Search for third-generation scalar leptoquarks with the CMS experiment

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

vorgelegt von

Marc Stöver

Hamburg
2017

Gutachter der Dissertation:	Prof. Dr. Johannes Haller Prof. Dr. Elisabetta Gallo
Gutachter der Disputation:	Prof. Dr. Johannes Haller Prof. Dr. Elisabetta Gallo Prof. Dr. Günter Sigl Dr. Andreas Meyer Dr. Christian Sander
Datum der Disputation:	29.01.2018
Vorsitzender des Prüfungsausschusses:	Prof. Dr. Günter Sigl
Vorsitzender des Promotionsausschusses:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs Physik:	Prof. Dr. Michael Potthoff
Dekan der Fakultät für Mathematik, Informatik und Naturwissenschaften:	Prof. Dr. Heinrich Graener

Abstract

A search for pair produced scalar leptoquarks decaying into a top quark and a tau lepton using proton-proton collision data recorded in 2016 at a center-of-mass energy of 13 TeV by the CMS experiment, corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ is presented. The search is performed in final states with an electron or a muon, one or two hadronically decaying tau leptons, and additional jets. The data are found to be consistent with the Standard Model predictions. Upper limits on the production cross section are set as a function of the leptoquark mass. Leptoquark masses below 900 GeV can be excluded at 95% confidence level, assuming a unity branching fraction of leptoquarks decaying into a top quark and a tau lepton. These results represent the most stringent limits for pair produced scalar leptoquarks in the top quark and tau lepton decay channel to date.

The presented search is based on precisely calibrated jets. In the context of this thesis, jet energy calibration measurements based on proton-proton collision data recorded in 2015 at the CMS experiment, corresponding to an integrated luminosity of $2.1\,\mathrm{fb}^{-1}$ are presented. The calibrations are extracted from data and simulated events, and utilize dijet events to determine residual $|\eta|$-dependent data-to-simulation correction factors. The correction factors differ from unity by less than 3% in the central regions and up to 17% in the endcap and hadron forward regions of the detector.

Kurzfassung

Eine Suche nach Paarproduktion von skalaren Leptoquarks, die in ein Top-Quark und ein Tau-Lepton zerfallen, wird präsentiert. Dazu wird der Datensatz, der in Proton-Proton Kollisionen im Jahr 2016 mit dem CMS Detektor aufgezeichnet wurde und einer integrierten Luminosität von $35.9\,\mathrm{fb}^{-1}$ entspricht, analysiert. Die Suche basiert auf Endzuständen mit einem Elektron oder Myon, einem oder zwei hadronisch zerfallenden Tau Leptonen und zusätzlichen Jets. Eine gute Übereinstimmung zwischen den gemessenen Daten und der Standardmodellerwartung wird beobachtet. Obere Ausschlussgrenzen auf den Wirkungsquerschnitt als Funktion der Leptoquarkmasse werden bestimmt. Leptoquarks mit Massen unter 900 GeV werden mit einem Konfidenzniveau von 95% ausgeschlossen. Dabei wird ein Verzweigungsverhältnis von 100% für den Leptoquarkzerfall in ein Top-Quark und ein Tau-Lepton angenommen. Dieses Ergebnis repräsentiert die bislang höchste Ausschlussgrenze für paarproduzierte skalare Leptoquarks, die in ein Top-Quark und ein Tau-Lepton zerfallen.

Die präsentierte Suche basiert auf präzise kalibrierten Jets. Dazu wird der Datensatz ausgewertet, der 2015 mit dem CMS Detektor aufgezeichnet wurde und einer integrierten Luminosität von $2.1\,\mathrm{fb}^{-1}$ entspricht. Die Kalibrierungen werden aus Daten und simulierten Ereignissen extrahiert. Ereignisse mit zwei Jets werden verwendet, um Korrektur-Faktoren zwischen Daten und Simulation in Abhängigkeit von $|\eta|$ des betrachteten Jets zu berechnen. Die Korrekturfaktoren weichen im inneren Detektorbereich um maximal 3% und im äußeren Detektorbereich um bis zu 17% von Eins ab.

Contents

1 Introduction

The Standard Model (SM) of particle physics is a very successful theory, which describes the fundamental forces and classifies the elementary particles. The SM was developed in the early 1970s and is formulated as a quantum field theory. It contains the set of all fundamental fermions and gauge bosons. The electromagnetic, weak, and strong forces are described by the SM. Almost all experimental results in particle physics can be predicted by the SM very precisely. However, the SM is an incomplete theory as it can not explain some phenomena nor provide answers to some open questions. Notably, a description of gravity is not included in the SM. Furthermore, dark matter and the matter-antimatter asymmetry observed in the universe can not be described. New findings from experiments may help to find answers to these open issues.

The Large Hadron Collider (LHC) located at the European Organization for Nuclear Research (CERN) in Geneva is the most powerful particle accelerator to date. Proton-proton collisions are performed at a center-of-mass energy of up to 13 TeV. The multipurpose detectors ATLAS and CMS have been constructed to measure the collision products.

Many scenarios with new physics beyond the SM are predicted to appear at the LHC. Many of the hypothetical new particles are expected to decay into jets, which are the experimental signatures of quarks and gluons. A deep understanding of jets is therefore essential for potential discoveries of new physics.

The CMS collaboration exploits a sequential calibration chain to correct the momenta of reconstructed jets to match that of particle level jets. In the context of this thesis, the jet energy scale in proton-proton collision data recorded in 2015 at $\sqrt{s} = 13\,\text{TeV}$ by the CMS experiment is measured in dijet event topologies. Jet responses are calculated in data and simulation as a function of pseudorapidity. Data-to-simulation correction factors are determined, which are provided to the CMS collaboration and applied in most CMS analyses performed with 2015 data that use jet objects.

A search for leptoquarks based on precisely measured jets is presented in this the-

sis. Leptoquarks are hypothetical particles that appear in many extensions of the SM, such as grand unified theories, technicolor, and composite models. They decay into a quark and a lepton and hence carry both lepton and baryon number, have fractional electric charge, and can either be scalar (spin-0) or vector (spin-1) particles. This thesis presents a search for pair-produced scalar leptoquarks decaying into a top quark and a tau lepton. For this search, the proton-proton collision data recorded in 2016 at $\sqrt{s} = 13\,\text{TeV}$ with the CMS detector, corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$ are analyzed. The final state that is studied contains three or more jets, an electron or muon, and one or two hadronically decaying tau leptons. A data-driven technique is implemented to estimate the main backgrounds that contain misidentified tau leptons. This analysis is the first search for pair produced leptoquarks decaying into a top quark and a tau lepton at $\sqrt{s} = 13\,\text{TeV}$ and resulted in a public preliminary result released by the CMS collaboration [1]. The results of an analysis performed earlier at $\sqrt{s} = 8\,\text{TeV}$ can be found in Ref. [2].

This thesis is organized as follows. Chapter 2 contains the theoretical foundations of the SM and physics beyond the SM, where the emphasis is given to leptoquarks. The LHC and the CMS detector are described in chapter 3. The event simulation using Monte Carlo (MC) techniques is introduced in chapter 4, while the algorithms to reconstruct and identify physics objects are presented in chapter 5. A measurement of the jet energy scale using dijet event topologies is presented in chapter 6. In chapter 7, a search for third-generation scalar leptoquarks decaying into a top quark and a tau lepton is presented. A summary is given in chapter 8.

2 Phenomenological Aspects of the Standard Model and Beyond

In this chapter the theoretical foundations relevant for this thesis are introduced. In Sec. 2.1, the Standard Model (SM) of particle physics is discussed. The SM describes the elementary particles and three of the four fundamental forces observed in nature. It is one of the most successful theories in physics. However, the SM is not yet able to explain all observed phenomena. A summary of the shortcomings of the SM is presented in Sec. 2.2 and possible extensions of the SM are described in Sec. 2.3. Many of the scenarios beyond the SM include hypothetical new particles, called leptoquarks (LQs). An introduction to the phenomenology of LQs is presented in Sec. 2.4.

2.1 Standard Model of Particle Physics

The SM of particle physics is a quantum field theory that describes the elementary particles and the strong, weak, and electromagnetic interactions between the particles. A gauge invariant quantum field theory describes the mathematics of the SM with an underlying gauge symmetry of the direct product of $SU(3)_C \times SU(2)_L \times U(1)_Y$. The SM is renormalizable, ensuring that infinities arising from higher-order corrections are treated, so that cross sections remain convergent. From an experimental perspective, the SM is a very successful theory, which predicts and describes the outcome of particle physics measurements with great accuracy and precision. A summary of the particles and interactions of the SM is given in the following and is based on Ref. [3, 4], unless explicitly stated otherwise.

Particles and Interactions of the SM

The SM includes elementary particles grouped into fermions and bosons. Fermions and bosons are distinguished via a quantum number called *spin*. The fermions are particles with half-integer spin, and are the constituents of the matter that

can be observed in nature. They are further grouped into six leptons and quarks (and their anti partners). The lepton sector contains three generations of charged leptons, namely the electron (e), the muon (μ), and the tau lepton (τ), as well as three generations of neutral neutrinos, the electron-, muon-, and tau-neutrino ($\nu_{e,\mu,\tau}$). The quark sector contains three generations of up-type quarks and down-type quarks. The up (u), charm (c), and top (t) quarks are up-type quarks with an electromagnetic charge of $+\frac{2}{3}e$. The down (d), strange (s), and bottom (b) quarks represent the down-type quarks with an electromagnetic charge of $-\frac{1}{3}e$.

The gauge bosons in the SM are the $W^{\pm}$ bosons, the Z^0 boson, the photon γ, and the gluon g. These bosons are spin-1 particles, and are the mediators of interactions between fermions.

A summary of the fermions and gauge bosons in the SM is presented in Fig. 2.1. The particle interactions described in the SM are discussed in the following.

The electromagnetic (EM) force is described by the theory of *quantum electrodynamics* (QED). The underlying gauge group is $U(1)_{EM}$. The massless photons are the mediators of EM interactions and couple to particles with electric charge. However, they are electrically neutral and no interactions between photons are possible.

The mediators of the weak interaction are the $W^{\pm}$ and Z bosons. The quantum number related to the weak interaction is referred to as *weak isospin*, which characterizes fermions. Fermions with left-handed (and anti fermions with right-handed) chiralities are grouped into into doublets with weak isospin of either $+\frac{1}{2}$ or $-\frac{1}{2}$, and right(left)-handed (anti-)fermions are grouped into singlets with weak isospin of 0.

The underlying gauge group of the weak interaction is $SU(2)_L$, where L denotes that the mediators of the weak interaction only couple to fermions with left-handed chiralities. However, it has been experimentally observed that the Z boson couples preferentially to left-handed fermions, but also to right-handed fermions. To resolve this discrepancy, the EM and weak interactions are unified, referred to as electroweak interaction.

Electroweak interactions are described by the $SU(2)_L \times U(1)_Y$ symmetry group, where Y denotes the hypercharge, defined as $Y = 2(Q - T_3)$. In this equation, Q denotes the electric charge and T_3 the weak isospin. The generators of the $U(1)_Y$ and $SU(2)_L$ groups are the B and the $W_{1,2,3}$ gauge fields, respectively. The $W^{\pm}$

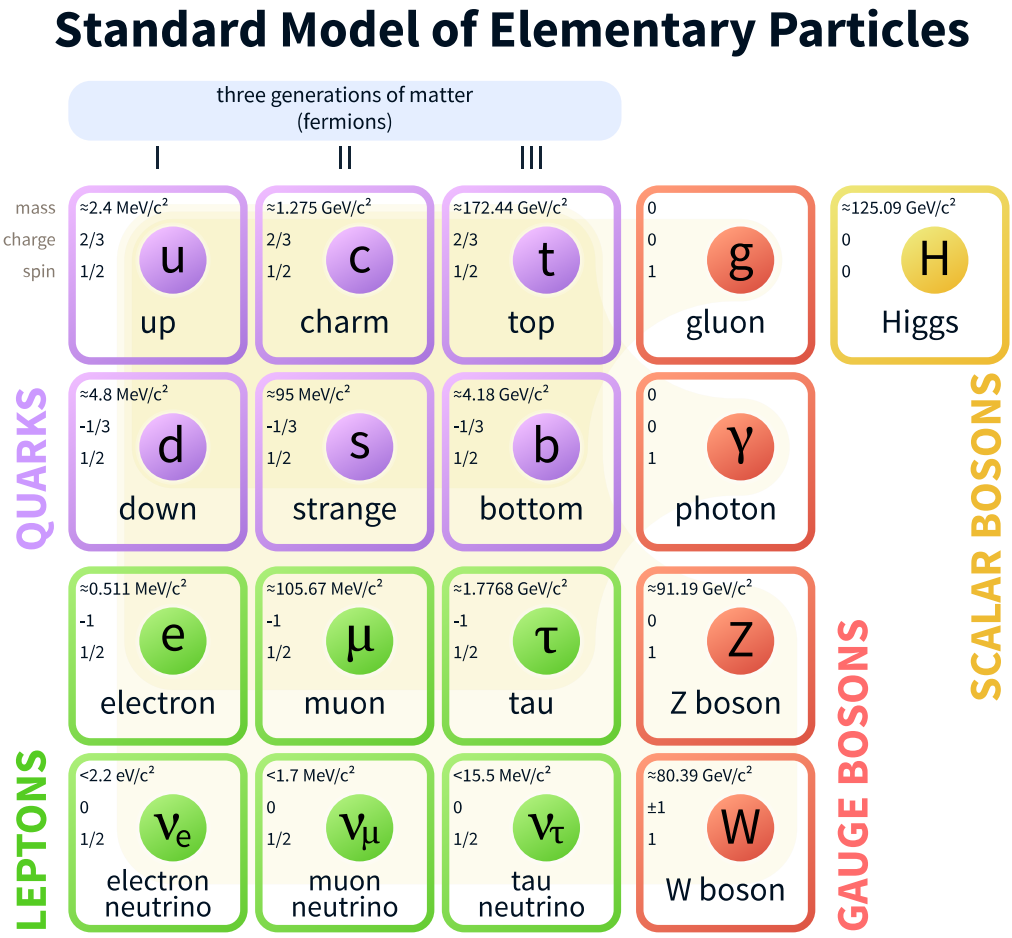

Figure 2.1: Elementary particles of the SM, grouped into fermions (quarks and leptons), gauge bosons and the scalar Higgs boson. The upper left values in each box correspond to the mass, charge, and spin, respectively. Taken from Ref. [5].

bosons are defined as linear combinations of W_1 and W_2:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2). \tag{2.1}$$

The Z boson and the photon γ are also mixtures of W_3 and B:

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}, \tag{2.2}$$

where θ_W denotes the weak mixing angle.

The $W^{\pm}$ bosons couple only to left-handed fermions and right-handed anti fermions. The coupling of W bosons to leptons occurs within the same generation of leptons, whereas for couplings to quarks intergenerational mixings have been observed. Each quark (and lepton) carries a quantum number referred to as *flavor*. The probability of flavor-changing of quarks via interactions with $W^{\pm}$ bosons is described by the amplitudes of couplings in the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix. The magnitudes of the CKM matrix are [6]:

$$V_{CKM} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.220 & 0.995 & 0.041 \\ 0.008 & 0.040 & 0.999 \end{pmatrix}, \tag{2.3}$$

where $|V_{ij}|^2$ is the transition probability of a quark with flavor i to a quark with flavor j. The diagonal matrix-elements are close to unity, and the contributions from off-diagonal elements are small.

The strong force completes the list of fundamental forces described in the SM. The massless gluons are the mediators of strong interactions, which are described by the theory of *quantum chromodynamics* (QCD). The underlying symmetry of the strong force is $SU(3)_C$, where C denotes the *color charge* that gluons couple to. Quarks carry color charge and the possible color states are denoted as *red*, *green*, *blue*, along with their respective anti-states. Gluons carry charge color themselves, so that self interactions between gluons are possible.

The self interaction between gluons leads to larger coupling strengths between colored particles with increasing distance scales, resulting in the fact that colored particles cannot exist freely. The latter phenomena is referred to as *color confinement*. Conversely, the coupling between colored particles becomes smaller with decreasing distance scales, such that quarks behave like free particles, known as *asymptotic freedom*.

A bound formation of color-charged particles is called *hadron*. The possible bound states consist of either three quarks with different color charges (*baryons*) or of a quark-antiquark-pair with color and anti-color charge (*mesons*). The formation of hadrons from individual quarks and gluons is referred to as *hadronization*.

Electroweak Symmetry Breaking

The SM is built upon gauge invariance of the Lagrangian to certain symmetries, which explicitly requires the gauge bosons to be massless. However, non-zero masses have been observed for all particles of the SM, except for the gluon and the photon.

The Higgs mechanism [7,8] was introduced to resolve this discrepancy and causes a spontaneous breaking of the electroweak symmetry group. A doublet Φ of complex scalar fields is added to the $SU(2)_L$ symmetry group, where

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{2.4}$$

and is referred to as the *Higgs fields*. The associated potential V of the Higgs fields is given by

$$V(\Phi) = \mu^2|\Phi|^2 + \lambda|\Phi|^4. \tag{2.5}$$

A spontaneous breaking of the electroweak symmetry is caused if a minimum of this potential occurs. In order to obtain a minimum, and to preserve a renormalizable theory, the parameter λ is chosen such that $\lambda > 0$. The parameter μ^2 is chosen such that $\mu^2 < 0$, as otherwise the symmetry is not broken and thus the model cannot generate the masses of the gauge bosons. The minimum v of the potential is given by

$$v = \frac{-\mu^2}{2\lambda} \approx 246\,\text{GeV}, \tag{2.6}$$

where v is referred to as vacuum expectation value.

Although the chiral structure of the $SU(2)_L$ symmetry requires the fermions to be massless particles, the Higgs mechanism also generates the masses of the fermionic fields by adding the mass term

$$m_f = c_f \frac{v}{\sqrt{2}} \tag{2.7}$$

to the SM lagrangian. In this equation, c_f is a free parameter, referred to as the Yukawa coupling. The Higgs mechanism is therefore not able to predict the values of m_f.

The Higgs mechanism induces a spin-0 boson called Higgs boson, which corresponds to the excitations of the Higgs field. In 2012, the ATLAS and CMS collaborations reported the discovery of a Higgs-like boson with a mass of around 125 GeV [9, 10]. To date, many of the properties of the new boson have been measured [11–14] and are in agreement with the properties of the SM Higgs boson.

2.2 Shortcomings of the Standard Model

Although the SM is a very successful theory that describes the electroweak and strong interactions with excellent experimental confirmations, it is still incomplete with open questions remaining. A summary of the most popular shortcomings of the SM is given in the following.

As already mentioned, three fundamental forces in the universe are described by the SM. However, the fourth fundamental force, gravity, is not included in the SM. Although gravity is successfully described by the general theory of relativity (GTR), the SM and the GTR are not unifiable.

The SM predicts an almost equal ratio of matter and antimatter in the universe. However, the universe consists of mostly matter, and thus this issue is referred to as matter-antimatter asymmetry. A mechanism is required that breaks the charge-conjugation-parity(CP) symmetry in the SM. The CKM matrix induces CP-violating contributions, but these are not sufficient enough to resolve the observed matter-antimatter asymmetry [15].

From observations [16–23] we know that the SM with the electroweak and strong interactions describes only approximately 5% of the matter in the universe, with approximately 25% consisting of dark matter and the remaining fraction arises from so-called dark energy. To date, no proper dark matter candidate has been observed experimentally, and a theory consistent with the SM describing dark energy does not exist.

The unification of the electromagnetic and weak forces as described in the previous section leads to the quest to show that the electroweak and strong interactions are also manifestations of one single force. A theory that links the forces is known as a grand unified theory (GUT), which requires that the coupling constants of the SM interactions meet at very high energies, in the order of $10^{16}\,\mathrm{GeV}$, referred to as the GUT scale.

One of the theoretical problems of the SM is that the mass of the Higgs boson is expected to be very large due to higher order corrections. These corrections depend on the so-called Planck scale Λ (with $\Lambda \sim 10^{19}\,\mathrm{GeV}$), up to which the SM is valid.

This leads to a natural mass of the Higgs boson in the order of $10^{14} - 10^{17}$ GeV. However, the experimentally measured mass of the Higgs boson is approximately 125 GeV. To explain this discrepancy, *unnaturally fine-tuned* cancellations between the high order corrections are required. This is referred to as the *hierarchy problem*.

2.3 Possible Extensions of the Standard Model

Many theories extending the SM have been proposed to resolve the shortcomings of the SM discussed in the previous section. A summary of the most common theories beyond the SM is given in the following.

The most well-known extension to the SM is supersymmetry (SUSY). SUSY models [24,25] predict so-called *super-partners* for each SM particle, with the same quantum numbers as their respective SM partner except for the spin, which differs by a half-integer unit. Thus, the super-partners of the SM quarks and leptons are bosons, called *squarks* and *sleptons*, respectively, and the super-partners of the SM gauge bosons and the Higgs boson are fermions, called *gauginos*. General SUSY models can solve many of the shortcomings of the SM. The contributions of the super-partners cancel out the higher order corrections to the mass of the Higgs boson in the SM, which could solve the hierarchy problem. Moreover, the neutralino predicted in SUSY models provides a dark matter candidate, and is referred to as the *lightest supersymmetric particle*. SUSY allows the strong and the electroweak gauge couplings to unify at high energies, and it could explain the matter-antimatter asymmetry as CP-violating contributions are added.

None of the super-partners has been observed experimentally, which implies that SUSY is a broken symmetry, possibly at the TeV scale. To conserve lepton and baryon number, many SUSY models impose a new quantum number called *R-parity*. The R-parity is defined as $R = (-1)^{3(\mathrm{B}-\mathrm{L})+2\mathrm{S}}$, where B, L and S denote the baryon, lepton and spin quantum numbers, respectively. All SM particles have $R = +1$, whereas SUSY particles have $R = -1$.

As already mentioned in the previous section, GUTs seek to unify the SM forces at the GUT scale. Moreover, GUTs aim for a possible unification of quarks and leptons as different states of one family. To realize GUTs, the SM symmetry group is embedded in one or more higher symmetry groups. The original proposals of GUTs

are based on $SU(4)$ and $SU(5)$ symmetries, pionieered by Pati and Salam [26], and Georgi and Glashow [27], respectively. However, both models have been ruled out for different reasons. In the Pati-Salam model the gauge couplings never unify, and the Georgi-Glashow model predicts that the proton should decay faster than already excluded by experimental lower limits on the proton lifetime.

Still, GUTs remain attractive models, and could be realized by embedding the SM symmetry group into even larger symmetries, e.g. $SO(10)$ or $E(6)$, to resolve the problems of the pionieer GUTs. Furthermore, an extension of already excluded $SU(5)$ GUTs to supersymmetric GUTs could still be realized [28].

The similarity between quarks and leptons in the SM implies that fermions might have a substructure, as considered by compositeness models [29, 30], in which the subcomponents of the fermions typically are referred to as *preons*. Such models could explain the remarkable symmetry between quarks and leptons in the SM. The matter-antimatter asymmetry could also be solved if the antimatter is contained by preons.

Furthermore, various composite Higgs theories have been proposed, e.g. in Ref. [31–34], in which the Higgs boson is not a fundamental particle, but a composite state. This state is bound by a new interaction, similar to the strong interaction in the SM, with a confinement scale in the order of several TeV. In such a model the mass of the Higgs boson does not depend on higher order corrections from scales greater than the new confinement scale. This leads to a mass of the Higgs boson compatible with the electroweak scale and thus could resolve the hierarchy problem.

Techni-color (TC) [35] models provide an explanation of the electroweak symmetry breaking without the requirement of the Higgs mechanism. TC models are based on a new strong gauge interaction that becomes significant at the electroweak scale. In the simplest TC model, a new set of fundamental fermions, called *techni-fermions*, is introduced. The scalar Higgs boson (and generally all scalars) is considered bound states of techni-fermions.

To predict the observed masses of the SM fermions, extended techni-color (ETC) models [36] are required, which assume that the TC-fermion-antifermion asymmetry is broken. ETC models include additional techni-fermions and gauge interactions, leading to interactions between techni-fermions and SM fermions.

2.4 Leptoquark Phenomenology

The remarkable symmetry between quarks and leptons in the SM hints towards more fundumental theories that allow transitions between them. Most such theories include new hypothetical particles called *leptoquarks* (LQs). LQs are color-triplet bosons and either scalar (spin-0) or vector (spin-1) particles and can turn quarks into leptons and vice-versa. They carry both lepton and baryon numbers and have fractional electric charge. LQs are often (but not strictly) assumed to couple to a single generation of quarks and leptons, referred to as first-, second-, or third-generation LQs.

Leptoquarks in BSM Theories

LQs appear in many scenarios beyond the SM. Here, the realization of LQs in the models introduced in the previous section is discussed.

In GUTs, some of the seemingly arbitrary features of the SM are contained within models with more symmetry [37]. The strong, weak, and electromagnetic forces are embedded in an underlying gauge theory, where the new symmetry also generates vector LQs. The masses of vector LQs are required to be very large and in the order of the GUT scale to avoid discrepancies with experimental limits on the proton lifetime and flavor changing neutral current processes, such as in the Pati-Salam SU(4) and Georgi-Glashow SU(5) models. Scalar LQ states appear in color triplets in the symmetry breaking sector. The expected range of scalar LQ masses depends on the model. Some unification models with extended symmetries as in Ref. [38,39] predict masses of scalar LQs of around a few TeV. Moreover, enhanced couplings of LQs to third-generation fermions are predicted in a four-color symmetry model [40].

In extended technicolor (ETC) models fermionic resonances can arise from techni-fermions, which are supposed to interact with SM fermions [41]. Such a model can contain LQs in the form of techni-mesons and would be built by a techni-quark and a techni-lepton. LQs with masses of around 1 TeV are expected in such an ETC model.

The existence of LQs has been predicted in several composite models, e.g. in Ref. [42–44]. In these models, leptons and quarks arise as bound states of more fundamental particles. Specific composite models include LQs and predict LQ masses

S	F	$SU(3)_C$	$SU(2)_W$	$U(1)_Y$	Allowed coupling
0	-2	$\bar{3}$	1	1/3	$\bar{q}_L \ell_L$, $\bar{u}_R e_R$
0	-2	$\bar{3}$	1	4/3	$\bar{d}_R e_R$
0	-2	$\bar{3}$	3	1/3	$\bar{q}_L \ell_L$
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	$\bar{q}_L \gamma^\mu e_R$, $\bar{d}_R \gamma^\mu \ell_L$
1	-2	$\bar{3}$	2	-1/6	$\bar{u}_R \gamma^\mu \ell_L$
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: Scalar and vector LQ states as defined in the BRW model. The fermion number is defined as $F = 3B + L$. Taken from Ref. [6] and modified.

around the TeV scale. They favor couplings to third-generation quarks and leptons [45, 46].

LQs can also be included in a supersymmetric SM, e.g. the scalar partners of each quark have the capability to couple to quarks and leptons similar to the coupling of LQs. This is realized in models such as in Ref. [47–49], in which the R-parity is required to be violated.

The Buchmüller-Rückl-Wyler Model

A general theory introducing LQs is presented by the Buchmüller-Rückl-Wyler (BRW) model [50], which is a model independent approach to classify the production and the decay of LQs. In this model, LQs conserve lepton and baryon number, and renomalization and gauge invariance is assumed for LQ interactions with the SM fermions and bosons. Tab. 2.1 shows the quantum numbers and allowed couplings of LQ states that are possible with the assumptions made in the BRW model. The spin S of LQs is either 0 or 1 and the fermion number can be either 0 or -2. The columns of $SU(3)_C$, $SU(2)_W$, and $U(1)_Y$ are the SM gauge group representations, respectively. The allowed couplings of LQ states are shown in the last column of Tab. 2.1. Fermion doublets are described by ℓ and q. Up- and down-type quarks and electrons are denoted by u, d, and e, respectively, where the indices L and R indicate left- and right-handed fermions. A global replacement of the respective

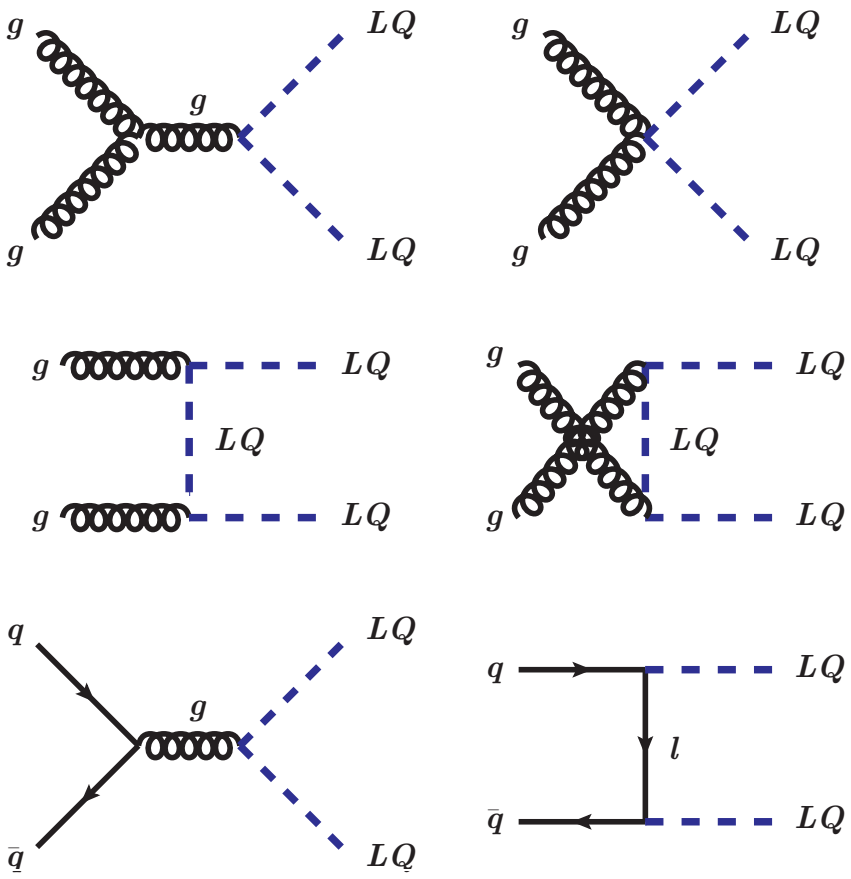

Figure 2.2: Leading-order Feynman diagrams of LQ pair production.

generation of fermions in Tab. 2.1 gives the corresponding generation of LQs.

Production of Leptoquarks in Proton-Proton Collisions

In proton-proton collisions, LQs would be predominantly produced in pairs through gluon-gluon fusion or quark-antiquark annihilation processes. Feynman diagrams of such processes are shown in Fig. 2.2. The dominant processes arise from the strong interaction sector. The contribution of the lepton mediated $t(u)$-channel is also shown in the bottom right diagram of Fig. 2.2, which depends on the unknown quark-lepton-LQ Yukawa coupling. For LQs decaying into third-generation quarks and leptons, this production mode is very unlikely at the LHC as third-generation quarks would be required in the initial state. Moreover, single production of LQs is generally possible at the LHC, but third-generation quarks would be required to produce the final state of interest in this analysis.

The leading order (LO) pair production cross section of LQs is [51, 52]

$$\sigma_{gg}^{\mathrm{LQ}} = \frac{\pi\alpha_s^2}{96\hat{s}}\left[\xi(41-31\xi^2)+(18\xi^2-\xi^4-17)\log\frac{1+\xi}{1-\xi}\right] \tag{2.8}$$

for gluon-gluon fusion processes, and

$$\sigma_{q\bar{q}}^{\mathrm{LQ}} = \frac{2\pi\alpha_s^2\xi^3}{27\hat{s}} \tag{2.9}$$

for quark-antiquark annihilation processes. The LQ velocity is described by $\xi = \sqrt{1-\frac{4M_{LQ}^2}{\hat{s}}}$, $\sqrt{\hat{s}}$ denotes the partonic center-of-mass energy, and α_s is the strong coupling constant. A more precise prediction of the cross section is obtained by considering higher-order corrections, including gluon loops and additional radiation of quarks. Including such corrections reduces the dependence of the production cross-section on the renomalization and factorization scales. In the presented search the next-to-leading order (NLO) cross sections at $\sqrt{s} = 13\,\mathrm{TeV}$ are determined based on Ref. [51].

Status of Searches for Leptoquarks

A variety of searches for singly and pair produced LQs has already been performed. In e^+e^-, ep, $p\bar{p}$, and pp colliders different LQ decay modes have been studied. LQ candidates have not been found in any of the experiments and thus exclusion limits on the cross sections and the masses of LQs have been set. The ATLAS and CMS experiments provide the most stringent exclusion limits to date. A summary of searches for pair-produced LQs at the LHC in different final states and their excluded LQ masses are shown in Tab. 2.2. Limits have been set for all generations of LQs. Third-generation searches have been performed for LQs decaying to $b\tau$, $b\nu$, $t\nu$ and $t\tau$. The analysis in Ref. [2] studies the decay channel of LQLQ $\rightarrow t\tau t\tau$ and has been performed at $\sqrt{s} = 8\,\mathrm{TeV}$. An analysis in the same channel performed at $\sqrt{s} = 13\,\mathrm{TeV}$ is presented in Chapter 7 of this thesis.

A summary of searches performed earlier in other experiments, e.g. at LEP, HERA, and Tevatron can be found in Ref. [53].

Recently, anomalies have been observed in measurements of the $\bar{B} \rightarrow D\tau\bar{\nu}$ and $\bar{B} \rightarrow D^*\tau\bar{\nu}$ decay rates by the BABAR [54, 55], Belle [56–58], and LHCb [59] collaborations. Deviations from the SM are found to be in the order of four standard

decay mode	experiment	lower mass limit [GeV]
$eejj$	ATLAS	1100 [67]
$eejj$	CMS	1130 [68]
$\mu\mu jj$	ATLAS	1050 [67]
$\mu\mu jj$	CMS	1165 [69]
$bb\tau\tau$	ATLAS	534 [70]
$bb\tau\tau$	CMS	850 [71]
$bb\nu\nu$	ATLAS	625 [72]
$bb\nu\nu$	CMS	450 [73]
$tt\nu\nu$	ATLAS	640 [72]
$tt\tau\tau$	CMS	685 [2]

Table 2.2: Summary of searches for pair-produced scalar LQs performed at the LHC. The most stringent limits on LQ masses in the respective decay mode to date from the ATLAS and CMS experiments are shown. The upper limits at 95% confidence level are shown, assuming a branching ratio of 100% to each decay channel.

deviations [60]. In order to explain these anomalies, flavor structure studies of LQ couplings have been performed, which show that predominant couplings to third-generation fermions are required [61–63].

Moreover, discrepancies in the ratio of $B^+ \to K^+\mu^+\mu^-$ to $B^+ \to K^+e^+e^-$ branching ratios observed by the LHCb collaboration [64–66] could be explained in a specific composite model, which predicts LQs decaying predominantly to third-generation fermions and LQs with masses around the TeV scale [46].

3 Experimental Setup

The analyses presented in this thesis are based on proton-proton collision data recorded by the Compact Muon Solenoid (CMS) detector [74, 75] at the Large Hadron Collider (LHC) [76] in 2015 and 2016. In the following, the LHC is introduced, followed by an overview of the detector components of the CMS experiment.

3.1 The Large Hadron Collider

The LHC is a circular accelerator and collider for protons and heavy ions with a circumference of 26.7 km located at CERN (**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire, European Organization for Nuclear Research) near Geneva. The LHC lies 45 m to 170 m below the surface and was designed to collide protons at a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$.

The protons are produced by the ionization of hydrogen gas using a duoplasmatron. Before the protons can be injected into the LHC, they are accelerated in a chain of various pre-accelerators. An overview of the CERN accelerator complex as well as the main experiments is shown schematically in Fig. 3.1. The accelerator chain consists of several linear accelerators (LINAC), the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). After the protons passed through the pre-accelerator chain, the energy of the proton beam has been increased to 450 GeV and the beams are injected into the LHC for further acceleration up to an energy of 14 TeV. Finally, the protons are brought to collisions at fixed interaction points, which are surrounded by particle detectors. The four main experiments at the LHC are ALICE, ATLAS, CMS, and LHCb. The multipurpose detectors ATLAS and CMS are designed for precision measurements of standard model processes, the search and the measurement of the standard model Higgs boson, and searches for physics beyond the standard model. The LHCb experiment is optimized for the measurement of general B-meson physics. At the ALICE experiment the focus lies on performing heavy ion collisions, to study the quark-gluon plasma.

One of the key advantages of the LHC is its high instantaneous luminosity $\mathcal{L}$. This is given by

$$\mathcal{L} = N_b f_r \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}, \tag{3.1}$$

where N_b is the number of bunches in a beam, f_r the revolution frequency of the bunches, and n_1 and n_2 the number of protons in the colliding bunches. The variables σ_x and σ_y are the transverse beam sizes in the horizontal and vertical direction, respectively. For the design luminosity the parameters are $N_b = 2808$ bunches per beam, $n_{1,2} = 1.15 \cdot 10^{11}$ protons per bunch, and a revolution frequency of $f_r = 11.25$ kHz. The value of the denominator of Eq. 3.1 depends on the bunch size and the beam crossing angle. With the design parameters, the LHC operates at $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for proton-proton collisions. The LHC has already run at higher luminosities, and reached a maximum of $\mathcal{L} = 1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ in 2016.

The integrated luminosity $L = \int \mathcal{L} dt$ represents a measure of the total number of collisions and total amount of data taken in a period. In Fig. 3.2, the evolution of the total integrated luminosity delivered by the LHC during the data taking periods between 2010 and 2016 is shown. The collisions during the run periods 2010/11 and 2012 were performed at a centre-of-mass energy of 7 TeV and 8 TeV, respectively. In this thesis, the collisions recorded in 2015 and 2016 at a centre-of-mass energy of 13 TeV are analyzed. The LHC delivered an integrated luminosity of $4.2\,\mathrm{fb}^{-1}$ in 2015. At the CMS experiment, a reduced amount of data ($2.3\,\mathrm{fb}^{-1}$) was recorded with full magnetic field and full detector operating. In 2016, the LHC delivered an integrated luminosity of $41.1\,\mathrm{fb}^{-1}$, out of which 37.8 fb^{-1} were recorded.

3.2 The Compact Muon Solenoid Detector

The CMS experiment is one of the two multipurpose experiments at the LHC. It has a length of 21.6 m, a diameter of 14.6 m and a mass of 14 000 t. The overall layout of the CMS detector is shown in Fig. 3.3. The detector is symmetric around the beam axis and covers almost the full solid angle of 4π around the interaction point. It consists of several layers, each of them optimized to detect different types of particles. The innermost part of the detector contains the silicon pixel tracker and the silicon strip tracker, surrounded by the electromagnetic calorimeter and then the hadronic calorimeter. All of these sub-detector systems are surrounded by a superconducting solenoid. In the outermost part of the detector are the muon chambers. Each sub-detector system will be discussed in more detail in later sections.

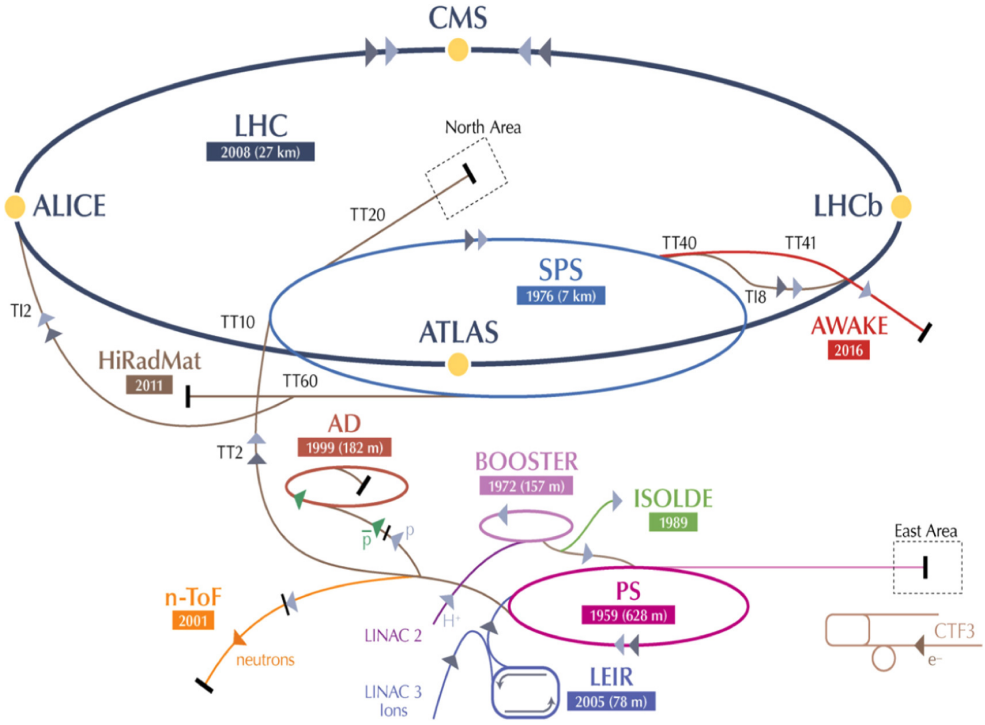

Figure 3.1: The accelerator complex at CERN. Taken from Ref. [77].

3.2.1 Coordinate Conventions

The CMS experiment uses a right-handed coordinate system with its origin centered at the nominal collision point. The x-axis points radially towards the center of the LHC ring, the y-axis points upwards vertically and the z-axis aligns with the counterclockwise beam direction. The azimuthal angle ϕ is measured in the x-y plane enclosed with the x-axis. The polar angle θ is the angle with respect to the z-axis. Instead of the polar angle θ, the pseudorapidity η, defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{3.2}$$

is used frequently because differences in η are conserved under Lorentz-transformations along the z-direction for massless particles. A value of $\eta = 0$ corresponds to the direction perpendicular to the beam and $|\eta| \to \infty$ points in the z-direction. The angular distance ΔR between two particles in the η-ϕ-plane is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.3}$$

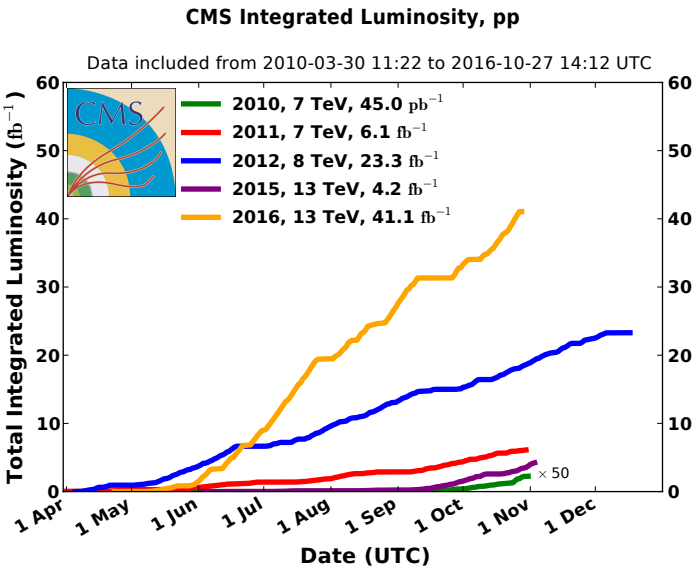

Figure 3.2: The total integrated luminosity delivered by the LHC during stable beams. It is shown for 2010 (green), 2011 (red), 2012 (blue), 2015 (purple) and 2016 (orange). Taken from Ref. [78].

and is also invariant under Lorentz-boosts in the z-direction.

3.2.2 Tracking System

The inner tracking system is the innermost part of the detector, directly surrounding the interaction point. It is designed to measure the charge-sign, the direction and, the momentum of charged particles. Moreover, it aims to identify primary and secondary vertices. It has a cylindrical shape, a length of 5.8 m and a radius of 1.1 m. It covers the region up to $|\eta| < 2.5$.

A diagram of the CMS tracking system is shown in Fig. 3.4. Closest to the collision point is the pixel detector. It consists of three layers at distances of 4.4 cm, 7.3 cm, and 10.2 cm to the beam. The size of one pixel cell is $100 \times 150\,\mu\text{m}^2$. In total, approximately 66 million pixel cells are installed in 1,440 modules. The resulting resolution is 10-20 μm for single hit positions. For the data taking starting in 2017 a fourth pixel layer has been installed.

The strip tracker is built around the pixel detector with a distance of 20-116 cm to the beam. It consists of 15,148 modules and more than 10 million readout channels. It is divided into various sectors. The barrel sector consists of the Tracker Inner Barrel (TIB), which spreads up to $|z| < 65$ cm and contains three Tracker Inner

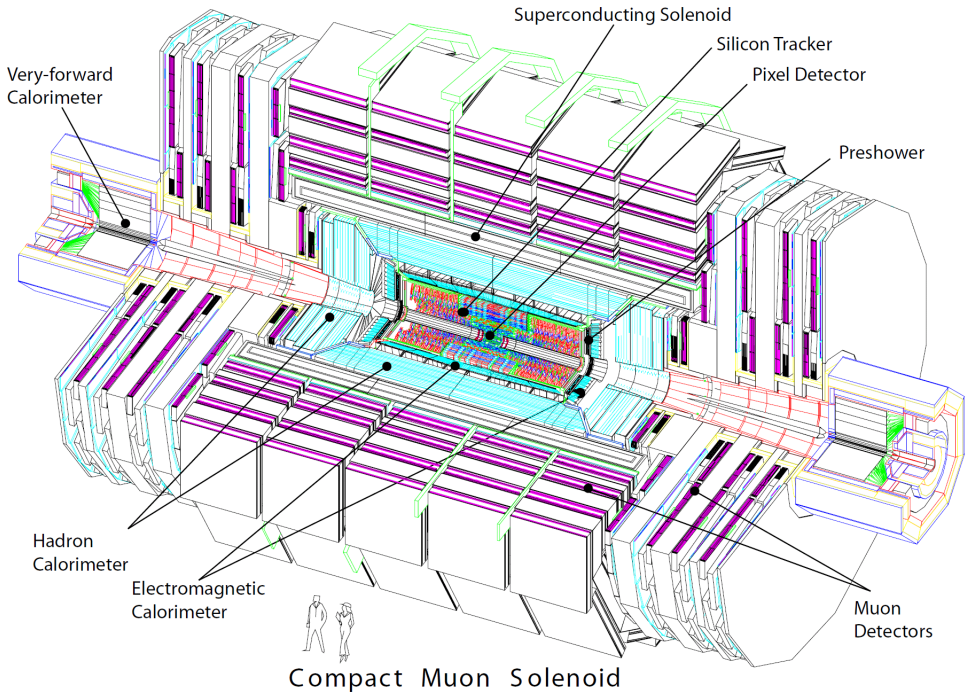

Figure 3.3: Overview of the CMS detector. Taken from Ref. [75].

Disks (TID). The TIB strip sensors have a thickness of 320 μm and the gap between the individual strips is in the range of 80-120 μm. The resulting resolution of the TIB is 23-24 μm. The barrel sector is extended by the Tracker Outer Barrel (TOB) with $|z| < 110$ cm. The TOB makes use of strip sensors, which have a thickness of 500 μm. The TOB is complemented by the Tracker End Cap (TEC), which contains nine further discs in the region of 120 cm$< |z| <$280 cm at each side of the tracking system. The thickness of the strip sensors in the TEC is 500 μm, where the gap between the sensors increased to 120-180 μm. The resulting resolution is 23-52 μm.

3.2.3 Calorimeters

The electromagnetic calorimeter (ECAL) surrounds the tracking system and is designed to provide excellent energy and position measurements of mainly electrons and photons. It is a homogeneous calorimeter and consists of 61 200 lead tungstate ($PbWO_4$) crystals in the central region and additional 7 324 crystals in each of the two endcaps. These crystals have a short radiation length of $X_0 = 0.89$ cm. The scintillators act with fast decays, with 80 % of the light emitted within the bunch

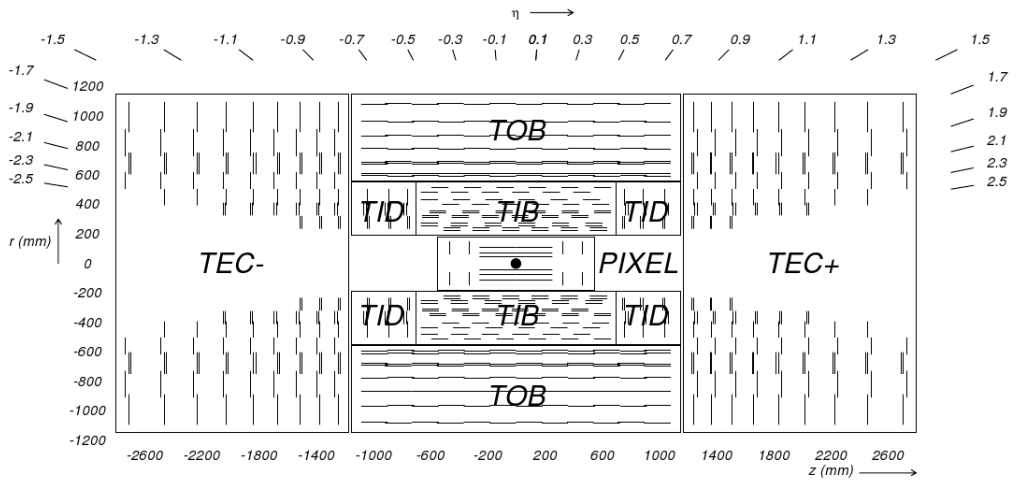

Figure 3.4: Diagram of the CMS tracking system. The pixel detector and the different sectors of the strip detector are shown in the rz-view. Taken from Ref. [75].

crossing spacing of 25 ns.

The layout of the CMS ECAL is shown schematically in Fig. 3.5. The barrel section (EB) covers a pseudorapidity range up to $|\eta| < 1.479$ and has a radius of 129 cm. It consists of crystals with a front face cross-section of approximately $22 \times 22\,\mathrm{mm}^2$ and a length of 230 mm, corresponding to 25.8 radiation lengths. The ECAL endcaps (EE) cover a pseudorapidity range of $1.479 < |\eta| < 3.0$ and have a distance of 314 cm from the vertex. The crystals in the endcaps have a front face cross-section of $28.6 \times 28.6\,\mathrm{mm}^2$ and a length of 220 mm, corresponding to 24.7 X_0. A preshower detector (ES) is placed in front of the endcap crystal calorimeter and covers a pseudorapidity range of $1.653 < |\eta| < 2.6$. It aims to identify neutral pions decaying to photon pairs, and consists of two layers of silicon strip detectors. The total thickness of the preshower detector is 20 cm, corresponding to $3\,X_0$.

The performance of the ECAL has been measured in a test beam. The energy resolution of electrons can be parametrized as follows [75]:

$$\frac{\sigma(E)}{E} = \frac{2.8\,\%}{\sqrt{E/\,\mathrm{GeV}}} \oplus \frac{12\,\%}{E/\,\mathrm{GeV}} \oplus 0.3\,\%. \tag{3.4}$$

The first term arises from statistical effects in the shower development in the ECAL, the second term describes the electronic noise, and the third, constant term arises from calibration errors and non-uniform light collection.

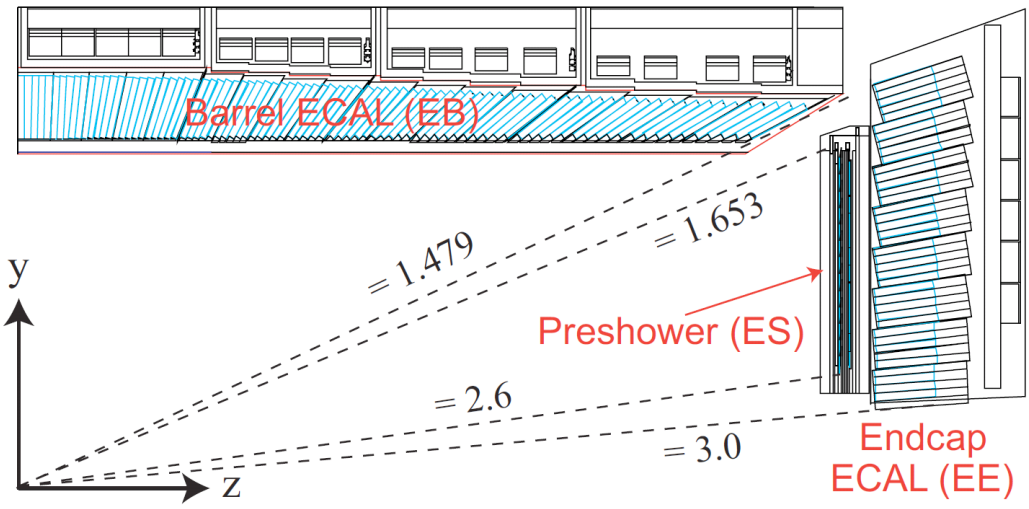

Figure 3.5: Diagram of one quarter of the CMS electromagnetic calorimeter. Taken from Ref. [74].

The hadron calorimeter (HCAL) is a sampling calorimeter with alternating layers of brass as absorber material and plastic scintillators as active medium. It completely surrounds the inner tracking system and the ECAL. The HCAL is built to provide precise energy measurements of hadron jets and to determine the missing transverse momentum.

The sub-detectors of the HCAL are shown in Fig. 3.6. The Hadron Barrel (HB) covers a pseudorapidity region of $|\eta| < 1.3$ and the Hadron Endcaps (HE) extend the coverage up to $|\eta| < 3.0$. Both contain brass plates with an interaction length of $\lambda_I = 16.4\,\text{cm}$ and are segmented into 2 304 towers covering 0.087×0.087 in η-ϕ for a pseudorapidity up to $|\eta| = 1.6$, and covering 0.17×0.17 in η-ϕ for larger pseudorapidities. The total thickness of the absorber material ranges from $5.8\,\lambda_I$ ($|\eta| = 0$) to a sufficient depth of about $10\,\lambda_I$ in the HE. The barrel region is completed by additional scintillator layers outside the magnet, referred to as Hadron Outer calorimeter (HO), which increase the thickness of the absorber material to about $10\,\lambda_I$ within $|\eta| < 1.26$.

The Hadron Forward calorimeter (HF) is located 11.2 m from the interaction point and slightly overlaps with the HE in η as it covers a pseudorapidity range of $2.9 < |\eta| < 5.2$. The HF makes use of steel as absorber material and quartz fibres as active medium. Due to the high particle fluxes, materials with a large radiation hardness are required. The HF is segmented and arranged in towers of 0.175×0.175 in η-ϕ, and the full depth of the absorber is 165 cm ($\approx 10\,\lambda_I$).

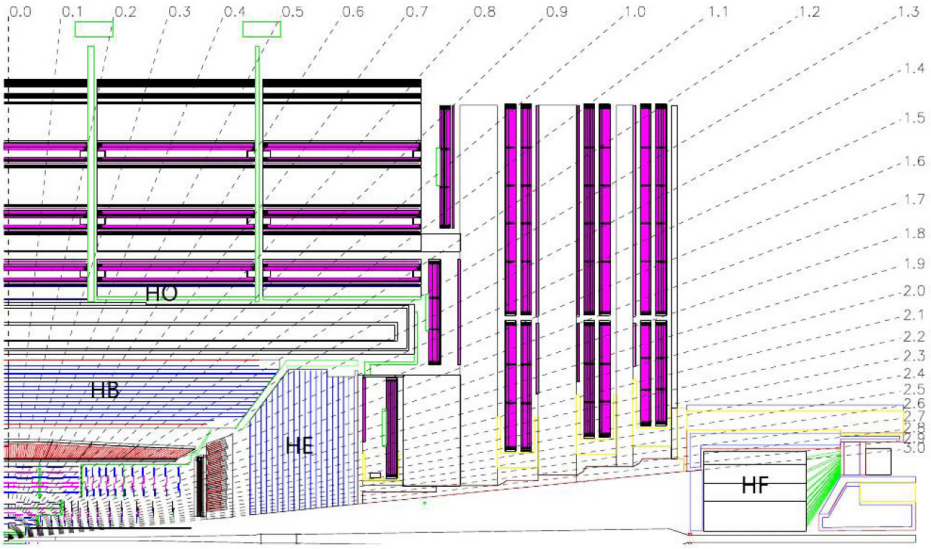

Figure 3.6: Diagram of one quarter of the CMS hadronic calorimeter. Taken from Ref. [75].

Similar to the ECAL, the performance of the HCAL has been measured in a test beam. Thus, a combined energy resolution of the ECAL and the HCAL system can be parametrized as [79]

$$\frac{\sigma(E)}{E} = \frac{115.3\,\%}{E/\,\mathrm{GeV}} \oplus 5.5\,\%, \tag{3.5}$$

which corresponds to an energy resolution of approximately 17 % for pions with an energy of $E = 50\,\mathrm{GeV}$.

3.2.4 Magnet System

The CMS experiment makes use of a superconducting solenoid which is designed to provide a field of 3.8 T. The solenoid surrounds the inner tracking system, the ECAL, and the HCAL. It has a length of 12.9 m and an inner diameter of 5.9 m. At full current an energy of 2.7 GJ is stored.

The bending power allows precise measurements of momenta and sign of the charge of charged particles. In addition, the return of the magnetic flux in the iron yokes provides a field strength of 2 T outside the solenoid. This leads to the trajectory of muons being bent in the opposite direction, allowing precise measurements of muon momenta.

3.2.5 Muon System

The muon system is the outermost part of the CMS detector. As muons are minimum ionizing particles, they pass through the inner tracking system and the calorimeters without significant energy loss. In contrast to other particle types, muons are the only particles that produce signals in the muon system.

Three types of gas detectors are used for the identification and measurement of muons. Drift tubes (DT) are used in the barrel region (MB) of $|\eta| < 1.2$, where the neutron induced background is low and the magnetic field is homogeneous. Each DT station is designed to provide a resolution of 100 μm in r-ϕ.

The endcap discs have a pseudorapidity range of $|\eta| > 0.9$ up to 2.4. Cathode strip chambers (CSC) are installed in this region. The CSCs are optimized for the large neutron backgrounds, the higher particle flux, and the stronger and less homogeneous magnetic field. They provide a spatial resolution of 75-150 μm in the r-ϕ-plane. Resistive plate chambers (RPC) complement the DTs and CSCs in the barrel and in the endcaps up to a pseudorapidity of $|\eta| < 1.6$. RPCs provide a fast response and a good time resolution.

The layout of the muon system is shown in Fig. 3.7. The barrel region is divided into four cylindrical stations. In each of the endcaps the CSCs and the RPCs are arranged in four discs perpendicular to the z-direction. In total, the muon system consists of 25,000 m^2 of active material and has almost a million readout channels.

3.2.6 Trigger System

The LHC provides a bunch crossing rate of 40 MHz, resulting in 10^9 interactions per second at the design luminosity. For technical reasons, it is not possible to store all events as the size of one event is in the order of 1 MB. Thus, only approximately 1000 events per second can be stored. To reduce the event rate a trigger system is employed. The trigger system selects potentially interesting events quickly and efficiently. This is performed in two steps.

First, the hardware based Level-1 (L1) trigger reduces the rate to 100 kHz by using information only from the calorimeters and the muon system. The time for a decision is limited to 3.2 μs. The trigger selects events if physics objects such as photons, electrons, muons, or jets pass a certain threshold in the transverse momentum.

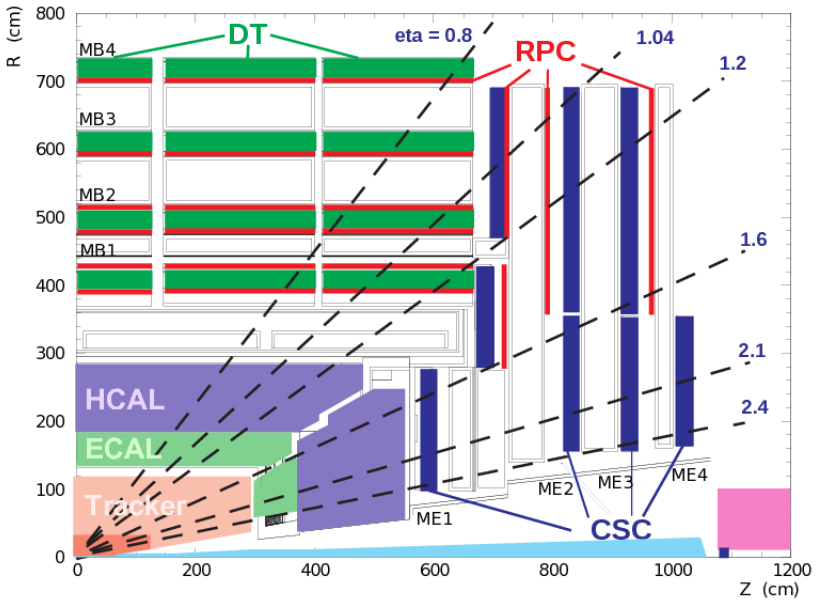

Figure 3.7: Layout of one quarter of the CMS muon system. Taken from Ref. [74].

For the events that pass the L1 trigger the rate is further reduced by the software based High-Level trigger (HLT). The HLT makes use of a farm of standard processors. Information from all sub-detectors fed into more sophisticated algorithms is used to select the events, similar to the offline reconstruction and selection. The time for a decision is limited to 50 ms per event. The HLT reduces the final output rate to the order of a few hundred Hz.

4 Event Simulation

The prediction of the outcome of the particle collisions such as those at the LHC is essential to validate theoretical models. Sophisticated event generators describe the structure of a proton-proton collision by a few main steps. The most common event generators and their corresponding simulation steps are discussed in Sec. 4.1. The background and signal samples used in this analysis are described in Sec. 4.2.

4.1 Event Generation

Monte-Carlo event generators are important tools for various purposes in particle physics, for example the design of the detector has been decided based on Monte-Carlo simulations. Moreover, event generators are employed to derive resolution and acceptance corrections, and they provide background estimates as well as a theoretical interpretation of results. Therefore, a sophisticated description of all aspects of the interaction is required.

The multi-purpose event generators HERWIG [80], PYTHIA [81], and SHERPA [82] cover the full chain composed of hard interaction, parton shower, and hadronization of simulated events. In some steps of the generation of events the multi-purpose generators can be interfaced to other generators.

The following description of each step of the event generation and the respective event generators is based on Ref. [83, 84].

Hard Interaction

The first step of event generators is the simulation of hard scatter processes of the proton-proton collisions. The hard interaction denotes the interaction between two partons a and b inside the colliding protons. In general, hard scatter processes describe the processes in the event with high transverse momentum transfer Q. The calculation of the cross section of these processes is based on perturbation theory.

At hadron colliders, the cross section for a scatter process $ab \to n$ is given by [85]

$$\sigma = \sum_{a,b} \int_0^1 \int_0^1 dx_a dx_b \int f_a(x_a, \mu_f) f_b(x_b, \mu_f) d\hat{\sigma}_{ab\to n}(\mu_f, \mu_r), \tag{4.1}$$

where $f(x, \mu)$ refers to the Parton Distribution Function (PDF), which describes the probability of the partons (a, b) taking part in this interaction. It depends on the momentum fraction x of a parton and on the factorization scale μ_f. The parton-level cross section for the production of the final state n is denoted by $\hat{\sigma}_{ab\to n}$ and depends on the factorization scale μ_f, the renormalization scale μ_r. In general, the choice of the renormalization scale and the factorization scale is not preassigned. The typical choice in most event generators for a specific hard scale Q^2 identified is $\mu_f = \mu_r = Q^2$.

The matrix element is evaluated as the sum over Feynman diagrams. All multi-purpose event generators calculate $2 \to n$, with n up to 3, matrix elements at LO. For higher order final states more dedicated generators are employed. The generators POWHEG [86,87] and MC@NLO [88] calculate specific processes at NLO, whereas the generators MADGRAPH [89–91] and aMC@NLO [91] can be used for general processes in NLO. The generator aMC@NLO performs calculations using the MC@NLO method for the matching to showers. All matrix element generators are able to calculate $2 \to n$ processes with unlimited values of n. However, the values of n are usually limited to a maximum of 9 to reduce the computing time. The simulated events of MC@NLO and aMC@NLO contain negative event weights in order to consider negative interferences due to double-counting of contributions in matching processes.

Parton Shower

The simulation of hard processes as discussed above describes the lower order of matrix elements. Higher order effects, i.e. a developing cascade caused by the radiation of gluons, additional initial state radiation (ISR) and final state radiation (FSR) are modeled by sophisticated parton shower algorithms. These algorithms simulate an evolution of parton showers downwards in momentum scale from high scales of hard processes to low scales (in the order of $\sim 1\,\text{GeV}$), at which the phase space is filled with soft gluons.

The parton shower algorithms are implemented in the event generators with dif-

fering approaches. These depend on the evolution variable, which is either the dipole mass (SHERPA), a certain angular requirement (HERWIG), or a transverse momentum requirement (PYTHIA). All multi-purpose event generators modeling the parton shower can get interfaced, i.e. with MADGRAPH, POWHEG, or MC@NLO. MADGRAPH includes an algorithm for interfacing, called *matching* in which the partons are clustered in jets with the k_T MLM scheme [92], POWHEG and MC@NLO are based on an extended subtraction scheme [93]. For aMC@NLO, the FxFx merging scheme [94] is used.

Hadronization

The partons modeled in perturbation theory as discussed in the previous steps can not be observed as free particles and thus they are not the final state particles. Instead, the dynamic enters a non-perturbative phase in which the partons are confined to hadrons. This process is referred to as hadronization. Two different models are used for the implementation of the hadronization procedure in the various event generators. One is the string model [95], which directly transforms the parton system to hadrons. This approach is implemented in PYTHIA. The alternative is the cluster model [96, 97], which is implemented in HERWIG and SHERPA. This method introduces an intermediate step of cluster objects with mass scales in the order of a few GeV.

In addition, unstable resonances occur due to many of the hadrons produced during hadronization. These resonances are simulated by the multi-purpose generators.

Underlying Event and Pileup

The preceding steps discussed above fully describe the process of the main parton-parton interaction. However, collisions between other partons in the protons that do not participate directly in the hard interaction lead to hadron remnants, whose potential hadronization and interaction with each other needs to be considered. This additional activity is referred to as the underlying event (UE) and is typically caused by gluon-gluon scattering processes, $gg \to gg$. The UE is taken into account by empirical models with parameters obtained from collision data, so-called *tunes*. These tunes are adapted in each multi-purpose event generator, respectively. In this analysis, the CUET Monash tune [98, 99] is used.

Additional activity of prior, simultaneous, or later bunch crossings can contribute to the event due to the limited resolution of the detector. This effect is referred to as pileup. Pileup effects are simulated by PYTHIA in this thesis.

Detector Simulation

The last step is to simulate how the generated particles interact with the detector material. The full simulation of the CMS detector is performed with the GEANT4 package [100]. GEANT4 is a toolkit that provides tools for detector geometry and sensitive detector response. It is also able to monitor the particle tracking. In general, the GEANT4 package is a common tool widely used in high energy physics.

4.2 Simulated Samples

The simulated signal and background samples used for the search for scalar leptoquarks have been generated in a MC production campaign of the CMS experiment and are summarized in this section. The simulated samples used for the jet energy calibration are described separately in chapter 6.

PYTHIA is used to simulate samples of pair produced scalar leptoquarks, where both leptoquarks are generated to decay into a top quark and a tau lepton. The signal samples are generated for leptoquark masses between 200 GeV and 2000 GeV. The signal samples used in this analysis are summarized in Tab. 4.1. The theory cross sections are determined at NLO accuracy [51].

The production of events with top quark pairs and the electroweak production of single top quark events in the tW-channel are simulated with POWHEG. The generator aMC@NLO is used to simulate single top quark events in the t-channel and in the s-channel. The simulation of W+jets events is also performed by aMC@NLO. The W+jets samples are produced in separate bins of $\hat{p}_{\mathrm{T}}$, which denotes the momentum transfer of the hard process. MADGRAPH is used to generate Drell-Yan production in association with jets and the MLM matching scheme is used to interface these processes with PYTHIA. The production of diboson (WW, WZ, and ZZ) events and the QCD multi-jet production is simulated with PYTHIA. All QCD samples are produced in separate bins of $\hat{p}_{\mathrm{T}}$. The QCD multijet samples include a *filter* that pre-selects events, which consist of at least one muon (μ enriched) or events in

Signal sample	Generator	σ [pb]	N_{gen}
LQLQ $\to t\tau t\tau$ (200GeV)	PYTHIA	60.6	74,350
LQLQ $\to t\tau t\tau$ (300GeV)	PYTHIA	8.05	74,997
LQLQ $\to t\tau t\tau$ (400GeV)	PYTHIA	1.74	74,452
LQLQ $\to t\tau t\tau$ (500GeV)	PYTHIA	0.496	74,675
LQLQ $\to t\tau t\tau$ (600GeV)	PYTHIA	0.169	74,988
LQLQ $\to t\tau t\tau$ (700GeV)	PYTHIA	0.0648	74,983
LQLQ $\to t\tau t\tau$ (800GeV)	PYTHIA	0.0273	74,983
LQLQ $\to t\tau t\tau$ (900GeV)	PYTHIA	0.0123	74,979
LQLQ $\to t\tau t\tau$ (1000GeV)	PYTHIA	0.00586	74,985
LQLQ $\to t\tau t\tau$ (1200GeV)	PYTHIA	0.00150	74,512
LQLQ $\to t\tau t\tau$ (1400GeV)	PYTHIA	0.000432	74,691
LQLQ $\to t\tau t\tau$ (1700GeV)	PYTHIA	0.0000774	74,158
LQLQ $\to t\tau t\tau$ (2000GeV)	PYTHIA	0.0000155	74,778

Table 4.1: Summary of the simulated signal samples. The cross sections are taken from Ref. [51].

which a final-state particle may be mis-reconstructed as an electron (EM enriched). Another QCD sample includes events with electrons arising from the decay of b or c quarks (BCtoE).

All background samples used in this analysis are summarized in Tab. 4.2. It shows the generators used for the simulation, the cross section σ and the number of generated events N_{gen} of each sample. To predict the number of selected data events, an event weight w is applied to each simulated event. It is defined as

$$w = \frac{\sigma \cdot L}{N_{\text{gen}}}, \tag{4.2}$$

where the integrated luminosity L corresponds to $35.9\,\text{fb}^{-1}$.

Background sample	Generators	σ [pb]	N_{gen}
$t\bar{t}$	POWHEG + PYTHIA	831.8	155,159,297
t (t-ch.)	aMC@NLO + PYTHIA	44.3	5,993,570
$\bar{t}$ (t-ch.)	aMC@NLO + PYTHIA	26.4	3,927,980
t and $\bar{t}$ (s-ch.)	aMC@NLO + PYTHIA	3.4	3,370,580
t (tW-ch.)	powheg + PYTHIA	35.85	6,942,907
$\bar{t}$ (tW-ch.)	powheg + PYTHIA	35.85	6,932,903
Z+jets ($Z \to \ell\ell$), H_{T} 100-200	MadGraph + PYTHIA	$147.4 \cdot 1.23$	10,606,926
Z+jets ($Z \to \ell\ell$), H_{T} 200-400	MadGraph + PYTHIA	$40.99 \cdot 1.23$	9,646,008
Z+jets ($Z \to \ell\ell$), H_{T} 400-600	MadGraph + PYTHIA	$5.678 \cdot 1.23$	10,008,141
Z+jets ($Z \to \ell\ell$), H_{T} 600-800	MadGraph + PYTHIA	$1.367 \cdot 1.23$	8,292,160
Z+jets ($Z \to \ell\ell$), H_{T} 800-1200	MadGraph + PYTHIA	$0.63 \cdot 1.23$	2,668,311
Z+jets ($Z \to \ell\ell$), H_{T} 1200-2500	MadGraph + PYTHIA	$0.1514 \cdot 1.23$	595,906
Z+jets ($Z \to \ell\ell$), H_{T} 2500-inf	MadGraph + PYTHIA	$0.003565 \cdot 1.23$	399,147
W+jets ($W \to \ell\nu$), $\hat{p}_{\text{T}}$ 100-250	aMC@NLO + PYTHIA	676.3	119,168,000
W+jets ($W \to \ell\nu$), $\hat{p}_{\text{T}}$ 250-400	aMC@NLO + PYTHIA	23.9	12,021,700
W+jets ($W \to \ell\nu$), $\hat{p}_{\text{T}}$ 400-600	aMC@NLO + PYTHIA	3.03	1,939,698
W+jets ($W \to \ell\nu$), $\hat{p}_{\text{T}}$ 600-inf	aMC@NLO + PYTHIA	0.45	1,914,241
WW	PYTHIA	118.7	994,017
WZ	PYTHIA	47.1	990,003
ZZ	PYTHIA	16.5	993,154
QCD μ enr., $\hat{p}_{\text{T}}$ 20-30	PYTHIA	$558{,}528{,}000 \cdot 0.0053$	31,475,095
QCD μ enr., $\hat{p}_{\text{T}}$ 30-50	PYTHIA	$139{,}803{,}000 \cdot 0.01182$	29,944,719
QCD μ enr., $\hat{p}_{\text{T}}$ 50-80	PYTHIA	$19{,}222{,}500 \cdot 0.02276$	19,806,515
QCD μ enr., $\hat{p}_{\text{T}}$ 80-120	PYTHIA	$2{,}758{,}420 \cdot 0.03844$	13,778,177
QCD μ enr., $\hat{p}_{\text{T}}$ 120-170	PYTHIA	$469{,}797 \cdot 0.05362$	8,042,660
QCD μ enr., $\hat{p}_{\text{T}}$ 170-300	PYTHIA	$117{,}989 \cdot 0.07335$	7,946,703
QCD μ enr., $\hat{p}_{\text{T}}$ 300-470	PYTHIA	$7820.3 \cdot 0.10196$	7,936,465
QCD μ enr., $\hat{p}_{\text{T}}$ 470-600	PYTHIA	$645.5 \cdot 0.12242$	3,850,466
QCD μ enr., $\hat{p}_{\text{T}}$ 600-800	PYTHIA	$187.1 \cdot 0.13412$	4,008,200
QCD μ enr., $\hat{p}_{\text{T}}$ 800-1000	PYTHIA	$32.3 \cdot 0.14552$	3,959,757
QCD μ enr., $\hat{p}_{\text{T}}$ 1000-inf	PYTHIA	$10.4305 \cdot 0.15544$	3,976,075
QCD EM enr., $\hat{p}_{\text{T}}$ 20-30	PYTHIA	$557{,}600{,}000 \cdot 0.0096$	9,218,839
QCD EM enr., $\hat{p}_{\text{T}}$ 30-50	PYTHIA	$136{,}000{,}000 \cdot 0.073$	4,730,140
QCD EM enr., $\hat{p}_{\text{T}}$ 50-80	PYTHIA	$19{,}800{,}000 \cdot 0.146$	23,473,892
QCD EM enr., $\hat{p}_{\text{T}}$ 80-120	PYTHIA	$2{,}800{,}000 \cdot 0.125$	35,841,321
QCD EM enr., $\hat{p}_{\text{T}}$ 120-170	PYTHIA	$477{,}000 \cdot 0.132$	77,770,116
QCD EM enr., $\hat{p}_{\text{T}}$ 170-300	PYTHIA	$114{,}000 \cdot 0.165$	11,539,879
QCD EM enr., $\hat{p}_{\text{T}}$ 300-inf	PYTHIA	$9{,}000 \cdot 0.15$	7,373,130
QCD BCtoE, $\hat{p}_{\text{T}}$ 20-30	PYTHIA	$557{,}627{,}000 \cdot 0.00059$	10,889,299
QCD BCtoE, $\hat{p}_{\text{T}}$ 30-80	PYTHIA	$159{,}068{,}000 \cdot 0.00255$	15,327,935
QCD BCtoE, $\hat{p}_{\text{T}}$ 80-170	PYTHIA	$3{,}221{,}000 \cdot 0.01183$	14,976,480
QCD BCtoE, $\hat{p}_{\text{T}}$ 170-250	PYTHIA	$105{,}771 \cdot 0.02492$	9,522,972
QCD BCtoE, $\hat{p}_{\text{T}}$ 250-inf	PYTHIA	$21{,}094.1 \cdot 0.03375$	9,772,915

Table 4.2: List of background MC samples with their corresponding generators, production cross sections σ, and number of generated events N_{gen}. The numbers 1.23 refers to the K-factor (defined as inclusive NLO/LO) of the Z+jets sample. The numbers of the QCD samples correspond to the cross section times filter efficiency.

5 Object Reconstruction and Jet Energy Corrections at CMS

Before the events from proton-proton collisions are analyzed, the detector signals must be transformed into physics objects. The particle-flow (PF) algorithm [101] exploits the information of all components of the detector and is employed for the global reconstruction of stable particles in the event. The PF algorithm is introduced in Sec. 5.1. The determination of the position of the proton-proton interaction point of the event is described in Sec. 5.2.

The physics objects of interest in this analysis are muons, electrons, tau leptons and hadronic jets. For the identification of each object criteria are required as presented in Sec. 5.3 to 5.6. The calibration of jets to the correct energy scale is important and discussed in Sec. 5.7. The reconstruction of the missing transverse energy and of a global variable called S_{T} is introduced in Sec. 5.8. Finally, the identification of boosted top quarks is reviewed in Sec. 5.9.

5.1 Event Reconstruction with the Particle-Flow Algorithm

The CMS experiment utilizes the PF algorithm for the reconstruction of stable particles in the events. As a global reconstruction approach, the PF algorithm combines the information from all sub-components of the detector. The particles that can be reconstructed are electrons, muons, photons, as well as charged and neutral hadrons. The PF algorithm proceeds in several steps. It relies on an efficient track reconstruction, a high granularity ECAL and hermetic HCAL, and a clustering algorithm. To connect the elements of different sub-detectors, a link algorithm is employed. Later on, a list of stable particles is used to construct the particle-based objects and observables (such as jets, missing transverse energy, tau leptons, etc.). All steps are described in more detail in the following.

The track reconstruction is an essential element to obtain precise measurements of charged particles. To provide a track reconstruction with high efficiencies and low fake rates an iterative-tracking strategy is used [102]. Hits forming short tracks in the inner layers of the tracker are used as seeds, and tracks are reconstructed with tight criteria, such as being produced near the interaction region, initially. In the next iterations, the hits associated with tracks from previous iterations are removed from the collection of all hits, and less stringent seeding criteria are used to increase the tracking efficiency. With this procedure, the efficiency of the track reconstruction of isolated muons is found to be 99.5%, and greater than 90% for charged hadrons in jets. Further iterations with less stringent constraints on the vertex of origin are used to reconstruct secondary charged particles that originate from long-lived particles.

In the calorimeters, a clustering algorithm is employed for a high detection efficiency and optimal position resolution. The clustering algorithm is used to detect and measure the energy and direction of stable neutral particles. Moreover, the neutral particles are separated from energy deposits from charged hadrons. Except for the hadron forward calorimeter, the clustering is applied for each sub-component of the ECAL and HCAL, separately. In the first step of the algorithm, calorimeter cells are considered if their local energy maximums are above a certain threshold. These cells are used as seeds for the cluster. In a second step, neighbouring cells are added to the cluster if their energy exceeds a given threshold represented by two standard deviations of the electronics noise in the ECAL and an amount of 800 MeV in the HCAL. The result of this procedure is called *topological clustering* and is repeated until there are no more neighbouring cells left that fulfill the criterion.

In general, a single particle gives rise to several PF-elements in the sub-detectors. A link algorithm is developed in order to connect these elements and fully reconstruct each particle. Simultaneously, potential double counting from different (sub)-detectors the algorithm also avoids. For each pair of elements in the event a distance between the linked elements is defined and the quality of the link is evaluated. First, the charged-particle tracks are linked to the calorimeter cluster if the extrapolated position matches the corresponding cluster within the cluster boundaries. Moreover, bremsstrahlung is taken into account by linking a cluster to the track if the extrapolated tangents to the track are positioned within the boundaries of the clus-

ter. Similarly, two (or more) calorimeter clusters are linked if the cluster in the more granular calorimeter matches the cluster in the less granular calorimeter. A charged-particle track can be identified as a global muon if a certain reconstruction criterion is fulfilled as described in Sec. 5.3.

The linked information, for example a linked track and cluster in the ECAL, is referred to as a block. The PF algorithm makes use of the collection of particles from each block of elements as follows. First, after a muon is identified, the corresponding track is removed from the block. Similarly, the corresponding track and ECAL clusters are removed from the block when an electron was identified. With more stringent criteria applied to the remaining tracks, the corresponding elements are used to identify charged hadron, photon, or neutral hadron candidates. The momentum and energy of charged hadrons are determined by a certain fit of measurements in the tracker and the calorimeters. Photon candidates are expected to produce signals only in the ECAL. The neutral hadron energy is obtained from a combination of clusters in both ECAL and HCAL.

Other relevant objects used in this analysis, reconstructed tau leptons, b tagged jets, or the missing transverse energy, can be reconstructed using the complete list of particles computed by the PF algorithm.

5.2 Reconstruction of Primary Vertices

The primary vertex (PV) reconstruction [102] aims to measure the position of the corresponding proton-proton interaction point for each event. It is very likely to identify more than one PV in the event due to the presence of pileup effects.

The PV reconstruction consists of three steps. In a first step the reconstructed tracks are selected based on the following requirements:

- tracks are allowed to have an impact parameter significance $\frac{|d_0|}{\sigma_{d_0}}$ of less than 5, where in this context the impact parameter is a measure of the distance between the track and the center of the beam spot (average position of interactions),
- at least five hits in the inner tracking system and at least two pixel hits need

to be associated with the track, and

- the normalized χ^2 from a fit to the trajectory has to be smaller than 20.

In a second step the selected tracks originating from the same interaction vertex are clustered to build vertex candidates. This is performed by using a *deterministic annealing* algorithm [103].

In a final step the position of each vertex is fitted using its associated tracks. For this purpose, the *adaptive vertex fitter* [104] is employed. It assigns a weight w between 0 and 1 for each track associated with the vertex. The value of the weight is close to 1 for tracks that can be connected to the position of the reconstructed vertex. The number of degrees of freedom provides a measure of the quality of the fitting procedure and is defined as:

$$n_{dof} = -3 + 2 \sum_{i=1}^{N_{\text{track}}} w_i, \tag{5.1}$$

where the sum runs over the tracks associated with the vertex.

In this analysis, a PV candidate is required to fulfill $n_{dof} > 4$. In addition, the PV candidate is required to lie within 24 cm in the z-direction and within 2 cm in the x-y-direction from the interaction point.

As described above more than one PV candidate in the event is expected. For the analyzed data that have been collected in 2015, the PV candidate with the highest sum of squared transverse momenta $\sum_i p^2_{\text{T,i}}$ associated to the tracks is taken to be the primary interaction vertex. For analyses with data collected in 2016, the PV candidate with highest $\sum_i p^2_{\text{T,i}}$ of the associated track-level physics objects is selected as the primary interaction vertex. The physics objects are obtained by a jet finding algorithm [105,106]. This algorithm takes all charged tracks associated with the PV candidate and their corresponding associated missing transverse momentum into account. Charged particles associated with other PVs are not considered.

5.3 Reconstruction of Muons

At the CMS experiment, muons are reconstructed from tracks in the tracking system (tracker muon) and from tracks in the muon system (standalone muon) [107, 108]. Thus, two reconstruction approaches for muons are used. *Global muons* are recon-

structed by matching the track of the standalone muon to a track from the tracker. A Kalman-filter technique [109] is used to combine the hits originating from the tracker track and the standalone muon track. For the reconstruction of *tracker muons* all tracks with a transverse momentum of at least 0.5 GeV are extrapolated to the muon system. In this procedure, the magnetic field and the expected energy losses are taken into account. A tracker muon is identified, if at least one hit in a segment in the muon chambers can be associated with the extrapolated track.

In this analysis, the muon candidates are required to be identified as global muons and have a transverse momentum of $p_{\mathrm{T}}^{\mu} > 30$ GeV and a pseudorapidity of $|\eta^{\mu}| < 2.4$. Moreover, the muon candidates have to fulfill certain identification criteria. For this purpose, the CMS experiment has developed different working points, where tighter working points lead to low misidentifaction rates but also to a lower efficiency. The selected muons in this analysis are identified with the *tight* working point. Tight muons have to pass the following requirements [107]:

- the muon candidate has been identified by the PF algorithm,
- a cut on the discriminator of the global muon track fit of $\chi^2/n_{dof} < 10$ is applied,
- at least one hit in the muon chamber is included in the global muon track fit,
- the muon candidate provides hits in at least two muon stations,
- the track in the tracker system has to have a transverse impact parameter of $d_{xy} < 2\,\mathrm{mm}$ with respect to the primary vertex,
- the longitudinal distance of the tracker track with respect to the primary vertex is $d_z < 5\,\mathrm{mm}$,
- at least one pixel hit is produced by the muon candidate, and
- hits in at least six different tracker layers are required.

In addition to these criteria, only isolated muons are selected. The relative isolation gives a measure of additional activity within the cone of $\Delta R < 0.4$ around the muon candidate. All particle-flow charged hadrons ($h^{\pm}$), neutral hadrons (h^{0}) and

photons (γ) are taken into account. The relative isolation is defined as

$$I_{\text{rel}} = \frac{\sum_{h^\pm}^{\text{PV}} p_{\text{T}} + max(0, \sum_\gamma p_{\text{T}} + \sum_{h^0} p_{\text{T}} - 0.5 \sum_{h^\pm}^{\text{PU}} p_{\text{T}})}{p_{\text{T}}^{\mu\ \text{cand.}}}. \tag{5.2}$$

In the first term of the numerator, only the hadrons $h^\pm$ from the leading PV are taken into account. Moreover, $\sum_{h^\pm}^{\text{PU}}$ denotes the charged hadrons that originate from a different primary vertex than the muon candidate and is a measure of neutral pileup. The relative isolation criterion for tight muon candidates is fulfilled, if the obtained value of I_μ is smaller than 0.15.

5.4 Reconstruction of Electrons

Information from tracks in the tracker system and the energy deposition in the ECAL is used for the reconstruction of electrons [108]. Two complementary algorithms are used: an ECAL-driven algorithm and a tracker-driven algorithm.

The ECAL-driven reconstruction forms *super-clusters*, which are built by combining the energy depositions within $|\Delta\eta| \leq 0.9$ and ± 0.3 rad in the azimuthal direction ϕ. To reconstruct the position of the electron track, changes in curvature due to bremsstrahlung emission is taken into account by a fitting procedure. The ECAL-driven approach is particularly well suited to reconstruct isolated electrons with high transverse momenta.

The tracker-driven approach is favored for the reconstruction of low p_{T} electrons that are not isolated. If the bremsstrahlung emission is negligible, the momentum of the electron candidate can be determined with good precision by using the Kalman-filter technique and the track can be reconstructed up to the ECAL by matching with the closest cluster. In case the energy loss due to bremsstrahlung emission is not negligible, track properties, for example the number of hits and the χ^2 of the fitted track, are used in a multivariate analysis for the reconstruction of the electron candidate. This re-fitting procedure is performed with a Gaussian sum filter algorithm [110].

Similar to the muon reconstruction, different working points for the identification of electrons have been developed by the CMS experiment. The selected electrons in this analysis are identified with the *medium* working point and have to fulfill the

following criteria [111]:

- A cluster-shape variable defined as $\sigma_{\eta\eta} = \sqrt{\frac{\sum(\eta_i - \bar{\eta})^2 w_i}{\sum w_i}}$ is taken into account and has to be smaller than a certain value depending on η, p_{T} and φ,
- certain criteria are applied on the variables $|\Delta\eta_{in}|$ and $|\Delta\phi_{in}|$; these variables measure the matching between the super-cluster and the extrapolated track,
- the ratio between the energy in the HCAL cluster directly behind the ECAL cluster and the energy in the ECAL cluster H/E (hadronic leakage) has to be smaller than a certain value,
- a criterion is applied to the relative isolation of the electron candidate,
- to validate the consistency of the electron candidate between the energy measured in the ECAL and the momentum measured in the tracker the variable $(|\frac{1}{E} - \frac{1}{p}|)$ has to pass certain criteria,
- the number of expected missing hits in different layers of the barrel is not allowed to be greater than 1,
- the electron candidates have to pass a conversion veto, which means that electrons are removed if they originate from photon conversion.

Furthermore, the electrons used in this analysis are required to have a transverse momentum of $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and a pseudorapidity of $|\eta| < 2.4$.

5.5 Reconstruction of Tau Leptons

The tau (τ) lepton is the heaviest of the charged leptons with a mass of $m_\tau = 1.78\,\mathrm{GeV}$ [6]. Its lifetime is $2.91 \times 10^{-13}\,\mathrm{s}$ [6] and several decay modes are possible. It is the only lepton that is heavy enough to decay into hadrons, which happens in approximately 65% of all cases. In Tab. 5.1, the most dominant decay modes of the τ lepton are summarized. More details can be found in Ref. [6].

The possible hadronic τ lepton (τ_{h}) decay modes are grouped into two decay modes, called *one-prong decay* if it contains one charged hadron and *three-prong decay* if it contains three charged hadrons. One or two neutral pions can also be involved in τ_{h} lepton decays.

The actual reconstruction of such decay modes is realized by the Hadron-Plus-Strip (HPS) algorithm [112–114]. Due to the high probability of the decay of the neutral pion into two photons ($\pi^0 \to \gamma\gamma$), the reconstruction of photons in strips is particularly important. The strips take all objects into account that are built out of electromagnetic particles (photons and electrons).

The one-prong decay mode without any strips includes $\tau \to \pi^{\pm}\nu$ decay modes, or $\tau \to \pi^{\pm}\pi^0\nu$ decay modes with π^0 at very low energy. Moreover, the one-prong decay mode $\tau \to \pi^{\pm}\pi^0\nu$ with more energetic neutral pions is considered by the HPS algorithm. In that case, one strip is reconstructed if both photons from the π^0 decay are collimated. Two strips are reconstructed if the two photons are well separated. Similarly, the reconstruction of $\tau^- \to h^-\pi^0\pi^0\nu_\tau$ decays is possible. The three-prong decay mode includes $\tau \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu$ decays, which occurs in an $a_1(1260)$ resonance.

The HPS algorithm is seeded by the reconstructed jets (see Sec. 5.6) and searches for τ_h lepton decay products of any of the hadronic decay modes listed in Tab. 5.1. After the jets are seeded, the HPS algorithm first attempts to center a strip on the most energetic electromagnetic particle within the jet. Within the window of 0.05×0.20 in the $\eta - \phi$ plane the algorithm searches for other electromagnetic particles. Once the most energetic particle was found and associated with the strip, the strip four-momentum is re-calculated and the procedure is repeated until the list of all electrons and photons associated with the strip is exhausted. Finally, the strips are combined with the charged hadrons to build the hadronically decaying τ_h lepton candidate in case the strips fulfill the transverse momentum requirement of $p_\mathrm{T}^\mathrm{strip} > 1\,\mathrm{GeV}$.

Due to the overwhelming production of jets at the LHC the jet-to-tau misidentification rate needs to be kept as low as possible. The isolation of the τ_h lepton candidate is the main tool to reduce such contamination. It is a measure of the activity around the cone of the τ_h lepton candidate and results in high values for jets and low values for τ_h leptons. In this thesis, the *cut-based* discriminator is used for the isolation. It is defined as

$$I_\tau = \sum p_\mathrm{T}^{charged} + \max(0, \sum p_\mathrm{T}^\gamma - \Delta\beta) \tag{5.3}$$

and includes the pile-up corrected sum of transverse momenta of all PF charged hadrons with $p_\mathrm{T} > 0.5\,\mathrm{GeV}$ and photons with $E_\mathrm{T} > 0.5\,\mathrm{GeV}$. The correction $\Delta\beta$

Decay mode	resonance	BR [%]
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \bar{\nu}_e \nu_\tau$		17.4
$\tau^- \to h^- \nu_\tau$ (1-prong)	$\pi^-(140)$	11.5
$\tau^- \to h^- \pi^0 \nu_\tau$ (1-prong)	$\rho(770)$	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$ (1-prong)	$a_1(1260)$	10.8
$\tau^- \to h^- h^+ h^- \nu_\tau$ (3-prong)	$a_1(1260)$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other hadronic decay modes		1.8
Hadronic modes		64.8

Table 5.1: Decay modes and branching ratios of the τ lepton and their corresponding resonances in case of hadronic τ lepton decay modes [6]. Charged hadrons are denoted by $h^\pm$, which typically correspond to pions or kaons.

compensates the effect of pileup on photon isolation and was determined to be 0.2 GeV by the CMS collaboration [114]. Various working points are available, corresponding to an upper threshold of the isolation I_τ. In this thesis, the medium working point is used. This criterion is fulfilled if the obtained value of the isolation of the τ_h lepton candidate is less than 1.5 GeV.

Furthermore, electrons and muons can be misidentified as τ leptons. Electrons can be misidentified as a 1-prong τ_h lepton, since, for example, both an electron and a charged pion can be associated to a track and calorimetric deposits. Moreover, electrons can emit bremsstrahlung and the emerging photon could be identified as originating from a π^0. In order to reject electrons an MVA discriminator is used. The information used by this discriminator is the amount of energy measured in the HCAL associated with the τ_h lepton candidate. The MVA discriminator also considers the amount of electromagnetic energy in a narrow strip around the leading track with respect to the total electromagnetic energy of the τ_h lepton candidate.

To avoid overlap double-counting between jets and τ_h leptons a process called *jet-tau-cleaning* is applied. Each jet (that fulfills the reconstruction criteria described in Sec. 5.6) is removed from the jet-list if the jet overlaps with a τ_h lepton candidate within a cone of the radius of $\Delta R = 0.4$.

5.6 Reconstruction of Jets

Due to the color confinement quarks and gluons can not be observed directly in the detector. Instead of measuring these fundamental particles in the detectors, hadronization processes produce a bunch of collimated hadrons, so-called jets. The main constituents of jets are charged hadrons, photons and neutral hadrons with an average makeup of 65 %, 25 % and 10 %, respectively. These constituents are reconstructed by the PF algorithm and used in specific algorithms to cluster them into jets. These algorithms start with a list of four momentum objects, e.g. the four momenta of PF particles or generator particles can be used as input.

Jet algorithms are required to fulfill two criteria: infrared and collinear (IRC) safety. The IRC safety requirement ensures that the modification of an event by collinear splitting or soft gluon emission does not affect the jet. Because perturbative QCD calculations diverge in the presence of such cases, theory and experiment can only be compared using IRC safe jet algorithms. The CMS experiment exploits several IRC safe jet clustering algorithms, including the k_T [115,116], Cambridge/Aachen [117], and anti-k_T [105] algorithms.

In this thesis, the anti-k_T algorithm, which belongs to a broad class of recombination jet algorithms, is used to cluster jets. The anti-k_T algorithm exploits a distance parameter d_{ij} between a pair of objects i and j and a distance parameter d_{iB} between object i and the beam axis. They are defined as

$$d_{ij} = \min(\frac{1}{k_{T,i}^2}, \frac{1}{k_{T,j}^2})\frac{\Delta R_{ij}^2}{R^2}, \tag{5.4a}$$

$$d_{iB} = \frac{1}{k_{T,i}^2}, \tag{5.4b}$$

where k_T denotes the transverse momentum of the respective object. The parameter ΔR_{ij} denotes the distance between the objects i and j in the rapidity-azimuth space and is defined as $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$. The parameter R defines the size of jets, where $R = 0.4$ is chosen in this analysis. The jets used in the analysis are henceforth referred to as AK4 jets.

The distance variables are utilized in an iterative procedure. First, the pair with the smallest d_{ij} between the object pair is determined. In case all d_{ij} are larger than

d_{iB} the object i is considered as a jet and removed from the list of objects. Otherwise, the pair of objects i and j with smallest distance variable d_{ij} is combined. This procedure is repeated until the list of objects is empty.

The CMS collaboration derives correction factors to correct the jet energy scale. These corrections factors are necessary as the measured jet energy differs from the original jet energy, due to various detector effects such as noise and misreconstructions. The calibration of jets will be discussed seperately in Sec. 5.7 and chapter 6.

Finally, jets have to fulfill certain identification requirements. The loose working point is chosen, which means that jets have to pass the following criteria:

- the fraction of neutral hadrons and neutral electromagnetic fraction has to be less than 0.99,
- the jet consists of at least two constituents.

In addition, jets within the pseudorapidity range of $-2.4 \leq \eta \leq 2.4$ have to fulfill further conditions:

- the charged energy hadron fraction as well as the charged multiplicity has to be greater than 0,
- the charged electromagnetic energy fraction of the jet has to be smaller than 0.99.

For the estimation of residual correction factors as presented in chapter 6 these identification criteria are not necessary and are therefore not applied. In the search for leptoquarks presented in chapter 7 jets have to fulfill the identification criteria. In addition, only jets with transverse momenta of at least 30 GeV within the pseudorapidity range of $|\eta| < 2.4$ are exploited in that search.

Identification of b Jets

As the final states in the model considered in this thesis naturally contain b quarks, it is of great importance to identify the jets originating from the hadronization of bottom quarks, in the following denoted as b jets.

To identify b jets, a b tagging algorithm [119] developed by the CMS experiment is used. The algorithm makes use of the properties of the production and the decay of b

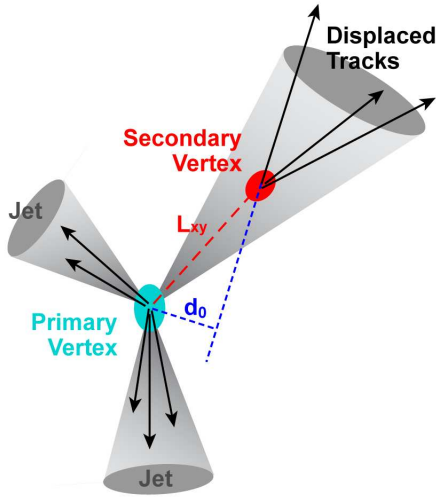

Figure 5.1: Reconstruction of a secondary vertex. The parameter L_{xy} denotes the distance between the primary vertex and the secondary vertex in the plane orthogonal to the beam direction. The impact parameter of a track with respect to the primary vertex is marked as d_0. Taken from Ref. [118].

hadrons. The most important property is the lifetime of b hadrons of approximately 1.5 ps, which is large compared to the lifetime of other particles. The relatively large lifetime results in a flight distance of a few cm and can be observed with the tracking detectors. Secondary vertices produced by the decay of b hadrons can be distinguished from primary vertices. The reconstruction of a secondary vertex is illustrated in Fig. 5.1. Other properties to identify b jets is the large mass of the b hadrons, the large multiplicity of charged particles from b hadron decays, and the fact that the b hadron in a b jet carries a large fraction of the jet energy.

In this analysis, the Combined Secondary Vertex (CSVv2) algorithm is exploited. It combines track-based variables as well as secondary vertex variables. The latter include the properties specified above. The mass of the secondary vertex candidate is required to be incompatible with the mass of the K^0 meson and has to be smaller than 6.5 GeV. Moreover, secondary vertices are not allowed to share more than 65 % of their associated tracks with the primary vertex. In case no secondary vertex is found, the CSVv2 algorithm uses track observables only. The variables related to the track-based approach are the number of tracks in the jet and the impact parameter significance, which is defined as the ratio of the impact parameter to its estimated

uncertainty.

Based on these variables, the discriminating CSVv2 variable is constructed to distinguish between b jets and jets arising from light-flavour quarks or gluons. In this analysis, the medium working point is used, which corresponds to a requirement on the discriminator such that the efficiency for identifying b jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is approximately 70% and the misidentification probability for light-flavor jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is 1 %.

5.7 Jet Energy Corrections at CMS

The aim of the jet energy corrections at the CMS experiment is to calibrate jets in order to have the correct energy scale with respect to the particle level jets clustered from stable and visible final state particles. Such corrections are necessary to correct for various effects like additional energy due to pileup, detector effects, or electronic noise. A factorized approach was applied by the CMS collaboration using several independent correction levels to correct for different effects. In Fig. 5.2 the sequence of the factorized approach is shown. After a pileup correction, non-linearity in p_{T} and non-uniformity in η, residual differences between data and simulation as a function of η and p_{T} are determined. The data-driven response is studied in channels of dijet, Z+jet, γ+jet, and multijet events. The optional jet-flavor corrections have not yet been considered at $\sqrt{s} = 13\,\mathrm{TeV}$.

Jet energy corrections are of essential importance for many physics analyses where jets are exploited. Moreover, systematic uncertainties arising from jet energy corrections can have a leading impact in physics analyses.

In the following, the consecutive steps of jet energy corrections are discussed. Distributions that illustrate the performance of the data taking period in 2015 and simulation can be found in Ref. [120]. Previous results of the jet energy corrections at $\sqrt{s} = 7\,\mathrm{TeV}$ and $\sqrt{s} = 8\,\mathrm{TeV}$ are published in Ref. [121] and [122], respectively.

5.7.1 L1 - Pileup Corrections

Multiple proton-proton collisions within each beam crossing happen due to the high instantaneous luminosity at the LHC. These additional proton-proton collisions at the same time as an event of interest are referred to as in-time pileup. The in-time pileup can be reduced by a method called charged hadron subtraction (CHS). PF jets with charge hadron subtraction are the most common jets used at CMS.

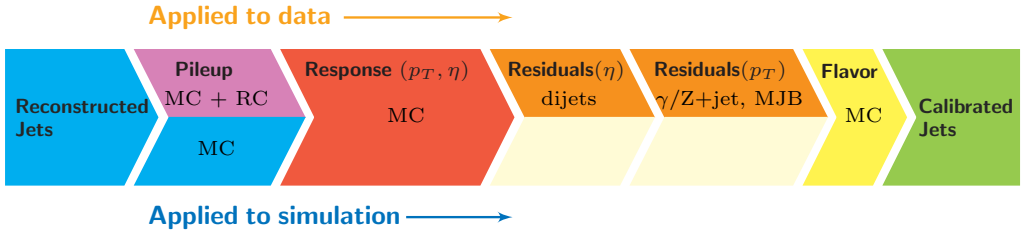

Figure 5.2: Schedule of the factorized approach of jet energy corrections at the CMS experiment. Corrections labeled with MC are derived from simulation. RC refers to random cone and MJB refers to the analysis using multijet events. Taken from Ref. [122].

A reduced list of PF candidates is employed to cluster PF+CHS jets by removing charge hadrons, which are connected to pileup vertices. A second effect in which the energy measured in previous or subsequent collisions overlaps with the energy deposition in the calorimeter from the event of interest is referred to as out-of-time pileup.

Pileup effects leading to additional contributions to the jet energy and to the jet momentum are referred to as the pileup offset. The offset is estimated with dependence on p_{T}, η, the jet area A, and the p_{T} density ρ by using the hybrid area method [123]. The area of a jet denotes the region in the (η, ϕ)-plane in which additional, infinitely soft particles (referred to as *ghost-particles*) are caught. The density ρ is defined as the median of the jet momentum divided by the jet area.

The offset in simulation is determined by the calculation of the pileup induced offset in the reconstructed jet p_{T}. For this purpose, the same event is reconstructed with and without pileup simulation and the reconstructed jets of the two samples are matched.

A scale factor between data and simulation is used to ensure that the pileup modeling in the simulation matches with the one in data. The scale factor is determined by a random cone method [121]. It is estimated from zero-bias data that does not contain any energy deposits from hard interactions. The average of the sum of the transverse momentum of PF candidates in random cones is used to estimate the offset due to pileup. Thus, the correction is determined by the ratio of the average pileup offset in data and the average pileup offset in simulation.

5.7.2 L2L3 - Response Corrections from Simulation

The simulated response corrections are derived to take several detector effects into account. After applying these corrections, the mean energy of reconstructed jets is the same as the mean energy of generated jets. As the name *L2L3* implies, the simulated response corrections are determined and applied for jets that are already corrected for the pileup as described in the previous section. The simulated jets from QCD multijet simulation used to derive these corrections are generated with PYTHIA [81] with the CUET Monash tune [124], and are passed through the full GEANT4 [100] detector simulation.

A reconstructed jet is matched to the closest generated jet. A high matching efficiency is achieved by matching reconstructed jets to generated jets within the half of a jet cone size. In this way, the simulated jet response is determined as the ratio of the transverse momentum of the reconstructed matched jet and the transverse momentum of the generated jet.

Utilizing simulation to derive the jet response provides several advantages. The simulation can cover regions of phase space, which lack sufficient events in data, for example at very low and at very high transverse momenta, and at low and at high pileup. Moreover, the simulation is not affected by biases that occur in data-based methods, e.g. due to the detector resolution.

5.7.3 L2L3Res - Residual Corrections for Data

Residual correction factors are determined after the jets are corrected for pileup and for simulated response. They are used to correct for differences in the mean response between data and simulation. First, the residual corrections are determined with a sample of dijet events, referred to as *L2Res corrections*. For this purpose, the jets are corrected relative to a jet with $|\eta| < 1.3$ since such a jet is detected within the well-understood barrel-region of the detector. The employed techniques, several studies, and the final results of the L2Res corrections are presented separately in Chapter 6.

Secondly, the response of jets is corrected by the combination of the channels with $Z(\rightarrow \mu\mu)$+jet, $Z(\rightarrow ee)$+jet, γ+jet, and multijet events in a transverse momentum range of approximately 30 GeV to more than 1 TeV. This is referred to as *L3Res corrections*.

To derive the L2Res and L3Res correction factors the p_{T} balance and the missing

transverse energy projection fraction (MPF) methods are used [122]. Both methods are briefly covered in this section, a detailed introduction is given in chapter 6.

For the p_{T} balance method the jet response is determined by comparing the momenta of the reconstructed probe jet $p_{\mathrm{T}}^{\mathrm{probe}}$ and the reference object $p_{\mathrm{T}}^{\mathrm{ref}}$. Depending on the selected channel, the reference object is either a Z boson, a photon γ, or the hadronic recoil of the multijet system. The response estimator is defined as

$$\mathcal{R}_{balance} = \frac{p_{\mathrm{T}}^{\mathrm{probe}}}{p_{\mathrm{T}}^{\mathrm{ref}}}. \tag{5.5}$$

The idea of the MPF method is that, at parton level, the reference object and the total hadronic recoil are perfectly balanced in the transverse plane. The response estimator is defined as

$$\mathcal{R}_{MPF} = 1 + \frac{\vec{\not{E}}_{\mathrm{T}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{ref}}}{(\vec{p}_{\mathrm{T}}^{\,\mathrm{ref}})^2}. \tag{5.6}$$

For each channel of the L3Res corrections, the resulting correction factors are input to a global fit in which the absolute jet p_{T} scale is fitted. The result of the global fit is shown in Fig. 5.3. The data to simulation ratio of the jet response is shown after the global fit is performed. All values of the response are found to be close to 1 and the uncertainties are smaller than 3%.

5.7.4 Jet Energy Resolution

Compared to the resolution of other objects, e.g. leptons, the resolution of the transverse momenta of jets is found to be inferior. Mismeasurements of the jet momenta for reasons of the limited detector resolution lead to a momentum imbalance in the event. Thus, the jet energy resolution (JER) needs to be measured in data and simulated events to determine correction factors. These correction factors are calculated in dijet events by exploiting the p_{T} balance method. The JER is measured after applying the jet energy corrections and the final correction factors are provided in bins of $|\eta|$.

The final correction factors of JER at 7 TeV, 8 TeV, and 13 TeV are shown in Fig. 5.4. Independent of the run period and center-of-mass energy, the resolution was found to be worse in data than in simulation. For the 2015 run period at 13 TeV, the difference between data and simulation is approximately 10 % in the barrel region and more than 20 % for larger values of $|\eta|$.

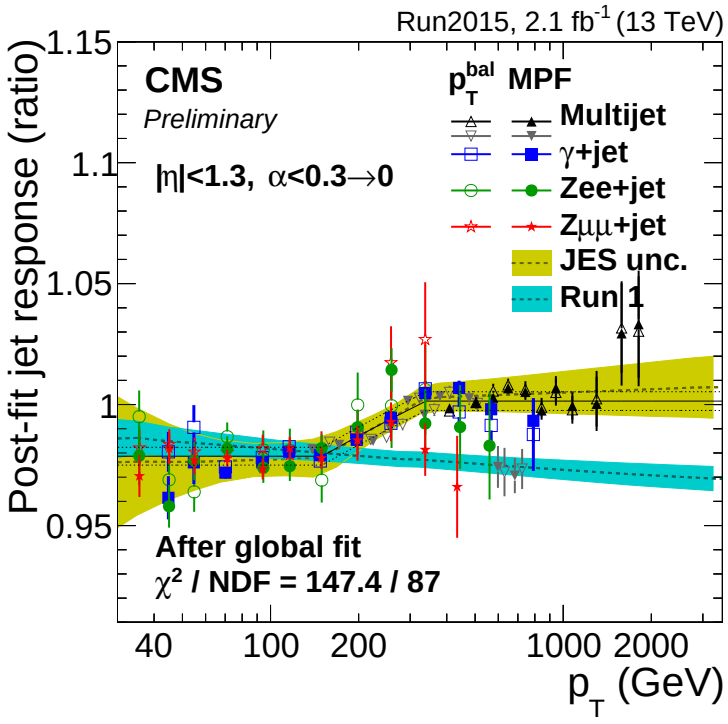

Figure 5.3: Results of the relative response in the Z($\rightarrow \mu\mu$)+jet, Z($\rightarrow ee$)+jet, γ+jet and multijet channels for the p_T balance and the MPF method. The result of the global fit is shown for the MPF method. The shaded area corresponds to the total uncertainty. As a comparison, the results from Run 1 are also shown. Taken from Ref. [120].

5.8 Missing Transverse Energy and Definition of the S_T Variable

Neutrinos pass through all detector components without leaving signatures. However, they can be detected indirectly by using the momentum imbalance that is induced in the transverse plane of the beam axis due to presence of neutrinos. This imbalance is called missing transverse energy and is defined as the negative vectorial sum of all visible PF-particles that pass a certain p_T threshold:

$$\not{E}_T = |\vec{\not{E}}_T| = \left| -\sum_i \vec{p}_{T,i} \right| . \tag{5.7}$$

To account for mismeasurements, the so-called type-1-corrected definition of the missing transverse energy [125] is employed. This means that the non-calibrated transverse momenta of PF-jets included in Eq. 5.7 are replaced with the transverse

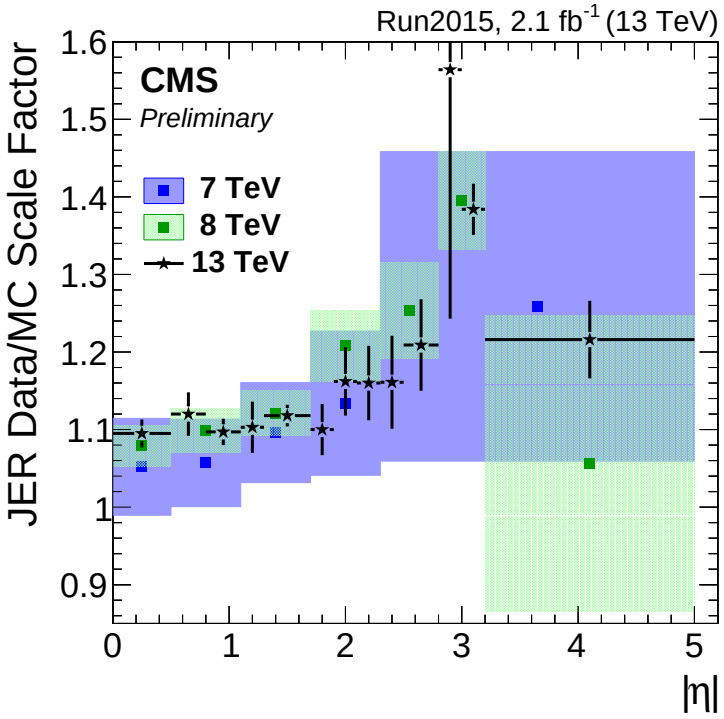

Figure 5.4: Corrections for the jet energy resolution at CMS in bins of $|\eta|$ at 7 TeV, 8 TeV, and 13 TeV. Taken from Ref. [120].

momenta of calibrated PF-jets.

Missing transverse energy can also be induced by mismeasurements of jets and hadronic recoil. In chapter 6 it is exploited in a sophisticated method to determine residual jet energy corrections. In the subsequent search for leptoquarks the missing transverse energy is primarily a measure of the vectorial transverse momentum of neutrinos arising from τ lepton decays, and leptonic W boson decays and is used as an input in the event selection.

Another relevant variable in the search for leptoquarks is the so-called S_T variable. It is defined as the scalar sum of the missing transverse energy and the transverse momenta of all electron, muons, τ leptons and jets in the event:

$$S_\mathrm{T} = \not{E}_\mathrm{T} + \sum_{\text{leptons,jets}} p_\mathrm{T}. \tag{5.8}$$

The S_T variable is a measure of the mass scale of the event produced in the collision. Therefore, it provides discrimination power between the Standard Model processes and the processes of leptoquarks of different mass hypotheses, and is also used as

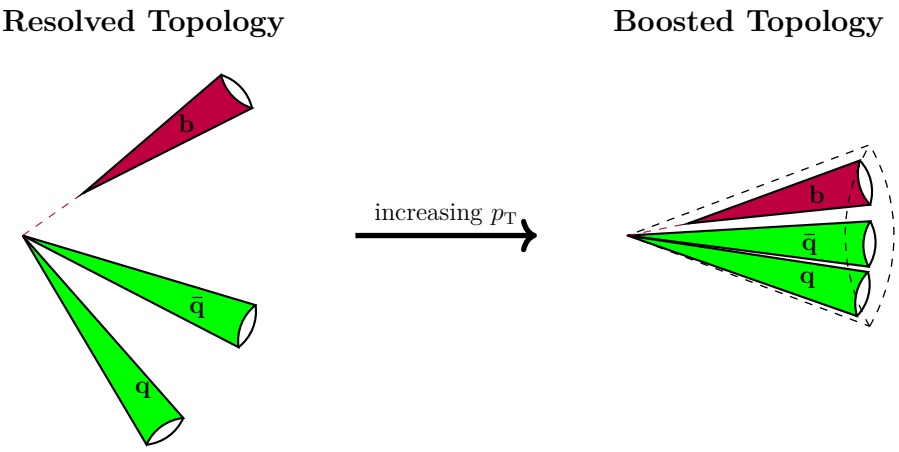

Figure 5.5: The decay of top quarks shown in the resolved topology (left) and in the boosted topology (right).

an input in the event selection.

5.9 Reconstruction of Boosted Top Quarks

The LHC can produce unstable particles with high kinematic energies. For such a particle with a high transverse momentum, its decay products are collimated in the direction of flight of the particle. The maximum distance between the decay products can be approximated by:

$$\Delta R = \frac{2m}{p_{\mathrm{T}}}, \tag{5.9}$$

where m is the mass and p_{T} the transverse momentum of the decaying particle.

In case of a top quark decaying hadronically with $p_{\mathrm{T}}^{\mathrm{top}} << m_{\mathrm{T}}^{\mathrm{top}}$, the decay products are found to be well resolved. However, for top quarks with high transverse momenta, the top quarks are *boosted* and the decay products of the top quark are collimated along its flight direction. In this scenario the decay products may be merged into a single jet, which has an internal substructure originating from the individual quarks from the top quark decay. In Fig. 5.5 the decay of a hadronically decaying top quark is shown in the resolved topology as well as in the boosted topology.

For the reconstruction of boosted top quarks new techniques called *top taggers* have been developed, e.g. the HEP Top Tagger [126, 127], the CMS Top Tag-

ger [126, 128], the OptimalR HEP Top Tagger [129], and the HOTVR Tagger [130].

In this analysis top quarks with both resolved and boosted topologies are taken into account. To consider both topologies simultaneously, none of the top tagging algorithms are applied as they would reduce the reconstruction efficiency of resolved top quark decays. Instead, hadronically decaying top quarks are reconstructed by assigning the permutation of AK4 jets in the event with invariant mass closest to the top quark mass. More details about this procedure can be found in Sec. 7.3.2.

6 Determination of Relative Residual Jet-Energy Corrections

Several consecutive stages of jet energy corrections need to be performed before jets are considered in physics analyses. Before data-driven residual correction factors can be determined, jets have been calibrated for pileup and simulated response corrections are applied as described in the factorized approach in Sec. 5.7. In principle, further corrections would not be required if the simulation described the data perfectly. However, slight differences in the jet response occur between data and simulation. To correct for these differences, residual corrections need to be derived.

In this chapter, the relative residual corrections using dijet events are presented, based on a data sample corresponding to an integrated luminosity of $2.1\,\text{fb}^{-1}$ recorded at $\sqrt{s} = 13\,\text{TeV}$ with the CMS detector. Correction factors derived with dijet events are the first step of the residual corrections. One advantage of the employment of dijet events is the huge jet production cross section that leads to results with small statistical uncertainties in this analysis. The response of jets is corrected over a wide range of the transverse momentum relative to a reference jet measured in the barrel region of the detector with pseudorapidity $|\eta| < 1.3$. Subsequently, the results of the dijet analysis are combined with the results from channels of Z+jet, γ+jet and multijet events, which measure the absolute residual scale in the barrel region.

In the following, the basic concept is described in Sec. 6.1. An introduction to the dijet balance method and the missing transverse energy projection fraction (MPF) method that are used to derive the correction scale factors is presented in Sec. 6.2 and Sec. 6.3, respectively. The data samples and the event selection can be found in Sec. 6.4. In Sec. 6.5, the definition of the residual correction is established that either uses the response from the dijet balance or from the MPF method. The final results of residual corrections of the 2015 data taking period are presented in Sec. 6.6. Stability tests and systematic uncertainties are discussed in Sec. 6.7 and a summary is given in Sec. 6.8. All studies and results presented in the following have

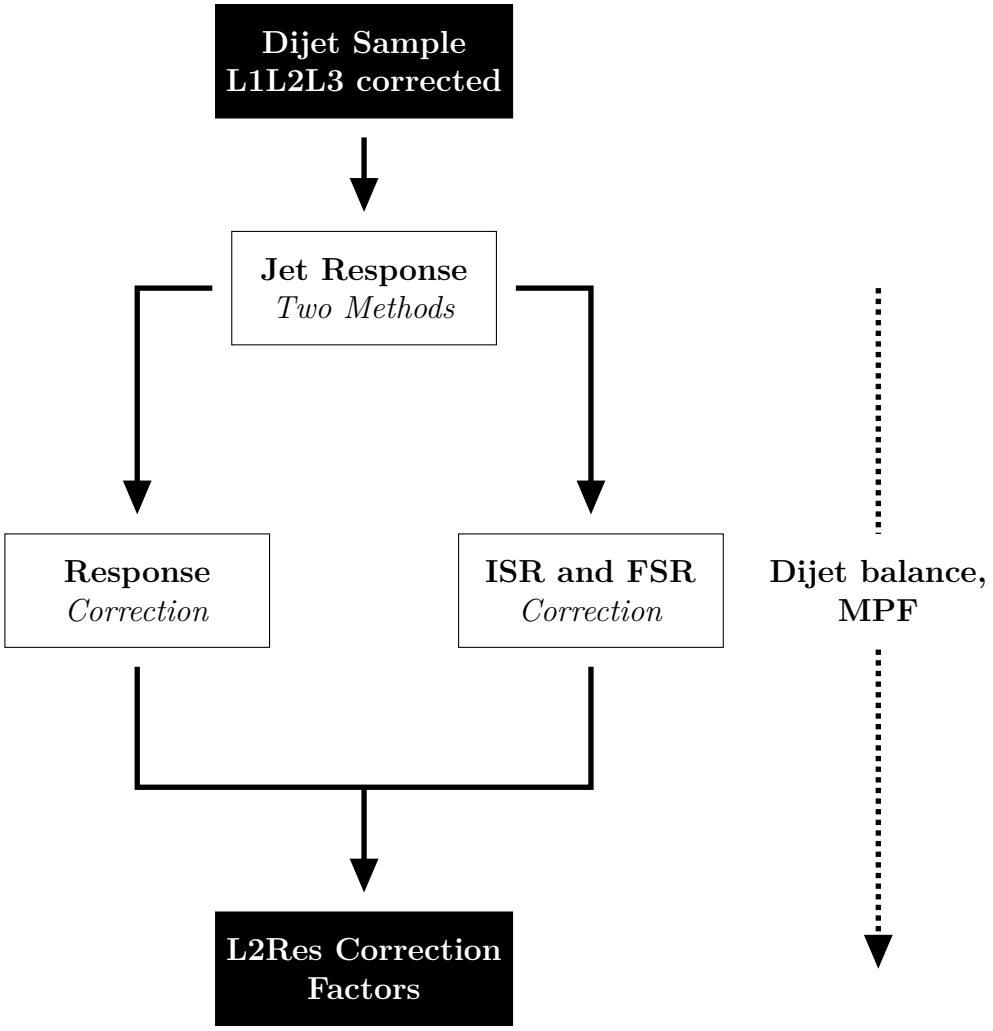

Figure 6.1: Determination of L2Res correction factors. The L1 and the MC truth corrections are applied to a dijet sample. Jet responses are calculated by using the p_{T} balance and the missing transverse energy projection fraction (MPF) methods. Jet response corrections as well as ISR and FSR corrections are considered to determine the L2Res correction factors.

been performed at the University of Hamburg by the author of this thesis together with Nataliia Kovalchuk [131]. The residual corrections obtained in this analysis have been applied to most CMS analyses analyzing data recorded in 2015. For data recorded in 2016 and 2017, the residual corrections are updated by other CMS colleagues.

Previous results of the relative jet energy corrections obtained at $\sqrt{s} = 7\,\mathrm{TeV}$ and $\sqrt{s} = 8\,\mathrm{TeV}$ are documented in Refs. [121, 132] and Refs. [122, 133], respectively.

6.1 Basic Concept

In this section, residual corrections using dijet events (so-called *L2Res corrections*) are determined. L2Res corrections are used to calculate scale factors between data and simulation. A summary of the analysis flow is presented in Fig. 6.1. For any level of residual corrections, jets corrected with simulation-based corrections as described in Sec. 5.7 are used. To derive the scale factors, a tag-and-probe approach is employed using the leading and the subleading jet. One of the two leading jets serves as reference 'tag' object (in the following referred to as barrel jet) within the well-understood region of the detector with $|\eta| < 1.3$, whereas the other (probe) jet has an unconstrained value of η. Thus, the response of all jets is corrected with respect to the response of all barrel jets. The results of the L2Res corrections are determined in bins of $|\eta|$. The absolute value of η is used to reduce statistical uncertainties in the endcap and hadron forward regions of the detector. The bin edges applied have been optimized to follow the structure of the CMS detector. In the following, values of $0 < |\eta| < 1.3$ are referred to as the barrel region, whereas values of $1.3 < |\eta| < 3.0$ are referred to as the endcap region. The hadron forward region overlaps with the endcap region as its pseudorapidity coverage is from $|\eta| = 2.9$ up to $|\eta| = 5.2$.

Two main biases can affect the measurements of the jet response, the resolution bias and the radiation bias. The resolution bias arises from the different energy resolution of the barrel jet and the probe jet. Thus, instead of using bins of the transverse momentum of the barrel jet, $p_{\mathrm{T}}^{\mathrm{barrel}}$, the average transverse momentum of the probe and the barrel jet, defined as

$$\bar{p}_{\mathrm{T}} = \frac{p_{\mathrm{T}}^{\mathrm{barrel}} + p_{\mathrm{T}}^{\mathrm{probe}}}{2}, \tag{6.1}$$

is taken. This observable minimizes the bias from the jet energy resolution that would appear in bins of $p_{\mathrm{T}}^{\mathrm{barrel}}$ as the resolution of the barrel (and the probe) jet has a nonzero width. In case the two leading jets have the same resolution, the bias cancels out on average. The radiation bias appears due to the imbalance in the transverse momentum caused by initial state radiation and final state radiation in the event and can also be reduced.

6.2 Asymmetry and Dijet Balance Method

The dijet balance method is one of two basic methods used to determine the residual correction factors. It was developed at the SP$\bar{\text{P}}$S-collider [134] and has been improved at the Tevatron experiment [135, 136]. At the CMS experiment, the method was used as the central method for the jet energy calibration in 2010 [121]. Generally, the dijet balance method exploits momentum conservation and the fact that the transverse momentum in (ideal) dijet events is balanced back-to-back in the azimuthal angle φ.

A fundamental observable for this method is the asymmetry $\mathcal{A}$, which is defined as

$$\mathcal{A} = \frac{p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{barrel}}}{p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{barrel}}}. \tag{6.2}$$

The relative response $\mathcal{R}_{\mathrm{rel}}$ of the probe jet with respect to the barrel jet is defined as

$$\mathcal{R}_{\mathrm{rel}} = \frac{1 + \langle \mathcal{A} \rangle}{1 - \langle \mathcal{A} \rangle}, \tag{6.3}$$

where $\langle \mathcal{A} \rangle$ is the average value of the asymmetry of all events. The response obtained with Eq. (6.3) is sensitive to ISR and FSR as these bias the balance of the dijet topology. In addition, the bias due to ISR and FSR effects can be corrected by a correction factor as described in Sec. 6.5.

6.3 Missing Transverse Energy Projection Fraction Method

The missing transverse energy projection fraction method has been studied comprehensively at the Tevatron experiments [135–137]. At the CMS experiment, the calibration of jets with Z+jets and γ+jet events during the data taking period in 2010 was derived with the MPF method [121]. For the data taking period in 2011 and 2012, the MPF method was also used to derive the residual correction factors with dijet events.

The basic idea of the MPF method is that dijet (as well as Z+jet and γ+jet) events have no intrinsic missing transverse energy $\vec{\not{E}}_{\mathrm{T}}$. Thus, at parton level, the reference jet, and the total hadronic activity in the event that recoils versus the reference jet (referred to as *total hadronic recoil*), should be perfectly balanced in

the transverse plane:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{barrel}} + \vec{p}_{\mathrm{T}}^{\,\mathrm{recoil}} = 0. \tag{6.4}$$

The total hadronic recoil is used as a reference object to derive the calorimeter response to jets. In case the objects could be reconstructed, Eq. (6.4) can be written as

$$\mathcal{R}_{\mathrm{barrel}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{barrel}} + \mathcal{R}_{\mathrm{recoil}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{recoil}} = -\,\vec{\not{E}}_{\mathrm{T}}, \tag{6.5}$$

where $\mathcal{R}_{\mathrm{barrel}}$ denotes the detector response of the reference jet and $\mathcal{R}_{\mathrm{recoil}}$ the hadronic recoil. Solving the equations above for $\mathcal{R}_{\mathrm{recoil}}$ leads to

$$\mathcal{R}_{\mathrm{recoil}} = \mathcal{R}_{\mathrm{barrel}} + \frac{\vec{\not{E}}_{\mathrm{T}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{barrel}}}{(\vec{p}_{\mathrm{T}}^{\,\mathrm{barrel}})^2}. \tag{6.6}$$

In the limit of no further activity apart from the dijet topology and assuming in addition $\mathcal{R}_{\mathrm{barrel}} = 1$, the final MPF response is given by

$$\mathcal{R}_{\mathrm{recoil}} = 1 + \frac{\vec{\not{E}}_{\mathrm{T}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{barrel}}}{(\vec{p}_{\mathrm{T}}^{\,\mathrm{barrel}})^2} \equiv \mathcal{R}_{\mathrm{MPF}}. \tag{6.7}$$

In the MPF method the barrel jet is balanced against the hadronic recoil. The radiation bias on the response due to transverse momentum imbalance is caused by gluon radiation in the event. The MPF response is sensitive to soft radiation only through differences in the response of the leading jet and additional jets, whereas the dijet balance response does not take additional radiation into account. Thus, the MPF method is less sensitive to gluon radiation than the dijet balance method. Nonetheless, the bias due to soft radiation is corrected for the MPF method similarly as in the dijet balance method.

The final results for L2Res corrections used by the CMS collaboration are determined with the MPF method and the results obtained with the dijet balance method serve as a cross-check.

6.4 Data Samples and Event Selection

In the following, the data samples used in this analysis, trigger studies, and the event selection will be discussed and summarized.

Data set	L [pb^{-1}]
/JetHT/Run2015C	16.3
/JetHT/Run2015D(v3)	552.7
/JetHT/Run2015D(v4)	1521.0
Total	2090.0

Table 6.1: Certified data sets and the corresponding integrated luminosity.

6.4.1 Samples

The dataset of the year 2015 recorded at a center-of-mass energy of 13 TeV and with a bunch spacing of 25 ns is analyzed. The total integrated luminosity used in this analysis is 2.1 fb^{-1}. Information about the data sets is listed in Tab. 6.1. A QCD-multijet MC sample is used. In total, this MC sample contains 10 million events generated with PYTHIA 8. The simulated events are generated with a flat p_{T} spectrum, which covers a large kinematic phase space with small statistical uncertainties.

6.4.2 Trigger Thresholds

The dijet sample is collected by dedicated dijet triggers. The events pass the trigger if the average transverse momentum $\bar{p}_{\mathrm{T}}$ is above a threshold, depending on the respective trigger. The triggers used in this analysis utilize a PF reconstruction algorithm with simplified tracking. So-called *prescale* factors are applied to each trigger, such that only every n-th triggered event is stored. Prescale factors are used to keep the event rate at the desired value in the respective region of $\bar{p}_{\mathrm{T}}$.

Only events with average transverse momentum $\bar{p}_{\mathrm{T}}$ in the region where a trigger is fully efficient are selected. To determine the trigger thresholds, the method as described in Ref. [138] is used. This approach introduces the trigger efficiency

$$\epsilon = \frac{N(\text{passing } A)}{N(\text{passing } B)}, \tag{6.8}$$

where A and B denote two successive triggers as listed in the left column of Tab. 6.2. The efficiency has been measured as a function of $\bar{p}_{\mathrm{T}}$ for each trigger. To obtain the trigger thresholds, each efficiency distribution was fitted with the smooth function

$$f(\bar{p}_{\mathrm{T}}) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\bar{p}_{\mathrm{T}} - p_0}{\sqrt{2}\cdot p_1}\right)\right), \tag{6.9}$$

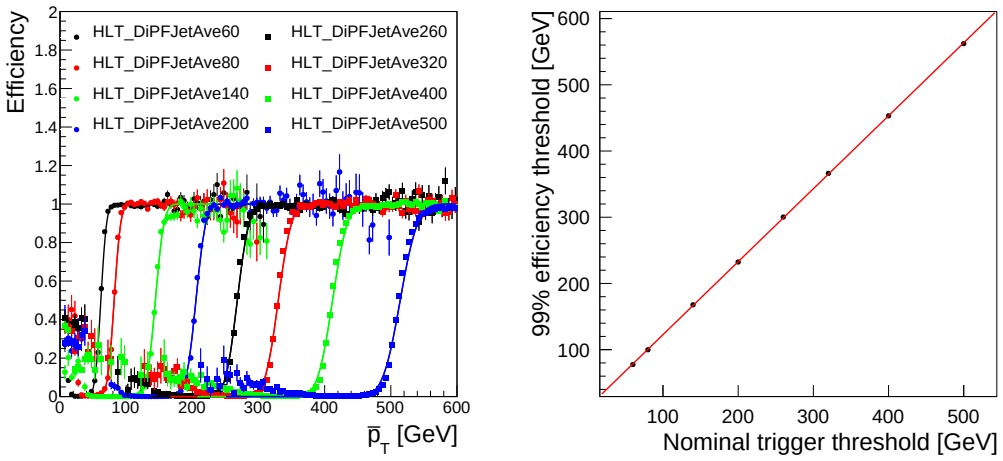

Figure 6.2: Trigger efficiency turn-on curves (left) and linear extrapolation of the threshold of the lowest dijet $\bar{p_T}$ trigger (right).

where p_0 and p_1 denote the fit parameters. For each trigger path, the trigger threshold is determined from the point at which the respective trigger reaches 99% efficiency. The corresponding $\bar{p}_T$ value is obtained from $f(\bar{p}_T) = 0.99$.

In Fig. 6.2 (left), the efficiency curves of different triggers used in this analysis are shown. Each efficiency curve has been normalized such that the plateau is at one. As different prescale factors are applied to the triggers, the plateau is not strictly at one by default and efficiencies greater than unity are possible.

For any combination of two successive triggers the threshold for the trigger with the higher value of $\bar{p}_T$ is calculated. Thus, the threshold of the lowest trigger can not be estimated by applying this method. To obtain the value of the lowest trigger a linear extrapolation of all other thresholds is used. The linear extrapolation of the thresholds can be found in Fig. 6.2 (right). The evaluated trigger thresholds are listed in Tab. 6.2.

6.4.3 Event Selection

Several selection criteria are applied in order to enrich the fraction of the events with dijet topologies. In the following, the criteria that an event has to fulfill to pass the dijet event selection are listed.

- At least two AK4 jets are required. In data, the average transverse momentum $\bar{p}_T$ has to be greater than the respective trigger threshold.

Trigger path	Threshold (GeV)
HLT_DiPFJetAve40	56
HLT_DiPFJetAve60	78
HLT_DiPFJetAve80	100
HLT_DiPFJetAve140	168
HLT_DiPFJetAve200	232
HLT_DiPFJetAve260	300
HLT_DiPFJetAve320	366
HLT_DiPFJetAve400	453
HLT_DiPFJetAve500	562

Table 6.2: Trigger thresholds of $\bar{p}_\mathrm{T}$ dijet triggers used for the determination of L2Res corrections of the data-set from 2015.

- At least one of the two leading jets serves as a reference and lies within the control region of $|\eta| < 1.3$ (barrel jet) whereas the other jet has an arbitrary value of $|\eta|$ (probe jet).
- To enrich the fraction of ideal dijet topologies, the two leading jets are required to be back-to-back. Therefore, a requirement on the azumithal angle with $\Delta\varphi(\mathrm{j}_1, \mathrm{j}_2) > 2.7$ is applied.
- In case the dijet balance is biased, large values for the asymmetry $\mathcal{A}$ can occur. Extreme asymmetry values are rejected by the requirement of $|\mathcal{A}| < 0.7$.
- The dijet topology can be biased by additional jets in the event. For this reason, a requirement on the relative third jet fraction defined as $\alpha = \frac{p_{\mathrm{T,j3}}}{\bar{p}_\mathrm{T}}$ is applied and events have to pass $\alpha < 0.2$ to be considered.

The impact of the requirements on $\Delta\varphi$ and α is shown in Fig. 6.3. It can be observed that for most of the events the leading and subleading jet end up back-to-back and with a third jet fraction α of less than 0.4. For greater values of α, the azimuth angle between the leading and subleading jet seems to scatter in any direction as the third leading jet deforms the optimal dijet topology. More control distributions are presented after applying the reweighting procedures described as follows.

6.4.4 Event Reweighting Procedures

Due to pileup effects, prescale factors in the triggers, and MC event weights, data and MC distributions are not comparable at this stage of the event selection. Thus,

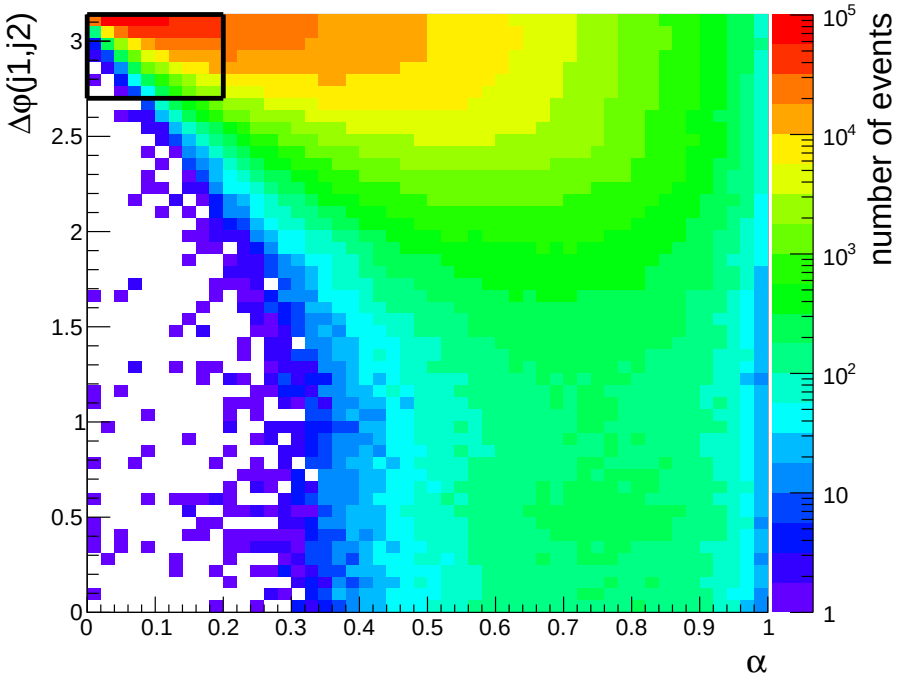

Figure 6.3: Two-dimensional distribution of the third jet fraction α and the difference of the azimuthal angle $\Delta\varphi$ between the leading and the subleading jet in simulation. The region of events that fulfill the selection is marked.

additional event weights for the simulation as described in the following are required.

Pileup Weights

The pileup distribution of the MC simulation is known from the simulation setup. The pileup distribution in data is calculated from the instantaneous luminosity of all colliding bunch pairs as well as the total inelastic cross section for proton-proton interactions, which was determined to be 69 mb by the CMS collaboration.

Both the maximum and the tail of the pileup distribution differ between data and simulation. Therefore, event weights as a function of the number of pileup events from the data to simulation ratio are determined. These event weights are applied to simulated events. The weights are normalized such that the total number of simulated events does not change. The outcome of the pileup weighting procedure is illustrated in the shape comparison of Fig. 6.4 (left). The agreement between data and simulation improves after applying pileup reweighting is applied and only slight differences can be observed. The impact of pileup on the relative response and the results of the residual correction factors is discussed in Sec. 6.7.

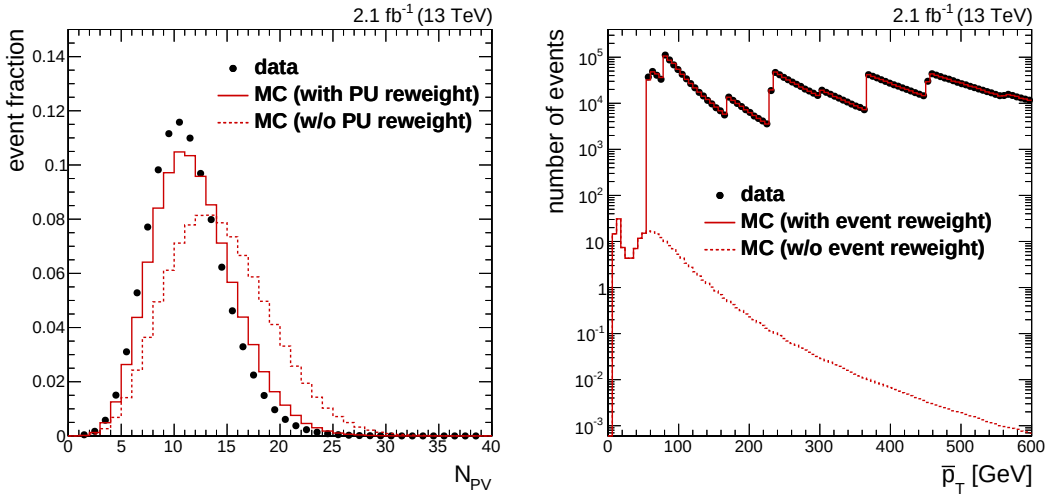

Figure 6.4: The impact of reweighting procedures. PU reweighting is illustrated in the left distribution. Simulated events with PU reweighting and without PU reweighting applied, and data events are shown as a function of the number of primary vertices N_{PV}. The distributions are normalized to their integral, respectively. In the right distribution, data events, and simulated events with event reweighting and without event reweighting are shown as a function of $\bar{p}_T$.

Dijet $\bar{p}_T$ Event Weights

Events in simulation are weighted as a function of $\bar{p}_T$ in order to match the number of events between data and MC in the respective trigger thresholds. This procedure takes the respective trigger prescale factors into account. The impact of the dijet event weighting procedure is illustrated in Fig. 6.4 (right). For simulated events without event weighting applied the nominal MC weights are used. These weights are very small, but the number of generated MC events is sufficient to apply the weighting procedure. A good agreement between data and simulation can be observed after event weighting is applied.

6.5 Determination of the Relative Residual Correction

The responses defined in Eq. (6.3) and Eq. (6.7) are evaluated in data and simulation in order to estimate the relative residual correction. Example distributions of the asymmetry and the MPF response after applying the event selection and

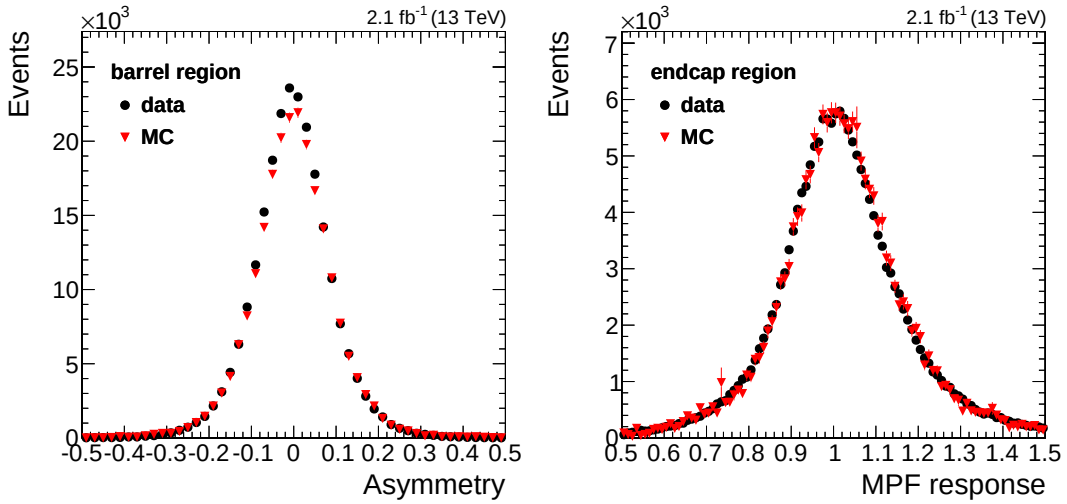

Figure 6.5: Example distributions of the asymmetry measured in the barrel region (left) and the MPF response measured in the endcap region (right).

reweighting procedures as described in the previous section are shown in Fig. 6.5. These kinds of distributions are evaluated in bins of $|\eta|$ and $\bar{p}_{\mathrm{T}}$ to derive the relative correction factors as described in the following.

In an ideal dijet topology two jets would show up exactly back-to-back in φ. However, the momentum balance of the two leading jets can be violated due to the radiation of additional jets. Thus, to enrich the fraction of the ideal dijet events, a requirement on the third jet fraction α is applied. Both ISR and FSR can bias the dijet balance. The correction factor k_{FSR}[1] is defined as follows:

$$k_{\mathrm{FSR}} = \frac{\left\langle \frac{\mathcal{R}^{\mathrm{MC}}}{\mathcal{R}^{\mathrm{data}}} \right\rangle_{\alpha \to 0}}{\left\langle \frac{\mathcal{R}^{\mathrm{MC}}}{\mathcal{R}^{\mathrm{data}}} \right\rangle_{\alpha < 0.2}}. \tag{6.10}$$

In this equation, $\mathcal{R}$ corresponds to the response either from the MPF or the dijet balance method. In order to determine the value of k_{FSR}, events with a third jet fraction up to $\alpha < 0.4$ are taken into account. In general, only events with $\alpha < 0.2$ are selected in this analysis to reduce the activity of radiation.

The extrapolation factor k_{FSR} is derived in bins of $|\eta|$ by an extrapolation of $\alpha \to 0$. Example visualizations of the k_{FSR} extrapolation using the MPF method for one bin in the barrel region and for one bin in the endcap region is shown in

[1]For reasons of brevity, the correction factor k_{FSR} is subscripted with FSR only, but ISR is also taken into account.

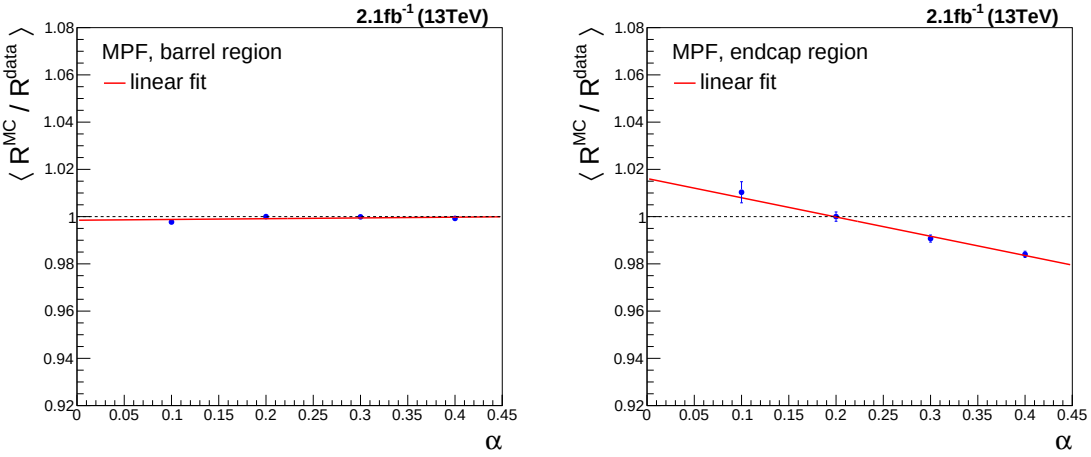

Figure 6.6: Measurements of the k_{FSR} parameter for the MPF method for one bin in the barrel (left) and for one bin in the endcap (right) region.

Fig. 6.6. A linear fit is applied to determine the extrapolation factor. The factor k_{FSR} is found to be close to 1 in the barrel region. In general, the extrapolated factor deviates from unity by less than 2% in all bins of $|\eta|$ in case the MPF method is used. As expected, the bias from radiation effects is larger for the dijet balance method. The distributions of the k_{FSR} extrapolations for each bin of $|\eta|$ for the MPF and the dijet balance method can be found in the appendix in Fig. A.1 and Fig. A.2, respectively.

The relative residual correction is generally defined as the data-to-simulation ratio of the response:

$$\mathrm{L2Res} = k_{\mathrm{FSR}} \cdot \left\langle \frac{\mathcal{R}^{\mathrm{MC}}}{\mathcal{R}^{\mathrm{data}}} \right\rangle_{\alpha<0.2} . \tag{6.11}$$

The correction factor k_{FSR} is taken into account. However, this equation is independent of the transverse momentum. To take a p_{T} dependency into account, the ratios of responses in data and simulation are evaluated in bins of the average transverse momentum. A log-linear fit to the data-to-simulation ratio of the jet response is used to determine the central results of the L2Res corrections:

$$\mathrm{L2Res}(p_{\mathrm{T}}, |\eta|) = k_{\mathrm{FSR}}(|\eta|) \cdot [k_0(|\eta|) + k_{p_{\mathrm{T}}}(|\eta|) \cdot \log(\bar{p}_{\mathrm{T}})]. \tag{6.12}$$

In this equation, k_0 and $k_{p_{\mathrm{T}}}$ are the fit parameters of the log-linear fit.

As a cross-check, a constant fit to the data to simulation ratio of the responses is

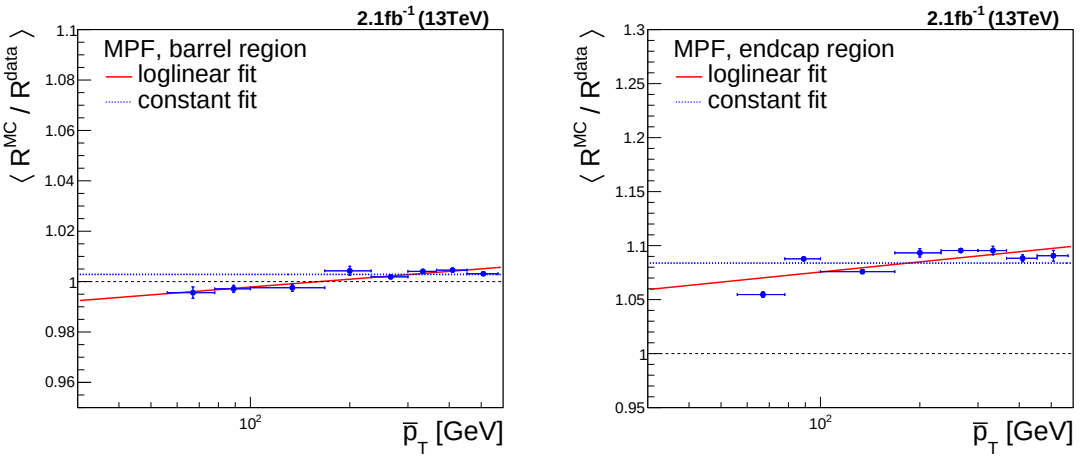

Figure 6.7: Log-linear and constant fits for the MPF method for one bin in the barrel (left) and for one bin in the endcap (right) region.

also performed. With this approach, the correction factor is defined as

$$\text{L2Res}(|\eta|) = k_{\text{const}}(|\eta|). \tag{6.13}$$

Examples of the fits in Eq. 6.12 and 6.13 for one bin in the barrel region and for one bin in the endcap region are shown in Fig. 6.7. The transverse momentum dependence is found to be almost flat in the barrel region, whereas a significant dependence is observed in the endcap region. All distributions of $\bar{p}_{\text{T}}$ dependent extrapolations for the the MPF and dijet balance method can be found in the appendix in Fig. A.3 and A.4, respectively.

6.6 L2Res Correction Results of the 2015 Data-taking Period

In this section, the final results for the L2Res corrections are presented. The studied triggers, the dijet event selection and reweighting procedures have been applied. Both methods, MPF and dijet balance, are compared to validate the results.

In Fig. 6.8, the resulting k_{FSR} and relative residual correction factors are shown. The k_{FSR} correction factors are found to be negligible small in the barrel region for the MPF method. Also, for the endcap and hadron forward region the resulting factors differ from unity by less than 2%. The k_{FSR} correction factors for the dijet

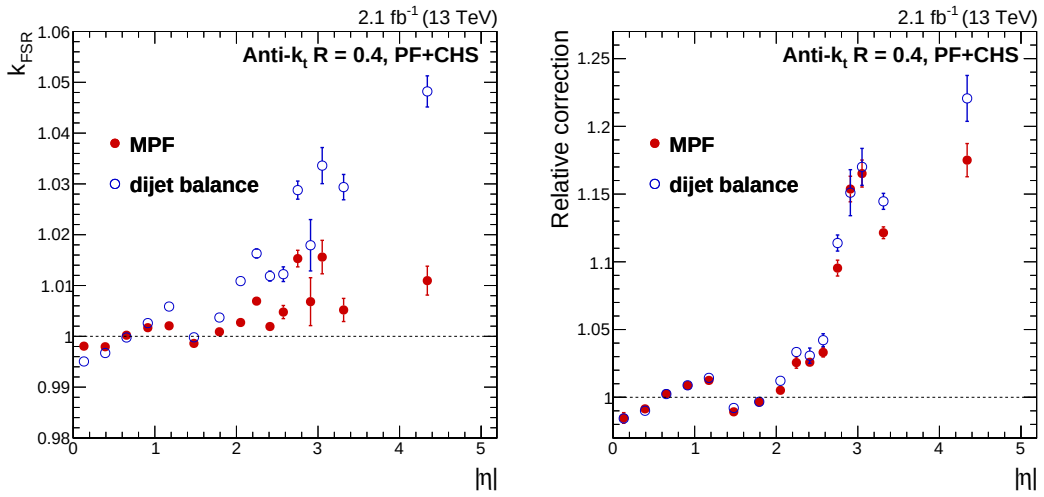

Figure 6.8: Final results for both methods, MPF in red and dijet balance in blue, of the k_{FSR} correction factors (left) and relative residual correction factors with k_{FSR} correction factors applied (right).

balance method are slightly different, with the values increasing for higher values of $|\eta|$, and vary across $|\eta|$ by up to 5%.

The resulting relative residual correction factors differ from unity by less than 3% for $|\eta| < 2.5$. The correction factors increase in the hadron forward region where they vary across $|\eta|$ by up to 17% for the MPF method and by up to 22% for the dijet balance method. This discrepancy in the hadron forward region is due to the difference in the k_{FSR} correction factor. However, the results obtained with the MPF method match very well with the ones from the dijet balance method in general and slight differences are covered by the systematic uncertainties described in the following.

6.7 Uncertainties and Stability Tests

The resulting L2Res correction factors can be biased due to uncertainties arising from several sources. Thus, both statistical and systematic uncertainties are evaluated. Systematic uncertainties are estimated by repeating the analysis procedure for each source of uncertainty and varying the respective considered quantity within one standard deviation. The resulting difference between the nominal result and the result with a variation is taken as the systematic uncertainty arising from the considered source. In the following, the sources of systematic uncertainties are described

and their impact is determined. Moreover, several stability tests are presented to validate the performance of the analysis.

Statistical Uncertainties

Statistical uncertainties on the L2Res correction factors are obtained by Gaussian error propagation of the statistical uncertainties on the fit parameters of the $\bar{p}_{\mathrm{T}}$ dependent fit and the fit on the third jet fraction α via Eq. (6.12). In total, the statistical uncertainties are found to be almost negligible in the barrel region. Also for the region of $1.3 < |\eta| < 2.7$ the statistical uncertainties remain below 0.5%. For higher values of $|\eta|$, the uncertainties are in the order of 1-2%.

Jet Energy Resolution

The jet energy resolution (JER) in data is found to be worse than in simulation. Based on this disagreement, the resolution is smeared in simulation such that the resolution in simulation matches the one in data. The smearing procedure is taken from Ref. [122]. The smeared transverse momentum of the jet is based on the difference of the transverse momentum between the reconstructed and the generated jet. The definition is given by

$$p_{\mathrm{T}}^{\mathrm{rec}} \to \max[0, p_{\mathrm{T}}^{\mathrm{gen}} + C(p_{\mathrm{T}}^{\mathrm{rec}} - p_{\mathrm{T}}^{\mathrm{gen}})], \tag{6.14}$$

where $p_{\mathrm{T}}^{\mathrm{rec}}$ denotes the transverse momentum of the reconstructed jet and $p_{\mathrm{T}}^{\mathrm{gen}}$ the transverse momentum of the generated jet. The factor C is the core resolution scale factor measured in ratios of the resolution between data and simulation for different bins of $|\eta|$. The values for C are taken from Ref. [120] and propagated to the analysis. The uncertainty on the jet energy resolution is determined by varying the scale factor C within its uncertainty. The impact of this uncertainty is shown in Fig. 6.9. The variation is insignificant for values of $|\eta| < 2.5$ and in the order of 1% (3%) for the MPF (dijet balance) method for higher values of $|\eta|$.

Dependence on Pileup

In order to derive systematic uncertainties on the pileup modeling and weights, the total inelastic cross section is varied. In previous results of correction factors in the relative response obtained at $\sqrt{s} = 8\,\mathrm{TeV}$ the systematic uncertainty due to

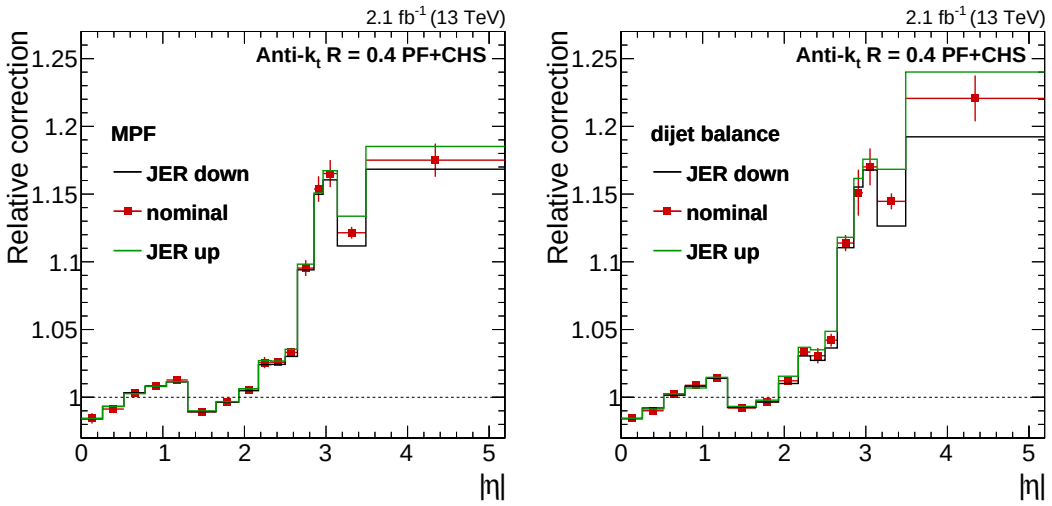

Figure 6.9: Relative L2Res correction factors with and without the variation on the jet energy resolution for the MPF (left) and dijet balance (right) method. The red lines correspond the nominal results whereas the results in black and green include the variation on JER. The difference to the nominal result is taken as a systematic uncertainty.

pileup was found to be negligible. In this analysis, a total inelastic cross section of 80 mb instead of the nominal value of 69 mb is used to verify a possible pileup dependence on the residual corrections. The impact of this variation is presented in Fig. 6.10. The uncertainty is found to be negligible in all bins of $|\eta|$ and is therefore not considered in the final results.

Moreover, the dependence of the residual corrections on the number of reconstructed vertices is validated to cross-check the simulation of pileup. This cross-check can also prove whether the previous correction steps of L1 corrections for pileup correct for the influence of pileup on the jet response. The response in data and simulation as a function of the number of primary vertices has been evaluated in bins of $|\eta|$. An example for one bin in the barrel and for one bin in the endcap region is shown in Fig. 6.11, with distributions of each bin of $|\eta|$ presented in the appendix in Fig. A.5. It is observed that the relative response is flat in data and in simulation. No dependency on the number of reconstructed vertices can be found.

Time Dependence

Over the duration of the data taking period the energy response in the ECAL and the HCAL may change, e.g. due to radiation damage in the endcaps. In order to

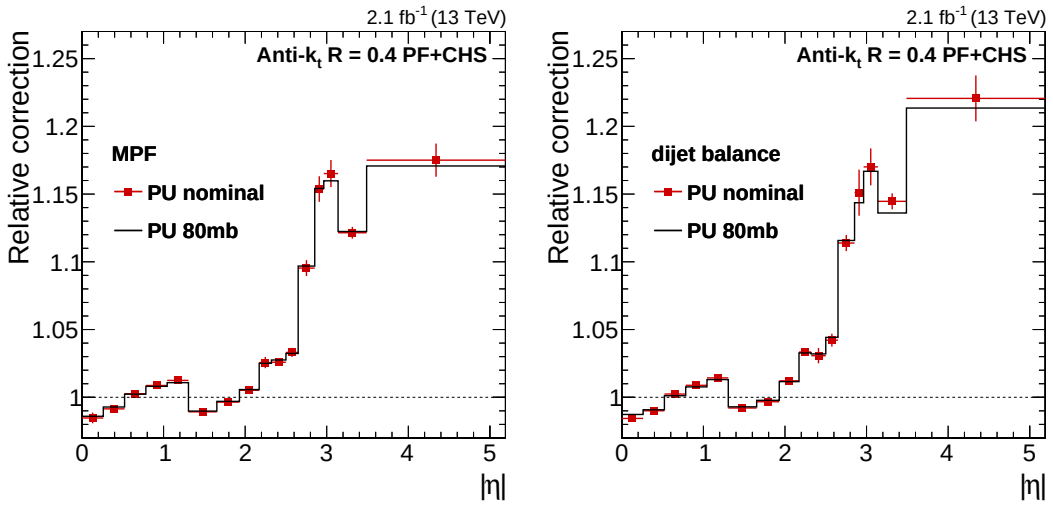

Figure 6.10: Relative L2Res correction factors with different PU profiles for the MPF (left) and dijet balance (right) method. The total inelastic cross section (red) is 69mb whereas the varied total inelastic cross section (black) is 80mb.

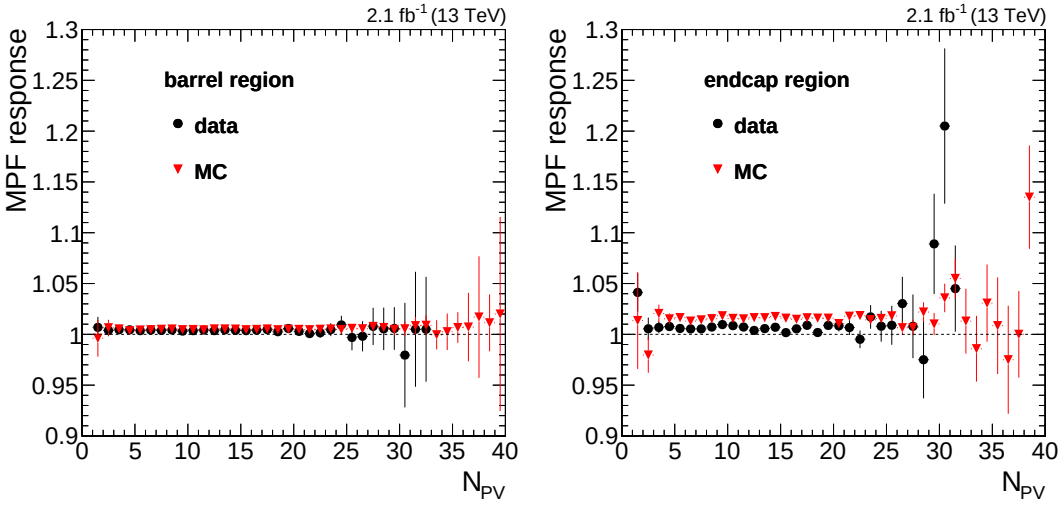

Figure 6.11: Relative response of data and simulation as a function of the number of primary vertices for one bin in the in the barrel (left) and for one bin in the endcap (right) region.

include a possible impact on the jet response, an uncertainty on the time dependence is considered by splitting the data set into multiple sets with approximately equal integrated luminosities. The divided data sets with their corresponding run numbers and integrated luminosities are summarized in the appendix in Tab. A.1. The uncertainty is then given by the root mean square of the nine obtained relative

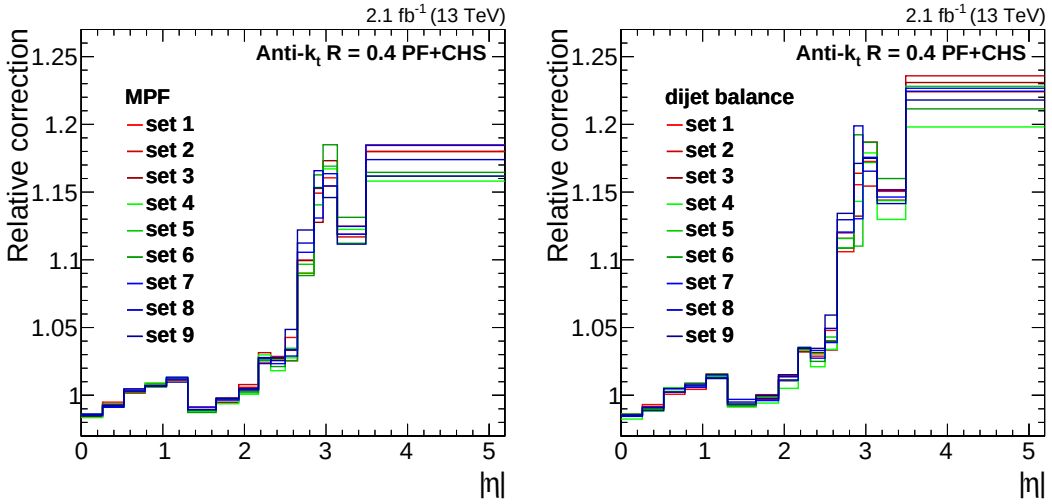

Figure 6.12: Relative residual correction factors for several data sets sorted in the order of time of the run period. The result is shown for the MPF (left) and dijet balance (right) method.

correction factors from multiple data sets. In Fig 6.12, the relative corrections for a set of nine run ranges is presented. The impact of this uncertainty turns out to be almost negligible for $|\eta| < 2.5$. For higher values of $|\eta|$ the uncertainty is in the order of 2-3%.

Transverse Momentum Dependence

The dependence of the L2Res corrections on the average transverse momentum $\bar{p}_{\mathrm{T}}$ is taken into account by Eq. 6.12. As the origin of the $\bar{p}_{\mathrm{T}}$ dependence is unknown, half of the difference between the corrections derived with the log-linear and the constant fit method (see Eq. 6.12 and 6.13) is taken as a systematic uncertainty.

Furthermore, a possible impact of the $\bar{p}_{\mathrm{T}}$ dependence is illustrated in Fig. 6.13. Here, the average transverse momentum $\bar{p}_{\mathrm{T}}$ was varied up and down by a factor of 2. The impact of the $\bar{p}_{\mathrm{T}}$ variation is quite small in the barrel and the endcap region and only on the percent-level in the hadron forward region.

Relative Balance

By exploiting the MPF and dijet balance methods, two complementary methods have been considered to obtain residual correction factors. A good agreement between the results of both methods has been observed and only minor difference are

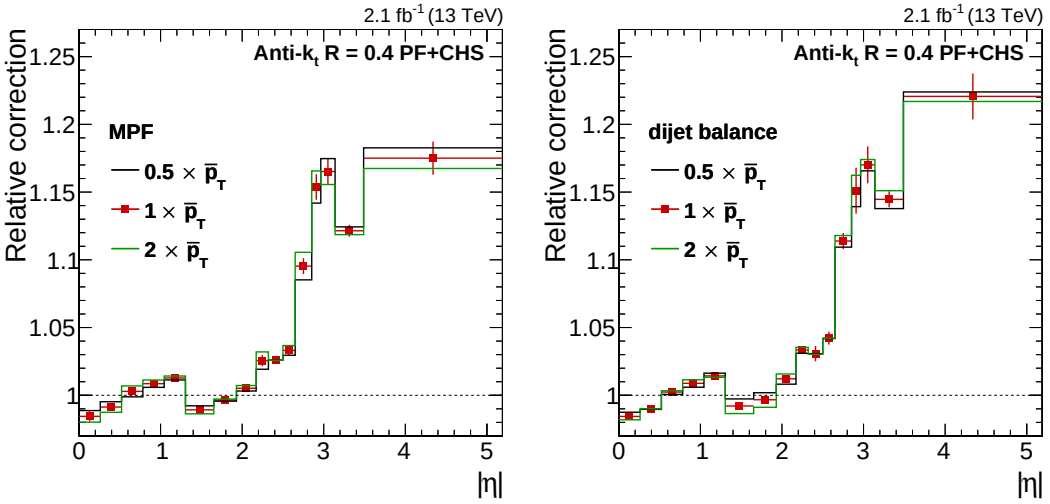

Figure 6.13: Relative residual correction factors with nominal and varied average transverse momentum $\bar{p}_{\mathrm{T}}$ for the MPF (left) and dijet balance (right) method.

found in the hadron forward region of the detector. As the origin of these differences is unknown, an additional uncertainty is taken into account by determining the $|\eta|$-dependent bin-by-bin ratios between the results of the two methods. This uncertainty has not been considered in earlier analyses.

6.8 Summary

The jet energy scale was measured in a dijet sample. The two complementary MPF and dijet balance methods have been exploited, where the latter was used to cross-check the results obtained with the MPF method. Data-to-simulation ratios of the jet response in dependence of $|\eta|$ have been calculated to extract correction factors. The correction factors differ from unity by up to 3% in the detector region up to $|\eta| < 2.5$ for both methods. In the endcap and hadron forward region of the detector, the correction factors vary across $|\eta|$ by up to 17% for the MPF method and up to 22% for the dijet balance method. Statistical uncertainties are found to be small and less than 2% over the full detector range. The dominant systematic uncertainties arise from time dependencies with uncertainties and differences in the results between the two methods with up to 5% in the hadron forward regions.

The final results obtained with the MPF method are shown in Fig. 6.14. The

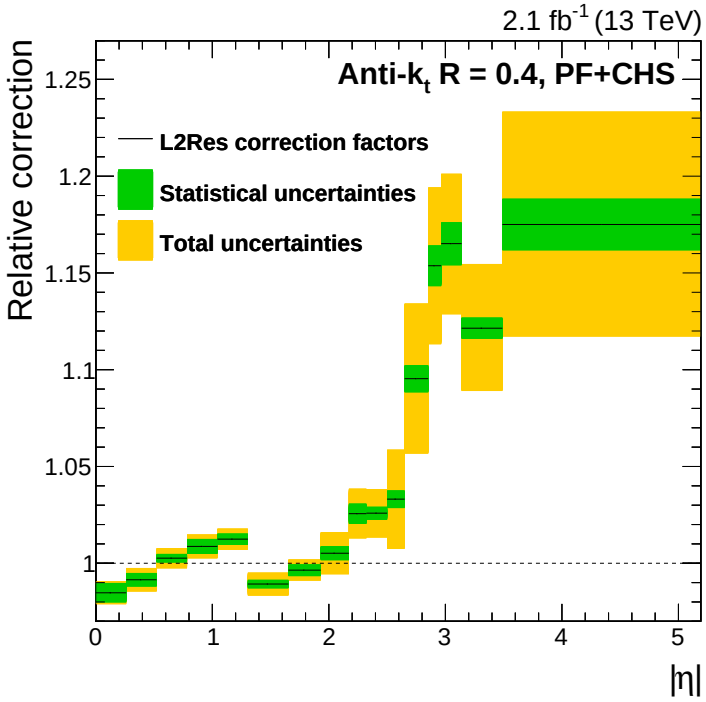

Figure 6.14: Relative residual correction factors determined on the 2015 dataset. The green band represents the statistical uncertainties and the orange band corresponds to the total uncertainties.

obtained correction factors have been provided to the CMS collaboration and applied to most CMS analyses analyzing data recorded in 2015.

7 Search for Scalar Leptoquarks

This chapter contains a search for pair produced scalar leptoquarks (LQs) decaying into a top quark and a tau lepton. A description of the analysis strategy is given in Sec. 7.1. All data samples and triggers used in this analysis are summarized in Sec. 7.2. The optimization steps of the event selection are presented in Sec. 7.3, and a data-driven estimation of the background due to misidentified τ_h leptons is described in Sec. 7.4. Systematic uncertainties are discussed in Sec. 7.5, and the final results are presented in Sec. 7.6. An outlook is given in Sec. 7.7.

7.1 Analysis Strategy

The decay of pair produced LQs provides several possible final states as discussed in Sec. 2.4. A sketch of the final state searched for in this analysis is shown in Fig. 7.1. We consider the pair production of LQs, where both LQs decay into a top quark and a hadronically decaying tau (τ_h) lepton. In this analysis, the $\ell + \tau_\mathrm{h} +$ jets final state is studied, where ℓ denotes an electron or muon. In the context of pair produced LQs decaying into a top quark and a τ lepton, candidate events containing at least one τ_h lepton are selected. The majority of LQ candidate events in this final state features one hadronically decaying W boson from a top quark, and another W boson decaying leptonically with either a muon or an electron in the final state. However, a small fraction of events are from the all-hadronic decay of the top quarks, with the τ lepton originating from an LQ decaying into a muon or an electron.

A sketch of the analysis strategy is shown in Fig. 7.2. First, a baseline selection is applied to create a validation region. For the optimization of the event selection two categories are adopted. In category A, exactly one τ_h lepton is required, whereas in category B events with at least two τ_h leptons are selected. Category A employs additional sub-categories based on the charge of the $\ell\tau_\mathrm{h}$ pair and on the value of S_T. By splitting into various sub-categories, maximum sensitivity to a wide range of LQ masses can be achieved. A template-based statistical evaluation of the results

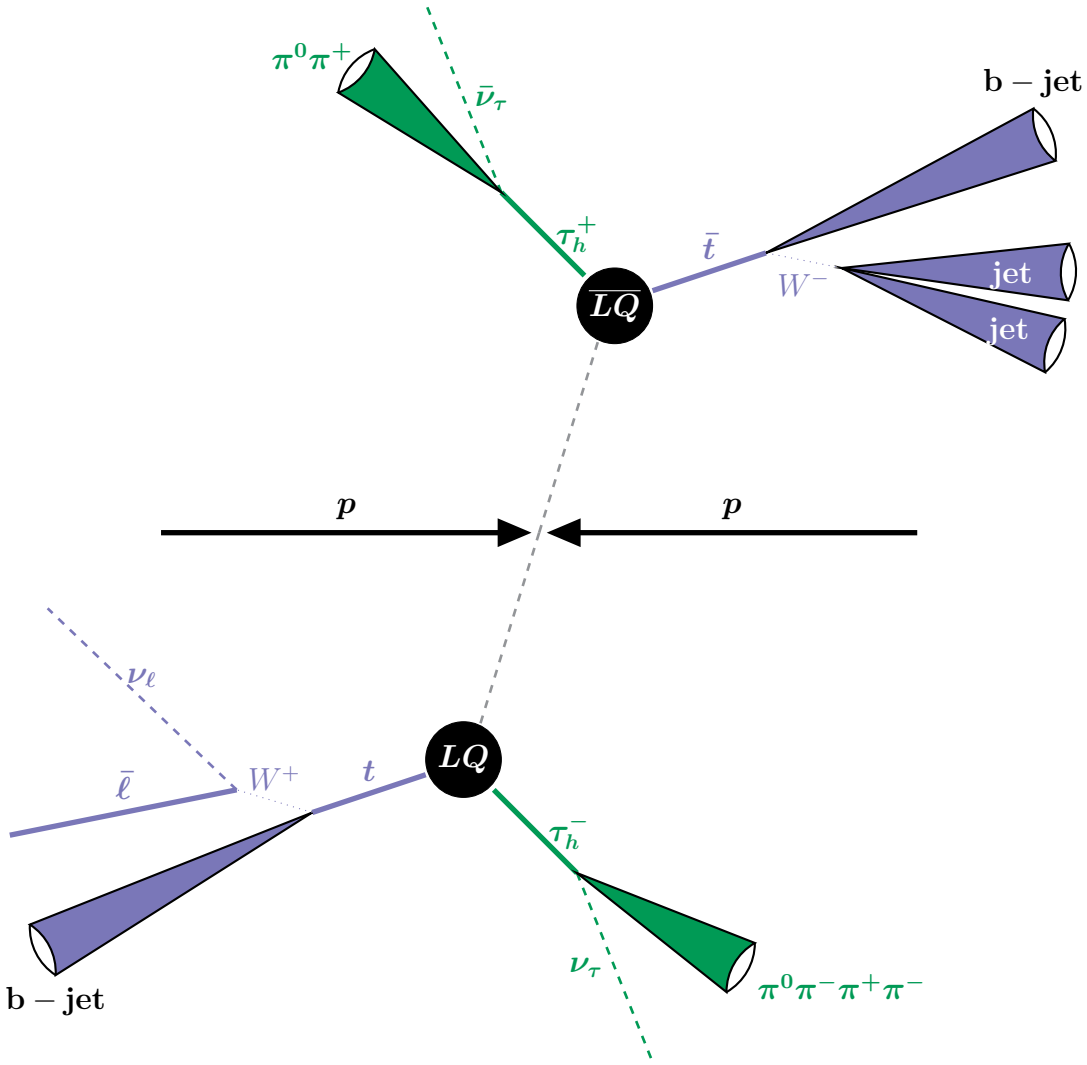

Figure 7.1: Sketch of the production of an LQ pair. Both LQs decay into a top quark and a τ_h lepton. The upper top quark decays hadronically and the lower one leptonically. For the upper τ_h lepton a 1-prong decay is illustrated, whereas a 3-prong decay is shown for the lower τ_h lepton. The lepton ℓ denotes an electron or a muon.

using the distribution of the transverse momentum of the reconstructed hadronically decaying top quark is performed in category A.

In category B, the requirement of at least two τ_h leptons and at least one additional lepton removes most events from background processes. This category is well-suited to provide good sensitivities to low LQ masses below 500 GeV. For higher LQ masses, the signal efficiency is too small to separate signal events from events of the remaining background processes. A counting experiment is performed in this

category as the number of expected background events is too small to benefit from a shape-based analysis.

The are two main sources of background events in this analysis. One consists of events passing the selection due to jets falsely identified as τ_h leptons, and another is from SM processes with prompt τ_h leptons. For the former, the main backgrounds are estimated using a data-driven method, whereas the backgrounds with prompt τ_h leptons, along with rarer background processes, are determined from simulation. Control regions (CRs) are defined for the estimation of the shape and normalization of the dominant background processes. The CRs consist of events with nonisolated τ_h leptons to enrich the fraction of misidentified τ_h leptons, while in the signal region (SR) only events that contain isolated τ_h leptons are selected. In category A, contributions from $t\bar{t}$ and W+jets containing misidentified τ_h leptons are derived in control region $\mathrm{CR_A}$ in which the isolation requirement of all τ_h leptons is inverted. In category B, the $t\bar{t}$ background contains either only misidentified τ_h leptons (in the following referred to as $t\bar{t}_f$), or at least one misidentified τ_h lepton and at least one prompt τ_h lepton (in the following referred to as $t\bar{t}_{p+f}$). Both background estimates are derived in separate control regions. The $t\bar{t}_f$ background is determined in control region $\mathrm{CR_{B1}}$ in which events contain only nonisolated τ_h leptons. In control region $\mathrm{CR_{B2}}$ events contain at least one isolated τ_h lepton and at least one nonisolated τ_h lepton to obtain the $t\bar{t}_{p+f}$ background.

7.2 Data Samples and Trigger

The dataset recorded in proton-proton collisions in 2016 at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ with a bunch spacing of 25ns is analyzed. The datasets are listed in Tab. 7.1. The total integrated luminosity is 35.9 fb^{-1} in both the single electron and single muon data sets [139]. The samples include only the data that have been certified as applicable for physics analysis with all sub-detectors of the CMS experiment functioning in good condition.

In the $\mu + \tau_h +$ jets final state, the single isolated muon triggers `HLT_IsoMu24` and `HLT_IsoTkMu24` have been used to select events. These triggers require at least one isolated muon candidate with $p_T > 24\,\mathrm{GeV}$ and differ only in the reconstruction of the muon in the HLT. Both triggers have been studied by the CMS collaboration

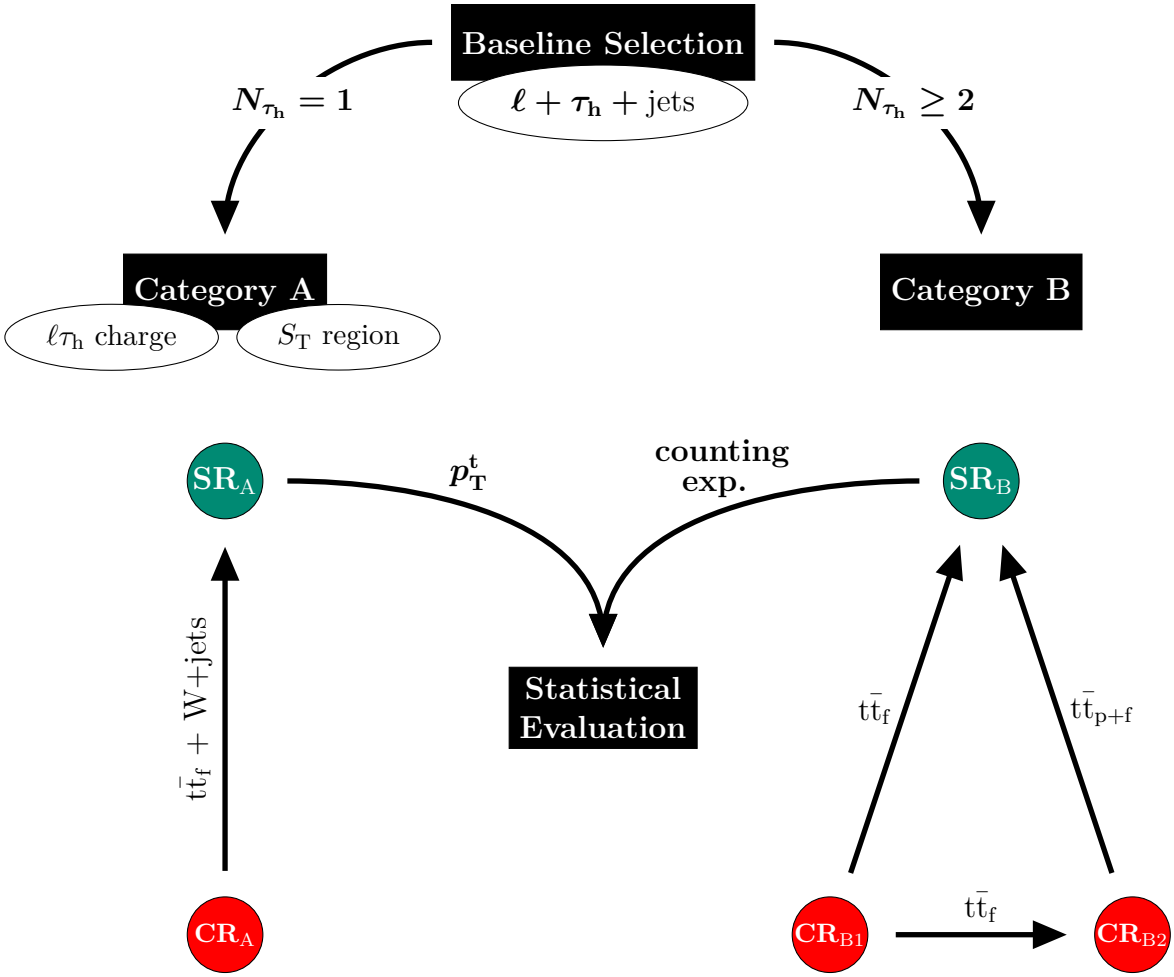

Figure 7.2: Sketch of the analysis strategy. The $\ell + \tau_h +$ jets final state is separated into two categories in which the events contain exactly one τ_h lepton or at least two τ_h leptons. In the signal region (green) the events are required to contain only isolated τ_h leptons, whereas the events in the control regions (red) contain nonisolated τ_h leptons. The dominant background processes are derived from the control regions.

and trigger efficiencies have been measured [140].

The data studied in the $e + \tau_h +$ jets final state have been collected by single isolated electron triggers. For the data taking periods B to G the trigger `Ele27_eta2p1_WPLoose` has been used, whereas the trigger `Ele27_eta2p1_WPTight` was used in period H. Both triggers require at least one isolated electron candidate with $p_T > 27\,\text{GeV}$ and $|\eta| < 2.1$. They differ in the working point (WP) of the electron identification criteria that is used in the HLT. The efficiencies of both triggers have been measured by the CMS collaboration [141].

Data set	$L\,[\mathrm{fb}^{-1}]$
/SingleMuon(SingleElectron)/Run2016B	5.78 (5.79)
/SingleMuon(SingleElectron)/Run2016C	2.57 (2.57)
/SingleMuon(SingleElectron)/Run2016D	4.25 (4.25)
/SingleMuon(SingleElectron)/Run2016E	4.01 (4.01)
/SingleMuon(SingleElectron)/Run2016F	3.10 (3.10)
/SingleMuon(SingleElectron)/Run2016G	7.54 (7.54)
/SingleMuon(SingleElectron)/Run2016H	8.61 (8.61)
Total	35.86 (35.87)

Table 7.1: Certified data sets and the corresponding integrated luminosity.

7.3 Event Selection

To select candidate events of our final state of interest, the requirement of at least one τ_h lepton candidate is mandatory. The transverse momentum p_T^τ of the τ_h lepton is, on average, expected to be larger in signal events than in background events. This is particularly true for high LQ masses. Requirements on p_T^τ are therefore used to reduce the amount of background events.

At least one lepton candidate (electron or muon) is required to select events in which one of the top quarks decays leptonically. As the decay products of leptonic top quarks and τ_h leptons contain neutrinos, a considerable amount of missing transverse energy $\not{E}_\mathrm{T}$ is expected in the signal events.

The signal events consist of an $\ell\tau_\mathrm{h}$ pair with either opposite sign (OS) charge or same sign (SS) charge. Under the assumption that exactly one τ_h lepton and exactly one lepton ℓ was selected in the event, the probability for events with an OS (SS) $\ell\tau_\mathrm{h}$ pair is approximately 70% (30%). In general, most background processes in an $\ell + \tau_\mathrm{h} +$ jets final state are expected to consist of an OS $\ell\tau_\mathrm{h}$ pair. However, the fraction of background events passing the SS requirement is relatively large. This is mainly due to τ_h leptons that are mimicked by a jet. As the charge of misidentified τ_h leptons is coincidental, such events can contribute to the SS $\ell\tau_\mathrm{h}$ pair category. Background events passing the OS requirement consist either of prompt τ_h leptons or misidentified τ_h leptons. Due to the different background components two categories based on the charge of the $\ell\tau_\mathrm{h}$ pair are considered.

Several jets are expected from both the decay of the top quarks, as well as additional radiation of gluons. In general, the signal events are expected to contain more jets than events from the dominant SM background. Similar to the transverse momenta of the τ_h lepton, the transverse momentum of jets in LQ events is also expected to be larger than for background events, in particular for high LQ masses. In addition, signal events contain a pair of b quarks, and thus b tagging plays an important role in this analysis.

For all signal processes, the S_T distribution (with S_T defined as in Sec. 5.8) is expected to peak at approximately $2M_{LQ}$ as this variable takes the transverse momenta of all reconstructed objects into account. The factor of 2 occurs because two LQs are generated in the signal events. Background events are characterized by significantly lower values of S_T. Thus, the S_T variable has good discrimination power between the SM background and the LQ process, with the exception of LQ samples with low masses ($M_{LQ} \lesssim 400\,\text{GeV}$), since such events have similar attributes as events from $t\bar{t}$ processes. Two categories based on different S_T regions are considered to achieve high selection efficiencies of LQs from low to high masses.

All selection criteria and categorization requirements are discussed in detail in the following.

7.3.1 Baseline Selection

A baseline selection is used to select as many signal events as possible, while reducing the number of background events. This defines a validation sample, which is used for further studies on the optimization of the event selection. The baseline selection criteria that events have to fulfill are summarized in the following:

- $\mu + \tau_h +$ jets final state: at least one muon candidate with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ passing the requirements described in Sec. 5.3,
- $e + \tau_h +$ jets final state: at least one electron candidate with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$ passing the requirements described in Sec. 5.4. A veto against muon candidates is used to avoid overlap between the two final states,
- at least one τ_h lepton candidate with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$ passing the requirements described in Sec. 5.5,

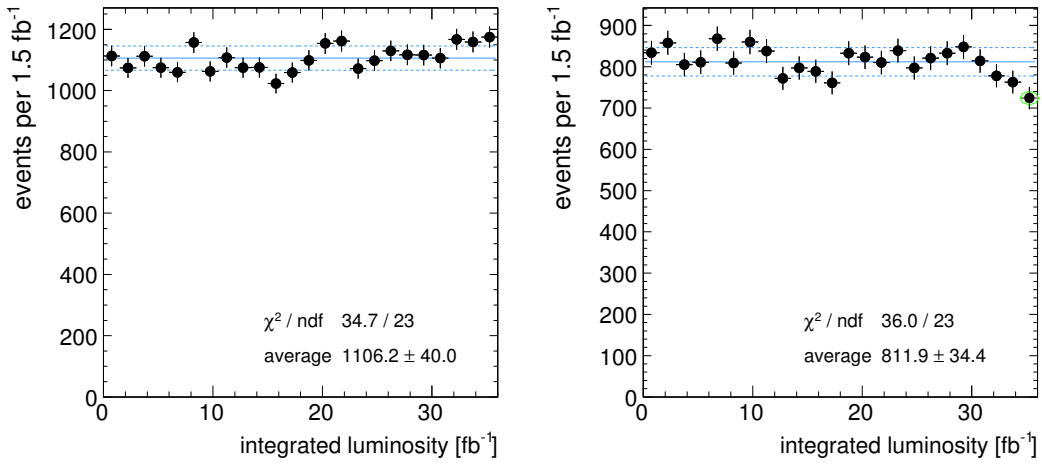

Figure 7.3: Number of events per $1.5\,\mathrm{fb}^{-1}$ passing the baseline selection as function of the integrated luminosity in the $\mu + \tau_\mathrm{h} +$ jets final state (left) and in the $e + \tau_\mathrm{h} +$ jets final state (right). The solid blue lines correspond to a constant fit to the data events. The dashed blue lines represent one standard deviation to the fit.

- at least two AK4 jets with $p_\mathrm{T} > 50\,\mathrm{GeV}$ and $|\eta| < 2.4$ passing the requirements described in Sec. 5.6,
- a third AK4 jet with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$,
- a requirement on the missing transverse energy of $\not{E}_\mathrm{T} > 50\,\mathrm{GeV}$ is applied to all events, where $\not{E}_\mathrm{T}$ fulfills the requirements described in Sec. 5.8, and
- a requirement on the S_T variable of $\mathrm{S_T} > 350\,\mathrm{GeV}$ is applied.

In Fig. 7.3 the number of events passing the baseline selection as a function of the integrated luminosity in bins of $1.5\,\mathrm{fb}^{-1}$ is shown for the $\mu + \tau_\mathrm{h} +$ jets final state and for the $e + \tau_\mathrm{h} +$ jets final state. The distributions allow us to check whether the collection of data events has any time dependence. The solid blue line shows the result of a constant fit to the data events and the dashed blue lines correspond to one standard deviation of the fit. In both final states, no trend is observed and the number of data events is stable as a function of time. This indicates that the conditions of the recording detector were robust and stable during all data taking periods.

The baseline selection is comparable to a dileptonic $t\bar{t}$ selection with additional jets, such that the dominant background in this analysis is from $t\bar{t} +$ jets production.

The $t\bar{t}$ background is split into two categories. For one category the $t\bar{t}$ events consist of at least one misidentified τ_{h} lepton, whereas in the other category the $t\bar{t}$ events consist of only prompt τ_{h} leptons. In all the following control distributions in this section, this separation is only shown for $t\bar{t}$ events as this is the main background in this analysis.

The second most important background is from W+jets events. Events from this process can pass the selection if a τ_{h} lepton is mimicked by a jet. The single top quark and Z+jets processes are minor backgrounds and events from these processes pass the baseline selection if the τ_{h} lepton was misidentified, except for single top quark processes in the tW-channel and Z+jets processes with $\tau_{\mathrm{h}}\tau(\rightarrow \ell\nu_{\tau}\nu_{\mu})$ decay modes. Events from diboson production that pass the selection contain mostly prompt τ_{h} leptons. However, only a few events from this process are expected as the production cross section is small compared to the cross sections from other SM background processes. The background from QCD-multijet production is almost negligible. This rejection power is from the requirement of at least one isolated lepton candidate with the tight identification criteria. In addition, the requirement on the missing transverse energy $\not\!\!E_{\mathrm{T}}$ also significantly reduces the QCD-multijet background.

Control distributions in the $\mu + \tau_{\mathrm{h}} + \mathrm{jets}$ ($e + \tau_{\mathrm{h}} + \mathrm{jets}$) final state after applying the baseline selection are shown in Fig. 7.4 (Fig. 7.5). In these control distributions, all background processes are determined from simulation. The dark grey bands represent the statistical uncertainties of the MC prediction. Many of the background events consist of misidentified τ_{h} leptons. In general, misidentified objects are not expected to be described well by simulation, notably the production of misidentification τ_{h} leptons can be different between data and simulation by approximately 20% [114]. Thus in the control distributions, an uncertainty of 20% is applied to the $t\bar{t}$ background that consists of misidentified τ_{h} leptons and is represented by the light grey bands. The $t\bar{t}$ background that contains misidentified τ_{h} leptons will be explicitly estimated from data in dedicated control regions (see Sec. 7.4).

In general, a good agreement of shape and normalization between data and the SM prediction is observed in all distributions.

Signal samples of LQs with masses of 300, 600, and 900 GeV are shown in the control distributions. As discussed before, the signal events tend to contain more

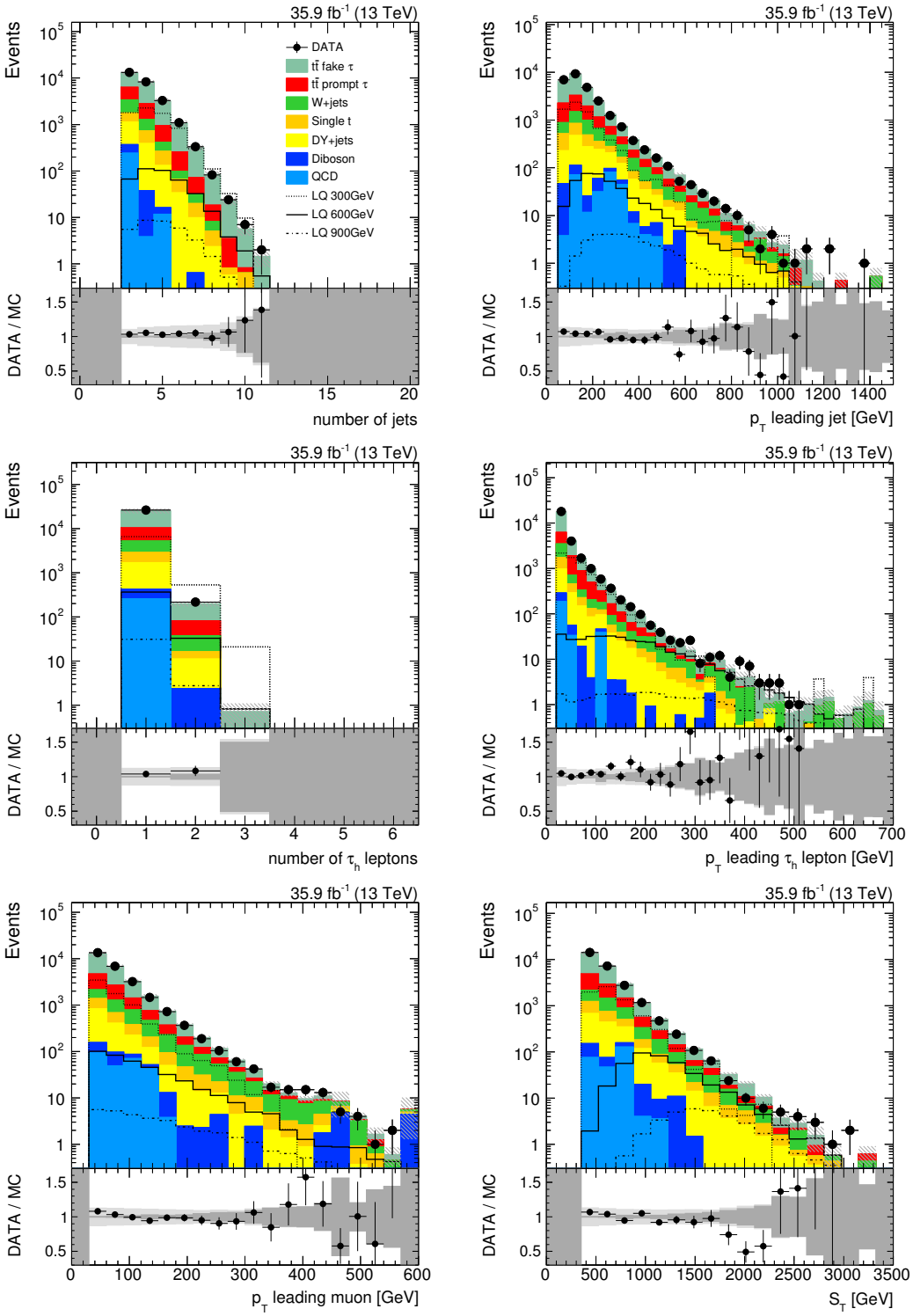

Figure 7.4: Control distributions in the $\mu + \tau_{\text{had}} +$ jets final state after applying the baseline selection. The distributions for the number of jets (top left), p_{T} of the leading jet (top right), the number of τ_{h} leptons (middle left), p_{T} of the leading τ_{h} lepton (middle right), p_{T} of the leading muon (bottom left) and S_{T} (bottom right) are shown. In the lower panels, the dark grey band corresponds to the statistical uncertainties of the background prediction, and the light grey band represents an uncertainty of 20% in the $t\bar{t}$ background that consists of at least one misidentified τ_{h} lepton.

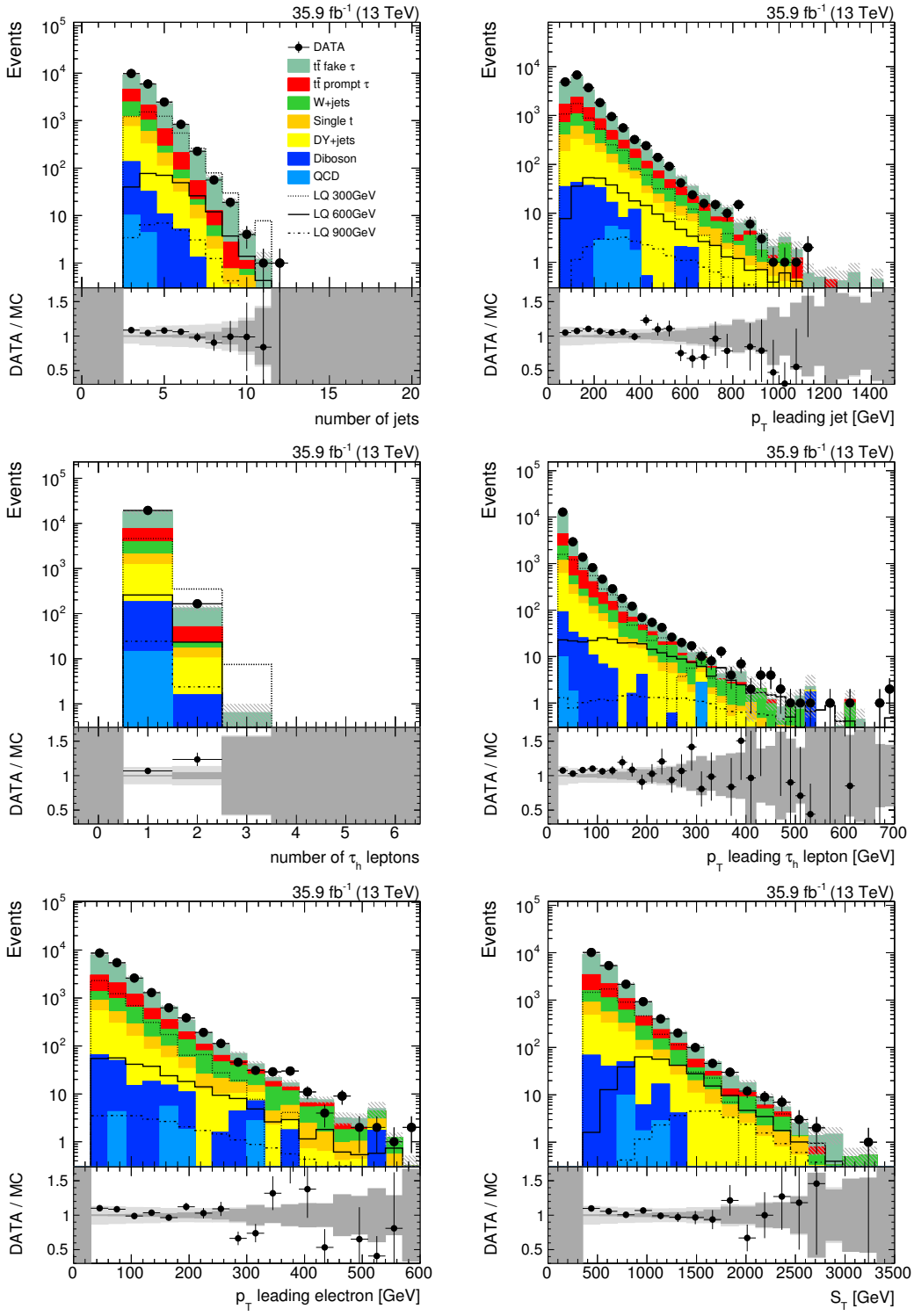

Figure 7.5: Control distributions in the $e + \tau_{\text{had}} +$ jets final state after applying the baseline selection. The distributions for the number of jets (top left), p_{T} of the leading jet (top right), the number of τ_{h} leptons (middle left), p_{T} of the leading τ_{h} lepton (middle right), p_{T} of the leading electron (bottom left) and S_{T} (bottom right) are shown. In the lower panels, the dark grey band corresponds to the statistical uncertainties of the background prediction, and the light grey band represents an uncertainty of 20% in the $t\bar{t}$ background that consists of at least one misidentified τ_{h} lepton.

jets than the background events. Moreover, the p_{T}^{τ} and $p_{\mathrm{T}}^{\mathrm{jet}}$ distributions exhibit harder spectra for signal processes compared to background processes from the SM.

In the distribution showing the number of τ_{h} leptons one can see that in most signal events only one τ_{h} lepton candidate is expected. This is because the reconstruction and identification efficiencies reduce the fraction of signal events that contain two or more τ_{h} leptons. In addition, at least one additional lepton is required. As the lepton can originate from a τ lepton decay, the probability to select two τ_{h} leptons decreases further.

The SM background can be reduced by imposing additional requirements on the variables discussed above. Furthermore, it is possible to optimize the selection for different LQ mass regions using different categories. The final selection criteria are discussed later.

7.3.2 Kinematic Reconstruction of the Hadronically Decaying Top Quark

A proper variable calculated from reconstructed objects in the detector is used as input for the statistical analysis of the search. This variable should provide good discrimination power between signal events and background events to benefit from the shape-based analysis that is performed. A common choice for the discriminating variable is the reconstructed mass of the hypothetical particle. However, in the studied final state of this analysis the reconstruction of the invariant mass of the LQs is unfeasible due to the presence of at least two and up to six neutrinos in the events. As the directions and the momenta of the respective neutrinos can not be determined, it is not possible to associate neutrinos to reconstructed objects in the reconstruction procedure.

A search for pair production of LQs decaying into a top quark and a tau lepton in the $\ell + \tau_{\mathrm{h}} +$ jets final state has already been performed at a center-of-mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$ [2]. In this analysis, the p_{T} of the leading τ_{h} lepton was used as the discriminating variable. As discussed above, the transverse momentum of the τ_{h} lepton is generally higher in the LQ signal events compared to the SM background events and is therefore a powerful variable in this analysis. To estimate the backgrounds in the analysis performed at 8 TeV, a control region was built to derive misidentification rates in the τ_{h} lepton identification, where MC simulation was used

to determine the backgrounds. The misidentification rates were then applied to the simulated events.

However, for the analysis in this thesis a data driven approach is used for the estimation of the backgrounds that contain misidentified τ_h leptons (see Sec. 7.4). The background estimation in this method would be inaccurate if the p_T of the leading τ_h lepton was used, because the kinematic differences between signal region and control region observed in this variable become too large. Thus, this variable is not utilized as the discriminating variable.

Beside tau leptons, top quarks are the other decay products of the LQs. It is also expected that the transverse momentum of top quarks takes, on average, higher values in signal events than in SM background events. For this reason, the hadronically decaying top quark is reconstructed and its transverse momentum p_T^t is used for the statistical interpretation of the results in this analysis. The leptonically decaying top quark can not be considered due to the presence of at least two neutrinos in signal events.

For the reconstruction of p_T^t, all AK4 jets in an event are considered. If there are more than seven AK4 jets in an event, the seven jets with highest p_T are taken into account. Out of these seven jets, a list of all possible combinations containing between one and seven jets is constructed. Each combination constitutes a top quark hypothesis. The four-momentum of a hypothesis is given by the sum of the four-momenta of the assigned jets. Finally, only one hadronic top quark hypothesis is chosen in each event. The selected hypothesis is the one in which the reconstructed top quark mass is closest to the mass of 172.5 GeV. In Fig. 7.6, the invariant mass of the selected top quark hypothesis and the number of jets used to reconstruct the hypothesis are shown. For all processes, the mass distributions peak at the top quark mass. However, this peak is found to be narrower in signal events as true top quarks in hadronic decay channels are reconstructed in these events. For the SM backgrounds, events only from $t\bar{t}$ processes can contain hadronically decaying top quarks, if the reconstructed τ_h lepton is misidentified.

Two jets create the top quark hypothesis in most of the background events. In Tab. 7.2 the fraction of the number of chosen hypotheses with a given number of contained jets is shown for the dominant $t\bar{t}$ background and for signal samples with different LQ masses. In all samples most of the hypotheses consist of two jets.

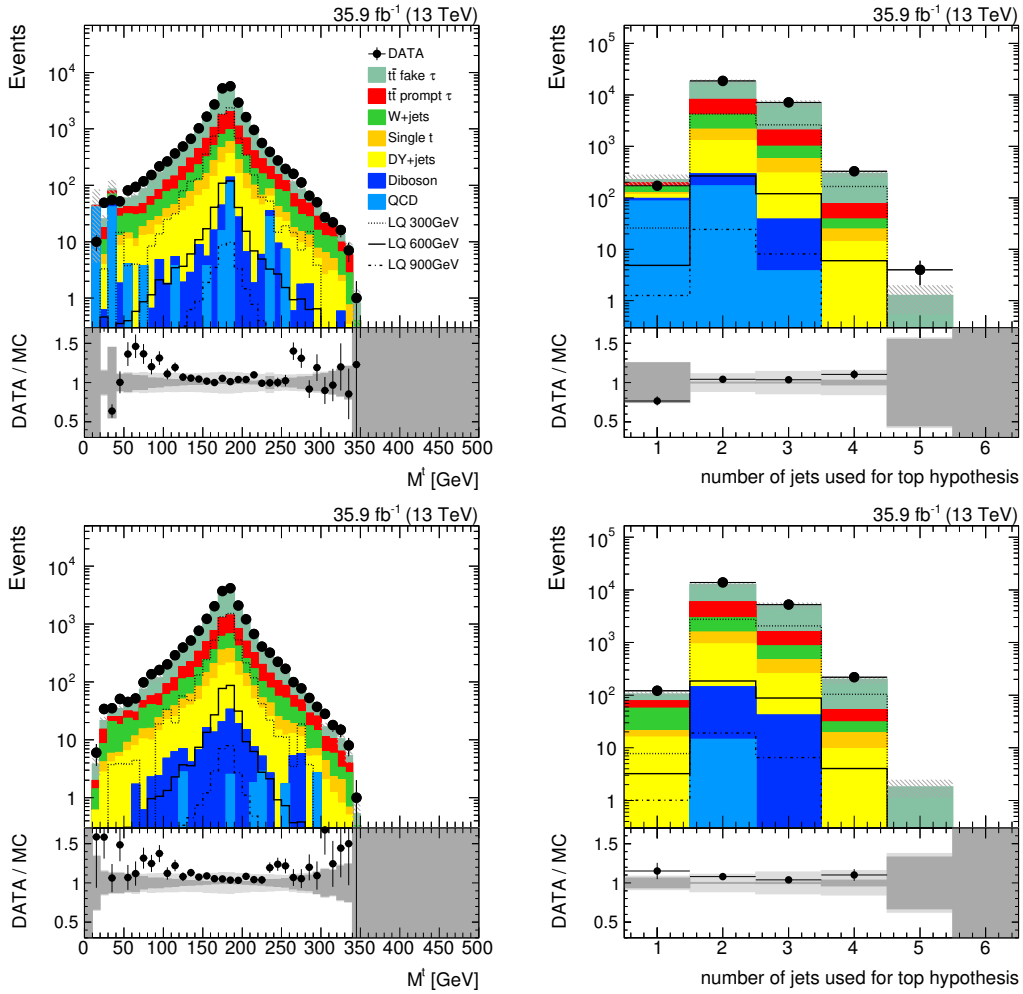

Figure 7.6: Control distributions of the kinematic reconstruction of the hadronically decaying top quark after applying the baseline selection. The distributions of the invariant mass (top left) and the number of jets used for the hypothesis (top right) in the $\mu + \tau_h +$ jets final state are shown. The corresponding distributions in the $e + \tau_h +$ jets final state are shown in the bottom. In the lower panels of each plot, the dark grey band corresponds to the statistical uncertainties of the background prediction, and the light grey band represents an uncertainty of 20% on the $t\bar{t}$ background that consists of at least one misidentified τ_h lepton.

The number of jets used to build the top quark hypothesis decreases for higher LQ masses. Due to the larger Lorentz-boost of top quarks in the signal events of high LQ masses, the decay products become more collimated.

The performance of the method used to reconstruct the hadronically decaying top

	$N^{\mathrm{t}}_{\mathrm{constituents}}$			
process	1	2	3	$\geq$4
$t\bar{t}$ SM	0.31%	69%	29%	1.3%
LQ(300 GeV)	0.37%	60%	37%	2.3%
LQ(400 GeV)	0.36%	63%	35%	1.5%
LQ(500 GeV)	0.66%	66%	32%	1.4%
LQ(600 GeV)	1.2%	67%	30%	1.5%
LQ(700 GeV)	1.6%	67%	30%	1.0%
LQ(800 GeV)	2.8%	70%	27%	0.7%
LQ(900 GeV)	3.7%	72%	24%	0.8%
LQ(1000 GeV)	5.5%	72%	22%	0.8%

Table 7.2: Fraction of the number of jets that are used to build the hypothesis of a hadronically decaying top quark for $t\bar{t}$ events and signal events of different masses.

quark is validated as follows. In Tab. 7.3, the fraction of *matchable* and *correctly-matched* events is listed for the $t\bar{t}$ background and several LQ samples of different masses.

Events are classified as *matchable* if a correct hypothesis of at least one hadronically decaying top quark exists. This is the case if at least one top quark was found to decay hadronically at generator level and if the decay products of the generator top quark can be matched with any of the reconstructed jets passing the event selection within the radius of $\Delta R = 0.4$. The fraction of matchable events is given with respect to the number of selected events of the baseline criteria. The fraction of matchable hypotheses is very low for the $t\bar{t}$ background and arises only from events with misidentified τ_{h} leptons. Dileptonic $t\bar{t}$ events are included if the reconstructed τ_{h} lepton and the reconstructed muon or electron are found to be prompt leptons. In that case no hadronically decaying top quark exists at generator level.

In the signal samples, the fraction of matchable events increases from 31 to 57% for increasing LQ masses. The generator partons of an LQ decay and their corresponding reconstructed jets have larger transverse momenta for higher LQ masses. Because a transverse momentum threshold of 30 GeV is required in order to consider a jet in an event (see Sec. 5.6), it is more likely to keep those reconstructed jets in signal events with large LQ masses and thus more of these events can be classified as matchable.

A *correctly-matched* hypothesis was found if each AK4 jet corresponding to the reconstructed top quark hypothesis can be matched to a generated final state parton

process	matchable [%]	correctly-matched [%]
$t\bar{t}$ SM	9.7	43.6
LQ (300 GeV)	31.2	32.6
LQ (400 GeV)	37.5	26.4
LQ (500 GeV)	41.9	29.3
LQ (600 GeV)	46.4	33.3
LQ (700 GeV)	50.2	37.4
LQ (800 GeV)	53.9	37.8
LQ (900 GeV)	54.8	41.5
LQ (1000 GeV)	57.4	44.7

Table 7.3: Fraction of matchable and correctly-matched events after applying the baseline selection. The fraction of correctly-matched events is determined with respect to the fraction of matchable events. The efficiencies are shown for the $t\bar{t}$ background and different LQ hypotheses.

from the top quark decay within the radius of $\Delta R = 0.4$. Each parton is required to be matched by an AK4 jet. To consider collimated top quark decay products, it is possible that more than one of the generated partons are matched by a single reconstructed jet. The fraction of *correctly-matched* events is given with respect to the number of *matchable* events. This fraction is in the range of 32 to 45% for the signal events with better performances for higher LQ masses.

The normalized distributions of $p_{\mathrm{T}}^{\mathrm{t}}$ for different LQ mass hypotheses and the $t\bar{t}$ background are shown in Fig. 7.7 after applying the baseline selection in the $\ell + \tau_h +$jets final states. A comparison with respect to the $t\bar{t}$ background is shown as this is the dominant background in this analysis. The shape comparison shows that the $p_{\mathrm{T}}^{\mathrm{t}}$ spectrum of the reconstructed hadronically decaying top quark gives good discrimination power between different LQ mass hypotheses as well as between the LQ samples and the $t\bar{t}$ background. In particular, the top quarks arising from LQs have higher transverse momenta compared to the top quarks in $t\bar{t}$ SM background processes.

7.3.3 Final Selection and Definition of Search Categories

The baseline selection represents a validation region and it is the basis for the optimization procedure of the event selection. All selection criteria presented in the following are applied in addition to the baseline selection.

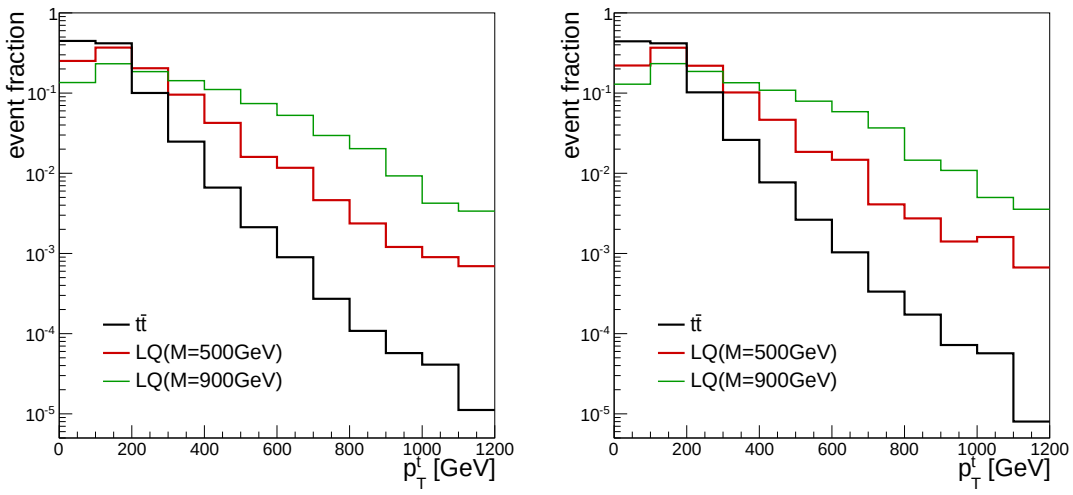

Figure 7.7: Shape comparison of $p_{\mathrm{T}}^{\mathrm{t}}$ for different LQ mass hypotheses and the $t\bar{t}$ background in the $\mu + \tau_{\mathrm{h}} +$ jets (left) and $e + \tau_{\mathrm{h}} +$ jets (right) final state.

In general, the requirements have been chosen such that the signal to background separation is optimized. A common variable for the optimization of the event selection is the $s/\sqrt{s+b}$ ratio, where s denotes the number of signal events and b denotes the number of background events. It has been found Ref. [142] that this value can diverge from true sensitivity optimization for cases with a low number of background events. As this is true in this analysis, an asymptotic formalism as described in Ref. [142] is used. Following this approach, the median significance med[Z_0], defined as

$$\mathrm{med}[Z_0] = \sqrt{2((s+b)\ln(1+s/b) - s)}, \tag{7.1}$$

is maximized to optimize the event selection.

To maximize med[Z_0], several combinations of different requirements to potentially sensitive variables have been varied. These input variables are the number of jets N_{jets}, the missing transverse energy $\not{E}_{\mathrm{T}}$, the number of b jets $N_{\text{b-jets}}$, the transverse momenta of jets $p_{\mathrm{T}}^{\mathrm{jet(s)}}$, the transverse momentum of the τ_{h} lepton $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$, the transverse momentum of the muon or electron p_{T}^{ℓ} and S_{T}. The modification of the requirements on p_{T}, $\not{E}_{\mathrm{T}}$, and S_{T} has been done in steps of 20-100 GeV. The event selection and categorization as presented in the following is found to provide the best sensitivity over a wide range of LQ masses. Top tagging algorithms as

described in Sec. 5.9 have also been investigated. However, these algorithms have not been found to improve the sensitivity in this analysis.

Two categories are considered. Events that contain exactly one τ_{h} lepton candidate are selected in category A, whereas in category B they must contain at least two τ_{h} lepton candidates. The fraction of selected signal events is independent of the LQ mass and approximately 90%(10%) in category A(B) after the baseline selection criteria are applied. Category A is hence employed to improve the sensitivity to a wide range of LQ masses. Tight requirements on kinematic variables are applied in addition to the baseline selection for this purpose.

In category B, the selection efficiency of the signal samples is independent of the LQ mass, but the branching fraction is low for increasing LQ masses. This leads to a low total number of expected signal events for LQ samples of large masses and results in poor sensitivity. However, category B is well suited to improve the sensitivity for low LQ masses. For these masses the total number of signal events is large and many background events are removed by the requirement of at least two τ_{h} leptons. Additional requirements to the baseline selection are also applied in category B, but the requirements are much looser with respect to the requirements applied in category A.

Category A

In this category, events containing exactly one τ_{h} lepton candidate are selected. In addition, two sub-categories based on the charge of the $\ell\tau_h$ pair are introduced. For this purpose, events with an opposite sign $\ell\tau_h$ pair are separated from the events that contain an $\ell\tau_h$ pair with same sign charge. If the events contain more than one electron or muon, the pair with the largest scalar p_{T} sum is chosen. Except for the rare SM background with diboson (WZ, ZZ) processes, background events are not supposed to enter the SS category since there is no other SM process that contains a $\ell\tau_h$ pair with SS charge. However, events from non-diboson background processes survive the requirement of one SS $\ell\tau_h$ pair. This is mainly caused by reconstructed τ_{h} leptons, which can be mimicked by a jet and the charge sign of misidentified τ_{h} leptons is arbitrary. A further but minor reason is that the charge of a prompt τ_{h} lepton can be misidentified. After applying the baseline selection, the fraction of events that contain a SS $\ell\tau_h$ pair is approximately 30%.

In the SS sub-category, an event has to fulfill the following requirements:

- The event contains a same sign $\ell\tau_{\mathrm{h}}$ pair,
- the p_{T} of the leading jet is at least 150 GeV,
- the p_{T} of the leading τ_{h} lepton is at least 100 GeV, and
- at least one b tagged AK4 jet is present passing the medium identification criteria.

In the OS sub-category, the following requirements are applied:

- The event contains an OS $\ell\tau_{\mathrm{h}}$ pair,
- a fourth AK4 jet with with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$ is present,
- the event has to have $\not{E}_{\mathrm{T}} > 100$ GeV,
- the p_{T} of the leading τ_{h} lepton is at least 100 GeV, and
- at least one b tagged AK4 jet is present passing the medium identification criteria.

The med[Z_0] significance is shown after applying each selection step successively in Tab. 7.4 to validate the performance of the final event selection. The values are presented in the μ+τ_{h}+jets and e+τ_{h}+jets final states for LQ samples with masses of 500 GeV and 900 GeV, respectively. One can observe that the med[Z_0] significance increases after applying each selection criterion. The largest improvements arise from the requirement in the transverse momentum of the leading τ_{h} lepton. The values are larger in the LQ mass samples of 500 GeV compared to the LQ sample with 900 GeV because the production cross section is higher. However, the relative improvements of the med[Z_0] significance are found to be larger for the LQ mass samples of 900 GeV.

It can also be observed that the sensitivity in the OS sub-category is better compared to the SS sub-category, although more background events contain an OS $\ell\tau_h$ pair after applying the baseline selection. This fraction of background events containing an OS $\ell\tau_{\mathrm{h}}$ pair is approximately 70%. However, the fraction of signal events containing an OS $\ell\tau_h$ pair is also found to be in the order of 70% and hence the OS category provides a large room to improve the med[Z_0] significance.

	$\mu + \tau_h + \text{jets}$		$e + \tau_h + \text{jets}$	
Selection	LQ 500 GeV	LQ 900 GeV	LQ 500 GeV	LQ 900 GeV
Baseline + OS	5.08	0.18	3.96	0.16
$p_T^{j1} > 150\,\text{GeV}$	5.98	0.27	4.71	0.24
at least four jets	6.82	0.31	5.58	0.29
$\not{E}_T > 100\,\text{GeV}$	7.31	0.40	5.90	0.37
$p_T^{\tau} > 100\,\text{GeV}$	12.45	1.11	9.24	1.00
at least one b tag	12.59	1.12	9.50	1.01
Baseline + SS	3.04	0.09	2.37	0.09
$p_T^{j1} > 150\,\text{GeV}$	3.63	0.14	2.92	0.13
$p_T^{\tau} > 100\,\text{GeV}$	6.94	0.42	4.73	0.32
at least one b tag	7.37	0.48	5.38	0.38

Table 7.4: Cut-flow table in Category A in the $\mu + \tau_h + \text{jets}$ and $e + \tau_h + \text{jets}$ final states. The med[Z_0] significance is shown for LQ signal samples of masses of 500 GeV and 900 GeV. The values are given after applying each requirement successively.

Finally, two additional sub-categories are considered based on the value of S_T. The S_T variable includes the transverse momenta of all reconstructed objects in an event. For the background processes, most of the events are observed at low values of S_T as shown in the bottom right distributions of Fig. 7.4 and Fig. 7.5. In processes of pair produced LQs, most events in the S_T distribution are found at $S_T \approx 2M_{LQ}$. Thus, by using a tight requirement on S_T many background events can be removed while signal events with high LQ masses are not affected. However, the signal efficiency in signal samples with low LQ masses would be reduced by such a requirement. In order to improve the sensitivity on a wide range of LQ masses the selection is further divided into a low S_T and a high S_T category. In the low S_T category events with $S_T \leq 1200\,\text{GeV}$ are selected, in the high S_T selection a requirement of $S_T > 1200\,\text{GeV}$ is applied. This means that category A contains a total of four sub-categories, based on the charge of the $\ell\tau_h$ pair and based on the value of S_T.

Category B

In category B, events are required to contain at least two τ_h lepton candidates. This requirement removes a large fraction of the SM background processes as these are usually only selected because they contain other objects misidentified as one or more

prompt τ_h leptons. The exception to this are diboson production events, which may contain one or more τ_h leptons, but the cross sections for these processes are small. The selection criteria in this category are adapted to provide high sensitivities to low LQ masses. For this reason rather loose criteria are applied in addition to the baseline selection.

The following requirements are applied in category B:

- The events contain at least two τ_h leptons and the leading τ_h lepton and second leading τ_h lepton are required to have opposite sign charge,
- the p_T of the leading τ_h lepton is at least 65 GeV, and
- the p_T of the second leading τ_h lepton is at least 35 GeV.

In the $\mu + \tau_h +$ jets ($e + \tau_h +$ jets) final state, the values of med[Z_0] are 9.30 (7.42) and 0.52 (0.43) for LQ masses of 500 GeV and 900 GeV, respectively. In this category, a counting experiment is performed as the number of expected background events is too small to benefit from a shape-based analysis.

The selection efficiency on the signal samples is shown in Fig. 7.8 for both categories and the combination of them. The efficiencies are determined by the ratio of events passing the respective event selection divided by all generated events. Signal events with lower LQ masses have lower values in p_T^τ and p_T^{jet}. Since requirements on these variables are applied, lower LQ masses have lower effiencies and the selection efficiency increases as a function of the LQ mass. The selection efficiencies are shown in the $\mu + \tau_h +$ jets and $e + \tau_h +$ jets final states, respectively. They are found to be larger in the $\mu + \tau_h +$ jets final state because a veto against muons is applied in the $e + \tau_h +$ jets final state. Moreover, a tighter requirement to the pseudorapidity of electrons ($\eta < 2.1$) is applied in the $e + \tau_h +$ jets final state due to the limited trigger acceptance.

7.4 Estimation of the Background due to Misidentified Tau Leptons

In this analysis, only hadronically decaying τ leptons are considered. Background events that pass the event selection contain either misidentified or prompt τ_h leptons (or both in some cases).

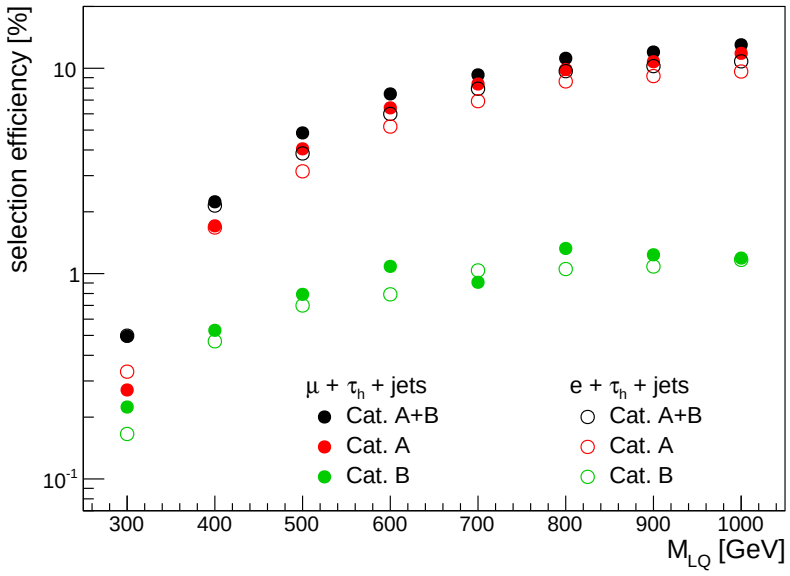

Figure 7.8: Selection efficiency of the signal processes as a function of the LQ mass in category A and B individually and inclusively.

Mimicked objects like falsely identified τ_h leptons are very difficult to describe in simulation. A reliable prediction of events containing misidentified objects can be determined in a data-driven approach.

In the following, the dominant backgrounds that contain at least one misidentified τ_h lepton are obtained from dedicated control regions defined for the two categories A and B. In this procedure, the backgrounds are derived and extrapolated from data in separate control regions.

An extrapolation method, called α-method, is utilized for the background estimation. In Fig. 7.9, the scheme of the α-method is illustrated. The first step is to define a control region. In general, similar selection criteria between the signal region and the control region are chosen to keep the extrapolation procedure simple. The common procedure is to invert one of the selection criteria of the signal region. The signal contamination in the control region is required to be very low, otherwise, potential signal events could bias the background of interest by the extrapolation procedure. Simultaneously, the fraction of the background of interest is required to be large to reduce statistical uncertainties in the extrapolation procedure.

The difference in the event selection between signal region and control region can potentially lead to kinematic differences between the two regions. Thus, the second

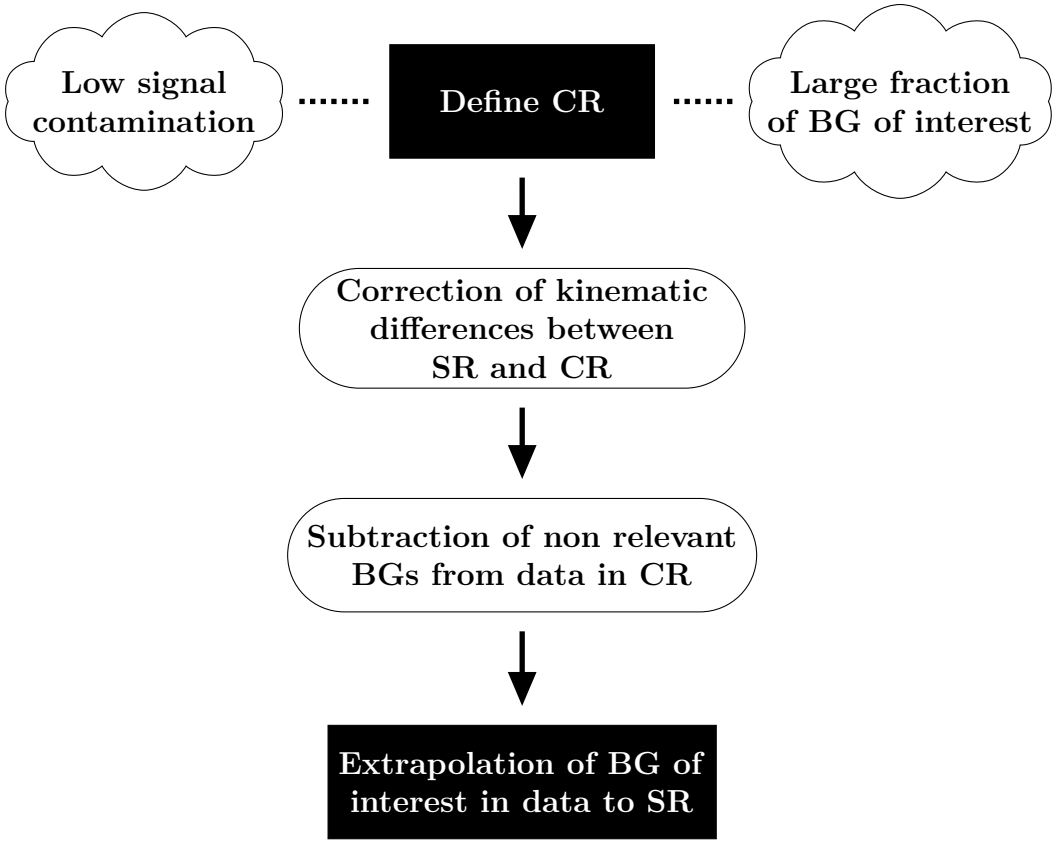

Figure 7.9: Strategy of the α-method schematically. A control region (CR) is defined in which the signal contamination is low and the fraction of the background of interest is large. Once the kinematic differences between signal region and control region are corrected and the non relevant backgrounds are subtracted from the data in the control region, the remaining data is extrapolated from the control region to the signal region.

step of the α-method is to correct for those kinematic differences.

The third step is to subtract the non relevant backgrounds from data in the control region to make sure that the data consist of only the background of interest. Finally, the remaining data events are extrapolated from the control region to the signal region.

As will be discussed in the following the $t\bar{t}$ background is derived in different control regions separately for different multiplicities of misidentified τ_h leptons and prompt τ_h leptons in an event. The following declarations are considered:

- the $t\bar{t}$ background that consists of only misidentified τ_h leptons is referred to as $t\bar{t}_\mathrm{f}$,

- the $t\bar{t}$ background that consists of at least one prompt τ_h lepton and at least one misidentified τ_h lepton (only used in Category B) is referred to as $t\bar{t}_{p+f}$, and
- the $t\bar{t}$ background that consists of only prompt τ_h leptons is referred to as $t\bar{t}_p$.

The $t\bar{t}_p$ background is estimated in simulation in category A. In category B, none of the simulated $t\bar{t}_p$ events pass the event selection.

Category A

In each search region of category A, the fraction of events from $t\bar{t}$ processes is the largest with more than 75% with respect to the total number of background events. The $t\bar{t}_f$ background events are derived from a dedicated control region. It has been verified whether the τ_h lepton is a prompt lepton or falsely identified by using a matching criteria between reconstructed and generated τ_h leptons. A prompt τ_h lepton is identified if the distance ΔR between reconstructed and generated τ_h lepton is smaller than 0.4. The fraction of falsely identified τ_h leptons in $t\bar{t}$ events depends on the search region. In the SS category, the $t\bar{t}$ events are generally not expected to contain prompt τ_h leptons. However, some of the $t\bar{t}$ events passing the SS requirement contain a prompt τ_h lepton because the charge sign of τ_h leptons can be misidentified. The fraction of falsely identified τ_h leptons in the SS category is 95%. The OS category basically corresponds to a dileptonic selection of $t\bar{t}$ + jets processes, but due to jets faking a τ_h lepton the fraction of misidentified τ_h leptons is 51%. In both categories, the remaining $t\bar{t}_p$ events are taken from simulation.

The W+jets background is derived simultaneously with the $t\bar{t}_f$ background from the control region. The W+jets background events surviving the selection requirements contain a misidentified τ_h lepton. The fraction of falsely identified muons and electrons is found to be negligible due to the tight identification and isolation requirements of selected muons and electrons.

The isolation is the main tool to reduce the misidentification rate of τ_h leptons. Thus, by inversion of the isolation requirement most of the events contain falsely identified τ_h leptons. The control region in category A (CR_A) is defined by the same selection requirements as in the SR, but the isolation requirement of the τ_h lepton is inverted. The isolation variable of all τ_h leptons in the signal region after applying the baseline selection is shown in Fig. 7.10 (left). As introduced in Eq. (5.3), the requirement of $I_\tau < 1.5\,\text{GeV}$ is applied. The peak at zero arises from events that

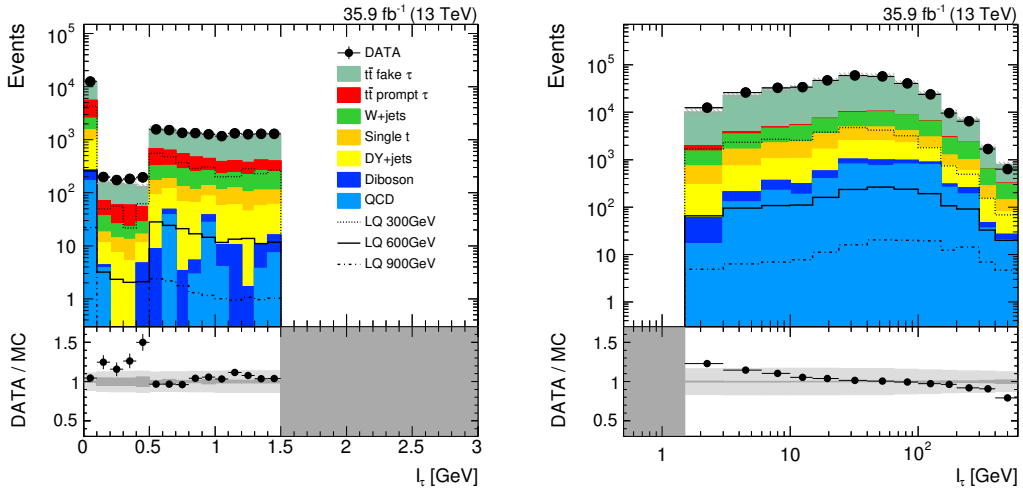

Figure 7.10: Distribution of the isolation of τ_h lepton candidates, after the baseline selection was applied. In the signal region, the τ_h lepton candidates are isolated (left), whereas in the CR (right) the isolation requirement of τ_h lepton candidates is inverted.

do not contain any further charged hadron or photon within the isolation cone of $\Delta R < 0.4$ of the τ_h lepton candidate. The step at 0.5 GeV appears because of events that contain at least one charged hadron (those charged hadrons are required to fulfill $p_T > 0.5\,\text{GeV}$) within the radius of the τ_h lepton candidate. In case of events containing photons but no charged hadrons within the radius of the τ_h lepton candidate, the resulting value of the τ_h lepton isolation can be between 0 and 0.5 GeV. The distribution of the τ_h isolation variable in the control region is shown in the right distribution of Fig. 7.10.

Events with at least one τ_h lepton candiate are selected in the control region CR_A (whereas in the signal region exactly one τ_h lepton is required). Due to the inverted isolation requirement in the control region CR_A most events contain more than one misidentified τ_h lepton. The statistical uncertainty of the extrapolation procedure of the α-method is significantly reduced if those events are kept.

In Fig. 7.11, the p_T^t distributions in the control region CR_A are shown for the OS and SS categories. To validate the quality of the definition of the control region CR_A the fraction of the $t\bar{t}_f$ and W+jets backgrounds as well as the signal contamination are determined. The purity of $t\bar{t}_f$ events with respect to the total number of background events is 84% (82%) in the SS (OS) category. By adding the W+jets events, the total fraction of the background of interest ends up with 89% (90%). The fraction of misidentified τ_h leptons of all backgrounds in the control region CR_A

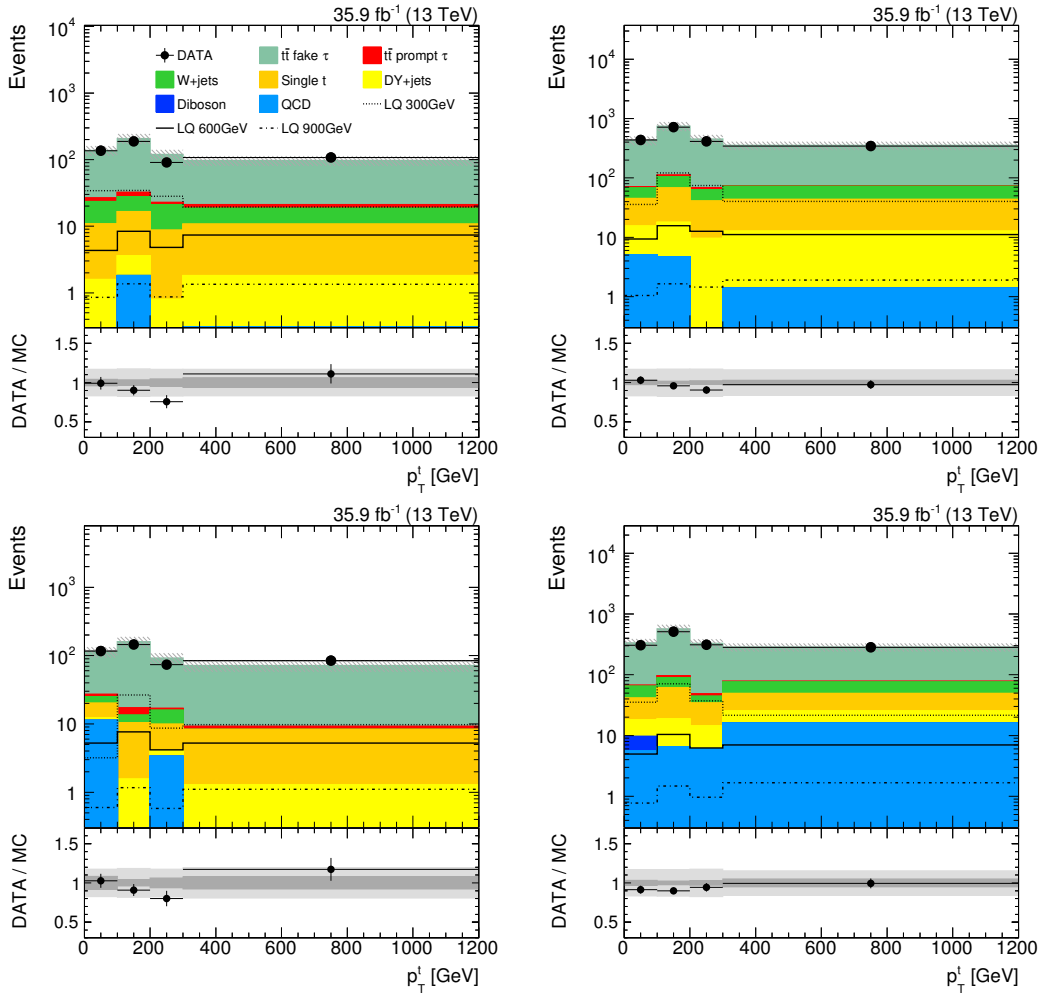

Figure 7.11: Distributions of $p_{\mathrm{T}}^{\mathrm{t}}$ in the control region for the OS (top left) and SS (top right) search regions in the $\mu + \tau_{\mathrm{h}} +$ jets final state and in the $e + \tau_{\mathrm{h}} +$ jets final state (bottom).

is 99% (98%) in the SS (OS) category.

Furthermore, by inversion of the isolation requirement of the τ_{h} lepton the LQ signal contamination in the control region $\mathrm{CR_A}$ is suppressed significantly. The signal contamination is less than 10% for low LQ masses and less than 2.5% for LQ masses greater than 700 GeV. The fraction of misidentified τ_{h} leptons in the signal processes in the control region $\mathrm{CR_A}$ is in the order of 60-70% (41-57%) in the SS (OS) category and depends on the LQ mass. The fraction of misidentified τ_{h} leptons increases for higher LQ masses. As a comparison, the fraction of misidentified τ_{h} leptons in signal processes in the signal region is low and less than 3% (1%) in the SS (OS) category.

The fraction of events from $t\bar{t}_{\mathrm{f}}$ processes and W+jets processes is required to be similar between signal region and control region $\mathrm{CR_A}$ to derive both backgrounds simultaneously. Otherwise, a different fraction of background events may be extrapolated from the control region to the signal region. The ratio of $t\bar{t}_{\mathrm{f}}$ to W+jets events has been derived from simulated events. In the OS search region, this ratio is 0.104 ± 0.013 in the signal region and 0.113 ± 0.004 in the control region. In the SS search region, the values of this ratio are 0.121 ± 0.014 in the signal region and 0.127 ± 0.002 in the control region. The ratios are in good agreement between signal region and control region in both search regions.

Following the strategy of the α-method, the next step is to investigate potential kinematic differences between signal region and control region. For this purpose, the shape of the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions for simulated $t\bar{t}_{\mathrm{f}}$ and W+jets events is compared between both regions. This shape comparison is shown in Fig. 7.12. A trend can be observed in the OS and SS search regions as the $p_{\mathrm{T}}^{\mathrm{t}}$ distribution is slightly harder in the signal region than in the control region, respectively.

This trend is found to depend on the jet multiplicity. For this reason, the shape comparison is studied for different jet multiplicities. These studies have been performed for events passing the baseline selection. A detailed description of the studies can be found in App. B.1. Correction factors have been determined as a function of the jet multiplicity and $p_{\mathrm{T}}^{\mathrm{t}}$. The final correction factors in the $\mu + \tau_{\mathrm{h}} +$ jets final state are shown in Fig. 7.13. One can observe that the correction factors are close to 1 for events that contain three jets. The values of the correction factors increase to up to 1.5 for high jet multiplicities and for high values of $p_{\mathrm{T}}^{\mathrm{t}}$. In the following, the correction factors are applied to all events in the control region $\mathrm{CR_A}$.

The shape comparison of the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions between signal region and control region for simulated $t\bar{t}_{\mathrm{f}}$ and W+jets events after the reweighting procedure is applied is shown in Fig. 7.14. Here, all events passed the final selection criteria. It can be observed that the shapes of the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions are in a good agreement between the signal region and the control region.

For the extrapolation of $t\bar{t}_{\mathrm{f}}$ and W+jets events from the control region to the signal region, a corresponding extrapolation factor has to be determined. A small value of the extrapolation factor indicates that the number of $t\bar{t}$ and W+jets events passing the selection in the control region is larger compared to the signal region. Thus, more data events are expected in the control region than in the signal region.

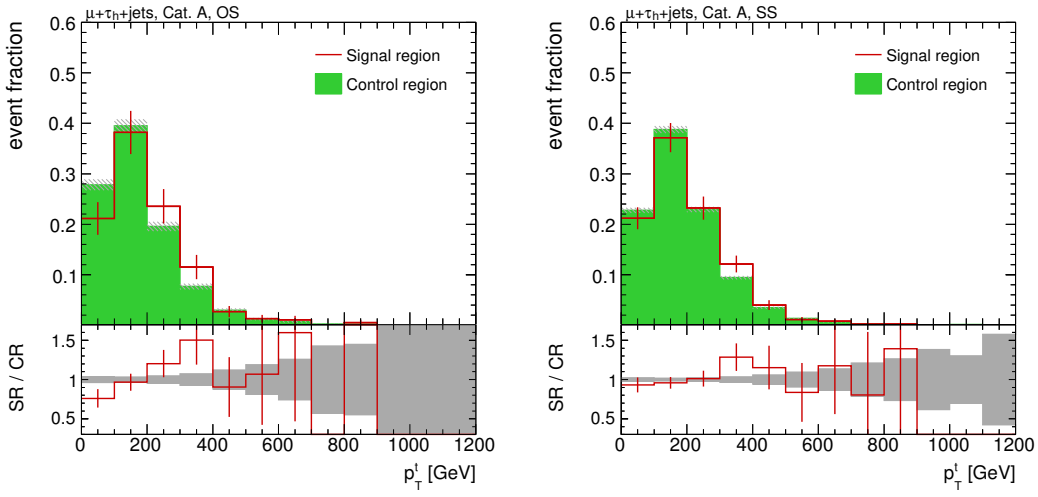

Figure 7.12: Shape comparison between signal region ($\mathrm{SR_A}$) and control region ($\mathrm{CR_A}$) of the p_T^t distribution. The comparison is shown inclusively in S_T after applying the full selection in the OS (left) and the SS (right) category in the $\mu + \tau_\mathrm{h} +$ jets final state.

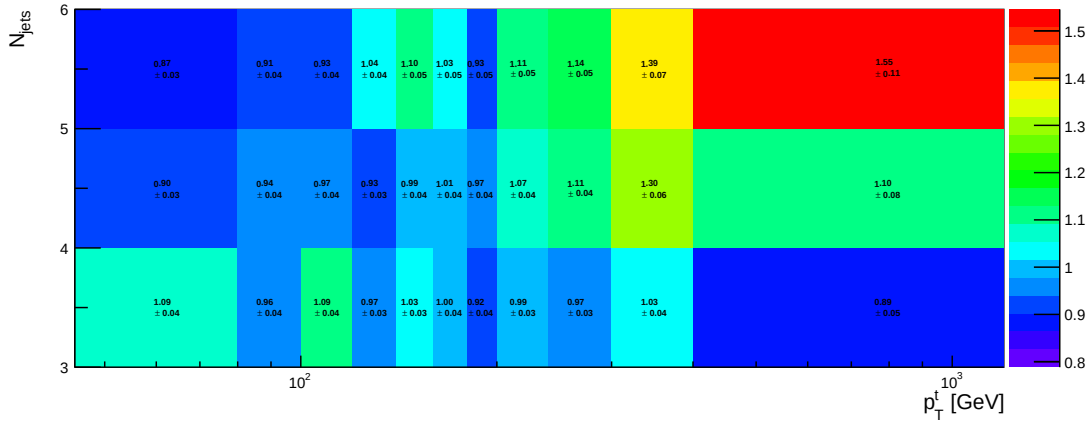

Figure 7.13: Correction factors applied in the control region ($\mathrm{CR_A}$) of the μ+τ_h+jets final state as a function of p_T^t and the jet-multiplicity.

A large number of data events in the control region leads to smaller statistical uncertainties in the final $t\bar{t}_\mathrm{f}$ and W+jets background as the data events will be extrapolated from control region to signal region. In category A, the extrapolation factor α_A is determined using simulated $t\bar{t}_\mathrm{f}$ and W+jets events and is defined as

$$\alpha_\mathrm{A} = \frac{\sum t\bar{t}_\mathrm{f}(\mathrm{SR_A}) + \sum W\text{+jets}(\mathrm{SR_A})}{\sum t\bar{t}_\mathrm{f}(\mathrm{CR_A}) + \sum W\text{+jets}(\mathrm{CR_A})}, \tag{7.2}$$

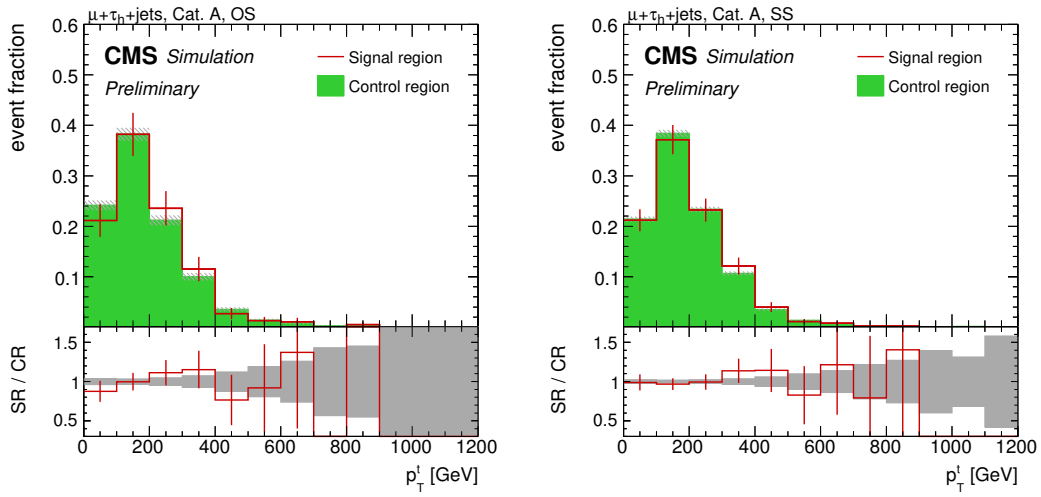

Figure 7.14: Shape comparison between signal region ($\mathrm{SR_A}$) and control region ($\mathrm{CR_A}$) of the p_T^t distribution after applying the correction factors. The comparison is shown inclusively in S_T after applying the full selection in the $\mu + \tau_\mathrm{h} +$ jets final state in the OS (left) and the SS (right) category. Published in Ref. [1].

where the sums denote the total number of events in $t\bar{t}_\mathrm{f}$ or W+jets processes in the signal region or control region. Values of $\alpha_\mathrm{A} = 0.052 \pm 0.002$ (0.090 ± 0.004) and $\alpha_\mathrm{A} = 0.060 \pm 0.002$ (0.084 ± 0.005) are determined in the SS (OS) category of the $\mu + \tau_\mathrm{h} +$ jets and $e + \tau_\mathrm{h} +$ jets final state, respectively.

Finally, after the kinematic differences between the signal region and the control region have been corrected, the following equation is used to derive the $t\bar{t}_\mathrm{f}$ and W+jets background in the $\mathrm{SR_A}$ region:

$$N_{\mathrm{SR_A}}^{t\bar{t}_\mathrm{f}\,+\,W\text{+jets}} = (N_{\mathrm{CR_A}}^{\mathrm{data}} - N_{\mathrm{CR_A}}^{\mathrm{other,MC}}) \cdot \alpha_\mathrm{A}. \tag{7.3}$$

In this equation, N is the total number of events for the respective process in the signal region or control region. Other denotes all non-$t\bar{t}_\mathrm{f}$ and non-W+jets background processes estimated from simulation. The other backgrounds are taken into account by subtracting them from data in the control region $\mathrm{CR_A}$. This background subtraction is required because the fraction of the other backgrounds is in the order of 10% and would falsely be extrapolated to the signal region.

In App. B.2, the background estimation procedure in category A is validated in

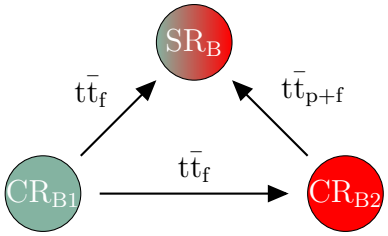

Figure 7.15: Sketch of the background estimation strategy in category B. The $t\bar{t}_{\mathrm{f}}$ background in the signal region is derived from the control region $\mathrm{CR_{B1}}$. The $t\bar{t}_{\mathrm{p+f}}$ background in the signal region is derived from the control region $\mathrm{CR_{B2}}$. To obtain an estimate of the $t\bar{t}_{\mathrm{f}}$ background in the control region $\mathrm{CR_{B2}}$, the control region $\mathrm{CR_{B1}}$ is used.

a data control sample with $p_{\mathrm{T}}^{\tau_{\mathrm{h}}} < 100\,\mathrm{GeV}$ and $S_{\mathrm{T}} < 1200\,\mathrm{GeV}$, which has negligible signal contamination. Good agreement between the data-driven background estimates and observed yields in the control sample is found for the whole $p_{\mathrm{T}}^{\mathrm{t}}$ range.

Category B

In category B, the dominant background originates again from $t\bar{t}$ production. The fraction of misidentified electrons and muons is found to be negligible in this analysis so that at least one of the τ_{h} leptons is mimicked by a jet. Thus, $t\bar{t}$ events contain either only misidentified τ_{h} leptons, or at least one prompt τ_{h} lepton and at least one misidentified τ_{h} lepton. For both, the $t\bar{t}_{\mathrm{f}}$ and the $t\bar{t}_{\mathrm{p+f}}$ background, separated control regions are defined and the α-method is utilized to estimate these backgrounds.

In Fig. 7.15, the background estimation strategy in category B is illustrated. The $t\bar{t}_{\mathrm{f}}$ background is estimated in a control region referred to as $\mathrm{CR_{B1}}$. Background events from $t\bar{t}_{\mathrm{p+f}}$ processes are derived from a control region referred to as $\mathrm{CR_{B2}}$. As the contamination of $t\bar{t}_{\mathrm{f}}$ events in the control region $\mathrm{CR_{B2}}$ is large, this background is extrapolated from control region $\mathrm{CR_{B1}}$ to control region $\mathrm{CR_{B2}}$.

A correction for kinematic differences between signal region and control region is not necessary in category B because a counting experiment is performed. All other background processes are found to have only small contributions. They are estimated from simulation.

Estimation of the $t\bar{t}_{\mathrm{f}}$ background

Similar to the control region in category A, an inverted isolation requirement of the τ_{h} leptons is the key to enrich the events from $t\bar{t}$ processes with falsely identified

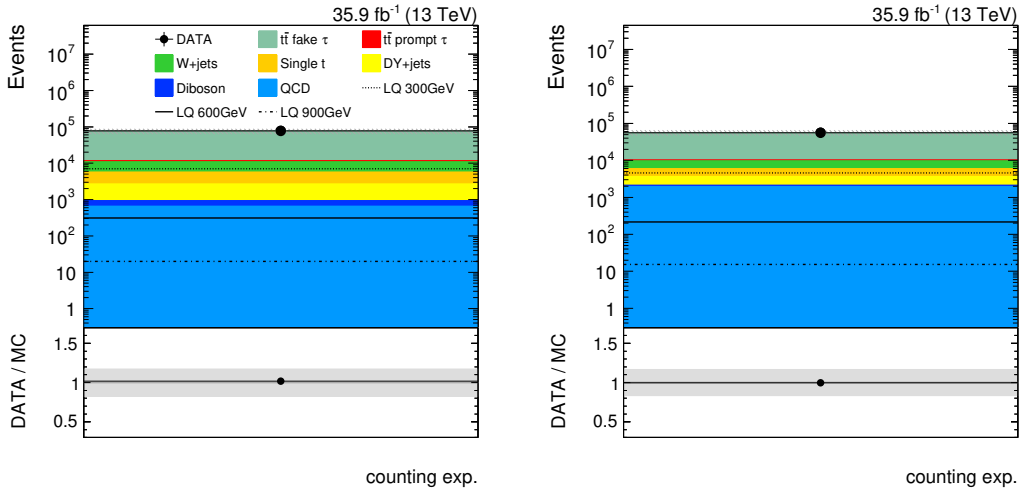

Figure 7.16: Number of events in the control region $\mathrm{CR_{B1}}$ in the $\mu + \tau_\mathrm{h} +$ jets final state (left) and in the $e + \tau_\mathrm{h} +$ jets final state (right).

τ_h leptons. In category B, a first control region $\mathrm{CR_{B1}}$ is defined in which the isolation requirement of all τ_h leptons is inverted with respect to the applied isolation requirement in the signal region. The control region $\mathrm{CR_{B1}}$ is used to derive the $t\bar{t}_\mathrm{f}$ background. The following requirements are changed and added with respect to the selection criteria in the signal region:

- at least two τ_h lepton candidates are present, but without any charge criterion on the leading and subleading τ_h lepton, and
- a requirement on the transverse momenta of τ_h leptons of $p_\mathrm{T}^{\tau_\mathrm{h}} < 100\,\mathrm{GeV}$ is fulfilled.

As the charge sign of a pair of misidentified τ_h leptons is arbitrary, this requirement is removed. The second requirement is applied in order to avoid overlap with the control region $\mathrm{CR_A}$ and to reduce the signal contamination. The final number of events in the control region $\mathrm{CR_{B1}}$ is shown in Fig. 7.16. The fraction of events from $t\bar{t}_\mathrm{f}$ processes is 85% (82%) in the $\mu + \tau_\mathrm{h} +$ jets ($e + \tau_\mathrm{h} +$ jets) final state. The signal contamination for the LQ sample of a mass of 300 GeV is 9% (8%) and drops with increasing LQ masses.

A new factor α_B1 is defined in the extrapolation procedure:

$$\alpha_\mathrm{B1} = \frac{\sum t\bar{t}_\mathrm{f}(\mathrm{SR_B})}{\sum t\bar{t}_\mathrm{f}(\mathrm{CR_{B1}})}. \tag{7.4}$$

The value of α_{B1} is $(4.24 \pm 1.13)\cdot 10^{-5}$ ($(4.62 \pm 1.40)\cdot 10^{-5}$) in the $\mu + \tau_{\mathrm{h}} +$ jets ($e+\tau_{\mathrm{h}}+$jets) final state. Equivalent to Eq. 7.3, the non-$t\bar{t}_{\mathrm{f}}$ background is subtracted from data in the $\mathrm{CR_{B1}}$ before extrapolating to the signal region. The final equation

$$N_{\mathrm{SR_B}}^{t\bar{t}_{\mathrm{f}}} = (N_{\mathrm{CR_{B1}}}^{\mathrm{data}} - N_{\mathrm{CR_{B1}}}^{\mathrm{other,MC}}) \cdot \alpha_{\mathrm{B1}} \tag{7.5}$$

is used to extrapolate the $t\bar{t}_{\mathrm{f}}$ events from control region $\mathrm{CR_{B1}}$ to the signal region. In contrast to category A, the W+jets background is taken from simulation and considered by $N_{\mathrm{CR_{B1}}}^{\mathrm{other,MC}}$.

Estimation of the $t\bar{t}_{\mathrm{p+f}}$ background

A second control region $\mathrm{CR_{B2}}$ is defined to estimate the $t\bar{t}_{\mathrm{p+f}}$ background. For this purpose, the following selection in addition to the baseline selection is applied:

- At least one isolated τ_{h} lepton and at least one nonisolated τ_{h} lepton is present in an event,
- no charge criterion on the leading and subleading τ_{h} lepton is applied,
- a requirement on the transverse momenta of τ_{h} leptons of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}} < 100\,\mathrm{GeV}$ is applied,
- the leading lepton ℓ + leading τ_{h} lepton ($\ell^1\tau_{\mathrm{h}}^1$) pair is required to have OS charge in an event,
- a requirement on the transverse mass $M_{\mathrm{T}}(\ell, \not{E}_{\mathrm{T}}) > 100\,\mathrm{GeV}$ is fulfilled, where $M_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\ell}\not{E}_{\mathrm{T}}(1 - \cos(\Delta\varphi(\vec{\ell}, \vec{\not{E}}_{\mathrm{T}})))}$, and
- at least one b tagged AK4 jet is present passing the medium identification criteria is present.

The requirement on the transverse momentum of τ_{h} leptons leads to low signal contaminations.

The $\ell^1\tau_{\mathrm{h}}^1$ pair in $t\bar{t}_{\mathrm{p+f}}$ events has OS charge if both the lepton ℓ and the τ_{h} lepton are prompt leptons. In case the leading τ_{h} lepton is mimicked by a jet, the charge of the $\ell^1\tau_{\mathrm{h}}^1$ pair is arbitrary. Thus, approximately 75% of the $t\bar{t}_{\mathrm{p+f}}$ events contain an $\ell^1\tau_{\mathrm{h}}^1$ pair with OS charge. This fraction is approximately 50% for other background processes as events from those processes contain mainly misidentified τ_{h} leptons.

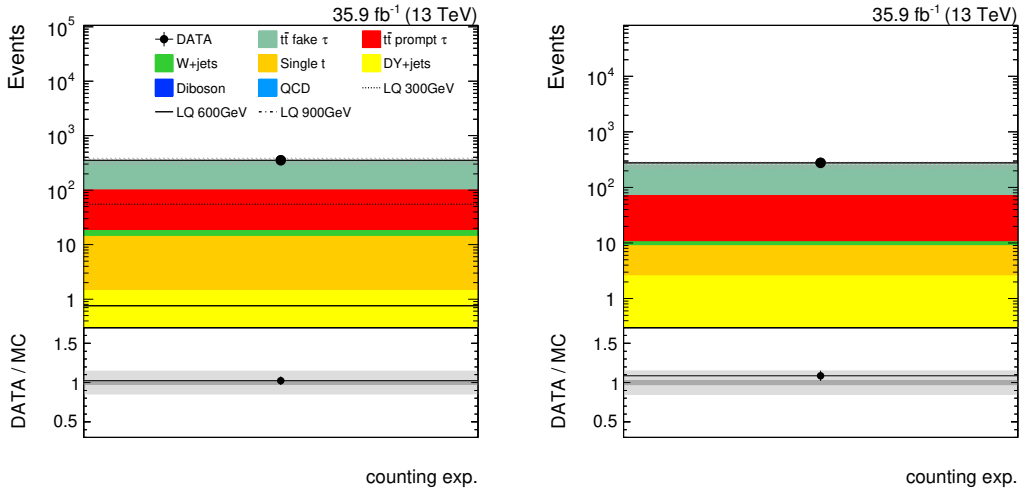

Figure 7.17: Number of events in the control region $\mathrm{CR_{B2}}$ in the $\mu + \tau_\mathrm{h} +$ jets final state (left) and in the $e + \tau_\mathrm{h} +$ jets final state (right).

The value of the M_T variable is low if the difference of the azimuth angle between the lepton ℓ and the missing transverse energy projection $\Delta\varphi(\vec{\ell}, \vec{\not{E}}_\mathrm{T})$ is small. Events with low contributions of missing transverse energy $\not{E}_\mathrm{T}$ are also expected to have low values of M_T. Events of $t\bar{t}_\mathrm{p+f}$ processes contain at least two neutrinos. This leads to high values of $\not{E}_\mathrm{T}$ and arbitrary values of $\Delta\varphi(\vec{\ell}, \vec{\not{E}}_\mathrm{T})$, therefore large values of M_T are expected in $t\bar{t}_\mathrm{p+f}$ events. For other background processes low values of M_T are expected, because most of those events consist only of misidentified τ_h leptons. E.g. in $t\bar{t}_\mathrm{f}$ events this leads to exactly one neutrino in the final state, which results in low values of $\not{E}_\mathrm{T}$. As in that case the neutrino and the charged lepton ℓ arise from the decay of a W boson, small angles of $\Delta\varphi(\vec{\ell}, \vec{\not{E}}_\mathrm{T})$ are expected. Therefore, a requirement is applied to the M_T variable.

Finally, at least one b tag is required to reduce the contamination of events from non-$t\bar{t}$ processes in the control region $\mathrm{CR_{B2}}$.

The resulting number of events in the $\mu + \tau_\mathrm{h} +$ jets and $e + \tau_\mathrm{h} +$ jets final states after applying the selection in the control region $\mathrm{CR_{B2}}$ are shown in Fig. 7.17. The signal contamination for the LQ sample of a mass of 300 GeV is 16% and decreases for higher LQ masses. The purity of events from $t\bar{t}_\mathrm{p+f}$ processes is 23% with respect to the total amount of expected SM background events. This fraction can not be enriched sufficiently by additional selection criteria because the selection efficiency of $t\bar{t}_\mathrm{p+f}$ events would become too low otherwise. By performing the α-

method under this condition, a large fraction of simulated events from non relevant background processes would be subtracted from data, leading to large uncertainties in the extrapolation procedure. Moreover, the prediction of the $t\bar{t}_{\mathrm{p+f}}$ background would not be reliable. Thus, the α-method can not be utilized unless modifications are applied.

The largest fraction of contamination comes from events of $t\bar{t}_{\mathrm{f}}$ processes and is 72%, whereas most events from other background processes are rejected. A reliable prediction of the $t\bar{t}_{\mathrm{f}}$ background is determined by utilizing the control region $\mathrm{CR_{B1}}$ and the α-method again. The $t\bar{t}_{\mathrm{f}}$ events in the control region $\mathrm{CR_{B2}}$ are derived from control region $\mathrm{CR_{B1}}$ with this approach. For this purpose, another extrapolation factor is introduced:

$$\alpha_{\mathrm{B1toB2}} = \frac{\sum t\bar{t}_{\mathrm{f}}(\mathrm{CR_{B2}})}{\sum t\bar{t}_{\mathrm{f}}(\mathrm{CR_{B1}})}. \tag{7.6}$$

A value of $\alpha_{\mathrm{B1toB2}} = (3.8 \pm 1.1) \cdot 10^{-3}$ $((4.0 \pm 1.2) \cdot 10^{-3})$ is determined in the $\mu + \tau_{\mathrm{h}} + \mathrm{jets}$ $(e + \tau_{\mathrm{h}} + \mathrm{jets})$ final state. Eq. 7.5 is used for the extrapolation from $\mathrm{CR_{B1}}$ to $\mathrm{CR_{B2}}$, by replacing the extrapolation factor α_{B1} with α_{B1toB2}. This extrapolation leads to the estimation of the $t\bar{t}_{\mathrm{f}}$ background in control region $\mathrm{CR_{B2}}$.

Finally, the events from $t\bar{t}_{\mathrm{p+f}}$ processes are extrapolated from the $\mathrm{CR_{B2}}$ to the signal region. The corresponding extrapolation factor is defined as

$$\alpha_{\mathrm{B2}} = \frac{\sum t\bar{t}_{\mathrm{p+f}}(\mathrm{SR_B})}{\sum t\bar{t}_{\mathrm{p+f}}(\mathrm{CR_{B2}})} \tag{7.7}$$

and yields 0.025 ± 0.008 (0.025 ± 0.009) in the $\mu + \tau_{\mathrm{h}} + \mathrm{jets}$ $(e + \tau_{\mathrm{h}} + \mathrm{jets})$ final state. The final event yield of the $t\bar{t}_{\mathrm{p+f}}$ background is obtained by

$$N_{\mathrm{SR_B}}^{t\bar{t}_{\mathrm{p+f}}} = (N_{\mathrm{CR_{B2}}}^{\mathrm{data}} - N_{\mathrm{CR_{B2}}}^{\mathrm{other}}) \cdot \alpha_{\mathrm{B2}}. \tag{7.8}$$

In this equation, $N_{\mathrm{CR_{B2}}}^{\mathrm{other}}$ includes the $t\bar{t}_{\mathrm{f}}$ background derived from data in the control region $\mathrm{CR_{B1}}$, and the remaining background events obtained from simulation.

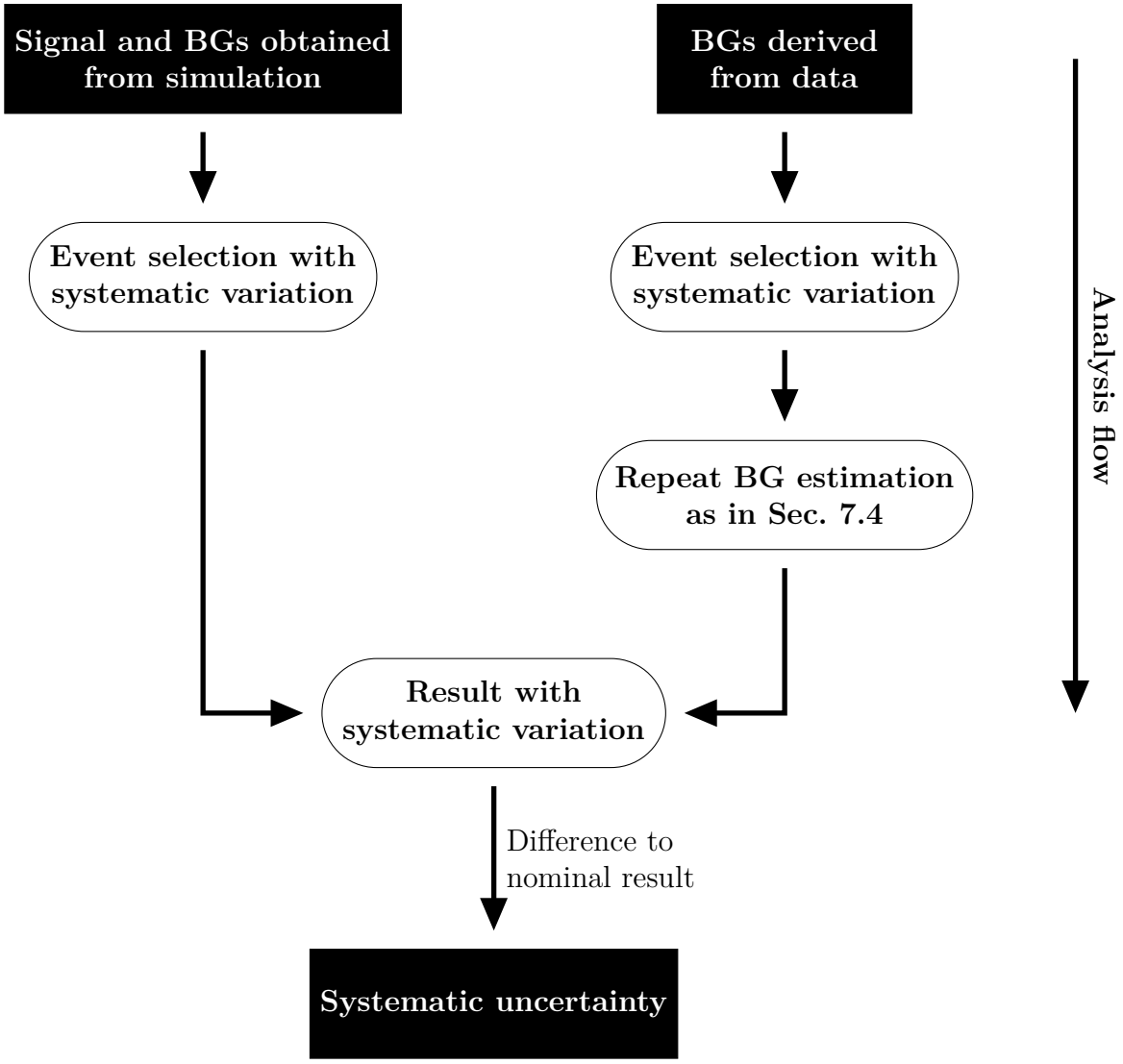

Figure 7.18: Diagram of the determination of systematic uncertainties.

7.5 Systematic Uncertainties

In this section, the systematic uncertainties affecting the results of the presented analysis are discussed. Systematic uncertainties can vary the normalization as well as the shape of the distribution of physics observables.

Fig. 7.18 shows a sketch how the systematic uncertainties are obtained in this analysis. For each source of systematic uncertainty, the full analysis chain is performed. This means that the full event selection is used, but a systematic variation is applied to all simulated events. For the backgrounds that are derived from data

control regions, the complete background estimation procedure is repeated. Thus, the systematic variations enter the extrapolation factors and the minor backgrounds in the α-method. Finally, a new result is obtained in which a systematic variation is applied and the difference to the nominal result without applying a variation corresponds to the systematic uncertainty.

A list of all systematic uncertainties taken into account in this analysis is given in the following.

- The uncertainty in the luminosity measurement recorded with the CMS detector is 2.5% [139].
- Depending on the background process, the following uncertainties in the production cross sections are taken:
  - 5.6% [143] in the $t\bar{t}$ production cross section for $t\bar{t}$ events that include only prompt τ_{h} leptons.
  - 10% in the production of single top quark [144], W+jets, and Z+jets [145].
  - 20% in the production of diboson events [146, 147].
- The estimation of pileup effects is based on the total inelastic cross section. This cross section is determined to be 69.2 mb. The uncertainty is taken into account by varying the total inelastic cross section by $\pm 5\%$ [148, 149].
- Muon scale factors derived in Ref. [107] are considered. The simulated events are reweighted as a function of $|\eta|$ and p_{T} and the corresponding systematic uncertainty is taken into account by varying each scale factor within its uncertainty. Uncertainties are applied to:
  - the muon trigger efficiency scale factor,
  - the muon scale factor addressing the relative isolation of the muon,
  - the muon ID scale factor, and
  - the muon scale factor due to the tracking efficiency.
- Electron scale factors [111] are considered and taken into account by utilizing the same procedure applied on the uncertainties in the muon scale factors. Uncertainties are applied to:
  - the electron trigger efficiency scale factor,

  - the electron reconstruction scale factor, and
  - the electron ID scale factor.

- The jet energy scale (JEC) corrections are determined as a function of $|\eta|$ and p_{T} [120]. This uncertainty is considered by varying the energy scale within their uncertainties. This variation is propagated to the calculation of $\not\!\!E_{\mathrm{T}}$. A variation of the jet energy scale is made before applying selection criteria to the transverse momentum and η.

- The jet energy resolution (JER) is determined as a function of $|\eta|$ and scale factors have been determined in Ref. [120]. The uncertainty of the jet energy resolution is determined by varying the scale factors within their uncertainties. This variation is also propagated to the calculation of $\not\!\!E_{\mathrm{T}}$. A variation of the jet energy resolution is made before applying selection criteria to the transverse momentum and η.

- Scale factors for the b tagging efficiencies [150] are applied. These scale factors are measured as a function of the jet p_{T}. The corresponding uncertainty is taken into account by varying the scale factors within their uncertainties.

- Following the recommendations as in Ref. [114], three uncertainties regarding the τ_{h} lepton reconstruction are considered:
  - An uncertainty of 5% in the τ_{h} lepton identification is applied to each prompt τ_{h} lepton. Due to the occurrence of boosted τ_{h} leptons in this analysis, an additional uncertainty of $20\% \cdot p_{\mathrm{T}}/1\,\mathrm{TeV}$ is applied.
  - An uncertainty of 3% in the τ_{h} lepton energy scale is applied to each prompt τ_{h} lepton, scaling the four momentum of the τ_{h} lepton by a factor of $k = 1 \pm 0.03$.
  - Since the information of the τ_{h} lepton charge is used due to the categorization into OS and SS events, an uncertainty in the charge misidentification rate of 2% is applied to each prompt τ_{h} lepton.

- Parton density functions (PDFs) are used to generate simulated events for background and signal samples. The PDF uncertainties are computed following the recommendations of the PDF4LHC group [151]. A set of 100 varied PDFs is provided, which leads to 100 variations of the final distribution in each search region. The standard deviation of the 100 distributions is determined

	Category A			Category B		
Uncertainty	$t\bar{t}_{\mathrm{p}}$	$t\bar{t}_{\mathrm{f}} + W$+jets	LQ	$t\bar{t}_{\mathrm{f}}$	$t\bar{t}_{\mathrm{p+f}}$	LQ
μ_{f}, μ_{r}	$26-42\%$	$1-7\%$	–	$5-7\%$	$2-6\%$	–
τ ID	$8-9\%$	$0-1\%$	$9-11\%$	0%	$5-6\%$	$18-20\%$
BG estimation	–	$6-18\%$	–	$26-30\%$	$30-38\%$	–

Table 7.5: Summary of largest systematic uncertainties for the $t\bar{t}_{\mathrm{p}}$ background obtained from simulation, for the $t\bar{t}_{\mathrm{f}}$ (and W+jets) and $t\bar{t}_{\mathrm{p+f}}$ backgrounds derived from data, and for an LQ signal sample with a mass of 700 GeV.

in each bin with respect to the nominal result. The PDF uncertainty is taken into account by varying the nominal result within its standard deviation. The associated PDF uncertainties in the signal acceptance are estimated following the PDF uncertainty prescription for the LHC [151]. For the background taken from simulation, uncertainties in the acceptance and the normalization are determined.

- Uncertainties in the renormalization (μ_r) and factorization (μ_f) scales are considered by varying the respective scales by factors between 0.5 and 2. All possible combinations of variations of μ_r and μ_f are taken into account and the envelope of the different variations is used. These uncertainties are only considered for the background processes.

- An uncertainty in the background estimation method is applied by varying the extrapolation factors α_{A}, α_{B1}, α_{B1toB2} and α_{B2} within their uncertainties. An uncertainty due to kinematic differences between the signal region $\mathrm{SR_A}$ and the control region $\mathrm{CR_A}$ is taken into account by varying the correction factors within their uncertainties.

In Tab. 7.5 the systematic uncertainties causing the largest effects on the most important background processes and on a signal sample with a mass of 700 GeV are summarized. A range of the uncertainty values is provided since the uncertainties can differ between the search regions in category A and between the $\mu + \tau_{\mathrm{h}} +$ jets and $e + \tau_{\mathrm{h}} +$ jets final states.

The impact of the different sources of uncertainties varies for different processes. In category A, the dominant backgrounds come from $t\bar{t}_{\mathrm{p}}$ and $t\bar{t}_{\mathrm{f}}$ (and W+jets) processes. The events from $t\bar{t}_{\mathrm{p}}$ processes are taken from simulation and the uncer-

tainty in the scale variation has a large impact on this background. For the $t\bar{t}_{\mathrm{f}}$ and W+jets backgrounds obtained from the control regions most systematic uncertainties are small. In the procedure of the α-method, the uncertainties are propagated to the extrapolation factor. As the impact of systematic variations is similar in signal region and control region most uncertainties are reduced significantly due to the background estimation method. However, the uncertainty applied to the extrapolation factor is the largest due to the limited number of simulated events that pass the event selection.

The dominant backgrounds in category B arise from $t\bar{t}_{\mathrm{f}}$ and $t\bar{t}_{\mathrm{p+f}}$ processes. The uncertainty of the extrapolation factor has the largest impact to both background processes for the same reasons as in category A. In both categories, the uncertainty in the τ_{h} lepton identification scale factor has the largest impact on the LQ sample. This uncertainty has larger values in category B as at least two τ_{h} leptons are selected, whereas exactly one τ_{h} lepton is required in category A.

The final distributions of the SM background processes are determined by performing a maximum likelihood estimation (MLE) fit to constrain the normalization of the background processes to their best fit-values. The `theta` software package [152] is used in this procedure. Each systematic uncertainty is accounted for as a nuisance parameter to the MLE fit and log-normal prior distributions are used for rate uncertainties and Gaussian prior distributions otherwise. The final background estimates are determined by fitting the background-only hypothesis simultaneously in all search regions of the $\mu + \tau_{\mathrm{h}} +$ jets and $e + \tau_{\mathrm{h}} +$ jets final states to the measured $p_{\mathrm{T}}^{\mathrm{t}}$ distributions in data in category A and to the number of events in data in category B. The deviations of the post-fit parameters from their input values in units of the corresponding prior uncertainty are shown in Fig. 7.19. All nuisance parameters taken into account in the MLE fit are shown. The post-fit parameters are within 1 standard deviation (SD) of their prior uncertainty or at the edge of 1 SD and 2 SD.

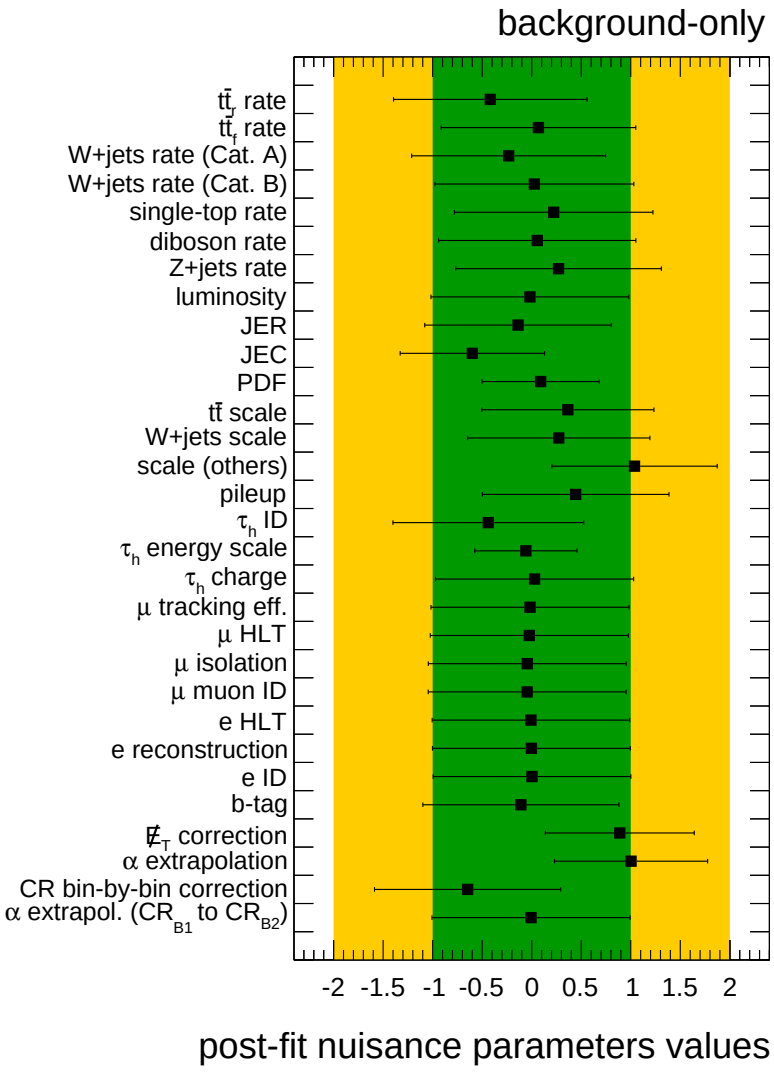

Figure 7.19: Post-fit values of the nuisance parameters of the background-only model. The post-fit parameters are shown as deviations from their pre-fit values in units of the corresponding prior uncertainty.

7.6 Results

The final post-fit distributions under background-only hypothesis in the $\mu+\tau_h+$jets and $e+\tau_h+$jets final states in category A are shown in Fig. 7.20 and Fig. 7.21, respectively. The distributions correspond to the OS and SS and to the high and low S_T search regions in category A. The contributions from $t\bar{t}$ and W+jets processes that contain a misidentified τ_h lepton are obtained from the control region CR_A. The SM backgrounds that include a prompt τ_h lepton and other minor backgrounds are

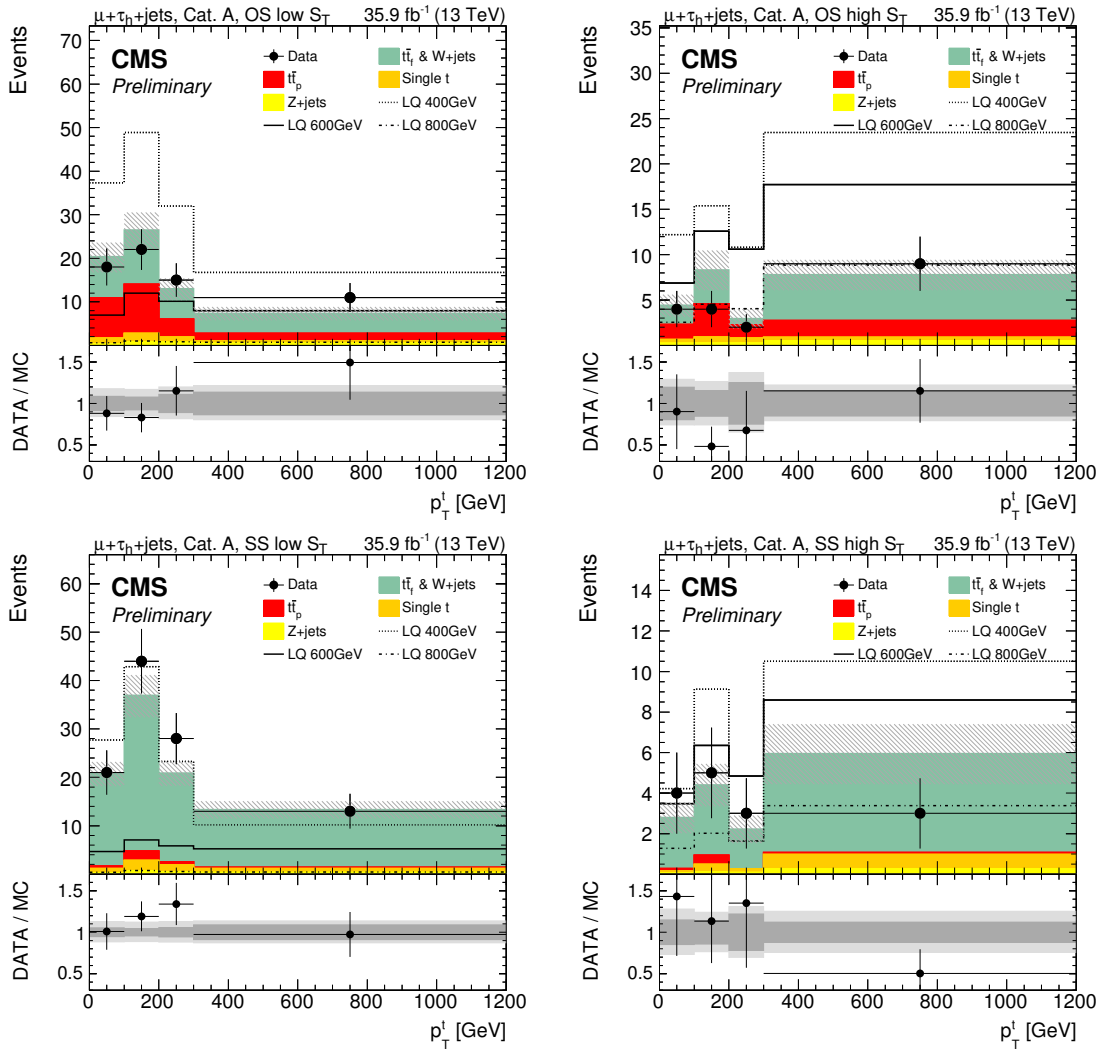

Figure 7.20: Post-fit distributions of $p_{\mathrm{T}}^{\mathrm{t}}$ in the $\mu + \tau_{\mathrm{h}} +$ jets final state after applying the full selection in category A. The events are separated into OS (top), SS (bottom), low S_{T} (left) and high S_{T} (right) search regions. In the bottom panel, the dark gray shaded area represents the statistical uncertainties and the light gray shaded area corresponds to the total uncertainties. Published in Ref. [1].

estimated from simulation. Good agreement between data and the SM background is observed. The total event yields of the collected data, the expected SM backgrounds, and the signal processes are summarized for both final states in Tab. 7.6 and Tab. 7.7, respectively. No significant deviation from the SM prediction is observed in data.

In Tab. 7.8, the total number of events in data, background processes and signal processes in category B is summarized. A counting experiment is performed as the

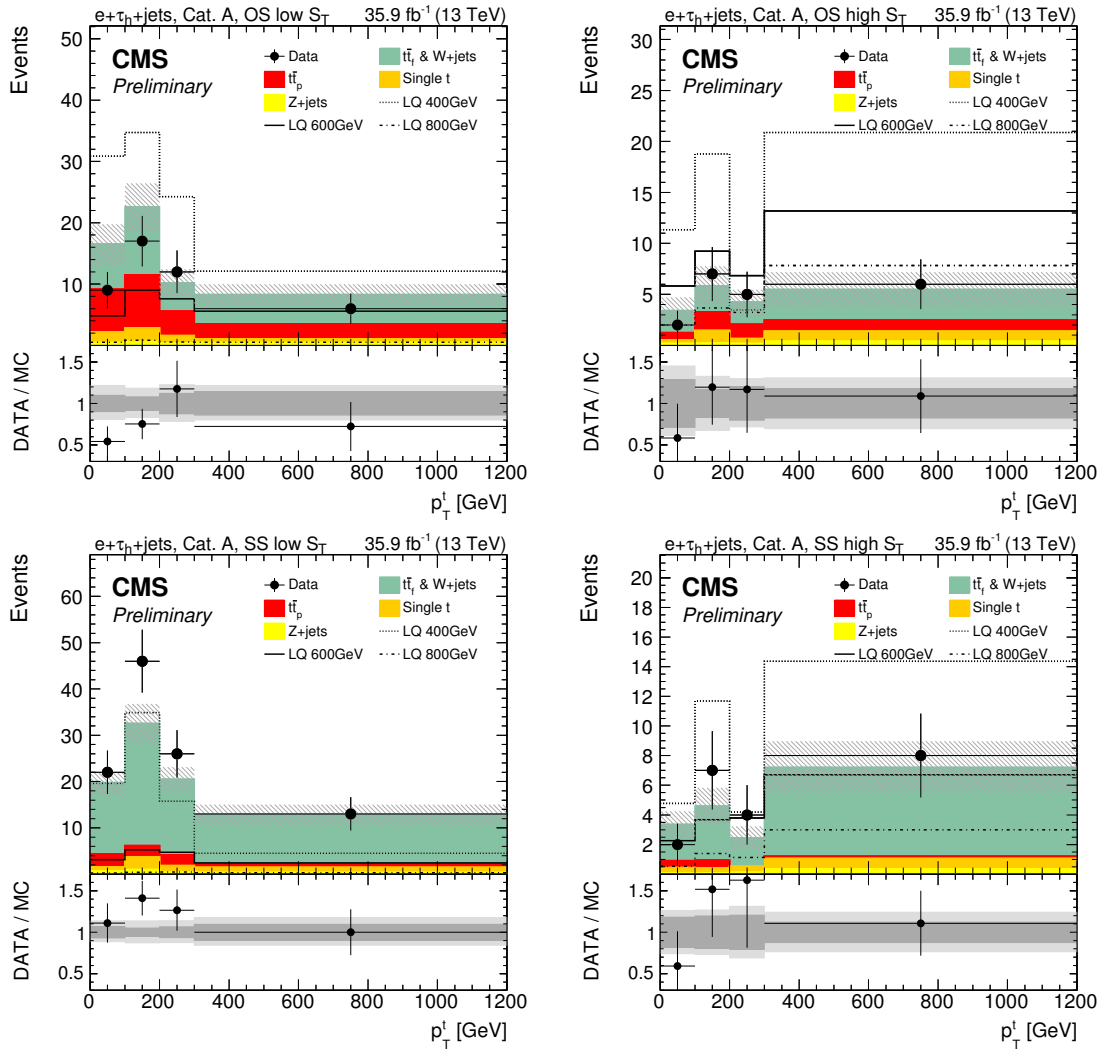

Figure 7.21: Post-fit distributions of $p_{\mathrm{T}}^{\mathrm{t}}$ in the $e + \tau_{\mathrm{h}} +$ jets final state after applying the full selection in category A. The events are separated into OS (top), SS (bottom), low S_{T} (left) and high S_{T} (right) search regions. In the bottom panel, the dark gray shaded area represents the statistical uncertainties and the light gray shaded area corresponds to the total uncertainties. Published in Ref. [1].

number of expected SM background events is too low to benefit from a shape-based analysis. Again, no significant deviation from the SM prediction is observed in data.

A bayesian method [153] is used to calculate the 95% confidence level (CL) exclusion limits on the cross section times branching ratio for pair produced scalar LQs decaying into a top quark and a tau lepton. The measured $p_{\mathrm{T}}^{\mathrm{t}}$ distributions

Sample	N (OS high S_{T})	N (OS low S_{T})	N (SS high S_{T})	N (SS low S_{T})
LQ (300 GeV)	31^{+12}_{-12}	90^{+21}_{-23}	13^{+8}_{-7}	74^{+22}_{-19}
LQ (400 GeV)	62^{+10}_{-11}	135^{+18}_{-19}	26^{+6}_{-6}	104^{+15}_{-15}
LQ (500 GeV)	57^{+8}_{-10}	94^{+11}_{-11}	26^{+4}_{-5}	59^{+7}_{-7}
LQ (600 GeV)	48^{+6}_{-7}	37^{+4}_{-5}	23^{+3}_{-4}	22^{+3}_{-3}
LQ (700 GeV)	34^{+4}_{-5}	11^{+1}_{-2}	15^{+2}_{-2}	$7.5^{+1.0}_{-1.1}$
LQ (800 GeV)	20^{+2}_{-3}	$3.1^{+0.4}_{-0.5}$	$8.3^{+1.1}_{-1.2}$	$2.2^{+0.3}_{-0.3}$
LQ (900 GeV)	11^{+1}_{-2}	$0.9^{+0.1}_{-0.1}$	$4.3^{+0.6}_{-0.7}$	$0.7^{+0.1}_{-0.1}$
LQ (1000 GeV)	$5.9^{+0.8}_{-0.9}$	$0.23^{+0.04}_{-0.04}$	$2.4^{+0.3}_{-0.4}$	$0.20^{+0.04}_{-0.04}$
$t\bar{t}_{\mathrm{f}}$ + W+jets	11^{+2}_{-2}	31^{+4}_{-5}	11^{+3}_{-3}	74^{+8}_{-7}
$t\bar{t}_{\mathrm{p}}$	12^{+5}_{-4}	33^{+10}_{-8}	$0.9^{+0.5}_{-0.5}$	$3.8^{+1.4}_{-1.2}$
Single top	$2.5^{+1.0}_{-0.8}$	$7.0^{+1.5}_{-1.7}$	$2.0^{+0.7}_{-0.7}$	$6.6^{+1.6}_{-1.5}$
Z+jets	$1.6^{+0.5}_{-0.4}$	$2.8^{+0.8}_{-0.7}$	$0.6^{+0.2}_{-0.2}$	$1.9^{+0.6}_{-0.5}$
Diboson	$0.0^{+1.2}_{-0.0}$	$0.0^{+1.2}_{-0.0}$	$0.0^{+1.2}_{-0.0}$	$0.0^{+1.2}_{-0.0}$
total bg	27^{+6}_{-5}	74^{+11}_{-10}	15^{+3}_{-3}	87^{+8}_{-8}
DATA	19	66	15	106

Table 7.6: Final event yield in each search region of category A in the $\mu + \tau_{\mathrm{h}} +$ jets final state for different LQ mass hypotheses, the background processes and data. The total uncertainties for the background and signal processes are shown.

in category A and the events measured in category B are used to perform a statistical template-based shape analysis by using the `theta` software package. Each systematic uncertainty is treated as an additional nuisance parameter in the likelihood fit. A uniform prior distribution is used for the signal cross section parameter. Log-normal prior distributions are used for the other nuisance parameters. Pseudo-experiments are performed by varying these parameters within their prior distributions to estimate the 95% CL expected limits.

A comparison of the expected 95% CL upper limits on the cross section times branching ratio squared for all categories in both final states is shown in Fig. 7.22. In category A, the low S_T search regions give the best sensitivity for LQ masses up to almost 500 GeV, whereas for higher signal masses the high S_{T} categories are dom-

Sample	N (OS high S_{T})	N (OS low S_{T})	N (SS high S_{T})	N (SS low S_{T})
LQ (300 GeV)	34^{+12}_{-18}	92^{+22}_{-21}	11^{+6}_{-6}	60^{+18}_{-17}
LQ (400 GeV)	55^{+9}_{-11}	102^{+15}_{-16}	35^{+7}_{-7}	75^{+12}_{-11}
LQ (500 GeV)	39^{+6}_{-6}	58^{+7}_{-7}	22^{+4}_{-4}	37^{+5}_{-5}
LQ (600 GeV)	35^{+5}_{-5}	27^{+3}_{-4}	16^{+2}_{-3}	16^{+2}_{-2}
LQ (700 GeV)	25^{+3}_{-4}	$8.6^{+1.1}_{-1.2}$	11^{+2}_{-2}	$5.1^{+0.7}_{-0.7}$
LQ (800 GeV)	17^{+2}_{-3}	$2.4^{+0.3}_{-0.4}$	$6.1^{+0.8}_{-0.9}$	$1.4^{+0.2}_{-0.2}$
LQ (900 GeV)	$8.6^{+1.1}_{-1.4}$	$0.6^{+0.1}_{-0.1}$	$3.5^{+0.5}_{-0.5}$	$0.4^{+0.1}_{-0.1}$
LQ (1000 GeV)	$4.5^{+0.6}_{-0.7}$	$0.20^{+0.04}_{-0.04}$	$1.8^{+0.3}_{-0.3}$	$0.11^{+0.03}_{-0.02}$
$t\bar{t}_{\mathrm{f}}$ + W+jets	$8.9^{+2.9}_{-2.5}$	25^{+4}_{-4}	12^{+3}_{-3}	63^{+8}_{-7}
$t\bar{t}_{\mathrm{p}}$	$6.6^{+3.2}_{-2.3}$	26^{+8}_{-6}	$1.4^{+0.7}_{-0.6}$	$9.3^{+2.4}_{-2.2}$
Single top	$3.5^{+1.0}_{-1.2}$	$5.9^{+1.5}_{-1.3}$	$1.9^{+0.6}_{-0.7}$	$5.9^{+1.3}_{-1.5}$
Z+jets	$1.2^{+0.3}_{-0.3}$	$2.3^{+0.6}_{-0.5}$	$0.7^{+0.2}_{-0.2}$	$2.7^{+0.8}_{-0.7}$
Diboson	$0.0^{+1.2}_{-0.0}$	$0.0^{+1.2}_{-0.0}$	$0.0^{+1.2}_{-0.0}$	$0.0^{+1.2}_{-0.0}$
total bg	20^{+5}_{-4}	60^{+9}_{-8}	16^{+3}_{-3}	81^{+8}_{-8}
DATA	20	44	21	107

Table 7.7: Final event yield in each search region of category A in the $e + \tau_{\mathrm{h}} +$ jets final state for different LQ mass hypotheses, the background processes and data. The total uncertainties for the background and signal processes are shown.

inant. The results in category B improve the sensitivity over the whole mass range, but most significantly for low signal masses, notably the mass point at 300 GeV is improved by a factor of approximately 3 due the implementation of this category.

The observed exclusion limits in the $\mu + \tau_{\mathrm{h}} +$ jets and $e + \tau_{\mathrm{h}} +$ jets final states are shown in Fig. 7.23. In these distributions, the observed 95% CL upper limits on the cross section times branching ratio squared as a function of the LQ mass are presented. The dashed lines correspond to the theoretical pair-production cross section of scalar LQ pairs at NLO accuracy [51]. The uncertainty of the cross section due to the PDFs and variations of the renormalization and factorization scales by factors of 0.5 and 2 is indicated by the dotted lines. The observed limits agree well with the expected limits and lie within the uncertainty band of 1 SD of the

Sample	N ($\mu + \tau_h +$ jets)	N ($e + \tau_h +$ jets)
LQ (300GeV)	172^{+38}_{-38}	97^{+25}_{-24}
LQ (400GeV)	101^{+19}_{-18}	74^{+14}_{-14}
LQ (500GeV)	$46.1^{+8.7}_{-8.1}$	$34.6^{+6.7}_{-6.3}$
LQ (600GeV)	$22.1^{+4.3}_{-4.0}$	$14.3^{+2.9}_{-2.7}$
LQ (700GeV)	$7.3^{+1.6}_{-1.5}$	$7.5^{+1.6}_{-1.5}$
LQ (800GeV)	$4.5^{+1.0}_{-0.9}$	$3.3^{+0.7}_{-0.7}$
LQ (900GeV)	$1.9^{+0.4}_{-0.4}$	$1.6^{+0.4}_{-0.3}$
LQ (1000GeV)	$0.9^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$
$t\bar{t}_{\mathrm{f}}$	$3.2^{+1.5}_{-1.2}$	$2.4^{+0.8}_{-1.1}$
$t\bar{t}_{\mathrm{p+f}}$	$2.1^{+0.8}_{-0.9}$	$1.6^{+0.7}_{-0.8}$
single t	$0.00^{+0.2}_{-0.0}$	$0.3^{+0.3}_{-0.3}$
W+jets	$0.4^{+0.7}_{-0.4}$	$0.5^{+1.2}_{-0.5}$
Z+jets	$1.0^{+0.4}_{-0.4}$	$1.4^{+0.5}_{-0.5}$
Diboson	$1.7^{+1.8}_{-1.7}$	$1.6^{+1.7}_{-1.6}$
total bg	$8.4^{+2.6}_{-2.4}$	$7.8^{+2.4}_{-2.5}$
DATA	11	9

Table 7.8: Post-fit event yields in category B in the $\mu+\tau_h+$jets and $e+\tau_h+$jets final states for different LQ mass hypotheses, the background processes and data. The total uncertainties for the background and the signal processes are shown.

expectation over the whole LQ mass range.

The results in the $\mu + \tau_h +$ jets final state have been combined with the results in the $e + \tau_h +$ jets final state. The outcome of the combination is shown in Fig. 7.24 (top). In Fig. 7.24 (bottom) the 95% CL observed and expected exclusion limits on the LQ mass as a function of the branching ratio is shown. Under the assumption of a branching ratio of 100% for LQs decaying into a top quark and a tau lepton pair, production cross sections of 0.6 pb for LQ masses of 300 GeV and about 0.01 pb for masses up to 1.5 TeV are excluded at 95% CL. Pair-produced third-generation scalar LQs are excluded up to masses of 900 GeV. This result improves the limits obtained

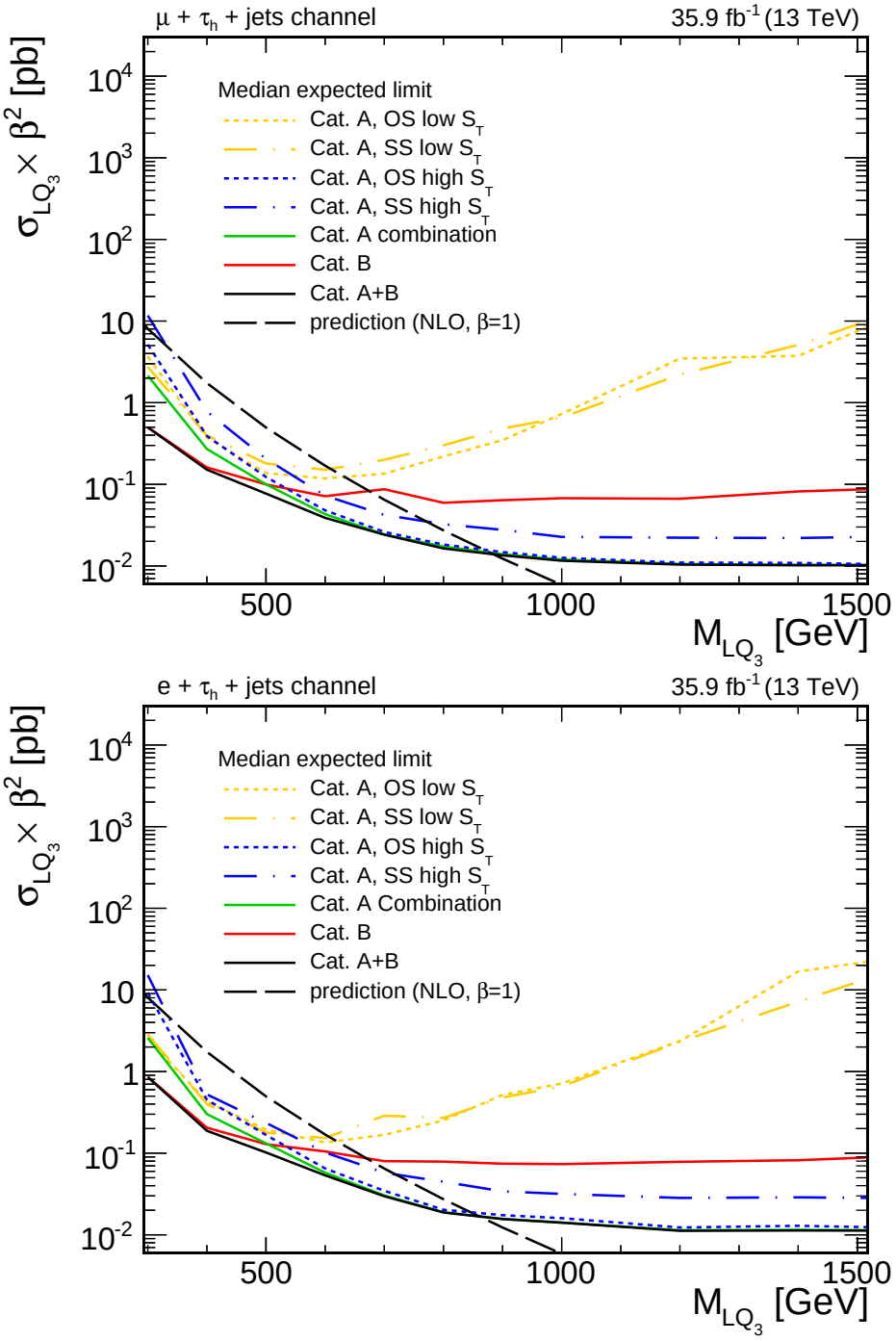

Figure 7.22: Comparison of expected limits at 95% CL in the search regions of category A and in category B in the muon (top) and electron (bottom) channels.

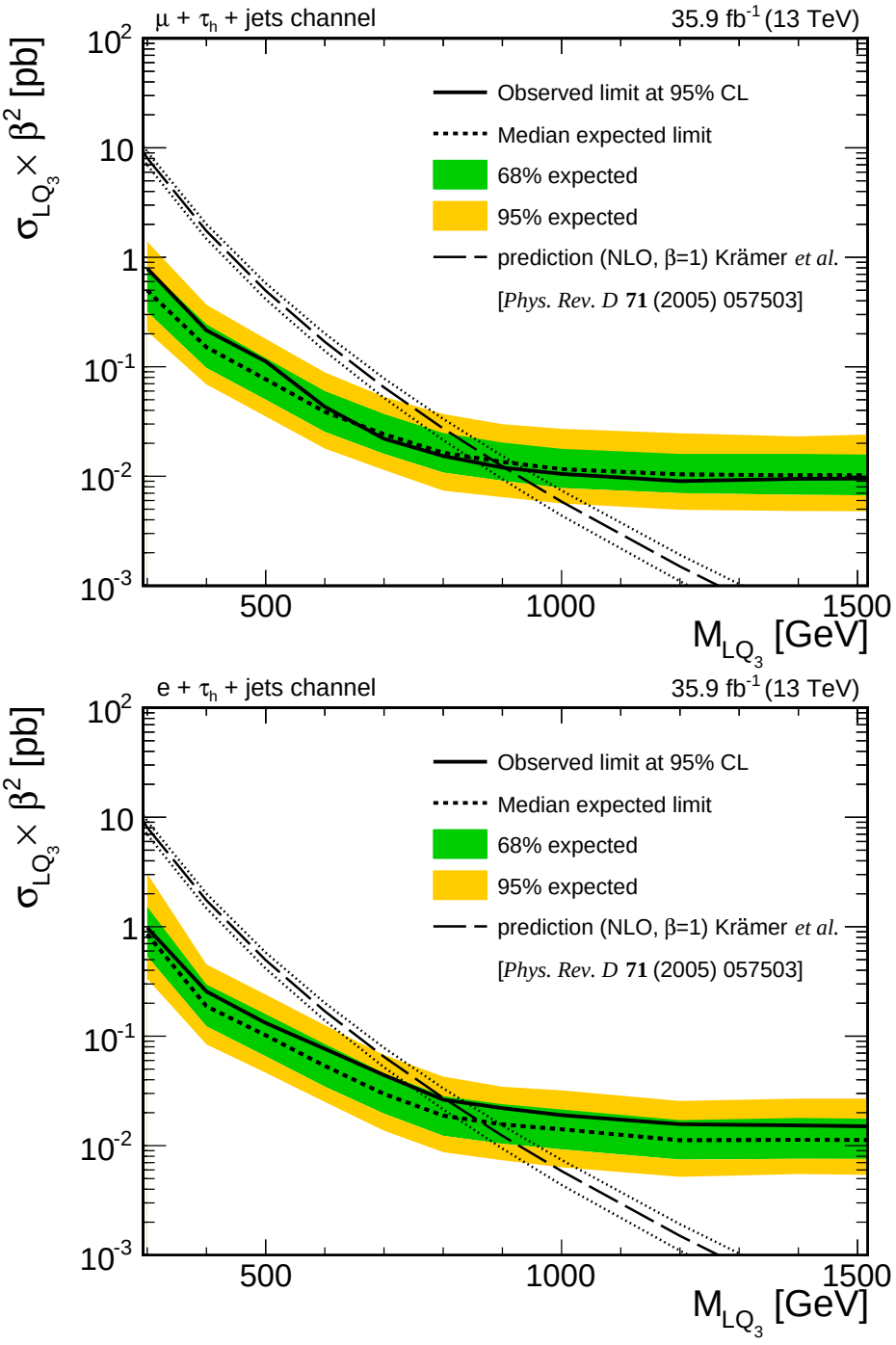

Figure 7.23: Upper limits at 95% confidence level on the cross section (σ) times branching fraction squared (β^2) on the pair production of scalar leptoquarks decaying to a top quark and a tau lepton. The results include all search categories in the μ+τ_h+jets (top) and e+τ_h+jets (bottom) channels. The theory curve corresponds to the NLO cross section with uncertainties from PDF and scale variations [51].

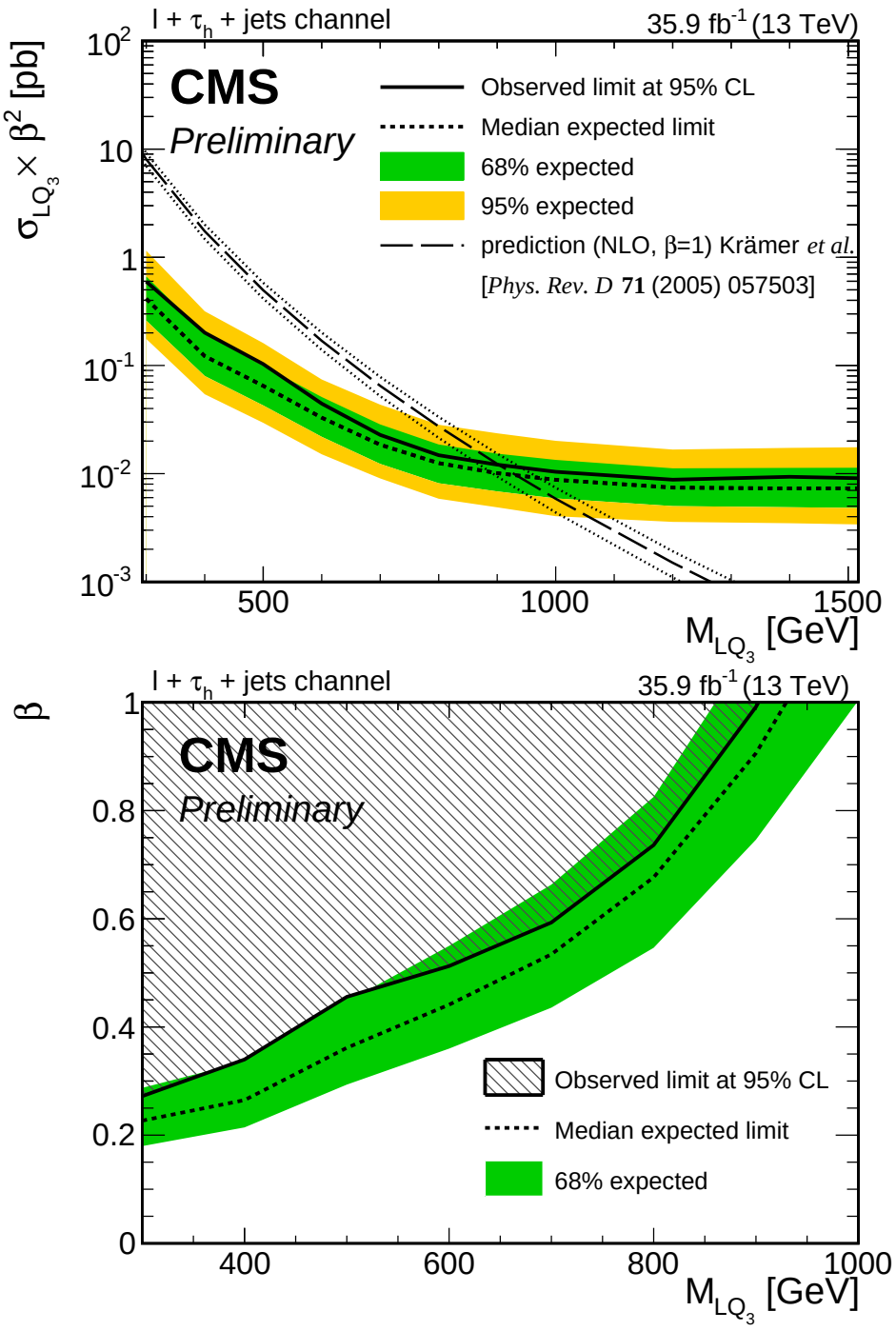

Figure 7.24: Upper limits at 95% confidence level on the cross section times branching fraction squared (top) and the LQ mass as a function of the branching ratio (bottom) on the pair production of scalar LQs decaying to a top quark and a tau lepton. The results include all search categories in the ℓ+τ_h+jets channels. The theory curve corresponds to the NLO cross section with uncertainties from PDF and scale variations. Published in Ref. [1].

by the analysis of 8 TeV data [2] by 215 GeV and represents the most stringent limit for pair-produced scalar LQs in the top quark and tau lepton decay mode to date. Furthermore, models that predict scalar leptoquarks to explain recently observed flavor anomalies observed in B meson decays (cf. Sec. 2.4) can be constrained due to the limits set in this thesis.

7.7 Outlook

The LHC has collected data during the runs in 2017 and is going to collect more data in 2018. By combining each run at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ a total integrated luminosity of more than $100\,\text{fb}^{-1}$ is expected. With this large amount of data sensitivities towards higher LQ masses can be achieved. The expected exclusion limits with prospective integrated luminosities that can be reached in the future are shown in Fig. 7.25. For the determination of each expected upper limit, the same strategy as in the analysis presented in this thesis is used and systematic uncertainties are taken into account, and all signal and background events are scaled to the respective integrated luminosity. One can observe that the expected limits improve with increasing integrated luminosity only marginally.

For further improvements, the event selection is required to be re-optimized. In terms of tuning the selection criteria improvements in the lepton isolation are mandatory. The fraction of isolated muons and electrons is expected to become small for LQ masses greater than 1 TeV as the decay products are more collimated for those masses. In this regime, conventional isolation requirements reduce the signal efficiency drastically. More sophisticated isolation criteria are necessary to increase the signal efficiency and to reduce SM background processes, e.g. QCD multijet processes in which leptons are usually found to be nonisolated.

The $\ell + \tau_\text{h} +$ jets final state contains 65.9% of the total branching ratio of pair produced LQs decaying into a top quark and a tau lepton. A final state without any muons or electrons ($0\ell + \tau_\text{h} +$ jets final state) has a branching ratio of 23.4%. By including this final state, further improvements in the sensitivity to third-generation LQs can be achieved. Prospect studies [154] for a search in the $0\ell + \tau_\text{h} +$ jets final state are presented in App. B.3.

Irrespective of the exact final state, boosted techniques will become important for

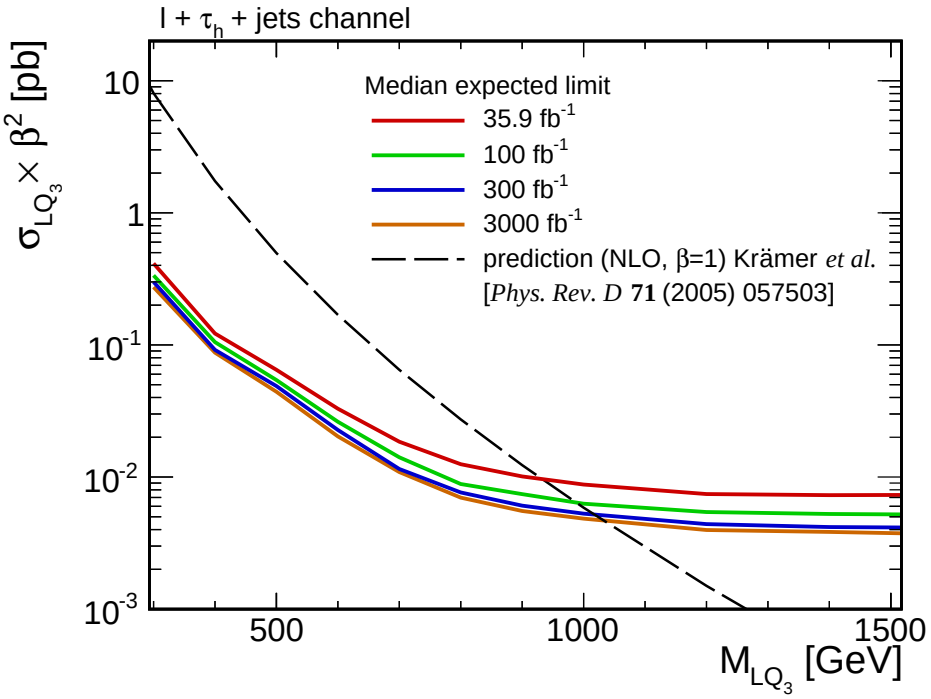

Figure 7.25: Expected exclusion limits on the cross section times (σ) branching fraction squared (BR^2) on the pair production of scalar LQs decaying to a top quark and a tau lepton at different integrated luminosities.

future analyses. Prospective analyses are expected to be sensitive to LQ masses of more than 1 TeV. The decay products of the top quark and the tau lepton become more collimated at high LQ masses. Sophisticated techniques are required for the reconstruction of those particles. Some of these techniques have been discussed in Sec. 5.9 concerning the identification of hadronically decaying top quarks. For the analysis presented in this thesis, these techniques have been found to provide worse sensitivity compared to the event selection described in Sec. 7.3. However, for future analyses tools like top tagging algorithms could play an important role to achieve sensitivities towards higher LQ masses.

Dedicated τ_h lepton reconstruction algorithms for the regime of high transverse momenta have been studied [155]. In this approach, the standard τ_h lepton reconstruction as described in Sec. 5.5 is applied to the subjets of large radius jets. This technique might also become important in a future analysis.

Third-generation scalar LQs are of theoretical interest as already discussed in Sec. 2.4. They could explain anomalies in B meson decay rates recently observed

by the BABAR [54, 55], BELLE [56–58] and LHCb [59] collaborations. In theories aiming to explain the observed anomalies LQ couplings to third-generation quarks and leptons are required typically. However, in a specific compositeness model pair production of LQs decaying into different generations of quarks and leptons is considered [46]. For instance, LQs decaying into a top quark and a muon would be allowed in this model. Depending on the parameter space, the branching ratio in this decay channel could be as large as the branching ratio for LQs decaying into a top quark and a tau lepton. Thus, decay channels as e.g. $\mathrm{LQLQ} \rightarrow \mathrm{t}\mu\mathrm{t}\mu$ will become of interest to search for at the LHC in the future [156].

8 Conclusions

Albeit the Standard Model of particle physics is a very successful theory that describes many of the phenomena observed in the universe, a couple of unanswered questions are still remaining. To resolve the open questions, many extensions to the Standard Model have been proposed and leptoquarks are among the hypothetical particles, which are predicted by many of the possible extensions. Leptoquarks decay into a lepton and a quark, and jets appear in any possible decay mode. A deep understanding of jets is therefore essential for a potential discovery of leptoquarks.

Jet energy scale measurements in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ collected in 2015 at the CMS experiment have been presented with the emphasis given to dijet event topologies. The CMS collaboration uses a factorized jet energy calibration approach to correct reconstructed jets up to particle level. The calibrations are extracted from data and simulation events and employ dijet events to determine residual $|\eta|$-dependent data-to-simulation correction factors. The two complementary dijet balance and MPF methods have been exploited and the results of both methods have been found in a good agreement. The correction factors differ from unity by less than 3% in the detector region up to $|\eta| < 2.5$ and by up to 17% in the endcap and hadron forward region of the detector. The obtained data-to-simulation correction factors are applied to most CMS analyses selecting jet objects performed with data recorded in 2015.

In the second part of this thesis, a search for pair produced third-generation scalar leptoquarks decaying into a tau lepton and a top quark has been conducted. The full proton-proton collision data set recorded in 2016 at a center-of-mass energy of 13 TeV at the CMS experiment, corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$ has been analyzed. The search was extracted in several search regions with an electron or muon, one or two hadronically decaying tau leptons, and additional jets in the final state to reach maximal sensitivity. In each search region, dedicated data control regions have been used to derive the main backgrounds that contain

misidentified hadronically decaying tau leptons. The measured transverse momentum distributions of the reconstructed hadronically decaying top quark candidate is evaluated in the search regions that contain exactly one hadronically decaying tau lepton. In the search region with two hadronically decaying tau leptons, the measured number of data events is compared to the background prediction. A good agreement between data and background prediction is found in each search region and upper cross section limits have been set. Assuming a unity branching ratio of leptoquarks decaying into a top quark a tau lepton pair, production cross sections of 0.6 pb for leptoquark masses of 300 GeV and 0.01 pb for leptoquark masses of 1.5 TeV are excluded at the 95% confidence level. These limits were compared with next-to-leading order cross section, so that third-generation scalar leptoquarks have been excluded up to masses of 900 GeV. This result represents the most stringent limits for pair produced scalar leptoquarks in the top quark and tau lepton channel to date.

A Residual Jet Energy Corrections

A.1 Control Plots - kFSR Extrapolations

The k_{FSR} extrapolations in each bin of $|\eta|$ are shown in Fig. A.1 for the MPF method and in Fig. A.2 for the dijet balance method.

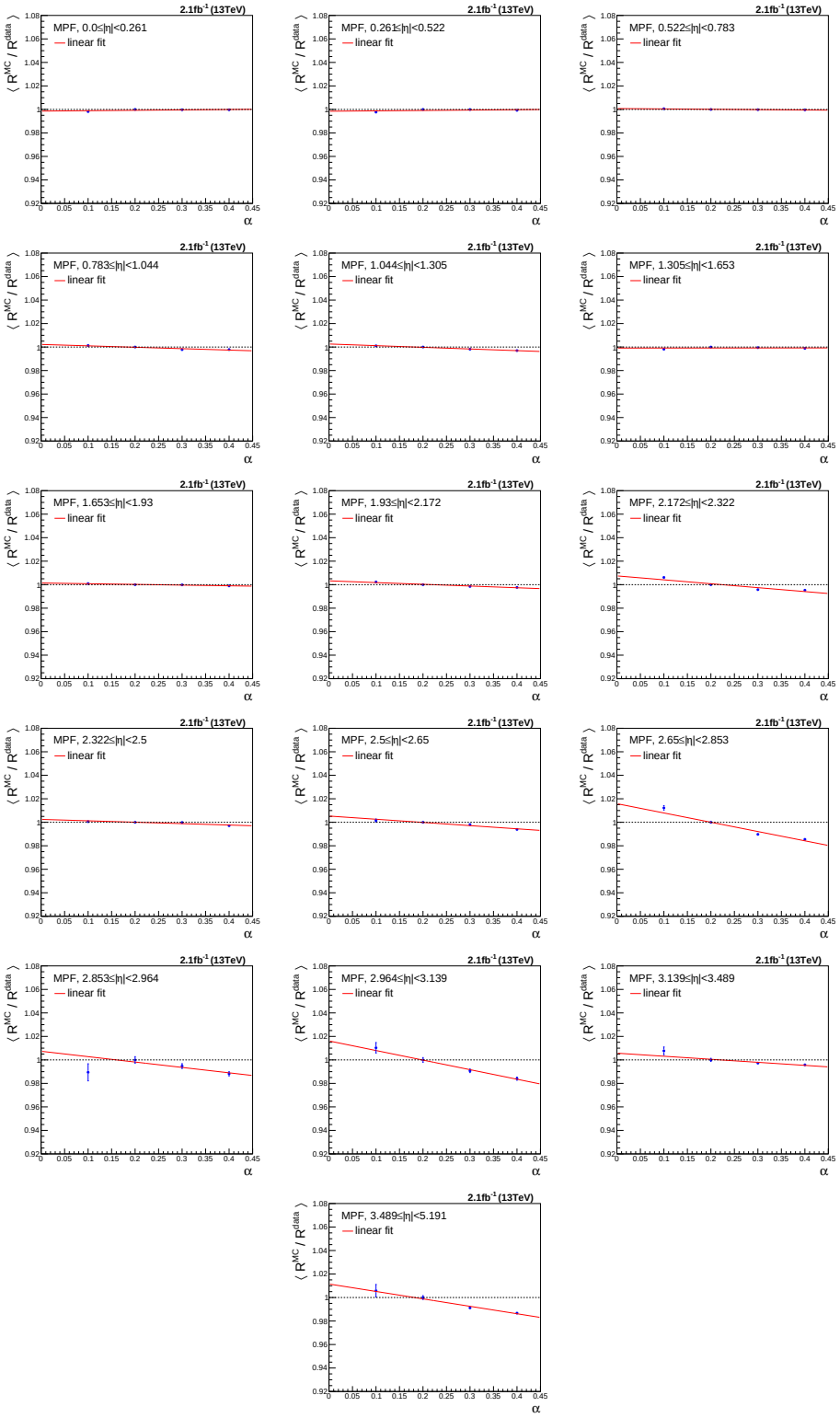

Figure A.1: Linear k_{FSR}-extrapolations in bins of $|\eta|$ as a function of the third jet fraction α with the MPF method. A linear fit is performed.

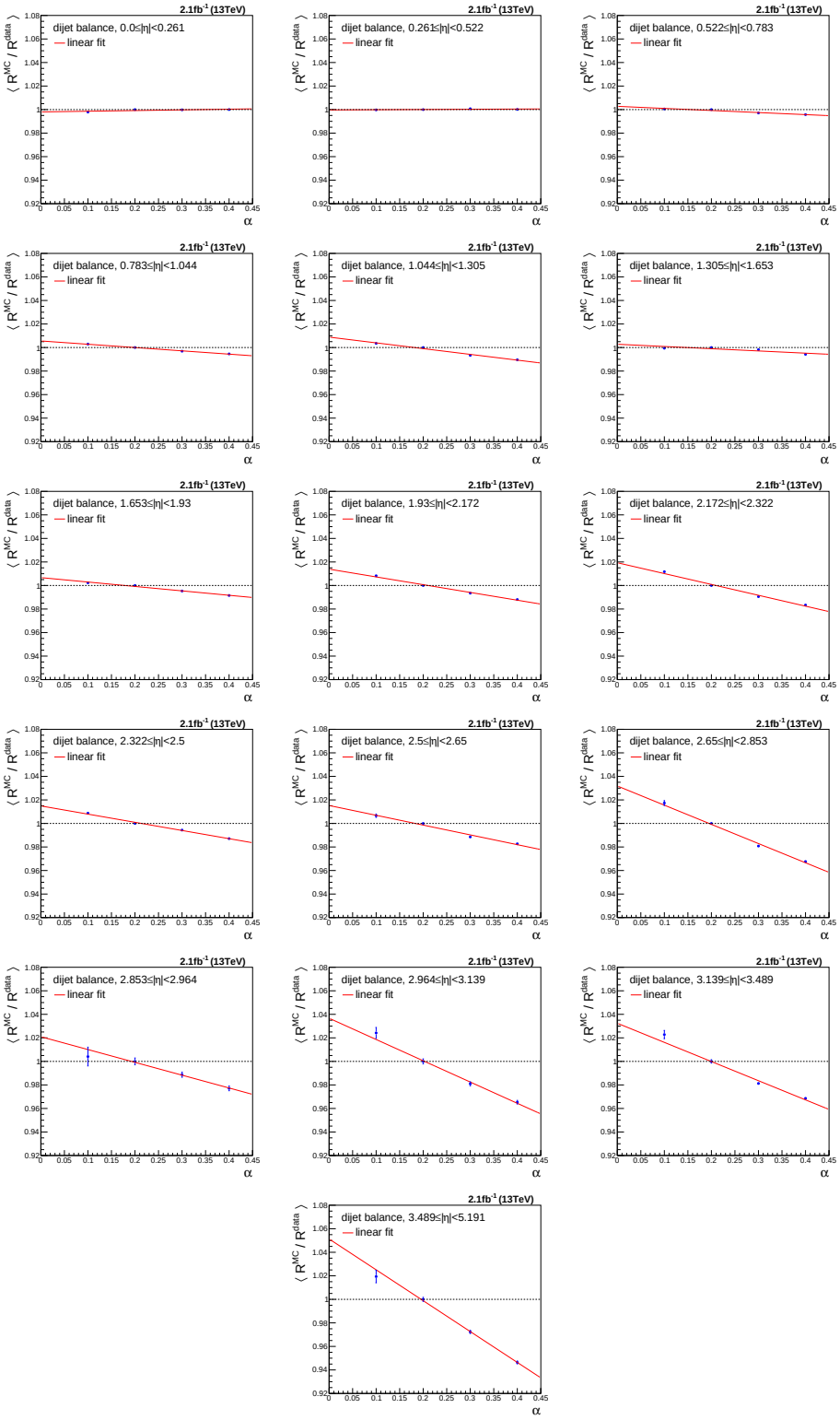

Figure A.2: Linear k_{FSR}-extrapolations in bins of $|\eta|$ as a function of the third jet fraction α with the di-jet-balance method. A linear fit is performed.

A.2 Control Plots - $\bar{p}_{\mathrm{T}}$ Extrapolations

The $\bar{p}_{\mathrm{T}}$ extrapolations in each bin of $|\eta|$ are shown in Fig. A.1 for the MPF method and in Fig. A.2 for the dijet balance method.

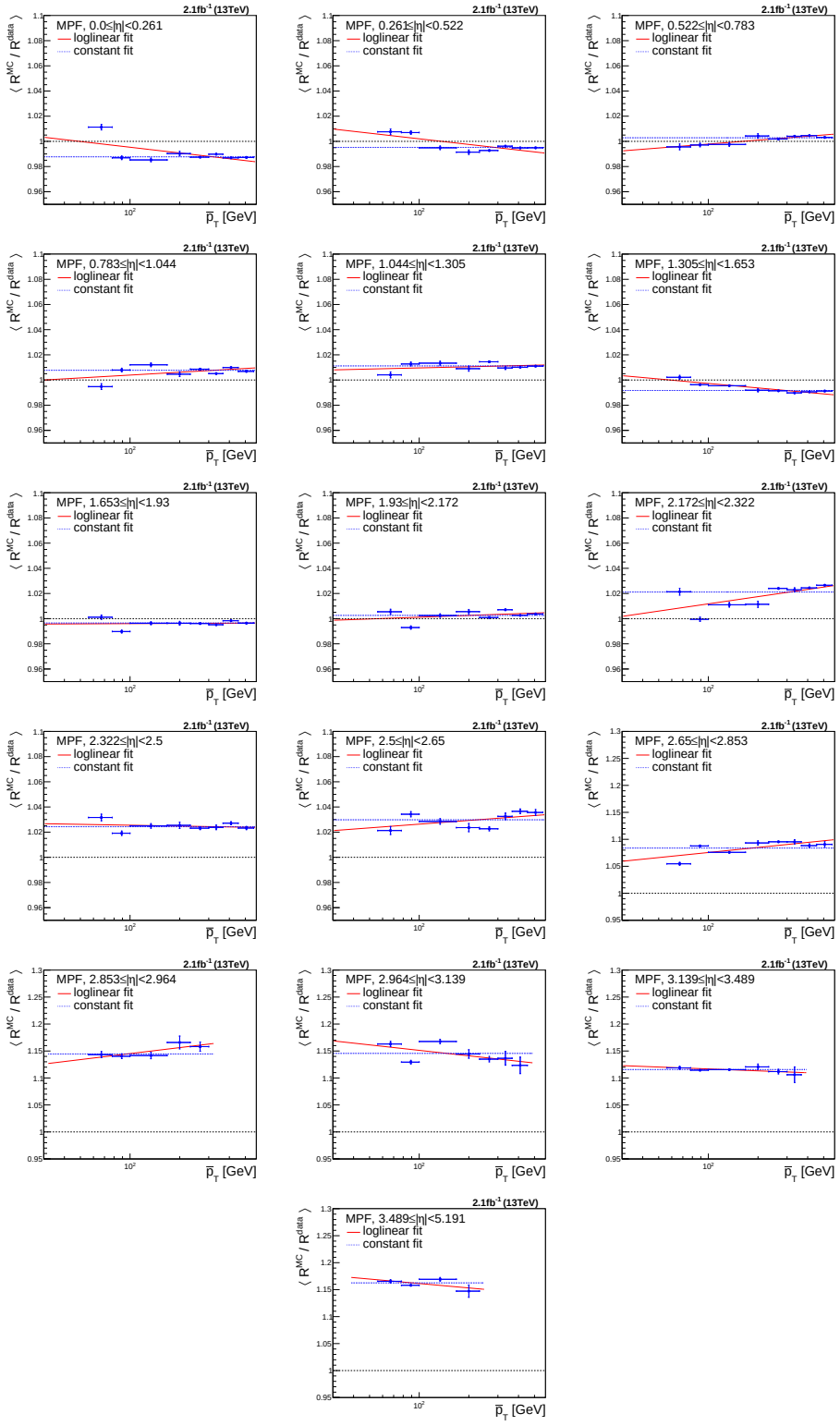

Figure A.3: Linear and log-linear extrapolations in bins of $|\eta|$ as a function of the average transverse momentum $\bar{p}_T$ with the MPF method. A log-linear (red) and a constant (blue) fit is performed.

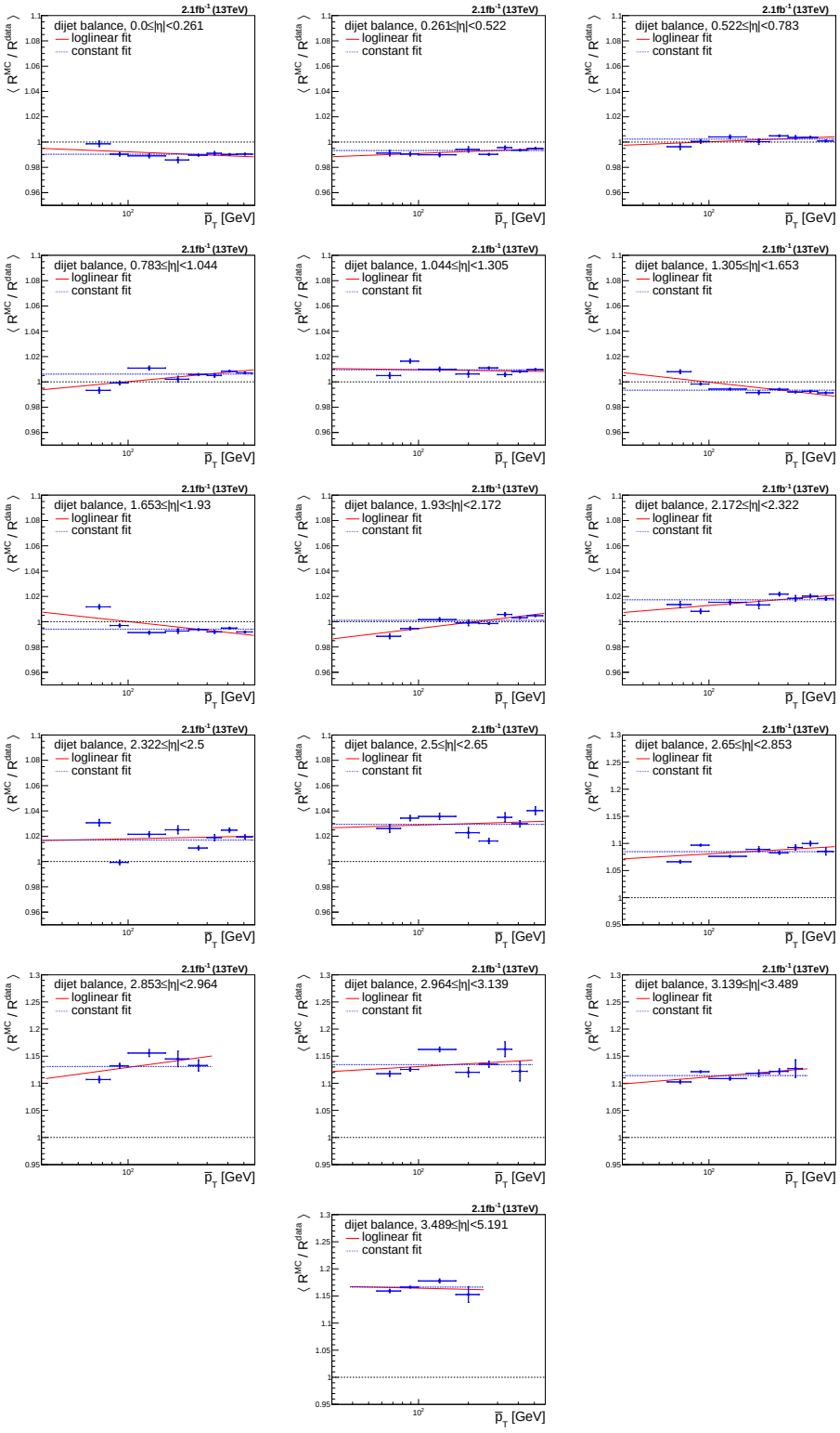

Figure A.4: Linear and loglinear extrapolations in bins of $|\eta|$ as a function of the average transverse momentum $\bar{p}_T$ with the di-jet-balance method. A loglinear (red) and a constant (blue) fit is performed.

A.3 Control Plots - Relative Response vs. Number of Primary Vertices

In Fig. A.5, the MPF response in data and simulation as a function of the number of primary vertices is shown for each bin of $|\eta|$ that has been analyzed. For each bin of $|\eta|$ the response is flat as a function of the number of primary vertices. Moreover, the response has been found to be close to one for the central detector region and in the order of up to 20% in the hadron forward region.

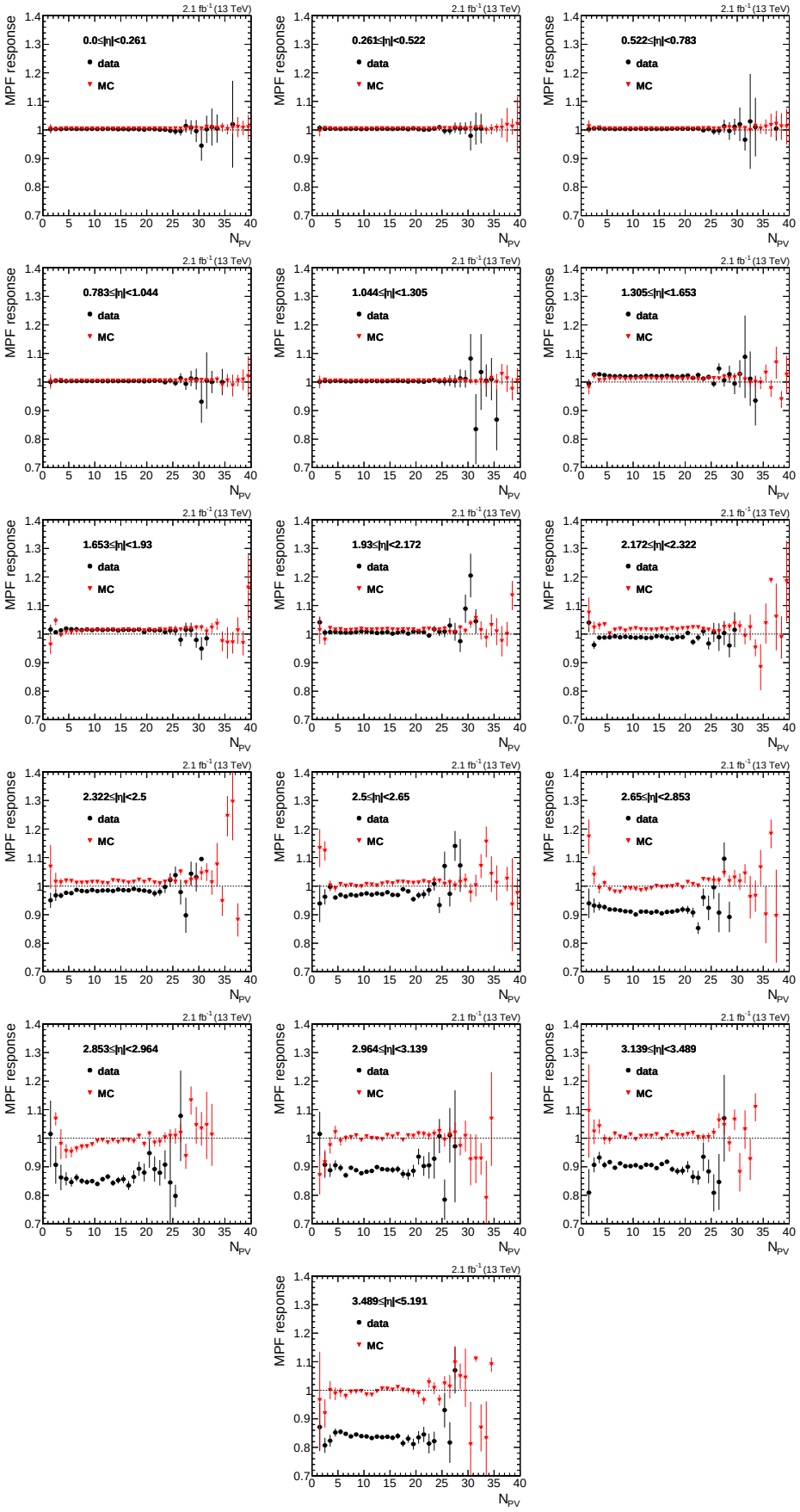

Figure A.5: Relative response as a function of the number of primary vertices for each bin of $|\eta|$.

A.4 Run Numbers of Time Dependence Studies

All run numbers used to derive the L2Res relative correction factors for the data taking period of 2015 are summarized in Tab. A.1. The data was divided into nine self-contained sets in order to verify a possible time dependence in the jet response.

	RunNr	L (1/pb)
set 1	254231-257613	210.282
set 2	257614-258136	232.832
set 3	258157-258177	237.130
set 4	258211-258448	229.872
set 5	258655-258741	214.421
set 6	258742-259685	225.970
set 7	259686-260373	219.400
set 8	260424-260534	266.200
set 9	260536-260627	258.137

Table A.1: Run regions and corresponding luminosity used for the studies of the time dependence.

B Leptoquark Analysis

B.1 Kinematic Differences between Signal Region and Control Region

It has been discussed in Sec. 7.4 that the shapes of the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions between signal region $\mathrm{SR_A}$ and control region $\mathrm{CR_A}$ are slightly different. These kinematic differences only affect category A as in category B a counting experiment is performed. The $p_{\mathrm{T}}^{\mathrm{t}}$ distributions in the signal region are found to be harder than in the control region. The main difference of the selection criteria between signal region and control region is the definition of the isolation of the τ_{h} lepton. In the signal region, the τ_{h} leptons are required to pass isolation criterion of $I_\tau < 1.5\,\mathrm{GeV}$, which means that the τ_{h} leptons are isolated. This requirement on the τ_{h} lepton isolation is inverted to define the control region. Moreover, exactly one τ_{h} lepton is required in the signal region, whereas at least one τ_{h} lepton in the control region is selected in the events.

Due to the different isolation criteria the event kinematics can be different in the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions between signal region and control region. Nonisolated τ_{h} lepton candidates are expected to have, on average, higher transverse momenta than isolated τ_{h} lepton candidates. The activity around the τ_{h} lepton candidates is larger if no isolation requirement is applied, which leads to higher transverse momenta of nonisolated τ_{h} lepton candidates. This means that the transverse momenta of τ_{h} leptons are expected to be greater, on average, in the control region than the transverse momenta of τ_{h} leptons in the signal region ($p_{\mathrm{T}}^{\tau}(\mathrm{CR}) > p_{\mathrm{T}}^{\tau}(\mathrm{SR})$).

Following the jet-tau-cleaning as described in Sec. 5.5, each jet is removed from the jet list if the jet overlaps with a τ_{h} lepton candidate within a cone of the radius of $\Delta R = 0.4$. Thus, the jets removed in the procedure of the jet-tau-cleaning have higher transverse momenta in the control region than in the signal region. In Fig. B.1, the p_{T} distribution of the jets that are removed in the jet-tau-cleaning

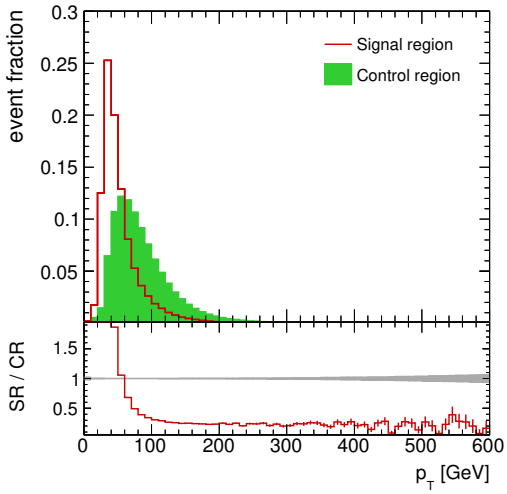

Figure B.1: Shape comparison of the p_{T} of the jet that is removed in the jet-tau-cleaning procedure for the signal region (red) and the control region (green) in the $\mu + \tau_{\mathrm{h}} +$ jets final state.

procedure are shown for signal region and control region. It can be observed that the removed jets in the signal region are rather soft compared the to jets erased in the control region. By removing jets with higher transverse momenta, the transverse momentum of the reconstructed hadronically decaying top quark is expected to be softer in the control region.

In Fig. B.2, the shape comparison is shown for jet-multiplicities of 3, 4, and at least 5. It can be seen that the shapes of the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions differ between signal region and control region depending on the jet multiplicity. In these distributions, a looser selection has been applied to increase the number of expected events and to decrease statistical uncertainties. Only the baseline selection criteria are applied in addition to the requirement of at least one OS or SS $\ell\tau_h$ pair in the event. For events that contain exactly three jets the shapes of the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions are similar between signal region and control region. With increasing jet-multiplicities, the $p_{\mathrm{T}}^{\mathrm{t}}$ distribution in the signal region gets harder with respect to the $p_{\mathrm{T}}^{\mathrm{t}}$ distribution in the control region. For this reason, correction factors are determined to correct the $p_{\mathrm{T}}^{\mathrm{t}}$ distribution in the control region in dependence of the jet multiplicity and of $p_{\mathrm{T}}^{\mathrm{t}}$. This is performed with events that passed the baseline selection criteria for OS and SS events inclusively.

The correction factors are applied to all events in the control region $\mathrm{CR_A}$. In the

$\mu + \tau_{\mathrm{h}} +$ jets final state, the correction factors are shown in Fig. 7.13. A similar procedure is applied in the $e + \tau_{\mathrm{h}} +$ jets final state as discussed in the following.

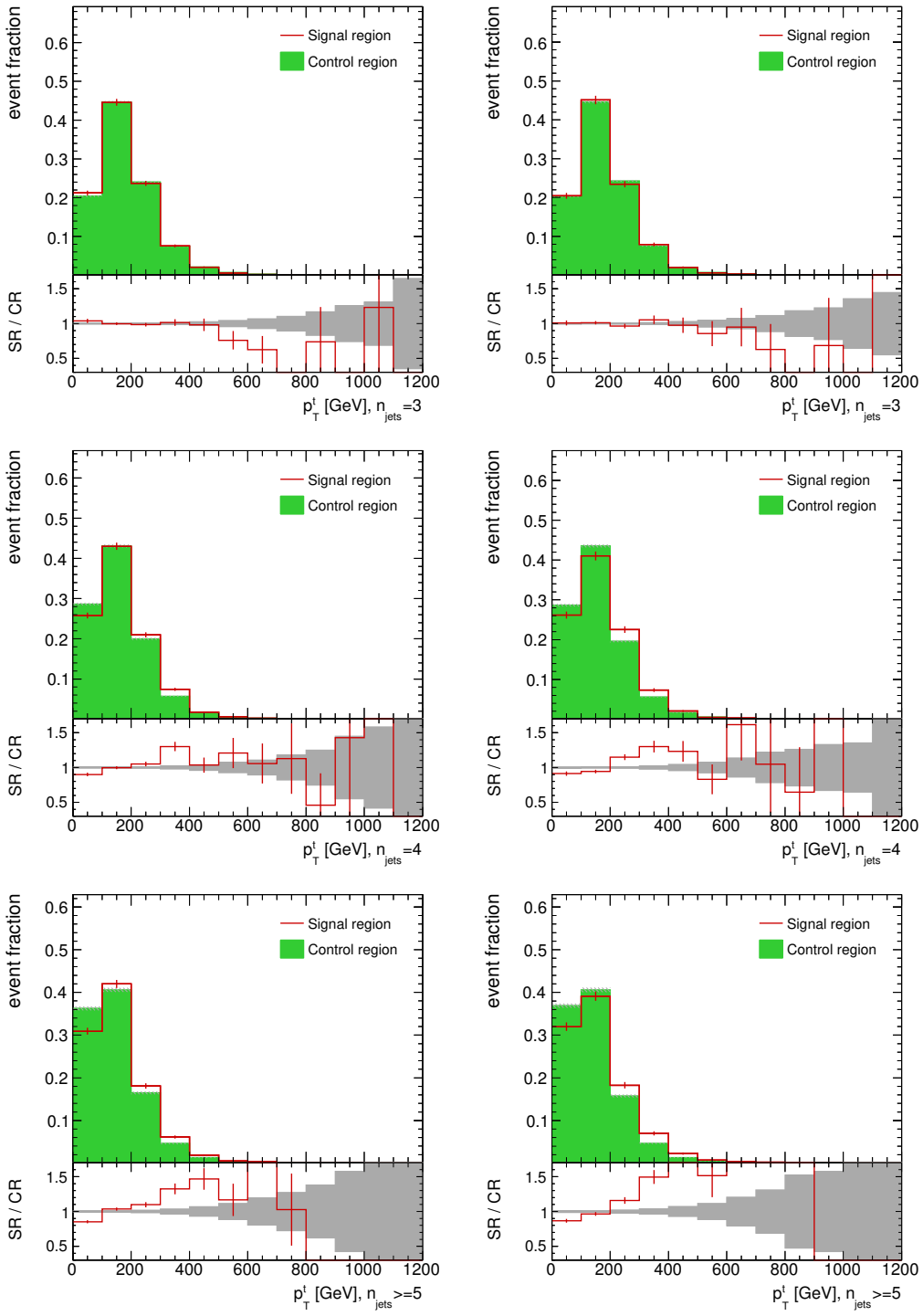

Figure B.2: Shape comparison between signal region (SR_A) and control region (CR_A) of the p_T^t distribution for jet-multiplicities of 3 (top), 4 (middle), and at least 5 (bottom) in the $\mu + \tau_h +$ jets final state. The baseline selection criteria and the OS (left) and SS (right) $\mu\tau_h$ pair requirements are applied, respectively.

Reweighting in the e+tau+jets final state

The shape comparison of the $p_{\mathrm{T}}^{\mathrm{t}}$ distribution between signal region and control region for simulated $t\bar{t}_{\mathrm{f}}$ and W+jets events in the $e + \tau_{\mathrm{h}} +$ jets final state is shown in Fig. B.3. A similar trend as shown in Sec. 7.4 in the $\mu + \tau_{\mathrm{h}} +$ jets final state can be observed. Individual correction factors have been determined in the $e + \tau_{\mathrm{h}} +$ jets final state to correct for kinematic differences between signal and control region. The correction factors in dependence of the jet multiplicity and $p_{\mathrm{T}}^{\mathrm{t}}$ are shown in Fig. B.4. One can see that the correction factors derived in the $e + \tau_{\mathrm{h}} +$ jets final state are compatible with the ones obtained in the $\mu + \tau_{\mathrm{h}} +$ jets final state.

The $p_{\mathrm{T}}^{\mathrm{t}}$ shape comparison between signal and control region after applying the correction factors is shown in Fig. B.5. The correction factors have been applied to all events in the control region $\mathrm{CR_A}$ in the $e + \tau_{\mathrm{h}} +$ jets final state. It can be observed that the shapes of the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions between signal and control region are in a good agreement.

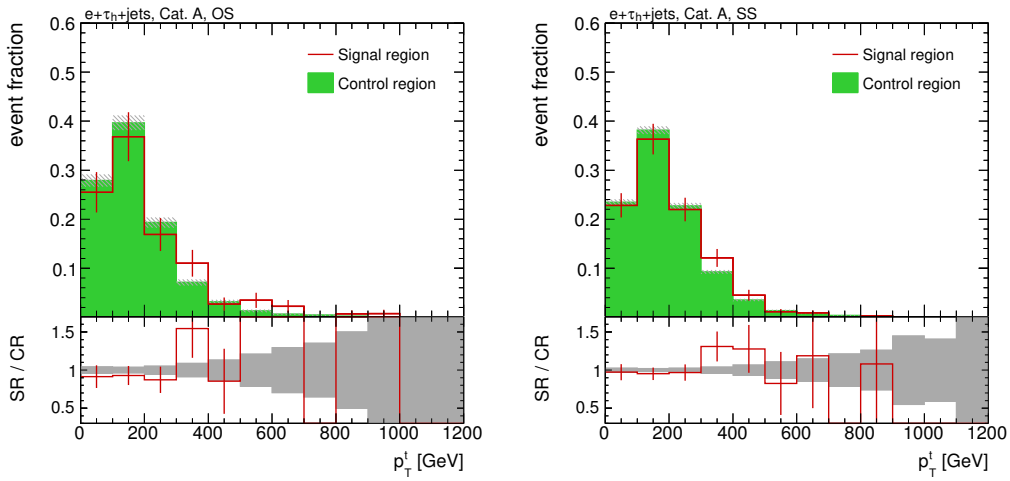

Figure B.3: Shape comparison between signal and control region ($\mathrm{CR_A}$) of the $p_{\mathrm{T}}^{\mathrm{t}}$ distribution. The comparison after applying the full selection in the OS (left) and the SS (right) category in the $e + \tau_{\mathrm{h}} +$ jets channel is shown.

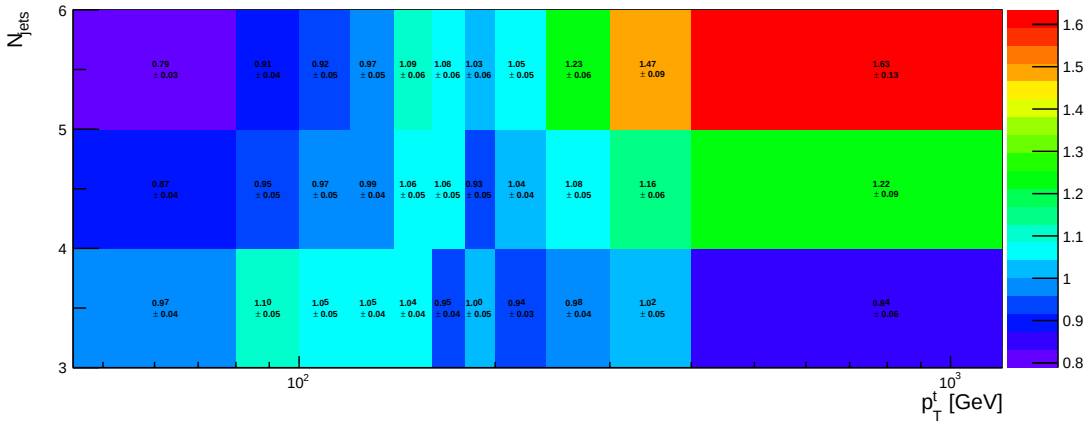

Figure B.4: Correction factors in the $e + \tau_\mathrm{h} +$ jets channel. In dependence from p_T^t and the jet-multiplicity and are applied in the control region ($\mathrm{CR_A}$).

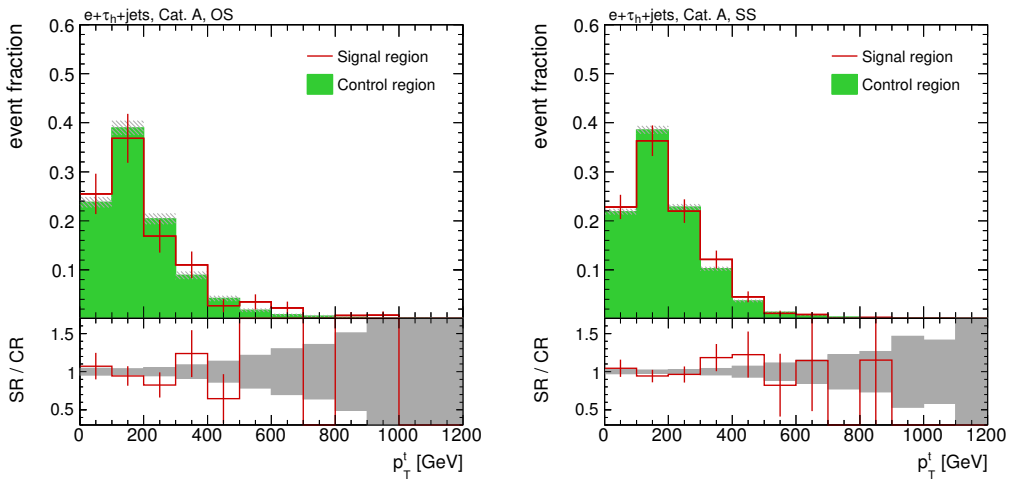

Figure B.5: Shape comparison between signal and control region ($\mathrm{CR_A}$) of the p_T^t distribution after applying the correction factors. The comparison after applying the full selection in the $e + \tau_\mathrm{h} +$ jets final state and in the OS (left) and the SS (right) category is shown.

B.2 Validation of the Background Estimation Method

An additional test to validate the background estimation method is presented in this section. The background estimation relies on simulation to normalize the yields in the control region. In case of a positive discovery, a further validation with data helps to make the background estimate more reliable.

A new control region (in the following referred to as $\mathrm{CR}_{\mathrm{test}}$) is defined with selection criteria close to the ones applied in the signal region SR_{A}. This means that all selected events contain one isolated τ_{h} lepton. In this test only the $\mu + \tau_{\mathrm{h}} +$ jets final state is used. To reduce the signal contamination and to avoid overlap with the signal region the $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ requirement is inverted and the events are required to pass $p_{\mathrm{T}}^{\tau_{\mathrm{h}}} < 100\,\mathrm{GeV}$. Moreover, only the low S_{T} regions of category A are considered as the number of expected background events in the high S_{T} regions become low due to the criterion applied to $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$.

The full analysis flow of category A is performed and the control region $\mathrm{CR}_{\mathrm{test}}$ is treated as the signal region. Thus, the $t\bar{t}_{\mathrm{f}}$ and W+jets backgrounds are derived from the control region CR_{A} and extrapolated to the control region $\mathrm{CR}_{\mathrm{test}}$. New extrapolation factors α are determined for the OS and SS regions. Correction factors to eliminate kinematic difference in the $p_{\mathrm{T}}^{\mathrm{t}}$ distributions between the control regions $\mathrm{CR}_{\mathrm{test}}$ and CR_{A} are calculated. Both, the extrapolation factors and the correction factors are different to the ones obtained in the main analysis as the phase space is different between signal region SR_{A} and control region $\mathrm{CR}_{\mathrm{test}}$.

All systematic uncertainties as described in Sec. 7.5 are taken into account in this test. The final $p_{\mathrm{T}}^{\mathrm{t}}$ distributions in the control region $\mathrm{CR}_{\mathrm{test}}$ are shown in Fig. B.6. In these distributions, the $t\bar{t}_{\mathrm{f}}$ and W+jets backgrounds are derived from data events in control region CR_{A}. It can be observed that the data to (mainly)data-driven background ratio is in a good agreement in both the OS and SS category. Thus, the background estimation method is validated successively.

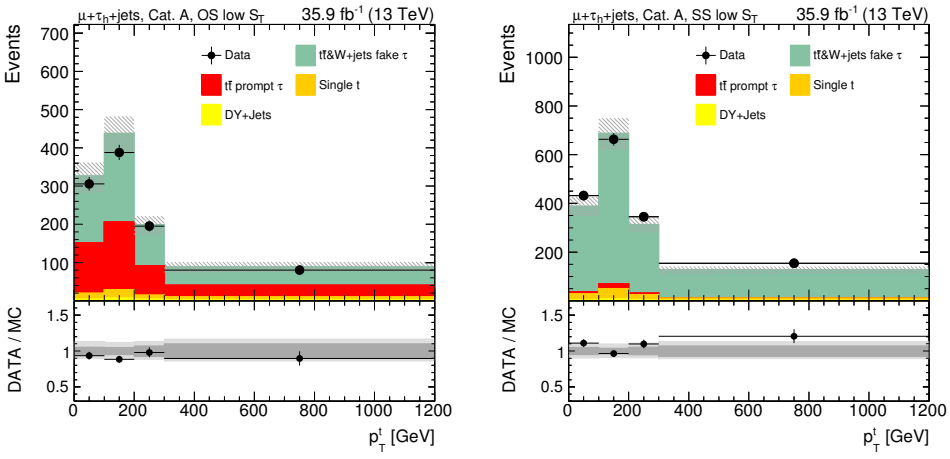

Figure B.6: Distributions of $p_{\mathrm{T}}^{\mathrm{t}}$ in the control region $\mathrm{CR}_{\mathrm{test}}$. The events are separated into OS (left) and SS (right) regions. In the bottom panel, the dark grey shaded area represents the statistical uncertainties and the light grey shaded area corresponds to the total uncertainties.

B.3 Prospect Studies in a Hadronic Final State

In this section, prospective studies for a search for pair produced LQs decaying into a top quark and a tau lepton in an all-hadronic final state are presented. This final state has a branching ratio of 23.9%. Thus, a combination with the $\ell + \tau_h$ + jets final state would improve the total sensitivity. The following studies are based on Ref. [154] and only studies based on simulation will be presented. The simulated processes are normalized to an integrated luminosity of $35.9\,\text{fb}^{-1}$, corresponding to the dataset recorded in 2016.

Baseline Selection

A baseline selection is applied at first. A veto against electrons and muons fulfilling the reconstruction criteria described in chapter 5 is applied to avoid overlap with the $\ell + \tau_h$ + jets final state. The dominant background in all-hadronic final states typically arises from QCD multijet processes. At least two τ_h leptons are required to reduce the number of QCD multijet events. For this purpose, a di-τ_h lepton trigger is used, which requires at least two τ_h lepton candidates with medium isolation criterion, $p_T > 35\,\text{GeV}$, and $|\eta| < 2.1$.

The baseline selection requirements are:

- at least two τ_h lepton candidates with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.1$,
- a veto against electrons and muons,
- at least two AK4 jets with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.4$,
- $\not{E}_T > 50\,\text{GeV}$, and
- $S_T > 350\,\text{GeV}$.

Distributions after applying the baseline selection are shown in Fig. B.7. Although two τ_h lepton candidates are selected, the QCD multijet background is dominant. However, this background is reduced further in the following, particularly by the requirement of at least one b tag. The final selection criteria are described later.

Discrimating Variable

The p_T^t distribution is used as the discriminating variable in the analysis of the $\ell + \tau_h$ + jets final state as the reconstruction of the LQ mass is not possible due to

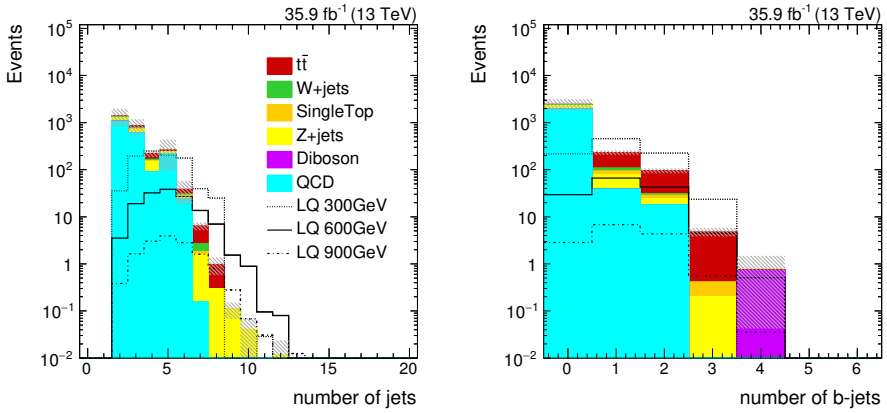

Figure B.7: Control distributions in the all-hadronic final state after applying the baseline selection. The distributions for the number of jets (left) and number of b-jets (right) are shown.

a potentially large number of neutrinos. In an all-hadronic final state, however, the number of occurring neutrinos is reduced. For the reconstruction of the LQ mass the reconstruction of the top quark and τ_h lepton masses is required. Hadronically decaying top quarks are reconstructed by using the same approach described in Sec. 7.3.2. In the reconstruction of the mass of τ_h leptons only the visible fraction of the decay is used by the HPS algorithm. In general, visible reconstructed tau lepton masses are smaller than the true mass of the tau lepton.

The reconstructed mass of the τ_h lepton is corrected under the assumption that the τ_h leptons have a large lorentz-boost. In that case, a collinear approximation method can be utilized [157]. Under the assumption that the event contains exactly two τ_h leptons, it is expected that the neutrinos are collinear to each τ_h lepton, respectively. This leads to the following equations

$$\nu_1 = x_1 \tau_{h_1}, \;\; \nu_2 = x_2 \tau_{h_2}, \tag{B.1}$$

where ν_1, ν_2, τ_{h_1} and τ_{h_2} are the four momenta of the respective neutrinos and τ_h leptons. The values of the factors x_1 and x_2 are between 0 and 1. These factors are determined by the following linear system of equations:

$$\begin{pmatrix} \not{E}_{T,x} \\ \not{E}_{T,y} \end{pmatrix} = \begin{pmatrix} p_{T,\nu_{1,x}} \\ p_{T,\nu_{1,y}} \end{pmatrix} + \begin{pmatrix} p_{T,\nu_{2,x}} \\ p_{T,\nu_{2,y}} \end{pmatrix}. \tag{B.2}$$

In this equation, the missing transverse energy $\not{E}_T$ and the transverse momenta of

the neutrinos are split into their x and y components, respectively. The resulting masses of the τ_h leptons are obtained by

$$\begin{aligned} m_{\tau_{\mathrm{h},1}} &= (1 + x_1) m_{\tau_{\mathrm{h},1}^{\mathrm{rec}}} \\ m_{\tau_{\mathrm{h},2}} &= (1 + x_2) m_{\tau_{\mathrm{h},2}^{\mathrm{rec}}}, \end{aligned} \tag{B.3}$$

where $m_{\tau_{\mathrm{h},1}^{\mathrm{rec}}}$ and $m_{\tau_{\mathrm{h},2}^{\mathrm{rec}}}$ are the visible masses of τ_h leptons reconstructed by the HPS algorithm.

To reconstruct the four momenta of both LQs an assignment of the top quark hypothesis to the τ_h leptons is required. This leads to two sets of LQ hypotheses for each given pair of top quark hypotheses.

Additional hypotheses are included by the reconstruction of the hadronically decaying top quarks. Similar to the reconstruction described in Sec. 7.3.2, a list of all possible permutations of AK4 jets in the event is constructed to reconstruct both hadronic top quarks. The top quark four momenta are given by the sum of the four momenta of all reconstructed jets assigned to the top quarks in each hypothesis. A discriminator is introduced by

$$\chi^2_\mathrm{t} = \left(\frac{m_{\mathrm{t1,hyp}} - m_\mathrm{t}}{\sigma_\mathrm{t}} \right)^2 + \left(\frac{m_{\mathrm{t2,hyp}} - m_\mathrm{t}}{\sigma_\mathrm{t}} \right)^2 \tag{B.4}$$

and the selected hypothesis in the one with the lowest value of χ^2_t. The values m_t and σ_t correspond to top quark mass and width determined in simulation with correct hypotheses. These values are taken from Ref. [158].

For the reconstruction of the τ_h leptons the missing transverse energy $\not{E}_\mathrm{T}$ is distributed to two components. The corresponding discriminator is defined as

$$\chi^2_{\tau_\mathrm{h}} = \left(\frac{(p_{\mathrm{T},\nu_1} + p_{\mathrm{T},\nu_2})_x - \not{E}_{\mathrm{T}x}}{\sigma_{\not{E}_\mathrm{T},x}} \right)^2 + \left(\frac{(p_{\mathrm{T},\nu_1} + p_{\mathrm{T},\nu_2})_y - \not{E}_{\mathrm{T}y}}{\sigma_{\not{E}_\mathrm{T},y}} \right)^2 . \tag{B.5}$$

The variables $\sigma_{\not{E}_\mathrm{T},x}$ and $\sigma_{\not{E}_\mathrm{T},y}$ correspond to the resolution of the missing transverse energy and are taken from Ref. [125].

In addition, LQ hypotheses are favored if their reconstructed masses are close to

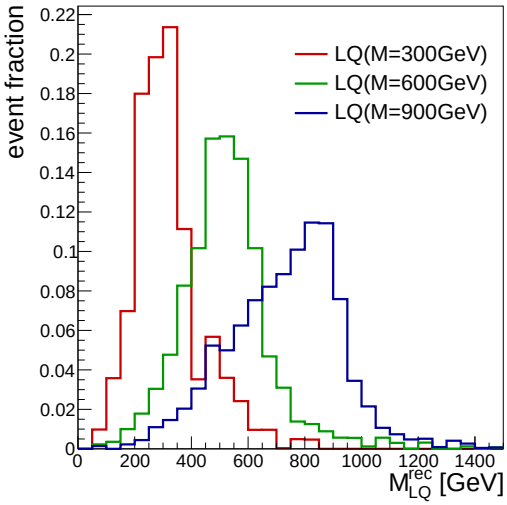

Figure B.8: Reconstructed LQ mass after applying the baseline selection. Each distribution is normalized to its integral.

each other. This leads to a another discriminator:

$$\chi^2_{\mathrm{LQ,diff}} = \left(\frac{\Delta M^{\mathrm{rel}}_{\mathrm{LQ}} - \overline{\Delta M}^{\mathrm{rel}}_{\mathrm{LQ}}}{\sigma_{\Delta M}} \right), \tag{B.6}$$

where $\Delta M^{\mathrm{rel}}_{\mathrm{LQ}}$ is the relative difference between the masses of the two LQ hypotheses. The values $\overline{\Delta M}^{\mathrm{rel}}_{\mathrm{LQ}}$ and $\sigma_{\Delta M}$ are determined from simulation. The values are taken from hypotheses containing the reconstructed objects that truly originate from the decay of the LQ pair.

In total, a final discriminator to build the LQ hypotheses is defined by

$$\chi^2_{\mathrm{tot}} = \chi^2_{\mathrm{t}} + \chi^2_{\tau_{\mathrm{h}}} + \chi^2_{\mathrm{LQ,diff}}. \tag{B.7}$$

For each event the set of LQ hypotheses with the smallest χ^2_{tot} is chosen. Since each event then contains two LQ candidates, the average mass of both LQs is used. In Fig. B.8 the distributions of the reconstructed average LQ mass is shown for different signal samples. Each distribution is normalized to its integral. One can see that the distributions peak at the value according to the mass of the generated LQ. The width of the peak increases for higher LQ masses, but the relative width is compatible for each mass point.

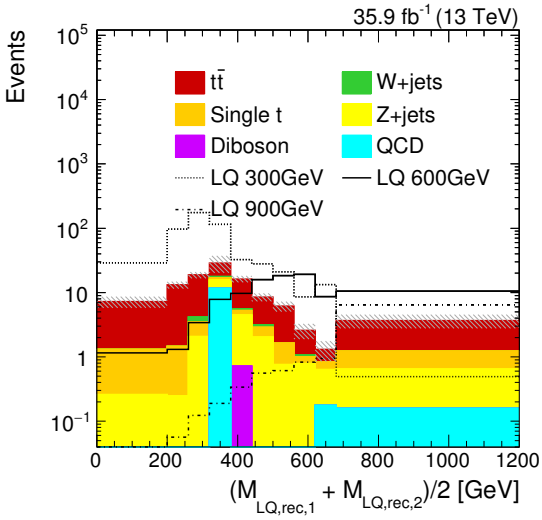

Figure B.9: Reconstructed LQ mass after applying the full selection.

Final Selection and Expected Limits

The signal to background ratio is small after applying the baseline selection. In particular, events from QCD multijet processes are still dominating and additional selection criteria are required to get rid of the SM background events. Many of the QCD events do not contain a b tagged jet. Thus, b tagging may become a very important tool in this analysis. In addition, events from signal processes consist, on average, of more jets than events from SM BG processes.

The final selection criteria are

- at least four AK4 jets with $p_{\mathrm{T}} > (50, 50, 30, 30)\,\mathrm{GeV}$ and $|\eta| < 2.4$,
- $S_{\mathrm{T}} > 400\,\mathrm{GeV}$,
- at least one b tagged AK4 jet with the medium working point.

The reconstructed LQ mass is shown in Fig. B.9 after applying these requirements. One can observe that the signal over background ratio has improved significantly. The background fraction of QCD multijet processes is reduced. The reconstructed LQ mass peaks at low masses for most of the background events. Thus, this variable separates well between background events and events from signal processes with higher masses.

Expected limits are calculated after the full selection criteria have been applied. The expected limits are shown in Fig. B.10. Here, only statistical uncertainties and

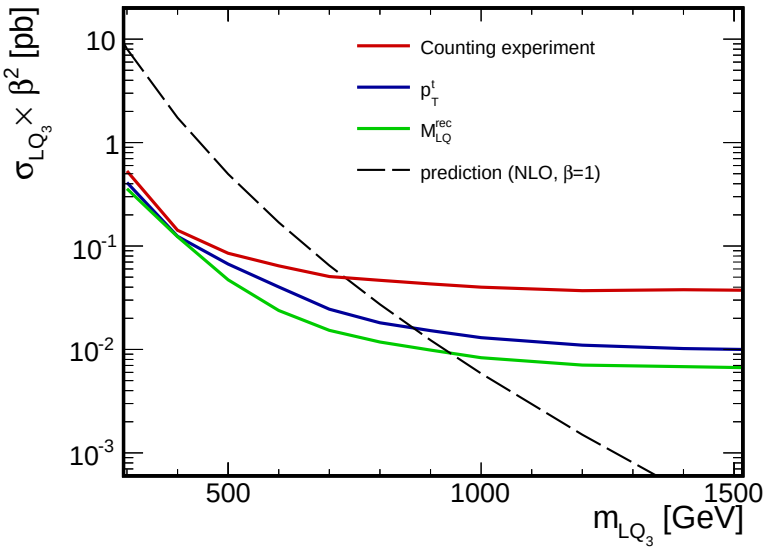

Figure B.10: Expected exclusion limits in the hadronic decay channel after applying the full selection. A counting experiment has been performed (red line) and the transverse momentum of the top quark (blue line) as well as the reconstructed LQ mass (green line) have been used as the input to the limit calculation.

the uncertainties in the normalization of the background processes as described in Sec. 7.18 are taken into account. In addition, a rate uncertainty of 100% has been applied to the QCD background events. The expected limits have been determined using the reconstructed LQ mass and the transverse momentum of the reconstructed top quark $p_{\mathrm{T}}^{\mathrm{t}}$ with the statistical method described in Sec. 7.3.2. Additionally, a counting experiment has been performed. The counting experiment gives the worst results as no shape separation between background processes and signal processes is taken into account. However, for low LQ masses the counting experiment performs almost as good as the other variables but the mass exclusion is rather low. The $p_{\mathrm{T}}^{\mathrm{t}}$ distribution performs better than the counting experiment over the full mass range and the excluded mass is at almost 900 GeV. The reconstructed LQ mass provides the best exclusion limits and can exclude LQ masses of more than 900 GeV.

Outlook

The results of a search for pair produced LQs decaying into a top quark and a tau lepton in the hadronic final state are promising. The expected exclusion limits are comparable to the results in the $\ell + \tau_{\mathrm{h}} +$ jets final state, but some improvements are

required to be added to the main analysis.

Further optimization studies of the event selection can improve the sensitivity. As the studies shown in this section are performed inclusively, one could include different categories similar to the categorization procedure in the $\ell + \tau_{\mathrm{h}} +$ jets final state.

More importantly, a sophisticated background estimation is required. The main background comes from $t\bar{t}$ processes. Thus, $t\bar{t}_{\mathrm{f}}$, $t\bar{t}_{\mathrm{p+f}}$ and $t\bar{t}_{\mathrm{p}}$ processes are possible. For the background events that consist of at least one misidentified τ_{h} lepton a data driven estimation is required. This concerns particularly background events from QCD multijet processes, which consist of misidentified τ_{h} leptons only. A data driven QCD background estimation is also necessary as the number of simulated events is not sufficient after applying the full selection.

Furthermore, systematic uncertainties have not been taken into account in these studies.

In conclusion, a combination of an all-hadronic final state with $\ell + \tau_{\mathrm{h}} +$ jets final state would lead to a significant improvement to the analysis. For future searches of third-generation pair-produced LQs decaying into a top quark and a tau lepton this should be taken into account.

Bibliography

[1] **CMS** Collaboration, "Search for third generation scalar leptoquarks decaying to a top quark and a tau lepton at sqrt(s) = 13 TeV," *CMS Physics Analysis Summary* **CMS-PAS-B2G-16-028** (2017) . https://cds.cern.ch/record/2297047.

[2] **CMS** Collaboration, V. Khachatryan *et al.*, "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 [hep-ex].

[3] C. Berger, *Elementarteilchenphysik*. SpringerLink : Bücher. Springer Spektrum, Berlin, Heidelberg, 3. aufl. 2014 ed., 2014. http://dx.doi.org/10.1007/978-3-642-41753-5.

[4] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008. https://cds.cern.ch/record/111880.

[5] Cush, "Standard Model of Elementary Particles." https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. [accessed October 18, 2017].

[6] **Particle Data Group** Collaboration, K. A. Olive *et al.*, "Review of Particle Physics," *Chin. Phys.* **C38** (2014) 090001.

[7] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.* **13** (1964) 508–509.

[8] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.* **13** (1964) 321–323.

[9] **ATLAS** Collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett.* **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[10] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett.* **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[11] **ATLAS** Collaboration, G. Aad *et al.*, "Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC," *Phys. Lett.* **B726** (2013) 88–119, arXiv:1307.1427 [hep-ex]. [Erratum: Phys. Lett.B734,406(2014)].

[12] **ATLAS** Collaboration, G. Aad *et al.*, "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector," *Phys. Rev.* **D91** no. 1, (2015) 012006, arXiv:1408.5191 [hep-ex].

[13] **ATLAS** Collaboration, G. Aad *et al.*, "Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector," *Phys. Rev.* **D90** no. 11, (2014) 112015, arXiv:1408.7084 [hep-ex].

[14] **CMS** Collaboration, V. Khachatryan *et al.*, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV," *Eur. Phys. J.* **C75** no. 5, (2015) 212, arXiv:1412.8662 [hep-ex].

[15] A. G. Cohen, D. B. Kaplan, and A. E. Nelson, "Progress in electroweak baryogenesis," *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 27–70, arXiv:hep-ph/9302210 [hep-ph].

[16] A. Refregier, "Weak gravitational lensing by large scale structure," *Ann. Rev. Astron. Astrophys.* **41** (2003) 645–668, arXiv:astro-ph/0307212 [astro-ph].

[17] A. Vikhlinin, A. Kravtsov, W. Forman, C. Jones, M. Markevitch, S. S. Murray, and L. Van Speybroeck, "Chandra sample of nearby relaxed galaxy clusters: Mass, gas fraction, and mass-temperature relation," *Astrophys. J.* **640** (2006) 691–709, arXiv:astro-ph/0507092 [astro-ph].

[18] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A direct empirical proof of the existence of dark

matter," *Astrophys. J.* **648** (2006) L109–L113, arXiv:astro-ph/0608407 [astro-ph].

[19] N. Jarosik *et al.*, "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results," *Astrophys. J. Suppl.* **192** (2011) 14, arXiv:1001.4744 [astro-ph.CO].

[20] **WMAP** Collaboration, E. Komatsu *et al.*, "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophys. J. Suppl.* **192** (2011) 18, arXiv:1001.4538 [astro-ph.CO].

[21] **Supernova Search Team** Collaboration, A. G. Riess *et al.*, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.* **116** (1998) 1009–1038, arXiv:astro-ph/9805201 [astro-ph].

[22] **Supernova Cosmology Project** Collaboration, S. Perlmutter *et al.*, "Measurements of Omega and Lambda from 42 high redshift supernovae," *Astrophys. J.* **517** (1999) 565–586, arXiv:astro-ph/9812133 [astro-ph].

[23] P. J. E. Peebles and B. Ratra, "The Cosmological constant and dark energy," *Rev. Mod. Phys.* **75** (2003) 559–606, arXiv:astro-ph/0207347 [astro-ph].

[24] S. P. Martin, "A Supersymmetry primer," arXiv:hep-ph/9709356 [hep-ph]. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[25] I. J. R. Aitchison, "Supersymmetry and the MSSM: An Elementary introduction," arXiv:hep-ph/0505105 [hep-ph].

[26] J. C. Pati and A. Salam, "Lepton Number as the Fourth Color," *Phys. Rev.* **D10** (1974) 275–289. [Erratum: Phys. Rev.D11,703(1975)].

[27] H. Georgi and S. L. Glashow, "Unity of All Elementary Particle Forces," *Phys. Rev. Lett.* **32** (1974) 438–441.

[28] A. Pomarol, "Beyond the Standard Model," in *Proceedings, High-energy Physics. Proceedings, 18th European School (ESHEP 2010): Raseborg, Finland, June 20 - July 3, 2010*, pp. 115–151. 2012. arXiv:1202.1391 [hep-ph]. https://inspirehep.net/record/1088246/files/arXiv:1202.1391.pdf.

[29] B. Schrempp and F. Schrempp, "LIGHT LEPTOQUARKS," *Phys. Lett.* **B153** (1985) 101–107.

[30] H. Harari, "A Schematic Model of Quarks and Leptons," *Phys. Lett.* **86B** (1979) 83–86.

[31] M. J. Dugan, H. Georgi, and D. B. Kaplan, "Anatomy of a Composite Higgs Model," *Nucl. Phys.* **B254** (1985) 299–326.

[32] S. Dimopoulos and J. Preskill, "Massless Composites With Massive Constituents," *Nucl. Phys.* **B199** (1982) 206–222.

[33] D. B. Kaplan, H. Georgi, and S. Dimopoulos, "Composite Higgs Scalars," *Phys. Lett.* **136B** (1984) 187–190.

[34] T. Banks, "CONSTRAINTS ON SU(2) x U(1) BREAKING BY VACUUM MISALIGNMENT," *Nucl. Phys.* **B243** (1984) 125–130.

[35] E. Farhi and L. Susskind, "Technicolor," *Phys. Rept.* **74** (1981) 277.

[36] E. Eichten and K. D. Lane, "Dynamical Breaking of Weak Interaction Symmetries," *Phys. Lett.* **90B** (1980) 125–130.

[37] I. Dorsner, S. Fajfer, J. F. Kamenik, and N. Kosnik, "Can scalar leptoquarks explain the f(D(s)) puzzle?," *Phys. Lett.* **B682** (2009) 67–73, `arXiv:0906.5585 [hep-ph]`.

[38] P. H. Frampton, "Light leptoquarks as possible signature of strong electroweak unification," *Mod. Phys. Lett.* **A7** (1992) 559–562.

[39] I. Dorsner and P. Fileviez Perez, "Unification without supersymmetry: Neutrino mass, proton decay and light leptoquarks," *Nucl. Phys.* **B723** (2005) 53–76, `arXiv:hep-ph/0504276 [hep-ph]`.

[40] P. Yu. Popov, A. V. Povarov, and A. D. Smirnov, "Fermionic decays of scalar leptoquarks and scalar gluons in the minimal four color symmetry model," *Mod. Phys. Lett.* **A20** (2005) 3003–3012, `arXiv:hep-ph/0511149 [hep-ph]`.

[41] S. Davidson, D. C. Bailey, and B. A. Campbell, "Model independent constraints on leptoquarks from rare processes," *Z. Phys.* **C61** (1994) 613–644, `arXiv:hep-ph/9309310 [hep-ph]`.

[42] L. F. Abbott and E. Farhi, "Are the Weak Interactions Strong?," *Phys. Lett.* **101B** (1981) 69–72.

[43] B. Schrempp and F. Schrempp, "A Confining SU(2)-l X SU(2)-r Gauge Model of the Weak Interactions," *Nucl. Phys.* **B231** (1984) 109–138.

[44] J. C. Pati, "A Model for a Unification of Scales: From M(planck) to M(ν)," *Phys. Lett.* **B228** (1989) 228–234.

[45] B. Gripaios, A. Papaefstathiou, K. Sakurai, and B. Webber, "Searching for third-generation composite leptoquarks at the LHC," *JHEP* **01** (2011) 156, arXiv:1010.3962 [hep-ph].

[46] B. Gripaios, M. Nardecchia, and S. A. Renner, "Composite leptoquarks and anomalies in B-meson decays," *JHEP* **05** (2015) 006, arXiv:1412.1791 [hep-ph].

[47] L. J. Hall and M. Suzuki, "Explicit R-Parity Breaking in Supersymmetric Models," *Nucl. Phys.* **B231** (1984) 419–444.

[48] S. Dawson, "R-Parity Breaking in Supersymmetric Theories," *Nucl. Phys.* **B261** (1985) 297–318.

[49] G. F. Giudice and R. Rattazzi, "R-parity violation and unification," *Phys. Lett.* **B406** (1997) 321–327, arXiv:hep-ph/9704339 [hep-ph].

[50] W. Buchmüller and D. Wyler, "Constraints on SU(5)-type leptoquarks," *Phys. Lett. B* **177** (1986) 377.

[51] M. Krämer, T. Plehn, M. Spira, and P. M. Zerwas, "Pair production of scalar leptoquarks at the CERN LHC," *Phys. Rev.* **D71** (2005) 057503, arXiv:hep-ph/0411038 [hep-ph].

[52] A. Belyaev, C. Leroy, R. Mehdiyev, and A. Pukhov, "Leptoquark single and pair production at LHC with CalcHEP/CompHEP in the complete model," *JHEP* **09** (2005) 005, arXiv:hep-ph/0502067 [hep-ph].

[53] I. Dorsner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Kosnik, "Physics of leptoquarks in precision experiments and at particle colliders," *Phys. Rept.* **641** (2016) 1–68, arXiv:1603.04993 [hep-ph].

[54] **BaBar** Collaboration, J. P. Lees *et al.*, "Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays," *Phys. Rev. Lett.* **109** (2012) 101802, arXiv:1205.5442 [hep-ex].

[55] **BaBar** Collaboration, J. P. Lees *et al.*, "Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons," *Phys. Rev.* **D88** no. 7, (2013) 072012, arXiv:1303.0571 [hep-ex].

[56] **Belle** Collaboration, A. Matyja *et al.*, "Observation of B0 —> D*- tau+ nu(tau) decay at Belle," *Phys. Rev. Lett.* **99** (2007) 191807, arXiv:0706.4429 [hep-ex].

[57] **Belle** Collaboration, A. Bozek *et al.*, "Observation of $B^+ \to \bar{D}^{*0}\tau^+\nu_\tau$ and Evidence for $B^+ \to \bar{D}^0\tau^+\nu_\tau$ at Belle," *Phys. Rev.* **D82** (2010) 072005, arXiv:1005.2302 [hep-ex].

[58] **Belle** Collaboration, M. Huschle *et al.*, "Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle," *Phys. Rev.* **D92** no. 7, (2015) 072014, arXiv:1507.03233 [hep-ex].

[59] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$," *Phys. Rev. Lett.* **115** no. 11, (2015) 111803, arXiv:1506.08614 [hep-ex]. [Erratum: Phys. Rev. Lett.115,no.15,159901(2015)].

[60] B. Dumont, K. Nishiwaki, and R. Watanabe, "LHC constraints and prospects for S_1 scalar leptoquark explaining the $\bar{B} \to D^{(*)}\tau\bar{\nu}$ anomaly," *Phys. Rev.* **D94** no. 3, (2016) 034001, arXiv:1603.05248 [hep-ph].

[61] M. Tanaka and R. Watanabe, "New physics in the weak interaction of $\bar{B} \to D^{(*)}\tau\bar{\nu}$," *Phys. Rev.* **D87** no. 3, (2013) 034028, arXiv:1212.1878 [hep-ph].

[62] Y. Sakaki, M. Tanaka, A. Tayduganov, and R. Watanabe, "Testing leptoquark models in $\bar{B} \to D^{(*)}\tau\bar{\nu}$," *Phys. Rev.* **D88** no. 9, (2013) 094012, arXiv:1309.0301 [hep-ph].

[63] I. Dorsner, S. Fajfer, N. Kosnik, and I. Nisandzic, "Minimally flavored colored scalar in $\bar{B} \to D^{(*)}\tau\bar{\nu}$ and the mass matrices constraints," *JHEP* **11** (2013) 084, arXiv:1306.6493 [hep-ph].

[64] **LHCb** Collaboration, R. Aaij *et al.*, "Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$," *JHEP* **08** (2013) 131, arXiv:1304.6325 [hep-ex].

[65] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$," *Phys. Rev. Lett.* **111** (2013) 191801, arXiv:1308.1707 [hep-ex].

[66] **LHCb** Collaboration, R. Aaij *et al.*, "Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays," *Phys. Rev. Lett.* **113** (2014) 151601, arXiv:1406.6482 [hep-ex].

[67] **ATLAS** Collaboration, M. Aaboud *et al.*, "Search for scalar leptoquarks in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment," *New J. Phys.* **18** no. 9, (2016) 093016, arXiv:1605.06035 [hep-ex].

[68] **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) . https://cds.cern.ch/record/2205285.

[69] **CMS** Collaboration, "Search for pair-production of second-generation scalar leptoquarks in pp collisions at $\sqrt{s} = 13$ TeV with the CMS detector," *CMS-PAS-EXO-16-007* (2016) . https://cds.cern.ch/record/2139349.

[70] **ATLAS** Collaboration, G. Aad *et al.*, "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 [hep-ex].

[71] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Search for third-generation scalar leptoquarks and heavy right-handed neutrinos in final states with two tau leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP* **07** (2017) 121, arXiv:1703.03995 [hep-ex].

[72] **ATLAS** Collaboration, G. Aad *et al.*, "Searches for scalar leptoquarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *Eur. Phys. J.* **C76** no. 1, (2016) 5, arXiv:1508.04735 [hep-ex].

[73] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for third-generation leptoquarks and scalar bottom quarks in pp collisions at $\sqrt{s} = 7$ TeV," *JHEP* **12** (2012) 055, arXiv:1210.5627 [hep-ex].

[74] **CMS** Collaboration, G. L. Bayatian *et al.*, "CMS physics: Technical design report," (2006) .

[75] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC," *JINST* **3** (2008) S08004.

[76] L. Evans and P. Bryant, "LHC Machine," *JINST* **3** (2008) S08001.

[77] C. Lefevre, "LHC: the guide (English version). Guide du LHC (version anglaise)." Feb, 2009.

[78] **CMS** Collaboration, "CMS luminosity - public results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2016. [accessed October 18, 2017].

[79] **CMS** Collaboration, S. Abdullin *et al.*, "Design, performance, and calibration of CMS hadron-barrel calorimeter wedges," *Eur. Phys. J.* **C55** (2008) 159–171.

[80] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)," *JHEP* **01** (2001) 010, arXiv:hep-ph/0011363 [hep-ph].

[81] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159–177, arXiv:1410.3012 [hep-ph].

[82] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP* **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[83] A. Buckley *et al.*, "General-purpose event generators for LHC physics," *Phys. Rept.* **504** (2011) 145–233, arXiv:1101.2599 [hep-ph].

[84] 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*, pp. 287–319. 2013. arXiv:1304.6677 [hep-ph]. http://inspirehep.net/record/1229804/files/arXiv:1304.6677.pdf.

[85] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and collider physics," *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1–435.

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

[87] 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 [hep-ph].

[88] S. Frixione and B. R. Webber, "The MC and NLO 3.4 Event Generator," arXiv:0812.0770 [hep-ph].

[89] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156 [hep-ph].

[90] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond," *JHEP* **06** (2011) 128, arXiv:1106.0522 [hep-ph].

[91] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "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 [hep-ph].

[92] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions," *JHEP* **01** (2007) 013, arXiv:hep-ph/0611129 [hep-ph].

[93] S. Frixione, Z. Kunszt, and A. Signer, "Three jet cross-sections to next-to-leading order," *Nucl. Phys.* **B467** (1996) 399–442, arXiv:hep-ph/9512328 [hep-ph].

[94] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO," *JHEP* **12** (2012) 061, arXiv:1209.6215 [hep-ph].

[95] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.* **97** (1983) 31–145.

[96] B. R. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference," *Nucl. Phys.* **B238** (1984) 492–528.

[97] J.-C. Winter, F. Krauss, and G. Soff, "A Modified cluster hadronization model," *Eur. Phys. J.* **C36** (2004) 381–395, arXiv:hep-ph/0311085 [hep-ph].

[98] **CMS** Collaboration, V. Khachatryan *et al.*, "Event generator tunes obtained from underlying event and multiparton scattering measurements," *Eur. Phys. J.* **C76** no. 3, (2016) 155, arXiv:1512.00815 [hep-ex].

[99] **CMS** Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV," *CMS Physics Analysis Summary* **CMS-PAS-TOP-16-021** (2016) . https://cds.cern.ch/record/2235192.

[100] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[101] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Particle-flow reconstruction and global event description with the cms detector," *JINST* **12** (2017) P10003, arXiv:1706.04965 [physics.ins-det].

[102] **CMS** Collaboration, S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST* **9** no. 10, (2014) P10009, arXiv:1405.6569 [physics.ins-det].

[103] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," in *Proceedings of the IEEE*, pp. 2210–2239. 1998.

[104] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting," *J. Phys.* **G34** (2007) N343.

[105] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP* **04** (2008) 063, arXiv:0802.1189 [hep-ex].

[106] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J.* **C72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[107] **CMS** Collaboration, S. Chatrchyan *et al.*, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV," *JINST* **7** (2012) P10002, arXiv:1206.4071 [physics.ins-det].

[108] **CMS** Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV," *CMS Physics Analysis Summary* **CMS-PAS-PFT-10-003** (2010) . https://cds.cern.ch/record/1279347.

[109] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting," *Nucl. Instrum. Meth.* **A262** (1987) 444–450.

[110] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC," *eConf* **C0303241** (2003) TULT009, arXiv:physics/0306087 [physics.data-an]. [J. Phys.G31,N9(2005)].

[111] **CMS** Collaboration, V. Khachatryan *et al.*, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at 8 TeV," *JINST* **10** no. 06, (2015) P06005, arXiv:1502.02701 [physics.ins-det].

[112] **CMS** Collaboration, S. Chatrchyan *et al.*, "Performance of tau-lepton reconstruction and identification in CMS," *JINST* **7** (2012) P01001, arXiv:1109.6034 [physics.ins-det].

[113] **CMS** Collaboration, V. Khachatryan *et al.*, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS," *JINST* **11** no. 01, (2016) P01019, arXiv:1510.07488 [physics.ins-det].

[114] **CMS** Collaboration, "Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2," *CMS Physics Analysis Summary* **CMS-PAS-TAU-16-002** (2016) .

[115] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions," *Nucl. Phys.* **B406** (1993) 187–224.

[116] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266 [hep-ph].

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

[118] A. B. Galtieri, F. Margaroli, and I. Volobouev, "Precision measurements of the top quark mass from the Tevatron in the pre-LHC era," *Rept. Prog. Phys.* **75** (2012) 056201, arXiv:1109.2163 [hep-ex].

[119] **CMS** Collaboration, S. Chatrchyan *et al.*, "Identification of b-quark jets with the CMS experiment," *JINST* **8** (2013) P04013, arXiv:1211.4462 [hep-ex].

[120] **CMS** Collaboration, "Jet energy scale and resolution performances with 13TeV data," *CMS Detector Performance Summary* **CMS-DP-16-020** (2016) . https://cds.cern.ch/record/2160347.

[121] **CMS** Collaboration, S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST* **6** (2011) P11002, arXiv:1107.4277 [physics.ins-det].

[122] **CMS** Collaboration, V. Khachatryan *et al.*, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV," *Submitted to: JINST* (2016) , arXiv:1607.03663 [hep-ex].

[123] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Phys. Lett.* **B659** (2008) 119–126, arXiv:0707.1378 [hep-ph].

[124] P. Skands, S. Carrazza, and J. Rojo, "Tuning pythia 8.1: the monash 2013 tune," *European Physical Journal C* **74** no. 8, (2014) 1 – 39.

[125] **CMS** Collaboration, V. Khachatryan *et al.*, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV," *JINST* **10** no. 02, (2015) P02006, arXiv:1411.0511 [physics.ins-det].

[126] **CMS** Collaboration, "Boosted Top Jet Tagging at CMS," *CMS Physics Analysis Summary* **CMS-PAS-JME-13-007** (2014) . https://cds.cern.ch/record/1647419.

[127] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, "Stop Reconstruction with Tagged Tops," *JHEP* **10** (2010) 078, arXiv:1006.2833 [hep-ph].

[128] **CMS** Collaboration, "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging," *CMS Physics Analysis Summary* **CMS-PAS-JME-09-001** (2009) . https://cds.cern.ch/record/1194489.

[129] G. Kasieczka, T. Plehn, T. Schell, T. Strebler, and G. P. Salam, "Resonance Searches with an Updated Top Tagger," *JHEP* **06** (2015) 203, arXiv:1503.05921 [hep-ph].

[130] T. Lapsien, R. Kogler, and J. Haller, "A new tagger for hadronically decaying heavy particles at the LHC," *Eur. Phys. J.* **C76** no. 11, (2016) 600, arXiv:1606.04961 [hep-ph].

[131] **CMS** Collaboration, "Determination of residual jet-energy corrections using dijet events in 13 TeV data," *CMS Analysis Note (private)* **CMS-AN-15-254** (2015) . http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/254.

[132] H. Kirschenmann, *Jet Energy Scale Corrections and their Impact on Measurements of the Top-Quark Mass at CMS*. PhD dissertation, University of Hamburg, 2014.

[133] D. Rathjens, *Jet Energy Calibration and a Search for Supersymmetry with Vector Boson Fusion Channel Like Sign di-τ_h Final States*. PhD dissertation, University of Hamburg, 2015.

[134] **UA2** Collaboration, P. Bagnaia *et al.*, "Measurement of Production and Properties of Jets at the CERN anti-p p Collider," *Z. Phys.* **C20** (1983) 117–134.

[135] **D0** Collaboration, B. Abbott *et al.*, "Determination of the absolute jet energy scale in the D0 calorimeters," *Nucl. Instrum. Meth.* **A424** (1999) 352–394, arXiv:hep-ex/9805009 [hep-ex].

[136] **CDF** Collaboration, A. Bhatti *et al.*, "Determination of the jet energy scale at the collider detector at Fermilab," *Nucl. Instrum. Meth.* **A566** (2006) 375–412, arXiv:hep-ex/0510047 [hep-ex].

[137] **D0** Collaboration, V. M. Abazov *et al.*, "Jet energy scale determination in the D0 experiment," *Nucl. Instrum. Meth.* **A763** (2014) 442–475, arXiv:1312.6873 [hep-ex].

[138] M. Schröder, *Quality of Jet Measurements and Impact on a Search for New Physics at CMS*. PhD dissertation, University of Hamburg, 2012.

[139] **CMS** Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period," *CMS Physics Analysis Summary* **CMS-PAS-LUM-17-001** (2017) . https://cds.cern.ch/record/2257069.

[140] **CMS** Collaboration, "Single Muon efficiencies in 2012 Data," *CMS Detector Performance Summary* **CMS-DP-13-009** (2013) . https://cds.cern.ch/record/1536406.

[141] **CMS** Collaboration, "Search for dark matter produced in association with top quark pair in proton-proton collisions at sqrt(s) = 13 TeV," **CMS-EXO-16-049 (In preparation)** (2018) .

[142] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J.* **C71** (2011) 1554, arXiv:1007.1727 [physics.data-an]. [Erratum: Eur. Phys. J.C73,2501(2013)].

[143] **CMS** Collaboration, V. Khachatryan *et al.*, "Measurement of the $t\bar{t}$ production cross section using events in the eμ final state in pp collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J.* **C77** (2017) 172, arXiv:1611.04040 [hep-ex].

[144] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV," *Submitted to: Phys. Lett. B* (2016) , arXiv:1610.00678 [hep-ex].

[145] **CMS** Collaboration, "Measurement of inclusive W and Z boson production cross sections in pp collisions at sqrt(s)=13 TeV," *CMS Physics Analysis Summary* **CMS-PAS-SMP-15-004** (2015) . https://cds.cern.ch/record/2093537.

[146] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC," *JHEP* **07** (2011) 018, arXiv:1105.0020 [hep-ph].

[147] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhofer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi, "W^+W^- Production at Hadron Colliders in Next-to-Next-to-Leading-Order QCD," *Phys. Rev. Lett.* **113** (2014) 212001, arXiv:1408.5243 [hep-ph].

[148] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV," *Phys. Lett.* **B722** (2013) 5–27, arXiv:1210.6718 [hep-ex].

[149] **ATLAS** Collaboration, M. Aaboud *et al.*, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC," *Phys. Rev. Lett.* **117** no. 18, (2016) 182002, arXiv:1606.02625 [hep-ex].

[150] **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) . https://cds.cern.ch/record/2138504.

[151] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II," *J. Phys.* **G43** (2016) 023001, arXiv:1510.03865 [hep-ph].

[152] T. Müller, J. Ott, J. Wagner-Kuhr, "theta - a framework for template-based modeling and inference." http://www.theta-framework.org/. [accessed October 18, 2017].

[153] J. Ott, *Search for Resonant Top Quark Pair Production in the Muon+Jets Channel with the CMS Detector*. PhD dissertation, Karlsruher Institut für Technologie, 2012.

[154] T. Christensen, *Search for third-generation leptoquarks decaying fullhadronically into a top quark and a tau lepton*. Bachelor thesis, University of Hamburg, 2016.

[155] **CMS** Collaboration, "Tau identification in boosted topologies," *CMS Detector Performance Summary* **CMS-DP-16-038** (2016) . https://cds.cern.ch/record/2202971.

[156] **CMS** Collaboration, "Search for third-generation scalar leptoquarks decaying into a top quark and a muon," **CMS-B2G-16-027 (In preparation)** (2018) .

[157] R. K. Ellis, I. Hinchliffe, M. Soldate, and J. J. van der Bij, "Higgs Decay to tau+ tau-: A Possible Signature of Intermediate Mass Higgs Bosons at the SSC," *Nucl. Phys.* **B297** (1988) 221–243.

[158] **CMS** Collaboration, "Search for $t\bar{t}$ resonances in boosted semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV," *CMS Physics Analysis Summary* **CMS-PAS-B2G-15-002** (2015) . `https://cds.cern.ch/record/2138345`.

Danke!

Hiermit möchte ich mich bei allen bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Zunächst bedanke ich mich bei Johannes Haller für die Betreuung dieser Arbeit, den wöchentlichen Treffen sowie der Möglichkeit an Schulen, Konferenzen und Workshops teilnehmen zu können.

Bei Elisabetta Gallo bedanke ich mich für die Übernahme des Zweitgutachtens der Dissertation. Bei der gesamten Prüfungskommission bestehend aus Johannes Haller, Elisabetta Gallo, Andreas Meyer, Christian Sander sowie Günter Sigl bedanke ich mich für die Beurteilung der Disputation.

Ganz besonders bedanke ich mich bei Roman Kogler für die regelmäßigen Treffen, aus der zahlreiche Ideen zur Analyse-Strategie entstanden sind, und für die tatkräftige Unterstützung hinsichtlich der Veröffentlichung dieser Analyse.

Bei Thomas Peiffer bedanke ich mich für das Beantworten unzähliger technischer Fragen, für die Instandhaltung der gemeinsam genutzten Software und für das gründliche Lesen dieser Arbeit.

Many thanks to Robin Aggleton, Andreas Hinzmann, and Anastasia Karavdina for providing valuable modification proposals to this thesis.

Hartmut Stadie danke ich für die Betreuung der Service-Arbeit.

Für gute Stimmung im Büro bedanke ich mich bei Andreas Kell, Andreas Ditte, Alexander Fröhlich und Irene Zoi. Ganz besonders danke ich Tobias Lapsien und Arne Reimers, die mich seit meiner Masterarbeit (fast) dauerhaft begleitet haben. Danke für die zahllosen Diskussionen, das Suchen-Finden-Lösen von (meist däm-

lichen) Programmierfehlern und natürlich für den Spaß in den Mittagspausen.
Arne, danke für das gemeinsame Implementieren vieler Code-Segmente, für das sorgfältige Lesen dieser Arbeit und ganz besonders für das freundschaftliche Miteinander!

Meiner Familie danke ich für die mentale Unterstützung und ihren Glauben an mich. Ganz besonders danke ich meinen Eltern, die immer hinter mir stehen und mich während des gesamten Studiums tatkräftig unterstützt und mir Kraft gegeben haben. Danke, dass ihr immer an mich geglaubt habt, auch wenn ich es selbst nicht mehr tat.

Наталія, es war eine Freude mit dir die Service-Arbeit zu teilen ;-) Серьезно: Дякую за вашу поддержку и поощрение. Дякую за потрясающие моменты, которые у нас были и которые нам еще предстоит пережит. 🐓

Eidesstattliche Versicherung / Declaration on oath

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

Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium.

Die Dissertation wurde in der vorgelegten oder in einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Ort, Datum	Marc Stöver