



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

---

# Search for single vector-like quarks with the CMS detector

---

Dissertation zur Erlangung des Doktorgrades an der Fakultät für  
Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik der Universität Hamburg

VON

**Daniel González Vázquez**

**Hamburg**

**2019**

**Gutachter/innen der Dissertation:**

Prof. Dr. Johannes Haller

Prof. Dr. Elisabetta Gallo

**Zusammensetzung der Prüfungskommission:**

Prof. Dr. Günter Sigl

Prof. Dr. Johannes Haller

Prof. Dr. Elisabetta Gallo

PD Dr. Andreas B. Meyer

Dr. Roman Kogler

**Vorsitzender der Prüfungskommission:**

Prof. Dr. Günter Sigl

**Datum der Disputation:** 16.09.2019

**Vorsitzender Fach-Promotionsausschusses Physik:**

Prof. Dr. Günter Sigl

**Leiter des Fachbereichs Physik:**

Prof. Dr. Wolfgang Hansen

**Dekan der Fakultät:**

Prof. Dr. Heinrich Graener

Todo llega y todo pasa.

Nada eterno:

ni gobierno

que perdure,

ni mal que cien años dure.

-Tras estos tiempos vendrán  
otros tiempos y otros y otros...

(Antonio Machado)



# Abstract

In this thesis, proton-proton data is analysed in search of singly produced vector-like quarks. An integrated luminosity corresponding to  $35.9 \text{ fb}^{-1}$  was recorded with the CMS experiment in 2016 at the LHC at  $\sqrt{s} = 13 \text{ TeV}$ . The search is designed for vector-like quarks decaying into a W boson and a top quark with a muon or electron, hadronic jets, and missing transverse momentum in the final state. From the decay products, the invariant mass of the vector-like quark and the Standard Model background prediction, determined from control regions in data, are reconstructed. The event selection is optimised for the reconstructed invariant mass with simulation. Bottom quarks, boosted top quarks, as well as boosted W bosons reconstructed from hadronic decays and identified through algorithms, are exploited to increase the sensitivity of the search. No significant deviation from the background expectation of the Standard Model is observed, and exclusion limits at 95% confidence level are set for vector-like quark masses between 0.7 and 2 TeV. The exclusion limits on the product of the branching fraction and cross section range from 0.3 to 0.03 pb depending on the vector-like quark mass, the coupling, the relative width, and the type. Masses of up to 1660 GeV are excluded at a 95% confidence level for relative mass widths of 30%. The results of this analysis represent the most stringent results on single vector-like production, obtained in the  $tW$  channel.



---

# Zusammenfassung

Eine Suche nach vektorartigen Quarks am CMS-Experiment mit Daten aus Proton-Proton Kollisionen wird in dieser Arbeit vorgestellt. Die verwendeten Daten, die einer integrierten Luminosität von  $35.9 \text{ fb}^{-1}$  entsprechen, wurden bei einer Schwerpunktsenergie von 13 TeV aufgenommen. Die Suche fokussiert sich auf vektorartige Quarks als Signal, die in ein Top Quark und ein W Boson zerfallen und über elektroschwache Prozesse produziert werden. Die einzeln produzierten Signale enthalten im Endzustand: ein Myon oder Elektron, hadronische Jets und fehlende transversale Energie. Aus den Messungen im Detektor werden diese Teilchen und die invariante Masse der vektorartigen Quarks rekonstruiert. Um die Sensitivität der Suche zu steigern, werden multivariater Algorithmen zur Erkennung von Top Quarks, W Bosonen und Bottom Quarks auf Jets angewendet, vor allem bei hohem Impuls. Die Messungen in dieser Suche entsprechen der datengetriebenen Standardmodellvorhersage und kein Anzeichen für vektorartige Quarks ist vorhanden. Deshalb werden Ausschlussgrenzen auf das Produkt des Wirkungsquerschnitts und der Zerfallsrate der vektorartigen Quarks berechnet. Für den untersuchten Massenbereich der vektorartigen Quarks, zwischen 0.7 und 2 TeV, liegen die Ausschlussgrenzen zwischen 0.3 und 0.3 pb, bei einem Konfidenzniveau von 95%. Diese Ausschlussgrenzen sind abhängig von der Quarkkopplung, dem Quarktyp und der Zerfallsbreite. Die vorgestellte Suche schließt Massen bis zu 1660 GeV mit einer Zerfallsbreite von 30% aus und hat die beste Sensitivität im tW Kanal für die untersuchten Signale an Hochenergie-Experimenten.



# Contents

<b>1. Introduction</b>	<b>9</b>
<b>2. Standard Model and extensions</b>	<b>13</b>
2.1. Standard Model of particle physics . . . . .	13
2.1.1. Standard Model fermions . . . . .	14
2.1.2. Standard Model bosons . . . . .	15
2.1.3. Proton collisions at the LHC . . . . .	19
2.1.4. Limitations of the Standard Model . . . . .	21
2.2. Extensions of the Standard Model of particle physics . . . . .	23
2.2.1. Vector-like quarks . . . . .	25
<b>3. Experimental setup</b>	<b>29</b>
3.1. Large Hadron Collider . . . . .	29
3.2. CMS experiment . . . . .	31
3.2.1. Tracker . . . . .	33
3.2.2. Calorimeter . . . . .	35
3.2.3. Magnet . . . . .	38
3.2.4. Muon system . . . . .	39
3.2.5. Trigger . . . . .	40
<b>4. Event reconstruction and particle identification</b>	<b>43</b>
4.1. Particle flow algorithm . . . . .	43
4.1.1. Requirements . . . . .	44
4.1.2. Charged tracks . . . . .	45
4.1.3. Clustering . . . . .	47
4.1.4. Linking . . . . .	47

4.2. Charged leptons . . . . .	48
4.2.1. Muon identification . . . . .	48
4.2.2. Electron identification . . . . .	49
4.2.3. Lepton isolation . . . . .	50
4.3. Hadronic jets . . . . .	51
4.3.1. Jet clustering . . . . .	51
4.3.2. Pileup corrections . . . . .	53
4.3.3. Jet calibration . . . . .	55
4.3.4. Jet tagging . . . . .	58
4.4. Missing transverse momentum . . . . .	63
<b>5. Data and simulation</b>	<b>65</b>
5.1. CMS data sets . . . . .	65
5.2. SM simulation . . . . .	66
5.3. VLQ simulation . . . . .	70
5.4. HCAL simulation . . . . .	74
<b>6. Search for singly produced VLQs</b>	<b>81</b>
6.1. Top quark and W boson reconstruction . . . . .	84
6.1.1. Leptonic W boson . . . . .	84
6.1.2. Jet assignment and hypothesis selection . . . . .	85
6.1.3. Simulated VLQ reconstruction . . . . .	90
6.2. Event selection . . . . .	96
6.2.1. Trigger . . . . .	96
6.2.2. Selection optimisation . . . . .	99
6.2.3. Selected phase space . . . . .	104
6.2.4. Data and SM reconstruction . . . . .	107
6.3. Categorisation . . . . .	110
6.4. Signal region and background estimation . . . . .	114
6.4.1. Signal and sideband regions . . . . .	114
6.4.2. Background estimation . . . . .	117
6.4.3. Background validation . . . . .	118
<b>7. Systematic uncertainties</b>	<b>123</b>
<b>8. Results and interpretation</b>	<b>129</b>

---

<b>9. Summary &amp; outlook</b>	<b>139</b>
<b>A. Acronyms</b>	<b>161</b>
<b>B. Theory prediction</b>	<b>165</b>
<b>C. Signal injection tests</b>	<b>167</b>



# 1. Introduction

The pursuit of a self-consistent theory describing the fundamental quantities of elementary particle physics has driven the development of theory models, and their experimental verification. The Standard Model (SM) is the most successful theory describing elementary particles and their interactions, from the smallest to the largest scales in a plenitude of experiments. Although the SM incorporates many phenomena and the last predicted particle was discovered in 2012, some aspects are still unexplained and do not yet fit into a holistic picture. The most prominent imperfections are the missing integration of the gravitational force and the nature of dark matter and energy. Theories beyond the SM are developed to overcome these limitations, and tested experimentally, with the expectation of finding new phenomena which hint to a theory in which the SM is embedded.

So far the Large Hadron Collider (LHC) provides the source of the highest energies achieved in a particle accelerator laboratory experiment with a centre of mass energy of 13 TeV. Several detectors are operated at the LHC, where they record data in proton-proton and lead ion collisions. For theories beyond the SM, the typical energy scale up to which new phenomena can be probed and observed is a few TeV. A new quark generation is one of the phenomena that could hint to a specific set of theories, and could be found at the LHC. The models predicting additional quarks should not violate SM measurements such as; the couplings of the Higgs boson [1] or the exclusion limits from already performed searches. Quarks with left- and right-handed couplings, which are often referred to as vector-like, fulfil these requirements. In particular, their mass can be at the TeV scale or above, meaning that they may have thus far escaped detection. At a hadron collider such as the LHC, vector-like quarks (VLQs) could be produced either in pairs through the strong interaction or singly through their coupling to the weak interaction. The coupling of VLQs to the third quark generation

is expected to be dominant, due to their high mass and exclusion limits on the mixing with light quarks.

In this thesis, the data collected in 2016 with the Compact Muon Solenoid (CMS) detector with an integrated luminosity of  $36.9 \text{ fb}^{-1}$  is used to search for the production of single VLQs, which decay into a top quark and a W boson. These VLQs are produced in association with a quark that emits a Z or W boson and a heavy quark from a gluon splitting. Three charge conserving production channels are considered for two types of VLQs:  $B+b$ ,  $B+t$ , and  $X_{5/3}+t$ , where the B has a charge of  $-1/3$  and the  $X_{5/3}$  a charge of  $5/3$ . The studied final states contain several hadronic jets, one muon or electron, and missing transverse momentum. One of the jets is close to the beamline and originates from a light quark emitting a vector boson. This type of jet is rare in most SM processes and is utilised for the SM background prediction. Additionally, jet tagging is exploited to identify bottom quarks, boosted top quarks, and boosted W bosons in their fully hadronic decay modes. These identified quarks and bosons are used to classify the final state into several categories. The missing transverse momentum in the events is expected to arise from a neutrino leaving the experiment undetected. The invariant mass of the VLQ is reconstructed by combining the objects in the detector, using the decay properties of the VLQ, the W boson, and the top quark. The invariant VLQ mass is the most discriminative variable, and the sensitivity of the search is optimised for it in simulation. For the final evaluation, a data-driven background prediction is used to minimise the influence of uncertainties from simulation on the predicted background.

Many searches for VLQs have been performed at the LHC both by A Toroidal LHC Apparatus (ATLAS) and CMS [2–28], excluding VLQs up to masses of 1.3 TeV. While the production cross section of vector-like pairs depends on the mass, for singly produced VLQs the production cross section can be much larger for high masses, depending on the vector-like coupling to the electroweak sector [29]. The Feynman diagrams for the simulation of the signal are calculated at leading order for VLQs with narrow and wider widths. Previous results in this channel from ATLAS at 8 TeV are presented in Ref.[19]. The results presented in this thesis are published in Ref. [30] and are the first results in this channel at 13 TeV. They are the most stringent limits in this channel and extend the investigated phase space up to masses of 2 TeV. With the current experimental data, several model scenarios containing VLQ with widths

larger than 10% are already excluded. The additional data collected in the next years will make it possible to further explore wider as well as narrow width scenarios. The full data of the LHC after the high luminosity upgrade might even make it possible to measure or constrain the couplings of VLQs.

The structure of this thesis is as follows: in chapter 2 a brief description of the SM is provided including its limitations and possible extensions. The focus is on the phenomenological model-independent interpretation of VLQs. Thereafter, in chapter 3, the designs of the LHC and CMS experiment are presented. The reconstruction of the global event quantities and the identification of particles are described in chapter 4, while the details about the data and simulation used are explained in chapter 5. A description of the algorithms and selection used in the search for VLQs is provided in chapter 6. The uncertainties are covered in chapter 7, with the final statistical interpretation discussed in chapter 8.



## 2. Standard Model and extensions

The SM of particle physics has been developed and established during the last decades, as the theoretical framework predicting the behaviour of all known elementary particles, as well as their interactions. It is one of the cornerstones of human understanding of the laws of the universe and successfully explains almost all properties of the visible matter. The Higgs boson discovered in 2012 [31, 32] was the last missing elementary particle of the SM. Even though the picture of the SM was completed, a plethora of unanswered questions remain; examples include the inflation of the universe, the baryon asymmetry, and the inclusion of the gravitational force. Many extensions to the SM are developed and probed through experiments to overcome these limitations.

Reviews of the SM can be found in Refs. [33–35]. The cited values are taken from Ref. [36] if not otherwise indicated. The reference includes a short review of the SM, many models beyond it, and the description of many techniques currently used at elementary particle experiments.

### 2.1. Standard Model of particle physics

The SM is a renormalisable quantum field theory developed in the second half of the last century. The theoretical formulation was established in the 1970s, and experimental evidence of the last predicted elementary particle, the Higgs boson, was found in 2012. The usual Lagrangian formalism predicts the existence of two types of elementary particles: fermions and bosons. The bosons are the force carrying elementary particles, while the fermions are amongst other the constituents of baryonic matter. All elementary particles are defined by having no inner structure, thus being indivisi-

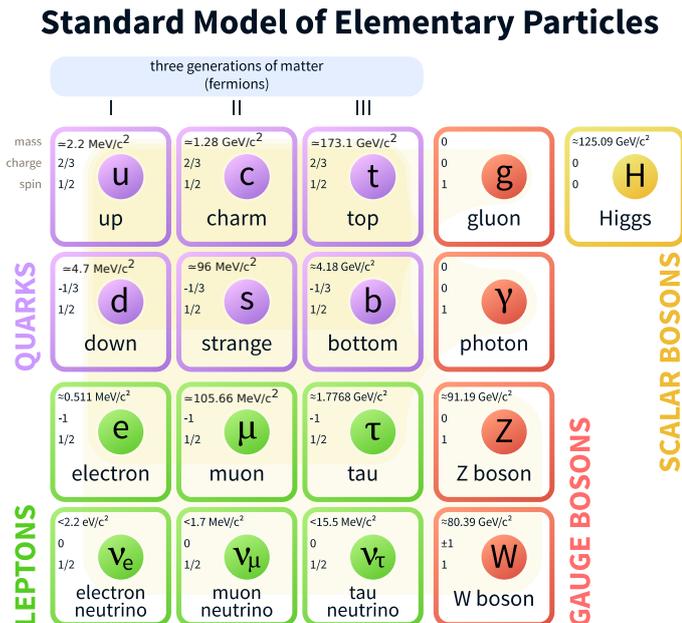


Figure 2.1.: Depiction of the particle content (fermions and bosons) of the SM. The mass, electric charge, and spin of the elementary particles are shown [37]

ble. The properties of the known elementary particles are briefly discussed in the next paragraphs. Their behaviour is essential for detecting new phenomena and exploring the data collected with the CMS detector.

### 2.1.1. Standard Model fermions

Each known elementary particle has an antiparticle that has the same quantum numbers, but for the opposite electric charge. For the photon, Higgs, and Z boson the particle coincides with its antiparticle. Some properties of the SM particle content are illustrated in Fig. 2.1, including the SM particle content.

The first set of elementary particles discussed, are the fermions, which have half-integer spin and therefore follow the Fermi-Dirac statistics. Two types of fermions are distinguished, namely quarks and leptons. In addition to the spin quarks also have other quantum numbers: electrical charge, colour charge, isospin, hypercharge, baryon number, charmness, strangeness, bottemness, and topness. The electric charge

of quarks is either  $2/3 e$  for up type or  $-1/3 e$  for down type quarks. The up and down quarks are the lightest quarks and were discovered first. They are usually referred to as the first generation. Today, three different generations of up (down) quarks are known: up (down), charm (strange) and top (bottom). A quantum number that only quarks have is the colour charge. Every quark has one of the three colour charges often called red, green, and blue or the corresponding anticolour. Composite free particles are colourless, which is achieved by, for example, combining three quarks each with a different colour (anticolour) like in baryons or by combining two quarks with colour and corresponding anticolour like in mesons. Every quark has a quantum number; topness is the quantum number of the top and antitop quarks, which is conserved under strong and electromagnetic interactions, but not weak interactions. Charmness, strangeness, and bottomness are the quantum numbers of the other corresponding quarks. As for the quarks, three generations of leptons exist, and every generation has one electrically charged (with the elementary charge  $e$ ) and one neutral lepton. The charged leptons are named electron, muon, and tau. The neutrinos are named according to the charged leptons: electron neutrino, muon neutrino, and tau neutrino. The set of quantum numbers for the leptons include the aforementioned electric charge and spin as well as isospin, weak hypercharge, and lepton number.

### 2.1.2. Standard Model bosons

Apart from the fermions there are also bosons in the SM, which are characterised by an integer spin and follow the Bose statistics. The bosons carry the forces and are often divided into the quantum chromodynamics (QCD) and the electroweak sector. Both sectors are governed by the gauge symmetry of the SM, which is the unitary product of

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \quad (2.1)$$

Here,  $SU(3)_C$  and  $SU(2)_L$  are the special unitary group, with  $C$  denoting the colour charge symmetry. The subscript  $L$  is added to indicate that the symmetry is only valid for left-handed leptons, making the SM a chiral theory. The  $U(1)_Y$  is the unitary group, and the  $Y$  indicates the symmetry with respect to the weak hypercharge.

The QCD sector is based on the  $SU(3)_C$  with a Lagrangian of the form

$$\mathcal{L}_{QCD} = -\frac{1}{4} \sum_{A=1}^8 F^{A\mu\nu} F_{\mu\nu}^A + \sum_{j=1}^{n_f} \bar{q}_j (iD_\mu \gamma^\mu - m_j) q_j. \quad (2.2)$$

Here,  $\gamma^\mu$  is the representation for the Dirac matrices,  $D_\mu$  is the covariant derivative of the QCD sector, and  $q$  denotes the quark fields. The quark masses are denoted  $m_j$ , and their colour charges  $n_f$ . The gluon field strength tensor is written as  $F^{A\mu\nu}$ , where  $A$  is the representation of the eight gluon fields. The gluons are the massless vector gauge bosons of the QCD with Spin 1 that carry colour charge. They are responsible for the strong interaction, which is the mechanism behind the nuclear force. The quarks are bound through the nuclear force into hadrons. Gluons can only interact with quarks or other gluons. At the LHC highly energetic gluons are produced at high rates. They decay into pairs of quarks (e.g., top or bottom quark pairs). Apart from the quark masses, the coupling strength  $\alpha_s$  is a fundamental parameter of the QCD where the QCD coupling strength depends on the energy scale. It is very strong for low energies and weak at high energies, which is known as asymptotic freedom. At energies in the order of 150 MeV, a quark can be considered free. For lower energies, coloured particles are confined. These coloured particles cannot exist individually, and additional quarks and gluons are spontaneously created from the vacuum, forming hadrons. This process is also known as hadronisation.

In the electroweak sector, the Lagrangian can be separated into a symmetric  $\mathcal{L}_{sym}$  and the Higgs  $\mathcal{L}_{Higgs}$  Lagrangian:

$$\mathcal{L}_{EW} = \mathcal{L}_{sym} + \mathcal{L}_{Higgs}. \quad (2.3)$$

The electroweak Lagrangian  $\mathcal{L}_{EW}$  represents the unification of the electromagnetic and the weak forces. The symmetrical part is a Yang-Mills theory including the fermions and the gauge bosons with a  $SU(2) \times U(1)$  symmetry and takes the form

$$\mathcal{L}_{sym} = -\frac{1}{4} \sum_{A=1}^3 F_{\mu\nu}^A F^{A\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\psi}_L i\gamma^\mu D_\mu \psi_L + \bar{\psi}_R i\gamma^\mu D_\mu \psi_R. \quad (2.4)$$

The  $\psi$  are the fermion fields, understood in this equation as the sum over all flavours of quarks and leptons. The  $\gamma^\mu$  and  $D_\mu$  follow the same definition as in Eq. 2.2 and the gauge strength tensors are given by

$$F_{\mu\nu}^A = \partial_\mu W_\nu^A - \partial_\nu W_\mu^A - g\epsilon_{ijk}W_\mu^iW_\nu^j, \quad (2.5)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (2.6)$$

Here the gauge field  $B_\mu$  is associated with the  $U(1)_Y$  group, while  $W_\mu^A$  corresponds to the three  $SU(2)$  generators. The antisymmetric Levi-Civita tensor is denoted  $\epsilon_{ijk}$ , the partial derivatives  $\partial_\mu$ , and the electroweak coupling  $g$ .

The other part of the Lagrangian is composed by the Higgs potential and dynamics. Where the potential allows the spontaneous symmetry breaking, giving rise to the mixing and the masses of the electroweak gauge bosons. The Higgs potential has the form

$$V(\phi^\dagger\phi)_{Higgs} = -\frac{1}{2}\mu^2\phi^\dagger\phi + \frac{1}{4}\lambda(\phi^\dagger\phi)^2. \quad (2.7)$$

Here  $\mu$  and  $\lambda$  are two real valued parameters, with  $\mu$  having the dimension of a mass and  $\lambda$  being dimensionless. The Higgs field  $\phi$  is complex. With the unitary gauge of the  $\mathcal{L}_{Higgs}$ , the non-vanishing vacuum expectation value of the Higgs field is

$$\langle\phi\rangle = \frac{\nu}{\sqrt{2}} = \sqrt{\frac{\mu^2}{2\lambda}}, \quad (2.8)$$

where  $\nu$  is approximately 246 GeV and is the only SM parameter associated with a unit. Many features of the Higgs boson are measured at the LHC, such as the decay rate into several elementary particles. Detailed reviews on the status and measurements can be found in Ref. [36].

Through the spontaneous symmetry breaking, the observed bosons obtain their mass and the vector fields are rotated by the Weinberg angle  $\theta_W$ . The measured bosons are  $W^+$ ,  $W^-$ , and  $Z$ , which carry the weak interaction. The electromagnetic interactions are carried by  $\gamma$ . The neutral bosons can be expressed by the Weinberg angle as the rotation of the three neutral  $W$  and the neutral  $B$  as

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos(\theta_W) & \sin(\theta_W) \\ -\sin(\theta_W) & \cos(\theta_W) \end{pmatrix} \begin{pmatrix} B \\ W \end{pmatrix}. \quad (2.9)$$

The masses of the W boson ( $m_P W$ ) and Z bosons ( $m_Z$ ) are connected to the coupling and the vacuum expectation value through the following:

$$m_W = \frac{1}{2}\nu g, \quad (2.10)$$

$$m_Z = \frac{m_W}{\cos(\theta_W)}. \quad (2.11)$$

The transformation probability or mixing of one quark into another quark is displayed as values in the Cabibbo-Kobayashi-Maskawa (CKM) matrix. In general, these probabilities are represented by three real values and a complex phase. The values are obtained assuming unitary and are often expressed as the normalised probabilities  $|V_{ij}|$ , where  $i$  and  $j$  are the quark flavours. The values for the CKM matrix are

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.974 & 0.2243 & 0.004 \\ 0.218 & 0.997 & 0.042 \\ 0.008 & 0.04 & 1.019 \end{pmatrix}. \quad (2.12)$$

The  $|V_{tb}|$  indicates that the top quark almost always decays into a b quark. To do so, it emits a W boson. The W boson further decays into elementary particles: either a charged lepton with the corresponding neutrino, or a pair of quarks. The branching ratios of the former are about 10% per lepton flavour ( $\mu, e, \tau$ ), and about 68% for the decay into hadrons.

For this thesis, the quarks of the first and second generations are treated as massless in comparison to the scale at the LHC. The top quark, bottom quark, and W boson are especially important, and their properties are used to identify them. Top quarks are the only quarks known to decay before hadronisation. For the search presented, it is assumed that  $|V_{tb}|^2 = 1$ , i.e. the top quark in this thesis is assumed to always decay to a bottom quark and a W boson. Bottom quarks have a much lower mass than the top quark or W boson but have a longer lifetime than the other quarks. The lifetime is used to distinguish bottom quarks from other quarks and gluons.

### 2.1.3. Proton collisions at the LHC

At the LHC, protons beams are accelerated to velocities near the speed of light and collided head on. Protons are made out of three valence quarks (uud), gluons, and sea quarks. The constituents of a proton are also known as partons.

A factorisation ansatz is used to calculate the cross sections  $\sigma_{pp}$  at proton-proton collisions, given by

$$\sigma_{pp} = \sum_{i,j} \int \int dx_1 dx_2 f_i(x_i, Q^2) f_j(x_j, Q^2) \hat{\sigma}_{ij}(x_1, x_2, Q). \quad (2.13)$$

The parton momenta fractions involved are denoted as  $x_1$  and  $x_2$ . Here,  $Q$  is the energy scale (momentum transfer) of the collision. The hard interaction matrix element is expressed as  $\hat{\sigma}_{ij}$ . It is calculated perturbatively and depends on  $Q$ , as well as  $x_1$  and  $x_2$ . In this formula  $i$  and  $j$  denote the possible parton flavours. For the two partons involved, the structure function  $f_i$  is required. Structure functions, also known as parton distribution functions (PDFs), are typically measured in deep inelastic scattering (DIS) experiments where an electron is used to probe a proton. The DIS results with the highest energies were obtained at the HERA accelerator [38] with a maximal proton energy of 920 GeV and an electron energy of 27.5 GeV. These energies are much smaller than the 13 TeV reached at the LHC. To use the measurements from HERA at higher energies, the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [39–42] equations of the QCD evolution are used.

At proton colliders, elementary particles are produced with high energies by the annihilation of two partons from the protons, which is also referred to as “hard process”. Before and after the hard interaction, additional gluons can be radiated which is also known as initial and final state radiation. By the decay of the highly energetic elementary particles the energy scale is reduced. The gluons and quarks with reduced energy begin to split and create additional partons. This process is referred to as parton shower [44]. The process of the shower evolution is strongly influenced by the PDFs, and the evolution can be calculated perturbatively down to about 1 GeV in  $Q^2$ . The particles from the parton showers at colliders are usually within a cone inside a small part of the detectors. The location of the particles is used to reconstruct the hard quarks and gluons in cones, which are often referred to as “jets”. For highly

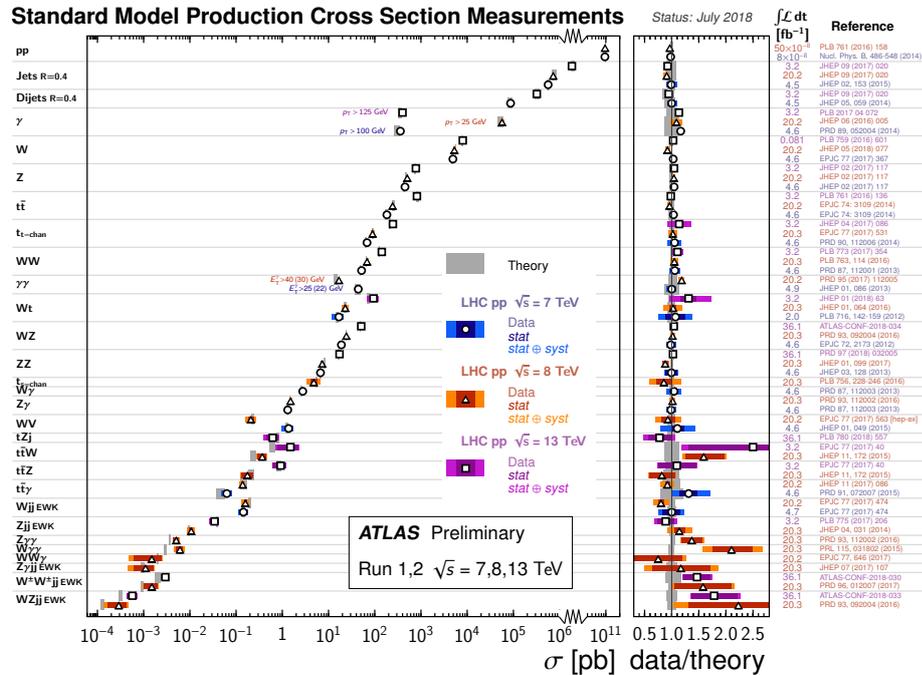


Figure 2.2.: Summary of several SM total production cross section measurements, corrected for leptonic branching fractions and compared to the corresponding theoretical expectations. All theoretical expectations were calculated at NLO or higher. The dark-coloured error bar represents the statistical uncertainty. The lighter-coloured error bar represents the full uncertainty, including systematics and luminosity uncertainties. The data/theory ratio, luminosity used, and reference for each measurement are also shown. Uncertainties for the theoretical predictions are quoted from the original ATLAS papers. They were not always evaluated using the same prescriptions for PDFs and scales. Not all measurements are statistically significant yet. Figure and text taken from Ref. [43]

energetic events these jets do not only contain the energy of single quarks or gluons but the energy of several particles. Typical examples include highly energetic top quarks decaying into a b quark and a W boson where the W boson decays to quarks. For high top quark momenta the three quarks could be reconstructed within a jet. These jets are different from a jet reconstructed from gluon emissions by an additional sub-structure in the energy distribution. The so-called “jet substructure” is associated with the quarks and often found by redoing the jet reconstruction to find smaller jets (“subjets”). With these jets, additional variables are calculated such as the combined invariant mass of the subjets, the energy flow axis, or others.

In Fig. 2.2 a summary of the measurements of cross sections conducted at the LHC by ATLAS [43] is presented (similar results are also available for CMS). The inelastic proton cross section with about 78 mb is the highest, followed by the production of jets, that is much smaller. Other SM processes like top quark pair production have a cross section, which is even more orders of magnitude smaller. To be able to measure the properties of SM top quark pairs, of which millions are produced at the LHC, the multijet and elastic scattering events must be filtered out with very high efficiency. In comparison, the analysis presented in this thesis has an expected signal cross section in the order of  $\mathcal{O}(100)$  fb.

#### 2.1.4. Limitations of the Standard Model

Following this brief summary of the particle content, some properties of the SM and its prominent limitations are discussed below. Even though the SM predicts the existence of all experimentally measured elementary particles and interactions, it cannot explain some observations. Also, considerations on the theoretical framework and its scales suggest that further mechanisms exist.

#### Dark matter and energy

From gravitational effects observed in distant galaxies such as gravitational lensing or the high rotation speed of the visible matter [45–48] it is deduced that apart from the

baryonic matter, additional matter and energy exist. These are called dark matter and dark energy since they do not interact electromagnetically and are therefore difficult to detect. There is no mechanism in the framework of the SM to explain dark matter and dark energy. If these effects are based on additional matter, the universe is made of 4.8 % baryonic matter, 26.8 % dark matter, and 68.3 % dark energy.

### **Gravitation**

Gravitation is not part of the SM, and it is not clear how to incorporate it. Any attempt to formulate it as quantum field theory has been fruitless. Additional terms in the SM Lagrangian usually lead to ultraviolet divergences and cannot be renormalised. Current and planned experiments are far away from the “Planck scale” (roughly  $10^{-35}$  m) where the gravitation would become important for highly energetic elementary particles. The missing description of the gravitation suggests that the SM is embedded in a more general theory and the current form is an effective field theory. A review can be found in Ref. [49].

### **Baryon asymmetry**

Another observation is that there is more matter than antimatter in the universe. This means that at the beginning of the universe there had to be a large amount of charge conjugation parity (CP) violation to create the baryon asymmetry now observed. Otherwise, the matter and antimatter would have annihilated and our galaxies would not exist. The amount of CP violation measured so far in experiments is not enough to explain this. A detailed review of this topic can be found in Ref. [50].

### **Naturalness**

One theoretical consideration is often termed the “naturalness problem” [51]. Theory parameters in a model are expected to have the same order of magnitude to be considered “natural”. These considerations are closely related to fine-tuning and the

hierarchy problem. In the SM, several observations violate this consideration by many orders of magnitude. One example is the difference in strength between the gravitation and the electroweak force. Where the latter is 24 order of magnitude stronger. An other is the mass of the Higgs boson, which is 17 orders of magnitude smaller than the Planck scale, or the size of the cosmological constant.

### **Vacuum stability**

The discovery of the Higgs boson also included the measurement of its vacuum energy. Analysis of the quantum corrections of the top quark, the Higgs boson, and the measured SM quantities have demonstrated that the electroweak vacuum might be metastable, leading to a possible phase transition of the universe to a stable state. No known principle forces the universe to be in the stable or even metastable state [52].

### **Neutrinos**

The measurements of neutrinos make it clear that they cannot be massless [36, 53], but that they have a small mass in contrast to the SM prediction. One of the other important discoveries of neutrinos is their oscillation [54, 55]. The measured neutrinos are always an admixture of the three known states. Apart from their mass, the nature of the neutrinos is also under experimental investigation and theoretical discussion. Nature refers to whether neutrinos behave like Dirac- or Majorana-spinors or are different from both.

## **2.2. Extensions of the Standard Model of particle physics**

Several classes of models have been proposed to overcome these limitations. Some act on an energy scale far beyond the reach of any planned collider, such as grand unified theories. It is not clear if these models can be verified by experiments to come.

Apart from these theories, many models exist that give rise to new particles at energies which could be far below the Planck scale. Some models that predict new particles are introduced below.

### **Supersymmetry**

One of the most prominent extensions that predicts new particles is called supersymmetry. Lectures can be found in, for example Refs. [56, 57]. For each particle the unbroken form of this model adds a supersymmetric partner with the same quantum numbers except for the spin, which differs by one half. Since such particles have not been observed, supersymmetry has to be broken. The supersymmetric partners can have higher masses or special configurations and thus escape detection. Such supersymmetric models have a large set of free parameters. Benchmark points or simplified versions with additional constraints are often considered in searches for these models. The simplified models often have an additional symmetry called “*R*-parity”. If this symmetry is not broken, supersymmetric particles always decay into at least another supersymmetric particle. A large list of results from ATLAS and CMS can be found at Refs. [58, 59].

### **Extra dimensions**

In the 1920s, Kaluza and Klein proposed that more than the known four (three space and one time) dimensions exist [60]. They assumed that the additional dimensions are curled up. Even though the original idea has been disproven, new models based on the idea of extra dimensions have been developed and could explain the differences in strength between the gravitational force and other interactions. Theories involving warped extra dimensions can be constructed at the TeV scale and provide a path to a grand unification [61]. Constraints on these models arise from cosmological, astrophysical, collider, and tabletop experiments, investigating sub-millimetre distances. The searches at the LHC dominate the constraints from collider experiments.

## Composite and little Higgs

The principal idea of composite Higgs models [62], is that the Higgs boson is a pseudo-Nambu-Goldstone boson made out of elementary particles. As in the QCD sector, a new scale  $\Lambda_{\text{EW}}$  for the effective Lagrangian is introduced, which could be as low as a few TeV. If this analogy holds, the light mass of the Higgs can be explained, addressing part of the naturalness, and additional elementary particles at this scale are expected.

In the little Higgs model scenarios as reviewed in Ref. [63, 64], the SM Higgs boson is also a pseudo-Nambu-Goldstone boson. It arises from spontaneous symmetry breaking at the TeV scale, and the main idea is that the product of several symmetry groups forms the gauge group. These symmetry groups are realised in additional space dimensions such that these models have 3+1 space dimensions while retaining many properties of extra dimension models. Similar to the composite Higgs models, new elementary particles would appear at the TeV scale.

### 2.2.1. Vector-like quarks

The idea to extend the quark content beyond the three generations has been around for a long time. The first approach was to add a fourth chiral generation, which has been studied thoroughly in theory and searched for at different colliders. With the discovery of the Higgs at a mass of 125 GeV, a simple addition of another generation into the SM Lagrangian is no longer possible [65]. New quarks need to be heavy since no additional quarks have been observed so far. Their contribution to the Higgs production and decay would be dominant and violate the measurements. Nevertheless, quarks with left- and right-handed couplings, often named VLQs [29, 66], can still exist without influencing the measurement of the Higgs properties. In contrast to their SM counterparts, their contribution in loop diagrams would cancel out. The name is derived from the vector-like structure, which is only present in beyond the Standard Model (BSM) theories. For these theories the electroweak currents  $j^\mu$  act independent

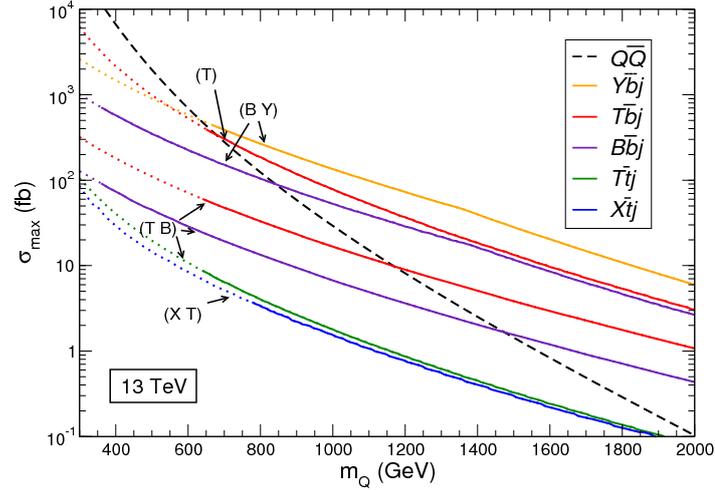


Figure 2.3.: Maximal allowed production cross section for several VLQs [29].

of the left- (LH) or right-handed (RH) chirality on the vector-like fermions  $f$ , which is formulated as

$$\begin{aligned}
 j_{\text{LH}}^\mu &= \bar{f}_{\text{LH}} \gamma^\mu f'_{\text{LH}}, & j_{\text{RH}}^\mu &= \bar{f}_{\text{RH}} \gamma^\mu f'_{\text{RH}}, \\
 j^\mu &= j_{\text{LH}}^\mu + j_{\text{RH}}^\mu = \bar{f} \gamma^\mu f'.
 \end{aligned}
 \tag{2.14}$$

As before,  $\gamma^\mu$  denotes the Dirac matrices.

Several models where VLQs arise have been proposed, such as non-minimal supersymmetry, extra dimensions, composite Higgs, and little Higgs [67]. At the LHC these additional quarks can be produced through either the strong or the electroweak force. In the former case they would be produced in pairs, while in the latter case they would be produced singly. Pair production is well studied and the production cross section, which depends on the mass of the new resonance, is well known. For single production, the coupling of the VLQ to the SM plays an important role, and the cross section is more model dependent. This leads to scenarios, especially at high VLQ masses, where the cross sections of singly produced VLQs are orders of magnitude larger than the cross sections of pair produced VLQs. The possibly higher cross sections of VLQs in the single production mode make them candidates to be detected at the LHC.

In Fig. 2.3 from Ref. [29], the maximal allowed production cross sections for selected single VLQs are illustrated. The constraints on the cross sections are taken from measurements of the oblique parameters and other SM measurements. From direct searches, constraints on VLQs exist for masses of about 1.3 TeV [68]. Since single VLQs

appear in many models with similar kinematical, production, and decay properties, several approaches have been proposed to search for them in a model-independent way. The approach presented in Ref. [69, 70] is followed in this thesis.

In general, VLQs can appear in singlets or multiplets; commonly, four different types are considered: B, T,  $X_{5/3}$ , and  $Y_{4/3}$ . The first two are the partners of the b and t quarks and share their electric charge. The  $X_{5/3}$  and  $Y_{4/3}$  have more exotic electric charges of  $5/3e$  and  $-4/3e$ . In this thesis, couplings to the heaviest quark generation are considered. The high mass of the VLQs and the mass constraints justify this assumption.

For the single VLQ production modes studied, it is assumed that the electric charge is conserved. At leading order single VLQs at the LHC, are mostly produced by the fusion of a top or bottom quark with a vector boson, usually a W or Z boson. The production with the Higgs is commonly suppressed by several orders of magnitude. The vector boson is usually emitted from a quark under a small scattering angle. The top or bottom quark from the fusion with a boson is normally generated by a gluon splitting. The quark or boson that does not take part in the hard interaction is expected to decay in the central part of the detector, yielding an additional bottom or top quark in the event, with low transverse momentum. The described production mechanism is very similar to the single top quark production and very different from most other SM processes.



## 3. Experimental setup

The typical experimental setup in high energy physics consists of a particle accelerator and one or several particle detectors. Particle accelerators bring two beams of particles, consisting of packages referred to as “bunches”, to velocities near the speed of light, and collide them “head on” at interaction points within the particle detectors. These detectors record data to measure the properties of elementary particles. The data analysed in this thesis were recorded with the CMS experiment at the LHC circular proton-proton accelerator at Conseil Européen pour la Recherche Nucléaire (CERN) with a centre of mass energy of 13 TeV. In the following sections the main features of the accelerator (Sec. 3.1) and the CMS detector (Sec. 3.2) are presented. A more detailed description can be found in the technical design reports [71, 72] that provide the basis for this chapter.

### 3.1. Large Hadron Collider

The Large Hadron Collider (LHC) is located at the European centre of high energy physics, CERN, near Geneva. It is the synchrotron accelerator with the largest circumference of about 27 km and has been operational since 2008. The LHC is designed to accelerate two beams consisting of protons or lead ions up to energies of 7 TeV per beam. The machine is currently operated at 13 TeV centre of mass energy for proton-proton collisions, which is the highest energy ever achieved at an accelerator facility. Apart from the high energy, the LHC also provides a very high instantaneous luminosity  $\mathcal{L}$  given by

$$\mathcal{L} = F \frac{\gamma f k_B N_p^2}{4\pi \epsilon_n \beta^*}. \quad (3.1)$$

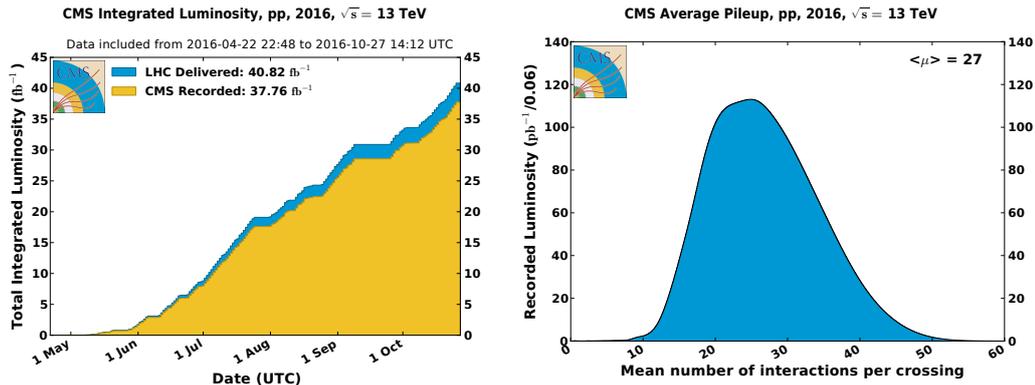


Figure 3.1.: Distributions of the integrated luminosity (left) and the number of interactions per bunch crossing (right) recorded with the CMS detector in 2016 [73].

In this equation  $\gamma$  is the Lorentz factor,  $f$  is the frequency at which the proton bunches pass,  $k_b$  is the number of bunches,  $N_p$  is the number of protons per bunch,  $\epsilon_n$  is the transverse beam emittance, and  $\beta^*$  is the beta function at the interaction point. Effects that reduce the luminosity (e.g., the angle between the beams) are taken into account through  $F$ . The quantities determining the luminosity are mostly the magnetic structure of the LHC. Both the emittance and the beta function at the interaction point are tuned by adjusting the focusing quadrupole magnets before the collision point. This tuning is a challenging task due to, for example, the electromagnetic repulsion of the protons in the bunches, or the billions of collisions the bunches experience during the regular operation of about 10 hours.

The proton bunches are accelerated to 450 GeV using a chain of accelerators including the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) before injecting them into the LHC. In the regular operation scheme, 2808 proton bunches circulate in opposite directions in the LHC with a spacing of 25 ns between the bunches. This corresponds to a bunch crossing frequency of 40 MHz where each bunch consists of about  $10^{11}$  protons. The peak instantaneous luminosity achieved in 2016 at the CMS interaction point is  $\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , while an integrated luminosity of  $37.76 \text{ fb}^{-1}$  was collected with an uncertainty of 2.5% [74]. On average, 27 interactions per bunch crossing (hereafter referred to as pileup) are observed in the 2016 data. In Fig. 3.1 the distributions of the integrated luminosity as a function of time (left) and the distribution of the number of pileup events (right) are shown.

The LHC has by now exceeded most design parameters except for the centre of mass energy and delivered the largest ever recorded data set of proton-proton collisions. It is supposed to run under similar conditions up to 2026. The LHC will then be further enhanced to achieve a higher instantaneous luminosity.

## 3.2. CMS experiment

The Compact Muon Solenoid (CMS) experiment is one of the particle detectors located at the LHC ring. Alongside ATLAS it is one of the two multipurpose detectors that discovered the Higgs boson. The CMS detector has an onion-like design with layers of varying detector types and a magnet. It can measure most of the SM particles with high precision and search for new elementary particles arising in BSM theories. To withstand the radiation and maintain its excellent performance, it is under permanent development.

The different parts of the detector are briefly discussed in this section, focusing on the detector's status in 2016. The innermost detector part is the tracking system (Sec. 3.2.1), built from silicon sensors. This is followed by an electromagnetic calorimeter (ECAL) and an hadronic calorimeter (HCAL) (Sec. 3.2.2). The ECAL is a homogeneous calorimeter built out of lead-tungsten crystals. In contrast, the HCAL is a sampling calorimeter composed of brass and plastic scintillators. These detectors are located inside the magnet (Sec. 3.2.3), build in a solenoid design, and have a magnetic field of up to 3.8 Tesla. A small fraction of the HCAL is situated outside of the solenoid. The outermost part of the CMS detector is the muon system (Sec.3.2.4). The triggering system is discussed in Sec. 3.2.5. It decides which interactions are stored. Most events are rejected based on variables calculated between collisions, which is necessary due to the large number of interactions. First, the coordinate conventions and the design goals of CMS are presented. The sketch of the CMS detector in Fig. 3.2 shows the detector systems and the dimension of the detector in comparison to a human.

The usual coordinate conventions are used where the origin is the interaction point at the CMS cavern, the  $z$ -axis points in the direction of the beam towards the Jura mountain, the  $x$ -axis points towards the centre of the LHC ring, and the  $y$ -axis points

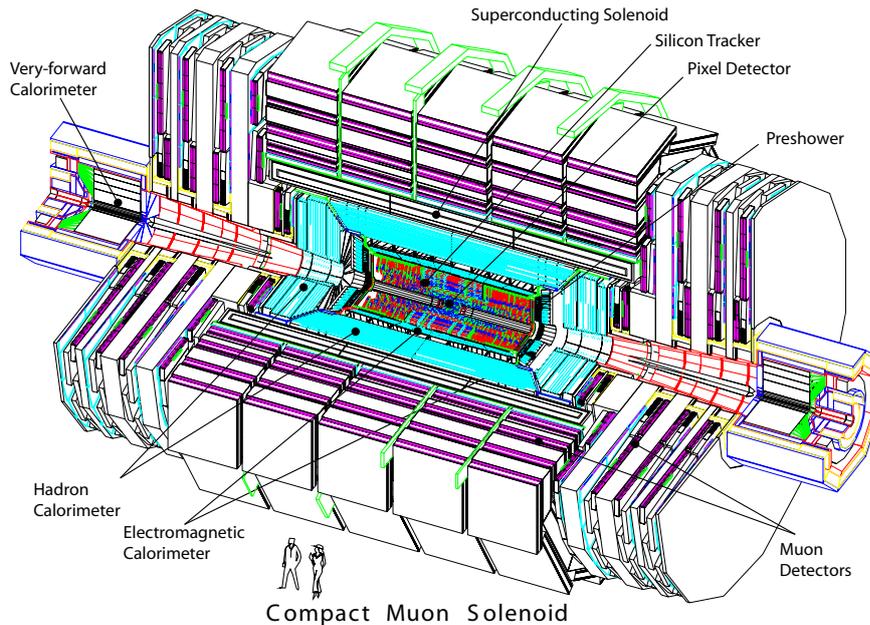


Figure 3.2.: Sketch of the CMS detector where part of the detector is missing for a better view of the parts. Humans are shown for scale [71].

vertically upwards. In most of this thesis, cylindrical coordinates are used where  $\phi$  is the azimuthal and  $\Theta$  is the polar angle. Since  $\Theta$  is not invariant under a Lorentz boost in  $z$ -direction, it is not often used at hadron colliders. Instead of  $\Theta$ , the pseudorapidity  $\eta = -\ln(\Theta/2)$  is employed, for which the difference  $\Delta\eta$  of two particles is invariant for a Lorentz boost in  $z$ -direction. For a distance measure in this context  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$  is mostly used hereafter, where  $\Delta\phi$  is the difference between two particles in  $\phi$ .

The detector design was chosen to target a plethora of different measurements and the reconstruction of physics objects. The most important benchmark scenario was the search for the Higgs boson and the measurement of its properties. The SM Higgs boson has branching fractions into all the elementary particles in the SM, either directly or through quark loops. The detector is designed to precisely measure the different elementary particles interacting with a Higgs boson with a hypothetical Higgs mass in the range of 114 GeV to  $\sim 1$  TeV. Apart from the Higgs boson, searches for supersymmetric models, massive new vector bosons, models with extra dimensions, and SM measurements were also considered during the design phase.

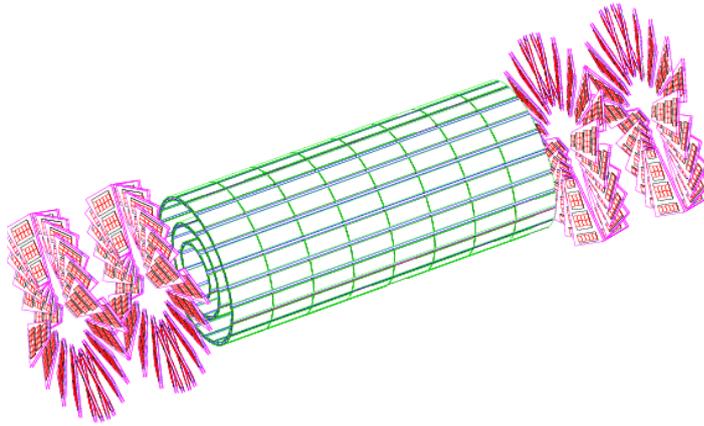


Figure 3.3.: Pixel detector layout in CMS, with three layers in the barrel and two layers in each end-cap [71].

### 3.2.1. Tracker

The main purpose of the tracker is the measurement of the energy and momentum of charged particles, their identification, and the reconstruction of primary vertices. It is the innermost part of the detector with the highest radiation and particle flux during data collection. To have a high resolution for charged particles and vertices the tracker needs to be highly granular with highly sensitive sensors. The tracking system used in CMS is based on silicon detector technology.

The CMS tracker has a cylindrical design with a radius of about 110 cm and a length of 540 cm. The occupancy of active channels per bunch crossing of the detector is designed to be around 1% to guarantee small dead times and stable conditions for data collection. The particle flux and occupancy in the design are limited to reasonable values and lead to a division of the tracker into three regions. The inner region has the highest particle flux and therefore the highest granularity. For the intermediate region of about 20 to 65 cm away from the nominal interaction point, as well as the outer region, the flux is much smaller, and it is possible to use silicon microstrip detectors. Apart from the flux and occupancy, the cost also needed to be feasible. These requirements are met by having the smallest sensors in the inner part and the larger sensors on the outer parts.

In Fig. 3.3 the design is sketched for the innermost tracker part, which is made out of pixel detectors. Three pixel layers are installed in the barrel with two additional layers on each end-cap (corresponding to the lids in the cylindrical design). The barrel layers are located at distances from the nominal interaction point of 4.4, 7.3, and 10.2 cm, while the end-caps are placed at a distance of 34.5 and 46.5 cm with radii of 6 and 15 cm. The design of the end-caps has a turbine-like geometry with blades tilted by  $20^\circ$ . The pixel sensors employed have a size of  $100 \times 150 \mu\text{m}^2$ , of which 768 are arranged into four modules in the barrel. The other 672 are placed in seven different modules in the blades. Both the barrel and end-caps are designed to have a sizeable Lorentz angle and exploit the charge sharing between modules.

The strip sensors are further divided into an inner and an outer barrel part, as well as the end-caps. The inner microstrip barrel is made out of four layers and extends up to  $|z| < 65$  cm. The sensors have a thickness of  $320 \mu\text{m}$  and strip pitches between 80 and  $120 \mu\text{m}$ . The first two layers are made out of stereo modules, which have a combined resolution of about  $23\text{-}34 \mu\text{m}$  in the  $r - \phi$  direction and  $23 \mu\text{m}$  in  $z$ -direction. The outer six layers, which are arranged up to  $|z| = 110$  cm and have modules with a thickness of  $500 \mu\text{m}$  and a strip pitch between 120 and  $180 \mu\text{m}$ . The modules achieve a single point resolution of  $35\text{-}53 \mu\text{m}$  in  $r - \phi$  and  $52 \mu\text{m}$  in  $z$ -direction. Nine disks fill the end-cap region with a distance of 120 to 280 cm to the nominal interaction point. Additionally, three disks are placed between the end-cap and the tracker in the barrel to close the gap. A total of five layers are mounted with stereo modules in the end-caps. The sensor thickness is either  $320 \mu\text{m}$  in the inner parts or  $500 \mu\text{m}$  in the outer parts. Throughout the end-caps, a total of 15400 microstrip modules are used.

The tracker was upgraded during the technical stop 2016/2017. A description of the new tracking device can be found in Ref. [75]. For the high luminosity scenario, which imposes new challenges for track reconstruction, an upgraded tracker is foreseen. Details can be found in Ref. [76]. The foreseen upgrades include an extension of the coverage to 5 in  $|\eta|$ , which would enable new techniques and considerably improve the sensitivity for the production of single VLQs.

### 3.2.2. Calorimeter

The goal of calorimeters is to measure the energy of all particles interacting within the detector volume. Two types of calorimeters are used: ECAL and HCAL. The former is designed to contain and detect electromagnetic and the latter hadronic showers. In CMS an ECAL made out of lead tungsten crystals and an HCAL constructed out of brass plates interleaved with plastic scintillators are used. The ECAL is homogeneous and collects the energy of the interacting particles, while the HCAL as a sampling calorimeter detects only the energy deposits in the sensitive material. The ECAL is presented in the next paragraph before the HCAL is discussed.

To evaluate the performance of a calorimeter, the relative energy resolution

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (3.2)$$

is usually used and parametrised as a function of the particle energy  $E$ . Here  $S$  is the stochastic term,  $N$  is the noise, and  $C$  is the constant term. The resolution is measured in a test beam for several energies, and for each incident energy the reconstructed energy is fitted with a Gaussian.

#### Electromagnetic calorimeter

The particles that mostly decay within the ECAL are electrons and photons. The main showering processes are the pair-production of electrons and positrons, as well as photons emitted through Bremsstrahlung. These particle types leave a very distinct signature in the ECAL, and measurements in this detector are the primary way to identify them. To fulfil this task, the ECAL has to be radiation hard, operable at the 40 MHz collision rate, and contain most parts of the electromagnetic showers. Hadrons passing the ECAL also deposit a considerable energy fraction in it.

The ECAL is a homogeneous and hermetic calorimeter made out of 75848 lead tungsten crystals, that have a radiation length of  $X_0 = 0.89$  cm and a Moliere radius of 2.2 cm. The crystals are radiation hard and can withstand up to 10 Mrad. About 80%

of the light created by a particle passing through is emitted within 25 ns with a low light yield of  $30\gamma/\text{MeV}$ . Photodetectors with intrinsic gain, which are undisturbed by the magnetic field, are needed to detect this light. In the barrel part of the detector ( $|\eta| < 1.48$ ) silicon avalanche photodiodes and in the end-caps ( $1.48 < |\eta| < 3$ ) vacuum phototriodes are used. Stable performance of the avalanche diodes requires the temperature to be stable at the  $0.1^\circ\text{C}$  level.

The barrel part is structured into 36 “supermodules”, each of which cover half the barrel length. The crystals have a small tilt of  $3^\circ$  with respect to the nominal interaction point. The tilt minimises the number of gaps through which radiation can escape detection. The front of the crystal is  $22 \times 22 \text{ mm}^2$  with a length of 230 mm, which corresponds to 25.8 radiation lengths.

In the end-caps  $5 \times 5$  super-crystals form two “D” structures, supported by aluminium structures. The front is  $28.6 \times 28.6 \text{ mm}^2$  with a length of 220 mm corresponding to a radiation length of 24.7.

In terms of energy resolution the performance is good, especially because of the homogeneous and hermetic setup. The stochastic term, for example, is about a factor of four smaller than in the ATLAS liquid argon sampling calorimeter [77]. The higher stochastic term in the ATLAS design is compensated in the particle reconstruction by the higher granularity of the ECAL. The values for the resolution with the CMS detector are  $S = 2.83 \pm 0.01\%$ ,  $N = 124 \text{ MeV}$ , and  $C = 0.26 \pm 0.04\%$ . Overall, the usage of lead tungsten crystals allows the design to be compact, to have a good energy resolution, to have a high radiation tolerance, and to fit within the solenoid magnet. Plans for future upgrades usually foresee the usage of a sampling calorimeter with much higher granularity from which the particle flow (PF) algorithm (discussed in Sec. 4.1) benefits further such that the overall performance is increased.

### **Hadronic calorimeter**

In contrast to the ECAL, the HCAL is designed to measure the energy of hadrons. These particles usually leave only a small fraction of their energy in the ECAL. Hadrons are produced in great abundance at the LHC, and their showering is consid-

erably different from that of electrons or photons. Hadronic showers usually develop later as electromagnetic showers with a considerable fraction of the energy deposited by pions. Sampling calorimeters are usually used in high energy experiments to contain the hadronic showers.

The HCAL is a sampling calorimeter made out of brass interleaved with plastic scintillator material from which the light yield is collected by wavelength-shifting fibres. The light fibres are read out by multi-channel hybrid photodiodes. The design of the HCAL is driven by the goal to maximise the interaction length inside the magnet, providing a high hermeticity and containment of the hadronic activity. The hermeticity is important to reduce the non-Gaussian tails of the  $\vec{p}_T^{\text{miss}}$  measurements. The HCAL consists of four parts: the barrel ( $|\eta| < 1.4$ ), end-cap ( $1.3 < |\eta| < 3$ ), forward ( $3 < |\eta| < 5$ ), and outer ( $|\eta| < 1.26$ ) calorimeters. The outer calorimeter is the only one outside of the solenoid magnet.

The barrel is made of 32 identical wedges, and the outermost plates are made of steel for mechanical strength. The segmentation in  $(\eta, \phi)$  of the scintillator is  $(0.087, 0.087)$  with a total of 16  $\eta$  sectors. Particles arriving at the HCAL first encounter a 9 mm thickness scintillator plate. Afterwards, plates made of stainless steel are attached followed by 15 brass layers and one additional steel plate layer. Between the several steel and brass layers, scintillator plates of 3.7 mm thickness are mounted. Further information on the HCAL in the barrel is found in Ref. [78].

For the end-caps, a total of 2304 towers are used. The towers are segmented in  $\phi$  of  $5^\circ$  in the innermost towers, and  $\phi$  of  $10^\circ$  in the outer part. The segmentation in  $\eta$  varies between 0.087 and 0.35.

The coverage for the highest  $\eta$  is achieved by the hadronic forward calorimeter, which is constructed from steel and quartz fibres, with a total of 1800 channels in two modules. The total depth of the absorber is 1.65 m. The steel plates have 1 mm square grooves filled with quartz fibres with a diameter of 0.6 mm. These are placed 5 mm apart in a square grid. The Cerenkov light produced in the fibres through the passing particles, is guided to photomultipliers. The segmentation in  $\phi$  is  $10^\circ$ , apart from the outermost part, where it is  $20^\circ$ . In  $\eta$  the segmentation varies between 0.1 and 0.3.

The magnet constrains the volume of the HCAL. Additional scintillator plates with a thickness of 10 mm, are mounted onto the exterior of the magnet to enhance the amount of collected hadronic energy. Part of the radiation tail of high momentum particles can be measured with the exterior plates. They are located in the muon system and match the  $\phi$  segmentation of the drift tubes with a geometry adapted to the muon system.

For pions the energy response  $S = 84.7 \pm 1.6\%$  and  $C = 7.4 \pm 0.8\%$  are measured in a test beam. The resolution is much lower than for any ECAL but contains most of the hadronic energy fraction of an event. Upgrades foresee a higher granularity for the HCAL, as with the ECAL, to maximise the performance of the PF algorithm.

### 3.2.3. Magnet

The magnet is essential for the identification and measurement of particles. Through the Lorentz force the trajectory of electrically charged particles is bent from which the momentum is reconstructed. Particles that do not bend are identified as electrically neutral. The radius of the trajectory depends on the magnetic field, the momentum of the particle, and the electric charge. With a precisely known magnetic field and electric charge, the momentum is reconstructed in the tracker with high precision. The magnet of CMS has a magnetic field of 3.8 T with a length of 12.9 m and an inner bore of 5.9 m in a solenoid design. It is the largest magnet with such a high magnetic field in any high energy particle detector. The superconducting niobium-titanium magnet features a high purity aluminium-stabilised conductor with indirect cooling and epoxy impregnation over the full structure. Due to this design the magnet can withstand an outward pressure of 64 atm, and during operations it stores roughly 18kA. Apart from the coils a “return yoke” that reaches out 14 m in diameter is also part of the structure. It guides the magnetic field, provides a filter for most particles, and is in large part responsible for the structural support of the detector systems.

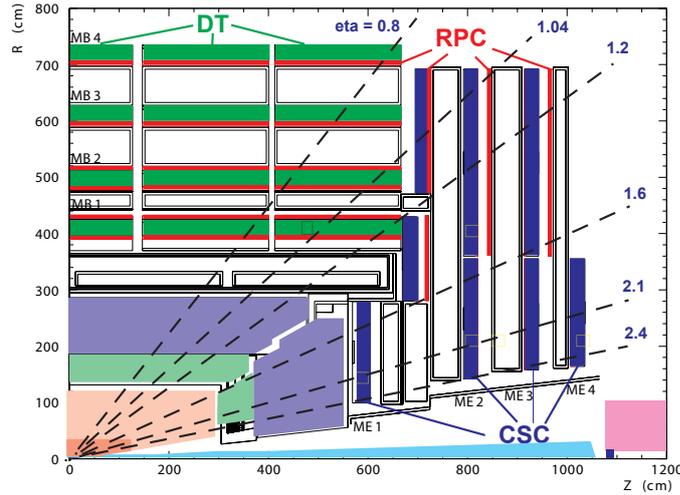


Figure 3.4.: Pictorial representation of one-quarter of the muon system in CMS [71].

### 3.2.4. Muon system

Compared to other particles from the SM, muons have a long lifetime and traverse the CMS detector depositing only a small fraction of their energy. They are often referred to as minimal ionising particles (MIP) and deposit about  $1\text{-}2 \text{ MeV}/g \text{ cm}^{-2}$  in the detector. Detectors are mounted outside the magnet to identify muons where no signal of other SM particles from the interactions is expected. The measurements from these detectors are combined with the measurements of the tracker to obtain a precise muon momentum measurement. Three types of detectors filled with gas are used to identify and measure the muon momentum outside the magnetic coil. Because of the very distinct signature, the identification of muons is usually the most accurate of all the SM particles.

In the design of the muon system the expected radiation, the amount of surface to cover, the muon rate, as well as the time, momentum, and position resolution are taken into account. Drift tubes are employed for the central part of the detector ( $|\eta| < 1.2$ ). Cathode strip chambers are used in the end-caps ( $|\eta| < 2.4$ ) which can handle a higher muon rate, radiation, and magnetic field. Additionally, resistive plate chambers are used in both regions. These have a better time resolution but coarser position resolution compared to the other two detector types. The high time resolution improves the identification of the correct bunch crossing.

A sketch of a quarter of the muon system in CMS is shown in Fig. 3.4. It provides about 25 000 m<sup>2</sup> of active detection plane with nearly 1 million channels. Compared to electrons and hadrons, muons often have lower trigger thresholds since the reconstruction in the muon system is less time-consuming. The fraction of misreconstructed particles is low (at the 1% level).

### 3.2.5. Trigger

From the bunch crossing rate of 40 MHz only about 50 Hz can be stored to disk due to the readout rate and disk capacities. The system deciding which events are of interest and are stored is referred to as the “trigger” [79]. It is designed to achieve a rejection rate of about 10<sup>6</sup>. The rejection is done in CMS in a two-step approach. The reconstruction of the event in the trigger system during two collisions is referred to as “online” and “level 1 (L1)”. Events that pass the online stage are processed further in the high level trigger (HLT) using more sophisticated and CPU consuming algorithms. Custom hardware processors are operated at level 1 to decide which events are interesting, which are then passed on to the HLT. The HLT is run on a computing farm and partly reconstructs the objects of interest. If an event also passes these steps all the objects in the event are reconstructed in the so-called “offline” reconstruction. Events that pass the offline reconstruction still have to fulfil further offline quality criteria.

The “level 1” custom hardware consists of a range of integrated circuits and programmable gate arrays that are used to calculate “trigger primitives”. These are physics objects like muons, electrons, jets, and photons reconstructed with a reduced resolution in a small part of the detector. Thresholds on these objects are applied, and these selections reduce the rate to about 100 kHz. During the first reconstructions and decisions, the data are held in memory pipelines.

The data from the memory pipelines are read out and passed to a computing farm. During the readout, the detector data are transferred to several hundred front-end buffers and combined to an event with an average size of about 1.5 MB, for a *pp* collision. During this process the data are compressed and additional signal processing is done. On the computer farm the physics objects are further reconstructed following

---

the guideline of reconstructing only those objects and regions that are necessary for a trigger decision. Events are partially reconstructed and rejected if they do not fulfil the trigger requirements. The rate is further reduced to 400 Hz through the HLT.

For events passing the HLT, all physics objects are reconstructed with the highest possible precision and resolution. These objects are later further calibrated and corrected. While the “online” decisions are irrevocable, the “offline” reconstruction is often improved over time, and data events are reprocessed to use the latest reconstruction algorithms and calibrations.



## 4. Event reconstruction and particle identification

The recorded collisions of the CMS detector are stored as collections of digitised sub-detector measurements for each event. These measurements are combined to reconstruct physics objects used for analysis and are performed for data and simulated events. An event usually refers to the stored collection of digitised measurements recorded during one bunch crossing. The first step in the reconstruction of an event is to combine the information of the several sub-detectors to form particle candidates with the so-called particle flow (PF) algorithm [80]. These candidates are used to identify muons, electrons, photons, and hadrons. The transverse momentum of particles not interacting with the detector material is calculated from the negative vectorial sum of all reconstructed PF candidates. Based on these reconstructed quantities, the analysis described in this thesis is performed; this includes the selection, reconstruction of the VLQ, and the statistical interpretation of the results.

This chapter is organised as follows: first, the PF algorithm is presented in Sec. 4.1. In Sec. 4.2, the criteria for identifying electrons and muons and their calibration are described. In Sec. 4.3 the jet reconstruction and calibration are discussed. The methods for tagging quarks and bosons within hadronic jets are explained there as well.

### 4.1. Particle flow algorithm

Traditionally, the distinct signatures of SM particles in the sub-detectors are used to reconstruct and identify them; for example, electrons are primarily reconstructed from

energy clusters of neighbouring cells in the ECAL. In contrast to an approach focusing on the reconstruction in one sub-detector, the holistic PF algorithm is used in CMS. This algorithm was first developed at the Large Electron Positron Collider (LEP) for the Apparatus for LEP Physics (ALEPH) experiment. The PF algorithm reconstructs particles in the different parts of the detector and matches them geometrically (linking) to PF candidates, which are used to identify leptons and hadrons. The combination of the available detector signal produced from different sub-detectors significantly improves the measurement of the energy and direction when the various sub-detectors gather complementary information. The PF algorithm also provides a good way of cross-calibrating sub-detectors. Because of the significant improvements in particle reconstruction, many future detectors are designed fully to exploit the PF algorithm [81–83]. The description of the PF algorithm in CMS follows the detailed article in Ref. [84].

### 4.1.1. Requirements

Various features are crucial in an experiment to achieve good performance of the PF algorithm. These features are based on the properties of the SM particles and are as follows:

- A strong magnetic field that allows the separation of neutral and charged particles in the tracker and calorimeters.
- A very finely grained tracker that provides reliable energy measurements of particles with low transverse momenta, which represent the majority of jet constituents.
- A highly segmented ECAL for a detailed reconstruction of the particle shower profile.
- A hermetic HCAL able to reconstruct most neutral and charged hadrons produced in the event.
- An efficient muon system.

The CMS detector was not originally designed for the deployment of the PF algorithm but was found to fulfil the aforementioned requirements to a high degree. The measurement of single particles with the PF algorithm especially benefits from the high magnetic field of 3.8 T, the high granularity of the tracker, and the efficient muon system. Especially for highly energetic particles the segmentation of the HCAL is not fine enough to avoid the merging of several particles in one cell. The HCAL was designed to fit within the magnet, which also delimits the hermeticity but is recovered partly by the HCAL on the outside of the magnet. A higher granularity in both calorimeters could improve the performance of the detector. Especially at low momentum, the combination of the information from several detectors allows for a highly accurate reconstruction of the momentum and direction of particles. Because of its performance, the PF algorithm is used for most publications in CMS since its commissioning in 2010.

### 4.1.2. Charged tracks

In order to reconstruct the trajectory of the charged particles the hits in several layers of the tracker are combined to reconstruct the trajectory of particles. This reconstruction is challenging due to a large number of hits and possible combinations. An iterative approach is used to achieve a high reconstruction efficiency and low misreconstruction rate. Because of their characteristics and their importance in physics analysis, muons, and electrons undergo a specialised reconstruction algorithm.

The combinatorial iterative track finder is based on Kalman filters with successive iterations targeting tracks coming from different decays. The seeds for the filters begin with triplets in the pixel detector to determine which seeds from the microstrip detector are added at later stages. Several quality criteria are applied to tracks and seeds (e.g., to the track fit  $\chi^2$  or to the compatibility to originate from a primary vertex). Thanks to the seeding algorithm, a much better reconstruction efficiency and a comparable misreconstruction rate are obtained in comparison to a one-pass algorithm (“single iteration”). Apart from higher efficiency, the iterative algorithm is twice as fast as the “single iteration” algorithm, which is essential since tracking is the most computationally demanding task of the reconstruction. Nuclear interactions,

which may occur in the tracker material, induce kinks in the trajectory of the hadrons, or produce secondary particles and vertices, are taken into account.

### **Muon tracks**

The muon tracking step is not inherent to the PF algorithm. The muon detectors enable a high identification efficiency (99%) over the full acceptance of the detector and high purity since most SM particles are absorbed in the calorimeters or magnet. The tracks in the muon detectors are built by default in the last iteration steps of the aforementioned track finder. For the analysis only muons are considered where tracks in the tracker and muon detector are matched to form global muon tracks. The resolution for large transverse momenta for global muons is superior in comparison to muons formed only with tracker or muon detector information. One of the few sources for misreconstructed muons are highly energetic hadrons which are not absorbed in the calorimeters or magnet and deposit energy in the muon detector; this also known as “punch through”.

### **Electron tracks**

Electrons deposit most of their energy in the ECAL; nonetheless, the measurements in the tracker improve their identification and position resolution. Especially for highly energetic electrons passing through the detector, the performance is deteriorated when only using ECAL cluster due to photons emitted by Bremsstrahlung in the tracker material. Additionally, the large amount of energy deposited in the ECAL by hadrons produced in the same event might overlay with the signature of electrons and make it challenging to identify them. A seeding and tracking algorithm is used based on the ECAL as well as the tracker system. For this, if track candidates exceed a  $p_T$  of 2 GeV the track finder is redone for track candidates with a simple and faster Gaussian-sum filter (GSF) algorithm [85]. The GSF algorithm is better suited for electrons since it allows for energy losses across the particle trajectory. After the reconstruction of the preselected tracks a boosted decision tree (BDT) is used to discriminate between electrons and misreconstructed hadrons. Only the electron candidates passing the BDT are reconstructed using the full GSF tracking. This procedure almost doubles

the efficiency of reconstructing an electron, keeping a reasonable computing time per event.

### 4.1.3. Clustering

The energy of particles in the calorimeter can be distributed over several calorimeter cells, and the algorithm used to associate the energy deposits from the same source with a single particle is called clustering. The energy clustering in the ECAL and HCAL is mainly done to measure neutral and charged hadrons, as well as electrons and photons. The algorithm is also exploited to separate and identify the different particle types.

The clustering algorithm was specially designed for the reconstruction with the PF algorithm. The sub-detectors are clustered separately, apart from the hadron forward calorimeter (HF) where each cluster seed gives rise to a cluster. First cluster seeds are identified that are calorimeter cells with an energy deposit above a threshold, and with an energy deposit higher than their neighbouring cells. Neighbouring cells share a side (HCAL) or at least a corner (ECAL and preshower) with the seed candidate. From these seeds topological clusters are grown. The cells are calibrated using radioactive sources, test beam, and data from cosmic muons traversing the detector taken before the operation of the LHC. This calibration is continually optimised with collision data to mitigate radiation damage accumulated over time. For the optimisation, the threshold levels are adjusted, and the energy from the clusters is corrected with  $\eta$  and  $p_T$  dependent factors. The calibration is essential since the calorimeters represent the only measurement of the energy of neutral particles and overlapping deposits with charged particles are only measured as an energy surplus in these clusters.

### 4.1.4. Linking

In general, reconstructed particles are expected to deposit their energy in several parts of the detector. Therefore, the PF algorithm proceeds by means of a linking algorithm that connects the different particle trajectories reconstructed in the sub-detectors. The

probability of finding the correct link between all detector parts is mostly limited by the granularity of the sub-detectors, and the material in front of the detectors, which may lead to secondary particles or kinks in the trajectory.

The linking is done in the  $\eta - \phi$  plane and is limited to the nearest neighbours to achieve a reasonable computing time for a single event. The produced chains of PF blocks are associated through a link or common linked elements. The quality of a link is assessed by the distance of the connected elements. First, the track from the innermost detector is extrapolated to the calorimeters. Afterwards, connections to clusters are constructed. The last step is to connect the tracker and muon detector elements. From these PF blocks first the muons are removed followed by the electrons and photons. The remaining PF candidates are treated as hadronic particles.

## 4.2. Charged leptons

From the three charged leptons, muons and electrons are considered in the presented analysis. The muon and electron candidates reconstructed with the PF algorithm have to pass the trigger and several quality criteria. These quality criteria improve the misidentification rate.

### 4.2.1. Muon identification

From the reconstructed PF muon candidates, only those with matched energy deposits are used in the tracker and muon system, which guarantees a low misidentification rate. Additionally, a muon candidate has to fulfil the following requirements:

- fraction of valid tracker hits greater than 0.49,
- normalised global track  $\chi^2 < 10$ ,
- at least two matched stations,

- transverse impact parameter  $d_{xy} < 0.2$  mm with respect to the primary vertex,
- at least one hit in the pixel detector
- at least five tracker layers with hits.

The global track  $\chi^2$  is calculated from the fit done in the muon track reconstruction. Particle flow candidates that pass these muon quality criteria have to fulfil a minimum  $p_T$  of 40 GeV and  $|\eta| < 2.4$ . The efficiency is at the 95% level with a mistag rate at the 0.01% level [86].

### 4.2.2. Electron identification

A multivariate algorithm is applied to the reconstructed PF electron candidates to identify them. Three types of variables are used to discriminate electrons from other background contributions such as multijet events. These variables are based on

- matching of tracker and ECAL information (both energy and geometrical information),
- shape of the energy deposits in the ECAL, HCAL and pre-shower detectors,
- differences between the Kalman filters and the GSF algorithm.

These types of variables are fed into a BDT. Based on the granularity of the clusters, three regions with different thresholds on the BDT are identified. For the two central regions of the detector with  $|\eta| < 0.8$  and  $0.8 < |\eta| < 1.44$ , the classifier has to be larger than 0.941 and 0.899, respectively. In the forward region, the classifier value has to be larger than 0.758. Identified electrons are only considered if they fulfil  $|\eta| < 2.4$  and  $p_T \leq 40$  GeV. The classifier values used correspond to a signal efficiency of 80% and a misidentification rate of less than 0.1% [87].

### 4.2.3. Lepton isolation

Processes with multiple hadronic jets dominate the SM background at the LHC. The VLQs are searched in decay channels with a prompt lepton in the final state. Prompt leptons are produced by the electroweak force during the decay of a quark or gluon before it can hadronise. For this analysis to be feasible the QCD background must be reduced by several orders of magnitude. Two out of many existing approaches are presented. The first relies on the lepton being away from hadronic activity in the  $\eta - \phi$  plane. The second algorithm is based on the lepton having a high  $p_T$  in comparison to the hadronic activity around it. The first algorithm is often used for SM measurements, while the second algorithm is often used to search for heavy resonances with jets close to the leptons [8, 88, 89].

#### Relative lepton isolation

The classical isolation [90] is a simple algorithm used to identify high-purity leptons from prompt decays. For this purpose the scalar sum of the transverse momentum of tracks within a cone of  $\Delta R = 0.3$  around the lepton axis is calculated, excluding the lepton candidate. This sum is divided by the transverse momentum of the lepton. Usually, leptons with isolation values smaller than 0.1 or 0.2 are selected. In the presence of additional radiation from hadronic activity in the event that is near the lepton, the efficiency of this simple algorithm drops considerably. Due to its simplicity, the approach is used at the trigger stage to reduce the threshold on the transverse momentum of the lepton. The amount of non-prompt leptons arising in multijet SM processes is typically reduced to a few per cent of the SM background prediction.

#### Boosted lepton isolation

The boosted lepton isolation is designed to achieve a higher efficiency than the relative isolation, for final states where the lepton has large transverse momentum and additional hadronic activity is present near the lepton. This is done by considering the relative momentum  $p_{T,\text{rel}} = \frac{|\vec{p}_{\text{jet}} \times \vec{p}_{\text{lep}}|}{|\vec{p}_{\text{jet}}|}$  between the lepton and the anti- $k_T$  jet with

$R = 0.4$  (ak4 jet) (as presented in the next section) closest in  $\Delta R$ , as well as their angular distance  $\Delta R(l, \text{jet})$ . The boosted lepton isolation selection requires  $\Delta R(l, \text{jet}) > 0.4$  or  $p_{T, \text{rel}} > 25$  GeV. Only ak4 jets with a  $p_T > 30$  GeV are considered. The lepton four-momentum is subtracted from the ak4 jet if it is one of the PF candidates clustered into the jet.

## 4.3. Hadronic jets

Jets aim at reconstructing the decay products of high-energy quarks and gluons. Because of the asymptotic freedom, these particles behave quasi-free for high energies  $\mathcal{O}(100 \text{ MeV})$ . In scattering experiments, they produce cascades of gluons and quark-antiquark pairs because of the hadronisation, with bremsstrahlung being an additional important factor. The reconstruction, simulation, and measurement of quarks and gluons is therefore challenging, from both computational and physical standpoints. Several algorithms have been developed to cluster all the decay products of a parton. These can be grouped into cone algorithms and sequential clustering algorithms. The latter is used at the LHC and discussed in Sec. 4.3.1. The corrections applied on the jet constituents before and after the clustering are discussed in Sec. 4.3.2. The calibration steps of the jets are outlined in Sec. 4.3.3. Detailed reviews on jets and related measurements can be found in Ref. [91, 92]. The development of algorithms to reconstruct jets and identify and tag quarks or bosons decaying into hadrons has extended the understanding of the QCD and SM; e.g., through the measurement of the jet substructure and the top quark mass [93, 94]. Details on the jet reconstruction, correction, and calibration in CMS, including the definition of the dedicated regions for their measurements are found in Ref. [95].

### 4.3.1. Jet clustering

At the LHC sequential jet clustering algorithms are mostly used to reconstruct the hadronic activity in the detector. The jet algorithms have to be infrared and collinear safe [96] to associate them with well-defined physics objects. This means soft radiation

or a collinear splitting of quarks or gluons does not affect the jet reconstruction. From the class of algorithms that fulfil these requirements, “generalised  $k_T$ ” algorithms are used at CMS. For this analysis, the PF particle candidates are used as input to the jet clustering. The clustering is done in three steps:

1. The distance between each pair of particles  $d_{i,j}$  is calculated, as well as the distance to the beam  $d_{iB}$ :

$$d_{ij} = \min(p_{T,i}^{2m}, p_{T,j}^{2m}) \frac{\Delta R_{i,j}^2}{R^2} \quad (4.1)$$

$$d_{iB} = p_{T,i}^{2p} \quad (4.2)$$

Here,  $R$  is a distance parameter,  $m$  is an integer number,  $\Delta R$  is the distance measure in the  $\phi-\eta$  plane used at the LHC, and  $p_{T,i}$  is the transverse momentum of the  $i$ th particle.

2. The minimum of all  $d_{iB}$  and  $d_{ij}$  is found. If the minimum corresponds to a  $d_{ij}$ , particles  $i$  and  $j$  are merged by summing their four-momenta. Otherwise, particle  $i$  is declared a jet and removed from the particle list.
3. The two previous steps are repeated until no particles are left.

The parameter  $m$  is usually set to -1, which corresponds to the anti- $k_T$  algorithm, or to 0, for the so-called Cambridge/Aachen algorithm. The Cambridge/Aachen algorithm first clusters particles that are closest in the  $\Delta R$  metric, whereas the anti- $k_T$  algorithm begins the clustering from particles with the highest  $p_T$ . All these algorithms are implemented in the FASTJET package [97, 98], which is used to find jets. In this analysis, two types of jets are used: anti- $k_T$  jets with  $R = 0.4$  (ak4 jets) and anti- $k_T$  jets with  $R = 0.8$  (ak8 jets).

### 4.3.2. Pileup corrections

In  $pp$  collisions at a centre-of-mass energy of 13 TeV in 2016 for each bunch crossing, on average about 25 interactions happened. The (primary) interaction with the highest energy is usually the most interesting because it allows the production of particles with masses and momentum beyond the reach of previous experiments. Other interactions in the same bunch crossing are called pileup. The interaction energy is calculated by summing the tracks of the corresponding vertices. The pileup is recorded with the primary interaction and leads to additional detected particles (mostly hadrons). Two types are distinguished, namely: “in-time” and “out-of-time” pileup. The first happens in the same bunch crossing and the second are bunch crossing remnants. Two methods are currently employed to eliminate the contribution of in-time pileup: the charged hadron subtraction (CHS) and the pileup per particle identification (puppi) algorithms. The out-of-time pileup is mostly already removed during the reconstruction. The algorithms are applied before the jet clustering is performed on the PF candidates. They do not remove the whole pileup contribution is removed, so that an additional step is applied after the jet clustering. This algorithm removes the expected average radiation per jet originating from pileup.

#### Charged hadron subtraction

For CHS [99] the tracks associated with secondary vertices are removed from the list of PF candidates. About 50% of the “in time” pileup is removed within the tracker acceptance of  $|\eta| < 2.4$  coming from roughly 30% of secondary vertices. The unreconstructed vertices are responsible for a large set of unassociated tracks. For high pileup scenarios with 80 or 140 pileup vertices per interaction as expected for high luminosity runs, this leads to a deterioration of the jet resolution. In a pileup scenario of approximately 25 secondary vertices, the resolution is at a reasonable level.

## Puppi

In the last years a new idea has emerged aiming at the mitigation of the pileup contribution. Apart from the information from secondary vertices and associated tracks, local shape information can also be used to remove the influence of pileup from jets. The puppi algorithm [100] calculates a weight per PF candidate based on the local information. The PF candidates from pileup usually get lower weights than PF candidates connected to the primary interaction. The weights are used in the jet clustering and reduce the influence of pileup on the final jet. The local shape variable  $\alpha_i$  is defined as

$$\alpha_i = \log \left( \sum_j \frac{p_{Tj}}{\Delta R_{ij}} \times \Theta(R_{min} \leq \Delta R_{ij} \leq R_0) \right), \quad (4.3)$$

where  $p_{Tj}$  is the  $p_T$  of the PF candidate  $j$  and  $\Delta R_{ij}$  is the usual distance metric, between two PF candidates. Here,  $R_{min}$  is effectively a regulator for the collinear splitting of PF candidates and  $R_0$  is the maximal allowed distance between two PF candidates. The notation  $\Theta(R_{min} \leq \Delta R_{ij} \leq R_0)$  is the short form for two Heaviside functions of the form  $\Theta(\Delta R_{ij} - R_{min}) \times \Theta(R_0 - \Delta R_{ij})$ . The logarithm is outside of the sum and rescales the variable.

In the original paper [100] the performance is studied in simulation, and it is demonstrated that for an idealised detector the effects of 80 or 140 pileup interactions per event are largely mitigated. In this analysis, puppi is only applied on the ak8 jets used for the W tag.

## Jet area subtraction

To further mitigate the influence of pileup on reconstructed jets, the expected average contribution is subtracted. The method is also called ‘‘hybrid because’’ it depends on the jet  $\eta$  and the number of vertices in the event apart from the jet area  $A$ . The correction factor

$$c(p_{T,uncorr}, \eta, A_j, \rho) = 1 - \left[ \frac{\rho_0(\eta) + \rho\beta(\eta)(1 + \gamma(\eta)\log(p_{T,uncorr}))}{p_{T,uncorr}} \right] \quad (4.4)$$

is multiplied to the jet momentum  $j$ . Here,  $p_{\text{T,uncorr}}$  is the uncorrected jet momentum and  $\rho$  is the energy density per event. The factors  $\rho_0(\eta)$ ,  $\beta(\eta)$ , and  $\gamma(\eta)$  model the  $\eta$  dependence of the correction. In simulation, they are obtained from the difference of the particle energies by reconstructing events with and without pileup. In the data  $\rho_0(\eta)$  and  $\beta(\eta)$  are estimated from minimum-bias data with a random cone method. There is no hard scattering and small noise contribution in these events, such that the average  $p_{\text{T}}$  in the cone of the PF candidates is assumed to be equal to the offset due to pileup. Both,  $\beta(\eta)$  and  $\gamma(\eta)$  parameterise the minimum-bias overlay simulated with PYTHIA, which matches data well.

### 4.3.3. Jet calibration

The response of the detector is non-linear as a function of particle energy or type. The energy scale and resolution are calibrated for jets and subjets to the true particle level jet  $p_{\text{T}}$  and resolution, which is described in detail in Ref. [95, 99]. Below, the most important calibration steps are outlined after the pileup corrections are applied. The residual corrections are applied on jets in data and are typically small. The residuals are calculated by using well-measured reference objects. Corrections dependent on the jet flavour are not discussed here. The differences in flavour mainly arise due to differences in the fragmentation, such that for different particles the energy within the jet varies.

Like leptons, jets also have to pass quality criteria, which has a small impact on this analysis. These criteria ensure that jets from detector noise and non-beam backgrounds are rejected. Jets have to be clustered from neutral and charged particles, as well as have more than one constituent. For the central part of the detector ( $|\eta| < 2.7$ ) the fraction of charged particles must also be smaller than 100%.

#### Jet energy scale

The jets are calibrated to obtain the correct energy, mitigating the shifts introduced into the energy scale by detector effects. The jet energy scale corrections (JEC) are

calculated from simulation in a multijet sample simulated with a flat  $p_T$  spectrum. With this approach, the full kinematic phase space available at the LHC is covered. The pileup profile of the simulation is reweighted to match the data.

The particle jets are matched to the jets in the event, if their separation in  $\phi - \eta$  is less than half the distance parameter with the usual distance metric  $\Delta R(\eta, \phi)$ . This provides a unique match per jet. The particle response  $r$  is defined as

$$r(\langle p_T \rangle, \eta) = \frac{\langle p_T \rangle}{\langle p_{T,\text{ptcl}} \rangle} [p_{T,\text{ptcl}}, \eta], \quad (4.5)$$

and depends on the average transverse jet momentum and  $\eta$  with  $p_{T,\text{ptcl}}$  being the particle level jet transverse momentum. The angle brackets  $\langle \rangle$  indicate the average of the variable, while variables with square brackets  $[]$  are binned. The response corrections are applied to both data and simulated jets.

### Jet energy resolution

The correction of the jet energy resolution (JER) is necessary because simulated events show a better jet resolution than jets in data. The idealised performance of the detector in simulation is compensated for by the JER smearing. A stochastic smearing is applied to the jet resolution with the correction factors given by

$$c = 1 + \mathcal{N}(0, \sigma_{\text{JER}}) \sqrt{\max(s_{\text{JER}}^2 - 1, 0)}. \quad (4.6)$$

A random number  $\mathcal{N}(0, \sigma)$  following a normal distribution with variance  $\sigma$  and centred around 0 is drawn. Here,  $s_{\text{JER}}^2$  and  $\sigma_{\text{JER}}$  encode the relative  $p_T$  resolution of the jet. After applying the smearing, the resolution of data and simulation agree.

### Residual jet corrections

After the correction of the jets for pileup and the simulated particle response, additional residual corrections are applied to the jet. The basic idea is to calibrate the energy of the jet to a reference object, to achieve a jet energy scale at unity. The

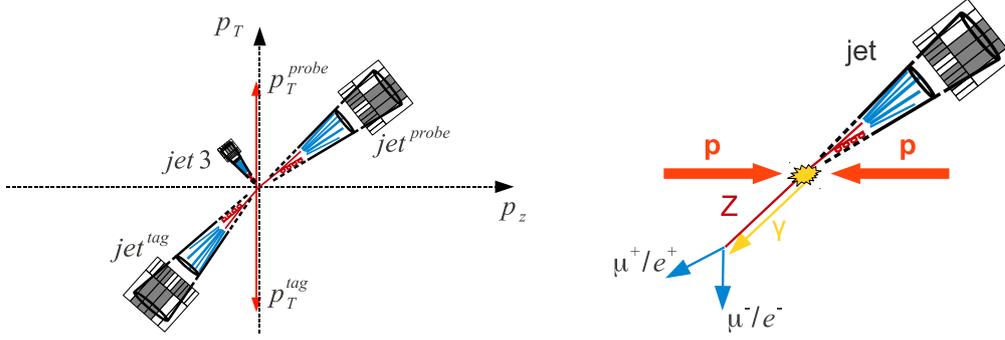


Figure 4.1.: Sketch of a multijet (left) and Z+jet (right) event, from Ref. [101].

reference objects are determined from measurements of  $Z(\rightarrow \mu\mu)+\text{jets}$ ,  $Z(\rightarrow ee)+\text{jets}$ ,  $\gamma+\text{jets}$  and multijet events. Figure 4.1 provides sketches of a multijet (left) and a Z+jet (right) event used for the corrections.

The corrections of the response are calculated using two methods: the  $p_T$  balance  $r_{\text{jet},p_T}$  and the missing transverse momentum projection fraction  $r_{\text{jet,MPF}}$ , defined as

$$r_{\text{jet},p_T} = \frac{p_{T,\text{jet}}}{p_{T,\text{ref}}}, \quad (4.7)$$

$$r_{\text{jet,MPF}} = 1 + \frac{\vec{p}_T^{\text{miss}} \cdot \vec{p}_{T,\text{ref}}}{(p_{T,\text{ref}})^2}, \quad (4.8)$$

where the  $\vec{p}_T^{\text{miss}}$  is explained in Sec.4.4. In both methods a tag and probe ansatz is made where the  $p_T$  of a reference (ref) object ( $p_{T,\text{ref}}$ ) is compared to the probe jet momentum  $p_{T,\text{jet}}$ . For the  $p_T$  balance, ideally dijet events are chosen from the multijet sample for the measurements. The statistics are too small to use only dijet events without additional activity. Events with more than two jets are therefore used to increase the statistics of the measurement and the results are extrapolated to the dijet event topology.

With these methods, relative and absolute corrections to the jet energy are derived. The relative corrections mitigate the remaining differences in the JER and the absolute corrections in the jet energy scale. Dijet events are used to derive relative corrections in  $\eta$  and  $p_T$ , as a function of the average of the two jets. The corrections also mitigate the effects of initial and final state radiation, to leading order. The  $p_T$  dependence is found to be rather small and of the form  $p_0 + p_1 \log(p_T)$ . Absolute corrections are

calculated from  $Z/\gamma$ +jets events, where the  $Z/\gamma$  is the reference object. As in dijet events, the initial and final state radiation is taken into account. A global fit of the corrections obtained with multijet and  $Z/\gamma$ +jets events is performed for the absolute corrections, which reduces the overall uncertainty.

#### 4.3.4. Jet tagging

Jet algorithms have been developed for jets originating from gluons or light quarks such as up and down quarks. For jets containing top quarks or W bosons, additional information from the (sub)structure of the jet is employed to differentiate between different quarks and bosons. In the last years, various innovative approaches have been developed in this research field; some examples can be found in Refs. [102–108]. The algorithms used in this analysis to identify bottom (b) and top quarks, as well as W bosons fully decaying into hadrons, are described below. Substructure information is employed to classify jets or events and to improve the reconstruction of the invariant VLQ mass.

The identification of b quarks (b tag) has been used for a long time. The tagging of jets originating from top quarks (t tag) and W bosons (W tag) have become relevant at LHC energies. Tagging of quarks or bosons usually relies on identifying a jet that contains all decay products. Bottom quarks are identified inside of ak4 jets or in the subjets of ak8 jets as part of the t tag algorithm. For the top quarks and W bosons identification ak8 jet are used.

#### Soft drop

The decay products of highly energetic quarks and bosons are often clustered into the same jet. The merging steps of the clustering are undone to identify the jet substructure. The resulting jets are referred to as subjets. For ak8 jets, the subjets are calculated using the soft drop [109] algorithm. These subjets are used to identify W bosons and t quarks.

First, the PF constituents of the ak8 jets are reclustered with the Cambridge/Aachen algorithm, to obtain an angular ordered pairwise clustering history. Afterwards, the following steps are applied, which depend on a soft threshold  $z_{cut}$  and an angular exponent  $\beta$ :

1. Undo the last clustering step of jet  $j$  and label the two resulting subjets  $j_1$  and  $j_2$ .

2. If  $j$  passes the soft drop criterion

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut} \left( \frac{\Delta R_{1,2}}{R_0} \right)^\beta, \quad (4.9)$$

it is the final subjet, where  $p_{T,1}(p_{T,2})$  is the transverse momentum of jet  $j_1(j_2)$ . In this equation  $R_0$  is the jet radius and  $R_{1,2}$  is the usual distance metric between  $j_1$  and  $j_2$ .

3. Otherwise,  $j$  is redefined to be the subjet with larger  $p_T$  and the procedure is iterated.

4. If  $j$  is a “singleton” and cannot be further unclustered, it is either removed (“tagging mode”) or it is the final subjet (“grooming mode”).

In the search presented the algorithm is operated in tagging mode and applied to ak8 jets with  $z_{cut} = 0.1$  and  $\beta = 0$ . The choice of  $\beta$  ensures that the algorithm is infrared and collinear safe. In each jet the unclustering is done maximally two times such that at most two subjets are found.

### N-subjettiness

The N-subjettiness ( $\tau_n$ ) [110] makes use of the difference in the energy flow in jets to differentiate between different decays by constructing  $n$  energy flow axes. The axes are calculated for the  $k$  PF candidates according to

$$\tau_n = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \dots, \Delta R_{N,k}), \quad (4.10)$$

$$d_0 = \sum_k p_{T,k} R_0, \quad (4.11)$$

were  $p_{T,k}$  is the transverse momentum of the particles found in a jet. As in Eq. 4.9,  $R_0$  is the radius parameter used to cluster the original jet and  $\Delta R$  is the usual distance metric. The metric  $\tau_n$  provides a measure of how likely a jet is to have  $n$  prongs. Usually ratios of  $\tau_n/\tau_{n-1}$  are used to discriminate  $n$  prong decays from multijet processes. This approach is usually orthogonal to other information like the reconstructed mass and provides additional discrimination power.

### b tagging

The B meson has a higher lifetime than other quarks and apart from the top quark the highest invariant mass. The long lifetime leads to a secondary vertex displaced from the primary interaction vertex in the order of a few millimetres. About 20% of the b hadrons decay into muons or electrons. These properties of b hadrons are exploited for the b tagging inside jets. In this analysis, the algorithm is applied to ak4 jets and subjets from ak8 jets clustered with the soft drop algorithm.

Algorithms for the identification of B mesons inside of jets (b tag) [111, 112] have been used in CMS for a long time. The employed combined secondary vertex algorithm (CSVv2) is the current version, which combines the information of secondary vertices with the information of displaced tracks through a multilayer perceptron with one hidden layer. The multivariate analysis (MVA) is trained with PF candidates found inside the identified b jets. Per jet, at least two PF candidates have to fulfil quality criteria, based on the reconstruction, the track impact parameter, and the angular

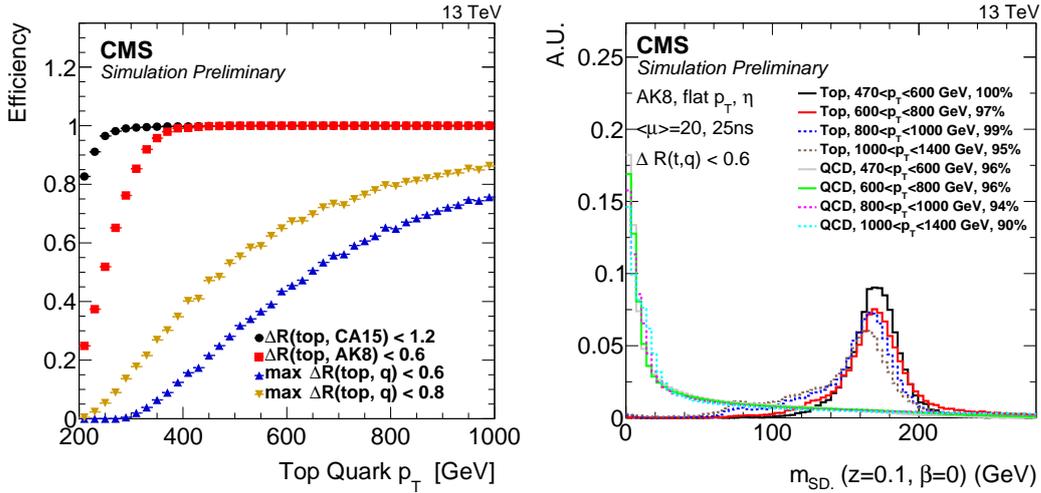


Figure 4.2.: Reconstruction efficiency (left) for a top in a jet as a function of the generated top quark  $p_T$ . The efficiency is defined as the fraction of top quarks for which a reconstructed jet with  $p_T > 200$  GeV can be found within  $\Delta R < 0.6$  ( $\Delta R < 1.2$ ) for ak8 jets (Cambridge-Aachen with a distance parameter 1.5). The fraction of merged top quarks is superimposed as a function of  $p_T$  for the two thresholds used: 0.8 (0.6) at low (high) boost. All distributions are made using hadronically decaying top quarks with  $p_T > 200$  GeV. The soft drop mass distributions (right) are shown for a fiducial selection with the merged-top requirement. The description and distributions are taken from Ref. [106].

distance to the jet axis. Combinations of tracks resulting in an invariant mass close to the one of the  $K_s^0$  are rejected. For jets from these PF candidates, the primary and secondary vertex information is used in several categories and an MVA algorithm is trained with a multitude of variables such as the number of secondary vertices, impact parameter significance, and the vertex mass. For heavy and light flavour jets the efficiency and misidentification rates are calculated separately.

### t tagging

The top quark is the heaviest particle in the SM and is produced at high rates at LHC energies. It decays before hadronising into a W boson and b quark (in more than 99% of the decays). In the fully hadronic decay mode it can have up to three quark (sub-)jets in the final state. The properties of jets are used to identify fully merged hadronically decaying top quarks [102]. In Fig. 4.2 the efficiency (left) to reconstruct

a top quark in an ak8 jet as a function of the top quark  $p_T$ . All top quarks with a  $p_T$  above 400 GeV can be reconstructed in an ak8 jet. The four-momenta sums of the subjects obtained with the soft drop algorithm are used to obtain the invariant mass of the top quark candidates (soft drop mass). The soft drop mass (right) is shown for events where the top quark is merged into one jet. High discrimination between multijet background and t quarks is achieved. Additionally, an excellent mass resolution is visible for the top quark. To tag t quarks, ak8 jets with a  $p_T > 400$  GeV are considered. The soft drop mass of t tagged jets has to fulfil  $105 \leq m_{\text{top}} \leq 220$  GeV. Furthermore, one of the soft drop subjects has to fulfil the b-tag criteria of the loose working point. The N-subjettiness ratio  $\tau_3/\tau_2$  of the ak8 jet has to be smaller than 0.5 such that fewer W bosons are misidentified and the invariant mass reconstruction of the top quark candidates is improved. Overall the misidentification rate is at the 0.1% level and the efficiency at approximately 15%.

## W tagging

Similar to t quarks, hadronically decaying W bosons are tagged using ak8 jets [102]. The W boson decay is a two-body problem. As a result, the  $p_T$  of the W boson at which its decay products merge into one ak8 jet are much smaller. For the tagging, ak8 jets with a  $p_T > 200$  GeV are used. The PF candidates used for clustering these jets, passed the puppi algorithm. As with the t tag, the invariant mass is the sum of the four-momenta of the subjects obtained from the soft drop algorithm. The invariant mass of the W boson is shifted to lower masses, which is corrected by applying  $\eta$  and  $p_T$  dependent correction factors. The correction factors are derived on the position of the W boson peak. The soft drop mass has to be in the range of  $65 \leq m_W \leq 220$  GeV. The N-subjettiness ratio  $\tau_2/\tau_1$  has been studied, but no improvement has been found when applying a selection to it. Overall, the misidentification rate is found to be smaller than 0.07% and the signal efficiency at the 80% level.

## 4.4. Missing transverse momentum

Most particles produced at the LHC leave distinct signatures in the CMS detector. Notable exceptions in the SM are weakly interacting neutrinos. Additionally, a broad set of hypothetical BSM particles with weak interactions have been predicted. The transverse momentum of particles not interacting with the detector is inferred from the  $p_T$  imbalance of the PF particles. The missing transverse momentum [113] is defined as

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{PFcandidates}} \vec{p}_{\text{PF}}, \quad (4.12)$$

which is the negative vectorial sum of the PF candidates. The considered PF candidates fulfil  $p_T > 1 \text{ GeV}$  and pass several quality criteria. The correction of the jet energy to particle level is propagated as

$$\vec{p}_T^{\text{miss,corr}} = \vec{p}_T^{\text{miss}} - \sum_{\text{ak4 jets}} (\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}). \quad (4.13)$$

Here, the difference of the corrected  $\vec{p}_{T,\text{jet}}^{\text{corr}}$  and uncorrected jets  $\vec{p}_{T,\text{jet}}$  is subtracted from the  $\vec{p}_T^{\text{miss}}$ , as defined in Eq. 4.12. In the following parts of the thesis  $\vec{p}_T^{\text{miss}}$  is used for the corrected missing transverse momentum if not denoted differently. In this analysis  $\vec{p}_T^{\text{miss}}$  is used to reduce the amount of multijet background and to reconstruct the neutrino expected in the signature of signal events.



---

## 5. Data and simulation

Proton proton collision data with a centre-of-mass energy of 13 TeV were recorded in 2016 with the CMS experiment, corresponding to an integrated luminosity of  $35.9 \text{ pb}^{-1}$ . The data have been collected using triggers based on muons and electrons. Photon triggers are also considered for events with a high electron momentum, where the electron is reconstructed as a photon at the online trigger level. This data is analysed for the search of the  $B+b$ ,  $B+t$ , and  $X_{5/3}+t$  production modes, where the VLQ decays into a top quark and W boson. Monte Carlo simulation is used to validate and optimise the analysis and to model the signal.

In Sec. 5.1 the data are presented, followed by the SM simulation in Sec. 5.2. Afterwards, the simulation of the VLQs is discussed in Sec. 5.3. Thereafter, in Sec. 5.3 the particular properties of the signal processes on the generator level are described and a selection of distributions shown.

### 5.1. CMS data sets

In 2016 the LHC delivered a total integrated luminosity of  $40.82 \text{ pb}^{-1}$ , of which CMS recorded  $37.76 \text{ pb}^{-1}$ . Of the recorded data,  $35.9 \text{ pb}^{-1}$  fulfils the quality requirements for this measurement. The reduction in recorded integrated luminosity is mostly, due to failures of sub-detector parts or the triggering system. These made it necessary to restart parts of the detector and discard recorded data. Data events are split into several sets according to the physics objects by which they are triggered. For the presented analysis the most important data events are in the sets with muons, electrons, and photons. The photon data is used to trigger electrons with high transverse

Table 5.1.: Triggers studied for the VLQ search with a lepton and jets in the final state. The minimum thresholds of the triggers are listed for the quantities calculated online during data taking. The photo trigger is employed to recover inefficiencies of the electron trigger at the highest  $p_T$ .

	lepton (photon) $p_T$ [GeV]	isolation	jet(s) $p_T$
single $\mu$	50	–	–
iso. $\mu$	24	✓	–
$\mu$ + jets	40	–	200 & 50
single e	115	–	–
iso. e	27	✓	–
e + jet	50	–	165
single $\gamma$	175	–	–

momentum. At trigger level, high momentum electrons and photons are very similar. The trigger demonstrate a slight (approximately 1%) inefficiency for electrons with a high momentum ( $p_T > 250$  GeV), which the trigger system instead identifies as photons. The quality criteria for data include at least one reconstructed primary vertex and that all major sub-detector parts be working. In Tab. 5.1 the requirements for the triggers studied for the VLQ search are summarised, illustrating their minimum thresholds of the corresponding quantities used for the online selection, during data taking. The values of the fully reconstructed leptons (photons) and jets are slightly different.

## 5.2. SM simulation

The most important SM processes are simulated using MC techniques based on SM calculations. Several “Monte Carlo event generators” exist which simulate the production and the decay of elementary particles according to their probability. A review of these generators and their current status can be found in Ref. [114]. In Fig. 5.1 the main components of the event simulation are illustrated. The calculations needed for a precise description include a wide range of effects from the PDFs, matrix elements (MEs) represented by Feynman diagrams, initial state radiation (ISR), and final state radiation (FSR) or multiple particle interactions (MPI) to the fragmentation inside jets. The events generated from these calculations are passed through a detector

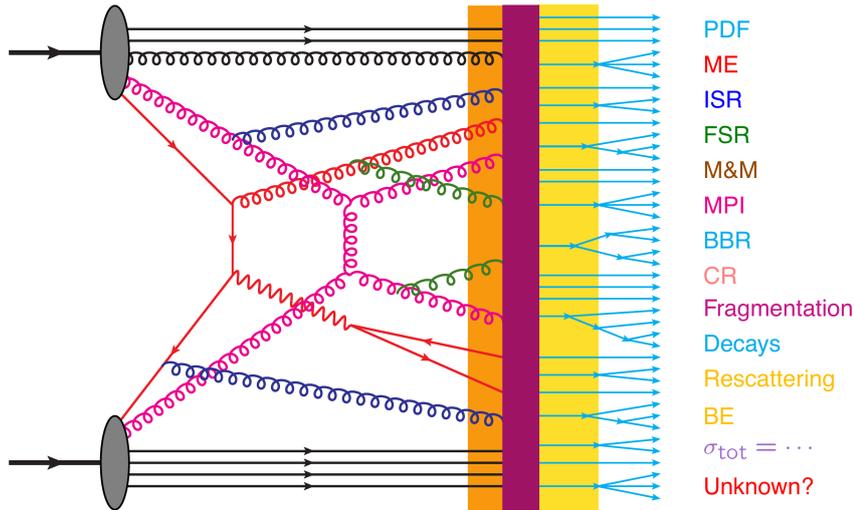


Figure 5.1.: The main components of the event simulation with a MC program [114].

simulation, based on GEANT4 [115] and capable of predicting the measurements in the sub-detector parts. Afterwards, the same algorithms used for data are applied on the simulation. The simulated events are used to evaluate the performance of the selection, optimise it, and validate the consistency of the employed algorithms.

The most important SM background processes for this search include top quarks. Events with a  $W$  boson decaying into a lepton and neutrino pair are the second major SM background. Processes with a  $Z$  boson or consisting of multiple jets from QCD (multijet) are also considered. Multijet processes are the most abundant at the LHC but are reduced by several orders of magnitude by the presence of a lepton and further by dedicated selections (discussed in Sec. 6.2).

Events are usually simulated for different particles and interactions separately, such that a large amount of simulated events is available for the statistical evaluation in the phase space of interest. MADGRAPH\_AMC@NLO (v2.2.2) [116] is used to generate single top quarks at next to leading order (NLO) in the  $s$ -channel. Single top quarks in the  $t$ -channel and single top quarks from the  $tW$  process, as well as top quark pairs produced via the strong force, are generated at NLO with POWHEG [117–119]. Both the events for the  $W$ +jets and  $Z$ +jets processes are generated with MADGRAPH\_AMC@NLO (v2.2.2) [116] using the FFXFX scheme [120] to match the parton shower (PS) emission for the former process and the MLM scheme [121] for the latter.

The Z+jets process is generated at leading order (LO) while the W+jets process is generated at NLO. The events for the multijet processes are generated using PYTHIA [122] at LO.

In Tab. 5.2 the SM background samples used for the VLQ search are presented. For each sample the generator, simulated process, and cross section are detailed. The highest order calculation available is used for the cross section. For most process being at NLO or next-to-next to leading order (NNLO) precision, apart from multijet events at LO. Most processes are further generated in bins, denoted by  $[\ ]$ , to achieve a sufficient number of events in the desired phase space. For top quark pairs, the invariant mass  $m_{t\bar{t}}$  of the generated top quark pair is used. The inclusive sample is used for events with  $m_{t\bar{t}} < 700$  GeV. For the W+jets sample only leptonic decays binned in the  $p_T$  of the W boson  $p_{T,\text{gen. W}}$  are considered. In the Z+jets samples the sum of the generated particles  $H_{T,\text{gen}}$  is used. For the multijet production process a large variety of samples is used. They are all binned according to the maximal transversal momentum of the generated particles  $\hat{p}_T$ . All samples with a muon in the final state are labelled with  $\mu$ . Multijet events with an electron or photon in the final state are labelled with  $e$ . Dedicated samples for the decay of a b or c quark into a electron are also considered and labelled  $bc \rightarrow e$ . The contributions of other processes are negligible for the VLQ search.

PYTHIA is used for the parton shower description and hadronisation of all the generated events, with HERWIG being the alternative approach. The PYTHIA tune has been studied in greater detail and is used in CMS for most MC simulation. The single top quark production in the  $t$ -channel uses as underlying event tune CUETP8M2T4 [123], while all other processes are simulated with the CUETP8M1 [124, 125] event tune. The NNPDF 3.0 [126] are used for the PDFs as recommended for the physics analysis at the LHC [127, 128]. The precision of the PDFs is matched to that of the ME calculation. Pileup from the same and previous bunch crossings are taken into account in the simulation, and the profile of primary vertices is reweighted to follow the one in the data.

Table 5.2.: Simulated MC samples of the SM background used for the evaluation of the selection and the agreement to data. For each sample the used cross section and generator are detailed. The highest order calculation available is used for the cross section. All events are interfaced with PYTHIA for the parton shower and to GEANT4, for the hadronisation and the detector simulation.

Background Process	Generators	Cross Section [pb]
Single top t-channel	POWHEG	136
Single top ass. with W boson	POWHEG	35.85
Single top s-channel	AMC@NLO	3.36
$t\bar{t}$ +jets inclusive	POWHEG	831.76
$t\bar{t}$ +jets $m_{t\bar{t}} \in [700, 1000]$ GeV	POWHEG	76.6
$t\bar{t}$ +jets $m_{t\bar{t}} > 1000$ GeV	POWHEG	20.578
W +jets $p_{T,\text{gen. } W} \in [100, 250]$ GeV	AMC@NLO FFXFX	676.3
W +jets $p_{T,\text{gen. } W} \in [250, 400]$ GeV	AMC@NLO FFXFX	23.94
W +jets $p_{T,\text{gen. } W} \in [400, 600]$ GeV	AMC@NLO FFXFX	3.031
W +jets $p_{T,\text{gen. } W} > 600$ GeV	AMC@NLO FFXFX	0.4524
Z +jets $H_{T,\text{gen}} \in [100, 200]$ GeV	MADGRAPH MLM	181.3
Z +jets $H_{T,\text{gen}} \in [200, 400]$ GeV	MADGRAPH MLM	50.4
Z +jets $H_{T,\text{gen}} \in [400, 600]$ GeV	MADGRAPH MLM	6.98
Z +jets $H_{T,\text{gen}} > 600$ GeV	MADGRAPH MLM	2.7
Multijet $\hat{p}_T \in [20, 30]$ GeV $\mu$	PYTHIA	2960198.4
Multijet $\hat{p}_T \in [30, 50]$ GeV $\mu$	PYTHIA	1652471.46
Multijet $\hat{p}_T \in [50, 80]$ GeV $\mu$	PYTHIA	437504.1
Multijet $\hat{p}_T \in [80, 120]$ GeV $\mu$	PYTHIA	106033.66
Multijet $\hat{p}_T \in [120, 170]$ GeV $\mu$	PYTHIA	25190.5
Multijet $\hat{p}_T \in [170, 300]$ GeV $\mu$	PYTHIA	8654.49
Multijet $\hat{p}_T \in [300, 470]$ GeV $\mu$	PYTHIA	797.35
Multijet $\hat{p}_T \in [470, 600]$ GeV $\mu$	PYTHIA	79
Multijet $\hat{p}_T \in [600, 800]$ GeV $\mu$	PYTHIA	25
Multijet $\hat{p}_T \in [800, 1000]$ GeV $\mu$	PYTHIA	4.7
Multijet $\hat{p}_T < 1000$ GeV $\mu$	PYTHIA	1.6
Multijet $\hat{p}_T \in [15, 20]$ GeV $e$	PYTHIA	2302200
Multijet $\hat{p}_T \in [20, 30]$ GeV $e$	PYTHIA	5352960
Multijet $\hat{p}_T \in [30, 50]$ GeV $e$	PYTHIA	9928000
Multijet $\hat{p}_T \in [50, 80]$ GeV $e$	PYTHIA	2890800
Multijet $\hat{p}_T \in [80, 120]$ GeV $e$	PYTHIA	350000
Multijet $\hat{p}_T \in [120, 170]$ GeV $e$	PYTHIA	62964
Multijet $\hat{p}_T \in [170, 300]$ GeV $e$	PYTHIA	18810
Multijet $\hat{p}_T < 300$ GeV $e$	PYTHIA	1350
Multijet $\hat{p}_T \in [20, 30]$ GeV $bc \rightarrow e$	PYTHIA	328999.9
Multijet $\hat{p}_T \in [30, 80]$ GeV $bc \rightarrow e$	PYTHIA	405623.4
Multijet $\hat{p}_T \in [80, 170]$ GeV $bc \rightarrow e$	PYTHIA	38104.43
Multijet $\hat{p}_T \in [170, 250]$ GeV $bc \rightarrow e$	PYTHIA	2635.8
Multijet $\hat{p}_T < 250$ GeV $bc \rightarrow e$	PYTHIA	711.9

### 5.3. VLQ simulation

Models that include phenomena beyond the SM are also often described within MC generator frameworks. For this search, a simplified model approach is used to give model-independent results.

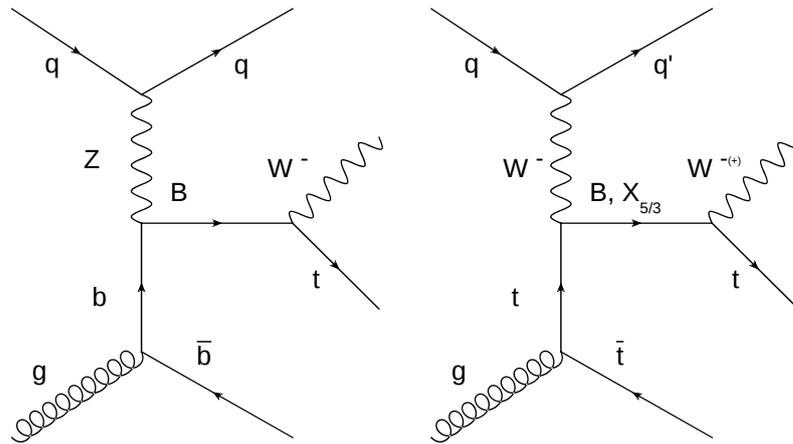


Figure 5.2.: Leading order Feynman diagrams for the production of a single vector-like  $B$  or  $X_{5/3}$  quark in association with a  $b$  (left) or  $t$  (right) and a light-flavour quark, and the subsequent decay of the VLQ to  $tW$  [30].

Two of the LO Feynman diagrams of interest are shown in Fig. 5.2. The production of single VLQ at the LHC is expected to occur mostly by the interaction of a  $b$  or  $t$  quark produced by a gluon splitting, and a boson radiated from a quark. The quark emitting the boson is only slightly deviated from the beam direction such that it is found close to the beam pipe. This quark is referred to as the forward quark, and forms the forward jet. Typically, it has a low transverse momentum and high energy. The quark from the gluon splitting is in the central part of the detector with low transverse momentum compared to the quarks from the VLQ decay.

MADGRAPH\_AMC@NLO (v2.2.2) is used to generate signal events with a modified version of the model from Ref. [61, 129, 130]. The  $W$  boson and  $t$  quark from the VLQ decay are treated by the MADSPIN [131, 132] package to retain the correct spin correlations. The parton shower and hadronisation are handled by PYTHIA using the CUETP8M1, and the simulation of the detector is done with GEANT4, as with the SM simulation. The predicted cross section is obtained from calculations used in

Table 5.3.: Summary of signal samples used, together with their mass ranges, mass widths, and either right-handed or left-handed VLQ couplings, in steps of 100 or 200 GeV.

Signal process	Mass range [GeV]	Mass steps [GeV]	Widths	Couplings
B+b	700 to 1800	100	1%	left & right
B+b	800 to 2000	200	10, 20, 30%	left
B+t	700 to 1800	100	1%	left & right
B+t	800 to 2000	200	10, 20, 30%	left & right
$X_{5/3}+t$	700 to 1600	100	1%	left & right
$X_{5/3}+t$	800 to 2000	200	10, 20, 30%	left

Ref. [69, 70, 133]. For the predicted cross section a simplified approach is used to provide a model-independent interpretation of the experimental results for relative mass widths between 1 and 30%. Three searches for single VLQ have been performed with this approach [10, 11].

Samples are generated for B and  $X_{5/3}$  produced in association with a t or b, right- or left-handed VLQ couplings, and decay widths relative to the VLQ mass of  $(\Gamma/m)_{\text{VLQ}} = 1, 10, 20,$  and  $30\%$ . While the B could be produced in association with both quarks, the  $X_{5/3}$  could only be produced in with a t as long as the charge is conserved. Samples with a relative mass width between 10 and 30% (1%) are simulated in mass steps of 200 GeV (100 GeV) between 0.8 (0.7) and 2 TeV (1.8 TeV). While samples with a 1% width are produced with both coupling types, samples with higher widths are mostly produced with left-handed VLQ couplings. The available and used samples are summarised in Tab. 5.3.

The forward jet distinguishes the signal sample from most SM processes. Due to the production mode, it has a low transverse momentum, high energy, and high  $|\eta|$  value. The high mass of VLQ candidates leads to a highly boosted W boson and t quark from the B or  $X_{5/3}$  decays located in opposite directions of the detector. Figure 5.3 illustrates the distributions on generator level of the  $\eta$  of the forward quark (upper left), the distance between the decay products (W boson and b quark) of the t quark originating from the VLQ decay chain (upper right), the transverse momentum of the t quarks (lower left) in the event, and the transverse momentum of the lepton (lower right). The signal samples are normalised to a cross section of 1 pb. The jet near the

beam pipe usually has an  $|\eta|$  around 3.5 and is used to define a signal region (SR), while events without such a jet are in a sideband (SB) region. Its properties make the forward jets challenging to distinguish from other radiation such as pileup. The distance of the t quark decay products, originating from the decay of the VLQ, are small and can often be reconstructed in a jet with a radius parameter of 0.8. Additionally, the transverse momentum distribution of all generated t quarks indicates that for most samples there are two t quarks. One carries about half the energy of the VLQ mass and the other has a relatively low transverse momentum. The t quark with low transverse momentum originates from the associated production through gluon splitting. On average, the lepton in the event has a low transverse momentum, which makes triggering these events challenging.

In Fig. 5.4 the “back to back” topology of the VLQ decay leads to a large distance in  $\Delta R$  between the W boson and t quark from the VLQ decay (left). Since the mass of the VLQ under consideration is much higher than the mass of either decay product, the momenta of the decay products are roughly the same. The momenta balance is confirmed by the ratio of the W boson and t quark momenta (right). The back to back topology is later exploited in the reconstruction, and both the distance, as well as the  $p_T$  ratio are taken into account. The distributions illustrated in Fig. 5.3 and 5.4 are similar for all samples.

In the typical signature, several quarks from the VLQ decay chain and quarks produced in association are present. This signature leads a high jet multiplicity, which increases the number of possible jet combinations in the reconstruction. A large number of jets also influence the triggering since most leptons are in close vicinity to jets. On the other hand, the high mass of the VLQ allows the usage of jet taggers and the back to back topology in the reconstruction.

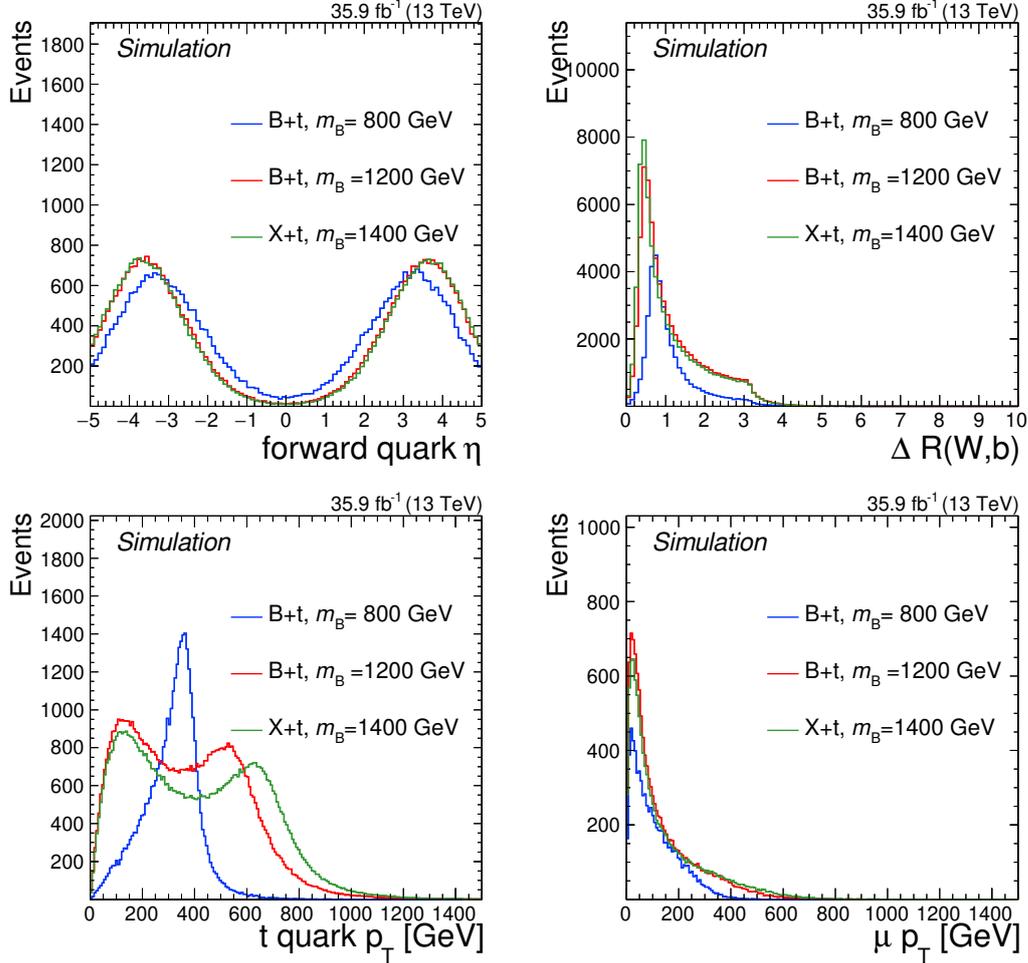


Figure 5.3.: Distributions from generator quantities for the muon channel and three different VLQ samples with masses of 800, 1200, and 1400 GeV and left-handed VLQ coupling in the three considered production modes. The  $\eta$  of the forward quark (upper left) and the distance of the decay products of the t quark from the VLQ decay (upper right) are shown, as well as the momentum distribution of all generated t quarks (lower left) and of the lepton (lower right). The samples are normalised to a cross section of 1 pb.

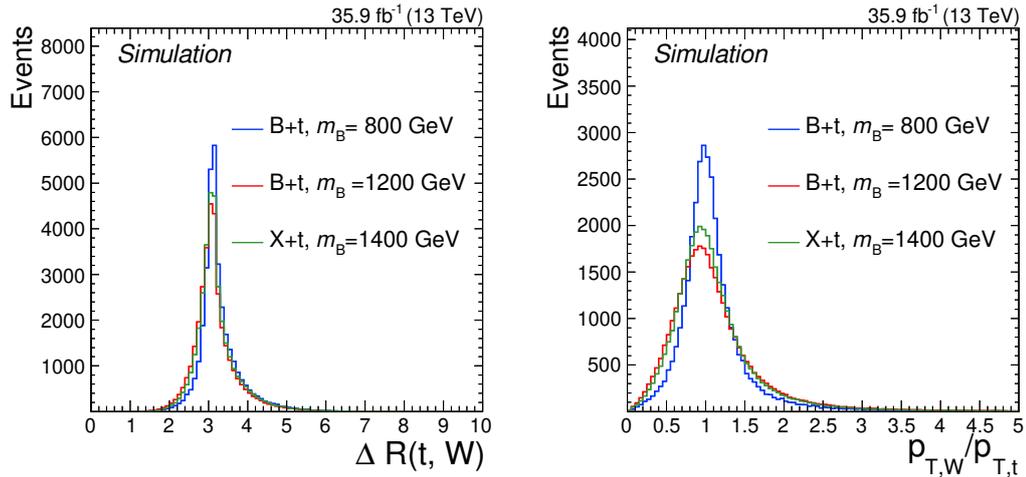


Figure 5.4.: Distributions from generator quantities for the muon channel and three different VLQ samples with masses of 800, 1200, and 1400 GeV and left-handed VLQ coupling in the three considered production modes. The distribution of the distance in  $\Delta R$  between the  $W$  boson and  $t$  quark from the VLQ (left) and the ratio of the  $W$  boson and  $t$  quark momenta (right) is shown. The samples are normalised to a cross section of 1 pb.

## 5.4. HCAL simulation

The simulation of the particles with GEANT4 is also referred to as “FullSim”. This type of simulation is used for most SM processes and many signal signatures. For searches with many signatures such as supersymmetry the simulation of all signal signatures with the FullSim approach is too time-consuming. Therefore an alternative approach that is referred to as “FastSim” [134–136] was developed to parameterise the most important effects of the particle decay in the detector. The parameterisation reduces the overall computing time for a single event by up to two orders of magnitude. FastSim is also used to evaluate additional uncertainties arising, for example, from theoretical uncertainties.

The most computationally demanding tasks in FullSim are the simulation of the interactions in the tracker and the calorimeter. In these two parts of the detector the multiplicity of interactions is high. A combinatorial approach is used for the reconstruction of the particles in the tracker while in the calorimeters the particles are clustered. In FastSim the information of the ME is exploited to considerably reduce the track reconstruction time. In Fig. 5.5 a sketch of the simulation of the hadronic

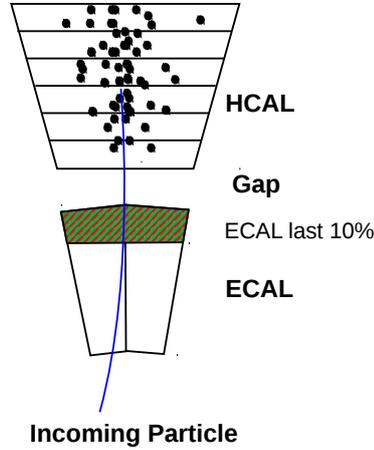


Figure 5.5.: Sketch of a hadronic shower simulated with FastSim in a small part of the CMS detector.

showers in the CMS detector with FastSim is provided. An incoming hadron is supposed to transverse most of the ECAL, leaving only a minimal amount of energy and beginning the shower in either the 10% of the ECAL most distant from the interaction or the HCAL. The energy deposited in the calorimeter by the shower in the longitudinal and transversal direction is sampled from probability distributions obtained from FullSim. Only those energy entries in the active calorimeter material are considered for the reconstruction.

Hadrons have many decay channels, resulting in a multitude of signatures in the detector. The multitude of channels make the parameterisation and simulation of the hadronic showers with the FastSim approach challenging. Typically, pions are the hadrons most present in the HCAL detectors at the LHC. In Fig. 5.6, four shower examples are shown as simulated with FullSim. These showers originate from a pion simulated in the barrel part of the CMS detector with an energy of 20 GeV. The starting point of the shower is always set to the origin of the coordinate system. The shower examples indicate that the pion decay can have a multipronged structure

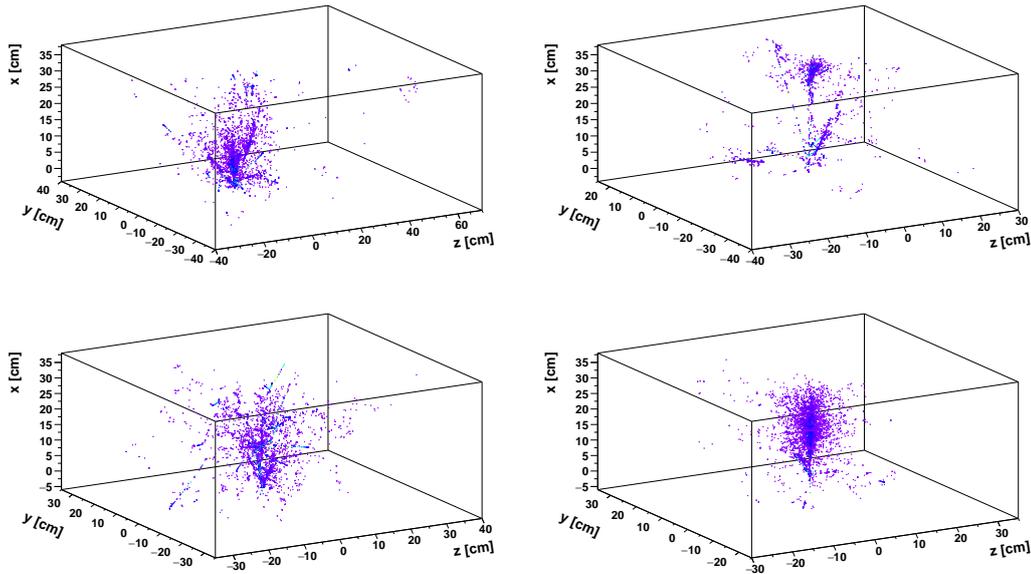


Figure 5.6.: Example of pion showers simulated with GEANT4 in the CMS detector. The pions have an energy of 100 GeV and are simulated in the central part of the detector. The shower start has been fixed to the origin of the coordinate system.

(upper and lower right) or be concentrated in a single prong (lower left), similar to an electron decay. The GFLASH [137, 138] ansatz is used to simulate the shower shape  $\mathcal{S}$ . The shower shapes are factorised into longitudinal  $\mathcal{L}(z)$  and transversal  $\mathcal{T}(r)$  probability density functions as

$$\mathcal{S}(r, z) = \mathcal{L}(z)\mathcal{T}(r). \quad (5.1)$$

Apart from depending on the radial  $r$  and longitudinal  $z$  distance from the interaction these probabilities also depend on the energy

$$\mathcal{L} : p(L|E), \quad (5.2)$$

$$\mathcal{T} : p(T|E). \quad (5.3)$$

To describe the longitudinal part, gamma distributions are used of the form

$$G(z, \alpha, \beta) = \frac{(\beta z)^{\alpha-1} e^{-\beta z}}{\Gamma(\alpha)} \quad (5.4)$$

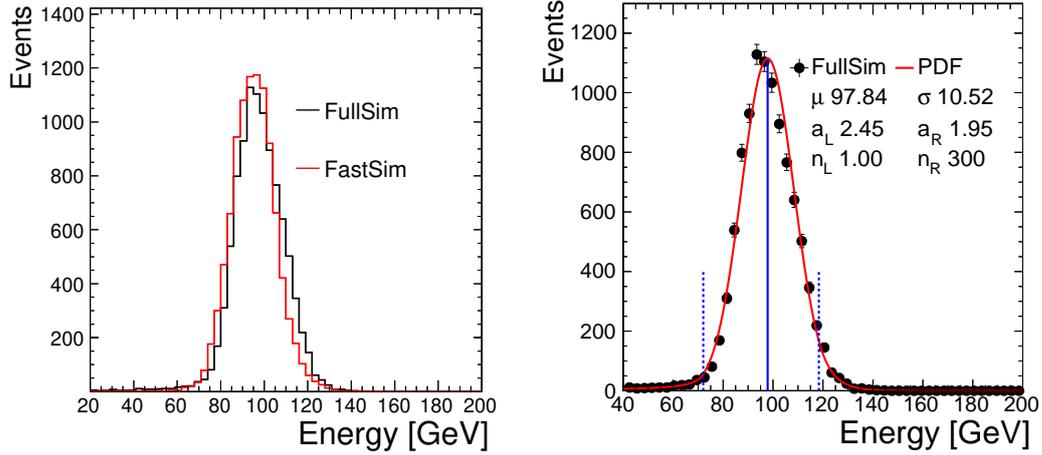


Figure 5.7.: Comparison of the energy distribution of 10000 pions between FullSim and FastSim (left). Fit of the FullSim showers with a Crystal Ball function as defined in Eq. (5.6) (right). The blue dashed lines mark the endpoint of the Gaussian core. The solid blue line shows the mean of the Gaussian core. The pions are generated with an energy of 100 GeV and are simulated in the central part of the detector.

Both,  $\alpha$  and  $\beta$  are free parameters obtained from fits to FullSim in a wide range of hadron energies. The gamma function is represented by  $\Gamma$ . The expected value is  $E(X) = \frac{a}{b}$  and the variance is  $Var(X) = \frac{a}{b^2}$ , which make it simple to find start values for  $\alpha$  and  $\beta$  in the fitting. The central limit theorem ensures that the total energy follows the normal distribution. For the transversal shower shape a less complicated ansatz is chosen with

$$T(r) = \frac{2rR^2}{(r^2 + R^2)^2}, \quad (5.5)$$

having only  $R$  as a free parameter.

Even though both the radial and transversal shower shapes have been tuned with events from FullSim, the energy distributions of the pion showers between FullSim and FastSim have residual differences. In Fig. 5.7 the energy distributions for FullSim and FastSim (left) for 10000 pions generated with an initial energy of 100 GeV is shown. The energy deposited in the detector is corrected by an additional probability density

function obtained from a fit to FullSim (right). For the fit a double sided Crystal Ball function is used defined as

$$f(E, \sigma, a_L, n_L, a_R, n_R) = \begin{cases} e^{-\frac{1}{2}a_L a_L (1 - \frac{1}{n_L}(a_L - (\frac{\mu - E}{\sigma})))^{-n_L}} \\ e^{\frac{1}{2}(\frac{\mu - E}{\sigma})^2} \\ e^{-\frac{1}{2}a_R a_R (1 - \frac{1}{n_R}(a_R + (\frac{\mu - E}{\sigma})))^{-n_R}}. \end{cases} \quad (5.6)$$

The function is named after the Crystal Ball Collaboration, and has six free parameters:  $\mu, \sigma, a_L, n_L, a_R$  and  $n_R$ . It consists of a Gaussian core described by the mean  $\mu$  and the variance  $\sigma$ . The other parameters correspond to power-law tails on either side. During this thesis the residual corrections were tuned to the latest FullSim version for energies between 1 GeV and 1 TeV in the central part of the detector. The updated correction improve the agreement of the energy of ak4 jets between FullSim and FastSim by up to 10%.

Apart from the residual corrections, FastSim can also be improved by extending the GFLASH used in CMS. For HCALs the original authors have suggested the usage of a longitudinal profile

$$\mathcal{L} = f_{em} G_{em}(z, \alpha_{em}, \beta_{em}) + (1 - f_{em}) G_{had}(z, \alpha_{had}, \beta_{had}). \quad (5.7)$$

The profile consists of two  $G(\alpha, \beta)$  functions that are responsible for the electromagnetic  $G_{em}$  and the hadronic  $G_{had}$  parts of the shower. In total, five free parameters are used:  $\alpha_{em}, \beta_{em}, \alpha_{had}, \beta_{had}$ , and  $f_{em}$ . The last is the energy fraction corresponding to the electromagnetic part of the shower profile. In Fig. 5.8 the longitudinal shower profiles from the showers presented in Fig. 5.6 are fitted with the extended ansatz. For most showers good fit parameters between the longitudinal shower profile and the  $G(\alpha, \beta)$  functions are found. The goodness of fit is expressed in terms of the  $\chi^2$  probability. Due to the large of amount of fluctuations possible, not all longitudinal shower shapes are captured by the extended ansatz. The accurate simulation of hadronic showers has gained importance in the last years with the rise of substructure techniques and is currently under further investigation in CMS for use in FastSim.

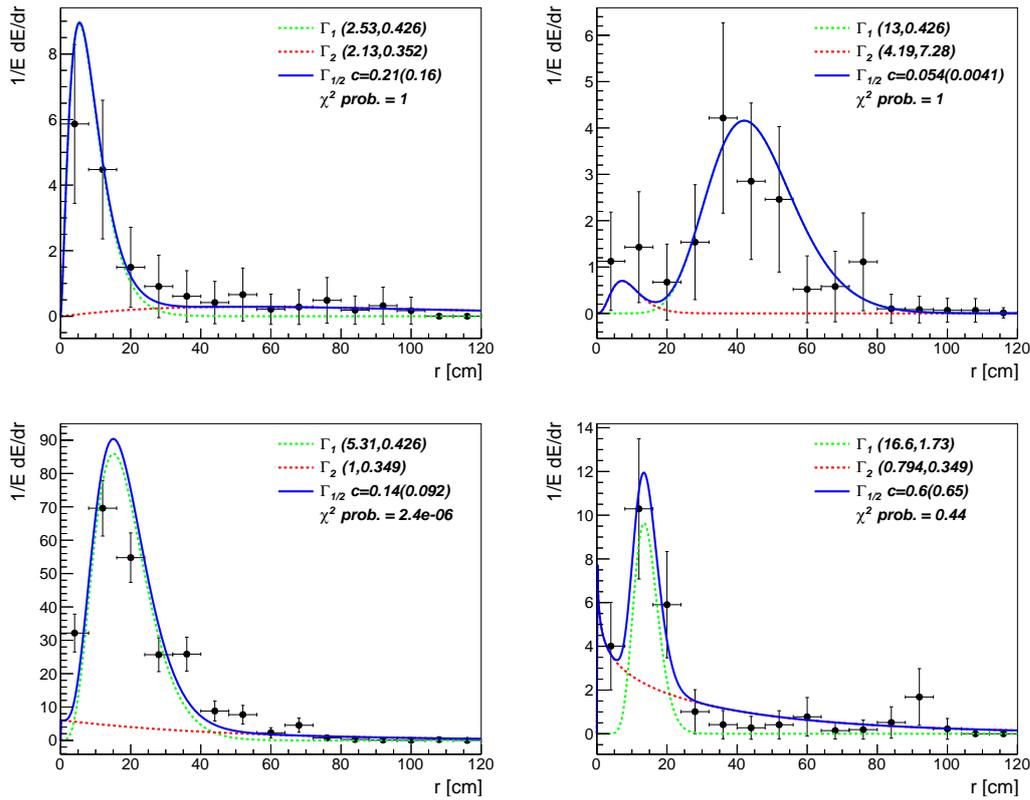


Figure 5.8.: Longitudinal showers profile of pions from Fig. 5.6 fitted by two  $G(\alpha, \beta)$  functions as in eq. (5.7). For most showers good fit parameters between the longitudinal shower profile and the  $G(\alpha, \beta)$  functions are found.



## 6. Search for singly produced VLQs

Vector-like quarks, as searched for in this analysis, are an additional generation of quarks predicted by many extensions of the SM (discussed in Sec. 2.2.1). A plethora of searches for VLQs have been performed at ATLAS and CMS [2–28]. While most of these searches focus on pair-produced VLQs, this search is designed for single vector-like B and  $X_{5/3}$  quarks with masses in the TeV region decaying to a t quark and a W boson with a muon or electron in the final state. This is the first search at a centre-of-mass energy of 13 TeV for B and  $X_{5/3}$  in the final state with one lepton and jets. Masses in the TeV region or above are very interesting for single VLQ production modes since their production cross sections can be significantly higher than for pair-produced VLQs. Single  $X_{5/3}$  and B are produced in association with a t quark or in association with a b quark. For the three production modes (B+b, B+t, and  $X_{5/3}$ +t) a light quark is also emitted close to the beam pipe under a small scattering angle (forward direction).

The analysis strategy to achieve the best sensitivity for B and  $X_{5/3}$  is schematically depicted in Fig. 6.1. Even though most of the VLQ signal is expected to decay into fully hadronic final states, at the LHC the production of multijet events has a cross section several orders of magnitude larger than any other processes. Because of the large background for most searches the sensitivity is smaller than in events with leptons. Additionally, the accurate simulation of these processes is challenging, and the uncertainties are usually larger than for final states with a lepton. Muons and electrons decay significantly differently than hadrons and are used here to trigger interesting events. Compared to the other charged leptons, taus decay more similarly to hadrons, and events consisting of a tau quark and jets are more difficult to distinguish from multijet events. Their contribution to the sensitivity with the current tau identification algorithms is expected to be small. For future analysis with better tau quark

discrimination, searches with tau quarks in final state might be important. For events that are triggered, the invariant mass  $m_{\text{reco}}$  of the VLQ is reconstructed. The  $m_{\text{reco}}$  distribution is expected to show a peak for a VLQ signal, and a falling invariant tW mass spectrum from SM backgrounds. For high VLQ masses this provides a good separation between SM backgrounds and B or  $X_{5/3}$ . The invariant mass of the tW system is reconstructed from jets, the lepton, and the missing transverse momentum ( $\vec{p}_{\text{T}}^{\text{miss}}$ ) in the event. The reconstruction is optimised for VLQ masses in the TeV region or above for VLQs decaying into t quarks and W bosons with high transverse momentum that are found in opposing directions of the detector (“back to back”). The event selection is optimised using simulation based on the  $m_{\text{reco}}$  distributions, which is later also used for the statistical interpretation. The sensitivity to VLQs is further maximised by taking advantage of the characteristic event properties based on t, W, and b tagged jets, as well as lepton flavour. The SM background prediction is conducted using a data-driven approach where the jet in the forward direction is exploited to form a SR and SB. The shape of the invariant VLQ mass distribution in the SB, which is dominated by SM background, is used as the background shape of the invariant mass distribution in the SR. By using a background prediction from data, systematic uncertainties from the SM simulations are reduced. The prediction of the background shape is validated both in simulation and in data using a dedicated region, which is not sensitive to B or  $X_{5/3}$  production.

The analysis strategy is described in detail below. First, the reconstruction of the tW system is discussed in Sec. 6.1, which describes the employed algorithms and the performance for signal, simulation, and data. Afterwards, in Sec. 6.2 the selection is explained, including its triggers and optimisation. Additionally, the agreement between data and simulation is investigated in the selected phase space. In Sec. 6.3 the categorisation based on the jet tags is specified. The usage of the forward jet for the SR, the data-driven background prediction, and the validation of the employed method are described in Sec. 6.4. The figures in this chapter illustrate distributions for events after the selection detailed in Tab. 6.3 and VLQ signal events with a relative width of 1% if not noted otherwise.

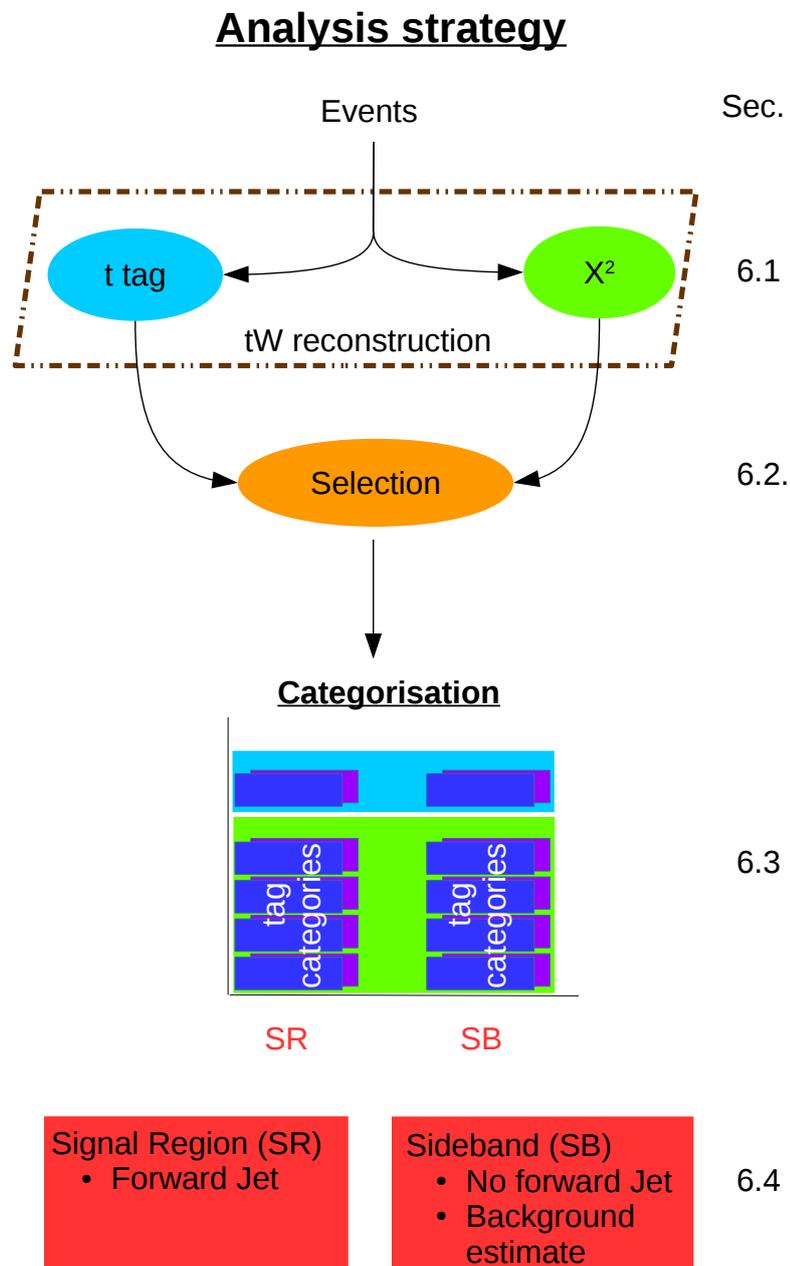


Figure 6.1.: Schematic representation of the analysis strategy. The mass is reconstructed for each event. Events that pass the selection are divided into categories where the requirement of a jet in forward direction is used to obtain a SR and SB. The background shape of the invariant mass distribution is estimated from the SB.

## 6.1. Top quark and W boson reconstruction

Reconstructing the mass of particles and quarks has been used extensively in the past, such as in the  $J/\psi$  discovery [139, 140], the discovery of the top quark [141, 142], and the Higgs discovery. The reconstructed invariant mass of the combined t quark and W boson system is the best discriminator between VLQ signal and SM background found in this analysis. A large number of variables have been inspected based on four-momentum considerations but showed less discrimination power. The t quark and W boson signature appears in the detector as jets,  $\vec{p}_T^{\text{miss}}$ , and a lepton. For the reconstruction it is assumed that the  $\vec{p}_T^{\text{miss}}$  and the lepton arise only from the decay chain of the VLQ. Particles (including neutrinos and leptons) from other decays in the events typically have a very low  $p_T$  (e.g. the associated t or b) and are not further considered. To reconstruct the hadronic part of the VLQ decay chain, combinations of jets are tested. For the tW system, since VLQ masses larger than 700 GeV are being considered, this implies a high Lorentz boost for the decay products of the t quark and W boson, which leads to them being grouped closely together. The reconstruction is designed to allow jets to contain the energy of several quarks and exploits the “back to back” topology of the t quark and W boson.

Two decay chains of the VLQ considered are: (i) a hadronically decaying t quark ( $t \rightarrow Wb \rightarrow q\bar{q} + b$ ) with a leptonically decaying W boson ( $W \rightarrow l + \nu$ ) and (ii) a leptonically decaying t quark ( $t \rightarrow Wb \rightarrow l\nu + b$ ) with a hadronically decaying W boson ( $W \rightarrow q\bar{q}$ ). The former case is called the hadronic t (i) and the latter the leptonic t (ii) hypothesis. Before the hypothesis selection it is ambiguous whether the lepton originates from the W boson or the t quark of the VLQ decay. Leptons and jets that are products of the decay of associated quarks are not considered in the reconstruction, because their  $p_T$  is very low and they are often not identified.

### 6.1.1. Leptonic W boson

All VLQ signatures of interest have a W boson in the decay chain which decays leptonically. The W boson, which decays into a neutrino and a lepton, is reconstructed from the  $\vec{p}_T^{\text{miss}}$  and the four-momentum of the lepton. The  $\vec{p}_T^{\text{miss}}$  is associated with the

transverse momentum of the neutrino. From four-momentum calculations, where the mass of the W boson ( $M_W$ ) is constrained to  $M_W = 80.4$  GeV, the neutrino momentum in z-direction is obtained as

$$p_{z,\nu} = \alpha \frac{p_{z,l}}{p_{T,l}} \pm \sqrt{E_l^2 \frac{\alpha^2 - p_{T,\nu}^2 p_{T,l}^2}{p_{T,l}^4}}, \quad (6.1)$$

$$\alpha = \frac{M_W^2 - M_l^2}{2} + p_{T,l} p_{T,\nu} \cos(\Delta\phi).$$

In the formulas  $p_{T,l}$  denotes the transverse momentum,  $E_l$  the energy, and  $M_l$  the invariant mass of the lepton. The azimuthal separation between the neutrino and lepton is given by  $\Delta\phi$ . The two solutions of the quadratic equation are taken into account in the hypothesis selection. If the reconstructed event has a  $p_{z,\nu}$  with an imaginary part, this part is minimised by varying the  $x$  and  $y$  components of  $\vec{p}_T^{\text{miss}}$ . The imaginary part arises due to additional neutrinos in the event or the mismeasurement of the  $\vec{p}_T^{\text{miss}}$ , which is mostly the limiting factor in the leptonic W boson mass and position resolution. A more detailed description on the minimisation of the imaginary part can be found in Ref. [143, 144]. The misidentification of leptons is a second source of uncertainty which in the muon channel is very small and for electrons is only a little larger.

### 6.1.2. Jet assignment and hypothesis selection

For the decay chain of B or  $X_{5/3}$ , at least half of the energy is associated to hadronic decays, which appear as jets in the detector. Since the LHC is a hadron collider, jets are abundant not only when the strong force is involved but also in electroweak interactions. This makes the association of jets to hadrons challenging. However, combinations of jets with the top quark mass, W boson mass, back to back topology with a high invariant mass, or jets with high  $p_T$  are far less common, and these requirements are exploited to select the right hypothesis.

The hadronic part of the VLQ decay chains (i) and (ii) are reconstructed using either a t tag or ak4 jets. If a t tag is found in the event, the decay chain with a hadronic

top ( $i$ ) is assumed for the event. The reconstruction with a  $t$  tag is preferred because the probability to correctly assign the jet, as well as the spatial and mass resolutions for  $t$  quarks with high  $p_T$ , is better in comparison to the reconstruction with  $ak4$  jets. In contrast, the algorithms with  $ak4$  jets can reconstruct both decay chains and have a higher acceptance, while the mass resolution of the VLQ is only 2–3% worse. The assignment of the correct jet hypothesis is the most difficult part of the VLQ reconstruction due to the number of combinatorial possibilities as a result of parts of the decay chain being outside of the acceptance of the detector and that there are additional jets due to pileup. Both algorithms presented here have been scrutinised in terms of parameters and criteria employed. The use of information from  $W$  tags and  $b$  tags has also been studied. The algorithm is chosen based on the expected sensitivity in terms of exclusion limits at 95% confidence level (CL). All other algorithms and parameters have a lower sensitivity and were dropped. The reconstruction of the  $W$  or  $t$  with quarks in the final state is crucial but more important is the resolution of the reconstructed mass of the VLQ since a smaller resolution leads to a higher discrimination between signal and background.

### Events with $t$ tag

Top quarks with  $p_T > 400$  GeV reconstructed within an  $ak8$  jet are relatively rare and dominated in the SM by pair produced top quarks with a very small contribution of events produced in multijet or  $W$ +jets processes. The  $t$  tag is associated to the  $t$  quark and the  $tW$  system is calculated by combining the four-vector of the  $t$  tagged jet and the four-vector of the leptonically decaying  $W$  boson candidates. The reconstructed  $tW$  system with the minimal value of  $|\Delta R(W, \text{top}) - 180^\circ|$  is chosen - reflecting the back to back topology of the signal. This reconstruction is referred to as “ $t$  tag reconstruction” in the remainder of this thesis.

### Events without $t$ tag

The VLQ signals under consideration often do not have a  $t$  tag either because of the leptonic top ( $i$ ) decay chain or because, for example, parts of the decay are not clustered within the  $ak8$  jet. For most events, instead of using an  $ak8$  jet at least one

ak4 jet is used to construct the VLQ hypothesis, with  $|\eta| < 2.4$  and  $p_T > 30$  GeV. For the hypothesis selection  $\chi^2$  functions are evaluated for the various possible jet assignments. The hypothesis with the highest  $\chi^2$  quantity is chosen based on the probability  $\text{Prob}(\chi^2, n_{dof})$  of the  $\chi^2$  functions.

The terms in the  $\chi^2$  functions compare the means and deviations of a given permutation with expected values as obtained from simulation. The reconstruction based on the  $\chi^2$  functions is hereafter called “ $\chi^2$  reconstruction”. The information about which decay chain is reconstructed is not used but might improve the sensitivity with higher statistics; for example, using the information in the categorisation. The reconstructed invariant mass is usually robust against misidentifying the decay chain since it is the sum of the assigned jets and in both reconstruction chains most of the energy of the simulated VLQ is reconstructed.

The parameters used in the  $\chi^2$  function are based on the reconstructed candidates. The angular distance  $\Delta R_{t,W}$  is calculated between the reconstructed t and W candidates. The transverse momentum of the t quark (W boson) candidate are denoted with  $p_{T,t}$  ( $p_{T,W}$ ). The combination of the angular distance and the  $p_T$  ratios encode the back to back topology. The  $\sigma$ 's denote the standard deviations. Overlined variables denote the expected values from simulation. For the t quark, the invariant mass was extracted for the final state with a lepton  $\overline{m}_{t,lep}$  and the final state consisting only of hadrons  $\overline{m}_{t,lep}$ . The measurements of the particles in both final states have different uncertainties that are propagated to the reconstructed t quark candidates. For the W boson only the mass of the hadronic final state is used since the mass is already employed for the reconstruction of the leptonic W candidates. The W and t candidate masses in the  $\chi^2$  functions are only used if the respective candidate is reconstructed from at least two jets. Otherwise, the invariant jet mass can fluctuate because of additional radiation, which is not considered. In a future version of this analysis, this could be changed by considering groomed ak4 jets, but the improvements are expected to be small. The  $\chi^2$  hypotheses are divided into several classes based on the number of ak4 jets. In the case of at least three ak4 jets, all quarks are expected to decay into separate jets. Topologies where two or more quarks of the final state are reconstructed in one ak4 jet are covered by the one or two ak4 jet class.

**3 or more ak4 jets** In the case where three or more ak4 jets are used, the back to back topology and the masses of the t quark and W boson are exploited,

$$\chi_{t,\text{had}}^2 = \frac{(m_{t,\text{had}} - \bar{m}_{t,\text{had}})^2}{\sigma_{t,\text{had}}^2} + \frac{(\Delta R_{t,W} - \pi)^2}{\sigma_{\Delta R}^2} + \frac{(1 - p_{T,W}/p_{T,t})^2}{\sigma_{p_{T,\text{ratio}}}^2}, \quad (6.2)$$

$$\chi_{t,\text{lep}}^2 = \frac{(m_{t,\text{lep}} - \bar{m}_{t,\text{lep}})^2}{\sigma_{t,\text{lep}}^2} + \frac{(m_{W,\text{had}} - \bar{m}_{W,\text{had}})^2}{\sigma_{W,\text{had}}^2} + \frac{(\Delta R_{t,W} - \pi)^2}{\sigma_{\Delta R}^2} + \frac{(1 - p_{T,W}/p_{T,t})^2}{\sigma_{p_{T,\text{ratio}}}^2}. \quad (6.3)$$

For the hadronic t (*i*) hypothesis Eq. (6.2) is used and for the leptonic t (*ii*) case Eq. (6.3) is used. The hadronic t candidates and the hadronically decaying W boson are formed by at least two ak4 jets and the leptonic t with at least one associated ak4 jet. This leads to four free parameters in the leptonic t case and three in the hadronic t case.

**2 ak4 jets** For the hypotheses with two ak4 jets, both the hadronic t, and the leptonic t hypothesis are considered. The respective  $\chi^2$  functions exploit the back to back topology, the  $p_T$  ratio, and the respective mass of the VLQ decay products. For the leptonic t hypothesis one ak4 jet would be associated with the t quark and one with the hadronic W boson. Since the hadronically decaying W candidate consists of one ak4 jet, its mass is not used. For the hadronic t hypothesis, the t candidate consists of the sum of four-momenta of two ak4 jets,

$$\chi_{t,\text{lep}}^2 = \frac{(m_{t,\text{lep}} - \bar{m}_{t,\text{lep}})^2}{\sigma_{t,\text{lep}}^2} + \frac{(\Delta R_{t,W} - \pi)^2}{\sigma_{\Delta R}^2} + \frac{(1 - p_{T,W}/p_{T,t})^2}{\sigma_{p_{T,\text{ratio}}}^2}, \quad (6.4)$$

$$\chi_{t,\text{had}}^2 = \frac{(m_{t,\text{had}} - \bar{m}_{t,\text{had}})^2}{\sigma_{t,\text{had}}^2} + \frac{(\Delta R_{t,W} - \pi)^2}{\sigma_{\Delta R}^2} + \frac{(1 - p_{T,W}/p_{T,t})^2}{\sigma_{p_{T,\text{ratio}}}^2}. \quad (6.5)$$

Both equations (6.4) and (6.5) have three free parameters.

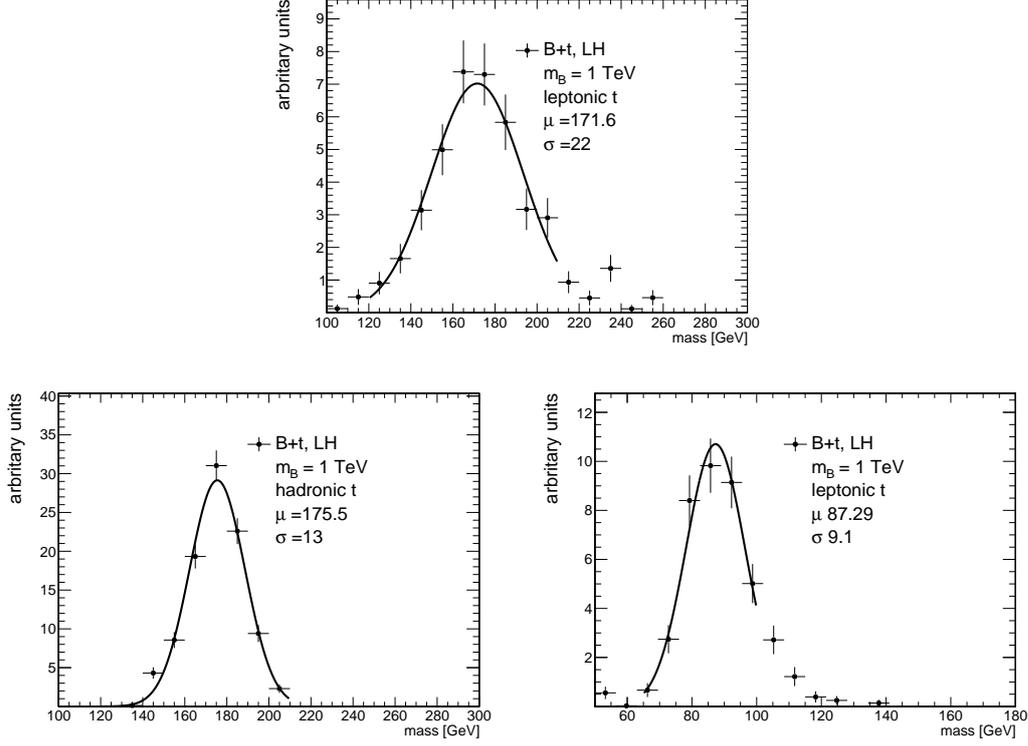


Figure 6.2.: Gaussian fits to simulated mass distributions of t quarks and the hadronically decaying W boson from the VLQ decay. The results are provided for the B+t production mode with a mass of 1 TeV, left-handed VLQ couplings, and a relative width of 1%. The reconstructed and matched leptonic t mass (upper), hadronic t mass (lower left) and leptonic W mass (lower right) are shown. The ak4 jets associated with the t quarks and W boson are matched to generator information with  $\Delta R < 0.4$  between the simulated and reconstructed t and W candidates.

**1 ak4 jet** The following equation  $\chi^2_{1 \text{ ak4}}$  for the hypotheses with one ak4 jet with 1 degree of freedom is used

$$\chi^2_{1 \text{ ak4}} = \frac{(\Delta R_{t,W} - \pi)^2}{\sigma_{\Delta R}^2} + \frac{(1 - p_{T,W}/p_{T,t})^2}{\sigma_{p_{T,\text{ratio}}}^2}. \quad (6.6)$$

It only employs the back to back topology.

**Expected values and deviations** The expected values and standard deviations that enter the  $\chi^2$  functions are obtained from generator information on the particle level matched to reconstructed jets for t quarks and W bosons in simulated signal events.

Table 6.1.: Mean values and resolution for the hadronic and leptonic t mass and the hadronic W boson as obtained from the fits from reconstructed ak4 jets matched to generator information. The values for the distance between the t and W candidates is taken from generator information and the variation is tested for values between 0.1 and 2. The values shown are used in the  $\chi^2$  functions to select the hypothesis.

$\overline{m}_{\text{top,had}}$	170.1 GeV	$\sigma_{\text{top,had}}$	14.29 GeV
$\overline{m}_{\text{top,lep}}$	172.6 GeV	$\sigma_{\text{top,lep}}$	19.1 GeV
$\overline{m}_{\text{W,had}}$	85.5 GeV	$\sigma_{\text{W,had}}$	8.7 GeV
$\sigma_{\Delta R}$	0.2	$\sigma_{p_{T,\text{ratio}}}$	0.8 GeV

The matching selects the hypothesis closest to the generated event by requiring a  $\Delta R < 0.4$  between the reconstructed and generated quantities for the simulated t quark and W boson. The combination of ak4 jets is chosen that matches both the candidates and takes into account the  $p_T$  between the t and W. In Fig. 6.2 the distribution of the reconstructed mass is provided for the B quark with a mass of 1 TeV, left-handed VLQ coupling, and a relative width of 1%. The same figure also presents the distributions of the reconstructed and matched leptonic t mass (upper), hadronic t mass (lower left), and W mass (lower right) which are used in the  $\chi^2$  functions. The quantities are extracted from Gaussian fits to these matched events. For the expected values and their standard deviations in the  $\chi^2$  functions the average over all available VLQ simulated samples is used ( given in Tab. 6.1). The values and deviations are independent of the mass or coupling of the VLQ. For the angular distance and the  $p_T$  ratio between the t quark and W boson in the  $\chi^2$  functions the expected values  $\pi$  and 1 are used. Their standard deviations are chosen by varying the uncertainty between 0.1 and 2 (GeV). The best expected cross section limits at 95% CL in simulation are achieved by choosing  $\sigma_{\Delta R} = 0.2$  and  $\sigma_{p_{T,\text{ratio}}} = 0.8$  GeV.

### 6.1.3. Simulated VLQ reconstruction

The results of the algorithm described in the previous sections are studied in signal simulation. The algorithm has to be able to reconstruct the SM t quark and W boson with high transverse momentum, as well as the invariant mass of the tW system, corresponding to the invariant mass of the VLQs.

In Fig. 6.3 the distribution of the reconstructed invariant masses are provided for signals with a leptonic t (upper left) and a hadronic t (upper right) t candidate. The distribution for the hadronically decaying W (lower left) candidates are also shown. These distributions display signal events from the B+t production mode with right-handed couplings, and B masses of 800, 1100, 1400, and 1700 GeV. The same figure also displays the distribution of the  $p_T$  of the leptonic W candidate (lower right). The events in the distributions are shown after the selection summarised in Tab. 6.3, and apart from differences in the acceptance, the shape of the invariant mass distributions for all shown signal masses is the same. The  $p_T$  distribution in Fig. 6.3 (lower right) depends on the mass of the VLQ and is shifted towards higher values for heavier signals. The peaks of the invariant masses are at the expected positions. The other production modes (B+b and  $X_{5/3}+t$ ) display the same behaviour.

The invariant mass of the tW system is calculated from the vector sum of the reconstructed t and W candidates' four-momenta and is denoted by  $m_{\text{reco}}$ . The  $m_{\text{reco}}$  distributions are used in the analysis for both the optimisation of the selection and the statistical evaluation of the search. It corresponds to the invariant mass of the B and  $X_{5/3}$ . In Figs. 6.4-6.5 the  $m_{\text{reco}}$  distribution is shown for both the  $\chi^2$  and the t tag reconstruction, for several VLQ masses and relative widths. In Fig. 6.4 the  $m_{\text{reco}}$  distributions are shown for the B+b (upper) and B+t (lower) production modes, for masses between 800 GeV and 1600 GeV with right-handed VLQ couplings, a relative width of 1%, for the  $\chi^2$  reconstruction (left) and the t tag reconstruction (right). The reconstructed  $m_{\text{reco}}$  distribution has its maximum around the simulated invariant VLQ mass. The width of the distributions is determined mostly by the experimental resolution of approximately 10% for the t tag reconstruction, and approximately 13% for the  $\chi^2$  reconstruction. The production in association with a b quark has a better resolution, while the production in association with a t has a higher signal efficiency. The coupling of the VLQ only has a minor influence on the shape of the  $m_{\text{reco}}$  distribution. The  $X_{5/3}+t$  production mode exhibits the same behaviour as the B+t production mode.

The  $m_{\text{reco}}$  distributions for widths of 1, 10, 20, and 30% for the B+b with a mass of 1200 GeV (upper) and B+t with a mass of 1400 GeV (lower) production modes are provided in Fig. 6.5 for left-handed VLQ couplings. The  $\chi^2$  reconstruction (left) has a wider resolution than the t reconstruction (right). Both peak at the expected value

for the shown signal events and the resolution widens for events with larger relative width. The  $X_{5/3}+t$  production mode resembles the  $B+t$  production mode in terms of resolution and acceptance. Signal events with other masses and relative widths behave in an analogous way.

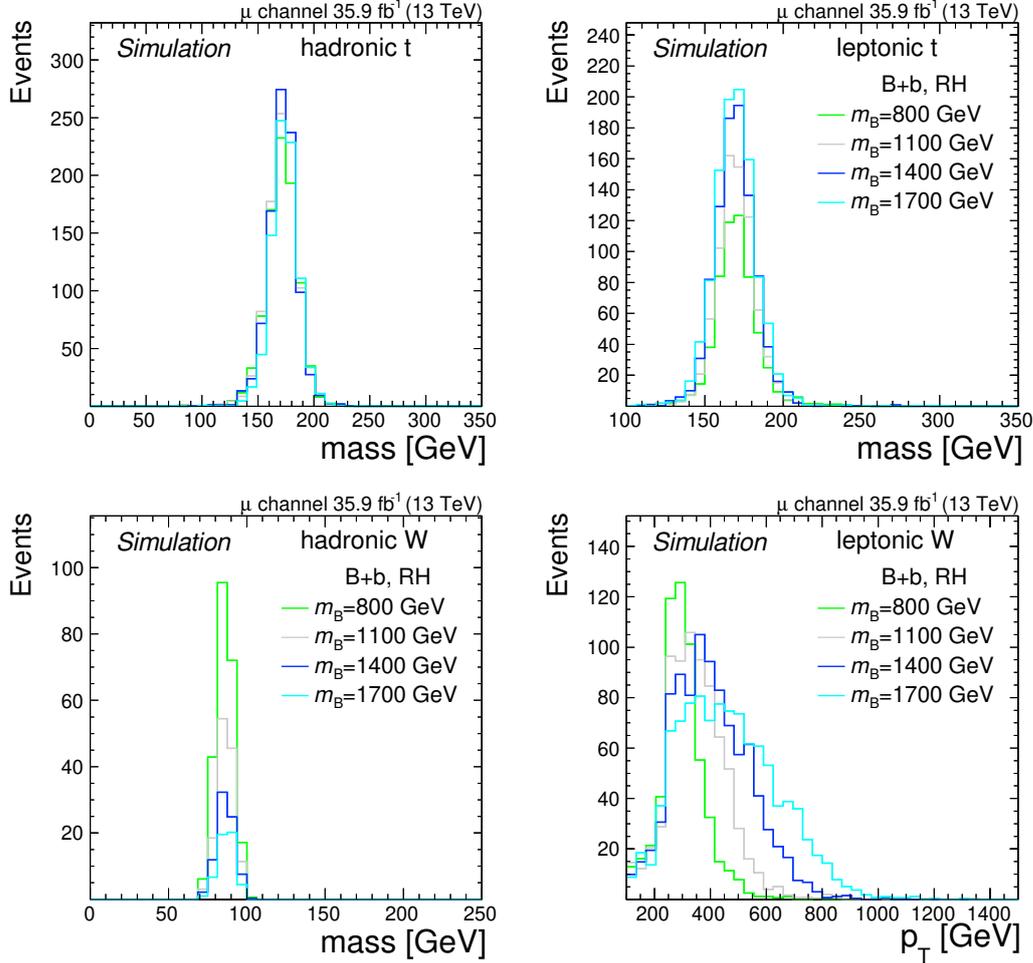


Figure 6.3.: Distribution of reconstructed t and W candidates for simulation signal events with masses of 800 GeV, 1100 GeV, 1400 GeV, 1700 GeV and right-handed VLQ couplings. The invariant mass of the hadronic (upper left) and leptonic (upper right) t, the hadronically decaying W (lower left) and the  $p_T$  of the leptonically decaying W (lower right) are shown. The invariant masses peak around the expected theory values and the leptonic W carries about half the momentum of the VLQ. The coloured lines in all shown distributions correspond each to the same VLQ signal.

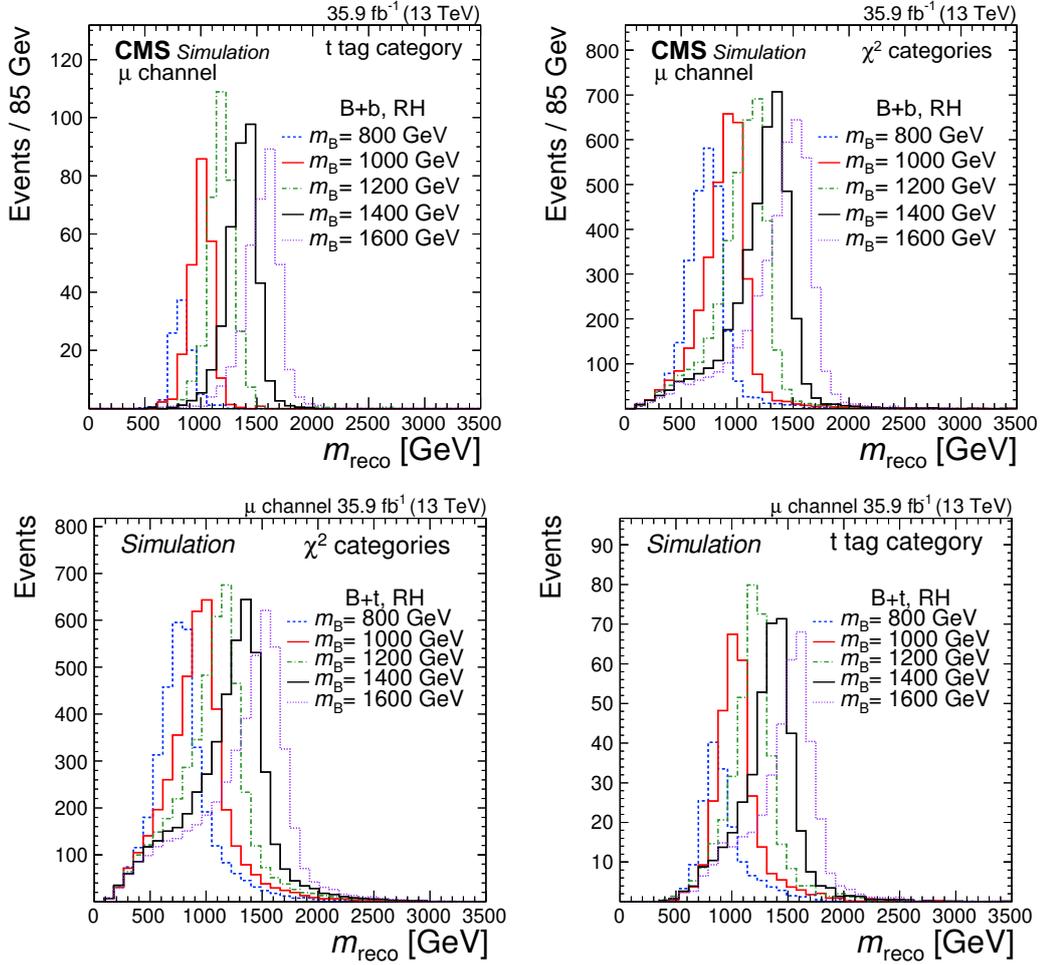


Figure 6.4.:  $m_{\text{reco}}$  distribution for signal events with masses of 800, 1000, 1200, 1400, 1600 GeV, and right-handed VLQ couplings in the muon channel. The B produced in association with a t quark (upper) and b quark (lower) are shown. The  $m_{\text{reco}}$  distribution for events with the  $\chi^2$  reconstruction (left) is wider compared with the t tag reconstruction (right) and shows a larger tail to lower masses. More events are reconstructed with the  $\chi^2$  method, and for both methods the signal events peak at the simulated B mass. The signal cross section is 1 pb for better visibility. The results for  $X_{5/3}$  are similar to the VLQ produced in association with a t quark. The differences between events with a muon and an electron is small. The upper distributions are taken from Ref. [30].

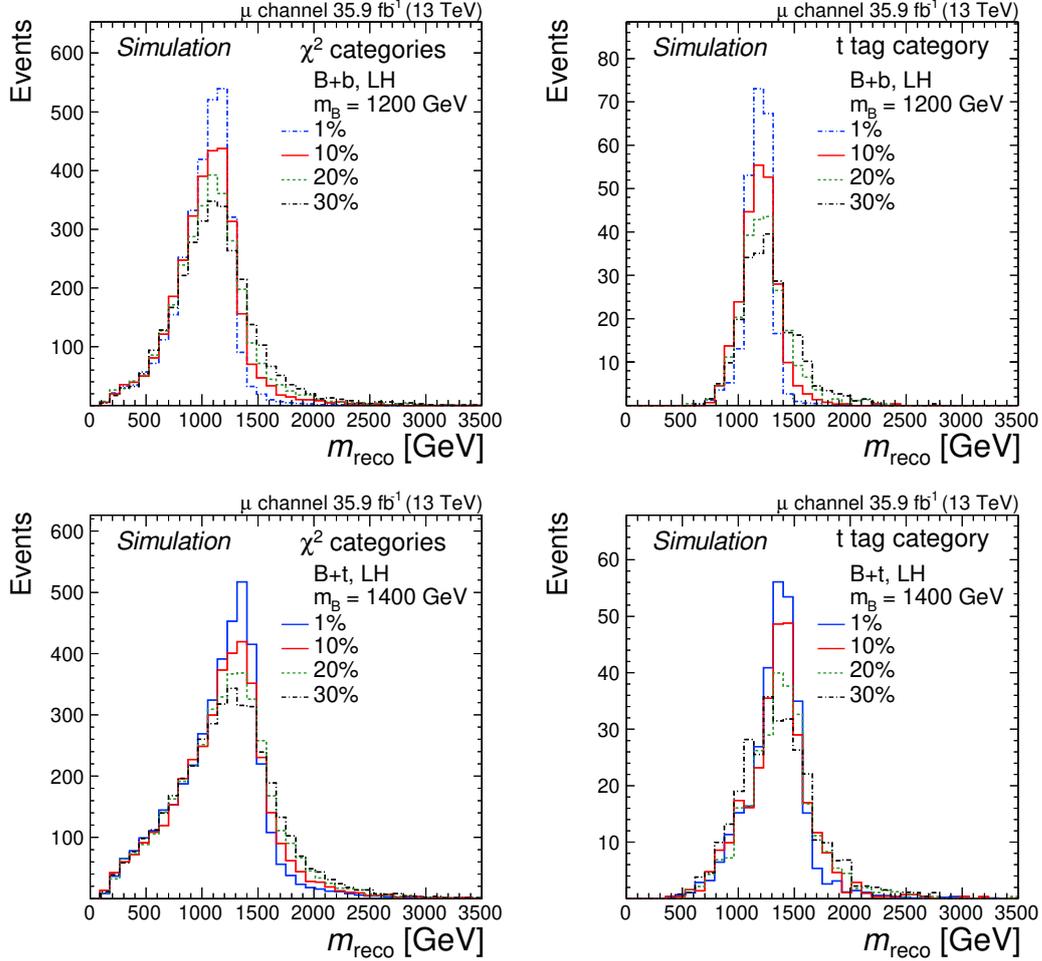


Figure 6.5.:  $m_{\text{reco}}$  distribution for signal events with a mass of 1200 GeV (upper) and 1400 GeV (lower) with widths of 1, 10, 20 and 30% for left-handed VLQ couplings in the muon channel. For corresponding widths scenarios, events reconstructed with a t tag (right) VLQ have a narrower width than events with the  $\chi^2$  reconstruction (left). The  $X_{5/3}$  behaves the same as the B produced in association with a t quark. The differences between the two lepton channels is small.

## 6.2. Event selection

The selection aims at removing events from SM processes while retaining as much of the signal as possible. This is essential, since some processes have a production cross section several orders of magnitude higher than the signal of interest. By selecting a small phase space of interest, the sensitivity to a signal is greatly enhanced. The event selection including the trigger, selection optimisation, and final selection is explained below

### 6.2.1. Trigger

The trigger is operated during data collection and has an important role in selecting interesting events depending on the activity in one or several sub-detector systems. As previously discussed in Sec. 3.2.5, not all events can be stored for later analysis. Therefore, a sensible first selection during the data collection is performed. The performance of several triggers is studied in a signal simulation for a single VLQs with a lepton and several jets in the final state. Since the  $p_T$  requirements for triggers based on jets are very high, they are discarded. Triggers based on the properties of leptons with much lower thresholds are used. In the electron channel, a trigger that considers both leptons and jets is used to enhance the acceptance. The trigger has the highest impact on the signal efficiency of all selection steps and therefore also on the sensitivity of this analysis. A trigger with forward jets or based on jets with lower thresholds might enhance the sensitivity of analysis further, although at present lepton triggers are the most important.

For the VLQ signal, jets are often found in the vicinity of the lepton. This occurs in situations where the t quark has a high Lorentz boost and the jets of the b quark are close to the lepton of the W boson decay products. Other sources for additional radiation found in the vicinity of the lepton can come from other quarks in the event, and final or initial state radiation. This non-isolated signature is challenging and leads to a low efficiency for isolated triggers, especially for masses around the TeV scale and above. To achieve the best sensitivity for the analysed VLQ signature, triggers based on single muons and electrons are studied both with and without online isolation.

Table 6.2.: Selection criteria on the denominator and numerator used to study the efficiency of the trigger as a function of the transverse momentum of the lepton as shown in Fig. 6.6. The selection applied to the denominator is also applied to the numerator. The additional quantities for the numerator correspond to the selection applied at the HLT level.

	single $\mu$ trigger	electron trigger	
		single e	e + jets
Denominator	$B \rightarrow tW \rightarrow l + \text{quarks}$ 2D isolation		ak4 jet $p_T > 185$ GeV
Numerator (HLT level)			$\mu p_T > 50$ GeV

Additionally, triggers that consider both a single electron or muon and jets are also investigated.

Triggers are compared in terms of their efficiency, calculated using

$$\epsilon = \frac{n_{num}}{m_{den}} \quad (6.7)$$

Here  $m_{den}$  denotes the events that pass the denominator selection and  $n_{num}$  the events that pass the numerator selection. The denominator selection requires the event to have a lepton that is a daughter of the B decay products. Additionally, the event has to pass the 2D isolation. Events counted for the numerator pass the selections of the denominator and a selection corresponding to one of the used online HLTs where the lepton and jet properties are calculated with lower precision. The requirements to calculate the numerator and denominator are summarised in Tab. 6.2.

In Fig. 6.6 the efficiency as a function of the lepton  $p_T$  for the muon trigger (upper left) and the electron triggers (upper and lower right) are shown for three B masses with left-handed VLQ couplings. The triggers efficiencies shown for the three VLQ masses do not have an online isolation requirement. The trigger efficiencies shown are independent of the masses of the B quark, and a very high efficiency for the interesting events with a high  $p_T$  lepton of about 95% is achieved. The trigger in the muon channel with isolation (lower left) is shown without online isolation requirements for comparison. The selection efficiency is highly correlated to the VLQ mass, and the number of additional events in comparison to the non-isolated trigger is negligible.

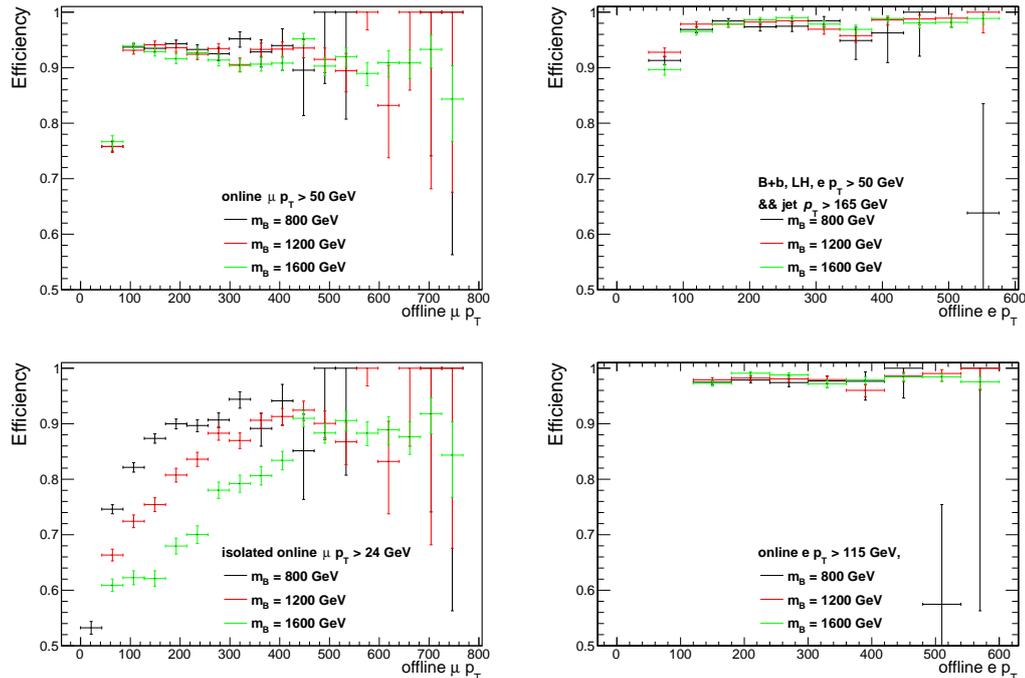


Figure 6.6.: Trigger efficiencies as a function of the transverse momentum of the lepton for three different B masses with left-handed VLQ couplings. The numerator and denominator are listed in Tab. 6.2. The trigger efficiency is independent of the transverse momentum of the lepton or the mass of the VLQ. The muon trigger (upper left) is based on an online muon  $p_T > 50$  GeV and shown as function of offline  $p_T$ . Two electron triggers are combined, one based on online electron  $p_T > 115$  GeV (upper right), as well as one based on electron  $p_T > 50$  GeV and an online ak4 jet with  $p_T > 165$  GeV (bottom right). Both are shown as function of the offline electron  $p_T$ . For comparison the trigger requiring isolation in the muon channel (bottom left) is shown, which displays a mass dependent drop in the signal selection efficiency. For the isolated trigger no offline isolation is required.

For the combined electron and jet trigger, along with an electron an ak4 jet with  $p_T > 185$  GeV is required in this analysis. The  $p_T$  requirement ensures that selected events are on the plateau of the trigger efficiency curve. The trigger based on a muon and a jet is not shown since the threshold for the muon in this trigger is very similar to the single muon trigger, and less than 1% of additional events are selected.

### 6.2.2. Selection optimisation

To obtain a selection with a high sensitivity to the signal, a basic selection is optimised by comparing expected asymptotic exclusion limits at 95% CLs for additional requirements on MC simulation. The sensitivity is enhanced by reducing the SM background through additional selections. The exclusion limits are calculated from the binned  $m_{\text{reco}}$  distribution where the binning is chosen such that the statistical uncertainty is smaller than 5% in any bin. Exclusion limits have the advantage that they do not depend on a production cross section but give independent results for each mass point while the full information from the  $m_{\text{reco}}$  distribution is used. The production cross section is not fixed and depends strongly on the mixing between SM and the new physics model, as well as the relative VLQ width. The amount of multijet events is reduced by several orders of magnitude by a loose first selection. The significance is used to cross check the results from the expected exclusion limits by using  $S/\sqrt{S+B}$  as the figure of merit where  $S$  is the expected signal and  $B$  the background in the  $m_{\text{reco}}$  distribution. Signal and background are counted in a window that contains 90% of the signal. The VLQ production cross section for a relative width of 10% is used.

The multijet background needs to be reduced since the cross section is several orders of magnitude larger than any other SM process, and with current MC methods it is difficult to simulate enough events to have a reliable background prediction with small uncertainties. Table 6.3 shows a summary of the first selection (pre-selection), where events passing fulfil:  $\vec{p}_{\text{T}}^{\text{miss}} > 40 \text{ GeV}$ ,  $H_{\text{T},1} > 100 \text{ GeV}$ ,  $p_{\text{T}} > 30 \text{ GeV}$  of the ak4 jet with the highest  $p_{\text{T}}$  (leading), the trigger, and the 2D isolation requirement (details in Sec. 4.2.3). The triggers include a lepton selection to avoid the trigger inefficiencies for low lepton transverse momenta. The criteria are based on the topology with either a muon or electron in the final state, requirements from the reconstruction algorithms, and an efficient multijet reduction. The characteristics of multijet events is exploited: the leptons are often soft and within hadronic decays, and the amount of  $\vec{p}_{\text{T}}^{\text{miss}}$  is small. The expected limits at 95% CL are calculated using only statistical and rate uncertainties on the simulated SM processes of 50% for W+jets, 100% for multijet, 50% for  $t\bar{t}$ , and 50% for single t quarks.

The cross section limits are shown in Fig. 6.7 for the muon channel as a function of the scanned variable:  $\vec{p}_{\text{T}}^{\text{miss}}$  (upper left),  $H_{\text{T},1}$  (upper right), leading ak4 jet  $p_{\text{T}}$  (lower

Table 6.3.: Summary of the selections applied on the data and simulation, split by lepton flavour. First, a loose selection (pre-selection) is applied to suppress most of the multijet background, and afterwards the additional selection steps applied are optimised as described in Sec.6.2.2.

Pre-selection	Muon	Electron
Lepton trigger	$\mu$ with $p_T > 55$ GeV	e with $p_T > 120$ GeV or with $p_T > 55$ GeV and ak4 jet with $p_T > 185$ GeV
ak4 jet 2D isolation $\vec{p}_T^{\text{miss}}$ $H_{T,1}$	1 with $p_T > 30$ GeV and $ \eta  < 2.4$ $p_{T,rel} > 25$    $\Delta R > 0.4$ 40 GeV 100 GeV	
Selection	Muon	Electron
$\vec{p}_T^{\text{miss}}$ $H_{T,1}$	50 GeV 250 GeV	60 GeV 290 GeV
ak4 jet ak8 jet	2 with $p_T > 50$ GeV and $ \eta  < 2.4$ at least 1 with $p_T > 175$ , $ \eta  < 2.4$	

left) and the number of ak4 jet (lower right). The variables are scanned for all possible selections and the best cross section limit is displayed as a function of the scanned variable for the B+b production mode, right-handed VLQ couplings, a relative mass of 1%, and masses of 800, 1000, 1200, and 1500 GeV. The highest sensitivity of the analysis corresponds to the smallest cross section limits. For the muon channel the  $H_{T,1}$  and the leading ak4 jet are mostly independent of the chosen selection, only showing fluctuations at very high values. For the  $\vec{p}_T^{\text{miss}}$  and number of jets, lower values are preferred. The selection with one jet was also tested, but no improvement was found in comparison to two jets. The selections were optimised for VLQ masses around 1 TeV but do not significantly affect the limit for lower or higher masses. The numerical values for the combinations of selections are compared and the final selection is found to have the highest sensitivity. The final selections applied on the events are shown as vertical lines. The results for the electron channel are shown in Fig. 6.8. The muon and electron channels are scanned independently and display a similar performance. For the  $\vec{p}_T^{\text{miss}}$ , slightly higher values are preferred in the electron channel, which is expected since the amount of multijet events is much larger in the electron channel. In general, the statistics in the electron channel are lower due to the higher thresholds.

For the expected significance, the most important selections are also scanned for the B+b production mode with right-handed VLQ couplings in Fig. 6.9. In the scan

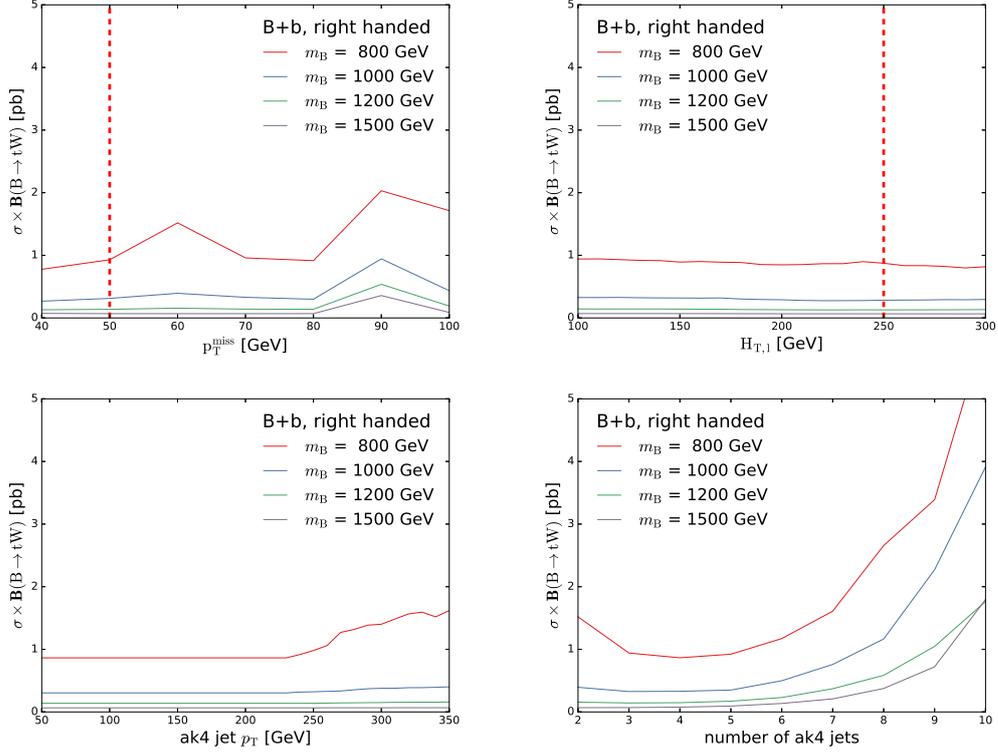


Figure 6.7.: Expected limits at 95% CLs as function of the scanned selection for the muon channel for the B+b production mode with masses of 800 to 1500 GeV, right-handed VLQ couplings, and a relative width of 1%. The  $\vec{p}_T^{\text{miss}}$  (upper left),  $H_{T,1}$  (upper right), ak4 jet  $p_T$  (lower left), and the number ak4 jets (lower right) are scanned, and the smallest expected limit at 95% CL is shown for each selection. The vertical dashed lines represent the selection with the overall highest sensitivity.

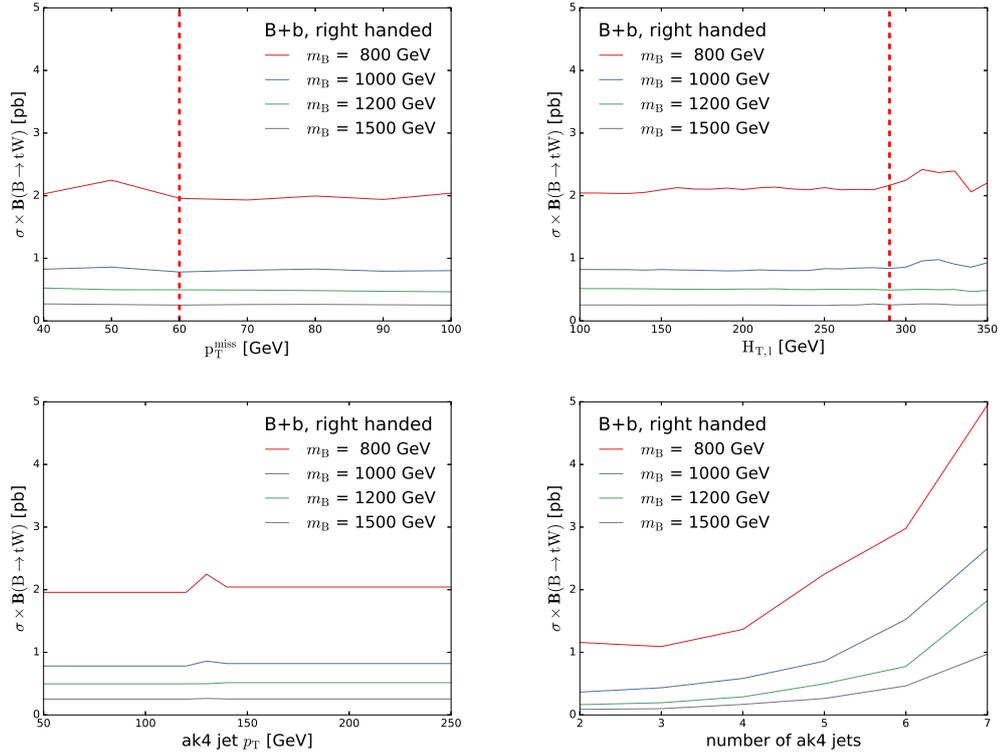


Figure 6.8.: Expected limits at 95%CLs as function of the scanned selection for the electron channel for the B+b production mode with masses of 800 to 1500 GeV, right-handed VLQ couplings, and a relative width of 1%. The  $\vec{p}_T^{\text{miss}}$  (upper left),  $H_{T,1}$  (upper right), ak4 jet  $p_T$  (lower left), and the number ak4 jets (lower right) are scanned, and the smallest expected limit at 95% CL is shown for each selection. The vertical dashed lines represent the selection with the overall highest sensitivity.

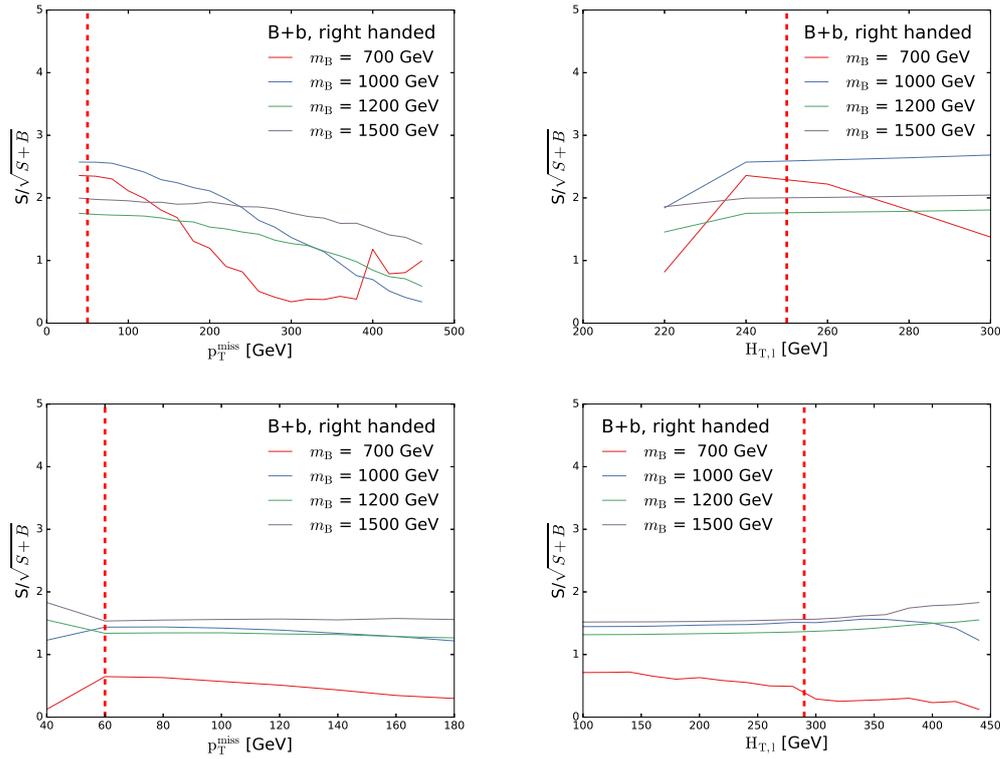


Figure 6.9.: Scanned  $S/\sqrt{S+B}$  as function of the  $\vec{p}_T^{\text{miss}}$  (left) and  $H_{T,1}$  (right), for the muon channel(upper) and electron channel(lower) for simulated background and signal events. The signal  $S$  and background  $B$  has been calculated in a windows in the  $m_{\text{reco}}$  distribution, which contains 90% of the signal. The scan is explained in detail in the main text.

a higher sensitivity in the analysis corresponds to larger values of the significance. The results are consistent with the results obtained for the exclusion limits. For both optimisation algorithms the best choice of jets is two ak4 jets, at least one ak4 jet with a  $p_T > 50$  GeV, and one ak8 jet. For the muon (electron) channel,  $H_{T,1} > 250(290)$  GeV and  $\vec{p}_T^{\text{miss}} > 50(60)$  GeV are found to be the values that yield the highest sensitivity for the selection. The higher values in the electron channel are the result of the higher misidentification rate of electrons in comparison to the muons, which results in a larger presence of multijet events and other SM backgrounds. The selection obtained in the optimisation process is summarised in Tab. 6.3 for the muon and electron channels.

### 6.2.3. Selected phase space

Good agreement between simulation and data is essential for validating the optimisation, reconstruction, and transfer factors calculated in Sec. 6.4. Most measurements at the LHC have shown very good agreement between data and simulated SM processes. The agreement between data and simulation is evaluated for events that meet the requirements discussed in Tab. 6.3. A small subset of distributions of the quantities employed in the optimisation and the reconstruction of  $m_{\text{reco}}$  are shown below.

In Fig. 6.10 the distributions of the muon (upper left) and electron (upper right)  $p_T$  are illustrated. Both distributions display a flat spectrum for  $p_T \lesssim 200$  GeV due to the  $H_{T,1}$  selection. The  $p_T$  distributions of the employed jet algorithms are also shown: ak4 jet in the muon channel (lower left) and ak8 jet in the electron channel (lower right). The distribution of  $\vec{p}_T^{\text{miss}}$  in the muon channel (upper left) and of  $H_{T,1}$  in the electron channel (upper right) are shown in Fig. 6.11. The plateau visible in the  $\vec{p}_T^{\text{miss}}$  distributions is also due to the selection applied on  $H_{T,1}$ . The distributions for the N-subjettiness ratio  $\tau_3/\tau_2$  for all ak8 jets in simulation and data in the muon (lower left) as well as in the electron (lower right) channel are shown in the same figure.

The dark grey bands in the ratios show the systemic uncertainties derived from the jet energy corrections (JEC), JER, PDF, scale, and renormalisation variations. The light grey bands show the statistical uncertainties. Within uncertainties the SM simulation describes the data well and no additional scale factors for simulation required. The sensitivity of the shown distributions for any VLQ signal are more than an order of

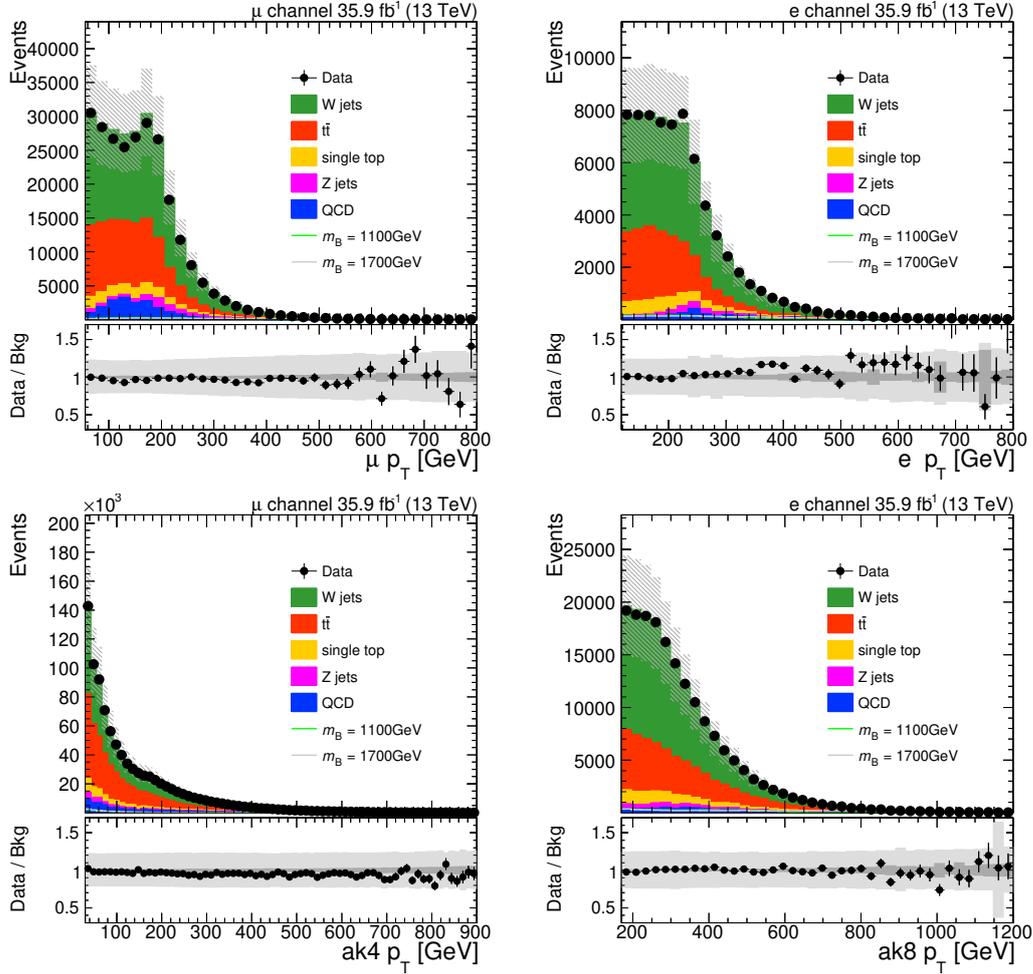


Figure 6.10.: Distributions of the muon (upper left) and electron (upper right)  $p_T$ , as well as for the  $p_T$  of  $ak_4$  jets for the muon channel (lower left) and for the  $p_T$  of  $ak_8$  jets in the electron channel (lower right). Events in the distributions pass the full selection. The plateau visible in both upper distributions is due to the selection applied on the  $H_{T,1}$ .

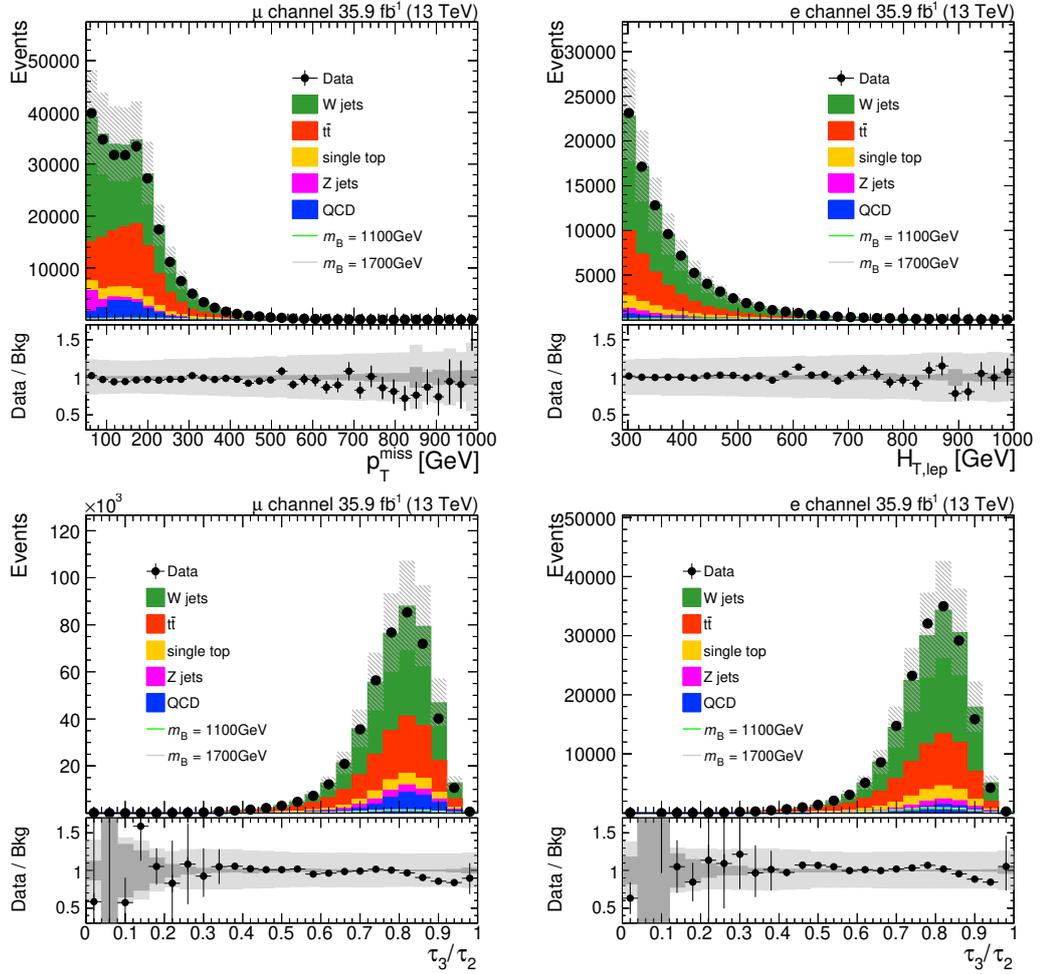


Figure 6.11.: Distributions of the  $\vec{p}_T^{\text{miss}}$  in the muon channel (left),  $H_{T,lep}$  in the electron channel (right) and of the N-subjettiness ratio  $\tau_3/\tau_2$  for all ak8 jets in the muon (lower left) and electron (lower right) channel. The full selection is applied on the events in the distributions.

magnitude smaller than the predicted cross sections. The presence of signal in these distributions would be negligible.

#### 6.2.4. Data and SM reconstruction

The agreement between data and simulation in the chosen phase space is crucial to ensure that the reconstruction algorithm is well-behaved and works as expected. The phase space after the selection consists mostly of events with a  $t$  and  $W$  and is used for this purpose. Furthermore, the reconstruction is designed for the signal masses above 1 TeV with decay products with high momentum and a large Lorentz boost. The distributions for the reconstructed  $t$  quark and  $W$  boson are shown and compared for different signal masses as well as between data and SM simulation. The invariant masses shown are the four-momenta sum of at least two  $ak_4$  jets or two subjets from an  $ak_8$  jet. Figure 6.12 illustrates the mass distribution for the simulated SM background prediction and data for the leptonic  $t$  (upper left), hadronic  $t$  (upper right), and hadronic  $W$  (lower left) candidates. The  $p_T$  of the leptonic  $W$  candidate (lower right) is also shown. The simulation and the data agree within uncertainties. The mass peaks in both the data and simulation at the expected  $t$  and  $W$  masses. For the  $p_T$  of the leptonic  $W$  candidate a peak at 250 GeV is visible due to the selection. In the SM the top quark pair production with a lepton in the final state has both  $t$  quarks decaying into jets or into jets or with a lepton; therefore, it is the most important SM background process. The simulated distribution of events from the  $W$ +jets process also has a peak at the  $t$  mass, because the reconstruction was optimised for high VLQ signals. In the reconstructed VLQ mass no such shift is visible, so this has no influence on the sensitivity of the analysis. The reconstructed  $t$  quarks and  $W$  bosons behave as expected for data as well as the signal and SM simulation. The  $m_{\text{reco}}$  distribution for the SM background processes is provided in Fig. 6.13 for the  $\chi^2$  reconstruction (left) and the  $t$  tag reconstruction (right) after the selection is applied. It is further divided by lepton flavour, for events with a muon (upper) or electron (lower). The data and simulation agree within uncertainties. The light and dark grey bands represent the systematic and statistical uncertainties of the simulation, respectively. The total uncertainties are in the range of 20 to 50%. Since the simulation yields very large uncertainties, a data-driven method that reduces the uncertainty drastically is used for the final evaluation and is detailed in section 6.4.

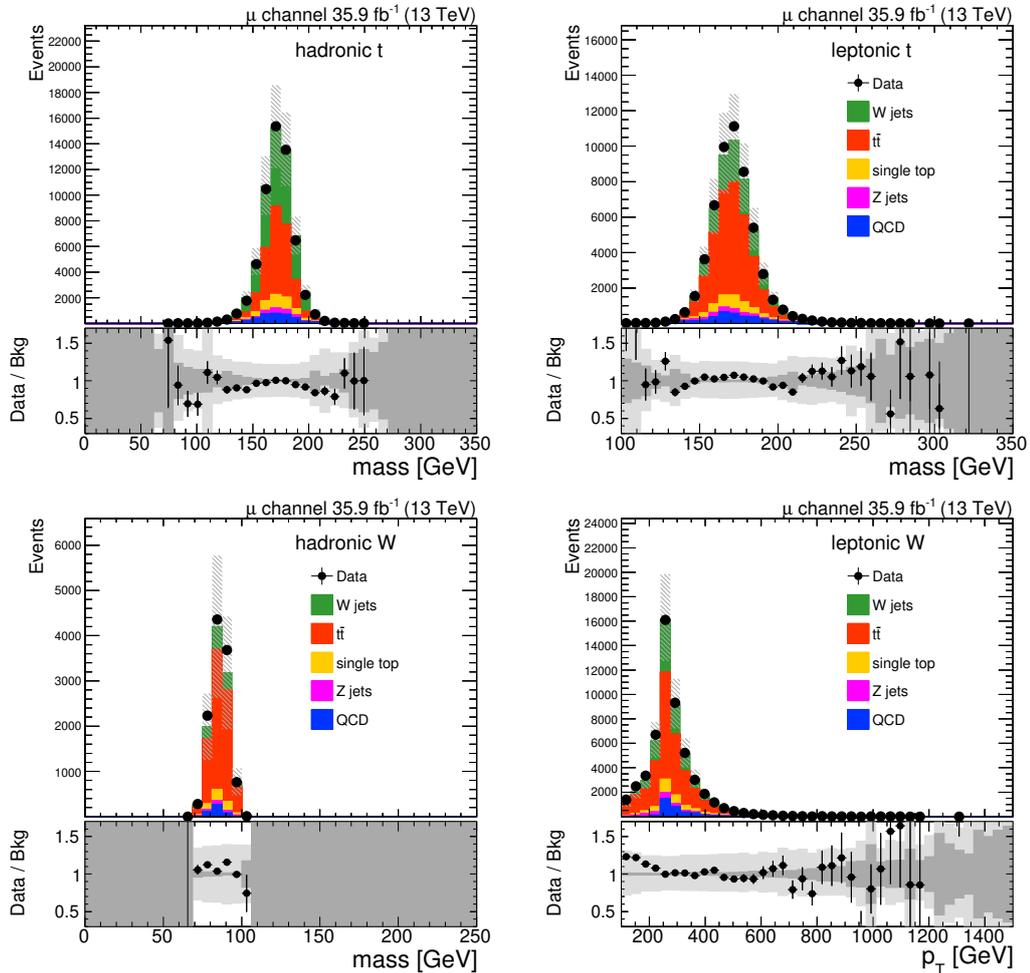


Figure 6.12.: Distribution of reconstructed  $t$  and  $W$  candidates in SM simulation and data for the  $\chi^2$  reconstruction. The invariant mass distributions of the hadronic (upper right) and leptonic (upper left)  $t$ , the hadronically decaying  $W$  (lower left) and the  $p_T$  of the leptonically decaying  $W$  (lower right) are in the muon channel. The invariant masses peak around the expected theory values. The simulation agrees with the data within the uncertainties.

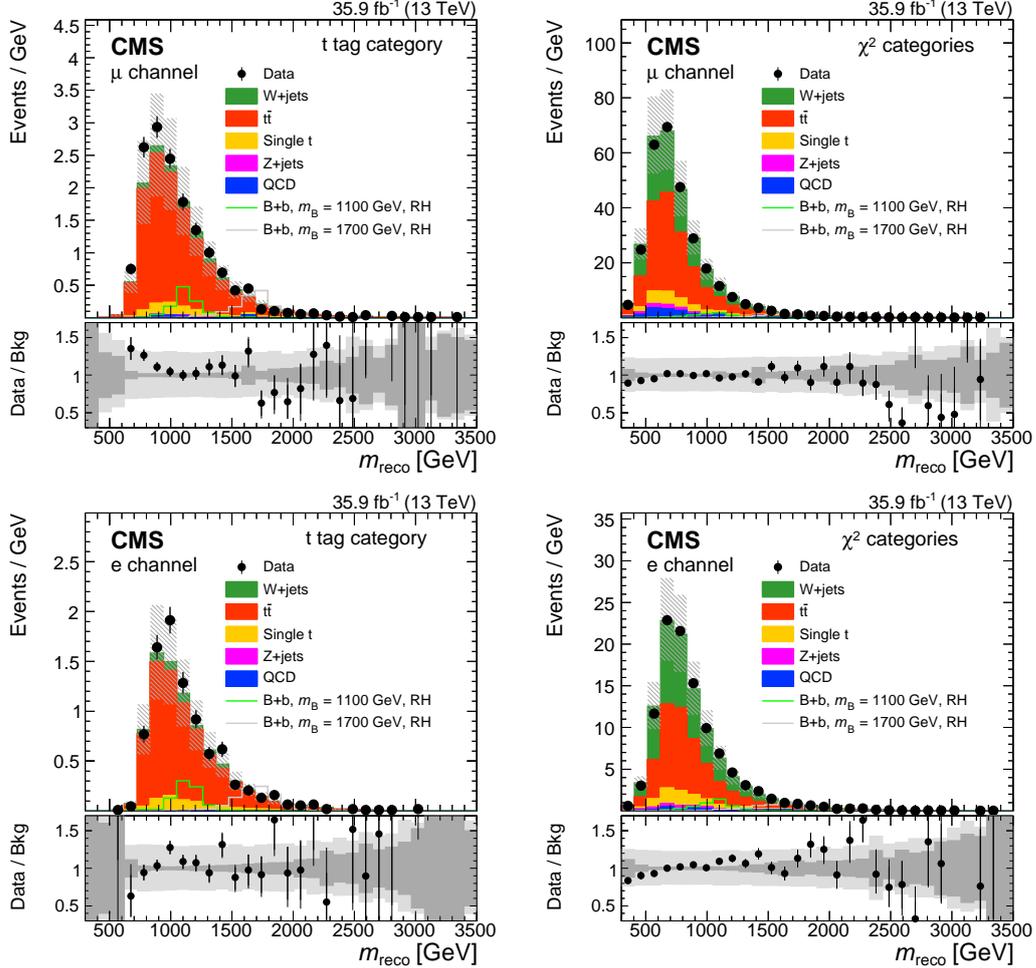


Figure 6.13.:  $m_{\text{reco}}$  distribution for the  $\chi^2$  (left) and  $t$  tag (right) reconstruction in the muon (upper) and electron (lower) channel for SM simulation and data after the selection. The data and simulation agree within the uncertainties. The light and dark grey bands show the systematic and statistical uncertainties of the simulation. The total uncertainties are in the range of 20% to 50%. The distributions are taken from Ref. [30].

### 6.3. Categorisation

In many analyses, events are split into several mutually-exclusive categories to maximise the sensitivity by exploiting additional information available in the event. These categories have higher relative signal fractions or lower uncertainties in comparison with the combined distributions. In this analysis the categories are chosen based on the presence of jets identified as hadronically decaying W bosons, t, b quarks, as well as based on the lepton flavour. The most sensitive categories here are those with a W tag or t tag. The categorisation scheme is presented below, and the properties of the categories are discussed. In addition, SM simulation is compared to data while the sensitivity of a single analysis category is studied by comparing expected exclusion limits at 95% CL.

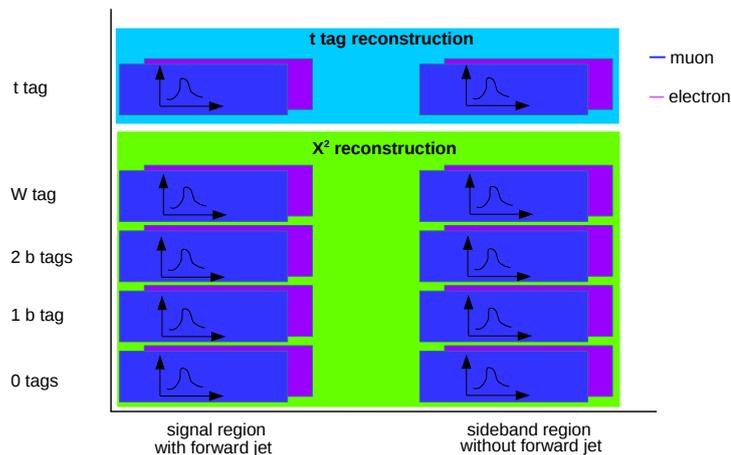


Figure 6.14.: Categories used for SR and SB used in the statistical evaluation. The muon (electron) channel is represented in dark blue (purple). The reconstruction with and without a t tag are shown in light blue and green.

Figure 6.14 illustrates the categorisation scheme including SRs and SBs (as explained in Sec. 6.4). The categorisation is based on the lepton flavour (muon or electron) and the number of t/W/b tags in an event. Based on the jet tagging, events with a t tag are grouped into one category (t tag) or, if no t tag is found, events with a W tag are collected in a second category (W tag). If neither a t nor a W tag is present, events are categorised depending on the b tag multiplicity. In total three type of b

tag categories: at least 2 b tags, exactly 1 b tag, and events with no tag (0 b tag). These five categories based on the jet tagging are subdivided into muon and electron flavours, resulting in a total of 10 categories.

Both the t tag as well as the W tag categories target signal events with a large Lorentz boost where either the decay products of the t quark or the W boson are merged into one ak8 jet. In the categories with b tags, VLQ mostly have a significantly lower Lorentz boost. In Fig. 6.15 the  $m_{\text{reco}}$  distributions in data for the electron channel, along with background predictions from MC simulation, are shown for events with a tagged jet. The SM background in the t tag (upper left) and 2 b tag (lower left) categories are dominated by  $t\bar{t}$  events. The categories with a W tag (upper right) or 1 b tag (lower right) have a higher W+jets contribution. The SM background in the 0 b tag category is dominated by the W+jets process. Apart from these two SM processes, processes with single t quarks also make a sizeable contribution. The best mass resolution for the  $\chi^2$  reconstruction is found in the 2 b tag category. Simulation and data agree within uncertainties, while the systematic uncertainty in the simulated backgrounds is of the order of 20%. The muon channel showcases the same features.

The impact of the categories on the sensitivity of the analysis is evaluated by comparing the expected exclusion limits at 95% CL on the product of branching fraction and cross section between the single categories and the combination of all categories. The expected limits are computed using the data-driven background estimation method explained in Sec. 6.4 and the systematic uncertainties explained in Sec. 7. For the B+b (left) and VLQ +t (right) production modes, the expected exclusion limits at 95% CL are illustrated in Fig. 6.16. The most sensitive categories contain a W or t tag, while the category with 0 b tags is the least sensitive. By including categories with b tags, the overall sensitivity of the expected CL is enhanced by approximately 20%. The categorisation itself yields an overall improvement of approximately 50%. The events are not split further based on the reconstruction modes (hadronic  $i$  and leptonic  $ii$  top), because the statistical uncertainties become dominant.

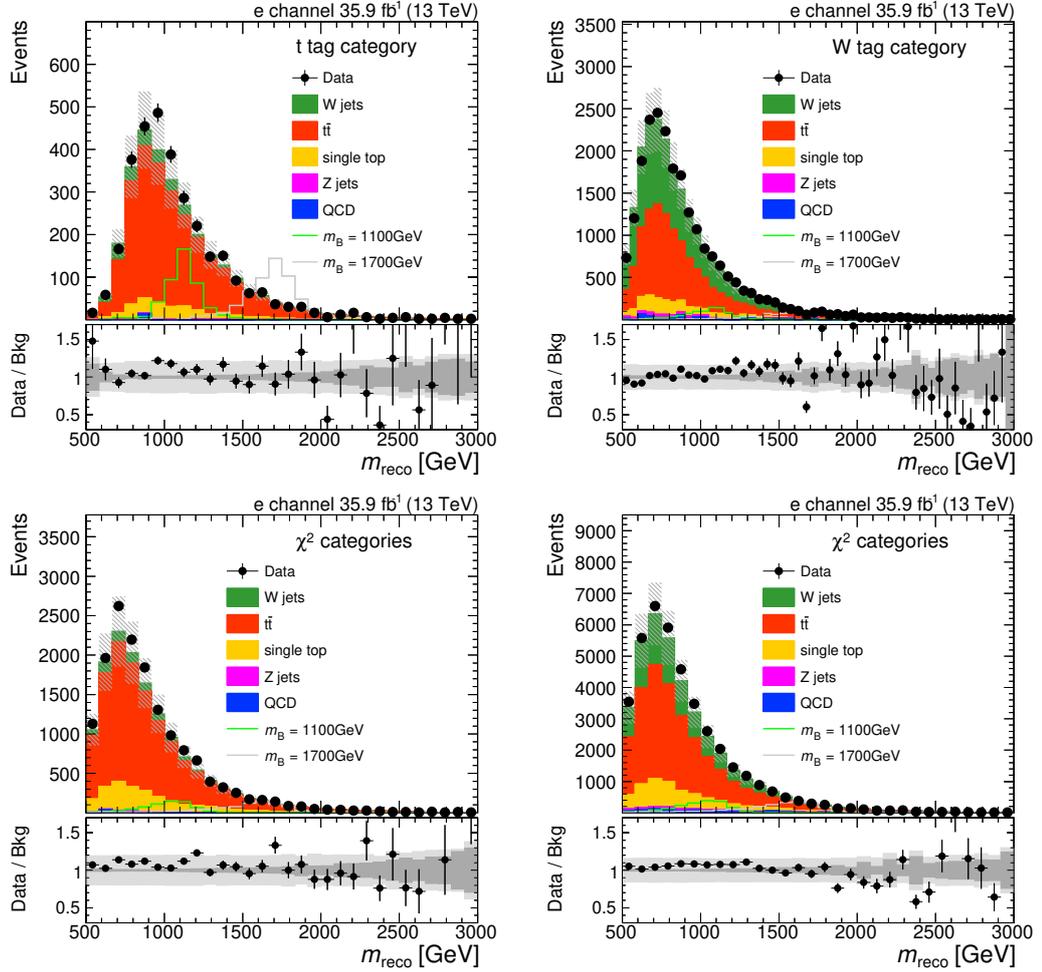


Figure 6.15.:  $m_{\text{reco}}$  distribution for the electron channel with background from simulation. The distribution for events with a t tag (upper left), a W tag (upper right), 2 b tags (lower left) and 1 b tag (lower right) are shown. Signal in the B+t production mode with right-handed VLQ couplings, a relative mass of 1%, and a cross section of 1 pb is displayed in the figure. The muon channel exhibits the same behaviour.

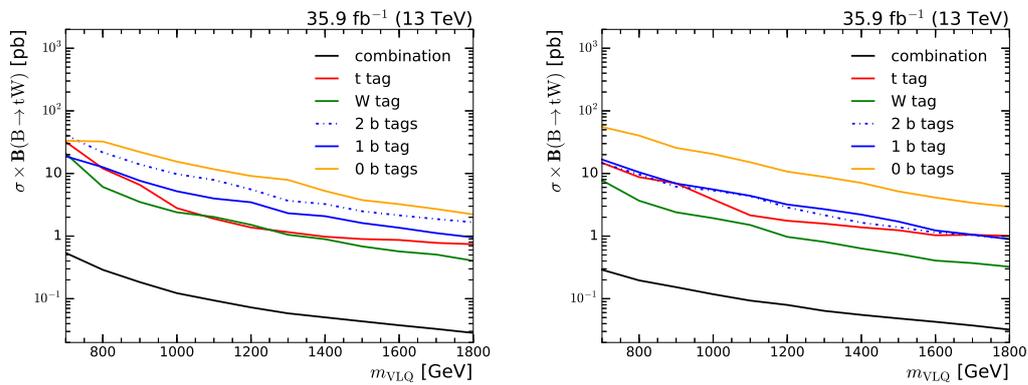


Figure 6.16.: Expected upper limits at 95% CL as function of the VLQ mass for the single categories and the combination in the B+b (left) and B+t (right) production modes with 1% relative widths and right-handed VLQ couplings. The results of the  $X_{5/3}+t$  production mode are similar to the B+t mode. The background prediction is taken from data (discussed in Sec. 6.4) and uncertainties are taken into account (explained in Sec. 7). The sensitivity of the analysis is improved by an order of magnitude by combining the several categories, compared to the results of a single category.

## 6.4. Signal region and background estimation

The forward jet accompanying the VLQ signal, which is not present in most SM processes, is used to construct both a region enriched with signal and a region dominated by SM backgrounds. The kinematics of the jet in the forward direction are independent of the categorisation and reconstruction of the invariant VLQ mass. Therefore, the kinematics for both regions are expected to be very similar, which is tested in simulation. Events with a forward jet are sorted into SRs, and those without a forward jet in SBs. The SBs are used to predict the background distributions in the corresponding SR categories. The method is validated both in simulation and with a subset of events in the SRs in data. By using a data-driven background estimate, the systematic and statistical uncertainties on the simulation of the background estimation are avoided.

### 6.4.1. Signal and sideband regions

The SR is defined using the quark jet in the forward direction, which is a feature of the single VLQ production at LO. The energy of this quark is high and its  $p_T$  usually low, making it difficult to distinguish them from pileup. In this analysis forward jets are defined as ak4 jets with  $2.4 < |\eta| < 4$  and  $p_T > 30$  GeV. In Fig. 6.17 distributions of the ak4 jet leading in  $|\eta|$  for events in the muon channel are provided for the  $\chi^2$  reconstruction (left) and the t tag reconstruction (right). The electron channel displays similar features. If no ak4 jet is found beyond  $|\eta| > 2.4$ , ak4 jets not employed in the  $\chi^2$  reconstruction or for the t tag reconstruction with at least  $\Delta R(\text{t tag, ak4 jet}) > 1.2$  are also shown. Most of the signal is located at  $|\eta| > 2.4$ . If no ak4 jet is found with  $|\eta| > 2.4$ , the event is used for the SM background estimate in the SB. For the SR with a W tag, an additional loose b tag is required, which improves the sensitivity in this region. On the other hand, the kinematics between SR and SB are changed, which increases the required correction factor. About 30% of the events in data have a forward jet, while about 80% of the signal events do. This is mostly independent of the category and a general feature of the SM predictions, with single t quark production being an exception. Single t quarks produced in association with a W boson and a quark in forward direction have LO Feynman diagrams almost identical to those of the VLQ signals. Due to the high mass of the VLQ signal, this

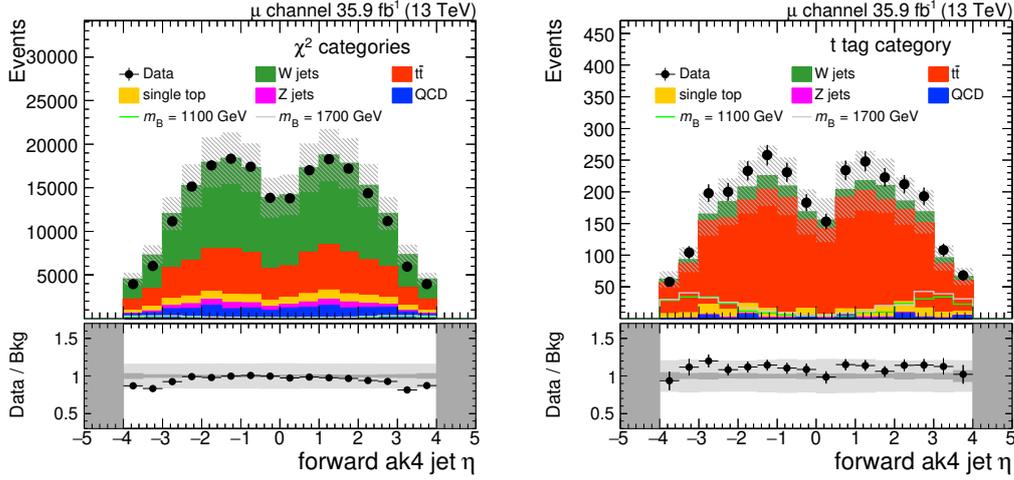


Figure 6.17.: Distribution of  $\eta$  for the most forward ak4 jet in the muon channel in simulation and data. Jets below  $|\eta| < 2.4$  have not been used for the reconstruction of  $m_{\text{reco}}$ . Both the  $\chi^2$  (left) and t tag (right) reconstructions are shown. The electron channel displays similar features.

SM process is only of minor importance, and its effect is taken into account in the background prediction by transfer parameters.

The event yields in the 10 SR categories are provided in Tab. 6.4 for simulation and data. In Fig. 6.18 the signal efficiency is shown as a function of the VLQ mass for the B+b (upper left), B+t (upper right), and  $X_{5/3}+t$  (lower) production modes. The denominator of the efficiency consists of signal events where the VLQ decays into a t quark and a W boson. The numerator is calculated from the subset of events in the SR. In this figure the efficiency is shown for the 1, 10, 20, and 30% relative widths for both the left-handed and right-handed VLQ couplings. The efficiency of all VLQs rises with their mass. Events with right-handed couplings and with a relative width of 1% of the VLQ have the highest efficiency of the B+b and  $X_{5/3}+t$  production modes, while for the B+t production mode the events with left-handed coupling have the highest efficiency. The efficiency is mostly independent of the width for relative widths of 10, 20, and 30% apart from the lowest mass samples. For the lowest VLQ masses the efficiency increases with the relative width, which indicates that the trigger is not optimal for the lowest mass points.

Table 6.4.: The event yields for SM processes in the SR are presented for the 10 categories used in the statistical evaluation. The event yields are higher for data than for simulation. Within an overall rate uncertainty of 15%, data and simulation agree.

Lepton	Process	t tag	W tag	2 b tags	1 b tag	0 b tags
$\mu$ channel	$t\bar{t}$ +jets	418	2497	2722	5755	3523
	W +jets	69	716	548	3226	22673
	Z +jets	2	33	23	169	1192
	Multijet	6	114	180	1482	2867
	single t	4	101	223	550	296
	data	552	3619	3834	10307	25523
e channel	$t\bar{t}$ +jets	238	825	939	1963	1185
	W +jets	42	315	257	1433	9259
	Z +jets	1	12	8	56	370
	Multijet	< 1	< 1	< 1	< 1	< 1
	single t	2	53	70	179	70
	data	301	1235	1376	3827	10143

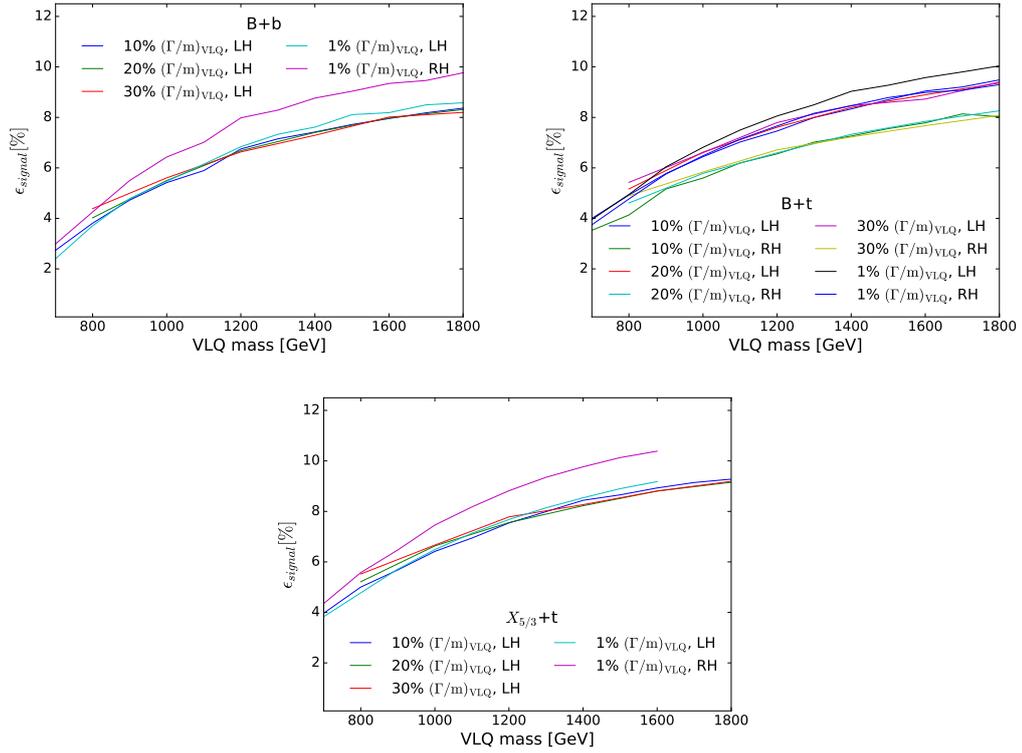


Figure 6.18.: The signal efficiency is shown as a function of the VLQ mass for the B+b (upper left), B+t (upper right) and  $X_{5/3}+t$ (lower) production modes. The denominator of the efficiency consists of signal events where the VLQ decays into a t quark and a W boson. The numerator is calculated from the subset of events in the SR.

### 6.4.2. Background estimation

The SB is defined as a phase space that is statistically independent of the SR and has very few signal events but very similar kinematics for SM processes. Therefore, the shape of the SB is used as the shape of the SM processes in the SR for the  $m_{\text{reco}}$  distribution. The normalisation for the background estimate is left free floating, assuming that the signal will appear as a peak in the  $m_{\text{reco}}$  distribution, with mild influence on the normalisation of the background prediction.

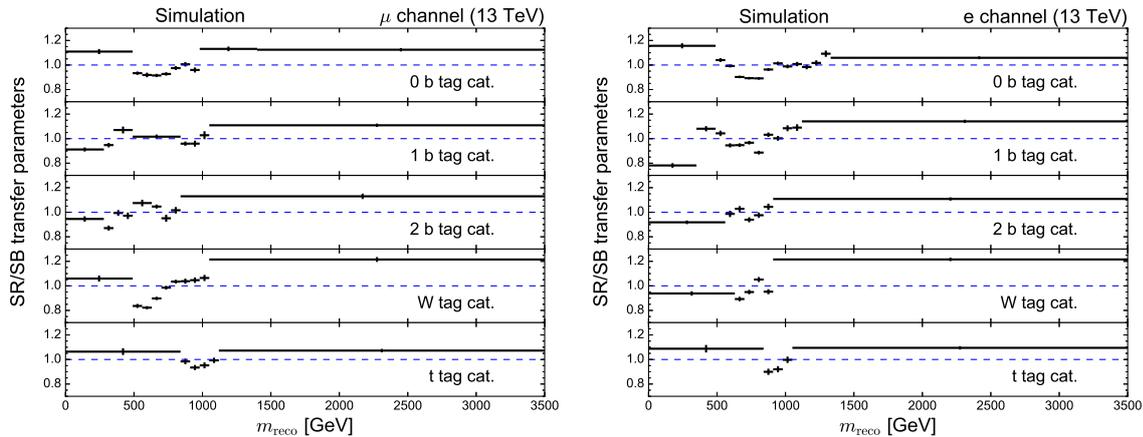


Figure 6.19.: Signal region to SB transfer parameters as function of  $m_{\text{reco}}$  in the muon (left) and electron (right) channel for the categories listed in Sec. 6.3. The transfer factor is the ratio of the SR and SB in simulation after rebinning the distributions to have less than a 5% statistical uncertainty in any bin. The statistical uncertainty is propagated to the scale factors and shown as error bars.

The SR and SB are compared in simulation for background composition and kinematical differences in the  $m_{\text{reco}}$  distribution. Apart from fluctuations, both regions only show residual differences, which are addressed by transfer parameters for the  $m_{\text{reco}}$  distribution derived from simulation. Figure 6.19 shows the transfer factors of the muon (left) and electron (right) channels. The transfer factors are calculated in each category by dividing the  $m_{\text{reco}}$  distribution in the SR by the corresponding distribution in the SB in simulation. The binning of the distributions is chosen such that the statistical uncertainty is smaller than 5%, and they are normalised to the same number of events. The statistical uncertainty is propagated and later added to the statistical uncertainty of the data-driven background estimate. The scale factors are generally within 10% of unity with no deviation larger than 20%.

The influence of the signal in the SB on the sensitivity of the analysis is negligible for both the upper exclusion limits at 95% CL and the normalisation. For the significance, the presence of signal in the SB is taken into account. Details are found in Appendix C, where the presence of signal is studied in simulated events and signal is injected into the simulation. The final background estimation is shown in Chapter 8.

### 6.4.3. Background validation

The background estimation method is validated by selecting events with a low  $\chi^2$  quantity (the region is hereafter called “validation region”), which leads to a small number of signal events. As with all other, the events in the validation region are sorted into a SR and SB. The background modelling is tested after applying the transfer parameters. The  $m_{\text{reco}}$  distributions in the different categories are used for a closure test as shown in Fig. 6.20 and 6.21.

The  $\chi^2$  quantity is based on the probability of obtaining the respective  $\chi^2$  given from Eq. 6.6 to 6.2. For events reconstructed with a t tag Eq. 6.2 is used. If the probability of the  $\chi^2$  quantity is  $p < 0.08$  for an event, it is sorted into the validation region.

The validation region is divided into categories as discussed in Sec. 6.3, and into a SR and SB as discussed in Sec. 6.4. The residual kinematic shape differences are corrected using the transfer parameters illustrated in Fig. 6.19. The background estimate is fitted using a maximum likelihood fit to obtain the normalisation. The results of the data-driven background estimate in the validation region are presented in Fig. 6.20 for the muon channel and in 6.21 for the electron channel for the categories: t tag (upper left), W tag (upper right), 2 b tag (middle left), 1 b tag (middle right) and 0 b tag (lower). The regions with and without a forward jet agree within the uncertainties and have a p-value of 80% for a combined least-squares test of all categories.

The sensitivity in the validation region, as studied in simulation, is about an order of magnitude smaller than that of the full phase space and well above theory predictions. The impact of the signal on the validation region is negligible, as detailed in Appendix C.

Using the validation region, the transfer parameters are verified by comparing them between data and simulation for the SR and SB. Within the propagated statistical uncertainties they agree. For the statistical evaluation of the transfer parameters a least-squares test is performed on the combined bins of all categories with a p-value of approximately 0.8.

The background estimation method works in simulation and in the validation region for both simulation and data.

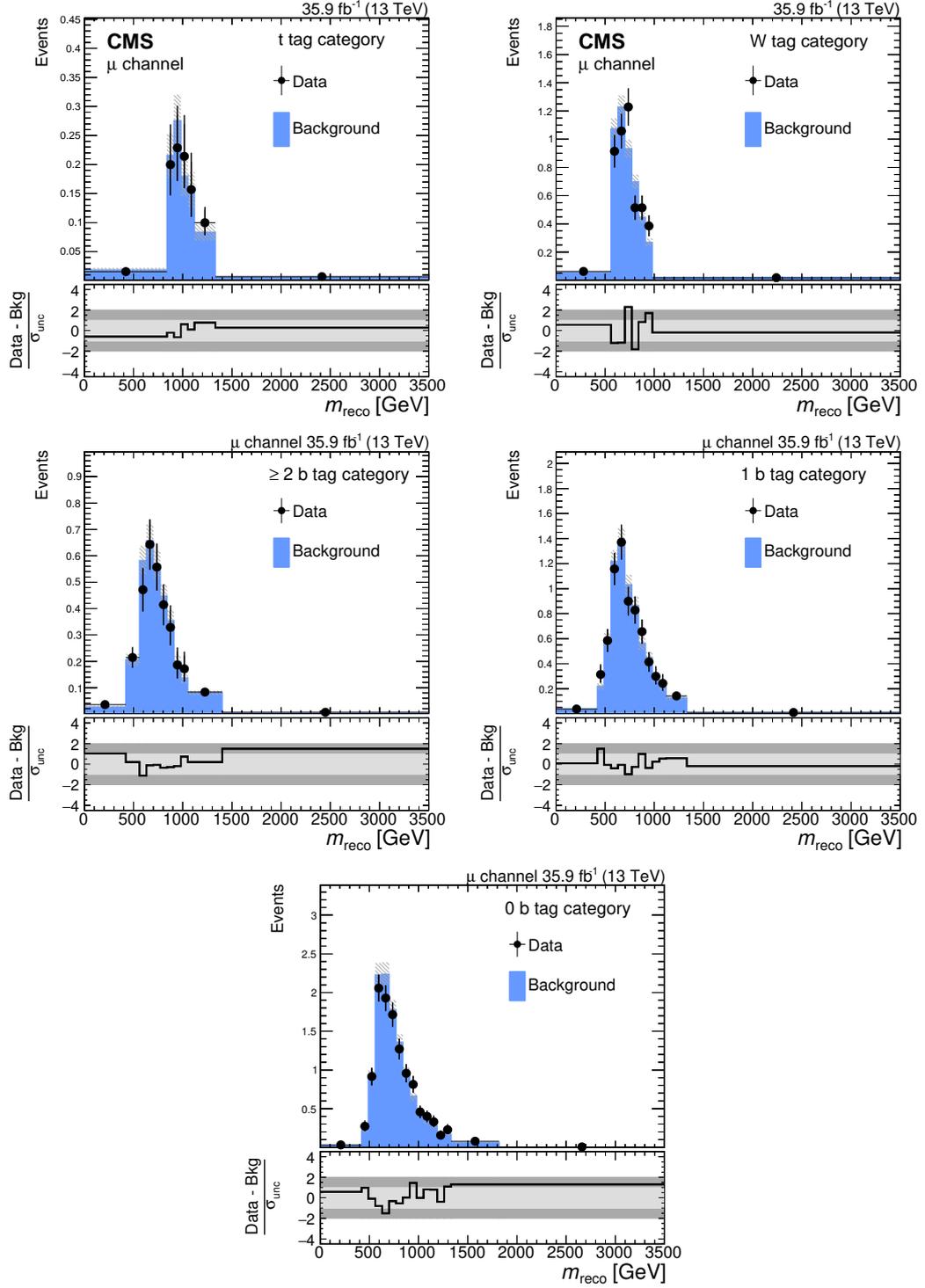


Figure 6.20.: Distributions of  $m_{\text{reco}}$  in the validation region of the categories in the muon channel. The validation region is a subset of the SR and SB and explained in the main text. The lower panels show the difference of data and background expectations in units of the total (stat. and sys.) uncertainty on the background estimate. The upper distributions and most of the caption are taken from Ref. [30]

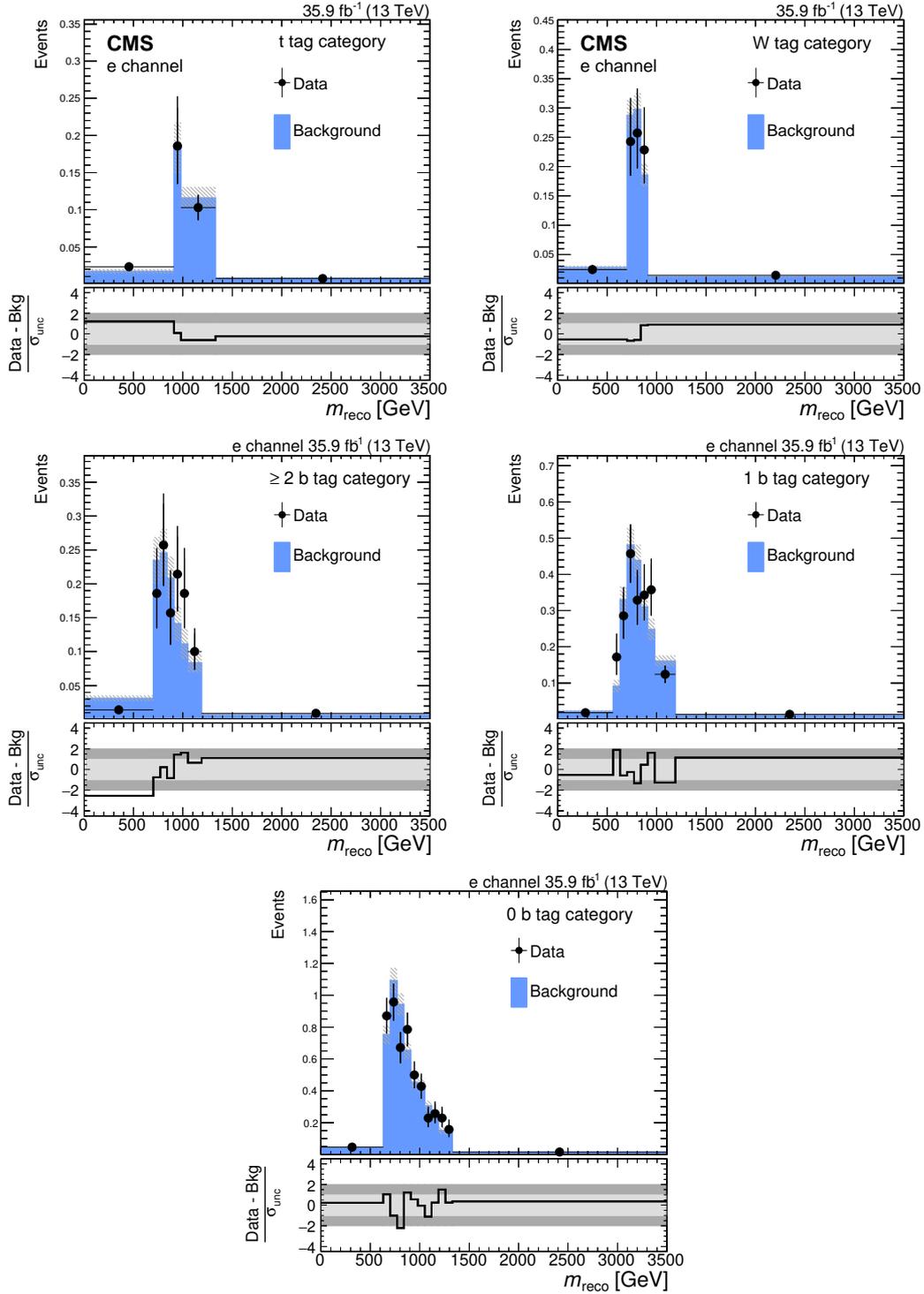


Figure 6.21.: Distributions of  $m_{\text{reco}}$  in the validation region of the categories in the electron channel. The validation region is a subset of the SR and SB and explained in the main text. The lower panels show the difference of data and background expectations in units of the total (stat. and sys.) uncertainty on the background estimate. The upper distributions and most of the caption are taken from Ref. [30]



## 7. Systematic uncertainties

The systematic effects considered affect both the shape as well as the normalisation of the  $m_{\text{reco}}$  distribution for the signal and the SM background predictions. Uncertainties taken into account include effects from the detector, theory predictions, and statistics. In this analysis, two distinct types of uncertainties are considered: uncertainties regarding the signal simulation and the data-driven SM background prediction. For the uncertainties, the muon and electron channels are assumed to be fully correlated.

The uncertainties regarding the simulated signal influence both the shape and normalisation. They are mostly derived from additional measurements or by comparing various theoretical predictions. The uncertainties considered are summarised below.

- The **luminosity** recorded with the CMS detector for the 2016 data taking period at  $\sqrt{s} = 13$  TeV is determined in an independent measurement with an uncertainty of 2.5% concerning the rate of events [74].
- Based on the total inelastic cross section the number of **pileup** interactions is matched to the instantaneous luminosity profile in the data with a minimum-bias cross section of 69.2 mb. The associated systematic uncertainty is obtained by varying the minimum-bias cross section by 4.6% [145].
- Correction factors for simulated events are applied for the **lepton trigger, identification, and isolation** efficiencies. The systematic uncertainty on the  $m_{\text{reco}}$  distribution is obtained by varying the  $\eta$ - and  $p_{\text{T}}$ -dependent correction factors within their uncertainties.

- **Jet energy correction and resolution:** correction factors on the jet energy are computed as functions of  $p_T$  and  $\eta$ . The uncertainty is estimated by simultaneously varying the factors within their uncertainties for ak4 and ak8 jets. The changes in ak4 jets are propagated to  $\vec{p}_T^{\text{miss}}$  [95].
- The differences in the **b tag** efficiencies and mistag rates between data and simulation are corrected by  $p_T$ -dependent scale factors applied to simulation. These are calculated on an event by event basis for light and heavy flavours. The scale factors for ak4 jets and subjets of ak8 jets are varied within their uncertainty. The heavy and light flavour jets are varied separately [146].
- The efficiency of the **t tag** has a rate uncertainty of 7% and  $-4\%$  [102].
- For the **W tag** two uncertainties are added in quadrature: the jet mass resolution and scale. The contribution of the jet mass resolution is derived from the differences of the hadronisation and shower models of PYTHIA and HERWIG [147]. This contribution depends on the  $p_T$  of the W boson, which is given by the mass of the VLQ. For a VLQ mass of 700 GeV the uncertainty is approximately 2%, and for a mass of 1800 GeV about 6%. The jet mass scale uncertainty was measured on the peak of the W boson and found to be 1%.
- The uncertainties originating from the choice of **PDF** are taken into account using a set of 100 replicas corresponding to the different PDF measurements from the NNPDF3.0 set. The uncertainties are calculated in accordance with the recommendations for the LHC in Ref. [127].
- **Factorisation  $\mu_f$  and renormalisation scale  $\mu_r$**  uncertainties are evaluated by varying the scales independently by factors of 0.5 and 2.

Table 7.1 lists the impact of the uncertainties on the signal. For the uncertainties of  $\mu_f$ ,  $\mu_r$ , and the PDFs only the change in acceptance and shape are taken into account. The uncertainty with the largest impact on the sensitivity from signal events is the determination of the jet energy scale. The effect of these uncertainties on the result is small in comparison to the uncertainties associated with the background prediction.

Table 7.1.: Uncertainties considered for simulated signal events in the B+b production mode ( $m_B = 900$  GeV) for right-handed VLQ couplings for the t tag and W tag categories. The uncertainties in the b tag categories are of comparable size to those in the W tag category. Text and numbers are from Ref. [30].

Uncertainty		t tag [%]	W tag [%]
W tagging	Rate	–	3.3
t tagging	Rate	$\begin{smallmatrix} +7 \\ -4 \end{smallmatrix}$	–
Luminosity	Rate	2.5	2.5
Pileup	Shape	1–3	0.2
Lepton reconstruction	Shape	2–3	2–3
b tagging	Shape	2.5	2.5
Jet energy scale	Shape	2–6	1–5
Jet energy resolution	Shape	1–2	1–2
PDF	Shape	2–3	0.5
$\mu_f$ and $\mu_r$	Shape	0.3	0.2

The uncertainties of the background prediction are one rate uncertainty per jet tag category (five in total) and shape uncertainties due to kinematic differences between the SR and SB. The rate uncertainty per jet tag category has an effect that is smaller than 4%. The prior normalisation of the background shapes is set to the event count in the corresponding SR category and the uncertainty is set to infinity (technically a large value). Kinematic differences between the SR and SB are taken into account through the transfer factors (as discussed in Sec. 6.4). Uncertainties associated with these kinematic differences are obtained from not applying the transfer factors and applying them twice. In Fig. 7.1 the  $m_{\text{reco}}$  distributions and their uncertainties derived from the transfer parameters are illustrated for the categories with W tag (upper) and t tag (lower) in the muon (left) and electron (right) channels. The uncertainty of the transfer parameters is in the order of 20% and larger than the statistical uncertainty. The uncertainties concerning the background prediction are considerably constrained during the statistical evaluation. The kinematic differences after the evaluation have an effect of 3–12% depending on the category. Because the SM background prediction is taken from data, only the statistical uncertainty, the uncertainty on the transfer parameters and the normalisation are taken into account. The main uncertainty is related to the kinematic differences between the SR and SB based on the transfer parameters.

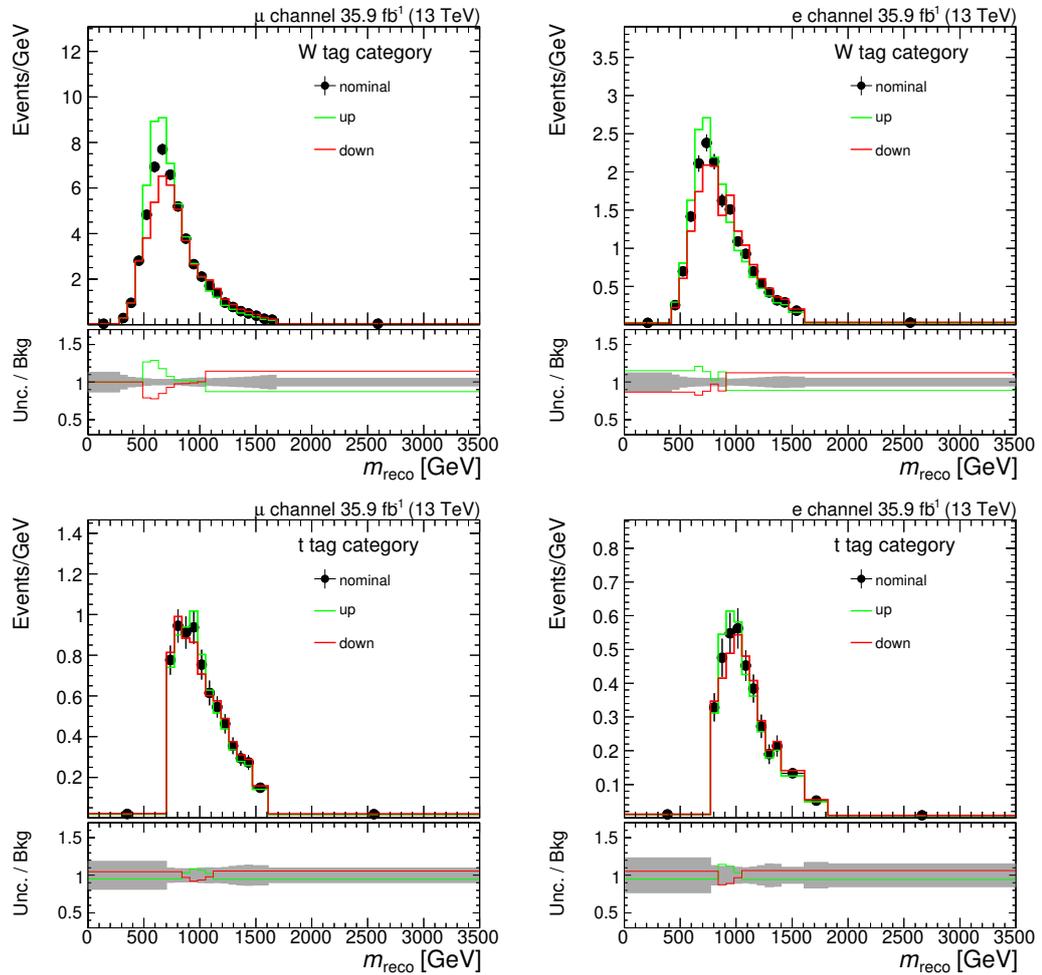


Figure 7.1.: Distributions of  $m_{\text{reco}}$  in the SB region in the muon (left) and electron (right) channels. The green and red lines show the upwards and downwards variations on the shape used as uncertainties for the W tag (upper) and t tag (lower) categories. The grey bands in the ratio display the statistical uncertainties on the background prediction.





## 8. Results and interpretation

The  $m_{\text{reco}}$  distributions are measured in the 10 SR categories with a data-driven SM background prediction from the SB. The SR and SB are defined by the presence or absence of an ak4 jet with  $|\eta| > 2.4$ . The distributions in the SR are found to have the highest discrimination power between SM processes and the B+b, B+t, and  $X_{5/3}+t$  production modes, where the VLQ is assumed to decay into a t quark and a W boson with a muon or electron and jets in the final state. The  $m_{\text{reco}}$  distributions in the SR, with the background prediction from the SB are used to calculate the standard score (defined later) for the B+b, B+t, and  $X_{5/3}+t$  production modes with masses between 0.7 and 2 TeV, right- and left-handed VLQ couplings, as well as relative widths of 1 to 30%. The standard score is used to classify the signal strength observed in the  $m_{\text{reco}}$  distributions. As a final step the upper limits at 95% CL on the product of production cross section and branching fraction for the signals are computed.

The binning of the  $m_{\text{reco}}$  distribution is chosen such that no bin has a statistical uncertainty higher than 20% for the background prediction, thereby avoiding large fluctuations. The resolution of the VLQ invariant mass is also taken into consideration for the binning and from simulation is found to be approximately 15%. Of the selected events, none exceeds the shown range.

In Fig. 8.1 and 8.2 the measured  $m_{\text{reco}}$  distributions are illustrated with the background prediction from the SB for the muon and electron channels. The normalisation of the background prediction is obtained from a maximum likelihood fit of the SB to the SR. The B+b production mode is shown for two masses with a cross section of 1 pb, left-handed VLQ couplings, and a relative width of 1% to illustrate the shape of a VLQ in the  $m_{\text{reco}}$  distribution. The five jet tag categories are shown for the two lepton flavours: t tag (upper left), W tag (upper right), 2 b tags (middle left), 1 b tag (middle right),

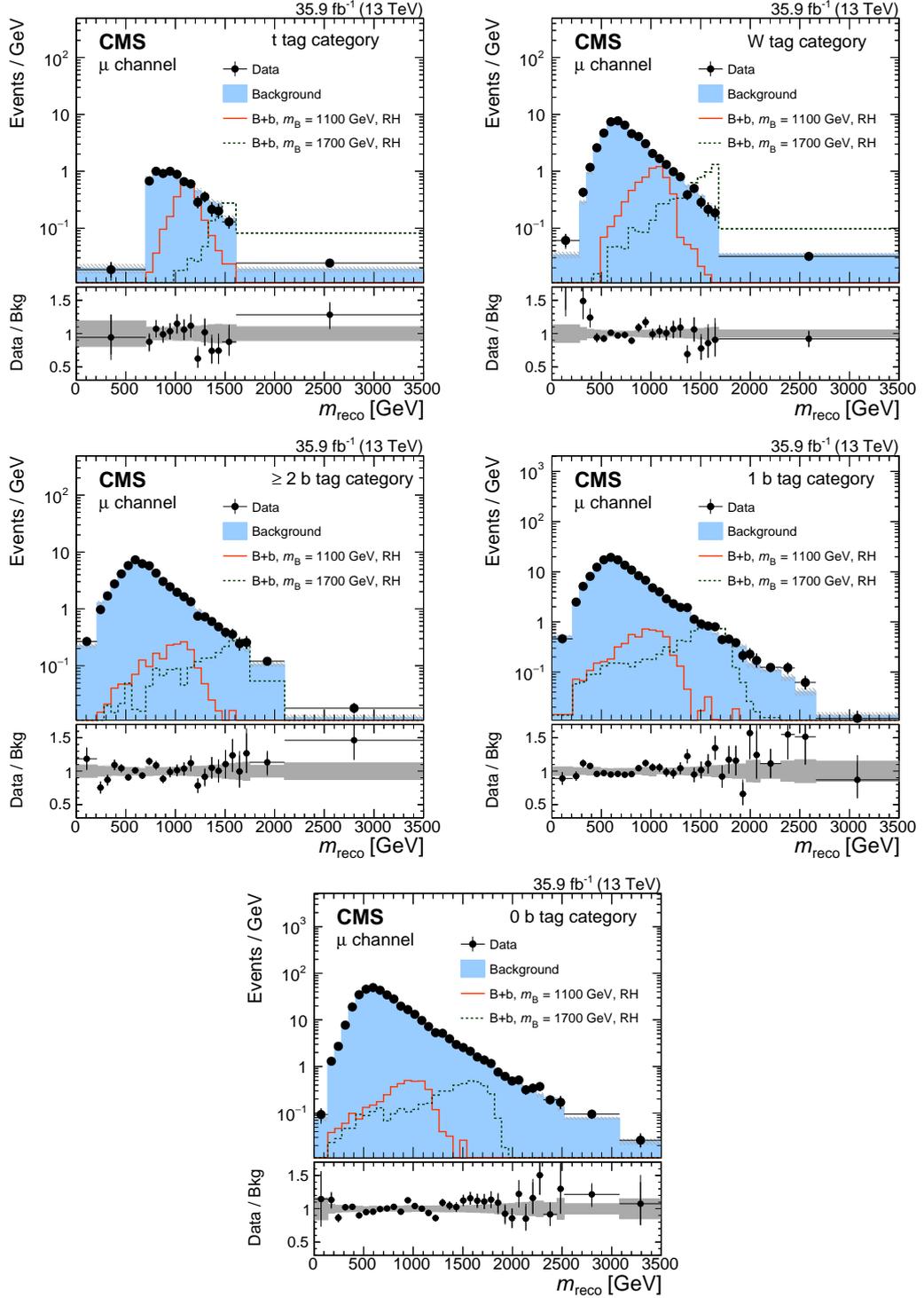


Figure 8.1.: Distributions of  $m_{\text{reco}}$  measured in the signal region for events with a jet in the forward direction with  $|\eta| > 2.4$  in the muon channel. Shown are the categories: t tag (upper left), W tag (upper right),  $\geq 2$  b tag (middle left), 1 b tag (middle right) and 0 b tag (lower). The background prediction is obtained from control regions as detailed in the main text. The distributions from two example signal samples for the B+b production mode with right-handed VLQ couplings with a cross section of 1 pb and a relative VLQ width of 1% are shown for illustration. Distributions and caption from Ref. [30].

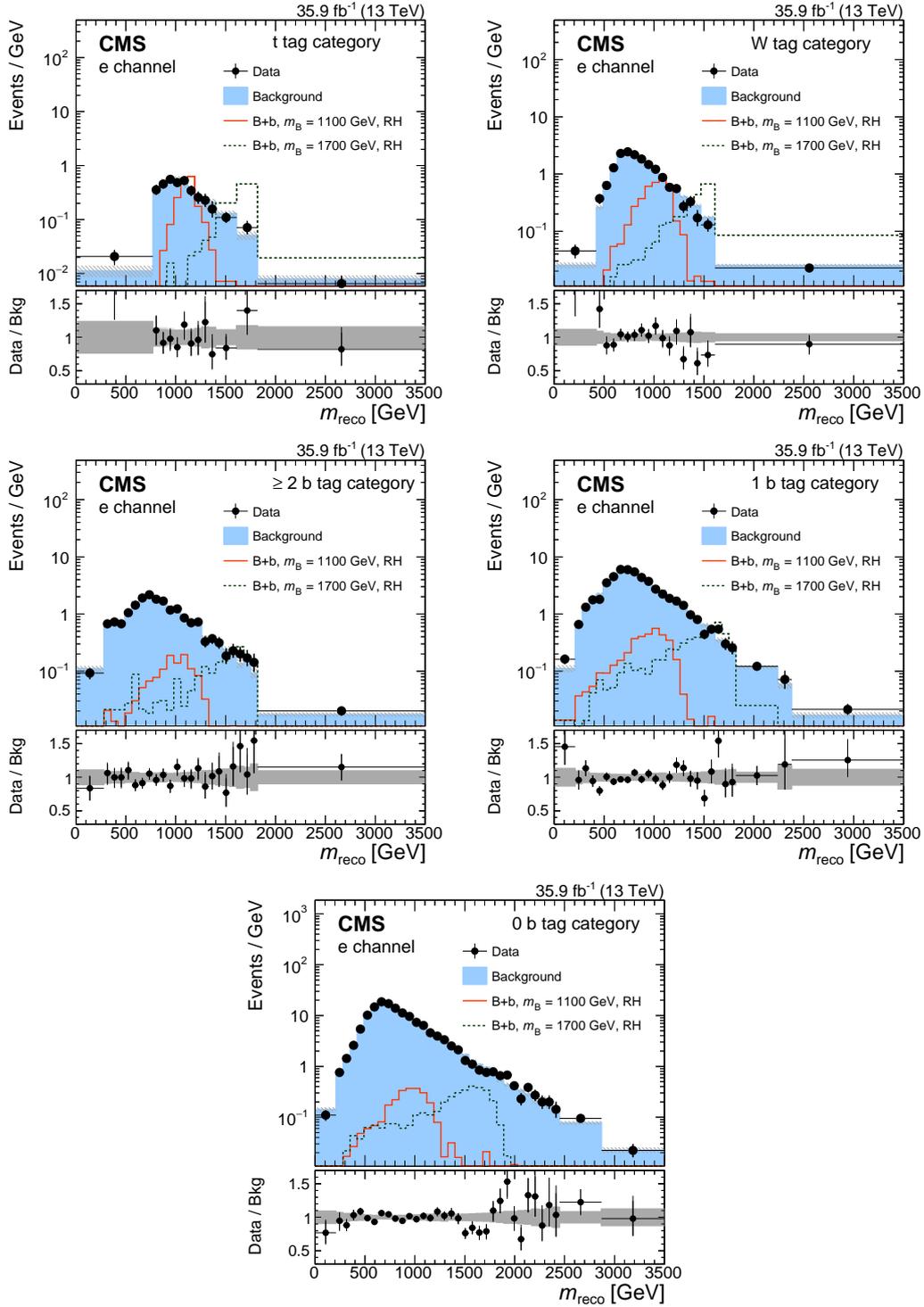


Figure 8.2.: Distributions of  $m_{\text{reco}}$  measured in the signal region for events with a jet in the forward direction with  $|\eta| > 2.4$  in the electron channel. Shown are the categories: t tag (upper left), W tag (upper right),  $\geq 2$  b tag (middle left), 1 b tag (middle right) and 0 b tag (lower). The background prediction is obtained from control regions as detailed in the main text. The distributions from two example signal samples for the B+b production mode with right-handed VLQ couplings with a cross section of 1 pb and a relative VLQ width of 1% are shown for illustration. Distributions and caption from Ref. [30].

Table 8.1.: Standard score derived from the likelihood ratio for production modes B+b and B+t with a relative mass width of 1% (10%), left- and right-handed VLQ couplings. The signal is subtracted from the SB as detailed in the main text. The numerical uncertainty is given in parentheses. No significant deviation from the SM prediction is observed.

TeV	B+b		B+t	
	1% RH	10% LH	1% RH	10% LH
0.8	-1.19 (0.08)	-0.60 (0.06)	1.28 (0.08)	0.86 (0.07)
1.0	1.30 (0.08)	1.08 (0.07)	1.03 (0.07)	0.93 (0.07)
1.2	0.45 (0.06)	0.81 (0.07)	-0.84 (0.06)	1.33 (0.08)
1.4	-0.75 (0.06)	-0.83 (0.06)	-0.59 (0.06)	-0.90 (0.07)
1.6	-0.86 (0.07)	-0.66 (0.06)	-0.98 (0.06)	0.59 (0.06)
1.8	0.79 (0.06)	1.52 (0.09)	1.27 (0.08)	1.50 (0.09)

and 0 b tags (lower). The distributions agree with the SM background prediction and show the same behaviour in both lepton channels. Small differences arise due to higher thresholds in the electron channel, as well as from statistical fluctuations. The distributions display the number of events in each bin divided by the bin width in GeV. The ratio of the background prediction and the SR data is presented below the distributions. The uncertainties in the ratio for data in the SR are displayed as vertical lines and for the background prediction as grey bands, with the dark grey band being the statistical and the light grey the systematic uncertainties. The uncertainties for the predicted SM background are very small compared to simulation and are driven by the amount of collected data. Overall, the data agrees with the background prediction and no sign of new physics is found, which is quantified by calculating the standard score.

The standard score is calculated exploiting a hybrid approach of Bayesian and frequentist inference in the THETA framework [148] using binned maximum likelihood estimators. In this approach pseudo data are randomly sampled according to the background prediction and its model uncertainties. The obtained distributions are minimised for models both with and without signal, and the standard score is obtained from the ratio of the binned likelihood estimators [149, 150]. The uncertainties of the background prediction and the signal are considered as nuisance parameters where uncertainties on cross sections are sampled from a uniform prior and log-normal priors are used for the other nuisance parameters. The two lepton channels are combined for the calculations. Even though the amount of signal in the SB is small, its presence would

Table 8.2.: Observed (expected) upper limits at 95% CL on the product of the cross section and branching fraction for the B+b and  $X_{5/3}+t$  production modes, for a set of VLQ masses, for VLQs widths of 1% and 10%, and for left-handed and right-handed couplings. The exclusion limits for the B+t production mode (not shown) are very similar to those for the  $X_{5/3}+t$  mode. Table and caption from Ref. [30].

TeV	$X_{5/3}+t$			B+b		
	1% LH	10% LH	1% RH	1% LH	10% LH	1% RH
0.8	0.31 (0.27)	0.32 (0.25)	0.21 (0.18)	0.29 (0.36)	0.27 (0.36)	0.25 (0.29)
1	0.25 (0.15)	0.25 (0.16)	0.15 (0.10)	0.29 (0.17)	0.29 (0.19)	0.21 (0.12)
1.2	0.10 (0.09)	0.10 (0.10)	0.06 (0.06)	0.10 (0.10)	0.11 (0.11)	0.07 (0.07)
1.4	0.05 (0.06)	0.05 (0.07)	0.03 (0.05)	0.07 (0.07)	0.06 (0.08)	0.03 (0.05)
1.6	0.04 (0.04)	0.05 (0.05)	0.03 (0.03)	0.05 (0.05)	0.05 (0.06)	0.03 (0.04)
1.8	—	0.05 (0.04)	—	0.04 (0.04)	0.05 (0.04)	0.03 (0.03)

change the results and is taken into account. In contrast to the fits of the SB to the SR (for Fig. 8.1 and 8.2), for the significance the VLQ signals are taken into account in both regions. Therefore, the SB is fitted to the SR, including in both regions one signal mass point with either left- or right-handed VLQs from one of the production modes: B+b, B+t, or  $X_{5/3}+t$ . Afterwards, the cross section obtained from the binned maximum likelihood fit is subtracted from the SB and the uncertainty of the signal cross section fit is propagated to the background prediction. The significance is then calculated considering all uncertainties and the results are shown in Tab. 8.1 for the B+b and B+t production modes with a relative width of 1% (10%) and right-handed (left) VLQ couplings. The results for the  $X_{5/3}+t$  production mode are very similar to B+t. Ideally, the signal would be taken into account within the maximum likelihood without an additional uncertainty. This is technically difficult and not realised in the statistical framework available. The standard score has values around one, and the fitted cross sections are compatible with the excluded cross sections discussed below. Based on the distributions in Fig. 8.1 and 8.2 as well as the standard score results, no significant deviation from the SM background prediction is observed.

Exclusion limits at 95% CL [151] on the product of branching ratio and cross section are computed using a hybrid approach of Bayesian and frequentist inference with THETA. Similar to the standard score, pseudo data, binned maximum likelihood estimators, and the same uncertainties are used. The exclusion limits at 95% CL are presented for single VLQ production in the B+b, B+t, and  $X_{5/3}+t$  modes. The

predicted theory cross sections are taken from Refs. [69, 70, 133], where results for the production of single  $t$  are adapted to VLQ.

For the  $B+b$  and  $X_{5/3}+t$  production modes with 1 and 10% relative widths, masses between 0.8 and 1.8 TeV, and both left- and right-handed VLQ couplings, the exclusion limits at 95% CL are shown in Tab. 8.2. The exclusion limits at 95% CL for the left-handed VLQ with 1 and 10% relative widths result in very similar limits, while the exclusion limits for the signals with right-handed VLQ couplings and 1% relative widths are more stringent. In Fig. 8.3 the exclusion limits at 95% CL for the  $B+b$  production mode are illustrated for 1% relative width (upper) with left- (left) and right-handed (right) VLQ couplings. The results for the exclusion limits at 95% CL for the  $B+b$  production mode with left-handed VLQ couplings with 10% relative width (lower left) is very similar to the 1% relative width scenario. The observed exclusion limits at 95% CL for this production mode are also shown for 10, 20, and 30% relative widths (lower right). The results of the calculation of the exclusion limits at 95% CL for the  $B+t$  and  $X_{5/3}+t$  are displayed in Fig. 8.4. The scenarios for 1% relative width and left-handed VLQ couplings have approximately the same performance for both production modes, while the predicted cross section for the  $B+t$  (upper left) production mode is much smaller than for the  $X_{5/3}+t$  (upper right) production mode. For the  $X_{5/3}+t$  production mode with a 10% relative width and left-handed VLQ couplings (lower left) the exclusion limits at 95% CL are also shown. The observed exclusion limits at 95% CL for the 10, 20, and 30% widths scenarios (lower right) are only presented for this production mode. The  $B+t$  production mode has lower cross section predictions, but as expected shows similar results.

While  $B (X_{5/3})$  with 1% relative width cannot be excluded,  $B (X_{5/3})$  with 10, 20, and 30% relative widths as well as left-handed VLQs couplings are excluded by the calculated exclusion limits at 95% CL for masses below 1490, 1590, and 1660 GeV (920, 1300, and 1450 GeV). The exclusion limits range from 0.3 to 0.03 pb for masses between 0.7 and 2 TeV.

Single VLQs have been explored in data with a centre-of-mass energy of 8 TeV in Ref. [19, 152] by ATLAS, with comparable results in terms of exclusion limits at 95% CL on the product of branching fraction and production cross section. The higher centre of mass energy is accompanied by more pileup and higher background cross

---

sections, especially at high  $p_T$ . This makes the analysis more challenging. A higher reach in term of VLQ masses is obtained from the higher production cross section predicted for the signal. To achieve similar exclusion limits on the product of production of cross section and branching fraction, a data-driven background method with much smaller uncertainties, as well as the more sophisticated VLQ reconstruction, are essential. Additionally, events with two leptons have been considered by the ATLAS search, which would improve the sensitivity of this search and are currently being studied in CMS. The inclusion of the two lepton final state would allow for a better differentiation between the  $B+t$  and  $X_{5/3}+t$  production modes, which create either leptons with the same charge sign or opposite charge signs. Because of the increase in the production cross section, the mass exclusion limits in this search are more stringent than the results at 8 TeV.

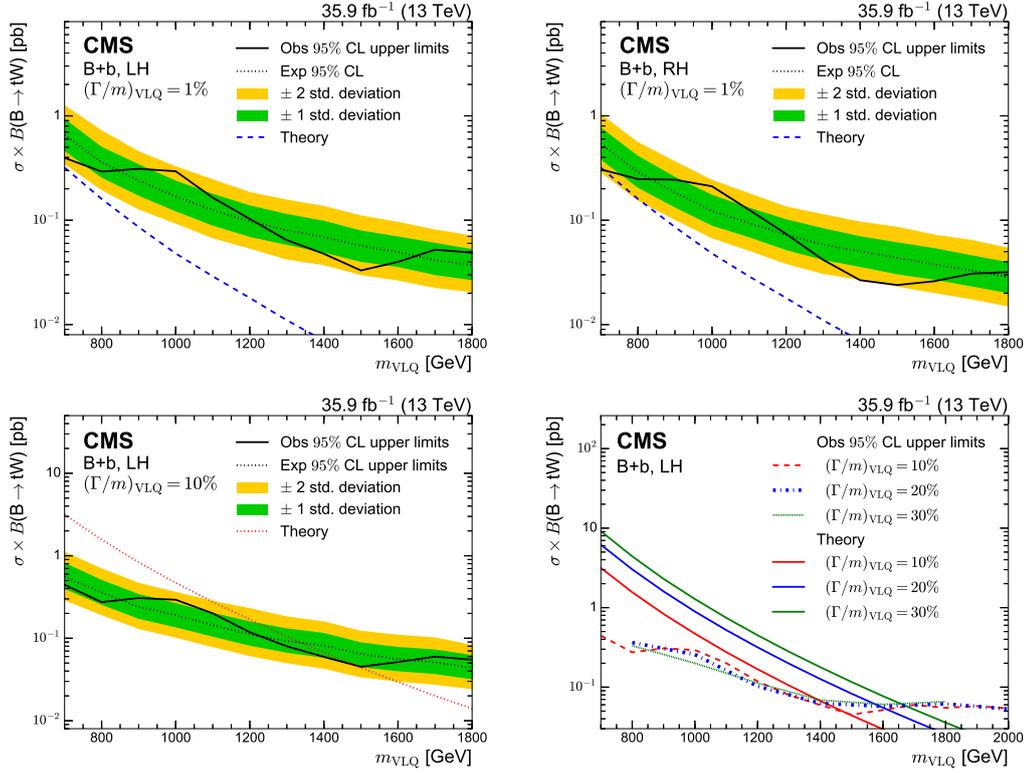


Figure 8.3.: Upper limits at 95% CL on the product of the VLQ production cross section and branching fraction for the B+b production mode for a relative VLQ width of 1% and left- and right-handed VLQ couplings (upper left and right), for 10% relative VLQ width and left-handed VLQ couplings (lower left), and a comparison of the observed exclusion limits for relative VLQ widths of 10, 20, and 30% for left-handed couplings (lower right). The dashed lines show the theoretical predictions. Distributions and caption from Ref. [30].

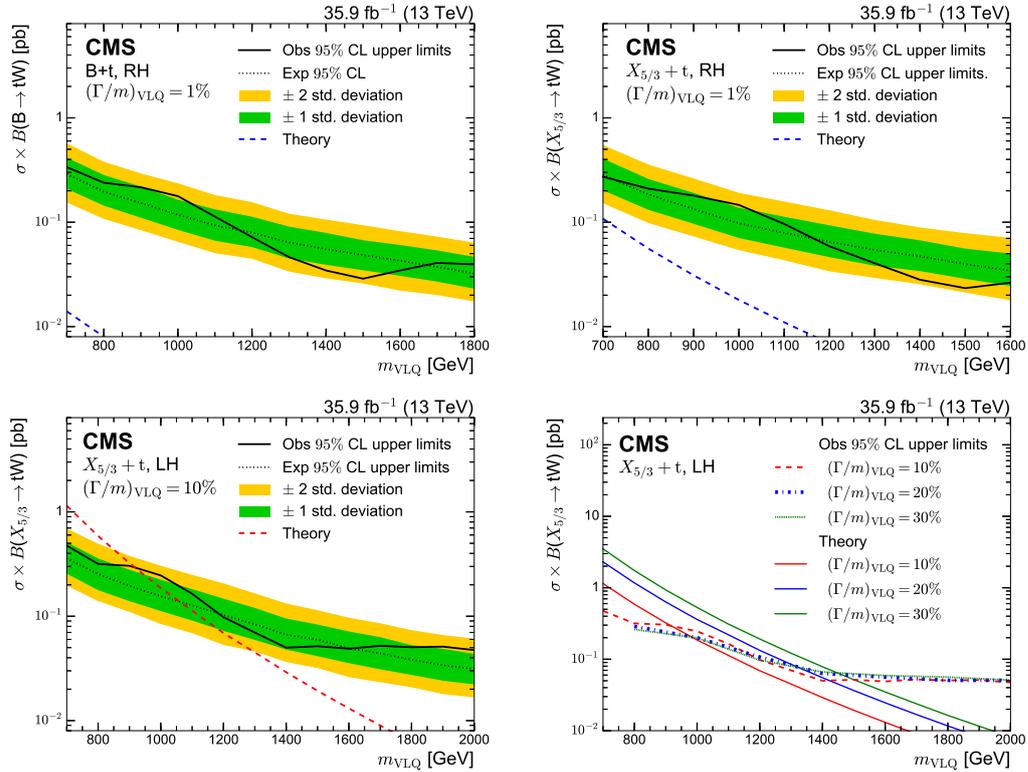


Figure 8.4.: Upper limits at 95% CL on the product of the VLQ production cross section and branching fraction for the  $B+t$  and  $X_{5/3}+t$  production modes for right-handed VLQ couplings assuming a relative VLQ width of 1% (upper left and right), for the  $X_{5/3}+t$  production mode with left-handed VLQ couplings and a 10% relative width (lower left) and a comparison of the observed exclusion limits for left-handed couplings for relative widths of 10, 20, and 30% (lower right). The dashed lines show the theoretical predictions. Distributions and text from Ref. [30].



## 9. Summary & outlook

In this thesis a search for singly produced vector-like quarks was performed using the data recorded by the CMS experiment in 2016 at the LHC, corresponding to an integrated luminosity of  $35.9 \text{ fb}^{-1}$ . The search is optimised for vector-like quarks with masses above 1 TeV decaying into a top quark and W boson. The targeted final state consists of a single muon or electron, at least two jets, and a significant amount of missing transverse momentum. The single vector-like quark is produced in association with a jet, detected near the beam pipe, and a top or bottom quark. The jet near the beam pipe is used to define a signal region since jets in this region are rare for processes at the LHC. The associated low momentum top or bottom quark is not considered further. The mass of the vector-like quark is reconstructed using jets, the lepton, and the missing energy in the event. In the simulation, the reconstructed mass shows a high discrimination power between Standard Model and new physics processes. The event selection is optimised in simulation to maximise the discrimination power of the mass reconstruction. In the interpretation of the signal region, the Standard Model prediction is taken from data by utilising the events without a jet near the beam pipe. The sensitivity is further enhanced by splitting the events into categories based on the quarks and bosons identified within hadronic jets: bottom quarks, boosted top quarks, and boosted W bosons. In the vector-like quark mass distributions, no significant deviation from the background prediction is observed, and exclusion limits at 95% confidence level are set on the product of branching fraction and cross section. The exclusion limits range between 0.3–0.03 pb for masses of the vector-like quarks between 0.7 and 2 TeV. The exact values of the exclusion limits depend on the mass, width, type, and Standard Model couplings of the vector-like quark. For relative widths of 30%, exclusion limits of up to 1.66 TeV are set on the vector-like quark mass, which are the world’s most stringent exclusion limits in this channel.

---

The presented analysis is mostly limited by statistics. With more data, the statistical uncertainties regarding the data-driven background can be reduced. Additionally, further improvements can be studied such as the use of deep neural networks to identify the jet near the beam pipe. This aspect may be crucial to lower the Standard Model background contribution in the search region and may be feasible with extended coverage of the tracker up to  $|\eta|$  of 5, as planned for future detector upgrades of CMS. In this detector scenario, the use of triggers based on jets in the region between  $2.5 < |\eta| < 5$  are important. These upgrades and more data may enable the discovery or exclusion of VLQs with relative mass widths of 1% over the presented mass range.





# Bibliography

- [1] ATLAS, CMS Collaboration, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC  $pp$  collision data at  $\sqrt{s} = 7$  and 8 TeV”, *JHEP* **08** (2016) 045, [arXiv:1606.02266](#).
- [2] CMS Collaboration, “Inclusive search for a vector-like T quark with charge  $\frac{2}{3}$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Lett. B* **729** (2014) 149, [arXiv:1311.7667](#).
- [3] CMS Collaboration, “Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure”, *JHEP* **06** (2015) 080, [arXiv:1503.01952](#).
- [4] CMS Collaboration, “Search for vector-like charge  $2/3$  T quarks in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Rev. D* **93** (2016) 012003, [arXiv:1509.04177](#).
- [5] CMS Collaboration, “Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state”, *Phys. Lett. B* **771** (2017) 80, [arXiv:1612.00999](#).
- [6] CMS Collaboration, “Search for single production of vector-like quarks decaying into a b quark and a W boson in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *Phys. Lett. B* **772** (2017) 634, [arXiv:1701.08328](#).
- [7] CMS Collaboration, “Search for top quark partners with charge  $5/3$  in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **08** (2017) 073,

arXiv:1705.10967.

- [8] CMS Collaboration, “Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **11** (2017) 085, arXiv:1706.03408.
- [9] CMS Collaboration, “Search for pair production of vector-like quarks in the  $bW\bar{b}W$  channel from proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *Phys. Lett. B* **779** (2018) 82, arXiv:1710.01539.
- [10] CMS Collaboration, “Search for single production of a vector-like T quark decaying to a Z boson and a top quark in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *Phys. Lett. B* **781** (2018) 574, arXiv:1708.01062.
- [11] CMS Collaboration, “Search for single production of vector-like quarks decaying to a b quark and a Higgs boson”, *JHEP* **06** (2018) 031, arXiv:1802.01486.
- [12] CMS Collaboration, “Search for a W’ boson decaying to a vector-like quark and a top or bottom quark in the all-jets final state”, Tech. Rep., CERN, Geneva, 2018. [cds.cern.ch/record/2631600](https://cds.cern.ch/record/2631600).
- [13] CMS Collaboration, “Search for Pair Production of Vector-Like Quarks in the Fully Hadronic Channel”, Tech. Rep., CERN, Geneva, 2019. [cds.cern.ch/record/2667230](https://cds.cern.ch/record/2667230).
- [14] ATLAS Collaboration, “Search for heavy vector-like quarks coupling to light quarks in proton-proton collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector”, *Phys. Lett. B* **712** (2012) 22, arXiv:1112.5755.
- [15] ATLAS Collaboration, “Search for pair production of a new quark that decays to a Z boson and a bottom quark with the ATLAS detector”, *Phys. Rev. Lett.* **109** (2012) 071801, arXiv:1204.1265.

- 
- [16] ATLAS Collaboration, “Search for pair and single production of new heavy quarks that decay to a  $Z$  boson and a third-generation quark in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *JHEP* **11** (2014) 104, [arXiv:1409.5500](#).
- [17] ATLAS Collaboration, “Analysis of events with  $b$ -jets and a pair of leptons of the same charge in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *JHEP* **10** (2015) 150, [arXiv:1504.04605](#).
- [18] ATLAS Collaboration, “Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *JHEP* **08** (2015) 105, [arXiv:1505.04306](#).
- [19] ATLAS Collaboration, “Search for vector-like  $B$  quarks in events with one isolated lepton, missing transverse momentum and jets at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *Phys. Rev. D* **91** (2015) 112011, [arXiv:1503.05425](#).
- [20] ATLAS Collaboration, “Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in  $\sqrt{s} = 13$  TeV  $pp$  collisions with the ATLAS detector”, *Phys. Rev. D* **94** (2016) 052009, [arXiv:1606.03903](#).
- [21] ATLAS Collaboration, “Search for single production of vector-like quarks decaying into  $Wb$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *EPJ. C* **76** (2016) 442, [arXiv:1602.05606](#).
- [22] ATLAS Collaboration, “Search for single production of a vector-like quark via a heavy gluon in the  $4b$  final state with the ATLAS detector in  $pp$  collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Lett. B* **758** (2016) 249, [arXiv:1602.06034](#).
- [23] ATLAS Collaboration, “Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in  $\sqrt{s} = 13$  TeV  $pp$  collisions with the ATLAS detector”, *JHEP* **08** (2017) 052, [arXiv:1705.10751](#).

- [24] ATLAS Collaboration, “Search for pair production of heavy vector-like quarks decaying to high- $p_T$  W bosons and b quarks in the lepton-plus-jets final state in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, *JHEP* **10** (2017) 141, [arXiv:1707.03347](#).
- [25] ATLAS Collaboration, “Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple  $b$ -jets with the ATLAS detector”, *JHEP* **07** (2018) 089, [arXiv:1803.09678](#).
- [26] ATLAS Collaboration, “Combination of the searches for pair-produced vector-like partners of the third-generation quarks at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, *Phys. Rev. Lett.* **121** (2018) 211801, [arXiv:1808.02343](#).
- [27] ATLAS Collaboration, “Search for single production of vector-like quarks decaying into  $Wb$  in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, *Submitted to: JHEP* (2018) , [arXiv:1812.07343](#).
- [28] ATLAS Collaboration, “Search for large missing transverse momentum in association with one top-quark in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, *JHEP* **05** (2019) 41, [arXiv:1812.09743](#).
- [29] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria, “Handbook of vectorlike quarks: Mixing and single production”, *Phys. Rev. D* **88** (2013) 094010, [arXiv:1306.0572](#).
- [30] CMS Collaboration, “Search for single production of vector-like quarks decaying to a top quark and a W boson in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, [arXiv:1809.08597](#).
- [31] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30, [arXiv:1207.7235](#).
- [32] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys.*

- Lett. B* **716** (2012) 1, [arXiv:1207.7214](https://arxiv.org/abs/1207.7214).
- [33] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008.
- [34] G. Altarelli, “The Standard model of particle physics”, [arXiv:hep-ph/0510281](https://arxiv.org/abs/hep-ph/0510281).
- [35] B. Martin and G. Shaw, *Particles Physics; fourth edition*. Physics textbook. Wiley, West Sussex, United Kingdom, 2017.
- [36] Particle Data Group Collaboration *Phys. Rev. D* **98** (Aug, 2018) 030001. [link.aps.org/doi/10.1103/PhysRevD.98.030001](https://link.aps.org/doi/10.1103/PhysRevD.98.030001).
- [37] “Particle content of the standard model of particle physics”, 2017. [commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg).
- [38] H1, ZEUS Collaboration, “HERA Inclusive Neutral and Charged Current Cross Sections and a New PDF Fit, HERAPDF 2.0”, *Acta Phys. Polon. Supp.* **8** (2015) 957, [arXiv:1511.05402](https://arxiv.org/abs/1511.05402).
- [39] V. N. Gribov and L. N. Lipatov, “Deep inelastic e p scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438.
- [40] L. N. Lipatov, “The parton model and perturbation theory”, *Sov. J. Nucl. Phys.* **20** (1975) 94.
- [41] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov. Phys. JETP* **46** (1977) 641.
- [42] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys.* **B126** (1977) 298.

- [43] ATLAS Collaboration, “Summary of SM measurements at ATLAS”, 2019.  
[atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM](http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM).
- [44] Z. Nagy and D. E. Soper, “What is a parton shower?”, *Phys. Rev. D* **98** (2018) 014034, [arXiv:1705.08093](https://arxiv.org/abs/1705.08093).
- [45] Supernova Search Team Collaboration, “Observational evidence from supernovae for an accelerating universe and a cosmological constant”, *Astron. J.* **116** (1998) 1009, [arXiv:astro-ph/9805201](https://arxiv.org/abs/astro-ph/9805201).
- [46] A. Refregier, “Weak gravitational lensing by large scale structure”, *Ann. Rev. Astron. Astrophys.* **41** (2003) 645, [arXiv:astro-ph/0307212](https://arxiv.org/abs/astro-ph/0307212).
- [47] 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, [arXiv:astro-ph/0507092](https://arxiv.org/abs/astro-ph/0507092).
- [48] 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, [arXiv:astro-ph/0608407](https://arxiv.org/abs/astro-ph/0608407).
- [49] C. Rovelli, “Quantum gravity”, *Scholarpedia* **3** (2008) . Revision 170369.
- [50] L. Canetti, M. Drewes, and M. Shaposhnikov, “Matter and Antimatter in the Universe”, *New J. Phys.* **14** (2012) 095012, [arXiv:1204.4186](https://arxiv.org/abs/1204.4186).
- [51] M. Dine, “Naturalness Under Stress”, *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 43, [arXiv:1501.01035](https://arxiv.org/abs/1501.01035).
- [52] S. Alekhin, A. Djouadi, and S. Moch, “The top quark and Higgs boson masses and the stability of the electroweak vacuum”, *Phys. Lett. B* **716** (2012) 214.  
[www.sciencedirect.com/science/article/pii/S0370269312008611](http://www.sciencedirect.com/science/article/pii/S0370269312008611).

- 
- [53] S. Mertens, “Direct Neutrino Mass Experiments”, *J. Phys. Conf. Ser.* **718** (2016) 022013, [arXiv:1605.01579](#).
- [54] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562, [arXiv:hep-ex/9807003](#).
- [55] SNO Collaboration, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* **89** (2002) 011301, [arXiv:nucl-ex/0204008](#).
- [56] S. P. Martin, “A Supersymmetry primer”, [arXiv:hep-ph/9709356](#). Adv. Ser. Direct. HEP.
- [57] U. Lindstrom, “Supersymmetry, a biased review”, *Rend. Circ. Mat. Palermo* **S71** (2003) 41, [arXiv:hep-th/0204016](#).
- [58] ATLAS Collaboration, “Summary of SUSY searches at ATLAS”, 2019. [twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults](#).
- [59] CMS Collaboration, “Summary of SUSY searches at CMS”, 2019. [cms-results.web.cern.ch/cms-results/public-results/preliminary-results/SUS/index.html](#).
- [60] O. Klein, “Quantentheorie und fünfdimensionale Relativitätstheorie”, *Zeitschrift für Physik* **37** (Dec, 1926) 895. [doi.org/10.1007/BF01397481](#).
- [61] A. Carvalho, “Gravity particles from warped extra dimensions, predictions for LHC”, [arXiv:1404.0102](#).
- [62] G. Panico and A. Wulzer, “The Composite Nambu-Goldstone Higgs”, *Lect. Notes Phys.* **913** (2016) 1, [arXiv:1506.01961](#).
- [63] M. Schmaltz and D. Tucker-Smith, “Little Higgs review”, *Ann. Rev. Nucl. Part. Sci.* **55** (2005) 229, [arXiv:hep-ph/0502182](#).

- [64] M. Perelstein, “Little Higgs models and their phenomenology”, *Prog. Part. Nucl. Phys.* **58** (2007) 247, [arXiv:hep-ph/0512128](#).
- [65] J. Bulava, K. Jansen, and A. Nagy, “Constraining a fourth generation of quarks: non-perturbative Higgs boson mass bounds”, *Phys. Lett.* **B723** (2013) 95, [arXiv:1301.3416](#).
- [66] A. De Simone, O. Matsedonskyi, R. Rattazzi, and A. Wulzer, “A First Top Partner Hunter’s Guide”, *JHEP* **04** (2013) 004, [arXiv:1211.5663](#).
- [67] Y. Okada and L. Panizzi, “LHC signatures of vector-like quarks”, *Adv. High Energy Phys.* **2013** (2013) 364936, [arXiv:1207.5607](#).
- [68] ATLAS Collaboration, “Combination of the Searches for Pair-Produced Vectorlike Partners of the Third-Generation Quarks at  $\sqrt{s} = 13$  TeV with the ATLAS Detector”, *Phys. Rev. Lett.* **121** (Nov, 2018) 211801. [link.aps.org/doi/10.1103/PhysRevLett.121.211801](https://link.aps.org/doi/10.1103/PhysRevLett.121.211801).
- [69] O. Matsedonskyi, G. Panico, and A. Wulzer, “On the Interpretation of Top Partners Searches”, *JHEP* **12** (2014) 097, [arXiv:1409.0100](#).
- [70] A. Carvalho, S. Moretti, D. O’Brien, L. Panizzi, and H. Prager, “Single production of vector-like quarks with large width at the Large Hadron Collider”, *Phys. Rev. D* **98** (2018) 015029, [arXiv:1805.06402](#).
- [71] CMS Collaboration, “Technical Design Report Volume 1: Detector Performance and Software”, Tech. Rep., CERN, Geneva, 2006. [cds.cern.ch/record/922757](https://cds.cern.ch/record/922757).
- [72] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. [cds.cern.ch/record/782076](https://cds.cern.ch/record/782076).
- [73] CMS Collaboration, “Luminosity and Pileup distributions for 2016  $pp$  collisions”, 2017. [twiki.cern.ch/twiki/bin/view/CMSPublic/](https://twiki.cern.ch/twiki/bin/view/CMSPublic/)

LumiPublicResults#2016\_proton\_proton\_13\_TeV\_collis.

- [74] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, Tech. Rep., CERN, Geneva, 2017. [cds.cern.ch/record/2257069](https://cds.cern.ch/record/2257069).
- [75] A. Dominguez *et al.*, “CMS Technical Design Report for the Pixel Detector Upgrade”, Tech. Rep., CERN, Sep, 2012. [cds.cern.ch/record/1481838](https://cds.cern.ch/record/1481838).
- [76] CMS Collaboration, “The Phase-2 Upgrade of the CMS Tracker”, Tech. Rep., CERN, Geneva, Jun, 2017. [cds.cern.ch/record/2272264](https://cds.cern.ch/record/2272264).
- [77] ATLAS Collaboration, “ATLAS liquid-argon calorimeter: Technical Design Report”, Tech. Rep., CERN, Geneva, 1996. [cds.cern.ch/record/331061](https://cds.cern.ch/record/331061).
- [78] CMS-HCAL Collaboration, “Design, performance, and calibration of CMS hadron-barrel calorimeter wedges”, *EPJ C* **55** (May, 2008) 159. [doi.org/10.1140/epjc/s10052-008-0573-y](https://doi.org/10.1140/epjc/s10052-008-0573-y).
- [79] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017) P01020, [arXiv:1609.02366](https://arxiv.org/abs/1609.02366).
- [80] ALEPH Collaboration, “Performance of the ALEPH detector at LEP”, *Nucl. Instrum. Meth. A* **360** (1995) 481.
- [81] T. C. collaboration, “Tests of a Particle Flow Algorithm with CALICE test beam data”, *JINST* **6** (2011) 7005. [stacks.iop.org/1748-0221/6/i=07/a=P07005](https://stacks.iop.org/1748-0221/6/i=07/a=P07005).
- [82] M. A. Thomson, “Particle Flow Calorimetry and the Pandora PFA Algorithm”, *Nucl. Instrum. Meth. A* **611** (2009) 25, [arXiv:0907.3577](https://arxiv.org/abs/0907.3577).
- [83] TLEP Design Study Working Group, “First Look at the Physics Case of TLEP”, *JHEP* **01** (2014) 164, [arXiv:1308.6176](https://arxiv.org/abs/1308.6176).

- [84] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, [arXiv:1706.04965](#).
- [85] CMS Collaboration, “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *J. Phys. G Nucl. Phys.* **31** (Sept., 2005) N9, [arXiv:physics/0306087](#).
- [86] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with  $pp$  collisions at  $\sqrt{s} = 13$  TeV”, *JINST* **13** (2018) P06015, [arXiv:1804.04528](#).
- [87] CMS Collaboration, “Electron and Photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 Conference”, Tech. Rep., CERN, May, 2018. [cds.cern.ch/record/2320638](#).
- [88] CMS Collaboration, “Search for  $t\bar{t}$  resonances in highly boosted lepton+jets and fully hadronic final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **07** (2017) 001, [arXiv:1704.03366](#).
- [89] CMS Collaboration, “Search for resonant  $t\bar{t}$  production in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Rev. D* **93** (2016) 012001, [arXiv:1506.03062](#).
- [90] CMS Collaboration, “Performance of CMS muon reconstruction in  $pp$  collision events at  $\sqrt{s} = 7$  TeV”, *JINST* **7** (2012) P10002, [arXiv:1206.4071](#).
- [91] A. Ali and G. Kramer, “Jets and QCD: A Historical Review of the Discovery of the Quark and Gluon Jets and its Impact on QCD”, *EPJ. H* **36** (2011) 245, [arXiv:1012.2288](#).
- [92] G. P. Salam, “Towards Jetography”, *EPJ. C* **67** (2010) 637, [arXiv:0906.1833](#).
- [93] CMS Collaboration, “Measurement of the jet mass in highly boosted  $t\bar{t}$  events from  $pp$  collisions at  $\sqrt{s} = 8$  TeV”, *EPJ. C* **77** (2017) 467, [arXiv:1703.06330](#).

- 
- [94] CMS Collaboration, “Measurement of jet substructure observables in  $t\bar{t}$  events from  $pp$  collisions at  $\sqrt{s} = 13$  TeV”, [arXiv:1808.07340](#).
- [95] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in  $pp$  collisions at 8 TeV”, *JINST* **12** (2017) P02014, [arXiv:1607.03663](#).
- [96] D. Yu. Grigoriev, E. Jankowski, and F. V. Tkachov, “Towards a standard jet definition”, *Phys. Rev. Lett.* **91** (2003) 061801, [arXiv:hep-ph/0301185](#).
- [97] M. Cacciari and G. P. Salam, “Dispelling the  $N^3$  myth for the  $k_T$  jet-finder”, *Phys. Lett. B* **641** (2006) 57, [arXiv:hep-ph/0512210](#).
- [98] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *EPJ. C* **72** (2012) 1896, [arXiv:1111.6097](#).
- [99] CMS Collaboration, “Pileup Removal Algorithms”, Tech. Rep., CERN, Geneva, 2014. [cds.cern.ch/record/1751454](#).
- [100] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification”, *JHEP* **10** (2014) 059, [arXiv:1407.6013](#).
- [101] CMS Collaboration, “Jet energy scale and resolution performances with 13 TeV data”, Tech. Rep., CERN, Jun, 2016. [cds.cern.ch/record/2160347](#).
- [102] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, Tech. Rep., CERN, Geneva, 2017. [cds.cern.ch/record/2256875](#).
- [103] ATLAS Collaboration, “Identification of high transverse momentum top quarks in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *JHEP* **06** (2016) 093, [arXiv:1603.03127](#).
- [104] ATLAS Collaboration, “Boosted hadronic top identification at ATLAS for early 13 TeV data”, Tech. Rep., CERN, Geneva, Dec, 2015. [cds.cern.ch/record/2116351](#).

- 
- [105] CMS Collaboration, “Boosted Top Jet Tagging at CMS”, Tech. Rep., CERN, Geneva, 2014. [cds.cern.ch/record/1647419](https://cds.cern.ch/record/1647419).
- [106] CMS Collaboration, “Top Tagging with New Approaches”, Tech. Rep., CERN, Geneva, 2016. [cds.cern.ch/record/2126325](https://cds.cern.ch/record/2126325).
- [107] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure”, *JHEP* **09** (2013) 029, [arXiv:1307.0007](https://arxiv.org/abs/1307.0007).
- [108] M. Dasgupta, A. Fregoso, S. Marzani, and A. Powling, “Jet substructure with analytical methods”, *EPJ. C* **73** (2013) 2623, [arXiv:1307.0013](https://arxiv.org/abs/1307.0013).
- [109] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146, [arXiv:1402.2657](https://arxiv.org/abs/1402.2657).
- [110] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **03** (2011) 015, [arXiv:1011.2268](https://arxiv.org/abs/1011.2268).
- [111] CMS Collaboration, “Performance of b-Tagging Algorithms in Proton Collisions at 13 TeV using the 2016 Data”, Tech. Rep., CERN, Jul, 2016. [cds.cern.ch/record/2202967](https://cds.cern.ch/record/2202967).
- [112] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, Tech. Rep., CERN, Geneva, 2016. [cds.cern.ch/record/2138504](https://cds.cern.ch/record/2138504).
- [113] CMS Collaboration, “Performance of missing energy reconstruction in 13 TeV  $p p$  collision data using the CMS detector”, Tech. Rep., CERN, Geneva, 2016. [cds.cern.ch/record/2205284](https://cds.cern.ch/record/2205284).
- [114] T. Sjöstrand, “Status and developments of event generators”, *PoS LHCP2016* (2016) 007, [arXiv:1608.06425](https://arxiv.org/abs/1608.06425).
- [115] GEANT4 Collaboration, “GEANT4—A simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250.

- 
- [116] 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](#).
- [117] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, [arXiv:hep-ph/0409146](#).
- [118] 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](#).
- [119] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, [arXiv:1002.2581](#).
- [120] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, [arXiv:1209.6215](#).
- [121] J. Alwall *et al.*, “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *EPJ C* **53** (2008) 473, [arXiv:0706.2569](#).
- [122] 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, [arXiv:1410.3012](#).
- [123] CMS Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of  $t\bar{t}$  at  $\sqrt{s} = 8$  and 13 TeV”, Tech. Rep., CERN, Geneva, 2016. [cds.cern.ch/record/2235192](#).
- [124] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *EPJ C* **76** (2016) 155, [arXiv:1512.00815](#).

- [125] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 tune”, *EPJ C* **74** (2014) 3024, [arXiv:1404.5630](#).
- [126] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, [arXiv:1410.8849](#).
- [127] J. Butterworth *et al.*, “PDF4LHC recommendations for LHC Run II”, *J. Phys. G* **43** (2016) 023001, [arXiv:1510.03865](#).
- [128] J. Rojo *et al.*, “The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II”, *J. Phys. G* **42** (2015) 103103, [arXiv:1507.00556](#).
- [129] M. Buchkremer, G. Cacciapaglia, A. Deandrea, and L. Panizzi, “Model independent framework for searches of top partners”, *Nucl. Phys. B* **876** (2013) 376, [arXiv:1305.4172](#).
- [130] B. Fuks and H.-S. Shao, “QCD next-to-leading-order predictions matched to parton showers for vector-like quark models”, *EPJ C* **77** (2017) 135, [arXiv:1610.04622](#).
- [131] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, “Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations”, *JHEP* **04** (2007) 081, [arXiv:hep-ph/0702198](#).
- [132] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”, *JHEP* **03** (2013) 015, [arXiv:1212.3460](#).
- [133] J. M. Campbell, R. K. Ellis, and F. Tramontano, “Single top production and decay at next-to-leading order”, *Phys. Rev. D* **70** (2004) 094012, [arXiv:hep-ph/0408158](#).
- [134] S. Abdullin, P. Azzi, F. Beaudette, P. Janot, and A. P. and, “The Fast Simulation of the CMS Detector at LHC”, *J. Phys. Conf. Ser.* **331** (Dec, 2011)

032049. doi.org/10.1088%2F1742-6596%2F331%2F3%2F032049.
- [135] A. Giammanco, “The Fast Simulation of the CMS Experiment”, *J. Phys. Conf. Ser.* **513** (Jun, 2014) . doi.org/10.1088%2F1742-6596%2F513%2F2%2F022012.
- [136] CMS Collaboration, “Recent Developments in CMS Fast Simulation”, in *Int. Conf. on High Energy Physics (ICHEP 2016)*, vol. ICHEP2016, 181. 2016. arXiv:1701.03850.
- [137] G. Grindhammer and S. Peters, “The Parameterized simulation of electromagnetic showers in homogeneous and sampling calorimeters”, in *Int. Conf. on MC Sim. in High-Energy and Nuc. Phys.* 1993. arXiv:hep-ex/0001020.
- [138] H1 Collaboration, “The Tracking, calorimeter and muon detectors of the H1 experiment at HERA”, *Nucl. Instrum. Meth.* **A386** (1997) 348.
- [139] J. J. Aubert *et al.*, “Experimental Observation of a Heavy Particle  $J$ ”, *Phys. Rev. Lett.* **33** (Dec, 1974) 1404. link.aps.org/doi/10.1103/PhysRevLett.33.1404.
- [140] J. E. Augustin *et al.*, “Discovery of a Narrow Resonance in  $e^+e^-$  Annihilation”, *Phys. Rev. Lett.* **33** (Dec, 1974) 1406. link.aps.org/doi/10.1103/PhysRevLett.33.1406.
- [141] D0 Collaboration, “Search for high mass top quark production in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV”, *Phys. Rev. Lett.* **74** (1995) 2422, arXiv:hep-ex/9411001.
- [142] CDF Collaboration, “Observation of top quark production in  $\bar{p}p$  collisions”, *Phys. Rev. Lett.* **74** (1995) 2626, arXiv:hep-ex/9503002.
- [143] D. Hirschbühl, *Measurement of the Charge Asymmetry and the W Boson Helicity in Top-Antitop Quark Events with the CDF II Experiment*. PhD thesis, KIT, 2005.

- [144] D. Gonzalez Vazquez, “Search for  $t\bar{t}$ -Resonances with the CMS-Detector”, Master’s thesis, Universität Hamburg, 2013.
- [145] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 7$  TeV”, *Phys. Lett. B* **722** (2013) 5, [arXiv:1210.6718](#).
- [146] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018) P05011, [arXiv:1712.07158](#).
- [147] M. Bähr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer, P. Richardson, M. H. Seymour, A. Sherstnev, and B. R. Webber, “Herwig++ physics and manual”, *EPJ. C* **58** (2008) 639, [arXiv:0803.0883](#).
- [148] J. Ott, J. Wagner-Kuhr, and T. Müller, “THETA — A framework for template-based modeling and inference” (2010), <http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto>.
- [149] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, *Ann. Math. Statist.* **9** (03, 1938) 60. <https://doi.org/10.1214/aoms/1177732360>.
- [150] A. O’Hagan and J. J. Forster, *Kendall’s advanced theory of statistics. Vol. 2B: Bayesian Inference*. Arnold, London, 2004.
- [151] A. L. Read, “Modified frequentist analysis of search results (The CL(s) method)”, in *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 January 2000: Proceedings*, 81. 2000. <http://weblib.cern.ch/abstract?CERN-OPEN-2000-205>.
- [152] ATLAS Collaboration, “Search for the production of single vector-like and excited quarks in the  $Wt$  final state in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *JHEP* **02** (2016) 110, [arXiv:1510.02664](#).





---

## A. Acronyms

**ak4 jet** anti- $k_T$  jet with  $R = 0.4$

**ak8 jet** anti- $k_T$  jet with  $R = 0.8$

**ALEPH** Apparatus for LEP Physics

**ALICE** A Large Ion Collider Experiment

**ATLAS** A Toroidal LHC Apparatus

**BDT** boosted decision tree

**b** bottom

**BG** background

**b tag** b quark identification algorithm

**BSM** beyond the Standard Model

**CERN** Conseil Européen pour la Recherche Nucléaire

**CHS** charged hadron subtraction

**CKM** Cabibbo-Kobayashi-Maskawa

**CL** confidence level

**CMS** Compact Muon Solenoid

**CP** charge conjugation parity

**CSC** cathode strip chambers

**CSVv2** combined secondary vertex algorithm

**DELPHI** Detector with Lepton, Photon and Hadron Identification

**DGLAP** Dokshitzer-Gribov-Lipatov-Altarelli-Parisi

**DIS** deep inelastic scattering

**DT** drift tubes

**ECAL** electromagnetic calorimeter

**FSR** final state radiation

**GSF** Gaussian-sum filter

**HCAL** hadronic calorimeter

**HERA** Hadron-Elektron-Ring-Anlage  
**HF** hadron forward calorimeter  
**HLT** high level trigger  
**ISR** initial state radiation  
**JEC** jet energy corrections  
**JER** jet energy resolution  
**LEP** Large Electron Positron Collider  
**LHC** Large Hadron Collider  
**LHCb** Large Hadron Collider beauty  
**LO** leading order  
**MC** Monte Carlo  
**ME** matrix element  
**MIP** minimal ionising particles  
**M&M** matching and merging  
**MPI** multiple particle interactions  
**MVA** multivariate analysis  
**NLO** next to leading order  
**NNLO** next-to-next to leading order  
**PDF** parton distribution function  
**PF** particle flow  
**PS** Proton Synchrotron  
**PS** parton shower  
 $\vec{p}_T^{\text{miss}}$  missing transverse momentum  
**puppi** pileup per particle identification  
**QCD** quantum chromodynamics  
**QED** quantum electrodynamics  
**RPC** resistiv plate chambers  
**SB** sideband  
**SM** Standard Model  
**SR** signal region  
**SPS** Super Proton Synchrotron  
**TC** Technicolour Model  
**t tag** top quark identification algorithm  
**VLQ** vector-like quark





## B. Theory prediction

The predicted cross sections of the  $B+b$ ,  $B+t$ , and  $X_{5/3}+t$  production modes are calculated based on the production cross section of the SM single top process. The calculation of the cross section values has been done in Ref. [69, 70, 133]. In Tab. B.1 the expected cross sections in pb of the  $B+b$ ,  $B+t$ , and  $X_{5/3}+t$  production modes are presented for several relative VLQ widths.

Table B.1.: Theory prediction for the production cross section of the B+b, B+t, and  $X_{5/3}+t$  production modes with relative mass widths between 1 and 30% in pb.

Mass [GeV ]	B+b [pb]				B+t [pb]				$X_{5/3}+t$ [pb]			
	1%	10%	20%	30%	1%	10%	20%	30%	1%	10%	20%	30%
700	0.319	3.16	6.18	9.13	0.014	0.159	0.321	0.485	0.108	1.146	2.3	3.5
800	0.16	1.563	3.03	4.42	0.008	0.084	0.169	0.251	0.056	0.588	1.17	1.74
900	0.086	0.833	1.6	2.3	0.0046	0.048	0.094	0.138	0.031	0.323	0.633	0.933
1000	0.048	0.47	0.893	1.283	0.0028	0.028	0.055	0.079	0.018	0.186	0.36	0.53
1100	0.029	0.276	0.522	0.745	0.0017	0.017	0.033	0.048	0.011	0.113	0.22	0.31
1200	0.018	0.168	0.3167	0.4495	0.001	0.011	0.021	0.0298	0.0069	0.069	0.133	0.192
1300	0.011	0.105	0.1976	0.2792	0.0007	0.0064	0.0132	0.019	0.0044	0.045	0.0849	0.1214
1400	0.007	0.067	0.1265	0.1783	0.0005	0.0046	0.0087	0.012	0.0029	0.0291	0.0551	0.0785
1500	0.0047	0.044	0.0827	0.1161	0.0003	0.0031	0.0058	0.008	0.0019	0.0193	0.0365	0.0518
1600	0.003	0.0295	0.0551	0.0771	0.0002	0.0021	0.0039	0.0055	0.0013	0.0131	0.0246	0.0348
1700	0.002	0.02	0.0373	0.0522	0.0001	0.0014	0.0027	0.0038	0.0009	0.009	0.0168	0.0238
1800	0.0014	0.014	0.0256	0.0358	0.0001	0.001	0.0019	0.0026	0.0006	0.0062	0.0117	0.0164
1900	0.001	0.0096	0.0178	0.0249	0.0001	0.0007	0.0013	0.0018	0.0004	0.0044	0.0082	0.0116
2000	0.0007	0.0067	0.0125	0.0175	0.0001	0.0005	0.0009	0.0013	0.0003	0.0031	0.0058	0.0082

---

## C. Signal injection tests

The impact of signal in the SB on the exclusion limits is studied in simulation. For this purpose, the SR and SB regions are taken from simulation (the samples used are described in Sec. 5.1). The same selections, reconstructions, and categorisations as in Tab. 6.3 are applied. The exclusion limits are calculated from SM simulation for VLQ masses from 0.7 to 1.8 TeV with a relative width of 1% and left-handed VLQ couplings for the B+b production mode. This is done with and without B signal in the SB. For both cases the same set of uncertainties for the SM simulation and B signal are taken into account, as presented in Sec. 7. Additionally, the exclusion limits for the SR in simulation with added signal are calculated (pseudo obs. CL). The cross section of the signal is much higher than expected from theory. The results of the study are summarised in Tab. C.1, as are the injected signal cross sections. For the presented exclusion limit with and without signal in the SB the results are very similar within the  $1\sigma$  bounds, indicating that the impact of the signal is small even for very large signal cross sections. As expected, the exclusion limits with signal in the SB are slightly worse. The pseudo observed limit is of the same order as the added signal but always slightly larger, showing that the method is robust and no signal would be excluded. Similar to the presented results, the presence of signal was also tested in the validation region with the same results.

Table C.1.: Summary of the results of the study in simulation on the impact of the signal in the SB on the exclusion limits. The expected upper CL limits without (w/o Exp. CL) and with (w. Exp. CL) signal in the SB are displayed. The signal cross section used and the upper limits in the SR with signal (pseudo Obs. CL) are also presented.

Mass [GeV]	w/o Exp. CL [pb]	$-1\sigma$	$+1\sigma$	B+b $\sigma$ [pb]	Injected signal	
					w. Exp. CL [pb]	Pseudo Obs. CL [pb]
700	0.85297	0.59939	7.45053	2.1695	1.0125	2.4819
800	0.68495	0.45004	3.89045	1.508	0.97736	1.5663
900	0.34535	0.19295	2.87485	1.1095	0.49239	1.1156
1000	0.17358	0.088246	1.62962	0.8265	0.20405	0.8705
1100	0.13699	0.078245	0.99111	0.596	0.1407	0.64807
1200	0.10268	0.050496	0.61213	0.448	0.10744	0.46274
1300	0.075233	0.041562	0.152607	0.3395	0.076335	0.38088
1400	0.058682	0.023406	0.107338	0.2645	0.065294	0.30847
1500	0.047473	0.027464	0.213847	0.2075	0.054106	0.24761
1600	0.06252	0.036972	0.23814	0.1595	0.048374	0.2005
1700	0.048293	0.022399	0.158907	0.124	0.048511	0.16527
1800	0.041746	0.016891	0.118454	0.0975	0.042741	0.13494





# Danksagung

Besonders möchte ich an dieser Stellen den vielen Menschen danken, ohne die diese Arbeit nicht möglich gewesen wäre.

Insbesondere möchte ich Herrn Prof. Johannes Haller danken, der mir das Thema für die vorliegende Arbeit anvertraut hat. Darüber hinaus möchte mich für die Diskussionen, das kritische Hinterfragen, das stetige Antreiben und vor allem für die Unterstützung in einigen langwierigen und schwierigen Diskussionen bedanken. Ohne seine Unterstützung und sein Engagement wäre die Arbeit nicht veröffentlicht worden.

Frau Prof. Elisabetta Gallo möchte ich danken, dass Sie sich bereitwillig als Zweitgutachterin zur Verfügung gestellt hat. Den Mitgliedern der Prüfungskommission möchte ich danken, dass Sie die Aufgabe ebenso bereitwillig übernommen haben.

Den Post-Docs aus der CMS Arbeitsgruppe insbesondere Roman Kogler, Thomas Pfeiffer und Alexander Schmidt gebührt mein besonderer Dank. Vielen Dank das ich euch mit meinen Fragen löchern durfte, eurer Unterstützung bei Diskussionen, technischen Fragen, Präsentationen und vielem mehr. In den letzten Jahren durfte ich vieles lernen und mich weiterentwickeln.

Bei den Kollegen, mit denen ich ein Büro geteilt habe, möchte ich mich auch bedanken. Ihr habt öfter meine (wahrscheinlich nicht ganz ungerechtfertigte) Frustration ertragen und wir haben viele lustige Diskussionen geführt. Ich hoffe, ihr habt die Zeit in ebenso guter Erinnerung wie ich. Den Kollegen der CMS Arbeitsgruppe möchte ich für ihre Offenheit, Hilfsbereitschaft und Kollegialität danken, die ich immer wieder erfahren durfte.

Meinen guten Freunden, die ich während des Studiums kennengelernt habe, möchte ich hier auch danken. Ohne euch wäre ich sicher nie so weit gegangen mich an eine Doktorarbeit zu trauen. Meiner Familie, insbesondere meine Eltern Dolores und Jose, meinem Bruder Jose Manuel und meiner Verlobtin Laura muss ich besonders danken. Ihr habt mich immer unterstützt, an mich geglaubt und viel Verständnis gezeigt dafür, dass ich oft sehr wenig Zeit hatte oder mich in meinen Gedanken verliere.





# Eidesstattliche Versicherung

Hiermit erkläre ich, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Arbeit beruht auf Daten welche mit dem CMS Detektor aufgenommen wurden. Die massgeblichen Ergebnissen dieser Arbeit sind in Ref. [30] veröffentlicht.

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

---

Daniel González Vázquez