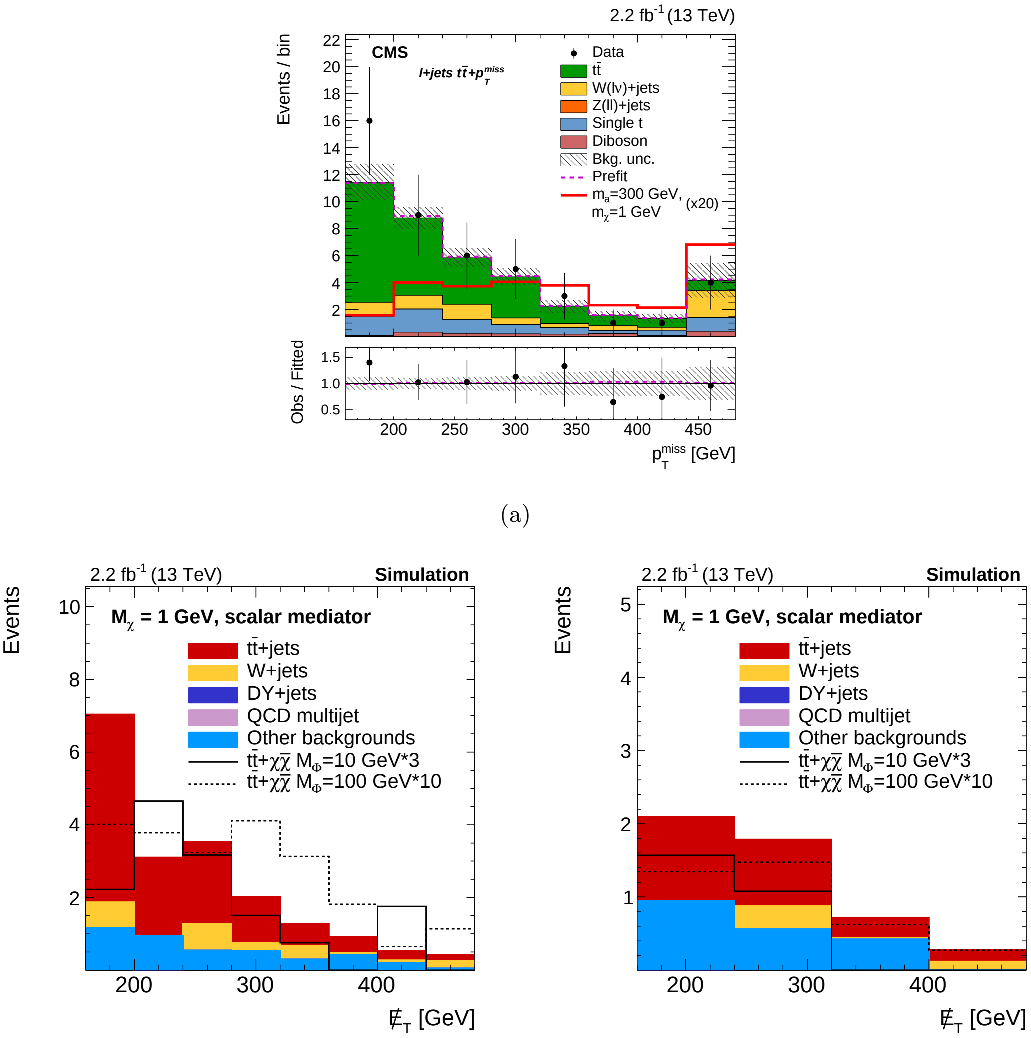
It has to be noted that only these basic selection requirements of EXO-16-005 have been adopted for this study. Neither a comparison of the event yields has been performed, nor the exact object definition and background estimation strategy has been copied. The study is based on simulated events only.

The main differences between the selection criteria applied in EXO-16-005 and in the boosted search are tighter requirements on $\not{E}_T$ and m_T as well as the application of selection criteria on m_{T2}^W and $\Delta\phi(\not{E}_T, \text{jet})$ in EXO-16-005. Additionally, tighter identification and isolation criteria are enforced on the muon candidates in the latter analysis, while no top tagging is applied. Instead the presence of three jets is enforced with looser p_T thresholds in comparison to the two jets required in the boosted search. Trigger requirements are not implemented for both analyses.

The $\not{E}_T$ distribution presented in EXO-16-005 [3, 4] and the $\not{E}_T$ spectrum obtained by rebuilding the selection criteria of the resolved analysis are shown in Figs. 8.16(a) and 8.16(b). Both distributions are scaled to an integrated luminosity of $2.2\,\text{fb}^{-1}$. Since for the result of EXO-16-005 the electron and muon channel have been combined, the observed difference in the event yield of the two analyses is expected. The difference is a bit less than a factor of two due to tight requirements imposed on electron candidates in the electron channel. Thus, the distribution in Fig. 8.16(b) looks reasonable and seems to roughly match the results obtained in EXO-16-005.

The sensitivity of the two analyses is compared by determining the expected upper 95% C.L. limit on the $t\bar{t} + \chi\overline{\chi}$ production cross section as a function of the mass of the mediator. The mass of the DM particles is set to 1 GeV. The comparison of the expected limits is presented in Fig. 8.17. A comparison of the expected limit obtained by rebuilding EXO-16-005 to the numbers provided by [3,4] (see Fig. 2.6) reveals that the expected limits differ by a factor of approximately two. This is reasonable since in this study only the muon channel is studied and high rate uncertainties (as listed in Section 6.8) on SM background processes are assumed, while in [3, 4] control regions were used to constrain the normalization of background processes more precisely. In conjunction with the comparison of the $\not{E}_T$ distributions discussed before (see Fig. 8.16), it is concluded that the implementation of EXO-16-005 in the described way is reasonable and enables a comparison of the sensitivity of the two analyses. The boosted analysis yields better sensitivity for low mediator masses, while for mediator masses higher than 60 GeV the resolved analysis gives comparable results. The resolved analysis starts to outperform the boosted search for masses above 100 GeV. This behavior of the resolved analysis with respect to the boosted search is mainly explainable with the high $\not{E}_T$ requirement implemented in EXO-16-005, which results in a very low selection efficiency for signal events with small mediator masses (see Fig. 8.2) and thus deteriorates the expected limit in the low mass region. Contrary, the boosted search can adopt a much looser $\not{E}_T$ requirement because of the application of top tagging. The latter requirement significantly suppresses SM background processes, while a high signal selection efficiency is retained. In the high mass region, the tight $\not{E}_T$ criterion of the resolved analysis drastically reduces the SM background processes, while the signal selection efficiency is still high. Thus, the sensitivity of the resolved analysis improves in comparison to the boosted search, which features a higher contamination with background processes due to the much lower $\not{E}_T$ criterion. The tight selection requirements on m_T and m_{T2}^W in the resolved analysis affect the sensitivity in a similar way.

In addition to the comparison of the sensitivity of the two analyses, it is estimated how



(a)

(b)

(c)

Figure 8.16: (a) $\not{E}_T$ distribution obtained by the analysis EXO-16-005, the figure is taken from [3, 4], (b) $\not{E}_T$ distribution obtained by rebuilding the main selection steps of the resolved analysis EXO-16-005 [3, 4] and (c) $\not{E}_T$ distribution obtained by rebuilding the main selection steps of the analysis EXO-16-005 [3, 4], whereby all events that contain a top tagged jet have been removed. In all distributions the SM background processes (*colored areas*) can be compared to the expected spectra of signal models (*red line in (a), dashed lines in (b) and (c)*). In (a) also data events are presented (*black dots*). All distributions are scaled to 2.2 fb^{-1}.

much improvement in sensitivity can be achieved by combining the two analyses. To make the two analyses statistically independent, the event overlap has to be removed. Thus, all events that contain a top tagged jet are discarded in the resolved analysis. The resulting $\not{E}_T$ distribution is presented in Fig. 8.16(c). Due to the high $\not{E}_T$ requirement in the resolved analysis, the selected events exhibit a relatively large Lorentz boost and many of them pass the applied top tagging criteria. Consequently, only few signal as well as background events are left in the resolved analysis after removing the overlapping events, which deteriorates the expected limit significantly with respect to the former resolved analysis (see Fig. 8.17). Hence, only a very small gain in sensitivity is achieved

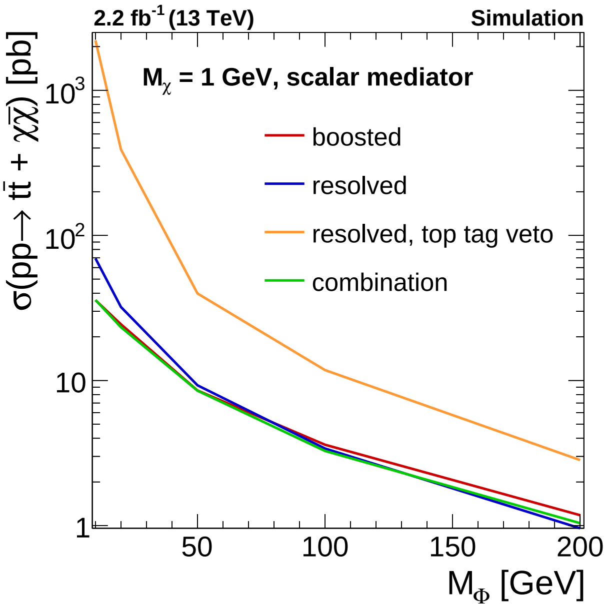


Figure 8.17: Comparison of the expected 95% C.L. upper cross section limit on the production of DM in association with a top quark pair obtained for the boosted search (*red line*), the analysis rebuilt according to [3, 4] (*blue line*), the rebuilt resolved analysis with top tag veto (*orange line*) and the combined resolved and boosted result (*green line*).

by combining the two analyses.

The same conclusions can be drawn by studying the sensitivity for signal models with DM particles with masses of 10 GeV, while for signal models with DM particles with masses of 50 GeV the resolved analysis slightly outperforms the boosted search. To check whether the same conclusions are valid with higher statistical precision, the same study has been performed with distributions scaled to 40 fb^{-1}. No differences in the behaviors of the two analyses have been observed.

In summary, these studies show that with the analysis presented in this thesis the sensitivity at low mediator masses can be significantly improved in comparison to published results. The loss in sensitivity of the boosted search with respect to the resolved analysis at high masses is quite low. A combination of the boosted search with an analysis like EXO-16-005 only yields small improvements with respect to the sensitivity of the boosted search only.

Analysis with 2016 data

8.6 Data samples and trigger

After the design of the analysis through simulation studies, the search for DM produced in association with a top quark pair is performed using pp collision data recorded at $\sqrt{s} = 13\,\text{TeV}$ by the CMS experiment. The runs B, C and D of the data acquisition periods of the year 2016 are analyzed. The amount of data corresponds to an integrated luminosity of $12.9\,\text{fb}^{-1}$ [81, 190]. The LHC machine parameters valid for the considered period are summarized in Section 3.3.

As discussed in Section 8.1, the studied search regions also contain muon candidates that fail the isolation criterion. Thus, only triggers that do not enforce isolation requirements on the muon candidates are used. Two different triggers are chosen. The first, denoted as *HLT_Mu50* within the CMS experiment, requires each event to contain a Global Muon candidate with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.4$ on trigger level. After data corresponding to an integrated luminosity of $622\,\text{pb}^{-1}$ had been collected in Run B, the second trigger, called *HLT_TkMu50*, became available. The trigger HLT_TkMu50 records events that contain a Tracker Muon candidate with $p_T > 50\,\text{GeV}$ on trigger level. To be considered in the analysis each event has to pass at least one of the two triggers. Since the studied simulated samples do not contain a simulation of the trigger menu, the trigger selection is only applied on data. The trigger performances have been studied by a group within the CMS experiment responsible for muon triggering, reconstruction and identification [191]. The trigger HLT_Mu50 reaches a plateau efficiency of approximately 85% for muon candidates with $p_T \gtrsim 50\,\text{GeV}$ when studied inclusively in η. By combining the two triggers, this plateau efficiency improves to a value of about 90%. In both cases, the efficiency drops significantly for high-p_T muon candidates that are found in high $|\eta|$ regions [191]. For muon candidates with $p_T > 500\,\text{GeV}$ the determination of correction factors to propagate these inefficiencies to simulation is not possible due to limited available statistical precision in data. Therefore, events that contain muon candidates with $p_T > 500\,\text{GeV}$ that have a pseudorapidity of $|\eta| > 1.2$ are discarded. For the remaining phase space, correction factors that model the data trigger efficiency in simulation are applied. They are discussed in Section 8.11.

The integrated luminosities corresponding to the three analyzed SingleMuon data streams are summarized in Table 8.1.

data set	run	L_{int} [fb^{-1}]
/SingleMuon/Run2016B-PromptReco-v2/MINIAOD	2016 B	5.892
/SingleMuon/Run2016C-PromptReco-v2/MINIAOD	2016 C	2.646
/SingleMuon/Run2016D-PromptReco-v2/MINIAOD	2016 D	4.353

Table 8.1: Summary of the integrated luminosities of the SingleMuon data streams in the three studied data acquisition periods of the CMS experiment at $\sqrt{s} = 13\,\text{TeV}$ [81, 190].

8.7 Simulated background and signal processes

Simulated samples are used to describe the SM background processes and to model the expected signal events. SM $t\bar{t}+$jets events as well as t-channel single top quark production and single top quark production in association with a W boson have been generated using POWHEG-BOX [95–100]. S-channel single top quark production has been simulated with MadGraph5_aMC@NLO(v5.2) [94], which has also been utilized to model $\mathrm{DY}(\rightarrow \ell\ell)+$jets events with an invariant mass of the two leptons $m(\ell\ell)$ between 10 GeV and 50 GeV, $t\bar{t}$ production in association with a W or Z boson and $W(\rightarrow \ell\nu_\ell)+$jets production. $\mathrm{DY}(\rightarrow \ell\ell)+$jets events with $m(\ell\ell) > 50\,\mathrm{GeV}$ and $W(\rightarrow \ell\nu_\ell)+$jets events have been generated, divided into different ranges of H_T[1] covering values larger than 100 GeV, using MadGraph (v5.2) [94]. To avoid overlap between the inclusive $W+$jets sample and the sample split in H_T, only events with $H_\mathrm{T} < 100\,\mathrm{GeV}$ are considered from the inclusive $W+$jets sample. QCD multijet processes, divided into different $\hat{p}_\mathrm{T}$ regions, and diboson events have been simulated with PYTHIA (v8) [101, 102].

The signal processes have been generated with up to one additional parton in the matrix element according to the simplified models discussed in [49, 53] (described in Section 2.5.2) using MadGraph (v5.2). Scalar mediators and pseudoscalar mediators have been simulated. Masses between 10 GeV and 500 GeV are considered for the mediator particles, while for the dark matter particle masses of 1 GeV, 10 GeV and 50 GeV are assumed. A summary of the studied mass points for scalar and pseudoscalar mediators as well as the respective production cross sections can be found in Table 8.2. To model the angular distributions between the decay products of the top quarks correctly MadSpin [103, 104] is utilized.

For all background and signal processes the parton shower and hadronization process is described using PYTHIA (v8). In events generated with MadGraph the MLM matching [108] scheme is used for the association of the partons emerging from the hard interactions to the particles produced in the hadronization process. In $\mathrm{DY}(\rightarrow \ell\ell)+$jets events, where $10\,\mathrm{GeV} < m(\ell\ell) < 50\,\mathrm{GeV}$, $t\bar{t}$ production in association with a W boson and in the inclusive $W+$jets sample the matching is performed with the FxFx merging scheme [109]. The pdf predictions for all samples are taken from the NNPDF3.0 pdf sets [90]. Geant4 [111] is employed to describe the interactions of all simulated particles with the detector material.

The simulated samples are scaled according to their cross section to the total integrated luminosity. For this re-weighting, the most accurate available cross section for the respective process is used. The cross sections are summarized for all background processes, together with the order of the calculation and the used event generator, in Tables 8.3 and 8.4.

8.8 Event filters

Only events certified as recorded with a completely functional detector are studied in the analysis. Nevertheless, it is particularly important for this analysis to apply event filters,

---

[1] H_T describes the scalar sum of the transverse momenta of all jets above a certain p_T-threshold.

		M_Φ [GeV] (scalar)								
		10	15	20	50	95	100	200	300	500
M_χ [GeV]	1	21.36 pb		10.95 pb	3.088 pb		0.7205 pb	0.1016 pb	30.45 fb	4.947 fb
	10	0.1011 pb	0.1279 pb		3.097 pb		0.7417 pb			
	50	2.078 fb			2.567 fb	7.202 fb		0.1003 pb	30.46 fb	

		M_Φ [GeV] (pseudoscalar)								
		10	15	20	50	95	100	200	300	500
M_χ [GeV]	1	0.4517 pb		0.4117 pb	0.308 pb		0.1932 pb	87.86 fb	39.5 fb	5.163 fb
	10	15.22 fb	19.5 fb		0.3091 pb		0.1976 pb			
	50	2.405 fb			2.928 fb	10.72 fb		84.76 fb	38.45 fb	

Table 8.2: Summary of the considered signal models assuming simplified models according to [49,53] and of the corresponding production cross sections at LO determined with MadGraph (v5.2.) [94]

process	generator	σ [pb]	order theory calc.
$t\bar{t} + \text{jets}$	POWHEG-BOX	831.76	NNLO [192]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, inclusive	MadGraph5_aMC@NLO	61526.7	NLO [94]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, $100\,\text{GeV} < H_\text{T} < 200\,\text{GeV}$	MadGraph	1627.45	LO [94], NNLO k factor of 1.21 included [193]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, $200\,\text{GeV} < H_\text{T} < 400\,\text{GeV}$	MadGraph	435.237	LO [94], NNLO k factor of 1.21 included [193]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, $400\,\text{GeV} < H_\text{T} < 600\,\text{GeV}$	MadGraph	59.1811	LO [94], NNLO k factor of 1.21 included [193]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, $600\,\text{GeV} < H_\text{T} < 800\,\text{GeV}$	MadGraph	14.5805	LO [94], NNLO k factor of 1.21 included [193]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, $800\,\text{GeV} < H_\text{T} < 1200\,\text{GeV}$	MadGraph	6.6562	LO [94], NNLO k factor of 1.21 included [193]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, $1200\,\text{GeV} < H_\text{T} < 2500\,\text{GeV}$	MadGraph	1.6081	LO [94], NNLO k factor of 1.21 included [193]
$W(\rightarrow \ell\nu_\ell) + \text{jets}$, $H_\text{T} > 2500\,\text{GeV}$	MadGraph	0.0389	LO [94], NNLO k factor of 1.21 included [193]
$Z(\rightarrow \ell\ell)+$ jets ($10\,\text{GeV} < m(\ell\ell) < 50\,\text{GeV}$)	MadGraph5_aMC@NLO	18610	NLO [94]
$Z(\rightarrow \ell\ell)+$ jets ($m(\ell\ell) > 50\,\text{GeV}$), $100\,\text{GeV} < H_\text{T} < 200\,\text{GeV}$	MadGraph	181.3020	LO [94], NNLO k factor of 1.23 included [193]
$Z(\rightarrow \ell\ell)+$ jets ($m(\ell\ell) > 50\,\text{GeV}$), $200\,\text{GeV} < H_\text{T} < 400\,\text{GeV}$	MadGraph	50.4177	LO [94], NNLO k factor of 1.23 included [193]
$Z(\rightarrow \ell\ell)+$ jets ($m(\ell\ell) > 50\,\text{GeV}$), $400\,\text{GeV} < H_\text{T} < 600\,\text{GeV}$	MadGraph	6.9839	LO [94], NNLO k factor of 1.23 included [193]
$Z(\rightarrow \ell\ell)+$ jets ($m(\ell\ell) > 50\,\text{GeV}$), $H_\text{T} > 600\,\text{GeV}$	MadGraph	2.7035	LO [94], NNLO k factor of 1.23 included [193]
Single t, s channel	MadGraph5_aMC@NLO	3.36	NLO [194, 195]
Single t, t channel	POWHEG-BOX	44.33	NLO [194, 195]
Single $\bar{t}$, t channel	POWHEG-BOX	26.38	NLO [194, 195]
Single t, tW production	POWHEG-BOX	35.6	approx. NNLO [196]
Single $\bar{t}$, tW production	POWHEG-BOX	35.6	approx. NNLO [196]

Table 8.3: Summary of the considered background processes. Shown are the event generator used to simulate the events, the cross sections taken for the weighting of the processes, and the order to which the calculation of the respective cross section has been performed.

process	generator	σ [pb]	order theory calc.
WW	PYTHIA	118.7	NNLO QCD [197]
ZZ	PYTHIA	16.523	NLO from MCFM [198–200]
WZ	PYTHIA	47.13	NLO from MCFM [198–200]
$t\bar{t}\,W$ ($W \to q\bar{q}'$)	MadGraph5_aMC@NLO	0.4062	NLO [94]
$t\bar{t}\,W$ ($W \to \ell\nu_\ell$)	MadGraph5_aMC@NLO	0.2043	NLO [94]
$t\bar{t}\,Z$ ($Z \to q\bar{q}$)	MadGraph5_aMC@NLO	0.5297	NLO [94]
$t\bar{t}\,Z$ ($Z \to \ell\ell$ and $Z \to \nu\nu$)	MadGraph5_aMC@NLO	0.2529	NLO [94]
QCD (muon enriched), 15 GeV $< \hat{p}_\mathrm{T} <$ 20 GeV	PYTHIA	3819570	LO, filter efficiency included [101, 102]
QCD (muon enriched), 20 GeV $< \hat{p}_\mathrm{T} <$ 30 GeV	PYTHIA	2960198.4	LO, filter efficiency included [101, 102]
QCD (muon enriched), 30 GeV $< \hat{p}_\mathrm{T} <$ 50 GeV	PYTHIA	1652471.46	LO, filter efficiency included [101, 102]
QCD (muon enriched), 50 GeV $< \hat{p}_\mathrm{T} <$ 80 GeV	PYTHIA	437504.1	LO, filter efficiency included [101, 102]
QCD (muon enriched), 80 GeV $< \hat{p}_\mathrm{T} <$ 120 GeV	PYTHIA	106033.6648	LO, filter efficiency included [101, 102]
QCD (muon enriched), 120 GeV $< \hat{p}_\mathrm{T} <$ 170 GeV	PYTHIA	25190.5151	LO, filter efficiency included [101, 102]
QCD (muon enriched), 170 GeV $< \hat{p}_\mathrm{T} <$ 300 GeV	PYTHIA	8654.4932	LO, filter efficiency included [101, 102]
QCD (muon enriched), 300 GeV $< \hat{p}_\mathrm{T} <$ 470 GeV	PYTHIA	797.3527	LO, filter efficiency included [101, 102]
QCD (muon enriched), 470 GeV $< \hat{p}_\mathrm{T} <$ 600 GeV	PYTHIA	79.0255	LO, filter efficiency included [101, 102]
QCD (muon enriched), 600 GeV $< \hat{p}_\mathrm{T} <$ 800 GeV	PYTHIA	25.0951	LO, filter efficiency included [101, 102]
QCD (muon enriched), 800 GeV $< \hat{p}_\mathrm{T} <$ 1000 GeV	PYTHIA	4.7074	LO, filter efficiency included [101, 102]
QCD (muon enriched), $\hat{p}_\mathrm{T} >$ 1000 GeV	PYTHIA	1.6213	LO, filter efficiency included [101, 102]

Table 8.4: Summary of the considered background processes. Shown are the event generator used to simulate the events, the cross sections taken for the weighting of the processes, and the order to which the calculation of the respective cross section has been performed.

which discard events where problems in the data acquisition have been found after the certification procedure. Otherwise, malfunctionings of the detector, mismeasurements of energies and misreconstructed particles can lead to the creation of spurious $\not{E}_T$, which can be confused with the $\not{E}_T$ produced by signal events. All applied event filters are listed in the following [155, 201, 202].

HBHENoiseFilter: Energy deposits in the HB and HE are classified as noise by exploiting time and shape information of the measured pulses as well as geometrical structures of the signals. The identified events are discarded.

HBHENoiseIsoFilter: By studying energy deposits in the ECAL, the HCAL and information from the tracking system, events with isolated noise deposits in the HB and HE are found and the corresponding events are omitted.

globalTightHalo2016Filter: Spurious signals in the detector are produced by interactions of the protons in the beam with residual gas molecules or machine parts of the LHC. These events very likely produce muons, which are in some cases detected within the experiment. Since the produced muons have a high probability to traverse the experiment on a trajectory parallel to the beam axis, such events are identified by multiple calorimeter energy deposits that share the same ϕ-coordinate and CSC signals associated to these.

ECALDeadCellTriggerPrimitiveFilter: In some towers of the ECAL a precise readout of the deposited energy is not possible. Instead, information from the trigger system is used to estimate the deposited energy. The measurements of the trigger system saturate at a certain energy level. Events that contain energy deposits in the broken towers close to this threshold are discarded.

eeBadScFilter: Certain arrays of 5×5 crystals in the endcaps of the ECAL are known to randomly produce anomalous high signals. Affected events are identified by imposing a requirement on the maximum allowed measured energy and removed if this threshold is exceeded.

chargedHadronTrackResolutionFilter: Events that contain high-p_T charged hadrons whose track resolution yields a very bad quality are removed if the $\not{E}_T$ significance [143] indicates that the measured $\not{E}_T$ in the event is mainly induced by the affected charged hadron.

muonBadTrackFilter: High-p_T muons with a low-quality track can be falsely identified as charged hadrons and thus be wrongly considered in the calculation of $\not{E}_T$. Affected events are identified through quality criteria on the tracks and removed.

8.9 Object quality criteria

All objects have to pass certain quality criteria to be considered in the analysis, which are based on the reconstruction algorithms and identification criteria described in Chapter 5.

They are listed below. In all cases, the recommendations of the responsible group within the CMS experiment are followed.

- The reconstruction of primary vertex candidates is performed as described in Section 5.2. All candidates have to be *good primary vertices*, meaning that the identification criteria mentioned in Section 5.2 have to be passed by all of them. Events that do not contain at least one good primary vertex candidate are discarded.
- All muon candidates have to pass the loose identification criteria for muon identification, $p_T > 10\,\text{GeV}$ and $|\eta| < 2.4$. No isolation criterion is imposed on the studied muon candidates.
- The electron identification criteria for Run-II described in Section 5.4 have to be passed by all studied electron candidates. Furthermore, only candidates with $p_T > 10\,\text{GeV}$ and $|\eta_{\text{SC}}| < 2.5$ are kept. No isolation criterion is imposed.
- Two different types of jets are studied. The pileup mitigation is performed using charge hadron subtraction in both cases. Furthermore, the JEC and the correction factors that align the JER in simulation and data are applied. The first jet collection contains jets clustered with the anti-k_T algorithm from PF candidates with a distance parameter of $R = 0.4$. Furthermore, these jets have to pass the loose identification criteria for PF jets, $|\eta| < 2.4$ and their corrected transverse momentum has to exceed $30\,\text{GeV}$. Jets passing these criteria will be denoted as *ak4-jets* in the following. Since the selected non-isolated light lepton candidates might have been clustered into these jets, a *jet-lepton cleaning* is performed. In this procedure, the four-momentum of a light lepton candidate that lies within $\Delta R < 0.4$ to a jet is subtracted from the uncorrected four-momentum of the jet if the lepton energy clustered into the jet exceeds a certain fraction of the jet energy. The jet energy corrections are reapplied afterwards.
  The second jet collection consists of jets clustered from PF particles with the HTT V2 algorithm as described in Section 7.3. A p_T criterion of $p_T > 150\,\text{GeV}$ and a requirement on the pseudorapidity of $|\eta| < 2.4$ is enforced on all these jets. Jets fulfilling the mentioned requirements are referred to as *HTT-jets.*
- The Type-1-PF description of $\not{E}_T$ is applied.

Since only particles produced in the hard interaction are of interest, objects whose tracks are not associated with the primary vertex candidate are discarded.

8.10 Event selection

The event selection follows the considerations and results of the studies using simulated events in the first part of this chapter. It is based on all objects that pass the quality criteria and is divided into a *pre-selection* and a *signal selection.* The pre-selection is designed with relatively loose requirements, such that the contribution of expected signal events is negligible compared to the total number of predicted background events at

this stage of the selection. Hence, the pre-selection allows a validation of the description of data by simulation (see Section 8.12). The final selection steps suppress the SM backgrounds considerably, while they enhance the search sensitivity significantly. Both selections are summarized below.
In order to pass the pre-selection all events have to satisfy the criteria hereafter.

- Each event contains exactly one muon candidate. This muon candidate has to pass the medium identification requirements, $p_T > 53\,\mathrm{GeV}$ and $|\eta| < 2.4$. The selection requirements ensure that the chosen muon candidate lies in the phase space where the trigger selection is fully efficient. No isolation criterion is imposed on the muon candidate. This selection step also implies a veto of events containing further muon candidates that pass the loose identification requirements, $p_T > 10\,\mathrm{GeV}$ and $|\eta| < 2.4$.
- At least two ak4-jets with $p_T > 50\,\mathrm{GeV}$ and $|\eta| < 2.4$ are present.
- Each event passes $\not{E}_T > 80\,\mathrm{GeV}$.
- If an event contains an electron candidate that fulfills the loose identification requirements $p_T > 10\,\mathrm{GeV}$ and $|\eta_{\mathrm{SC}}| < 2.5$, it is discarded.

The requirement on $\not{E}_T$ is tightened in comparison to the selection presented in Section 8.3 to reduce the QCD multijet background. The main backgrounds after the pre-selection are QCD multijet, $t\bar{t}+\mathrm{jets}$ and $W+\mathrm{jets}$ production.
The signal selection is defined by the following requirements:

- At least one b-tagged jet is found. To identify the b quark candidate jet the medium working point of the CSVv2 algorithm (see Section 5.5.5) is applied.
- The transverse mass m_T, defined in Section 8.3, exceeds $100\,\mathrm{GeV}$.

As discussed in Section 8.1, events passing these requirements are further categorized by studying the isolation of the selected muon candidate. Events are assigned to the *isolated lepton signal region* if they pass the criteria below:

- The muon candidate passes the tight isolation requirements (see Section 5.3).
- Each event contains at least one HTT-jet that is top tagged. For the top tag, the working point of the HTT V2 algorithm defined for a background efficiency of 3% without the subjet-b-tag requirement is utilized (see Table 7.2). The groomed n-subjettiness ratio ($z = 0.2$, $\beta = 1$) is used in the top tagging procedure.

To avoid double counting of objects, jets that lie within $\Delta R < 0.4$ to a muon candidate are discarded in the isolated lepton signal region.
The *non-isolated lepton signal region* comprises events that pass the criteria listed below:

- The muon candidate fails the tight isolation requirements.
- The event passes the 2D-cut as defined in Section 8.1.

8.11 Correction factors for simulation

The Standard Model background processes for this search are estimated based on simulated events only. The main background process for this search is $t\bar{t}+$jets production. Additionally, $W+$jets events and single top quark production constitute significant background processes in the signal regions, although their contribution is small compared to the $t\bar{t}+$jets processes. Since object properties as well as reconstruction and identification efficiencies of particle candidates differ between data and simulation, correction factors are applied on simulation to compensate for these differences. All applied corrections are introduced in this section. For the validation of the background estimation with simulated events multiple control regions are studied. Furthermore, the control regions are used to constrain the normalization of the $t\bar{t}+$jets and $W+$jets background processes. This is discussed in the next section (Section 8.12).

Pileup re-weighting

The simulated distribution of the number of pileup interactions does not agree with the observation in data. Hence, pileup re-weighting is performed as discussed in Section 6.6.1 using a minimum bias cross section of 69.2 mb [203].

Muon track reconstruction inefficiencies

In the 2016 data acquisition periods an inefficiency in the track reconstruction of charged particles has been observed [204, 205]. The inefficiencies are larger the higher the instantaneous luminosities are and are traced back to the so called *HIP (highly ionizing particles) effect*. The HIP effect describes nuclear interactions of highly ionizing particles with the detector material that lead to a saturation of the readout system of the tracking detector. As a result, the saturated readout channels are not able to process signals from the subsequent collision, and recover only slowly. Thus, the efficiency of the hit reconstruction and therefore also of the track reconstruction drops. This effect has been observed particularly in the muon track reconstruction. Since it is not simulated, correction factors of magnitudes up to 2% [204] as a function of η of the muon candidate are applied to model these inefficiencies.

Trigger efficiencies

Since the simulation of triggers is not included in the simulated events, the trigger selection can only be applied on data events. Thus, the data trigger efficiency is used to scale the simulation in order to match the trigger selection of data and simulation. The efficiency is determined from data events as a function of p_T and $|\eta|$ of the muon candidates using the Tag-and-Probe method [160] in a sample enriched in $Z \to \mu^+\mu^-$ events [191]. The trigger efficiencies range from approximately 64% to 93% for muon candidates with $p_T > 53\,\text{GeV}$ and $|\eta| < 2.4$ [191]. Since in the first $622\,\text{pb}^{-1}$ of data in Run B only the trigger HLT_Mu50 is considered, while for the remaining part of the data acquisition periods the combination of HLT_Mu50 and HLT_TkMu50 is applied, the trigger efficiencies are weighted according to the fraction of integrated luminosities for which they were available. As mentioned in Section 8.6, the available statistical precision of the efficiency measurements is not sufficient for the description of the drop in the trigger efficiency for muon candidates with $p_T > 500\,\text{GeV}$ and $|\eta| > 1.2$. Thus, events containing those muon candidates are discarded [191].

It has to be noted that the by the responsible group within the CMS experiment provided trigger efficiencies are determined by requiring the probe-muon to pass the high-p_T muon identification criteria [120], while in the presented analysis the medium identification working point is applied. Therefore, studies that validate the usage of the provided efficiencies in the presented search have been performed within the here presented work using data collected at $\sqrt{s} = 13\,\text{TeV}$ in the year 2015. The studies are presented in the appendix (see Appendix B). It is shown that the efficiencies determined using the medium identification criteria are consistent with the provided results. Furthermore, it is shown that the provided efficiencies, which are measured in a very different phase space than the one studied in this analysis, are also valid in a phase space close to the signal region of this search (see Appendix B).

Muon identification and isolation efficiencies

Correction factors as a function of p_T and $|\eta|$ of the muon candidate, which adjust the muon identification and isolation efficiencies in simulation to the ones in data, are considered. The factors have been estimated using the Tag-and-Probe method [160] applied in a sample dominated by $Z \rightarrow \mu^+\mu^-$ events [191]. For the identification efficiencies, correction factors between 0.92 and 1 are implemented. The correction factors for the isolation range between 0.99 and 1.

b tagging efficiencies and misidentification rates of c and light quark jets

In general, efficiencies of tagging a b jet correctly as well as a c jet or a jet originating from a lighter quark wrongly as a b jet can deviate between data and simulation. Therefore, re-weighting factors designed to diminish this effect, are taken into account [206, 207]. For this procedure, the efficiencies to identify a b jet, a c jet and a lighter quark jet as a b quark jet are determined for simulation in the signal regions of the analysis since the efficiencies depend on the characteristics of the studied events. These efficiencies are determined as a function of p_T and $|\eta|$. The flavor of the b-tagged jet on generator level is thereby defined by checking if a B or D hadron is found inside the examined jet. To receive the correct re-weighting factors, the resulting efficiencies are scaled with weights for each jet found in the event. Tagged jets are assigned the scaled efficiencies ϵ_i, while jets that are not tagged but should be tagged as a jet of the flavor i receive a scaled weight corresponding to $1 - \epsilon_i$.

Top tagging efficiencies

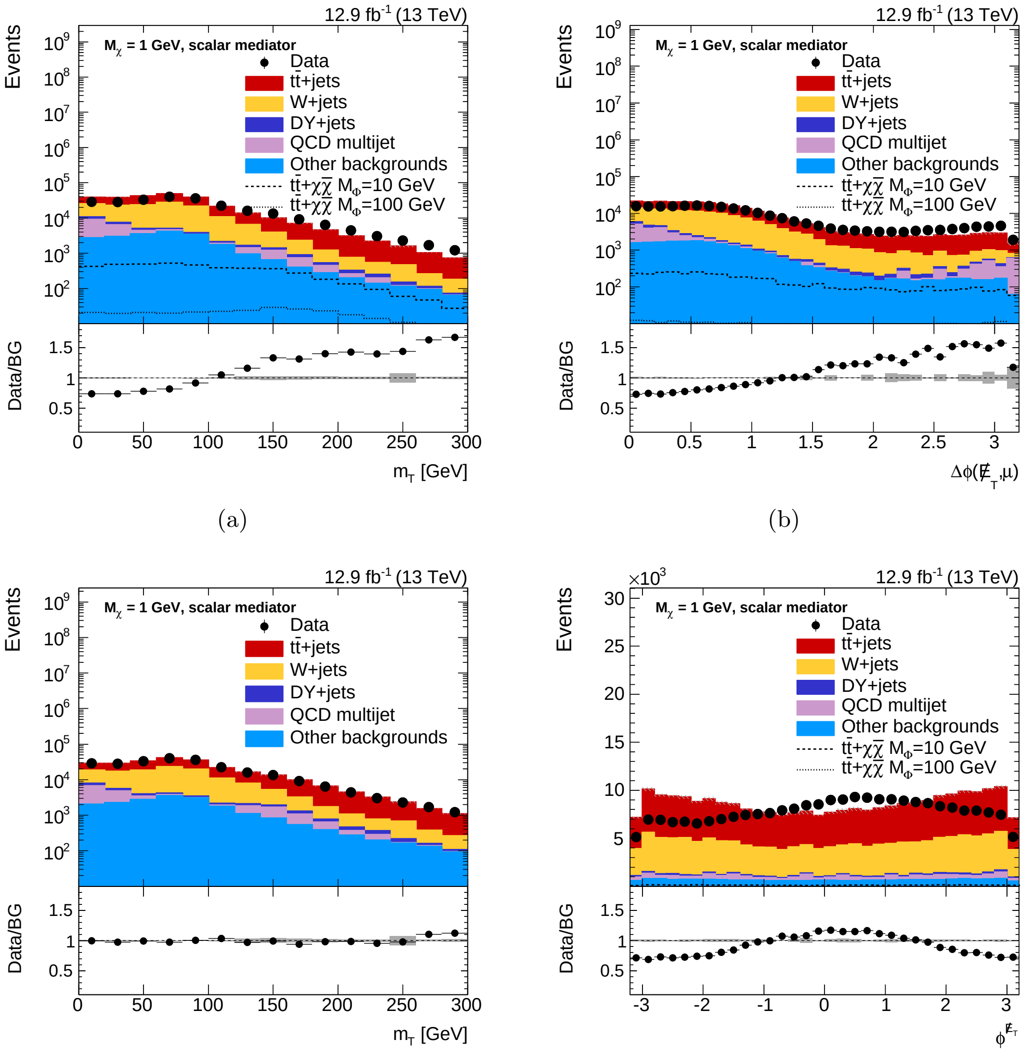
To compensate for differences between the top tagging efficiencies in data and simulation, a correction factor inclusive in p_T and η of the tagged HTT-jet is applied. The correction factor is determined in a data region enriched in lepton+jets $t\bar{t}$ decays and is found to be 0.98 for the applied working point [6]. This factor is applied on all tagged HTT-jets that can be matched within $\Delta R = 1.5$ to an hadronically decaying generator-level top quark whose decay products lie in a cone of $\Delta R = 1.5$ around it.

NLO k factors for simulated W + jets events

The LO W+jets samples are known to incorrectly model the jet-p_T distributions at high p_T. This effect has been observed already in analyses using 2015 data (see e.g. [208,209]). Its source is not yet fully understood, but can be most likely attributed to arise from NLO effects. The agreement between data and simulation is significantly improved by applying NLO/LO correction factors. These k factors have been derived by [3,4] using MadGraph [94] and are applied in this analysis. They comprise contributions arising from electroweak and QCD interactions to higher order processes and depend on the transverse momentum of the generated W boson. The factors are implemented such that only the shape of the LO W + jets simulation is changed, while its normalization is still derived from the LO cross section multiplied with the NNLO k factor (see Table 8.3).

Efficiency correction of the m_T selection requirement

While the measured shapes of other distributions are very well described by simulation (see Section 8.12), a large discrepancy between data and simulation is observed in the m_T distribution (see Fig. 8.18(a)). For values of m_T below 100 GeV, up to 25% more events are predicted than data events measured after the application of the pre-selection and the 2D-cut. Above values of 100 GeV, the observed data events exceed the prediction by up to 50%. The same trend is found in the spectrum of the difference in ϕ between the selected lepton and the measured $\not{E}_T$, referred to as $\Delta\phi(\not{E}_T, \mu)$ (see Fig. 8.18(b)). This is expected since m_T and $\Delta\phi(\not{E}_T, \mu)$ are strongly correlated (see Eq. (8.4)). To check if the discrepancy in $\Delta\phi(\not{E}_T, \mu)$ fully explains the discrepancy in the m_T distribution, a re-weighting of simulated events according to a fit of the ratio of data and simulation in the $\Delta\phi(\not{E}_T, \mu)$ distribution is performed. The fit function is of the form $f(x) = a + b\cos(cx + d)$ with parameters $a = 1.163$, $b = -0.426$, $c = 0.956$ and $d = -0.016$. Figure 8.18(c) shows the m_T distribution after the re-weighting of $\Delta\phi(\not{E}_T, \mu)$. Data and simulation agree very well. Hence, the disagreement found in the m_T distribution is fully attributed to the disagreement found in $\Delta\phi(\not{E}_T, \mu)$. The disagreement observed in $\Delta\phi(\not{E}_T, \mu)$ in turn, is traced back to a mismodeling of the ϕ distribution of the measured $\not{E}_T$, presented in Fig. 8.18(d). The latter is a known detector effect. It occurs most likely due to a mismodeling of the beam spot position, a misalignment of the tracker with respect to the calorimeter or due to both effects. A re-weighting of the ϕ distribution of $\not{E}_T$ in simulation has been tried, but resulted in little improvements in the description of the $\Delta\phi(\not{E}_T, \mu)$ and m_T distributions. This is explainable since the re-weighting procedure only introduces event weights, while correlations arising from the studied physics processes between the angle ϕ of the lepton and $\not{E}_T$ are not taken into account and are therefore not corrected by the re-weighting. The same effect has been observed by re-weighting $\Delta\phi(\not{E}_T, \mu)$, which on the one hand leads to a better agreement between data and simulation in m_T, but on the other hand deteriorates the agreement in other variables, like $\not{E}_T$. Hence, the source



(a) (b) (c) (d)

Figure 8.18: Comparison of data (*black dots*) and simulation (*colored areas*) after the application of the pre-selection and the 2D-cut. Presented are the distribution of (a) m_T, (b) $\Delta\phi(\not{E}_T, \mu)$, (c) m_T after re-weighting of $\Delta\phi(\not{E}_T, \mu)$ and (d) $\phi^{\not{E}_T}$.

of the discrepancy found in the m_T distribution is known and understood, but cannot be cured by re-weighting simulation. However, the analysis relies on the observable m_T only for the selection of signal and control regions. Thus, alone the overall normalization of simulation with respect to data is affected by the observed discrepancy. The shape of the m_T distribution does not influence the final result. Therefore, a correction is applied in the analysis that compensates for differences in the selection efficiency of the m_T requirement between data and simulation. The correction is determined after the application of the pre-selection and the 2D-cut. This is justified by the number of expected signal events being negligible compared to the number of selected background events at this stage of the selection. The 2D-cut is necessary to reduce the QCD multijet background to an insignificant amount. If a selection requirement of $m_T < 100\,\text{GeV}$ is imposed, the selection efficiency in data amounts to 67.44%, while the efficiency in simulation is 76.09%. The resulting correction factor $C_{m_T<100\,\mathrm{GeV}}$ to be applied on simulated events is 0.886. If a selection requirement of $m_T > 100\,\mathrm{GeV}$ is placed instead, the selection efficiency in data amounts to 32.56%, while it is 23.91% in simulation. This leads to a correction factor $C_{m_T>100\,\mathrm{GeV}}$ of 1.362. A systematic uncertainty due to the described efficiency correction is considered in the analysis (see Section 8.13).

8.12 Definition of background enriched control regions

The Standard Model background processes of this search are estimated from simulated events. Known discrepancies between simulation and data in reconstruction and identification efficiencies, selection requirements as well as in simulated event properties are accounted for by applying appropriate corrections (see Section 8.11). In order to validate the descriptions of the shape of the measured observables by simulation, two background enriched control regions are defined. Most importantly, these control regions are additionally used to constrain the normalization of the main Standard Model background processes of this search ($t\bar{t}$+jets and W+jets events) performing a background-only *maximum likelihood estimation (MLE)* fit in the control and signal regions simultaneously. In order to facilitate the determination of the normalization in this background-only fit, for each of the two processes one control region dominated by events of the respective background process is defined by inverting selection requirements of the signal selection before the event categorization. This way, orthogonality between the control regions and the signal regions is ensured. A region enhanced in $t\bar{t}$ + jets events is constructed by inverting the m_T selection requirement of the signal selection defined in Section 8.10. Thus, instead of selecting events with $m_T > 100\,\mathrm{GeV}$, events with $m_T < 100\,\mathrm{GeV}$ are studied. This control region is denoted as *$t\bar{t}$ control region* in the following. A region dominated by W + jets events, referred to as *W + jets control region*, is created by inverting the b tag requirement of the signal selection defined in Section 8.10. Hence, events that contain b-tagged jets are discarded. The choice of the W+jets control region might be confusing at first glance, since the transverse mass in W + jets events usually lies below 100 GeV, while the defined selection of the W + jets control region contains a requirement of $m_T > 100\,\mathrm{GeV}$. Thus, the control region comprises only misreconstructed W + jets events. However, these misreconstructed W + jets events represent the W + jets events that enter the signal regions. The proper modeling of these events has to be ensured, which justifies the definition of the control region adopted here.

In addition to these control regions, distributions after the pre-selection and the application of the 2D-cut are examined to check the agreement between data and simulation. The number of signal events expected at this stage of the selection is negligible compared to the large amount of expected background events such that the pre-selection applied together with the 2D-cut defines an additional validation region for the description of data by simulation. The application of the 2D-cut in addition is necessary due to the large contribution of the QCD multijet background still present after the pre-selection. The 2D-cut reduces this background to an amount that is comparable to the amount found after the application of the final signal selection and event categorization. Hence, the 2D-cut enables a validation of the background estimation by simulation after applying the pre-selection, which is then dominated by the same background processes as the signal region. Figure 8.19 presents the $\not\!\!E_T$ spectrum and the p_T distribution of the leading jet after the application of the pre-selection and the 2D-cut as well as in the $t\bar{t}$ and in the $W+$jets control regions. The main backgrounds after the pre-selection and the 2D-cut are $t\bar{t}+$jets and $W+$jets, with roughly equal portions. In the $t\bar{t}$ control region the main background is, with a fraction of approximately 72%, $t\bar{t}+$jets. Furthermore, roughly 13% $W+$jets are present. Assuming a signal model with $M_\chi = 1\,\text{GeV}$ and $M_\Phi = 10\,\text{GeV}$, which has the largest cross section of the studied signal models, the contamination with signal events is smaller than 2%. After the inversion of the b tag requirement, the control region contains approximately 68% $W+$jets and 21% $t\bar{t}+$jets events. The contribution from the signal model with $M_\chi = 1\,\text{GeV}$ and $M_\Phi = 10\,\text{GeV}$ amounts to roughly 1%. In all regions the shape of the $\not\!\!E_T$ spectra and the p_T distributions of the leading jet are well described by the simulation. No trends are observed. However, in all studied phase spaces discrepancies in the normalization between simulation and data are observed. These amount to roughly 10% after applying the pre-selection, while normalization differences of up to roughly 20% are observed in the $t\bar{t}$ control region. In the $W+$jets control region approximately 5% more events are observed than predicted. Thus, in the selected phase spaces, the normalization of the backgrounds is not well described by just scaling the background processes with the luminosity and the inclusive theory cross sections (see Section 8.7). Therefore, a background-only MLE fit is used to determine the best fit values of the cross sections of the main backgrounds using the defined control regions and the signal regions (see Section 8.14).

8.13 Systematic uncertainties

Before the background-only MLE fit can be performed, systematic uncertainties in the $\not\!\!E_T$ spectra in the isolated and non-isolated lepton signal regions as well as in the control regions are estimated. Uncertainties changing the normalization as well as the shape of the $\not\!\!E_T$ distributions are considered for all background and signal processes. If available, the recommendations of the responsible group within the CMS experiment are followed.

8.13.1 Normalization uncertainties

An uncertainty in the integrated luminosity of 2.5% [210] is applied. Furthermore, uncertainties in the cross sections of SM background processes are considered. The chosen prior uncertainties, listed below, are relatively large. However, they are significantly reduced during the statistical evaluation performed with the theta package [168]. This is especially true for the main backgrounds ($t\bar{t}+$jets and $W+$jets events), for which control regions are explicitly designed to constrain their normalization with high precision. The considered prior uncertainty in the cross section of SM $t\bar{t}+$jets production is 15% [166], while an uncertainty of 20% is assumed for $W+$jets production. For DY + jets as well as single top quark production prior uncertainties of 50% are considered. For all other SM background processes uncertainties of 100% are applied.

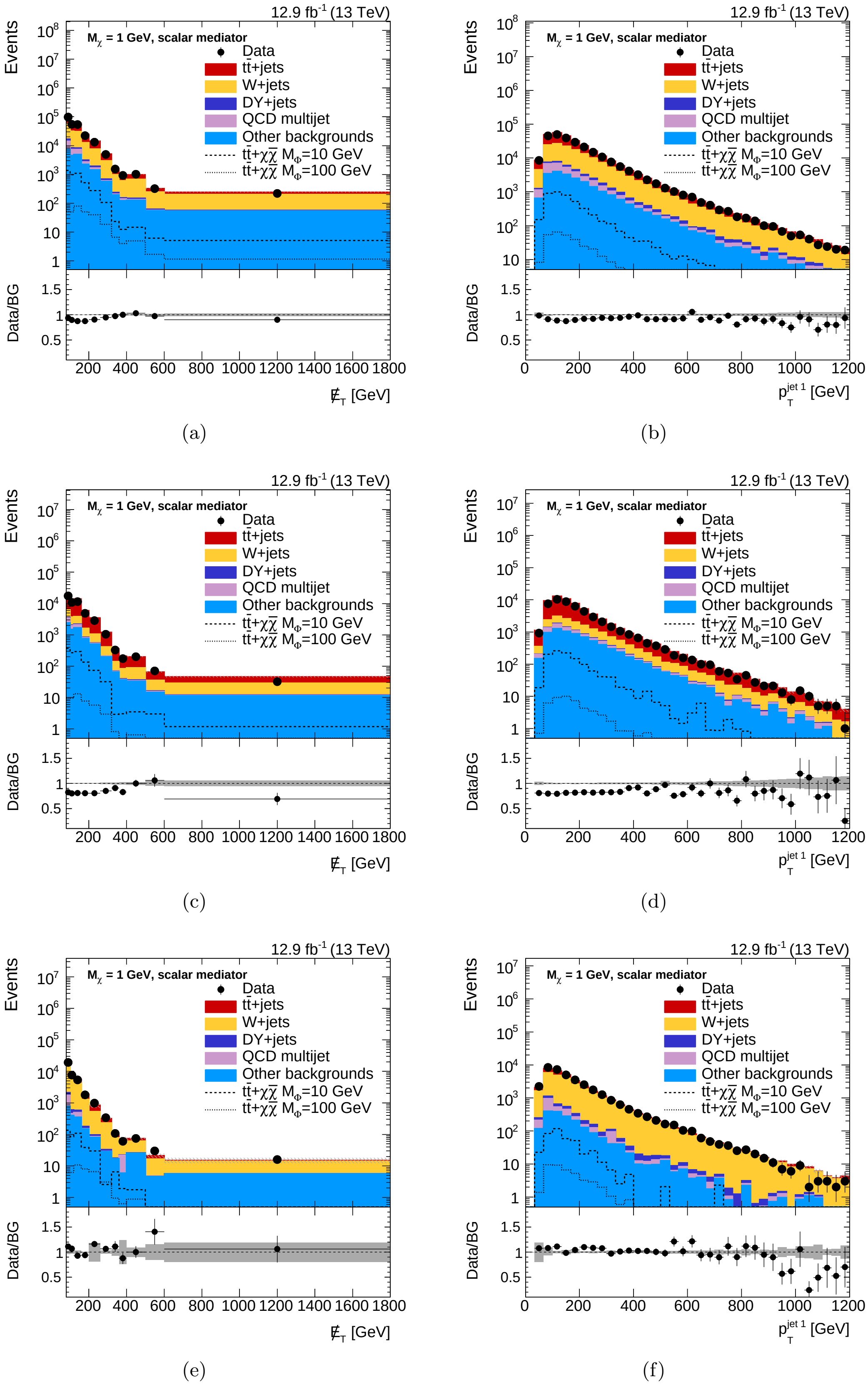


Figure 8.19: Comparison of data (*black dots*) and simulation (*colored areas*) in (a) the $\not{E}_T$ and (b) p_T spectra of the leading jet after the application of the preselection and the 2D-cut, (c) the $\not{E}_T$ and (d) p_T spectra of the leading jet in the $t\bar{t}$ control region and (e) the $\not{E}_T$ and (f) p_T spectra of the leading jet in the $W+$jets control region. The dark gray bands in the ratio plots and hatched areas in the figures illustrate statistical uncertainties of simulation.

8.13.2 Shape uncertainties

Uncertainties affecting not only the normalization of the $\not\!\!E_T$ spectra but also their shape are taken into account by varying the respective observable by one standard deviation up and down. The resulting changes in the $\not\!\!E_T$ distributions with respect to the nominal spectra yield the uncertainties.

Uncertainties in muon trigger, tracking, identification and isolation efficiencies

Uncertainties in the correction factors designed to diminish differences in the muon trigger, tracking, identification and isolation efficiencies between data and simulation are assigned [191, 205]. In order to take into account uncertainties in the trigger efficiencies, p_T and η dependent uncertainties ranging between below 1% and up to 13% in the high-p_T and high-$|\eta|$ regions are applied. These are the statistical uncertainties in the correction factors. Due to the observed drop in the trigger efficiencies for high-p_T muon candidates found in high-$|\eta|$ regions (see Section 8.6) an extra uncertainty has to be applied for muon candidates with $p_T > 100\,\mathrm{GeV}$ that lie within $|\eta| < 1.2$. This additional uncertainty amounts to 10% of the applied correction factor. The uncertainties in the correction factors for the track reconstruction inefficiencies are statistical uncertainties only, which depend on η of the muon candidate and lie below 0.5%. On the correction factors for the muon identification efficiencies a 1% uncertainty in the applied Tag-and-Probe method and a 0.5% uncertainty due to tracking inefficiencies caused by the HIP effect are assigned in addition to the η and p_T dependent statistical uncertainties. The statistical uncertainty amounts to up to 1% in great parts of the phase space and up to 8.5% in the high-p_T and high-$|\eta|$ region. p_T and η dependent statistical uncertainties at the per-mill level as well as an additional uncertainty of 1% are applied in the correction factors of the muon isolation efficiencies. The latter is added due to differences in the pileup conditions in the studied data set and the one used to determine the corrections, which is only a subset of the studied data set.

Uncertainties in the energy measurement of jets

The applied jet energy corrections are varied in dependence of p_T and η of the jets [211]. Uncertainties in the factors that adjust the jet energy resolution in simulation to the one measured in data of up to approximately 10% are considered as a function of η of the jet [212]. Both uncertainties are propagated to the calculation of $\not\!\!E_T$ and the resulting shifts yield the respective uncertainty in the $\not\!\!E_T$ spectra.

Generator and theory uncertainties

For signal processes and $t\bar{t}$ + jets, W + jets, DY + jets and $t\bar{t}$ + W as well as $t\bar{t}$ + Z production uncertainties in the renormalization and factorization scales used to model the respective processes are taken into account by determining the change in the $\not\!\!E_T$ spectra obtained after shifting both scales up or down by a factor of two at the same time. Additionally, the changes in the $\not\!\!E_T$ spectra resulting from shifting only one of

the scales up or down while the second scale is set to its nominal value are evaluated. This results in six combinations of variations of the two scales. The uncertainty due to the choice of the scales is then given by the biggest shift originating from the six variations with respect to the nominal spectra in each bin. For signal processes, only the uncertainty in the acceptance is determined. Uncertainties in the normalization due to scale variations are covered by uncertainties in the theory cross sections.
For other background processes than the ones mentioned above no scale uncertainties are considered, since their contribution to the total background is very small. Instead large normalization uncertainties in the cross sections of these processes are applied (see above), which cover uncertainties in the choice of the renormalization and factorization scales in the respective process.

Uncertainties in the pdfs are propagated to the $\not{E}_T$ spectra by evaluating the change in the $\not{E}_T$ distributions after applying the used pdfs shifted by their uncertainties. For the NNPDF3.0 pdf sets [90] adopted in this analysis 100 uncertainties are considered. From the resulting 100 variations of the $\not{E}_T$ spectra the root mean square (RMS) in each bin is calculated and taken as the uncertainty due to the choice of the pdf in the respective bin. Similar as for the scale uncertainties, also here only changes in acceptance are considered as systematic uncertainties for signal processes. Pdf uncertainties are taken into account when calculating the theory cross sections and result in a change of the normalization of the signal processes.

Since uncertainties in the factorization and renormalization scales used for the modeling of $W +$ jets events are applied, no additional uncertainty due to the application of the NLO/LO k factors is considered.

Uncertainties in pileup re-weighting

In order to take into account uncertainties in the integrated luminosity and the total inelastic pp cross section, the pileup re-weighting is repeated after varying the total inelastic pp cross section by $\pm 3.2\,\mathrm{mb}$ and the resulting change in the $\not{E}_T$ spectra is calculated [159].

Uncertainties in b tagging correction factors

To take into account uncertainties in the correction factors for different b tagging efficiencies and misidentification rates in data and simulation, the applied correction factors are varied by one standard deviation both up and down [206]. The uncertainties in the correction factors of the b tagging efficiency and of the misidentification rates of a c quark as b quark are assumed to be fully correlated, while the correction factors for the misidentification rates of a light quark as a b quark are assumed not to be correlated to the first two correction factors. For jets with transverse momenta beyond the phase space that is considered for the determination of the correction factors ($p_T > 670\,\mathrm{GeV}$ for the b quark identification efficiency as well as c quark misidentification rate and $p_T > 1000\,\mathrm{GeV}$ for the misidentification rate of light quarks as b quarks), the uncertainty in the highest considered transverse momenta is doubled and taken as the uncertainty in the corresponding correction factor.

Uncertainties in top tagging correction factors

An uncertainty of 1% in the correction factor that compensates for different top tagging efficiencies in data and simulation is applied [6]. Thus, the uncertainty of 1% is applied on all top tagged HTT-jets that are matched within $\Delta R < 1.5$ to an hadronically decaying generator-level top quark, whose decay products are found within $\Delta R < 1.5$ with respect to the initial top quark.

Uncertainties in the m_T efficiency correction

An uncertainty in the factors that correct the efficiency of the m_T selection requirement in simulation has to be taken into account. The statistical uncertainties in the efficiencies of the m_T selection are determined in both data and simulation and propagated to the correction factors. The resulting variations, which amount to 0.7% in the signal regions, are taken as the uncertainties on the applied correction.

All applied systematic uncertainties, their magnitude and the resulting effects on the total number of background events as well as on signal events assuming a scalar mediator with $M_\Phi = 10\,\mathrm{GeV}$ and $M_\chi = 1\,\mathrm{GeV}$ are summarized in Table 8.5. The numbers are presented averaged over the $\not\!\!E_T$ spectra in the isolated lepton and the non-isolated lepton signal. The uncertainties obtained for the background processes range from 0.2% (0.7%) to 30.5% (34.4%) in the isolated (non-isolated) lepton signal region, while for the signal process uncertainties between 0.2% (0.7%) and 4.5% (39.6%) are observed. For the background processes the biggest uncertainties arise from the propagation of uncertainties in the renormalization and factorization scales chosen in the modelling procedure of the processes. For the signal process the uncertainties in the jet energy corrections yield the highest effect. The impact of each systematic uncertainty on the $\not\!\!E_T$ spectra in both signal regions on the $t\bar{t}+$ jets background and the signal model mentioned above is presented in Appendix C.

8.14 Estimation of Standard Model background processes

After evaluating the systematic uncertainties, the final estimation of the SM background processes is performed by running an MLE fit in the control and signal regions using the theta package [168]. In the MLE fit, the background-only hypothesis is fitted simultaneously in the two control regions and the two signal regions to the measured $\not\!\!E_T$ spectra in data. The aim of this fit is to constrain the normalization of the $t\bar{t}+$jets and $W+$jets processes to their best-fit value, and decrease the assumed rate uncertainties in these processes. In the MLE fit, log-normal distributions are studied as prior constraints for all uncertainties listed in Section 8.13. In order not to introduce too many parameters in the fit and thus to prevent instabilities, the rates of the smallest background processes (QCD multijet, $t\bar{t}+$W/$t\bar{t}+$Z and diboson events) are fixed in the fit. Furthermore, these

		Isolated lepton signal region		Non-isolated lepton signal region	
Systematic uncertainty	Magnitude (%)	B (%)	S (%)	B (%)	S (%)
Integrated luminosity	2.5	2.5	2.5	2.5	2.5
Muon ID/iso/tracking & trigger	p_T, η dependent	2.3	2.1	2.1	1.8
b tagging (bc)	p_T, $\lvert\eta\rvert$, flavor dependent	2.9	2.9	3.6	2.8
b tagging (udsg)	p_T, $\lvert\eta\rvert$, flavor dependent	0.8	0.7	0.8	0.7
Jet energy scale	p_T, η dependent	+8.5	$^{+4.5}_{-2.3}$	+12.8	+39.6
Jet energy resolution	η dependent	+7.1	+2.3	+12.4	+14.4
m_T efficiency correction	0.7	0.7	0.7	0.7	0.7
Pileup	4.6	+2.5	+1.3	2.0	$^{+2.5}_{-2.0}$
Pdf (on acceptance)	—	—	1.6	—	3.3
Pdf (on background)	—	1.8	—	2.2	—
Scale (on acceptance)	$^{+100}_{-50}$	—	+2.1	—	+3.4
Scale (on background)	$^{+100}_{-50}$	$^{+30.5}_{-26.3}$	—	$^{+36.4}_{-30.7}$	—
Top tag correction	1	0.2	0.2	—	—

Table 8.5: Summary of the magnitudes of the considered uncertainties (except for the rate uncertainties in SM background processes) and their effect on the total number of background events B and signal events S (assuming a scalar mediator with $M_\Phi = 10$ GeV and $M_\chi = 1$ GeV) averaged over the $\not{E}_T$ spectra in the isolated lepton signal region and the non-isolated lepton signal region. Numbers without any sign indicate that the respective uncertainty affects the $\not{E}_T$ spectra symmetrically in the up and down directions. If only a positive value is given, both the up and the down variation increase the yields of the distributions. In these cases, the larger of the two shifts is given here.

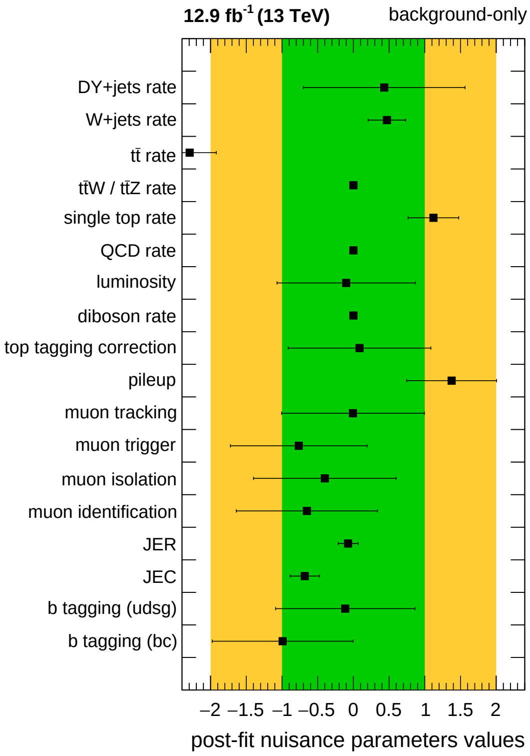


Figure 8.20: Differences between the post-fit uncertainties and the assumed prior uncertainties in the different nuisance parameter considered in the MLE fit in units of the respective prior uncertainties.

processes are studied without consideration of their uncertainties, since they can become quite large due to limited statistical precision of the simulated events. In Fig. 8.20 the deviations of the post-fit uncertainties from the assumed prior uncertainties are presented in units of the respective prior uncertainty. Except for the $t\bar{t}$ rate uncertainty, all post-fit uncertainties lie well within 2σ of the assumed prior uncertainties. The fit yields the following corrections for the normalization of the $t\bar{t}+$jets and $W+$jets background processes:

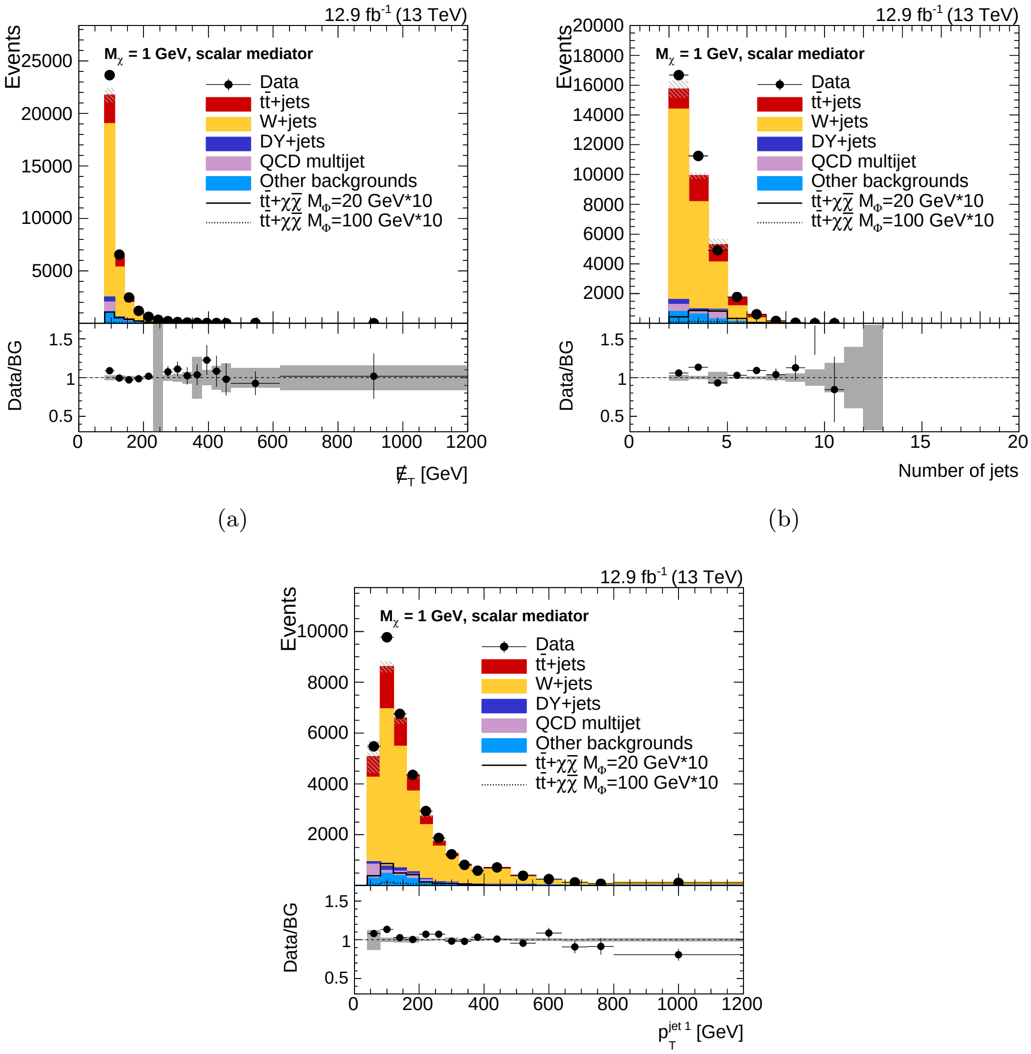
$t\bar{t}+$ jets: 0.724 ± 0.079,

$W+$ jets: 1.087 ± 0.074.

These correction factors are applied to the two processes and the scaled distributions are studied in the following. Control distributions of observables used in the final event selection are presented in Figs. 8.21 to 8.24. For the $W+$jets control region, the distributions of $\not{E}_T$, the number of jets and the p_T of the leading jet are presented in Fig. 8.21. The distribution of $\not{E}_T$, the number of jets, the p_T of the leading jet, and the the number of b-tagged jets applying the medium working point of the CSV algorithm in the $t\bar{t}$ control region are presented in Fig. 8.22. Figure 8.22 present the p_T of the HTT-jets, their mass m_{HTT}, f_{REC} and τ_{32}, which are the observables used in the top tagging procedure adopted for the isolated lepton signal region, in the $t\bar{t}$ control region. After the application of the corrections, not only the shape, but also the normalization of simulated events is well in agreement with data. No significant trends or large deviations between data and simulation are observed.

Since the region obtained after applying the pre-selection and the 2D-cut is not used in the background-only fit and since the number of signal events, at this stage of the selection, is small compared to the large amount of background events, it represents a validation region of the derived corrections. Figure 8.24 shows the distributions of the p_T of the leading muon candidate, $\not{E}_T$, the number of jets, the p_T of the leading jet, and the number of b-tagged jets applying the medium working point of the CSV algorithm after the application of the pre-selection and the 2D-cut. In this region, data and simulation are in good agreement as well after the application of the normalization corrections. No trends are observed in the ratio of data and simulated events.

In conclusion, the measured distributions in data are well described by simulation after the application of the corrections described in Section 8.11 and after deriving correction factors for the normalization of the $t\bar{t} +$ jets and $W +$ jets background processes. After these corrections, simulation gives a reliable estimate of the expected background events. Thus, the usage of simulated events to describe the background processes in the presented analysis is justified.

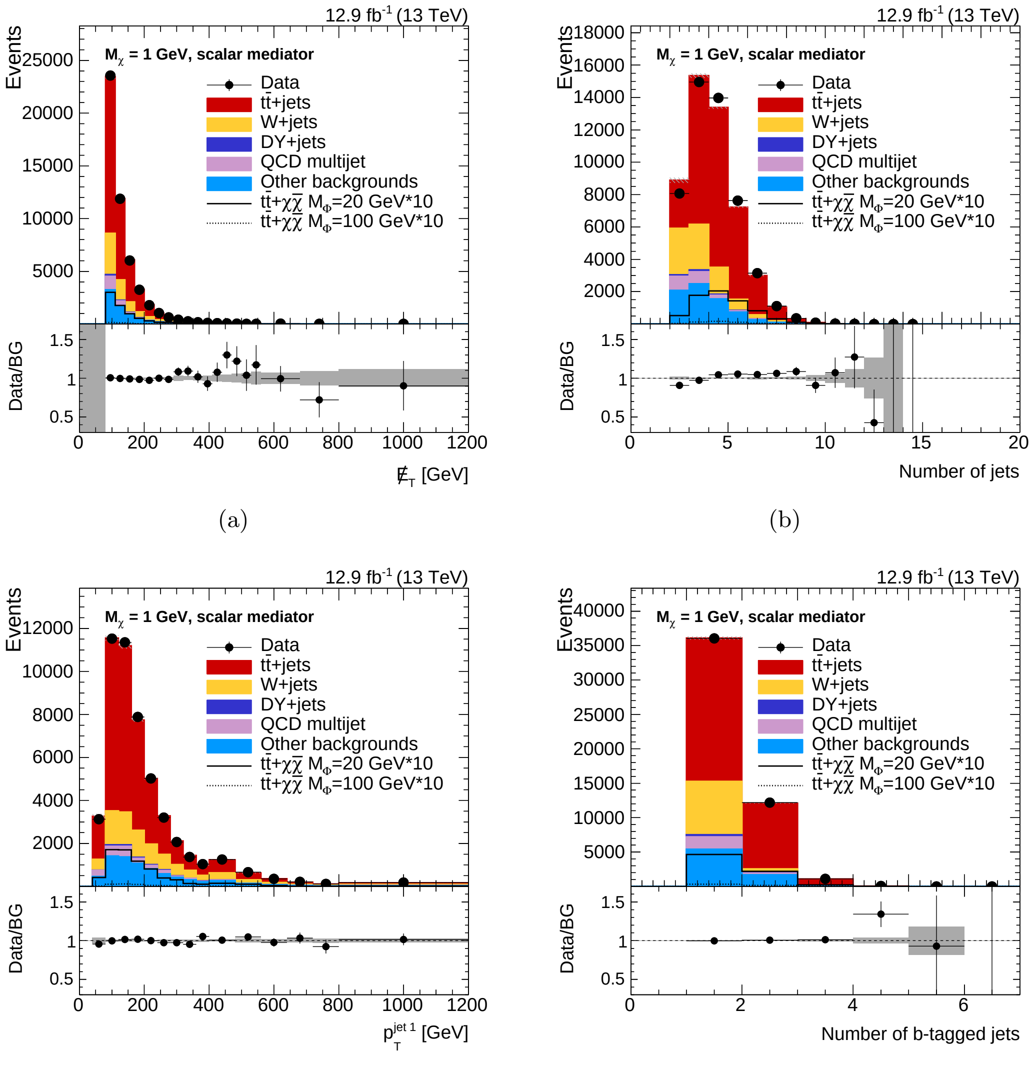


(a)

(b)

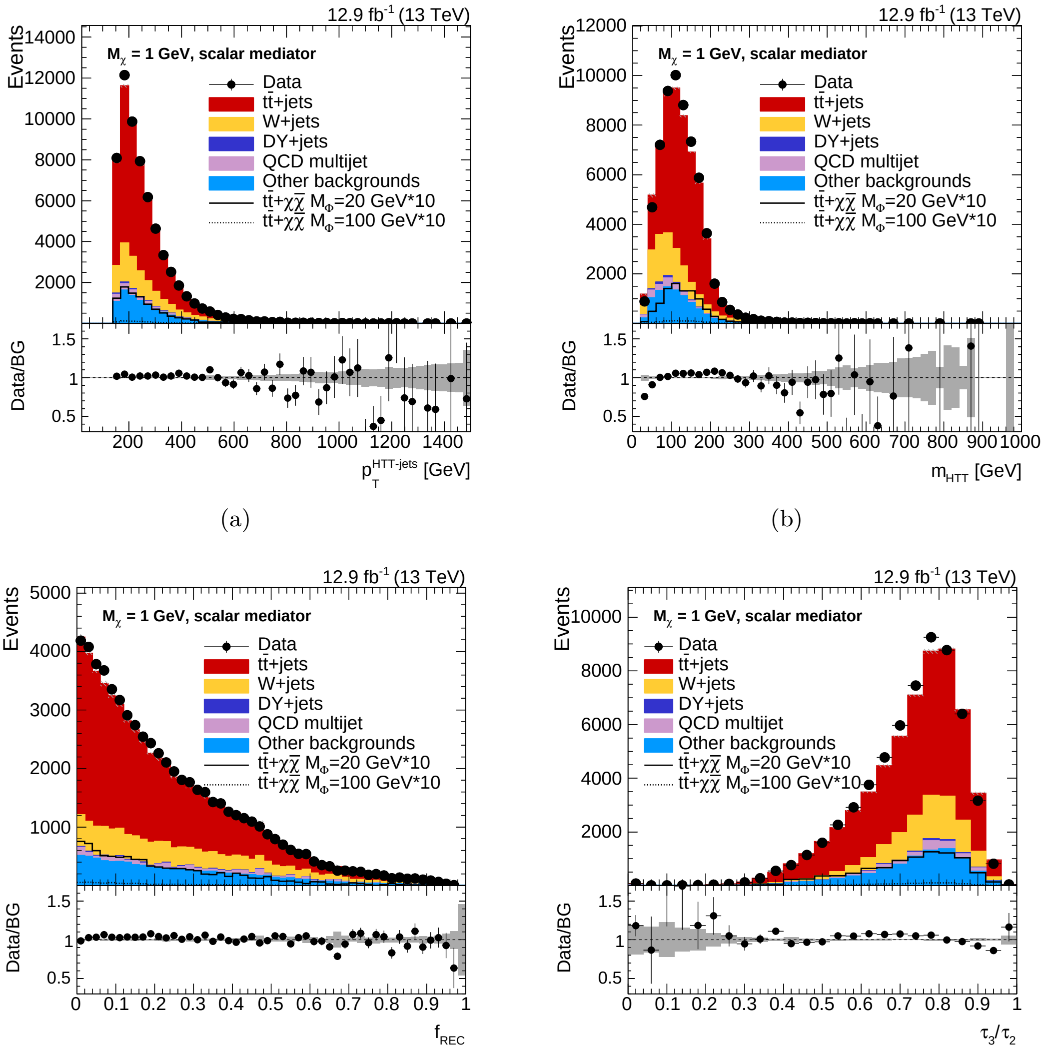
(c)

Figure 8.21: Comparison of data (*black dots*) and simulation (*colored areas*) in the $W+$ jets control region. Shown are the distribution of (a) $\not{E}_T$, (b) the number of jets, (c) the p_T of the leading.



(a) (b) (c) (d)

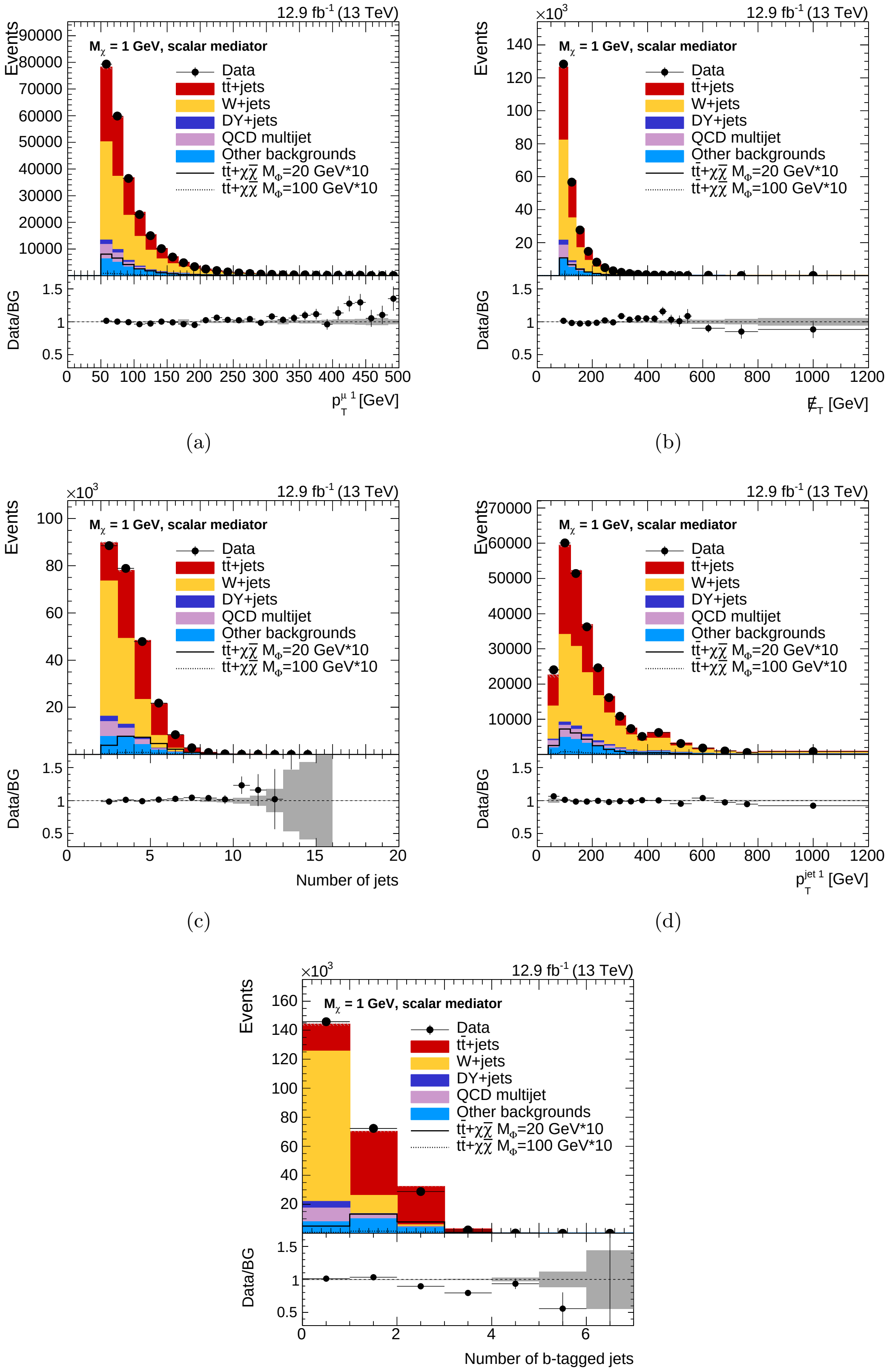
Figure 8.22: Comparison of data (*black dots*) and simulation (*colored areas*) in the $t\bar{t}$ control region. Shown are the distribution of (a) $\not{E}_T$, (b) the number of jets, (c) the p_T of the leading jet and (d) the number of b-tagged jet applying the medium working point of the CSV algorithm.



(a) (b)

(c) (d)

Figure 8.23: Comparison of data (*black dots*) and simulation (*colored areas*) in the $t\bar{t}$ control region. Shown are the distribution of (a) the p_T of the HTT-jets, (b) m_{HTT}, (c) f_{REC} and (d) τ_{32} for all HTT-jets.



(a)

(b)

(c)

(d)

(e)

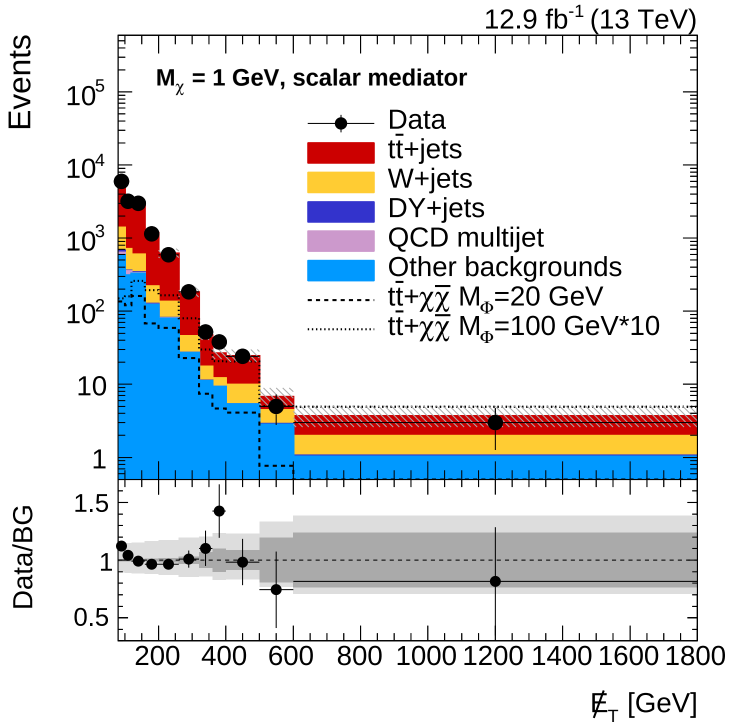
Figure 8.24: Comparison of data (*black dots*) and simulation (*colored areas*) after the application of the pre-selection and the 2D-cut in the distribution of (a) p_T of the leading muon candidate, (b) $\not\!\!E_T$, (c) the number of jets, (d) the p_T of the leading and (e) the number of b-tagged jet applying the medium working point of the CSV algorithm.

8.15 Results

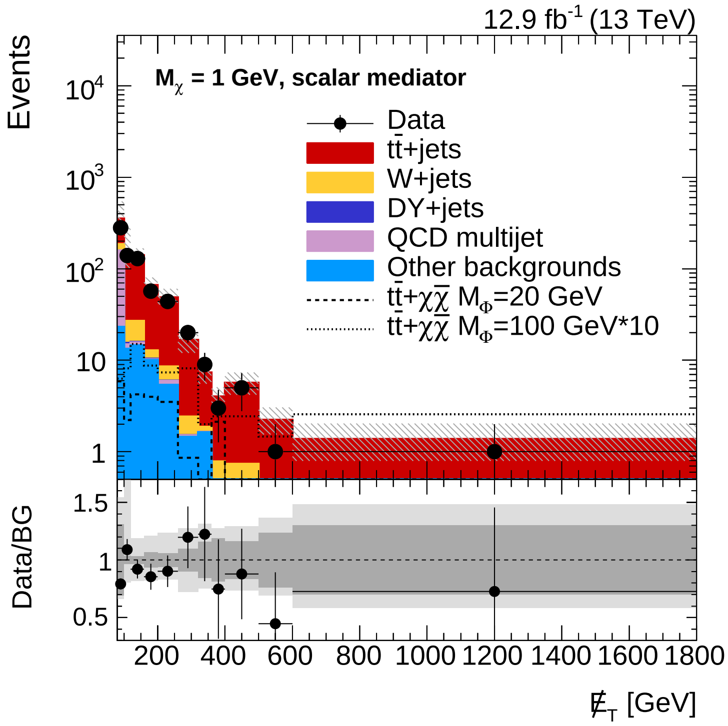
After the estimation of the SM background processes and the evaluation of the systematic uncertainties, the final results of the search can be assessed. The $\not\!\!E_T$ spectra obtained in the isolated lepton signal region and in the non-isolated lepton signal region are presented in Fig. 8.25. The main background process after the application of the full event selection and event categorization is $t\bar{t}$ + jets production. It makes up roughly 83% (73%) of the total background in the isolated (non-isolated) lepton signal region. The second largest background contribution arises from W + jets events. Approximately 8% (5%) of the total background is attributed to this process. Other background processes contribute only very little to the final $\not\!\!E_T$ spectra. In general, the total number of events is much higher in the isolated than in the non-isolated lepton signal region. Still, the non-isolated lepton signal region adds sensitivity to the search. While the background processes tend to have smaller values of $\not\!\!E_T$, the signal processes are expected to populate the high $\not\!\!E_T$ region. However, in both event categories data and simulation agree within the uncertainties over the whole $\not\!\!E_T$ range. No excess is seen in data. In total, 14161 data events are observed in the isolated lepton signal region, while 13500^{+4400}_{-3600} SM background events are expected. In the non-isolated lepton signal region, $770\pm^{+120}_{-110}$ SM background events are predicted, while 689 data events are observed. In the isolated lepton signal region the total uncertainty is dominated by the systematic uncertainties. Contrary, the statistical uncertainties dominate the total uncertainty in the non-isolated lepton signal region. It has to be noted that for the numbers given here the rate uncertainties in the cross sections of SM background events are not considered, since the assumed prior uncertainties are tightly constrained by the statistical interpretation of the result (Section 8.16). The corresponding post-fit uncertainties are much smaller.

The signal selection efficiencies in the two event categories are presented in Fig. 8.26. The efficiencies are calculated with respect to the total number of simulated events. Thus, events in all possible decay modes of the $t\bar{t}$ system are taken into account. No scaling with the branching ratio of the $t\bar{t}$ system into the μ+jets channel is performed. Assuming signal models with a scalar (pseudoscalar) mediator, efficiencies between 0.3% (1.0%) and 2.0% (1.8%) are achieved in the isolated lepton signal region, while the efficiencies range from 0.01% (0.06%) to 0.2% (0.15%) in the non-isolated signal region. In general, the signal selection efficiencies increase with the masses of the DM particles and the mediator masses. This behaviour is mainly connected to the $\not\!\!E_T$ spectra becoming harder with higher DM and mediator masses (see Fig. 8.2). Since the $\not\!\!E_T$ spectra in pseudoscalar models are on average higher than in scalar mediator models, especially for low DM and mediator masses, higher signal selection efficiencies are observed in pseudoscalar signal models than in scalar signal models. For signal models with large mediator and DM masses, where the differences between the $\not\!\!E_T$ spectra are less prominent, the efficiencies assuming a scalar mediator and a pseudoscalar mediator are comparable.

To conclude, the search revealed no deviations from the SM prediction. No hint for physics beyond the SM has been found. In the following section, the results of the search are used to set exclusion limits on physics beyond the SM.

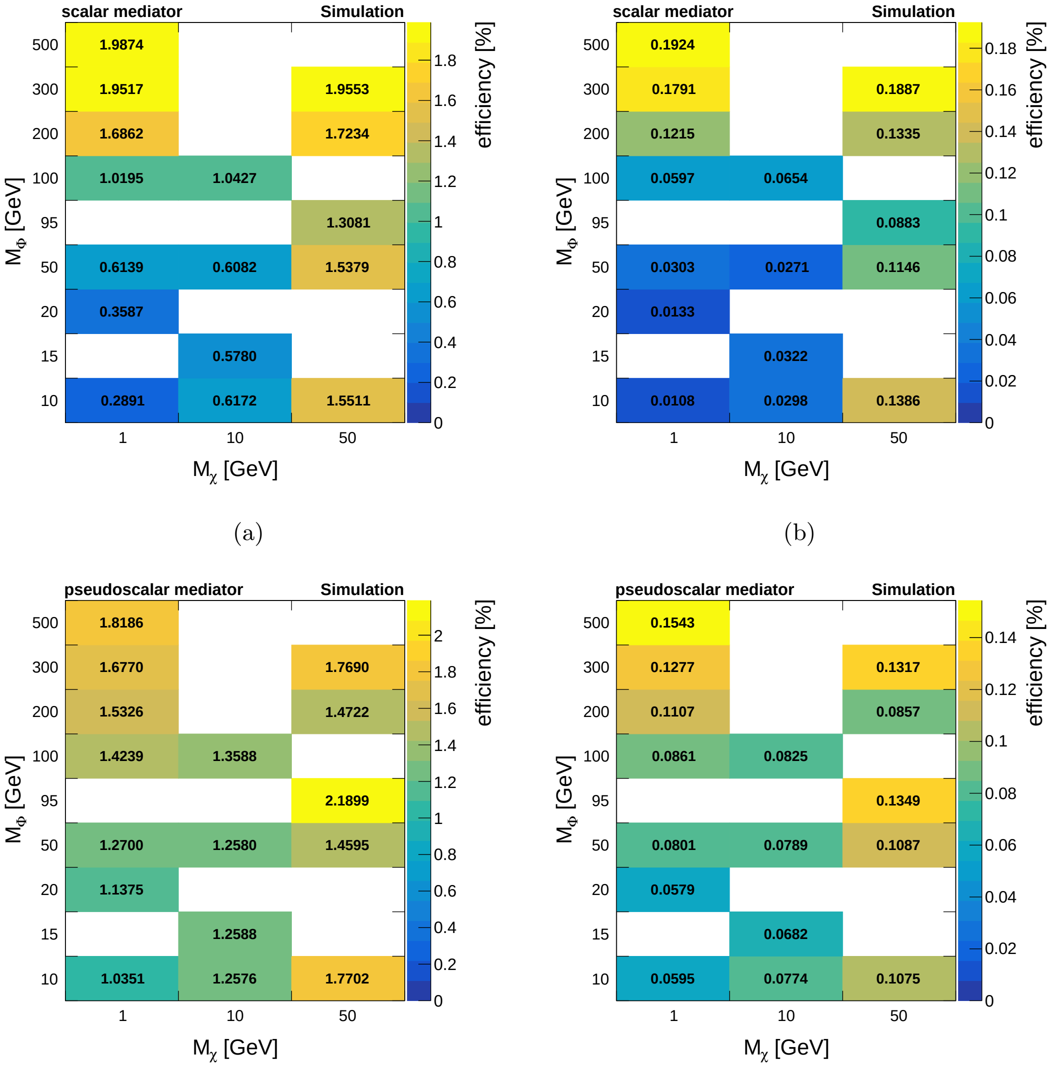


(a)



(b)

Figure 8.25: Comparison of data (*black dots*) and simulation (*colored areas*) in the $\not{E}_T$ spectra obtained in (a) the isolated lepton signal region and (b) the non-isolated lepton signal region. The two distributions are used for the final statistical interpretation of the search. Statistical and systematic uncertainties are represented by hatched areas in the figures. The dark gray bands in the ratio plots illustrate the statistical uncertainties only. The combined statistical and systematic uncertainties, except for the rate uncertainties on the cross sections of SM background processes, are represented by the light gray bands.



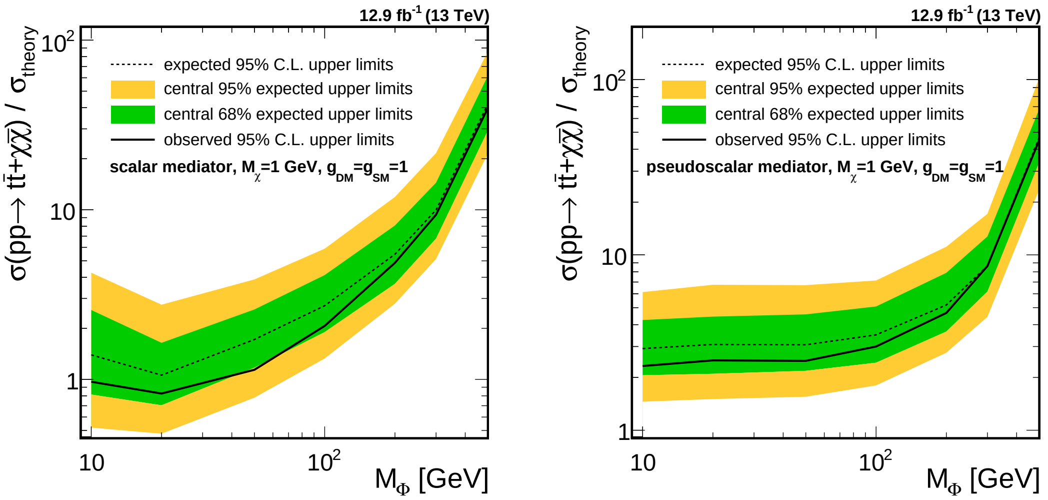
(a)

(b)

(c)

(d)

Figure 8.26: Signal selection efficiencies (*z-axis*) as a function of the mass of the DM particles (*x-axis*) and the mediator mass (*y-axis*) calculated with respect to all simulated events assuming signal models with scalar mediators in (a) the isolated lepton signal region and (b) the non-isolated lepton signal region, and assuming a pseudoscalar mediator in (c) the isolated-lepton signal region and (d) the non-isolated lepton signal region.



(a) (b)

Figure 8.27: Expected (*dashed line*) and observed (*solid line*) upper limits on the production cross section of DM produced in association with a top quark pair at 95% C.L. as a function of the mediator mass assuming $g_{\mathrm{DM}} = g_{\mathrm{SM}} = 1$, DM particles with masses of 1 GeV as well as (a) scalar and (b) pseudoscalar mediators. The upper limits are normalized to the simplified model theory cross sections (see Table 8.2). For the expected limit also the one (*green*) and two sigma bands (*yellow*) are shown.

8.16 Interpretation of the results

The results described in Section 8.15 are used to set 95% C.L. exclusion limits on the studied signal models. The limits are derived by performing a shape analysis based on the $\not\!\!E_T$ spectra in the two signal regions presented in Fig. 8.25. Furthermore, the $\not\!\!E_T$ spectra obtained in the $t\bar{t}$ and W+jets control regions (defined in Section 8.12) are added to the statistical interpretation of the result. This enables a precise determination of the normalization of the $t\bar{t}$+jets and W+jets processes in the studied phase space during the limit setting procedure (see Section 8.14). The $\not\!\!E_T$ spectra in the two control regions are depicted in Figs. 8.22(a) and 8.21(a). The exclusion limits are determined utilizing the Bayesian method implemented within the theta package [168]. A detailed description of the method and its implementation can be found in [169]. Nuisance parameters are used to handle the statistical as well as all systematic uncertainties listed in Section 8.13. All uncertainties, with the exception of the uncertainty on the top tagging correction factors, are treated as fully correlated among the two signal regions. The top tagging uncertainty has to be considered only in the isolated lepton signal region since no top tagging is applied in the non-isolated lepton signal region. The derived upper exclusion limits on the production cross section of DM in association with a top quark pair are presented in Fig. 8.27 as a function of the mediator mass for signal models assuming DM particles with masses of 1 GeV as well as scalar (Fig. 8.27(a)) and pseudoscalar mediators (Fig. 8.27(b)). Since theoretical models provide strong motivation especially for low-mass DM ($M_\chi < 10\,\mathrm{GeV}$) [213], this region is particularly interesting to study. Additionally, collider experiments feature an enhanced sensitivity in this parameter space compared to indirect or direct DM experiments [214]. The cross section limits in Fig. 8.27 are normalized to the simplified model theory cross sections (see Table 8.2) such that for values below unity the corresponding signal model is excluded. Thus, scalar mediators with masses below 36 GeV are excluded at 95% C.L. if $g_{\mathrm{DM}} = g_{\mathrm{SM}} = 1$ and DM particles with masses of 1 GeV are assumed. For models assuming a pseudoscalar mediator, upper limits on the production cross section between 1.05 pb and 0.233 pb are set. In general, the sensitivity of the search decreases with increasing mediator masses despite increasing signal selection efficiencies due to the low production cross section of signal models with high mediator masses. For models assuming DM particles with masses of 1 GeV, the observed limits at low mediator masses are significantly better for scalar than for pseudoscalar mediator models despite the harder $\not{E}_T$ spectra in the latter case. For mediators with masses of 200 GeV and above, where the production cross sections of the two models become comparable, the sensitivities are in a similar range. The expected and observed 95%C.L. upper cross section limits as well as the 1σ- and 2σ-uncertainty bands in the expected exclusion limits are presented for all examined signal points assuming scalar mediators in Table 8.6 and assuming pseudoscalar mediators in Table 8.7. For signal models with $M_\chi = 1\,\mathrm{GeV}$ or $M_\chi = 10\,\mathrm{GeV}$ and scalar mediators with $M_\Phi = 50\,\mathrm{GeV}$ or $M_\Phi = 100\,\mathrm{GeV}$ sensitivities close to the cross sections predicted by the simplified model framework are achieved. However, more data are needed to reach the sensitivity necessary to arrive at final conclusions about these signal models.

8.17 Comparison to other results

In Section 8.5 a comparison of the sensitivity of this search to the sensitivity of the resolved lepton+jets analysis has already been presented. That comparison has been performed without consideration of systematic uncertainties and control regions by rebuilding the main selection steps of EXO-16-005. Here, the final result of this thesis using $12.9\,\mathrm{fb}^{-1}$ of data is compared to the result of EXO-16-005 presented in [3,4]. The latter is a combination of searches in the full hadronic, electron+jets, muon+jets and di-leptonic decay modes of the $t\bar{t} + \chi\bar{\chi}$ system as well as a search for $b\bar{b} + \chi\bar{\chi}$ production using $2.2\,\mathrm{fb}^{-1}$ of data collected with the CMS experiment. The comparison of the expected upper cross section limit of the two results is presented in Fig. 8.28. For signal models assuming a scalar mediator, the two searches yield comparable sensitivity for low mediator masses. For high mediator masses the resolved analysis outperforms the boosted search, which is also true for models assuming a pseudoscalar mediator. However, it has to be noted that while a larger amount of data is analyzed in the boosted search, it is restricted to the muon channel, whereas in the resolved analysis all possible $t\bar{t}$ decay modes as well as $b\bar{b} + \chi\bar{\chi}$ production are considered. Furthermore, in EXO-16-005 only signal models assuming $M_\chi = 1\,\mathrm{GeV}$ are studied, while in the presented analysis further DM masses are considered.

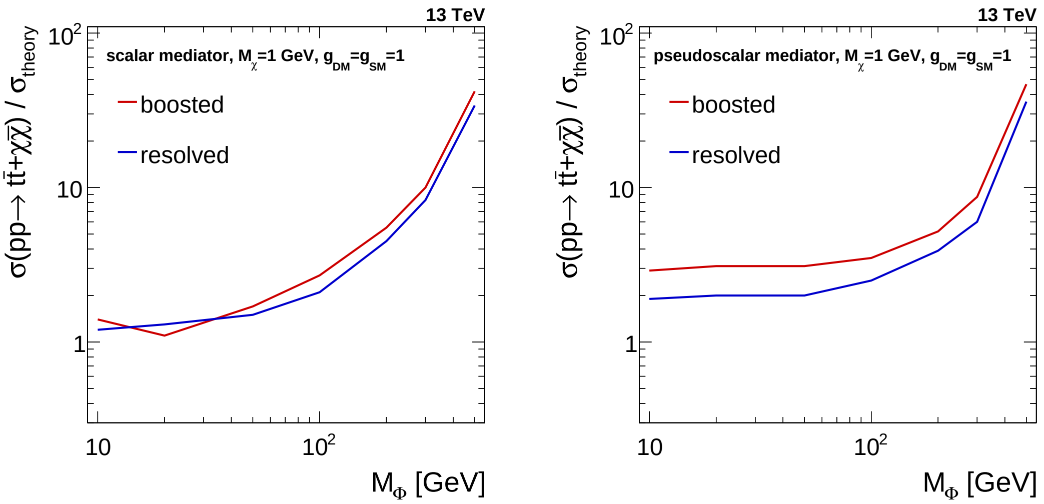
The most sensitive ATLAS result is the analysis in the lepton+jets channel. At 95% C.L., exclusion limits are set on scalar (pseudoscalar) mediator masses below 320 GeV (350 GeV). However, in contrast to the CMS results, DM particles with masses of 1 GeV

M_χ[GeV]	M_Φ[GeV]	exp. [pb]	obs. [pb]	$[-\sigma,+\sigma]$ [pb]	$[-2\sigma,+2\sigma]$ [pb]	$\sigma_{\mathrm{obs}}/\sigma_{\mathrm{theory}}$
1	10	29.8	20.7	[17.4, 54.8]	[11.1, 90.9]	0.97
1	20	11.6	9.04	[7.71, 18]	[5.24, 30.2]	0.83
1	50	5.31	3.52	[3.62, 7.98]	[2.41, 12]	1.14
1	100	1.96	1.49	[1.37, 2.97]	[0.957, 4.25]	2.07
1	200	0.556	0.495	[0.372, 0.821]	[0.284, 1.21]	4.87
1	300	0.304	0.285	[0.206, 0.44]	[0.156, 0.658]	9.36
1	500	0.208	0.197	[0.145, 0.305]	[0.106, 0.424]	39.82
10	10	4.83	3.09	[3.34, 7.32]	[2.46, 10.4]	30.56
10	15	4.86	3.17	[3.17, 7.56]	[2.26, 11.7]	24.78
10	50	5.51	4.29	[3.85, 8.16]	[2.77, 11.7]	1.39
10	100	1.85	1.37	[1.28, 2.77]	[0.976, 4.12]	1.85
50	10	0.546	0.472	[0.376, 0.804]	[0.271, 1.12]	227.14
50	50	0.605	0.489	[0.415, 0.9]	[0.308, 1.28]	190.49
50	95	0.926	0.771	[0.652, 1.37]	[0.472, 1.98]	107.05
50	200	0.538	0.469	[0.375, 0.795]	[0.265, 1.16]	4.68
50	300	0.308	0.272	[0.215,0.454]	[0.161,0.644]	8.93

Table 8.6: Summary of the 95% C.L. expected and observed upper exclusion limits on the cross of $t\bar{t} + \chi\overline{\chi}$ production assuming a scalar mediator together with the 1σ- and 2σ- uncertainties in the expected exclusion limits. Additionally, the ratio $\sigma_{\mathrm{obs}}/\sigma_{\mathrm{theory}}$ of the observed upper cross section limit and the theory cross sections (see Table 8.2) is given.

M_χ[GeV]	M_Φ[GeV]	exp. [pb]	obs. [pb]	$[-\sigma,+\sigma]$ [pb]	$[-2\sigma,+2\sigma]$ [pb]	$\sigma_{\mathrm{obs}}/\sigma_{\mathrm{theory}}$
1	10	1.32	1.05	[0.93,1.92]	[0.657,2.77]	2.32
1	20	1.27	1.03	[0.864,1.83]	[0.62,2.78]	2.50
1	50	0.946	0.766	[0.672,1.41]	[0.478,2.07]	2.49
1	100	0.676	0.58	[0.469,0.981]	[0.348,1.38]	3.00
1	200	0.455	0.41	[0.321,0.694]	[0.243,0.978]	4.67
1	300	0.342	0.341	[0.244,0.503]	[0.175,0.677]	8.63
1	500	0.241	0.233	[0.174,0.35]	[0.122,0.526]	45.13
10	10	0.862	0.737	[0.597,1.26]	[0.432,1.72]	48.42
10	15	0.859	0.738	[0.609,1.24]	[0.443,1.73]	37.85
10	50	0.925	0.769	[0.626,1.31]	[0.472,1.8]	2.49
10	100	0.735	0.608	[0.502,1.07]	[0.37,1.55]	3.08
50	10	0.447	0.421	[0.311,0.649]	[0.223,1.01]	143.78
50	50	0.471	0.423	[0.329,0.7]	[0.24,1.03]	144.47
50	95	0.593	0.51	[0.404,0.84]	[0.3,1.22]	47.57
50	200	0.487	0.436	[0.33,0.704]	[0.238,0.974]	5.14
50	300	0.382	0.463	[0.268,0.572]	[0.196,0.813]	12.04

Table 8.7: Summary of the 95% C.L. expected and observed upper exclusion limits on the cross of $t\bar{t}+\chi\overline{\chi}$ production assuming a pseudoscalar mediator together with the 1σ- and 2σ- uncertainties in the expected exclusion limits. Additionally, the ratio $\sigma_{\mathrm{obs}}/\sigma_{\mathrm{theory}}$ of the observed upper cross section limit and the theory cross sections (see Table 8.2) is given.



(a) (b)

Figure 8.28: Comparison of the sensitivity of the boosted search (*red*) and the resolved search (*blue*). Shown are the expected upper limits on the production cross section of DM produced in association with a top quark pair at 95% C.L. as a function of the mediator mass assuming $g_{\mathrm{DM}}=g_{\mathrm{SM}}=1$, DM particles with masses of 1 GeV as well as (a) scalar and (b) pseudoscalar mediators. The upper limits are normalized to the simplified model theory cross sections (see Table 8.2).

and couplings $g_{\mathrm{DM}}=g_{\mathrm{SM}}=3.5$ are assumed [5].

8.18 Outlook

The high instantaneous luminosities and the excellent overall performance of the LHC as well as well as the high data acquisition efficiency of the CMS experiment achieved during 2016 promise the recording of a data set as least as large as the 2016 data set in the year 2017. In combination with the remaining two thirds of the data collected in 2016, which roughly correspond to an integrated luminosity of 24 fb^{-1} and have not been studied in the presented analysis, the larger amount of available data will significantly enhance the sensitivity of the presented search for Dark Matter. This also holds true for models with pseudoscalar mediators, which are interesting since the sensitivity of collider searches for these models is much higher than the sensitivity of direct DM searches. Besides sensitivity improvements originating from the increased amount of data, further enhancement can be expected from improvements in the analysis. Several possibilities for improvements seem feasible. First of all, the likelihood reconstruction introduced in Section 8.2 promises enhanced sensitivity in comparison to results based solely on $\not{E}_T$ as higher DM particles and mediator masses come into reach through increased statistical precision of data. Additionally, as discussed in Section 8.4, the likelihood reconstruction could be improved through the optimization of the likelihood with a slightly lower $\not{E}_T$ requirement, the inclusion of the p_T of the lepton into the likelihood function or the dedicated study of the second minima in the χ^2, if present. Secondly, other aspects of the analysis, apart from the likelihood reconstruction, can be improved. Systematic uncertainties can be further reduced by estimating the main backgrounds entirely from control regions in data instead of deriving only the normalization of the backgrounds from these. Adding the electron channel can, depending on the p_T thresholds of the available electron triggers, almost double the amount of studied data, leading to an increase in sensitivity by a factor of up to roughly $\sqrt{2}$. Moreover, a combination with other analyses performed for DM production with a top quark pair can amplify the sensitivity of this search. A combination has already been performed for the full hadronic, lepton+jets resolved and di-leptonic channel for 2015 data in [3,4]. The resolved lepton+jets analysis would only have to add a veto on top tagged jets, as discussed in Section 8.5, to combine the search presented in this thesis with the other three channels. Orthogonality with the other two channels is already ensured by corresponding lepton vetoes in the corresponding channels. New analysis techniques promise improvements as well. Among these, the usage of PUPPI for the mitigation of pileup seems interesting since PUPPI corrected $\not{E}_T$ shows a better resolution [201] than type-1-PF-$\not{E}_T$ adopted in the presented search. Additionally, PUPPI corrected jets exhibit a more stable top tagging performance as a function of the number of reconstructed vertices in the event (see Chapter 7). With the high instantaneous luminosities already reached in the year 2016, a stable performance for various pileup conditions is crucial for the data acquisition in 2017.

The theoretical models used to interpret the results are currently being improved to enable a study of more realistic scenarios. This includes, among other aspects, models that also take into account a mixing of the new scalar mediator with the SM Higgs boson. The inclusion of the mixing leads, for example, to more realistic kinematic properties of the model as a whole and enables combinations with searches for other DM signatures like mono-jet searches. A discussion of the improvements performed for the theoretical models can be found e.g. in [55]. Furthermore, the $t\bar{t} + \chi\overline{\chi}$ search with a pseudoscalar mediator can also be interpreted in axion models like e.g. [215].

9 Summary

A search for third-generation leptoquarks decaying into a top quark and a tau lepton and a search for Dark Matter produced in association with a top quark pair were presented. While the leptoquark analysis used data corresponding to $19.7\,\mathrm{fb}^{-1}$ recorded at $\sqrt{s} = 8\,\mathrm{TeV}$, the DM search is based on data corresponding to $12.9\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13\,\mathrm{TeV}$. Both data sets were collected by the CMS experiment at the LHC.

The event signature of the leptoquark analysis is based on the presence of a muon or an electron candidate, one hadronically decaying tau lepton candidate, at least three jets and large values of S_T. The main reducible background processes of this search arise from jets that are misreconstructed as tau lepton candidates. Since the background prediction is obtained from simulated events, it has to be ensured that the misidentification rate of jets as tau lepton candidates is well modeled by simulation. Therefore, the misidentification rate has been measured in a sideband enriched in $W +$ jets events. It has been found to be in the order of a few percent, but different in data and simulation. Therefore, p_T dependent correction factors for simulation have been derived that range from 0.59 to 1.06. The derived correction factors depend slightly on η of the misidentified tau lepton, and on the flavor of the jet that is reconstructed as a tau lepton. To cover these dependencies, systematic uncertainties varying between 20% and 65% have been applied on the measured correction factors. These correction factors and the predictions for SM processes based on simulation have been verified in two independent control regions. The statistical interpretation of the search is based on the distribution of the p_T of the leading tau lepton candidate, which shows a good discrimination power between signal and background processes as well as between different leptoquark mass hypotheses. In total, after the final event selection 60 (63) events have been observed in the muon (electron channel), which is in agreement with the SM prediction of 60.4 ± 10.9 (57.4 ± 11.2) events. Hence, no excess over the SM prediction and therefore no sign for physics beyond the SM has been found. With the performed search, third-generation leptoquarks with masses between 279 GeV (269 GeV expected) and 678 GeV (668 GeV expected) have been excluded at 95% C.L. assuming a branching ratio of 100% into a top quark and a tau lepton. The presented search has been combined with a search in the same-sign $\mu\tau$-channel [145]. By the combination, third-generation leptoquarks with masses below 685 GeV (695 GeV expected) have been excluded at 95% C.L, again assuming a 100% branching-ratio into $t\tau$. By combining the presented search with a search for pair produced bottom squarks decaying into a b quark and a neutralino [171], which can be interpreted as a search for leptoquarks decaying into $b\nu_\tau$, the results have been generalized to arbitrary branching ratios. With this combination, exclusion of third-generation leptoquarks below 560 GeV (605 GeV expected) has been achieved at 95% C.L over the full range of branching ratios. These exclusion limits are also directly interpretable as a search for pair production of bottom squarks decaying via an R-parity violating coupling into a top quark and a tau lepton. The results in the combined muon channels have first been published in [1]. The combination of the electron channel, the muon channel and the same-sign analysis are published in [2]. These are the most stringent exclusion limits to date.

The second analysis performed within this thesis is a search for DM particles produced in association with a top quark pair. In contrast to already existing results in the $t\bar{t} + \chi\overline{\chi}$ channel, which are based on resolved decays of the $t\bar{t}$ system, the presented analysis focuses on boosted $t\bar{t}$ decays, where the leptons produced in $t\bar{t}$ decays are not isolated or the fully-hadronic decay of the top quark is not resolvable, but reconstructed within a large-cone jet. The new approach yields enhanced sensitivity at low mediator masses compared to previous studies, since the SM background processes can be greatly reduced due to the application of top tagging algorithms. As a result, the $\not{E}_T$ requirement can be significantly loosened. This translates directly into higher selection efficiencies for signal models with low mediator masses and thus to an enhanced sensitivity. Optimization studies of the event selection adopted in this search have been performed by comparing the expected sensitivity of many different selections. The final event selection is based on the requirement of a muon candidate without imposing isolation criteria, at least two jets, where one of them is b tagged, $\not{E}_T > 80\,\mathrm{GeV}$ and $m_T > 100\,\mathrm{GeV}$. The background processes of this search have been estimated using simulated events. By studying control regions enriched in the main SM backgrounds, which are $t\bar{t} + \text{jets}$ and $W + \text{jets}$ production, the normalization of these two background processes have been constrained during the statistical interpretation of the search. For the latter, the $\not{E}_T$ distributions obtained in two independent control regions, which have been constructed by either requiring an isolated muon candidate and a top tagged jet or a non-isolated muon candidate, have been used. After the application of the complete event selection, 14161 (689) data events have been observed, while 13500^{+4400}_{-3600} ($770\pm^{+120}_{-110}$) background events are predicted in the signal-region comprising isolated (non-isolated) muon candidates. The distributions of the $\not{E}_T$ spectra show no deviation from the SM prediction. Scalar mediators with masses below 36 GeV have been excluded at 95% C.L. if DM particles with masses of 1 GeV and $g_{\mathrm{DM}} = g_{\mathrm{SM}} = 1$ are assumed. The obtained sensitivity is comparable to the resolved analysis presented in [3, 4], which uses data corresponding to $2.2\,\mathrm{fb}^{-1}$ and is a combination of analyses covering the electron+jets, muon+jets, di-leptonic and fully-hadronic decay modes of the $t\bar{t}$ system, as well as a search for DM produced in association with two b quarks.

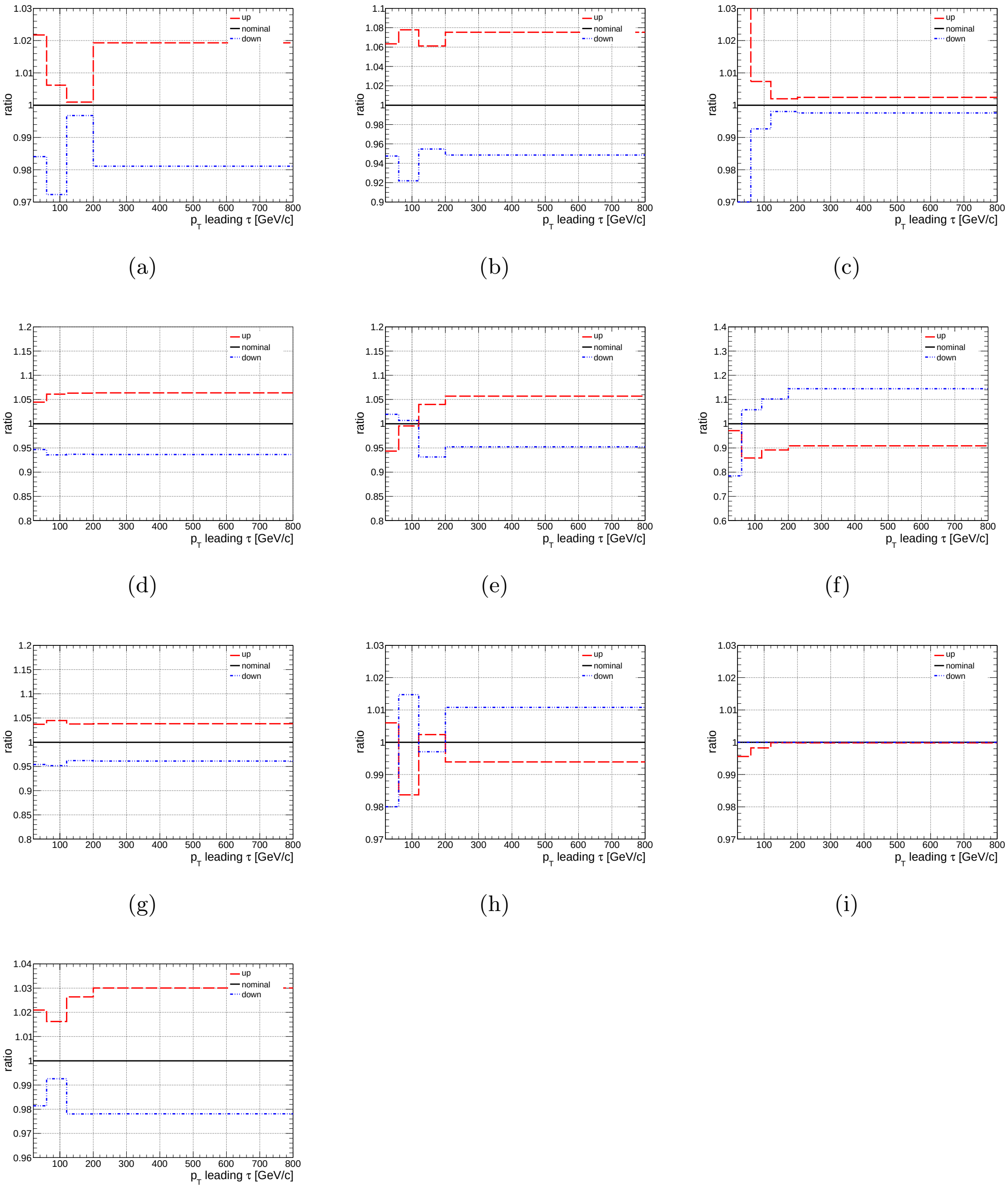
Since the main background, $t\bar{t} + \text{jets}$ production, has exactly the same visible final state as the signal processes, it is very hard to distinguish from signal processes. Within this thesis, a kinematical reconstruction method has been developed with the aim to find a variable with higher separation power between signal and background processes than $\not{E}_T$. The latter is usually used for the final statistical interpretation in $t\bar{t} + \chi\overline{\chi}$ collider searches. The likelihood reconstruction uses constrains from boosted $t\bar{t}$ events without considering the measured $\not{E}_T$ value to reconstruct the neutrino four-momentum in lepton+jets $t\bar{t}$ decays. By comparing the neutrino solution to the measured $\not{E}_T$ value, a variable with good separation power between lepton+jets $t\bar{t}$ events and signal processes has been found. The method works very well in the phase space it was designed for. However, after applying the optimized event selection better results have been achieved when using the $\not\!\!E_T$ distributions for the statistical interpretation of the results. Nonetheless, the likelihood reconstruction method promises enhanced performance especially for higher mediator and DM masses, although further optimization work is needed to be able to outperform the results obtained using $\not\!\!E_T$.

In addition to the two searches, the impact of two pileup mitigation techniques, PUPPI and CHS, on top tagging efficiencies and misidentification rates has been studied. The top tagging algorithms yield much more stable performances with increasing pileup events after the application of PUPPI than after the application of CHS. Thus, PUPPI will become the default algorithm to be used in top tagging at the CMS experiment. Additionally, existing top tagging algorithms have been reviewed and adopted within this thesis. Working points for usage of top tagging algorithms have been derived. These are currently the recommended working points to be used for analyses performed at the CMS experiment [6].

While no signs of new physics have been found with direct searches at the LHC so far, data acquisition at a center-of-mass-energy of 13 TeV has just restarted. The recording of a data set corresponding to at least $40\,\mathrm{fb}^{-1}$ is expected in 2017, which will significantly enhance the sensitivity of new physics searches.

Appendices

A Uncertainties leptoquark analysis



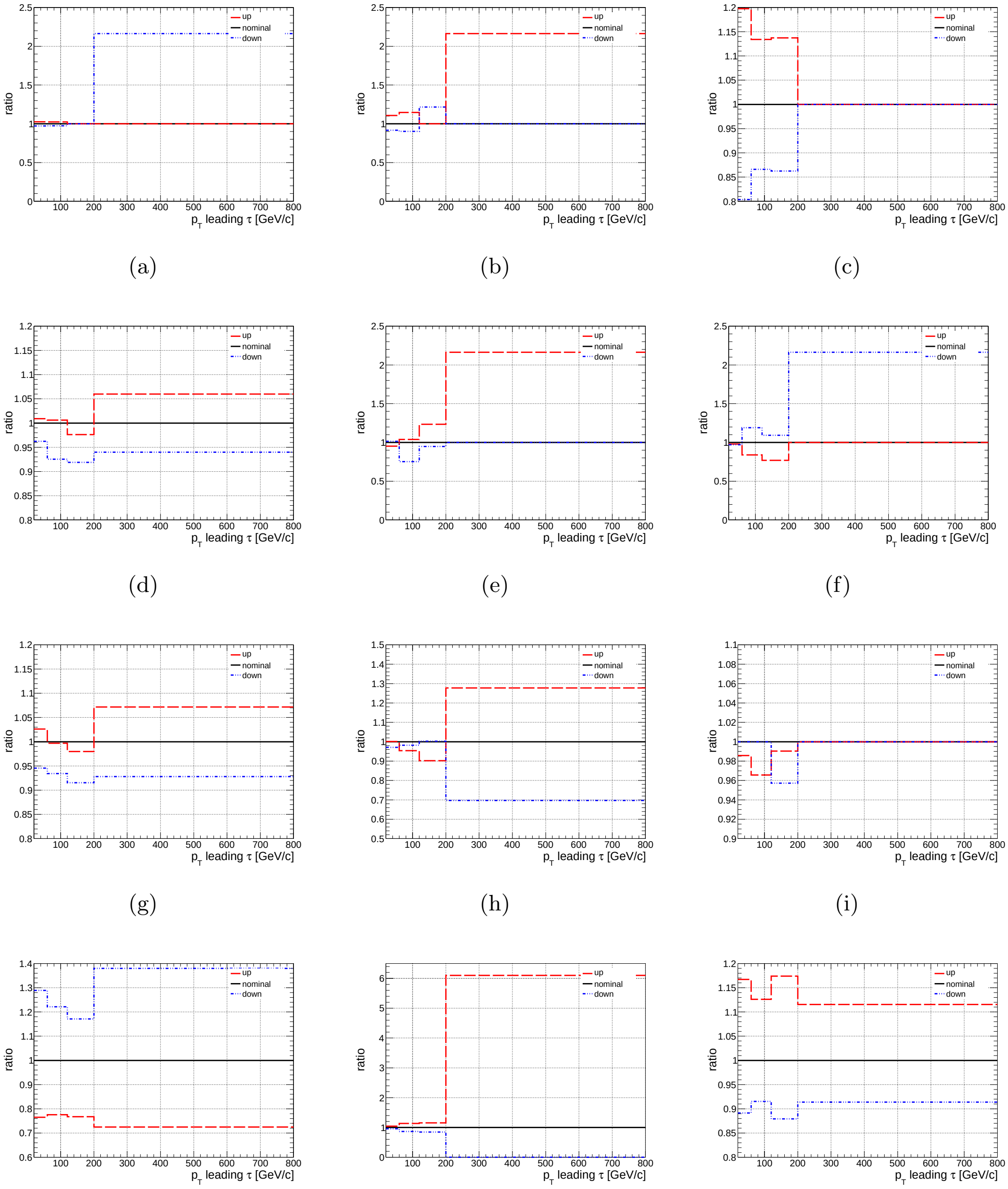
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j)

Figure A.1: Relative change in the $p_T^{\tau 1}$ distribution for the signal sample for a leptoquark with a mass of 500 GeV in the electron channel due to uncertainties in (a) JER, (b) JEC, (c) tau lepton misidentification rate, (d) tau lepton identification, (e) tau lepton energy scale, (f) tau lepton energy resolution, (g) electron identification and trigger, (h) pileup, (i) misidentification rate of an electron as a tau lepton candidate and (j) pdf.



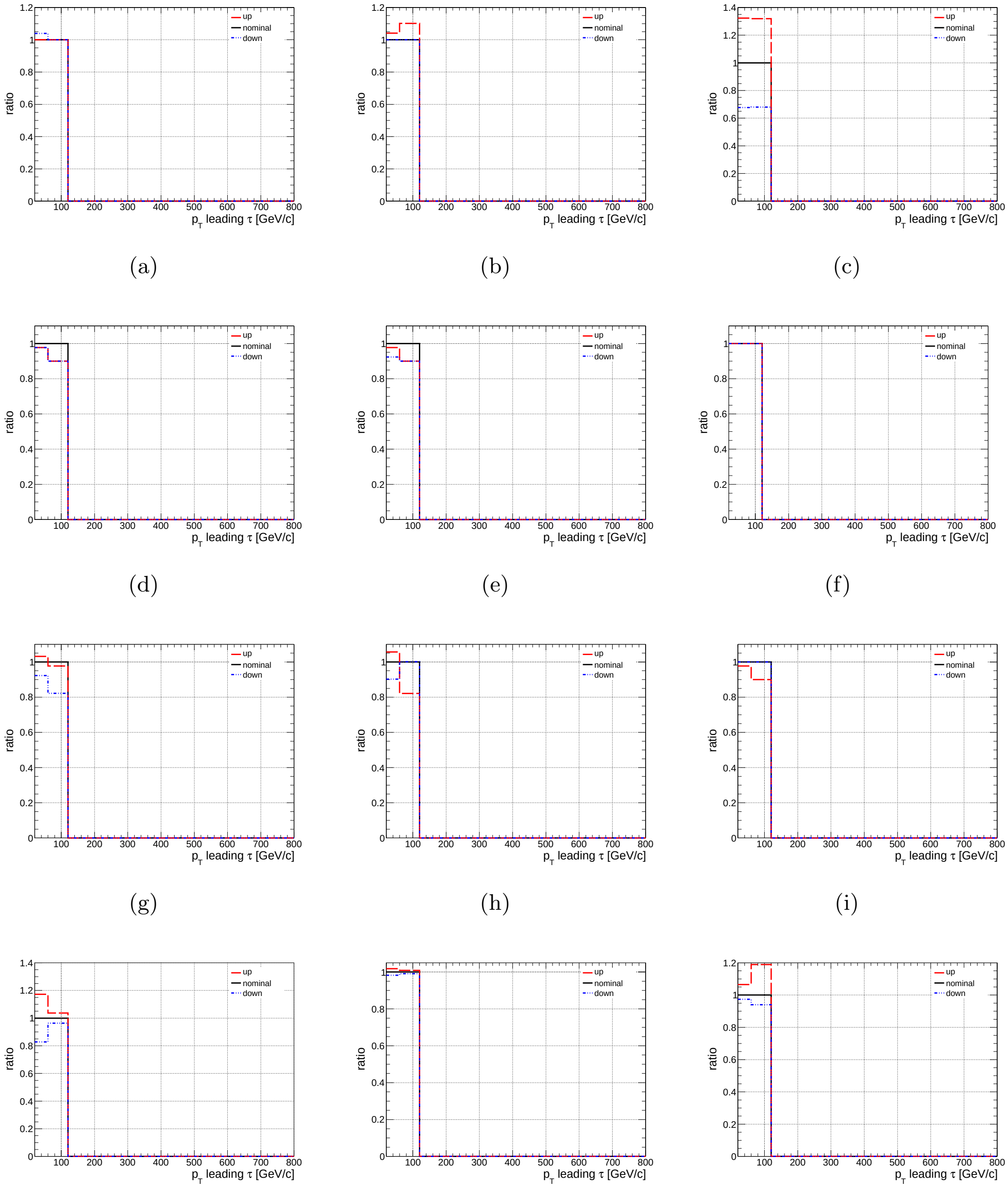
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.2: Relative change in the $p_T^{\tau 1}$ distribution for $t\bar{t}+$jets processes in the electron channel due to uncertainties in (a) JER, (b) JEC, (c) tau lepton misidentification rate, (d) tau lepton identification, (e) tau lepton energy scale, (f) tau lepton energy resolution, (g) electron identification and trigger, (h) pileup, (i) misidentification rate of an electron as a tau lepton candidate (j) the top p_T re-weighting (k) the scale uncertainty and (l) pdf.



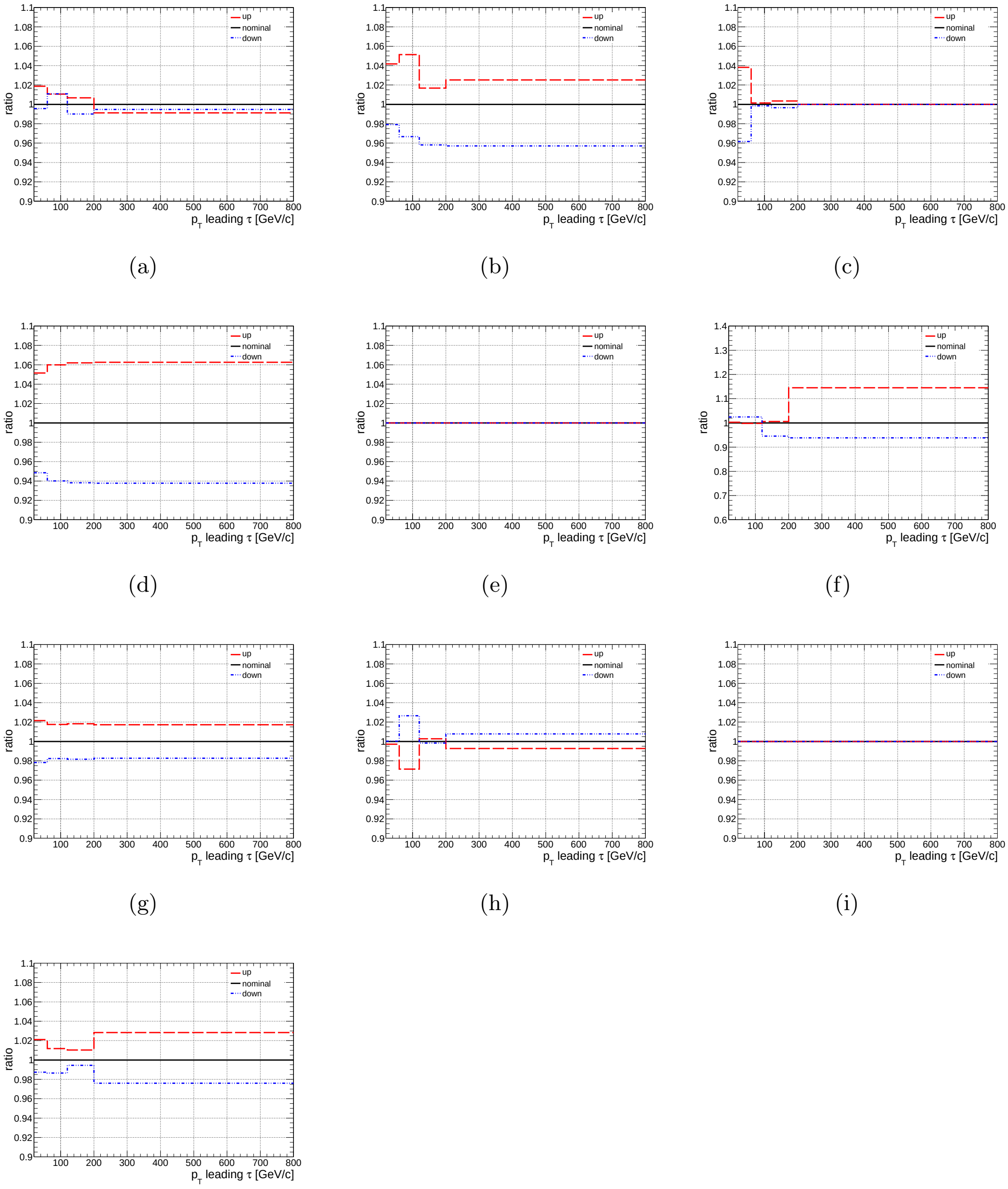
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.3: Relative change in the $p_T^{\tau 1}$ distribution for the $W+$jets processes in the electron channel due to uncertainties in (a) JER, (b) JEC, (c) tau lepton misidentification rate, (d) tau lepton identification, (e) tau lepton energy scale, (f) tau lepton energy resolution, (g) electron identification and trigger, (h) pileup, (i) misidentification rate of an electron as a tau lepton candidate, (j) the scale uncertainty, (k) the matching uncertainty and (l) pdf.



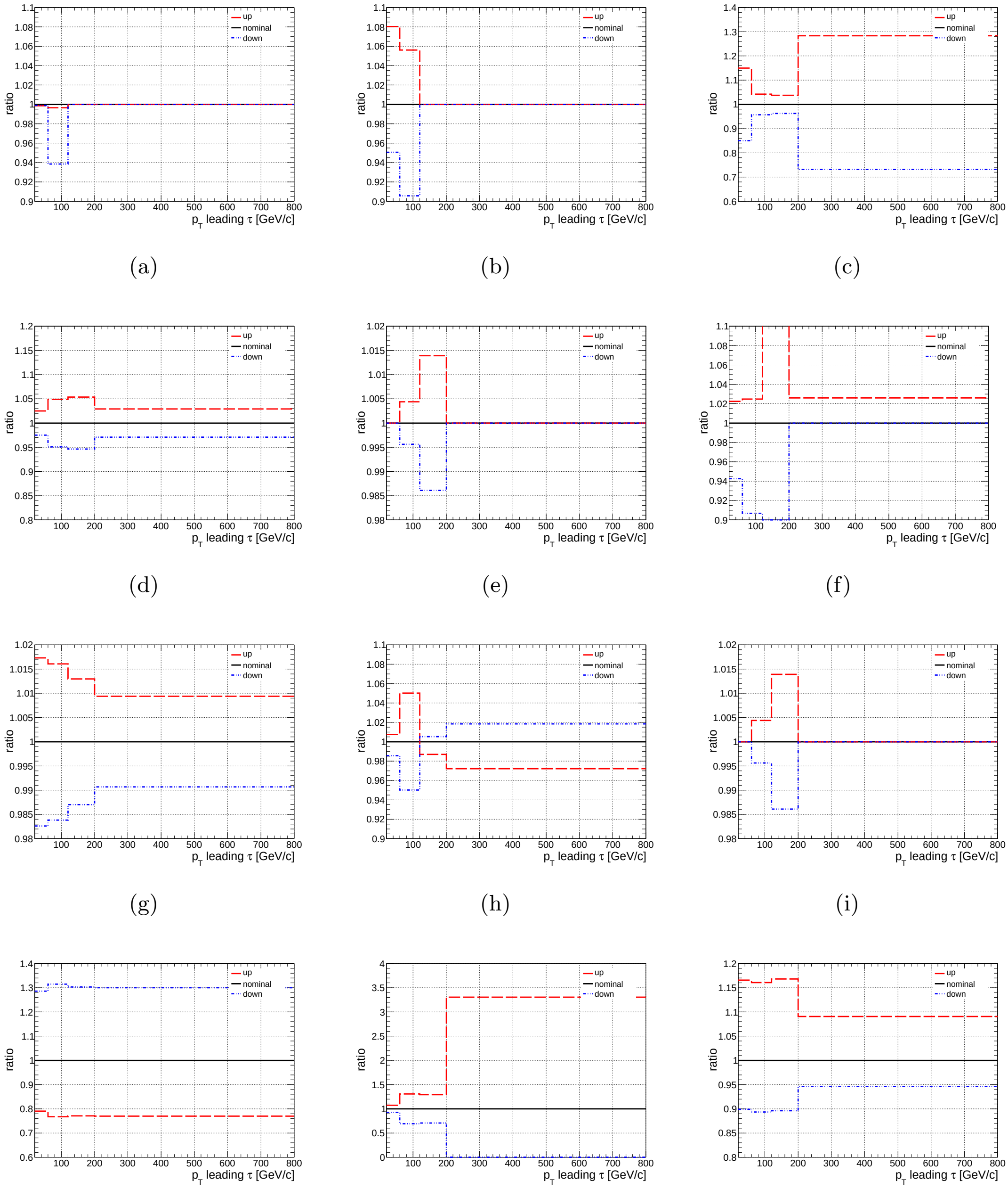
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j)

Figure A.4: Relative change in the $p_T^{\tau 1}$ distribution for the signal sample for a leptoquark with a mass of 500 GeV in the muon channel due to uncertainties in (a) JER, (b) JEC, (c) tau lepton misidentification rate, (d) tau lepton identification, (e) tau lepton energy scale, (f) tau lepton energy resolution, (g) muon identification and trigger, (h) pileup, (i) misidentification rate of an electron as a tau lepton candidate, and (j) pdf.



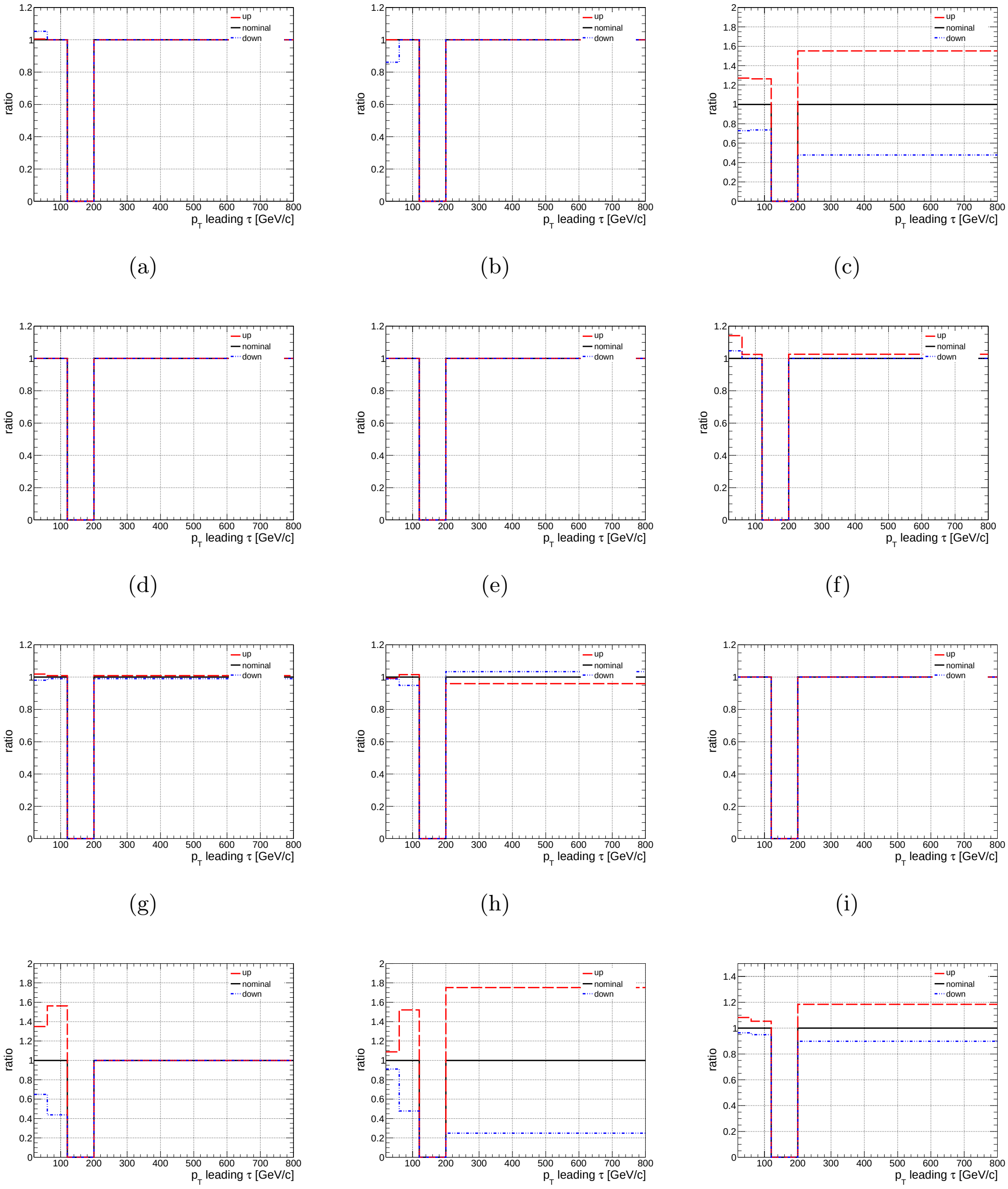
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.5: Relative change in the $p_T^{\tau 1}$ distribution for the $t\bar{t}$ + jets processes in the muon channel due to uncertainties in (a) JER, (b) JEC, (c) tau lepton misidentification rate, (d) tau lepton identification, (e) tau lepton energy scale, (f) tau lepton energy resolution, (g) muon identification and trigger, (h) pileup, (i) misidentification rate of an electron as a tau lepton candidate, (j) the top p_T re-weighting (k) the scale uncertainty and (l) pdf.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.6: Relative change in the $p_T^{\tau 1}$ distribution for the W+jets processes in the muon channel due to uncertainties in (a) JER, (b) JEC, (c) tau lepton misidentification rate, (d) tau lepton identification, (e) tau lepton energy scale, (f) tau lepton energy resolution, (g) electron identification and trigger, (h) pileup, (i) misidentification rate of an electron as a tau lepton candidate,(j) the scale uncertainty, (k) the matching uncertainty and (l) pdf.

B Determination of muon trigger efficiencies

Correction factors are applied to simulated events to model the data efficiency of the trigger selection (HLT_TkMu50 or HLT_Mu50) in the $t\bar{t} + \chi\overline{\chi}$ analysis presented in Chapter 8. Correction factors provided by the group responsible for muon triggering, reconstruction and identification within the CMS experiment are used [204]. They correspond to the trigger efficiency in data since no trigger emulation is implemented in simulation. The corrections are determined using the Tag-and-Probe method [160] in a sample enriched in $Z(\mu^+\mu^-)$+jets. However, the provided correction factors require the probe-lepton to pass the high-p_T muon identification criteria [120]. This identification working point has not been used in the presented analysis since it was not designed for usage in conjunction with PF objects. Therefore, studies have been performed in the course of this work to validate the usage of the corrections provided by the CMS experiment, which use the high-p_T muon identification criteria, in the $t\bar{t} + \chi\overline{\chi}$ analysis, which uses the medium working point for muon identification. Along with this, studies are performed that confirm the validity of the provided correction factors, which were determined in a phase space containing isolated muon candidates with small or moderate p_T, in the kinematic phase space of the $t\bar{t} + \chi\overline{\chi}$ analysis.

The determination of the trigger efficiencies is performed for the trigger HLT_Mu45_eta2p1 using data collected by the CMS experiment during the data acquisition periods of the year 2015. The data were recorded at $\sqrt{s} = 13\,\text{TeV}$ and correspond to an integrated luminosity of approximately $2.7\,\text{fb}^{-1}$. The objects used in this study are very similar to the ones used in the search presented in Chapter 8. Analogously to the trigger HLT_Mu50 utilized in the Dark Matter search, the trigger HLT_MU45_eta2p1 requires a Global Muon candidate on trigger level with $p_T > 45\,\text{GeV}$ and $|\eta| < 2.1$. Since the two triggers mainly differ in the imposed p_T thresholds on the muon candidates, the conclusions derived in the studies presented here are transferable to the search presented in Chapter 8. The trigger efficiency is determined with the Tag-and-Probe method in a data sample dominated by $Z(\mu^+\mu^-)$+jets events. The determination follows closely the approach pursued for the corrections provided by the CMS experiment. The general idea is to select a pair of muon candidates that have an invariant mass compatible with the mass of the Z boson. By requiring that one of the muon candidates, referred to as tag-muon, has fired a single muon trigger a sample of probe-muons is created. The probe-muon sample consists of the second of the two muon candidates in each event. Since the trigger decision made for the tag-lepton is independent from the trigger result for the probe-lepton, the probe-lepton sample can be used to determine the trigger efficiencies by counting the number of probe-muons that pass the trigger criteria under study with respect to all probe-muons. Thus, the trigger efficiency ϵ is given by

$$\epsilon = \frac{N(\text{probe-muons passing the trigger selection})}{N(\text{probe-muons})} \tag{B.1}$$

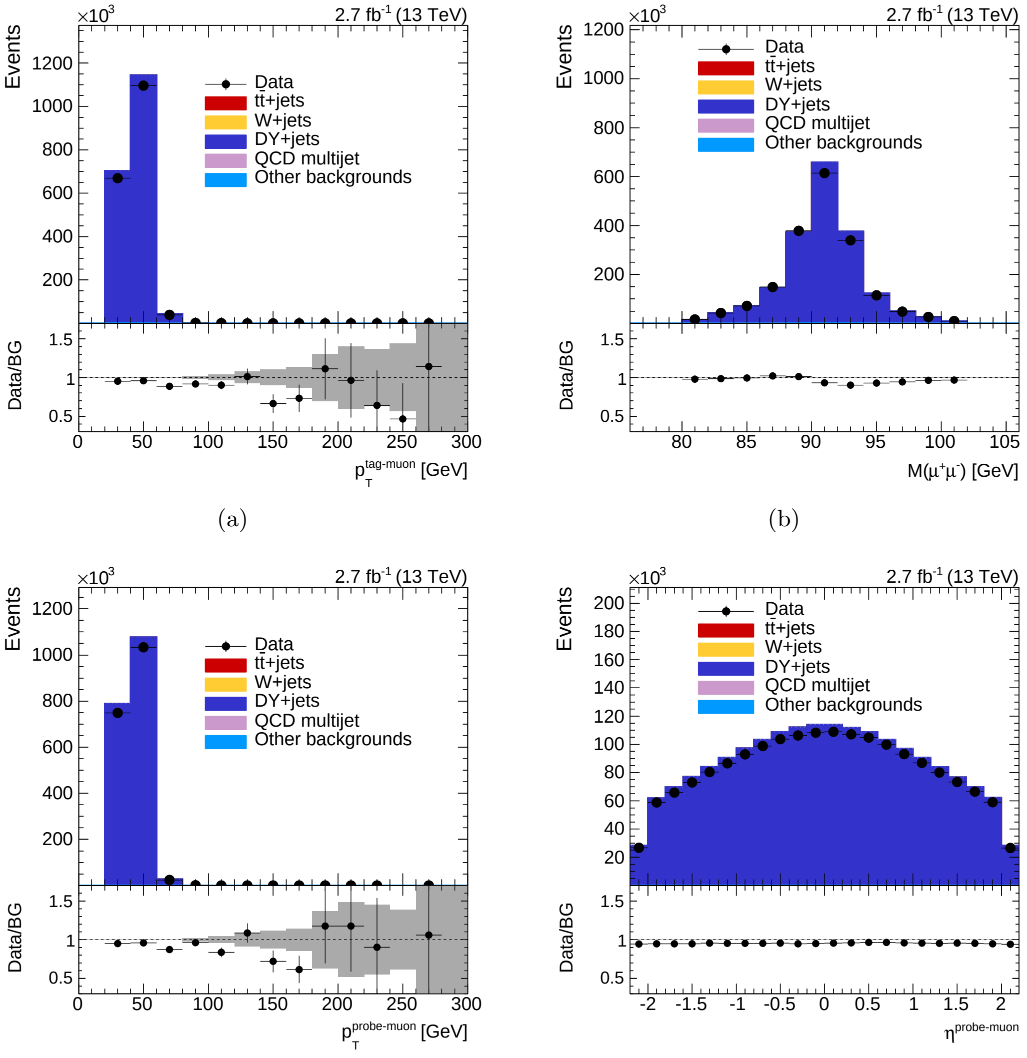
To check if a muon candidate passes the trigger selection, a matching of the reconstructed muon candidate to muon candidates on trigger-level that fulfill the trigger requirements is performed. If such a trigger-level object is found within $\Delta R < 0.1$ to the reconstructed muon candidate, the latter is considered to have fired the trigger. Additionally, the statistical precision of the efficiencies can be roughly doubled by sorting the leptons according to the sign of their charges and by considering each of them as a tag-and as a probe-lepton. In the second step the role of the two lepton candidates is switched.

With the described approach the trigger efficiency is determined in two different phase spaces in the following. In both cases, the background from non-DY events is negligible after the selections have been applied. Hence, no background subtraction is performed. Only data events certified by the CMS experiment to be used in analyses are considered and all recommended event filters are applied. Since also jet selections are applied, the jet energy corrections are applied to jets in data and simulation. Pileup re-weighting and a correction of the muon identification efficiencies are performed in simulation. Furthermore, each studied event has to contain at least one primary vertex candidate passing the identification requirements described in Section 5.2.

The first of the studied selections is very similar to the one applied for the determination of the scale factors provided by the CMS experiment. The main difference between the two selections is that the muon candidate has to pass the medium instead of the high-p_T muon identification criteria. The selection applied in the following aims at justifying that the provided corrections, which use the high-p_T muon identification for the probe-muons, are also applicable in analyses using the medium identification requirements. To pass the studied selection all events have to fulfill the requirements below. These are denoted as *jet veto selection*.

- Each event contains exactly two medium muon candidates with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.1$.
- Events with electron candidates that pass loose identification criteria, $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ or with additional jets fulfilling the loose PF identification requirements with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ are discarded.
- The invariant mass of the muon candidates has to lie around the mass of the Z boson ($81\,\text{GeV} < M(\mu\mu) < 101\,\text{GeV}$).

From the selected muon candidate pairs the tag-muon and the probe-muon candidates are defined. The *tag-muon* has to pass the tight identification and isolation criteria (see Section 5.3, the tight identification requirements are the same as those used in Run-I), $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$. Furthermore, the tag-muon has to be matched within $\Delta R < 0.1$ to a trigger-level object that passes the trigger HLT_IsoMu27, which requires an isolated Global Muon candidate with $p_T > 27\,\text{GeV}$ on trigger level. The *probe-muon* has to pass the medium identification requirement, $p_T > 25\,\text{GeV}$ and $|\eta| < 2.1$.



(a)

(b)

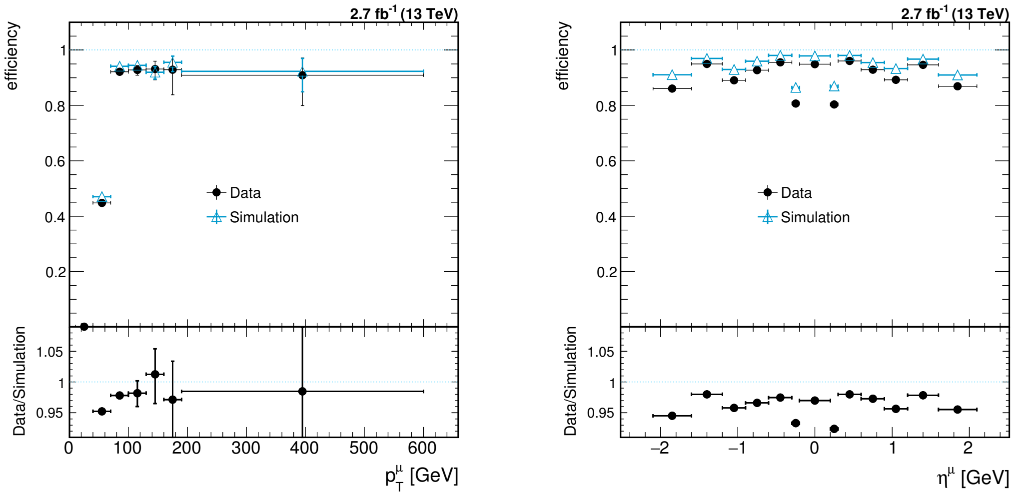
(c)

(d)

Figure B.1: Comparison of data (*black dots*) and simulation (*colored areas*) after the jet veto selection has been applied. Shown are the distributions of (a) the p_T of the tag-muon, (b) the invariant mass of the tag- and the probe-muon, (c) the p_T and (d) η of the probe-muon.

In Fig. B.1 control distributions after the selection and the assignment of probe- and tag-muons are presented. Shown are the p_T of the tag-muon, the invariant mass of the tag- and the probe-muon, and the p_T and η of the probe-muons. In all distributions data and simulation agree well. Furthermore, the sample is almost completely dominated by DY events. The contributions from other background processes is negligible.

With the selected probe-muons the trigger efficiency can be determined according to Eq. (B.1). The result is shown in Fig. B.2(a) as a function of p_T of the probe-muon. The trigger efficiency reaches a plateau efficiency above 90% for muon candidates with $p_T > 47\,\text{GeV}$. The trigger efficiency as a function of η is presented in Fig. B.2(b). Only muon candidates with $p_T > 47\,\text{GeV}$, meaning muon candidates in the plateau of the efficiency, are considered. The η dependency looks like expected including the two dips in the efficiencies for $|\eta| \approx 0.25$ and $|\eta|$ between 0.8 and 1.2, where transition regions between different sub-systems within the muon system of the CMS experiment are lo-



(a) (b)

Figure B.2: Comparison of the trigger efficiency in data (*black dots*) and simulation (*light blue triangles*) as a function of (a) p_T and (b) η (for muon candidates with $p_T > 47\,\text{GeV}$) of the probe-muon after the jet veto selection has been applied.

cated. Figure B.3 shows the 2D trigger efficiency divided in the same p_T and η ranges as the provided efficiencies [204]. They are compatible with the results determined by the group responsible for muon triggering within the CMS experiment. Thus, the efficiencies provided by the CMS experiment that are determined for the high-p_T identification requirements are also applicable for analyses using the medium identification requirements.

To check if the correction factors determined in a phase space with isolated muon candidates and without high-p_T jets can be transferred to the phase space of the analysis presented in Chapter 8, the determination of the trigger efficiencies is repeated with application of a selection closer to the one of the DM search presented in Chapter 8. The selections differ mainly in the requirement of the presence of additional jets. Instead of vetoing jets, each event has to contain at least one jet. Additionally, no isolation criterion is imposed on the muon candidate and the tag-muon candidate has to pass the trigger HLT_Mu45_eta2p1. To suppress the QCD multijet background the 2D-cut (see Section 8.1) is applied. Hence, events have to pass the following selection criteria, called *jet selection* in the following:

- Each event contains exactly two medium muon candidates with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.1$.

- Events with electron candidates $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ that pass loose identification criteria are not considered.

- At least one jet fulfilling the loose PF identification criteria with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ has to be present. The jet lepton cleaning described in Section 8.9 is applied.

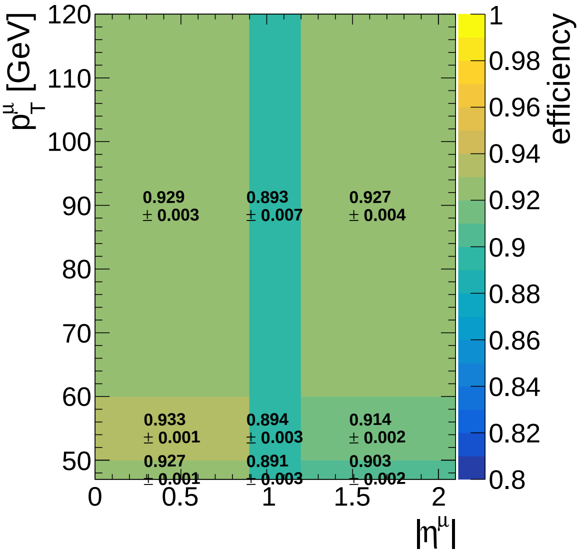


Figure B.3: Trigger efficiency in data as a function of η and p_T of the probe-muon after the jet veto selection has been applied.

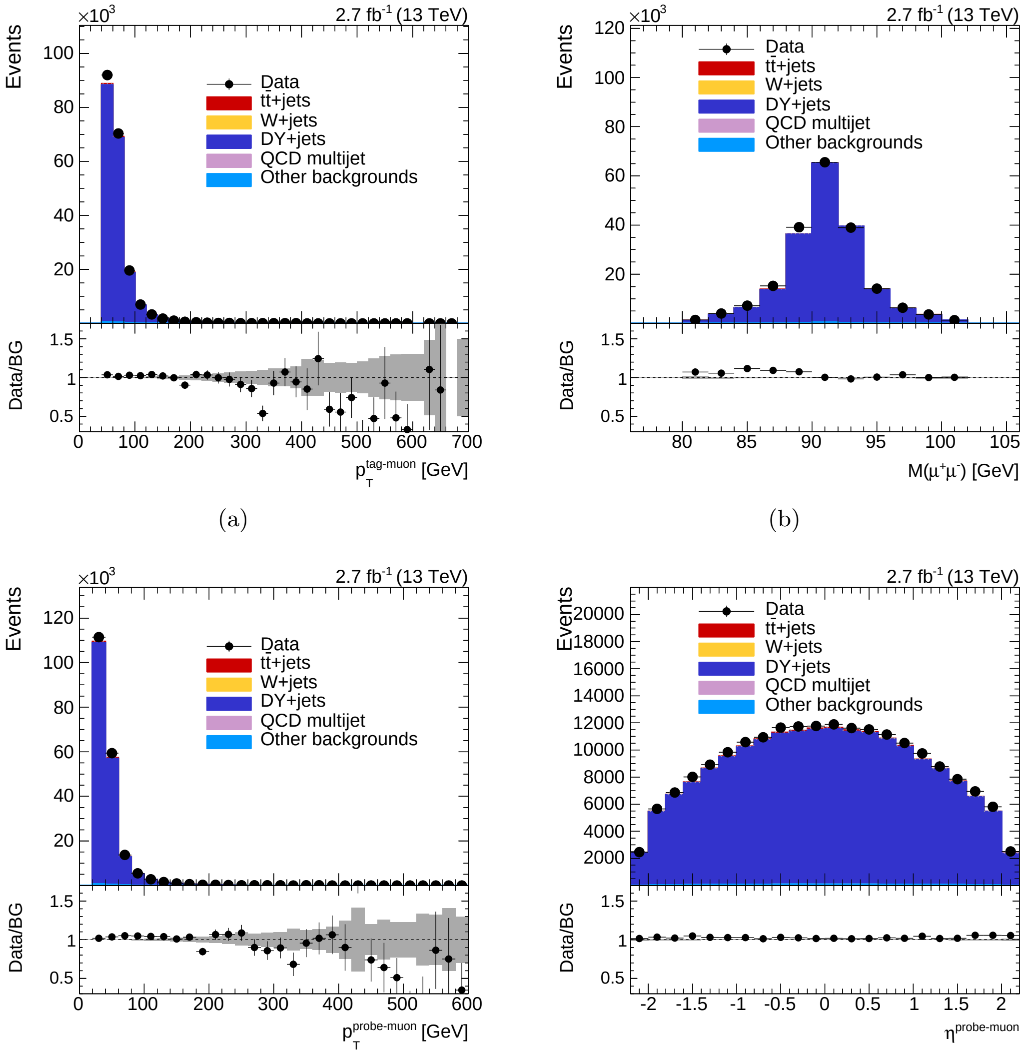
- The 2D-cut is applied (see Section 8.1), whereby the muon candidates have to pass either $\Delta R(\mu, \text{jet}) > 0.4$ or $p_{\mathrm{T}}^{\mathrm{rel}} > 20\,\mathrm{GeV}$ with respect to all jets with $p_T > 15\,\mathrm{GeV}$.
- The invariant mass of the muon candidates has to lie around the mass of the Z boson ($81\,\mathrm{GeV} < M(\mu\mu) < 101\,\mathrm{GeV}$).

The tag-muon has to fulfill the tight identification requirements, $p_T > 47\,\mathrm{GeV}$ and $|\eta| < 2.1$. Additionally, the tag-muon has to be matched within $\Delta R < 0.1$ to a trigger-level object of the trigger HLT_Mu45_eta2p1. The probe-muon has to pass the medium identification requirement, $p_T > 25\,\mathrm{GeV}$ and $|\eta| < 2.1$.

Control distributions after the selection are shown in Fig. B.4. The p_T of the tag-muon, the invariant mass of the tag- and the probe-muon, and the p_T and η of the probe-muon are presented. Data and simulation agree very well. Again, the sample is almost completely dominated by DY + jets events. Small contributions from other processes make up less than 2%. Since these 2% are dominated by $t\bar{t}+$jets and diboson events, which both contain real muons, no subtraction of background processes has to be performed for the determination of the trigger efficiencies.

The trigger efficiencies obtained with the jet selection according to Eq. (B.1) are shown in Fig. B.5. With this selection a better statistical precision can be achieved at high transverse momenta of the muon candidates than with the jet veto selection presented before. Since the jet selection and the jet veto selection are orthogonal and thus do not share selected events, the trigger efficiencies are independent from each other and can be compared within their statistical uncertainties. A comparison shows that they are compatible in the efficiency plateau. Figure B.6 illustrates the efficiencies as a function of η and p_T in the ranges as the efficiencies provided by the CMS experiment. These are compatible to the ones derived before and to the efficiencies provided by the CMS experiment.

In conclusion, the performed studies justify the application of the corrections provided by

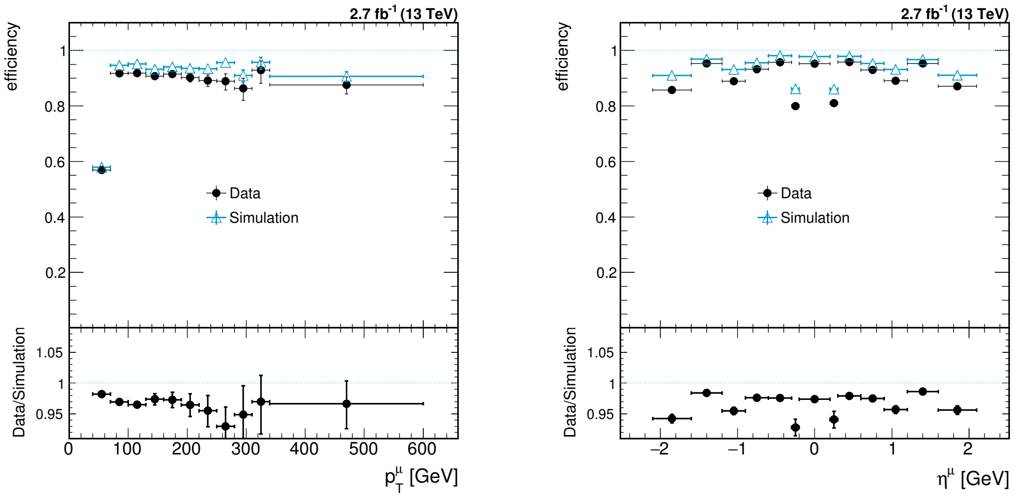


(a) (b)

(c) (d)

Figure B.4: Comparison of data (*black dots*) and simulation (*colored areas*) after the jet selection has been applied. Shown are the distributions of (a) the p_T of the tag-muon, (b) the invariant mass of the tag- and the probe-muon, (c) the p_T and (d) η of the probe-muon.

the CMS experiment [204] in analyses adopting the medium identification requirements. Additionally, the validity of the corrections provided by the CMS experiment in the phase space of the analysis searching for DM presented in Chapter 8 has been demonstrated.



(a) (b)

Figure B.5: Comparison of the trigger efficiency in data (*black dots*) and simulation (*light blue triangles*) as a function of (a) p_T and (b) η (for muon candidates with $p_T > 47$ GeV) of the probe-muon after the jet selection has been applied.

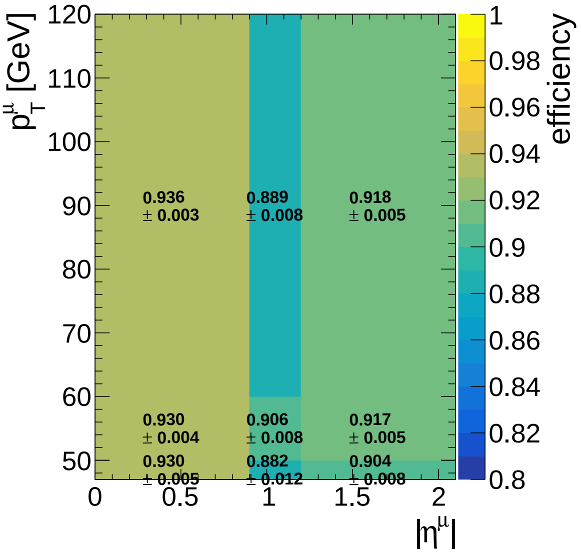
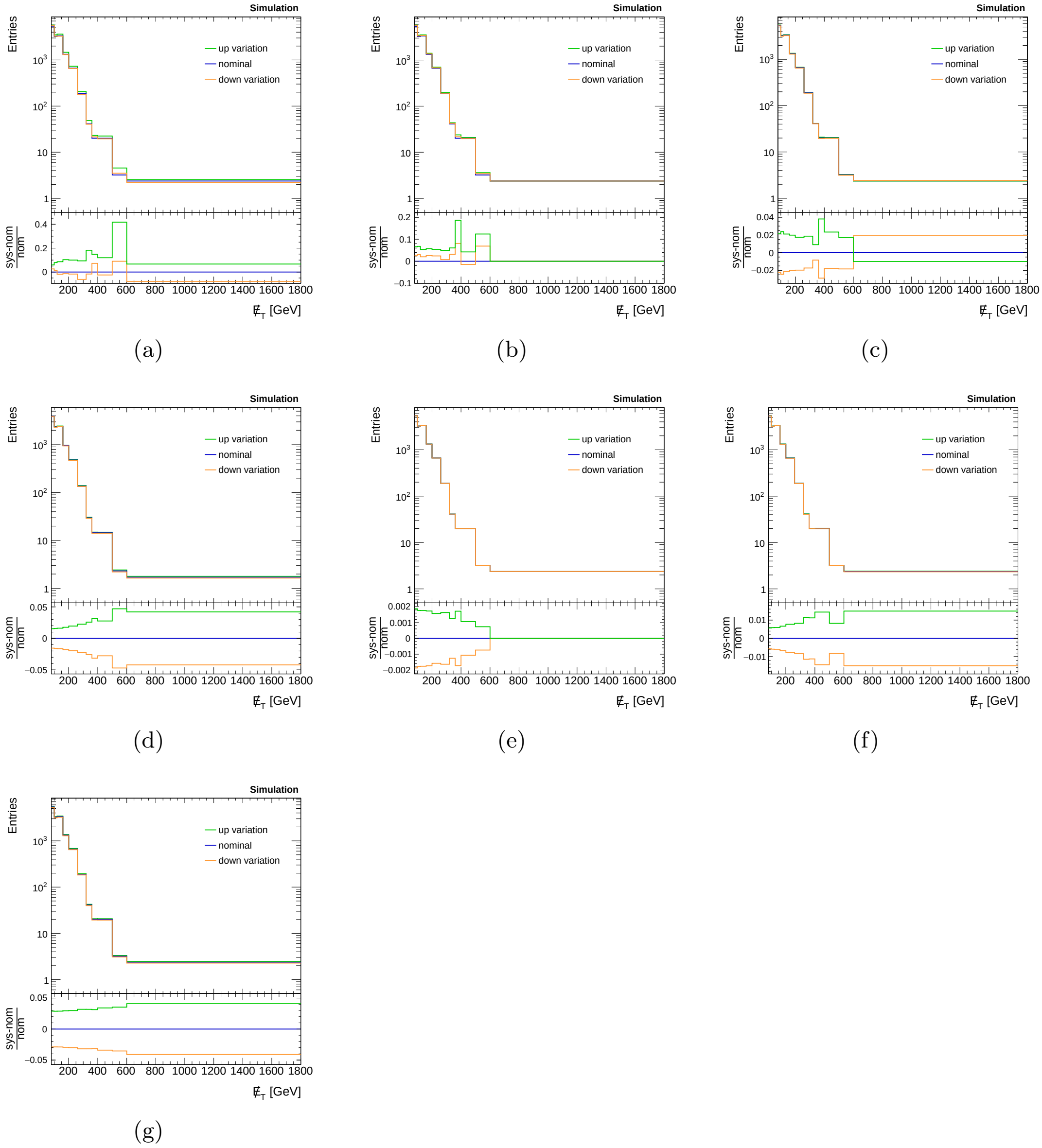


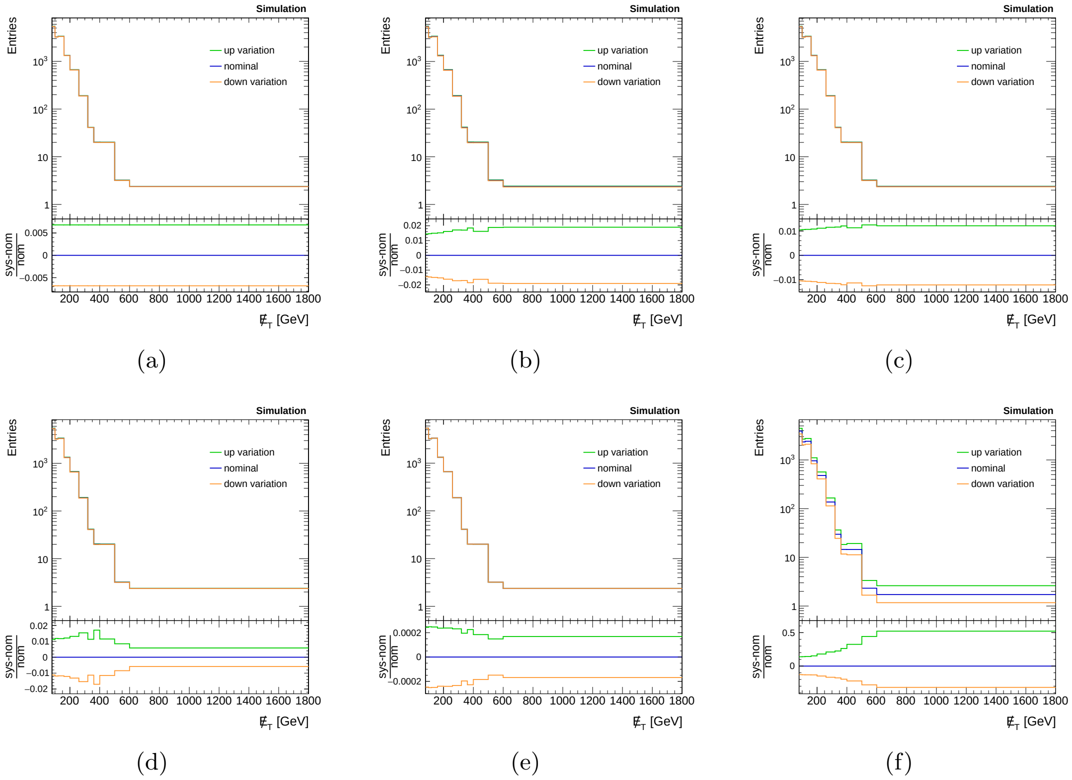
Figure B.6: Trigger efficiency in data as a function of η and p_T after the jet veto selection has been applied.

C Uncertainties $t\bar{t} + \chi\overline{\chi}$ analysis



(a) (b) (c) (d) (e) (f) (g)

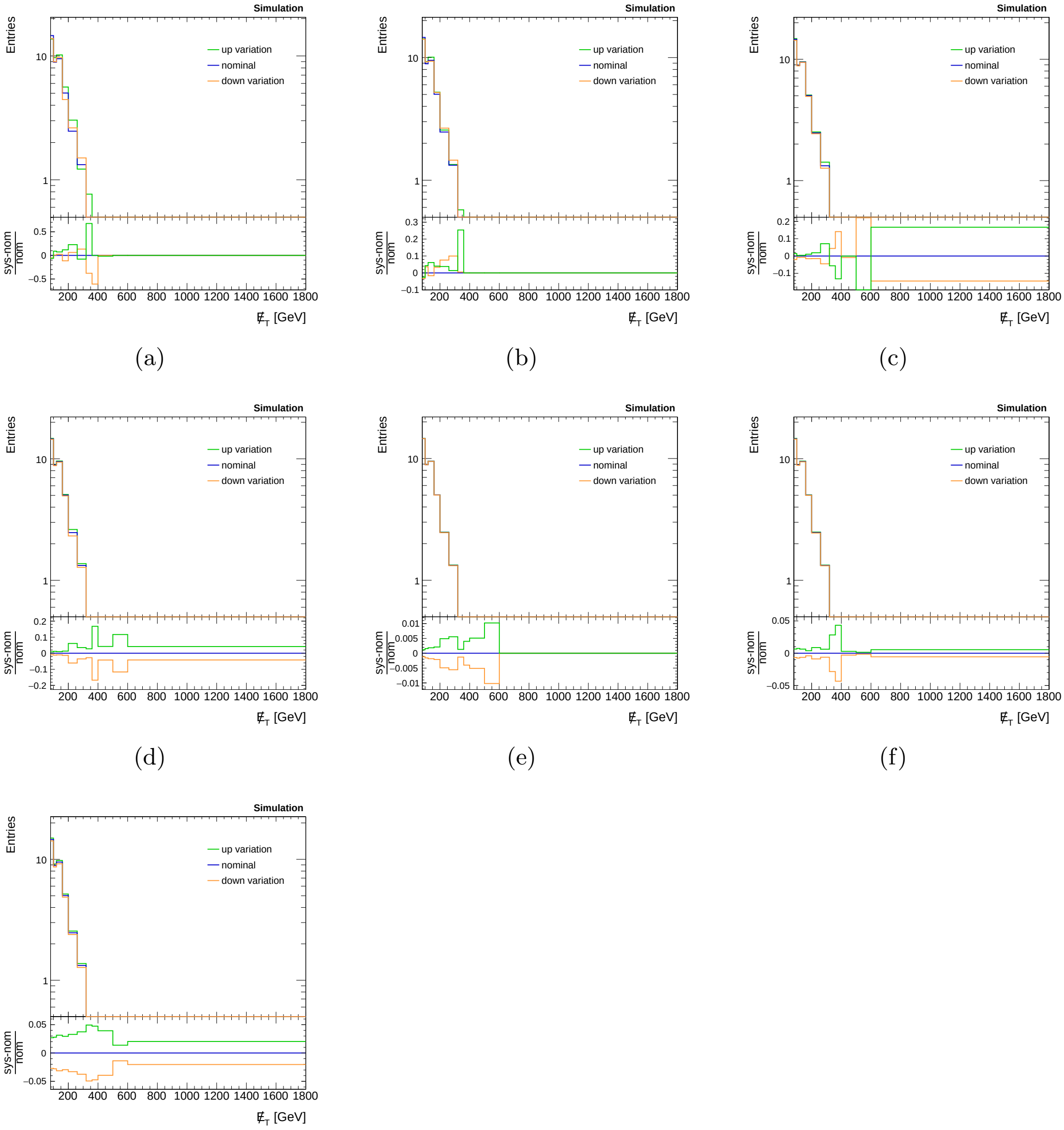
Figure C.1: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not{E}_T$ spectra in the isolated lepton signal region obtained for the $t\bar{t}$ + jets background processes. Shown are the systematic uncertainties in the (a) jet energy corrections, (b) jet energy resolution, (c) pileup re-weighting , (d) pdf, (e) top tag correction factor, (f) b tagging (udsg), and (g) b tagging (bc).



(a) (b) (c)

(d) (e) (f)

Figure C.2: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not\!\!E_T$ spectra in the isolated lepton signal region obtained for the $t\bar{t}+$jets background processes. Shown are the systematic uncertainties in the (a) m_T efficiency correction, (b) muon identification, (c) muon isolation, (d) muon trigger, (e) muon tracking, and (f) renormalization and factorization scale.

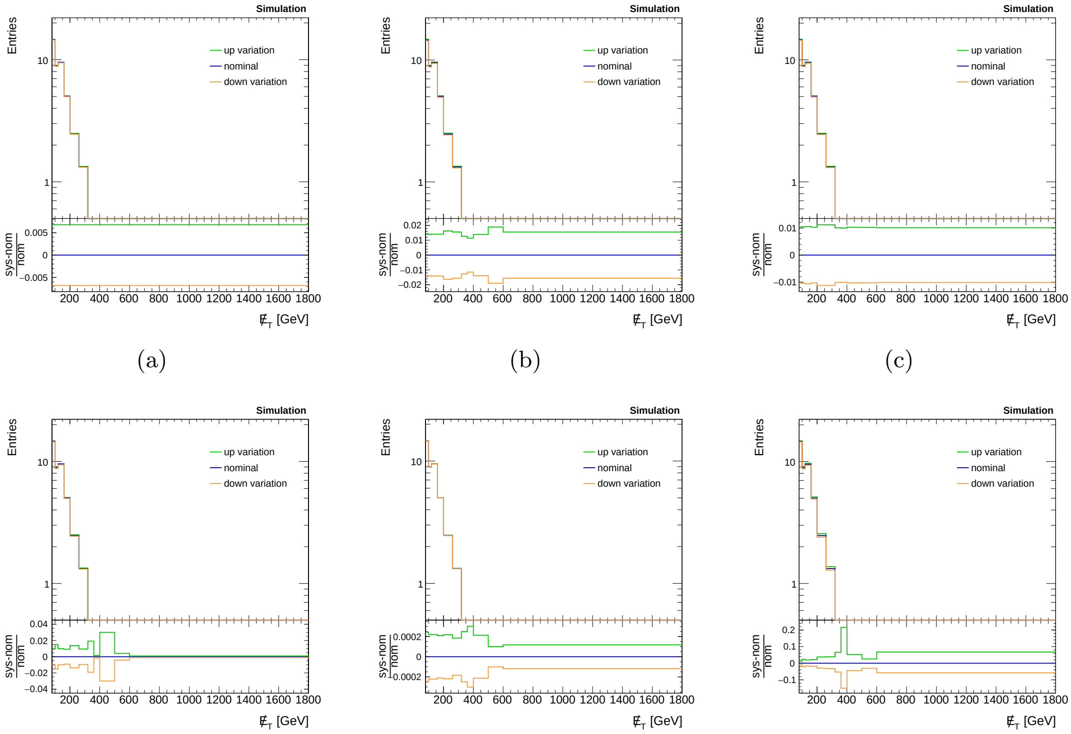


(a) (b) (c)

(d) (e) (f)

(g)

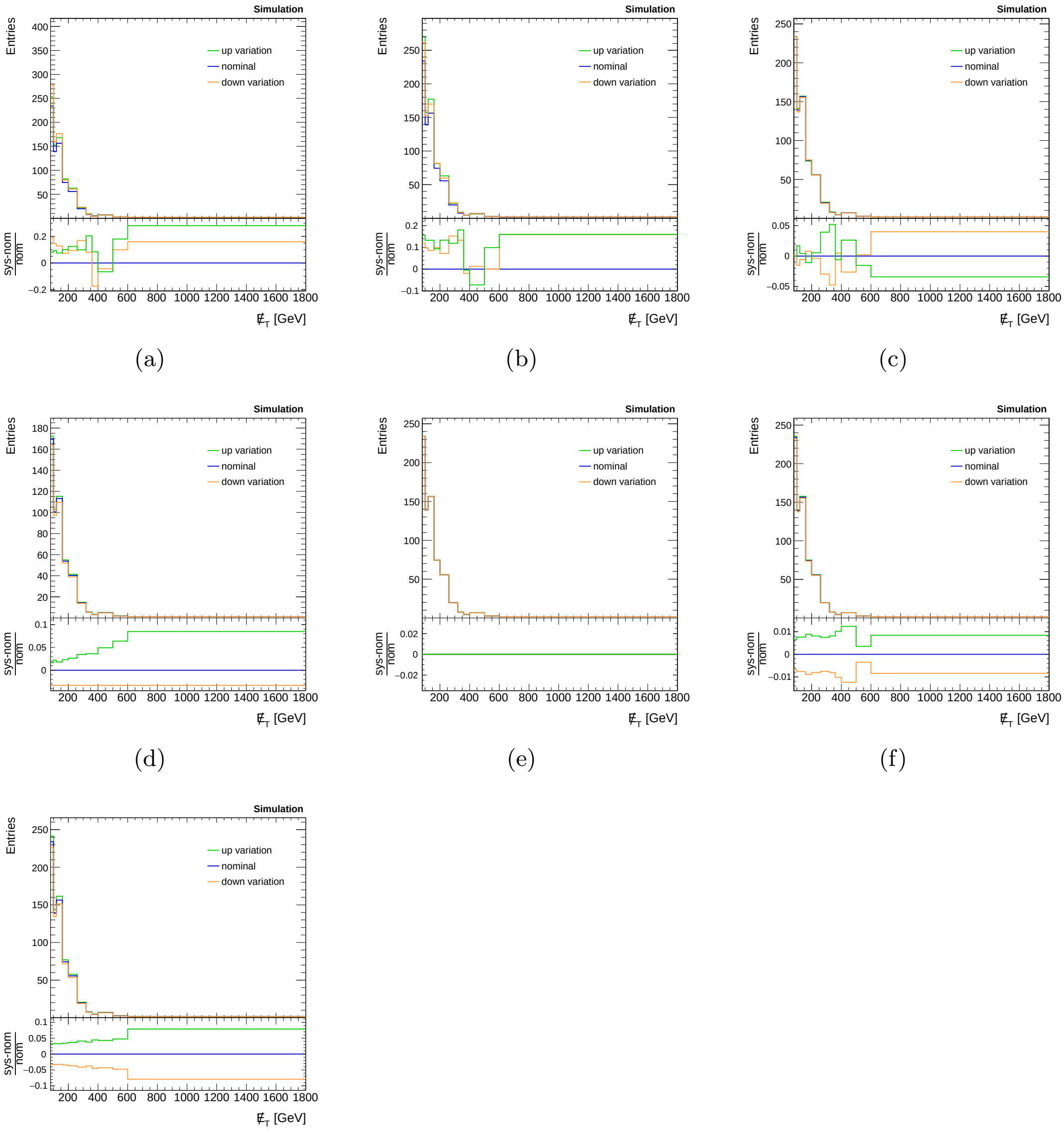
Figure C.3: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not\!\!E_T$ spectra in the isolated lepton signal region obtained for the signal model assuming a scalar mediator with $M_\Phi = 10\,\mathrm{GeV}$ and $M_\chi = 1\,\mathrm{GeV}$. Shown are the systematic uncertainties in the (a) jet energy corrections, (b) jet energy resolution, (c) pileup re-weighting , (d) pdf, (e) top tag correction factor, (f) b tagging (udsg), and (g) b tagging (bc). The production cross section of the studied signal process is scaled to 1 pb.



(a) (b) (c)

(d) (e) (f)

Figure C.4: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not{E}_T$ spectra in the isolated lepton signal region obtained for the signal model assuming a scalar mediator with $M_\Phi = 10\,\text{GeV}$ and $M_\chi = 1\,\text{GeV}$. Shown are the systematic uncertainties in the (a) m_T efficiency correction, (b) muon identification, (c) muon isolation, (d) muon trigger, (e) muon tracking, and (f) renormalization and factorization scale. The production cross section of the studied signal process is scaled to 1 pb.

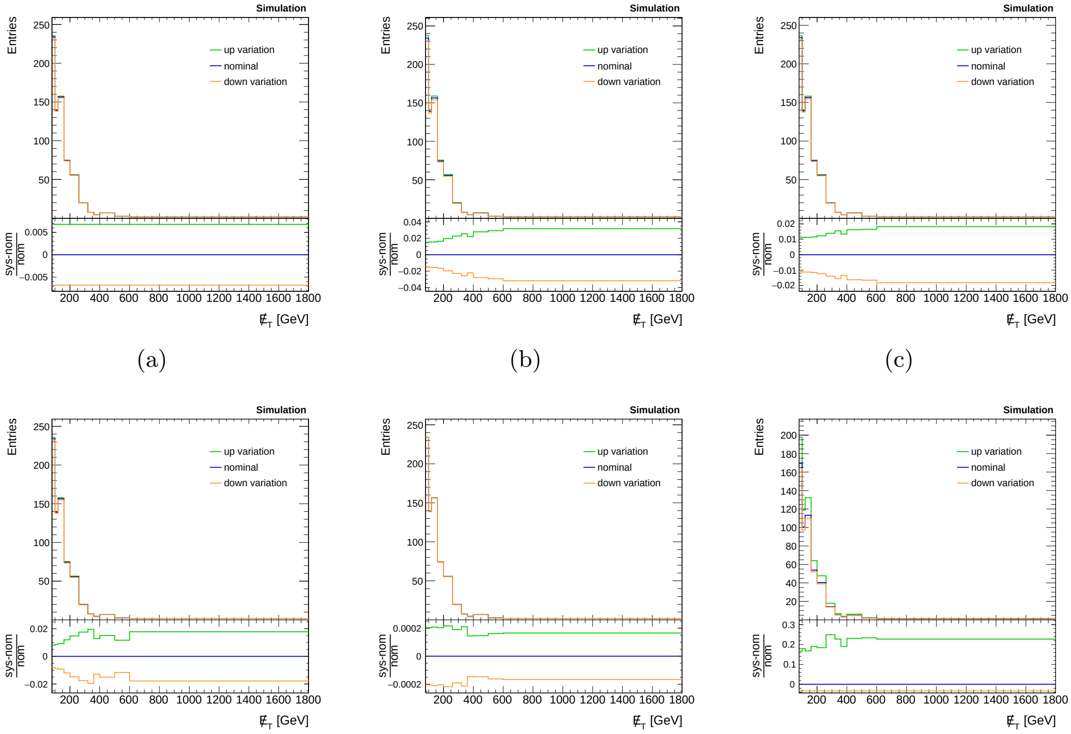


(a) (b) (c)

(d) (e) (f)

(g)

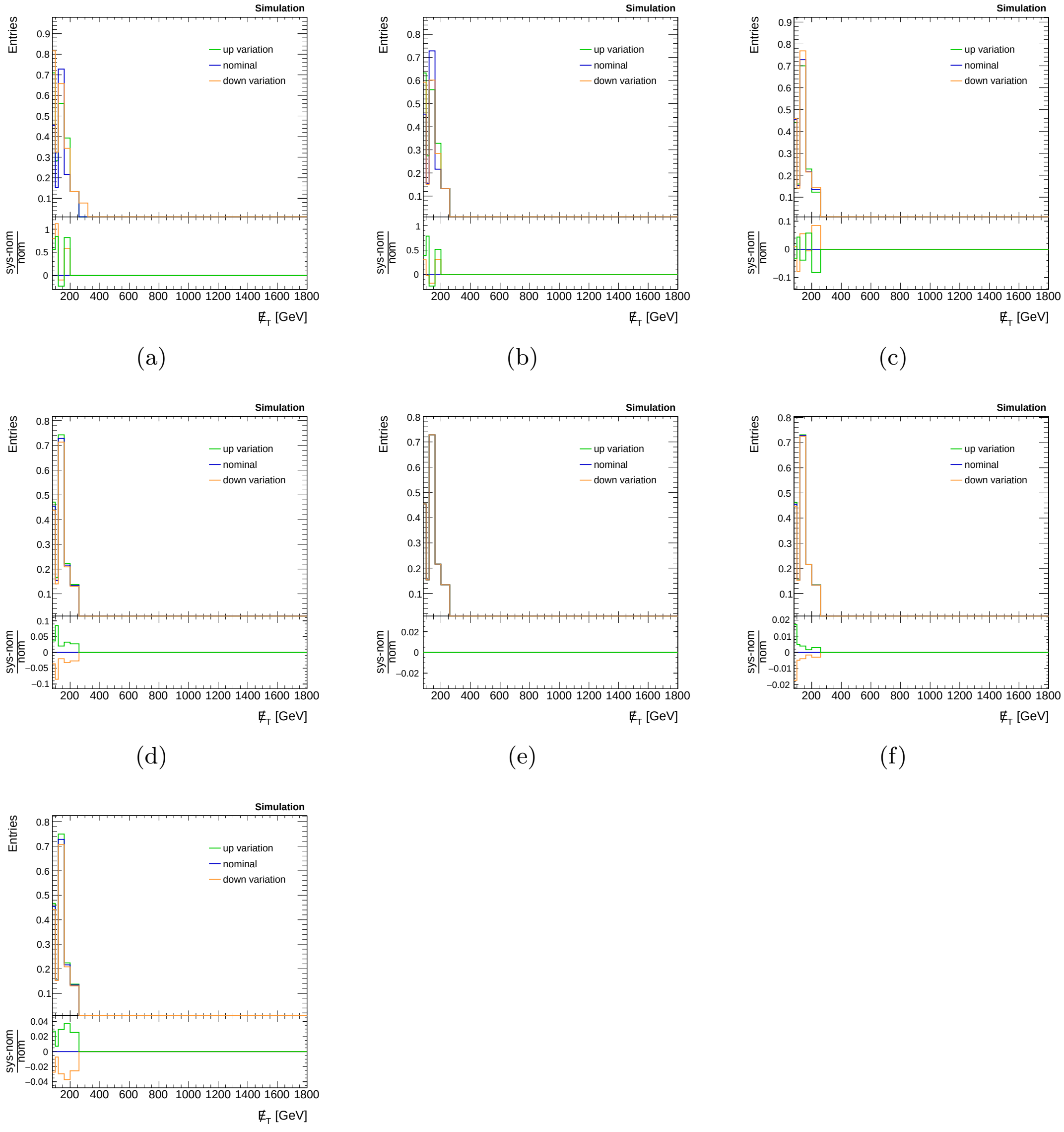
Figure C.5: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not{E}_T$ spectra in the non-isolated lepton signal region obtained for the $t\bar{t}+$jets background processes. Shown are the systematic uncertainties in the (a) jet energy corrections, (b) jet energy resolution, (c) pileup re-weighting , (d) pdf, (e) top tag correction factor, (f) b tagging (udsg), and (g) b tagging (bc).



(a) (b) (c)

(d) (e) (f)

Figure C.6: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not{E}_T$ spectra in the non-isolated lepton signal region obtained for the $t\bar{t}$ + jets background processes. Shown are the systematic uncertainties in the (a) m_T efficiency correction, (b) muon identification, (c) muon isolation, (d) muon trigger, (e) muon tracking, and (f) renormalization and factorization scale.

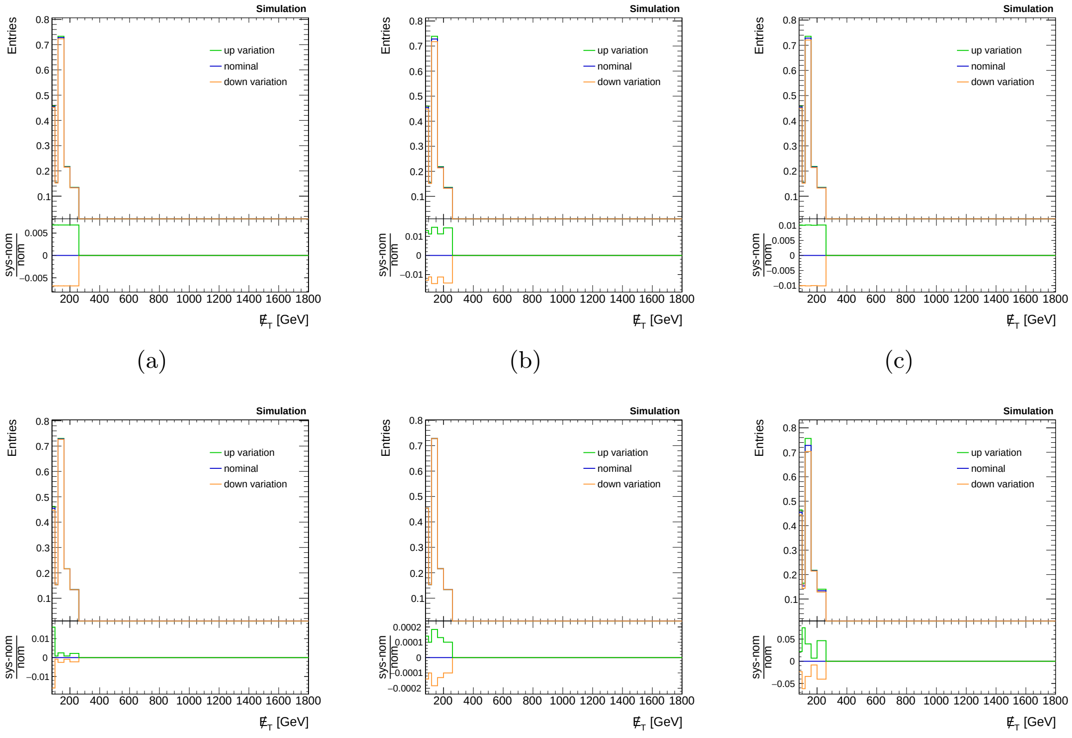


(a) (b) (c)

(d) (e) (f)

(g)

Figure C.7: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not{E}_T$ spectra in the non-isolated lepton signal region obtained for the signal model assuming a scalar mediator with $M_\Phi = 10\,\text{GeV}$ and $M_\chi = 1\,\text{GeV}$. Shown are the systematic uncertainties in the (a) jet energy corrections, (b) jet energy resolution, (c) pileup re-weighting , (d) pdf, (e) top tag correction factor, (f) b tagging (udsg), and (g) b tagging (bc). The production cross section of the studied signal process is scaled to 1 pb.



(a) (b) (c)

(d) (e) (f)

Figure C.8: Impact of all considered systematic uncertainties (see Section 8.13) on the $\not{E}_T$ spectra in the non-isolated lepton signal region obtained for the signal model assuming a scalar mediator with $M_\Phi = 10\,\text{GeV}$ and $M_\chi = 1\,\text{GeV}$. Shown are the systematic uncertainties in the (a) m_T efficiency correction, (b) muon identification, (c) muon isolation, (d) muon trigger, (e) muon tracking, and (f) renormalization and factorization scale. The production cross section of the studied signal process is scaled to 1 pb.

Bibliography

[1] CMS Collaboration, "Search for Third Generation Scalar Leptoquarks Decaying to Top Quark - Tau Lepton Pairs in pp Collisions", CMS Physics Analysis Summary CMS-PAS-EXO-13-010, 2014.

[2] CMS Collaboration, "Search for Third-Generation Scalar Leptoquarks in the tτ Channel in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JHEP* **07** (2015) 042, `arXiv:1503.09049`. [Erratum: JHEP11,056(2016)]. `doi:10.1007/JHEP11(2016)056, 10.1007/JHEP07(2015)042`.

[3] CMS Collaboration, "Search for dark matter in association with a top quark pair at $\sqrt{s} = 13\,\text{TeV}$", CMS Physics Analysis Summary CMS-PAS-EXO-16-005, 2016.

[4] CMS Collaboration, "Search for dark matter produced in association with heavy-flavor quarks in proton-proton collisions at sqrt(s)=13 TeV", `arXiv:1706.02581`.

[5] ATLAS Collaboration, "Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector", ATLAS Note ATLAS-CONF-2016-050, 2016.

[6] CMS Collaboration, "Jet algorithms performance in 13 TeV data", CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017.

[7] K. Olive and others (Particle Data Group), "Review of Particle Physics", *Chin. Phys.* **C38** (2014 and 2015 update) 090001. `doi:10.1088/1674-1137/38/9/090001`.

[8] A. Purcell, "Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules", BUL-NA-2012-269. 35/2012, Aug, 2012.

[9] LHCb Collaboration, "Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays", *Phys. Rev. Lett.* **118** (2017) 022003, `arXiv:1606.07895`. `doi:10.1103/PhysRevLett.118.022003`.

[10] LHCb Collaboration, "Amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays", *Phys. Rev.* **D95** (2017) 012002, `arXiv:1606.07898`. `doi:10.1103/PhysRevD.95.012002`.

[11] LHCb Collaboration, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.* **115** (2015) 072001, `arXiv:1507.03414`. `doi:10.1103/PhysRevLett.115.072001`.

[12] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579. `doi:10.1016/0029-5582(61)90469-2`.

[13] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264. `doi:10.1103/PhysRevLett.19.1264`.

[14] A. D. Sakharov, “Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe”, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32. [Usp. Fiz. Nauk161,61(1991)]. `doi:10.1070/PU1991v034n05ABEH002497`.

[15] Planck Collaboration, “Planck 2013 results. XVI. Cosmological parameters”, *Astron. Astrophys.* **571** (2014) A16, `arXiv:1303.5076`. `doi:10.1051/0004-6361/201321591`.

[16] G. Bertone, D. Hooper, and J. Silk, “Particle dark matter: Evidence, candidates and constraints”, *Phys. Rept.* **405** (2005) 279, `arXiv:hep-ph/0404175`. `doi:10.1016/j.physrep.2004.08.031`.

[17] K. G. Begeman, A. H. Broeils, and R. H. Sanders, “Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics”, *Mon. Not. Roy. Astron. Soc.* **249** (1991) 523.

[18] G. B. Gelmini, “TASI 2014 Lectures: The Hunt for Dark Matter”, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014) Boulder, Colorado, June 2-27, 2014*. 2015. `arXiv:1502.01320`.

[19] G. Kane, “Modern Elementary Particle Physics”. February, 2017.

[20] J. C. Pati and A. Salam, “Lepton Number as the Fourth Color”, *Phys. Rev.* **D10** (1974) 275. [Erratum: Phys. Rev.D11,703(1975)]. `doi:10.1103/PhysRevD.10.275, 10.1103/PhysRevD.11.703.2`.

[21] H. Georgi and S. L. Glashow, “Unity of All Elementary Particle Forces”, *Phys. Rev. Lett.* **32** (1974) 438. `doi:10.1103/PhysRevLett.32.438`.

[22] B. Schrempp and F. Schrempp, “Light Leptoquarks”, *Phys. Lett.* **B153** (1985) 101. `doi:10.1016/0370-2693(85)91450-9`.

[23] H. Harari, “A Schematic Model of Quarks and Leptons”, *Phys. Lett.* **B86** (1979) 83. `doi:10.1016/0370-2693(79)90626-9`.

[24] I. J. R. Aitchison, “Supersymmetry and the MSSM: An Elementary introduction”, `arXiv:hep-ph/0505105`.

[25] S. P. Martin, “A Supersymmetry primer”, `arXiv:hep-ph/9709356`. [Adv. Ser. Direct. High Energy Phys.18,1(1998)]. `doi:10.1142/9789812839657_0001, 10.1142/9789814307505_0001`.

[26] T. Ibrahim and P. Nath, "CP Violation From Standard Model to Strings", *Rev. Mod. Phys.* **80** (2008) 577, `arXiv:0705.2008`. `doi:10.1103/RevModPhys.80.577`.

[27] W. Buchmüller, R. Rückl, and D. Wyler, "Leptoquarks in Lepton - Quark Collisions", *Phys. Lett.* **B191** (1987) 442. [Erratum: Phys. Lett.B448,320(1999)]. `doi:10.1016/S0370-2693(99)00014-3, 10.1016/0370-2693(87)90637-X`.

[28] I. Doršner, S. Fajfer, A. Greljo et al., "Physics of leptoquarks in precision experiments and at particle colliders", *Phys. Rept.* **641** (2016) 1, `arXiv:1603.04993`. `doi:10.1016/j.physrep.2016.06.001`.

[29] S. Davidson, D. C. Bailey, and B. A. Campbell, "Model independent constraints on leptoquarks from rare processes", *Z. Phys.* **C61** (1994) 613, `arXiv:hep-ph/9309310`. `doi:10.1007/BF01552629`.

[30] M. Leurer, "A Comprehensive study of leptoquark bounds", *Phys. Rev.* **D49** (1994) 333, `arXiv:hep-ph/9309266`. `doi:10.1103/PhysRevD.49.333`.

[31] B. Gripaios, M. Nardecchia, and S. A. Renner, "Composite leptoquarks and anomalies in B-meson decays", *JHEP* **05** (2015) 006, `arXiv:1412.1791`. `doi:10.1007/JHEP05(2015)006`.

[32] S. Chakdar, T. Li, S. Nandi et al., "Top SU(5) Model: Baryon and Lepton Number Violations at the TeV Scale", arXiv:1206.0409. OSU-HEP-12-05. MIFPA-12-19, Jun, 2012.

[33] W. Kilian and J. Reuter, "Unification without doublet-triplet splitting", *Phys. Lett.* **B642** (2006) 81, `arXiv:hep-ph/0606277`. `doi:10.1016/j.physletb.2006.09.033`.

[34] B. Gripaios, A. Papaefstathiou, K. Sakurai et al., "Searching for third-generation composite leptoquarks at the LHC", *JHEP* **01** (2011) 156, `arXiv:1010.3962`. `doi:10.1007/JHEP01(2011)156`.

[35] M. Bauer and M. Neubert, "Minimal Leptoquark Explanation for the $\mathrm{R}_{D^{(*)}}$, R_K, and $(g-2)_g$ Anomalies", *Phys. Rev. Lett.* **116** (2016), no. 14, 141802, `arXiv:1511.01900`. `doi:10.1103/PhysRevLett.116.141802`.

[36] R. Barbier et al., "R-parity violating supersymmetry", *Phys. Rept.* **420** (2005) 1, `arXiv:hep-ph/0406039`. `doi:10.1016/j.physrep.2005.08.006`.

[37] M. Krämer, T. Plehn, M. Spira et al., "Pair production of scalar leptoquarks at the CERN LHC", *Phys. Rev.* **D71** (2005) 057503, `arXiv:hep-ph/0411038`. `doi:10.1103/PhysRevD.71.057503`.

[38] CMS Collaboration, "Search for pair-production of first generation scalar leptoquarks in pp collisions at $\sqrt{s} = 13$ TeV with 2.6 fb^{-1}", CMS Physics Analysis Summary CMS-PAS-EXO-16-043, 2016.

[39] ATLAS Collaboration, “Search for scalar leptoquarks in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment”, *New J. Phys.* **18** (2016) 093016, `arXiv:1605.06035`. `doi:10.1088/1367-2630/18/9/093016`.

[40] CMS Collaboration, “Search for single production of scalar leptoquarks in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev.* **D93** (2016) 032005, `arXiv:1509.03750`. `doi:10.1103/PhysRevD.93.032005`.

[41] CMS Collaboration, “Search for pair-production of second-generation scalar leptoquarks in pp collisions at $\sqrt{s} = 13$ TeV with the CMS detector”, CMS Physics Analysis Summary CMS-PAS-EXO-16-007, 2016.

[42] CMS Collaboration, “Search for the third-generation scalar leptoquarks and heavy right-handed neutrinos in $\tau_\ell\tau_h$jj final states in pp collisions at 13 TeV”, CMS Physics Analysis Summary CMS-PAS-EXO-16-023, 2016.

[43] ATLAS Collaboration, “Search for third generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *JHEP* **06** (2013) 033, `arXiv:1303.0526`. `doi:10.1007/JHEP06(2013)033`.

[44] CMS Collaboration, “Search for third-generation leptoquarks and scalar bottom quarks in *pp* collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **12** (2012) 055, `arXiv:1210.5627`. `doi:10.1007/JHEP12(2012)055`.

[45] ATLAS Collaboration, “Searches for scalar leptoquarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Eur. Phys. J.* **C76** (2016) 5, `arXiv:1508.04735`. `doi:10.1140/epjc/s10052-015-3823-9`.

[46] CMS Collaboration, “Search for pair production of third-generation scalar leptoquarks and top squarks in proton–proton collisions at $\sqrt{s}$=8 TeV”, *Phys. Lett.* **B739** (2014) 229, `arXiv:1408.0806`. `doi:10.1016/j.physletb.2014.10.063`.

[47] S. Rolli and M. Tanabashi, “Leptoquarks”, Chin. Phys. C 40, 100001 (2016), 2008.

[48] Snowmass 2013 Cosmic Frontier Working Groups 14 Collaboration, “Dark Matter in the Coming Decade: Complementary Paths to Discovery and Beyond”, *Phys. Dark Univ.* **7-8** (2015) 16, `arXiv:1305.1605`. `doi:10.1016/j.dark.2015.04.001`.

[49] G. Busoni et al., “Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter”, `arXiv:1603.04156`.

[50] CMS Collaboration, “Dark Matter Summary Plots from CMS for ICHEP 2016”,CMS-DP-2016-057, Aug, 2016.

[51] F. Kahlhoefer, “Review of LHC Dark Matter Searches”, *Int. J. Mod. Phys.* **A32** (2017) 1730006, `arXiv:1702.02430`. `doi:10.1142/S0217751X1730006X`.

[52] CMS Collaboration, “Search for the production of dark matter in association with top-quark pairs in the single-lepton final state in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **06** (2015) 121, `arXiv:1504.03198`. `doi:10.1007/JHEP06(2015)121`.

[53] D. Abercrombie et al., “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”, `arXiv:1507.00966`.

[54] G. Busoni, A. De Simone, T. Jacques et al., “Making the Most of the Relic Density for Dark Matter Searches at the LHC 14 TeV Run”, *JCAP* **1503** (2015) 022, `arXiv:1410.7409`. `doi:10.1088/1475-7516/2015/03/022`.

[55] M. Bauer et al., “Towards the next generation of simplified Dark Matter models”, `arXiv:1607.06680`.

[56] P. Harris, V. V. Khoze, M. Spannowsky et al., “Constraining Dark Sectors at Colliders: Beyond the Effective Theory Approach”, *Phys. Rev.* **D91** (2015) 055009, `arXiv:1411.0535`. `doi:10.1103/PhysRevD.91.055009`.

[57] M. Bauer, U. Haisch, and F. Kahlhoefer, “Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators”, *JHEP* **05** (2017) 138, `arXiv:1701.07427`. `doi:10.1007/JHEP05(2017)138`.

[58] T. Lin, E. W. Kolb, and L.-T. Wang, “Probing dark matter couplings to top and bottom quarks at the LHC”, *Phys. Rev.* **D88** (2013) 063510, `arXiv:1303.6638`. `doi:10.1103/PhysRevD.88.063510`.

[59] M. R. Buckley, D. Feld, and D. Goncalves, “Scalar Simplified Models for Dark Matter”, *Phys. Rev.* **D91** (2015) 015017, `arXiv:1410.6497`. `doi:10.1103/PhysRevD.91.015017`.

[60] C. Arina et al., “A comprehensive approach to dark matter studies: exploration of simplified top-philic models”, *JHEP* **11** (2016) 111, `arXiv:1605.09242`. `doi:10.1007/JHEP11(2016)111`.

[61] G. D’Ambrosio, G. F. Giudice, G. Isidori et al., “Minimal flavor violation: An Effective field theory approach”, *Nucl. Phys.* **B645** (2002) 155, `arXiv:hep-ph/0207036`. `doi:10.1016/S0550-3213(02)00836-2`.

[62] G. Isidori and D. M. Straub, “Minimal Flavour Violation and Beyond”, *Eur. Phys. J.* **C72** (2012) 2103, `arXiv:1202.0464`. `doi:10.1140/epjc/s10052-012-2103-1`.

[63] CMS Collaboration, “Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV”, `arXiv:1703.01651`.

[64] ATLAS Collaboration, “Search for direct top squark pair production and dark matter production in final states with two leptons in $\sqrt{s} = 13$ TeV pp collisions using 13.3 fb^{-1} of ATLAS data”, ATLAS Note ATLAS-CONF-2016-076, 2016.

[65] ATLAS Collaboration, "Search for the Supersymmetric Partner of the Top Quark in the Jets+E_T^{miss} Final State at $\sqrt{s} = 13$ TeV", ATLAS Note ATLAS-CONF-2016-077, 2016.

[66] M. Benedikt, P. Collier, V. Mertens et al., "LHC Design Report. 3. The LHC injector chain", CERN-2004-003-V-3, CERN-2004-003, 2004.

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

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

[69] O. S. Bruning, P. Collier, P. Lebrun et al., "LHC Design Report Vol.1: The LHC Main Ring", CERN-2004-003-V1, CERN-2004-003, CERN-2004-003-V-1, 2004.

[70] "LEP Design Report: Vol.2. The LEP Main Ring", CERN-LEP-84-01, 1984.

[71] C. De Melis, "The CERN accelerator complex. Complexe des accélérateurs du CERN", OPEN-PHO-ACCEL-2016-001, Jan, 2016. General Photo.

[72] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. `doi:10.1088/1748-0221/3/08/S08003`.

[73] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005. `doi:10.1088/1748-0221/3/08/S08005`.

[74] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002. `doi:10.1088/1748-0221/3/08/S08002`.

[75] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, `arXiv:1607.03663`. `doi:10.1088/1748-0221/12/02/P02014`.

[76] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, `arXiv:1405.6569`. `doi:10.1088/1748-0221/9/10/P10009`.

[77] CMS Collaboration, G. L. Bayatian et al., "CMS physics: Technical design report", CERN-LHCC-2006-001, CMS-TDR-008-1, 2006.

[78] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS-PAS-PFT-09-001, 2009.

[79] USCMS, ECAL/HCAL Collaboration, "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c", *Eur. Phys. J.* **C60** (2009) 359. [Erratum: Eur. Phys. J.C61,353(2009)]. `doi:10.1140/epjc/s10052-009-0959-5, 10.1140/epjc/s10052-009-1024-0`.

[80] Elias Mal, "LHC: Status, Prospects and Future Challenges". Conference talk at LHCP 2016. Accessed: 04/10/2017.

[81] CMS Collaboration, “CMS Luminosity - Public Results”. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 01/20/2017.

[82] “Long Shutdown 1: An opportunity for consolidation”. http://cms.web.cern.ch/news/long-shutdown-1-opportunity-consolidation. Accessed: 04/10/2017.

[83] Paolo Spagnolo on behalf of the CMS Collaboration, “CMS: First look at 13 TeV data”. Conference talk at LHCP 2015. Accessed: 04/10/2017.

[84] CMS Collaboration, “Technical proposal for the upgrade of the CMS detector through 2020”, CERN-LHCC-2011-006. LHCC-P-004, Jun, 2011.

[85] Jeremiah Mans on behalf of the CMS Collaboration, “CMS Phase 2 Upgrade: Scope and R & D goals”. talk at ECFA High Luminosity LHC Experiments Workshop - 2014. Accessed: 04/11/2017.

[86] A. Höcker, “Physics at the LHC Run-2 and Beyond”, in *2016 European School of High-Energy Physics (ESHEP 2016) Skeikampen, Norway, June 15-28, 2016*. 2016. `arXiv:1611.07864`.

[87] M. H. Seymour and M. Marx, “Monte Carlo Event Generators”, in *Proceedings, 69th Scottish Universities Summer School in Physics : LHC Phenomenology (SUSSP69): St.Andrews, Scotland, August 19-September 1, 2012*, p. 287. 2013. `arXiv:1304.6677`.

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

[89] S. Weinzierl, “Introduction to Monte Carlo methods”, `arXiv:hep-ph/0006269`.

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

[91] J. Pumplin, D. R. Stump, J. Huston et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **07** (2002) 012, `arXiv:hep-ph/0201195`. `doi:10.1088/1126-6708/2002/07/012`.

[92] H.-L. Lai, M. Guzzi, J. Huston et al., “New parton distributions for collider physics”, *Phys. Rev. D* **82** (2010) 074024, `arXiv:1007.2241`. `doi:10.1103/PhysRevD.82.074024`.

[93] J. Alwall, M. Herquet, F. Maltoni et al., “MadGraph 5: going beyond”, *JHEP* **06** (2011) 128, `arXiv:1106.0522`. `doi:10.1007/JHEP06(2011)128`.

[94] J. Alwall, R. Frederix, S. Frixione et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching

to parton shower simulations", *JHEP* **07** (2014) 079, `arXiv:1405.0301`. `doi:10.1007/JHEP07(2014)079`.

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

[96] 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, `arXiv:0907.4076`. [Erratum: JHEP02,011(2010)]. `doi:10.1007/JHEP02(2010)011, 10.1088/1126-6708/2009/09/111`.

[97] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J.* **C71** (2011) 1547, `arXiv:1009.2450`. `doi:10.1140/epjc/s10052-011-1547-z`.

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

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

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

[101] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, `arXiv:hep-ph/0603175`. `doi:10.1088/1126-6708/2006/05/026`.

[102] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, `arXiv:0710.3820`. `doi:10.1016/j.cpc.2008.01.036`.

[103] P. Artoisenet, R. Frederix, O. Mattelaer et al., "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* **03** (2013) 015, `arXiv:1212.3460`. `doi:10.1007/JHEP03(2013)015`.

[104] S. Frixione, E. Laenen, P. Motylinski et al., "Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations", *JHEP* **04** (2007) 081, `arXiv:hep-ph/0702198`. `doi:10.1088/1126-6708/2007/04/081`.

[105] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438. [Yad. Fiz.15,781(1972)].

[106] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum

Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[107] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys.* **B126** (1977) 298. `doi:10.1016/0550-3213(77)90384-4`.

[108] 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, `arXiv:0706.2569`. `doi:10.1140/epjc/s10052-007-0490-5`.

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

[110] Z. Wąs, "TAUOLA the library for tau lepton decay, and KKMC/KORALB/KORALZ/... status report", *Nucl. Phys. B, Proc. Suppl.* **98** (2001) 96. `doi:10.1016/S0920-5632(01)01200-2`.

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

[112] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth.* **A262** (1987) 444. `doi:10.1016/0168-9002(87)90887-4`.

[113] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P08010, `arXiv:1502.02702`. `doi:10.1088/1748-0221/10/08/P08010`.

[114] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, `arXiv:1211.4462`. `doi:10.1088/1748-0221/8/04/P04013`.

[115] K. Rose, "Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems", in *Proceedings of the IEEE*, p. 2210. 1998.

[116] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys.* **G34** (2007) N343. `doi:10.1088/0954-3899/34/12/N01`.

[117] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, `arXiv:1206.4071`. `doi:10.1088/1748-0221/7/10/P10002`.

[118] CMS Collaboration, "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *JINST* **5** (2010) T03022, `arXiv:0911.4994`. `doi:10.1088/1748-0221/5/03/T03022`.

[119] CMS Collaboration, "Baseline muon selections for Run-I". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId. Accessed: 01/05/2017.

[120] CMS Collaboration, "Baseline muon selections for Run-II". https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2. Accessed: 01/05/2017.

[121] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC", *JINST* **8** (2013) P11002, `arXiv:1306.6905`. `doi:10.1088/1748-0221/8/11/P11002`.

[122] CMS Collaboration, "Search for neutral Higgs bosons decaying to tau pairs in *pp* collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B713** (2012) 68, `arXiv:1202.4083`. `doi:10.1016/j.physletb.2012.05.028`.

[123] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, `arXiv:1502.02701`. `doi:10.1088/1748-0221/10/06/P06005`.

[124] W. Adam, R. Frühwirth, A. Strandlie et al., "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC", *eConf* **C0303241** (2003) TULT009, `arXiv:physics/0306087`. [J. Phys.G31,N9(2005)]. `doi:10.1088/0954-3899/31/9/N01`.

[125] CMS Collaboration, "CMS Top EGamma Coordination (Run1)". CMS Top EGamma Coordination (Run1). Accessed: 01/11/2017.

[126] CMS Collaboration, "Cut Based Electron ID for Run 2". https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2. Accessed: 01/12/2017.

[127] CMS Collaboration, "Electron Identification". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideElectronID. Accessed: 01/12/2017.

[128] A. Kell, "Suche nach skalaren Leptoquarks der dritten Generation im Zerfallskanal $LQ_3 \rightarrow t\tau$ am LHC", June, 2014. master thesis at the University of Hamburg.

[129] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-12-027, 2013.

[130] G. P. Salam, "Towards Jetography", *Eur. Phys. J.* **C67** (2010) 637, `arXiv:0906.1833`. `doi:10.1140/epjc/s10052-010-1314-6`.

[131] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone jet algorithm", *JHEP* **05** (2007) 086, `arXiv:0704.0292`. `doi:10.1088/1126-6708/2007/05/086`.

[132] G. Marchesini, B. R. Webber, G. Abbiendi et al., "HERWIG: A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons.

Version 5.1 - April 1991", *Comput. Phys. Commun.* **67** (1992) 465–508. `doi:10.1016/0010-4655(92)90055-4`.

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

[134] CMS Collaboration, "Study of Pileup Removal Algorithms for Jets", CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014.

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

[136] CMS Collaboration, "Jet Identification". https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID. Accessed: 01/06/2017.

[137] CMS Collaboration, "Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, ttbar and boosted topology events", CMS Physics Analysis Summary CMS-PAS-BTV-13-001, 2013.

[138] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.

[139] W. Waltenberger, "Adaptive vertex reconstruction", CERN-CMS-NOTE-2008-033, 2008.

[140] CMS Collaboration, "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV", *JHEP* **03** (2011) 136, `arXiv:1102.3194`. `doi:10.1007/JHEP03(2011)136`.

[141] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS", *JINST* **11** (2016) P01019, `arXiv:1510.07488`. `doi:10.1088/1748-0221/11/01/P01019`.

[142] CMS Collaboration, "TauID: recommendation from the Tau POG". https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendation. Accessed: 01/12/2017.

[143] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P02006, `arXiv:1411.0511`. `doi:10.1088/1748-0221/10/02/P02006`.

[144] M. Meyer, "Search for third-generation scalar leptoquarks decaying into a top quark and a tau lepton with the CMS detector", 2013. master thesis at the University of Hamburg.

[145] CMS Collaboration, "Search for Third Generation Scalar Leptoquarks Decaying to Top Quark - Tau Lepton Pairs in pp Collisions", CMS Physics Analysis Summary CMS-PAS-EXO-12-030, 2014.

[146] H. Saka and C. Tully, “Search for Third Generation Scalar Leptoquarks in the t-tau Channel with the CMS Detector”. PhD thesis, Princeton U., Apr, 2015. Presented 13 May 2015.

[147] CMS Collaboration, “The official CMS Luminosity Calculation Tools”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/LumiCalc. Accessed: 06/07/2017.

[148] CMS Collaboration, “2012 Analysis: purpose of this page”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV2012Analysis. Accessed: 06/07/2017.

[149] CMS Collaboration, “Reference muon id and isolation efficiencies”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs. Accessed: 06/07/2017.

[150] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in FEWZ”, *Phys. Rev. D* **86** (2012) 094034, `arXiv:1208.5967`. `doi:10.1103/PhysRevD.86.094034`.

[151] M. Czakon, P. Fiedler, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$”, *Phys. Rev. Lett.* **110** (2013) 252004, `arXiv:1303.6254`. `doi:10.1103/PhysRevLett.110.252004`.

[152] N. Kidonakis and R. Vogt, “Theoretical top quark cross section at the Tevatron and the LHC”, *Phys. Rev. D* **78** (2008) 074005, `arXiv:0805.3844`. `doi:10.1103/PhysRevD.78.074005`.

[153] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC”, *JHEP* **07** (2011) 018, `arXiv:1105.0020`. `doi:10.1007/JHEP07(2011)018`.

[154] M. V. Garzelli, A. Kardos, C. G. Papadopoulos et al., “$\mathrm{t\bar{t}W^{\pm}}$ and $\mathrm{t\bar{t}Z}$ hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects”, *JHEP* **11** (2012) 056, `arXiv:1208.2665`. `doi:10.1007/JHEP11(2012)056`.

[155] CMS Collaboration, “MET Optional Filters”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFilters. Accessed: 01/22/2017.

[156] CMS Collaboration, “Filters against Tracker and Tracking Odd Events”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/TrackingPOGFilters. Accessed: 06/07/2017.

[157] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013.

[158] CMS Collaboration, "Pileup Reweighting Utilities". https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities. Accessed: 06/07/2017.

[159] CMS Collaboration, "Estimating Systematic Errors Due to Pileup Modeling". https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors. Accessed: 06/07/2017.

[160] CMS Collaboration, "Tag and Probe". https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe. Accessed: 01/24/2017.

[161] CMS Collaboration, "Differential cross section measurements for the production of a W boson in association with jets in proton–proton collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B741** (2015) 12, `arXiv:1406.7533`. `doi:10.1016/j.physletb.2014.12.003`.

[162] CMS Collaboration, "Measurements of differential cross sections for associated production of a W boson and jets in proton-proton collisions at $\sqrt{s}$=8 TeV", `arXiv:1610.04222`.

[163] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in *pp* colisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J.* **C73** (2013) 2339, `arXiv:1211.2220`. `doi:10.1140/epjc/s10052-013-2339-4`.

[164] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur. Phys. J.* **C75** (2015) 542, `arXiv:1505.04480`. `doi:10.1140/epjc/s10052-015-3709-x`.

[165] CMS Collaboration, "pt(top-quark) based reweighting of ttbar MC". https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting. Accessed: 06/07/2017.

[166] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **02** (2014) 024, `arXiv:1312.7582`. [Erratum: JHEP02,102(2014)]. `doi:10.1007/JHEP02(2014)024, 10.1007/JHEP02(2014)102`.

[167] W. Ochs and R. Perez Ramos, "Particle Multiplicity in Jets and Sub-jets with Jet Axis from Color Current", *Phys. Rev.* **D78** (2008) 034010, `arXiv:0807.1082`. `doi:10.1103/PhysRevD.78.034010`.

[168] J. Ott. http://www.theta-framework.org/.

[169] J. Ott, T. Müller, and J. Wagner-Kuhr, "Search for Resonant Top Quark Pair Production in the Muon+Jets Channel with the CMS Detector". PhD thesis, Karlsruhe U., Dec, 2013. Presented 14 Dec 2012.

[170] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B722** (2013) 5, `arXiv:1210.6718`. `doi:10.1016/j.physletb.2013.03.024`.

[171] CMS Collaboration, "Searches for third-generation squark production in fully hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JHEP* **06** (2015) 116, `arXiv:1503.08037`. `doi:10.1007/JHEP06(2015)116`.

[172] CMS Collaboration, "Top Tagging with New Approaches", CMS Physics Analysis Summary CMS-PAS-JME-15-002, 2016.

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

[174] J. M. Butterworth, A. R. Davison, M. Rubin et al., "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001, `arXiv:0802.2470`. `doi:10.1103/PhysRevLett.100.242001`.

[175] T. Plehn, G. P. Salam, and M. Spannowsky, "Fat Jets for a Light Higgs", *Phys. Rev. Lett.* **104** (2010) 111801, `arXiv:0910.5472`. `doi:10.1103/PhysRevLett.104.111801`.

[176] T. Plehn, M. Spannowsky, M. Takeuchi et al., "Stop Reconstruction with Tagged Tops", *JHEP* **10** (2010) 078, `arXiv:1006.2833`. `doi:10.1007/JHEP10(2010)078`.

[177] G. Kasieczka, T. Plehn, T. Schell et al., "Resonance Searches with an Updated Top Tagger", *JHEP* **06** (2015) 203, `arXiv:1503.05921`. `doi:10.1007/JHEP06(2015)203`.

[178] J. M. Butterworth, A. R. Davison, M. Rubin et al., "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001, `arXiv:0802.2470`. `doi:10.1103/PhysRevLett.100.242001`.

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

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

[181] CMS Collaboration, "Boosted Top Jet Tagging". https://twiki.cern.ch/twiki/bin/view/CMS/JetTopTagging. Accessed: 02/25/2017.

[182] A. Höcker et al., "TMVA - Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, `arXiv:physics/0703039`.

[183] CMS Collaboration, "Search for $\mathrm{t\bar{t}}$ resonances in boosted semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-B2G-15-002, 2016.

[184] CMS Collaboration, "Search for resonant $t\bar{t}$ production in lepton+jets events in *pp* collisions at $\sqrt{s} = 7$ TeV", *JHEP* **12** (2012) 015, `arXiv:1209.4397`. `doi:10.1007/JHEP12(2012)015`.

[185] CMS Collaboration, "Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s}$=8TeV", *Phys. Rev. Lett.* **111** (2013) 211804, `arXiv:1309.2030`. [Erratum: Phys. Rev. Lett.112,no.11,119903(2014)]. `doi:10.1103/PhysRevLett.111.211804, 10.1103/PhysRevLett.112.119903`.

[186] D0 Collaboration, "Direct measurement of the top quark mass at D0", *Phys. Rev.* **D58** (1998) 052001, `arXiv:hep-ex/9801025`. `doi:10.1103/PhysRevD.58.052001`.

[187] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in *pp* collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-14-001, 2014.

[188] F. James, "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1", CERN-D-506, CERN-D506, 1994.

[189] Y. Bai, H.-C. Cheng, J. Gallicchio et al., "Stop the Top Background of the Stop Search", *JHEP* **07** (2012) 110, `arXiv:1203.4813`. `doi:10.1007/JHEP07(2012)110`.

[190] CMS Collaboration, "Luminosity Physics Object Group". https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM. Accessed: 01/20/2017.

[191] CMS Collaboration, "Work in progress (evolving results) and dedicated studies from PAGs". https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults. Accessed: 01/23/2017.

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

[193] R. Gavin, Y. Li, F. Petriello et al., "FEWZ 3.1 guide", Manual FEWZManual_3.1.pdf, 2012.

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

[195] P. Kant, O. M. Kind, T. Kintscher et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark

production in hadronic collisions", *Comput. Phys. Commun.* **191** (2015) 74, `arXiv:1406.4403`. `doi:10.1016/j.cpc.2015.02.001`.

[196] N. Kidonakis, "Top Quark Production", in *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013*, p. 139. 2014. `arXiv:1311.0283`.

[197] T. Gehrmann, M. Grazzini, S. Kallweit et al., "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD", *Phys. Rev. Lett.* **113** (2014) 212001, `arXiv:1408.5243`. `doi:10.1103/PhysRevLett.113.212001`.

[198] J. M. Campbell and R. K. Ellis, "An Update on vector boson pair production at hadron colliders", *Phys. Rev.* **D60** (1999) 113006, `arXiv:hep-ph/9905386`. `doi:10.1103/PhysRevD.60.113006`.

[199] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, `arXiv:1105.0020`. `doi:10.1007/JHEP07(2011)018`.

[200] J. M. Campbell, R. K. Ellis, and W. T. Giele, "A Multi-Threaded Version of MCFM", *Eur. Phys. J.* **C75** (2015) 246, `arXiv:1503.06182`. `doi:10.1140/epjc/s10052-015-3461-2`.

[201] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV *pp* collision data using the CMS detector", CMS Physics Analysis Summary CMS-PAS-JME-16-004, 2016.

[202] CMS Collaboration, "MET Filter Recommendations for Run II". https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2. Accessed: 01/22/2017.

[203] CMS Collaboration, "Utilities for Accessing Pileup Information for Data". https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData. Accessed: 03/09/2017.

[204] CMS Collaboration, "Reference muon id, isolation and trigger efficiencies for Run-II". https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffsRun2. Accessed: 01/23/2017.

[205] CMS Collaboration, "Loss of Hit Efficiency due to HIP". https://twiki.cern.ch/CMS/SiStripHitEffLoss. Accessed: 01/23/2017.

[206] CMS Collaboration, "Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC". https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80X. Accessed: 01/24/2017.

[207] CMS Collaboration, "Methods to apply b-tagging efficiency scale factors". https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods. Accessed: 01/24/2017.

[208] CMS Collaboration, “Search for pair production of vector-like T quarks in the lepton plus jets final state”, CMS Physics Analysis Summary CMS-PAS-B2G-16-002, 2016.

[209] CMS Collaboration, “Search for vector-like quark pair production in final states with leptons and boosted Higgs bosons at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-B2G-16-011, 2016.

[210] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[211] CMS Collaboration, “Recommended Jet Energy Corrections and Uncertainties For Data and MC”. https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC. Accessed: 04/10/2017.

[212] CMS Collaboration, “Jet Energy Resolution”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution. Accessed: 04/10/2017.

[213] T. Lin, H.-B. Yu, and K. M. Zurek, “On Symmetric and Asymmetric Light Dark Matter”, *Phys. Rev.* **D85** (2012) 063503, `arXiv:1111.0293`. `doi:10.1103/PhysRevD.85.063503`.

[214] Snowmass 2013 Cosmic Frontier Working Groups 1–4 Collaboration, “Dark Matter in the Coming Decade: Complementary Paths to Discovery and Beyond”, *Phys. Dark Univ.* **7-8** (2015) 16, `arXiv:1305.1605`. `doi:10.1016/j.dark.2015.04.001`.

[215] I. Brivio, M. B. Gavela, L. Merlo et al., “ALPs Effective Field Theory and Collider Signatures”, `arXiv:1701.05379`.

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich während meiner Doktorandenzeit unterstützt haben. Besonderer Dank gilt dabei:

- **Johannes Haller** für die Betreuung meiner Arbeit, die wöchentlichen Diskussionen und die vielen Schulen und Konferenzen an denen ich teilnehmen durfte.
- **Elisabetta Gallo** für die Übernahme des Zweitgutachtens.
- der gesamten Prüfungskommission bestehend neben **Johannes Haller und Elisabetta Gallo** aus **Caren Hagner, Gudrid Moortgat-Pick und Christian Sander** für die Bereitschaft Mitglied in meiner Prüfungskommission zu sein und die Dissertation sowie die Disputation zu beurteilen.
- **Roman Kogler** für die engagierte Hilfe und die Betreuung bei der Durchführung meiner Analysen, das geduldige Beantworten meiner Fragen und das Korrekturlesen meiner gesamten Arbeit.
- **Thomas Peiffer** für die Hilfe bei der Leptoquarkanalyse und bei technischen Problemen sowie für das Lesen meiner Arbeit.
- **Teresa Lenz, Dominik Nowatschin, Heiner Tholen, Simon Kurz, Rebekka Höing und Kristin Goebel** für die lustige gemeinsame Zeit inner- und ausserhalb des Instituts. Zusätzlich möchte ich mich insbesondere bei Teresa für die vielen Diskussionen sowie für das sorgfältige Korrekturlesen meiner Arbeit und die Hilfe bei der Vorbereitung auf meine Disputation bedanken.
- **Anna Benecke, Dennis Schwarz und Svenja Schumann** für die nette gemeinsame Zeit im Büro, viele unterhaltsame Mittagspausen und Spaziergänge um den Teich.
- **der gesamten Gruppe** für die nette Arbeitsatmosphäre. Ich bin immer gerne ans Institut gekommen.
- **meiner Familie und Nils** für die Unterstützung. Danke, dass ihr immer für mich da seid!