Vrije Universiteit Brussel

Faculteit Wetenschappen en Bio-ingenieurswetenschappen
Departement Natuurkunde

Search for new heavy quarks with the CMS detector at the Large Hadron Collider

Gerrit Van Onsem

Promotor
Prof. Dr. Jorgen D'Hondt
Copromotor
Dr. Petra Van Mulders

Proefschrift ingediend met het oog op het behalen van
de academische graad Doctor in de Wetenschappen

Mei 2014

Doctoral examination commission:

Prof. Dr. Alexander Sevrin (Vrije Universiteit Brussel, *chair*)
Prof. Dr. Freya Blekman (Vrije Universiteit Brussel, *secretary*)
Prof. Dr. Jorgen D'Hondt (Vrije Universiteit Brussel, *promotor*)
Dr. Petra Van Mulders (Vrije Universtiteit Brussel, *copromotor*)
Prof. Dr. Wolfgang De Meuter (Vrije Universiteit Brussel)
Prof. Dr. Christian Schwanenberger (Manchester University)
Prof. Dr. Shahram Rahatlou (INFN, Rome)
Prof. Dr. Fabio Maltoni (UCL)

Cover illustration: Different visualizations of the reconstructed particles in the CMS detector in a proton-proton collision event at a center-of-mass energy of 8 TeV. The signature is one that can be expected from an event involving the single production and the decay of a new heavy quark of mass 1.5 TeV.

Contents

Introduction **1**

1 New quarks beyond the three Standard-Model generations **3**
1.1 The Standard Model of particle physics . . . 3
1.1.1 The elementary particles and their interactions . . . 4
1.1.2 The theoretical framework of the Standard Model . . . 5
1.1.3 Shortcomings of the Standard Model . . . 13
1.2 A chiral fourth generation of quarks . . . 15
1.2.1 Motivations for a fourth generation . . . 16
1.2.2 The four-generation Standard Model . . . 17
1.2.3 Experimental and theoretical constraints . . . 19
1.3 Vector-like quarks . . . 24
1.3.1 Motivations to search for vector-like quarks . . . 24
1.3.2 Vector-like quark model . . . 26
1.3.3 Experimental constraints . . . 30

2 The Large Hadron Collider and the CMS experiment **35**
2.1 The Large Hadron Collider . . . 35
2.1.1 The LHC design and operation . . . 36
2.1.2 The physics programme at the LHC . . . 38
2.2 The Compact Muon Solenoid detector . . . 39
2.2.1 The CMS coordinate system . . . 40
2.2.2 The inner tracking system . . . 41
2.2.3 The electromagnetic calorimeter . . . 43
2.2.4 The hadronic calorimeter . . . 44
2.2.5 The muon system . . . 45
2.2.6 Data acquisition . . . 46
2.2.7 The CMS computing model . . . 47

3 Simulation and reconstruction of proton-proton collisions **49**
3.1 Collision event generation . . . 49
3.1.1 Parton distribution functions and the hard interaction . . . 51
3.1.2 Parton showering . . . 53
3.1.3 Hadronization and decay . . . 55
3.1.4 Underlying event . . . 56
3.2 Detector simulation . . . 56

3.3 Physics-object reconstruction and identification 59
3.3.1 The particle-flow event reconstruction method 59
3.3.2 Muon reconstruction and identification 60
3.3.3 Electron reconstruction and identification 63
3.3.4 Jet reconstruction 67
3.3.5 Identification of jets from b quarks 70
3.3.6 Missing transverse energy reconstruction 74

4 Search for the quarks of a sequential fourth generation 77
4.1 Model assumptions 78
4.1.1 Quark mixing 78
4.1.2 Fourth-generation processes 79
4.1.3 Masses and decay topologies 79
4.2 Data and simulation 80
4.2.1 Standard-Model background simulation 80
4.2.2 Fourth-generation signal simulation 82
4.2.3 Trigger requirements in data and simulation 82
4.3 Baseline event selection 82
4.3.1 Lepton selection criteria 85
4.3.2 Jet selection criteria 86
4.3.3 Missing transverse energy selection criteria 89
4.4 Event classification 89
4.4.1 The single-muon and single-electron decay channels 91
4.4.2 The same-sign dilepton and trilepton decay channels 95
4.5 The search variables 100
4.5.1 The S_T variable 100
4.5.2 The m_{bW} variable 101
4.6 Combined search 108
4.6.1 Construction of template distributions 111
4.6.2 Systematic uncertainties 115
4.6.3 Limit setting procedure 117
4.7 Results and discussion 120
4.7.1 Fitted background model 121
4.7.2 Exclusion limits for degenerate fourth-generation quark masses . 127
4.7.3 Exclusion limits for non-degenerate fourth-generation quark masses 129
4.7.4 Additional sensitivity tests 130
4.7.5 Summary 133

5 Search for vector-like quarks decaying to light quarks 135
5.1 Vector-like quark model considerations 136
5.1.1 Coupling parameters 136
5.1.2 Production processes 137
5.1.3 Signal topologies 138
5.2 Data and simulation 139
5.2.1 Standard-Model background simulation 140
5.2.2 Vector-like quark signal simulation 140

5.2.3 Trigger requirements in data and simulation 143
5.3 Baseline physics-object selection 144
5.3.1 Lepton selection criteria 144
5.3.2 Jet selection criteria 148
5.4 Subsamples optimized for vector-like quark processes 151
5.5 Data-driven background estimations 157
5.6 Search variables 164
5.6.1 Reconstruction of the mass of the heavy quark 164
5.6.2 The S_T variable and event counts 174
5.7 Inclusive search 176
5.7.1 Construction of template distributions 176
5.7.2 Systematic uncertainties 177
5.7.3 Limit setting procedure 181
5.8 Results and discussion 181
5.8.1 Fitted background model 183
5.8.2 Exclusion limits for down-type vector-like quarks decaying to light quarks 183
5.8.3 Summary 192

6 Conclusions and outlook **193**
6.1 Combined search for the quarks of a sequential fourth generation 194
6.2 Search for vector-like quarks decaying to light quarks 195

Bibliography **199**

Summary **213**

Samenvatting **215**

Acknowledgements **219**

Introduction

The biggest quest in elementary particle physics is to uncover which fundamental building blocks constitute our universe, and how they interact with each other. The Standard Model of particle physics describes the elementary particles and their interactions in a beautifully mathematical and consistent way. Most remarkably, this theory made an astonishing amount of predictions that were verified experimentally with unimaginable precision. The success of the Standard Model reached new heights with the discovery of the Brout-Englert-Higgs boson in July 2012 by the CMS and ATLAS experiments at the Large Hadron Collider (LHC) at the CERN laboratory in Geneva, Switzerland. The existence of this particle was predicted in the Standard Model about 50 years ago in order to explain the masses of elementary particles.

The Standard Model may provide an excellent description of subatomic particles and their interactions, but it cannot serve as a theory of everything. For instance, it does not incorporate gravity, and astronomical and cosmological observations of dark matter and dark energy revealed that we do not understand what 95% of the universe is made of. Moreover, and most relevant to the work presented in this thesis, the Standard Model does not predict exactly the fundamental fermion matter content; the quarks and leptons. Matter can be categorized in a definite structure, involving three copies, 'generations', of quarks and leptons. Each generation contains similar elementary particles, only differing in mass. The arrangement of matter in exactly three of such generations is considered to be a mystery in particle physics, since no mechanism prevents the existence of other quarks beyond the three known generations.

In the first chapter of this thesis, the known fundamental particles and forces and how they fit into the Standard Model, will be briefly described. The extension to a fourth chiral generation of fermions, as well as the less straightforward concept of vector-like quarks, will be introduced, emphasizing the motivation to search for these particles. Chapter 2 describes the experimental setup for the searches for new quarks; the LHC colliding protons at huge center-of-mass energies, and the Compact Muon Solenoid (CMS) detector recording the particles emerging from these collisions. Crucial in our search for new quarks is the understanding and simulation of the established Standard-Model and the hypothesized new-physics processes. The generation and simulation of proton-collision events and the actual reconstruction of the particles interacting with the detector into physics objects used for analysis, will be discussed in Chapter 3. Next, in Chapter 4, we will present a novel combined search for fourth-generation up- and down-type quarks using a data set of 5 fb^{-1} of integrated luminosity, collected by the CMS experiment in 2011 at a center-of-mass energy of 7 TeV. We move on in Chapter 5 to search for vector-like quarks decaying to first-generation quarks, using

about 20 fb^{-1} of 8 TeV collision data recorded in 2012. Finally, the main conclusions of these searches are summarized in Chapter 6, with an outlook to the future.

Chapter 1

New quarks beyond the three Standard-Model generations

For centuries, people try to find patterns in nature all around them. The biological classification of animals and plants into species helped Darwin to develop his theory of evolution. Grouping chemical elements in the periodic table of Mendeleev proved to have a large predictive power about new undiscovered elements. In the 20^{th} century, particle physicists have built their own classification system to categorize the elementary particles and their mutual interactions: the Standard Model (SM). As will be discussed in Section 1.1, the most fundamental building blocks of matter, quarks and leptons, can be subdivided into three chiral generations in the mathematical framework of the Standard Model as a quantum field theory. This formulation is highly succesful in describing experimental observations, but allows for a hypothetical extension to a fourth chiral generation of quarks, as will be explained in Section 1.2. More exotic 'vector-like' quarks, introduced in Section 1.3, provide another viable extension of the Standard Model. Both of these possible additions to the quark content are of significant interest, since they can address certain open questions in the field of fundamental interactions.

1.1 The Standard Model of particle physics

The Standard Model provides an elegant description of the fundamental particles and how they interact with each other. The main theoretical ideas were developed during the 1960's and 1970's [1–3], incorporating three of the four fundamental forces in nature in one consistent framework: the electromagnetic force, and the weak and the strong nuclear interactions. In Section 1.1.1, we will present the elementary-particle content of the Standard Model, while the mathematical formulation itself will be outlined in Section 1.1.2. Although the Standard Model withstood an incredible amount of experimental tests in the last decades, it falls short of being a complete theory, as we will discuss in Section 1.1.3.

1.1.1 The elementary particles and their interactions

Matter particles are made up of particles with half-integer spin, called fermions. There exist twelve different fermions; six quarks and six leptons, and they can be grouped into three generations, as summarized in Table 1.1. With only the first generation, one can build up all ordinary matter we see around us. Protons, consisting of two up quarks and one down quark, and neutrons, consisting of two down quarks and one up quark, form positively charged atomic nuclei, with the negatively charged electrons circulating around it. The electrically-neutral neutrino's interact extremely weakly with other matter, but they can be detected in large dedicated experiments. Interestingly, the masses of the quarks and charged leptons rise with each generation, the third-generation top quark being the heaviest elementary particle known to date, with a mass of about 173 GeV [4].[1]

A particle can be 'right-handed' or 'left-handed', depending on whether its spin is along the direction of motion or opposite to it. This so-called 'helicity' of a particle turns out to be a critical property.[2] For massive particles the fundamental property is the *chirality*, related to the 'chiral' nature of the so-called weak nuclear interaction: left-handed and right-handed particles interact in a different way. This phenomenon will be discussed in more detail in Section 1.1.2. It is also important to note that each fermion has an associated anti-fermion, only differing in electric charge and handedness of spin, and they will be indicated with a bar above the particle symbol. Matter and antimatter annihilate each other when they come into contact, and the resulting energy is transformed into other particles.

		Generation		
	Charge	1	2	3
quarks	+2/3	up u	charm c	top t
	-1/3	down d	strange s	bottom b
leptons	0	electron neutrino ν_e	muon neutrino ν_μ	tau neutrino ν_τ
	-1	electron e	muon μ	tau τ

Table 1.1: Fermions come in different 'flavors' in the Standard Model, and can be categorized into three generations. The electric charges are expressed as a multiple of the absolute value of the electron charge.

The quarks and leptons interact with each other through the exchange of force-carrier particles with integer spin, called gauge bosons. The electromagnetic interaction, mediated by massless photons, is one of the most familiar forces, as it is responsible for light, chemical reactions and electromagnetic fields. The weak nuclear interaction causes unstable radioactive elements to decay, and is propagated by two massive charged W and one neutral Z boson. Finally, the strong nuclear interaction,

[1] Throughout this thesis, we will use gigaelectronvolt ($1\ \mathrm{GeV/c^2} = 1.783 \times 10^{-27}$ kg) as the unit of mass. Also *natural units* will be employed, setting $c = \hbar = 1$.

[2] For a massive particle, the helicity is not a fundamental property however, since the direction of motion and therefore the helicity may change by considering the particle in a different reference frame. For massless particles the helicity does not change, and coincides with the terminology of chirality.

carried by massless gluons, holds the quarks inside a proton together, and binds the nuclei in an atom. These three fundamental forces and their corresponding gauge bosons are summarized in Table 1.2. As outlined in Section 1.1.2, the Brout-Englert-Higgs boson (commonly known as the Higgs boson or the H boson), discovered in 2012 at the Large Hadron Collider (LHC) [5, 6], plays a special role in the Standard Model, as it is responsible for the masses of the elementary particles. The fourth fundamental force in nature, gravity, is not incorporated in the Standard Model.

	Boson	Mass (GeV)
electromagnetic force	photon γ	0
weak force	W^+,W^-	80.385 ± 0.015 [4]
	Z	91.1876 ± 0.0021 [4]
strong force	gluons g	0
	Brout-Englert-Higgs H	125.7 ± 0.4 [7]

Table 1.2: An overview of the fundamental bosons in the Standard Model. The photon, $W^{\pm}$, Z and gluons are vector bosons (spin 1), while the H boson is scalar (spin 0).

A peculiar property of quarks is that they cannot be found isolated. Instead they form stable or unstable bound states (called 'hadrons') via the strong interaction. Hadrons formed by three quarks are called baryons, and quark-antiquark bound states are referred to as mesons. The top quark, however, has such a short lifetime that, when produced, it decays before *hadronizing*. Its decay is mainly to a W boson and a b quark, and the W boson, unstable as well, subsequently decays to two lighter quarks or a charged lepton and a neutrino. If new hypothetical quarks would exist with similar or higher masses than the top quark, they are expected to be highly unstable too, and cascade to lighter particles.

1.1.2 The theoretical framework of the Standard Model

The Standard Model is formulated mathematically as a quantum field theory. This framework provides a consistent description of sub-atomic particles travelling at high velocities close to the speed of light, thus in a domain where quantum-mechanical and relativistic effects are crucial to take into account. In a quantum field theory, each particle is considered to be an excitation of an associated field $\psi(x)$, where x is the spacetime coordinate. The dynamics and kinematics of this field are dictated by the Lagrangian density $\mathcal{L}$, a function of the field $\psi(x)$ and its first derivatives $\partial^\mu \psi(x)$, where μ represents the spacetime coordinate index. The action S can then be defined as the integral of the Lagrangian density over the spacetime coordinates

$$S = \int \mathcal{L}(\psi(x), \partial^\mu \psi(x)) d^4x. \tag{1.1}$$

Following the principle of least action, $\delta S = 0$, the equations of motion of the field can be obtained. The principles for the construction of the Standard-Model Lagrangian density will be outlined in this section, and we will illustrate how invariance of this Lagrangian density under abstract symmetries result in interactions between particles.

Origin of fermion interactions

Fermions can be described as complex relativistic spin-1/2 fields called Dirac spinors. The corresponding Lagrangian density of a free fermion field is given by the sum of a kinetic term and a mass term [8]:

$$\mathcal{L}_{Dirac} = i\overline{\psi}\gamma^\mu\partial_\mu\psi - m\overline{\psi}\psi. \tag{1.2}$$

Here γ^μ represent the Dirac matrices[3], and the adjoint field $\overline{\psi} = \psi^\dagger\gamma^0$ indicates the field associated to the anti-fermion.[4] Requiring $\delta S = 0$ would result in the Dirac equation of motion, $i\gamma^\mu\partial_\mu\psi - m\psi = 0$, which is in empirical agreement with relativistic moving electrons and the fine structure of the hydrogen spectrum. In the formalism of quantum physics, only the modulus squared $|\psi|^2$ of the complex field has a physical meaning (i.e. in the case of a normalized wave function it corresponds to the probability for the particle to be found at a certain position), which serves as a motivation to require the invariance of the Lagrangian density (1.2) under so-called local phase transformations

$$\psi \to \psi' = e^{i\vec{\epsilon}(x)\cdot\frac{\vec{\tau}}{2}}\psi. \tag{1.3}$$

Here $\vec{\epsilon}(x)$ represent rotation parameters depending on x (resulting in the 'local' character of the transformation), and $\vec{\tau}$ are the generators of a Lie group.[5] In a somewhat looser terminology, Equation (1.3) is said to express the 'transformation of the field ψ under a given Lie group'. Since the derivative ∂_μ in (1.2) spoils the invariance (i.e. 'symmetry') of the Lagrangian density under a local phase transformation, replacing it with a covariant derivate D_μ can restore invariance:

$$D_\mu = \partial_\mu - ig\frac{\vec{\tau}}{2}\cdot\vec{A}_\mu, \tag{1.4}$$

with $\vec{A}_\mu$ new vector 'gauge' fields, and g a corresponding coupling strength. By construction, the transformation properties of these vector fields are such that they compensate for invariance-breaking terms coming from the regular ∂_μ derivative. Introducing the covariant derivative D_μ in Equation (1.2) results in

$$\mathcal{L} = i\overline{\psi}\gamma^\mu\partial_\mu\psi - m\overline{\psi}\psi + g\overline{\psi}\gamma^\mu\frac{\vec{\tau}}{2}\cdot\vec{A}_\mu\psi. \tag{1.5}$$

Since an extra term in the Lagrangian density will influence the equations of motion of the fermion field, we can conclude from (1.5) that new *interacting* fields have been introduced when enforcing local gauge invariance. Depending on whether the generators $\vec{\tau}$ of the Lie group are Abelian (i.e. commutative) or non-Abelian (i.e. non-commutative), different properties of the new fields arise. In the case of an Abelian Lie group, only interactions between fermions and the gauge fields are allowed, while non-Abelian groups have the special property that also couplings between the gauge fields themselves appear.

[3]The Dirac matrices are defined by the anticommutator relation $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$ with $g^{\mu\nu}$ the Minkowski metric on spacetime. In addition, we use the Einstein summation convention that implies summation over a set of repeated indices, like in Equation (1.2).

[4]The † symbol indicates the hermitian conjugate.

[5]The notation $\vec{\epsilon}\cdot\vec{\tau}$ is short for the product $\epsilon_\alpha\tau^\alpha$, where summation over the α index, running from 1 to the number of generators in the Lie group, is implied.

The fundamental forces of the Standard Model

In Section 1.1.1 we introduced the electromagnetic, weak and strong interactions with their associated bosons. These interactions arise in the SM quantum-field theory description from gauge-invariance enforcing mechanisms as explained above. The full local gauge group associated to the Standard Model is

$$G_{SM} = SU(3)_c \times SU(2)_L \times U(1)_Y, \tag{1.6}$$

the direct product of three unitary groups[6] dictating the gauge-transformation rules of matter fields. The $SU(3)_c$ group and the $SU(2)_L \times U(1)_Y$ group define the strong and the unified electroweak interaction, respectively.

- **The strong interaction**: The theory of quantum chromodynamics (QCD) describes the interaction of particles with a so-called 'color charge', like quarks. Leptons do not have a color charge and as a result do not feel the strong interaction. More technically speaking, quarks are referred to as color triplets, since they transform under the fundamental representation of the strong $SU(3)_c$ gauge group, while leptons are color singlets, because they are invariant under $SU(3)_c$ transformations. Since the $SU(3)_c$ gauge group has eight generators, eight gluon fields G_μ^α ($\alpha = 1, ..., 8$) are introduced when requiring the Lagrangian density to be symmetric under $SU(3)_c$ transformations of the fermion fields. The covariant derivate in QCD thus becomes

  $$D_\mu = \partial_\mu - ig_s \frac{\lambda_\alpha}{2} G_\mu^\alpha, \tag{1.7}$$

  with λ_α the Gell-Mann matrices[7], and g_s a measure for the strength of the strong interaction. Since $SU(3)_c$ is a non-Abelian group, the gluon fields will self-interact; they reside in a $SU(3)$ color octet. Due to this gluon self-interaction, the strong-force field between two quarks tends to increase in strength when pulling them apart, causing the quarks to be confined within composite color-neutral hadrons.

- **The electroweak interaction**: The electroweak sector represents a unified framework of the electromagnetic and the weak interaction. Although these forces appear very different at low energy scales, they could be interpreted as two aspects of a unified interaction on a higher energy scale. The spontaneous electroweak symmetry-breaking mechanism causing the forces to be so different at everyday energies will be discussed later in this section. The covariant derivative associated to the $SU(2)_L \times U(1)_Y$ gauge group is given by

  $$D_\mu = \partial_\mu - ig \frac{\sigma_\alpha}{2} W_\mu^\alpha - ig' \frac{Y}{2} B_\mu, \tag{1.8}$$

[6] A unitary group $U(n)$ contains all $n \times n$ unitary matrices, while the matrices of the special unitary group $SU(n)$ have in addition a determinant equal to 1.

[7] The Gell-Mann matrices are defined by the commutation relation $[\lambda_\alpha, \lambda_\beta] = if^{\alpha\beta\gamma}\lambda_\gamma$, with $f^{\alpha\beta\gamma}$ the completely antisymmetric structure constants of the $SU(3)$ group.

where three new vector fields W_μ^α ($\alpha = 1, 2, 3$) are introduced by the $SU(2)_L$ group, and one new vector field B_μ by the $U(1)_Y$ group. The generators of the $SU(2)_L$ group are given by the three Pauli matrices[8] σ_α. The parameter Y, called 'weak hypercharge', can be identified in terms of the electric charge Q of the fermion and the third component of the weak isospin I_3, via the relation $Y = 2(Q - I_3)$. The interaction strengths associated to the gauge groups are denoted by g and g'. A crucial experimentally verified aspect of the electroweak force is that it is parity-violating [9]. A parity transformation inverses the spatial coordinates of the fermion field, and as a consequence, its helicity. The subscript L of the $SU(2)_L$ gauge group indicates that the corresponding gauge transformation only acts on left-handed particle fields. This means that left-handed fermion fields will transform as a $SU(2)_L$ doublet, while right-handed fields will transform as a $SU(2)_L$ singlet. As partly anticipated in Table 1.1, the quarks and leptons of the Standard Model can be grouped in the following left-handed $SU(2)_L$ doublets and right-handed singlets

$$\text{quarks :} \qquad \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L, u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R$$

$$\text{leptons :} \qquad \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \nu_{e,R},\ e_R,\ \nu_{\mu,R},\ \mu_R,\ \nu_{\tau,R},\ \tau_R.$$

The physical meaning of this structure is that particle components in the same doublet can transform into each other via emission or absorption of a W boson, while right-handed components cannot transform into each other. In the original formulation of the SM, neutrinos were considered to be massless, only allowing for one possible neutrino helicity (namely left-handed, as determined by laboratory experiments). However, since they appear to have a very small but non-zero mass, their helicity is not fixed, and right-handed neutrinos should exist in nature.

The observed physical fields of the electroweak gauge bosons are not the W_μ^α and B_μ gauge fields themselves, but rather a mixture of them. The W-boson, Z-boson and photon fields are obtained by the following linear combinations:

$$W_\mu^\pm = \sqrt{\frac{1}{2}}(W_\mu^1 \mp i W_\mu^2) \tag{1.9}$$

$$Z_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu \tag{1.10}$$

$$A_\mu = \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu, \tag{1.11}$$

where θ_W is the weak mixing angle, defined as

$$\tan\theta_W = \frac{g'}{g}. \tag{1.12}$$

[8] The Pauli matrices are $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

To express the chiral nature of the weak interaction, one can write the weak coupling of a (positively) charged current $J^{\mu+}$ of chiral SM quarks as

$$\frac{g}{\sqrt{2}} J^{\mu+} W_\mu^+ = \frac{g}{\sqrt{2}} (J_L^{\mu+} + J_R^{\mu+}) W_\mu^+ \tag{1.13}$$

$$= \frac{g}{\sqrt{2}} J_L^{\mu+} W_\mu^+ = \frac{g}{\sqrt{2}} \bar{u}_L \gamma^\mu d_L W_\mu^+$$

$$= \frac{g}{\sqrt{2}} \bar{u} \gamma^\mu (1 - \gamma^5) d W_\mu^+ \tag{1.14}$$

This formula reflects the $V-A$ (*Vector minus Axial-vector*) structure of the weak coupling in the Standard Model.

The photon is observed to be massless, but the W and Z bosons do have large masses. This difference is of uttermost importance to understand the distinctive behaviour of the electromagnetic and weak forces. While the former is long-ranged due to the force-carrier being massless, the high mass of the W and Z bosons lies at the basis of the weakness of the short-ranged weak force. To incorporate the masses of the W and Z bosons, one could introduce explicit mass terms of the form $\frac{m^2}{2} W_\mu W^\mu$ in the Lagrangian density, but this would break gauge invariance due to the gauge transformation properties of vector fields. In order to resolve this problem, as well as to explain the mixture of the electroweak gauge boson fields needed to obtain the observed physical fields, a spontaneous symmetry breaking mechanism has been proposed.

The Brout-Englert-Higgs mechanism of spontaneous symmetry breaking

In order to explain masses in the Standard Model, while retaining inherent gauge invariance of the theory, a mechanism to break the $SU(2)_L \times U(1)_Y$ symmetry spontaneously has been developed in the 1960's [10–12]. The so-called Brout-Englert-Higgs mechanism requires the introduction of a new scalar $SU(2)$ doublet field, with electrically charged and neutral complex components:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.15}$$

Using the covariant derivative from Equation (1.8), the gauge-invariant terms in the SM Lagrangian density corresponding to this scalar field can be written as

$$\begin{aligned} \mathcal{L}_H &= (D^\nu \Phi)^\dagger (D_\nu \Phi) - V(\Phi) \\ &= (D^\nu \Phi)^\dagger (D_\nu \Phi) - \mu^2 (\Phi^\dagger \Phi) - \lambda (\Phi^\dagger \Phi)^2 \end{aligned} \tag{1.16}$$

with μ^2 a mass parameter and $\lambda > 0$ a measure for the strength of the field's self-interaction. Taking $\mu^2 < 0$, the potential $V(\Phi)$ has a continuously degenerate minimum for

$$\Phi^\dagger \Phi = \frac{|\mu^2|}{2\lambda} = \frac{v^2}{2}, \tag{1.17}$$

with v the non-zero vacuum expectation value of the field Φ equal to 246 GeV. Although the potential $V(\Phi)$ is symmetric under gauge transformations, once the field will be in

the non-unique vacuum state (by definition the state where the potential is minimal), the gauge invariance will be lost. When choosing the vacuum state (the following particular choice is called the 'unitary gauge'), we can make an expansion of the field around the minimum:

$$\Phi_{\mathrm{v}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{1.18}$$

Since actual particles can be described as excitations of fields, or in other words, quantum fluctuations around the vacuum state, the $H(x)$ field expanded around the vacuum corresponds to an electrically neutral spin-0 particle: the famous Brout-Englert-Higgs boson. Identifying the mass term in (1.16), we can express the mass of the H boson as $m_H = v\sqrt{2\lambda}$. Since the original field in (1.15) had four degrees of freedom (two complex fields) and we are left with only one degree of freedom in (1.18), three other degrees of freedom are 'absorbed' by the W and Z bosons in the process of choosing a specific vacuum state. We can see this by writing the covariant-derivatives term in equation (1.16) in the unitary gauge:

$$\begin{aligned} (D^\mu \Phi_{\mathrm{v}})^\dagger (D_\mu \Phi_{\mathrm{v}}) = {} & \frac{1}{2}(\partial_\mu H)(\partial^\mu H) \\ & + \frac{1}{8}(v+H)^2 g^2 (W^1_\mu + iW^2_\mu)(W^{1\mu} - iW^{2\mu}) \\ & + \frac{1}{8}(v+H)^2 (g' B_\mu - g W^3_\mu)(g' B^\mu - g W^{3\mu}). \end{aligned} \tag{1.19}$$

By identifying the mixed gauge fields with mass terms of physical $W^\pm$ and Z boson fields, we find the relations

$$m_W = \frac{vg}{2}, \qquad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2}, \qquad \text{and} \qquad \frac{m_W}{m_Z} = \cos\theta_W. \tag{1.20}$$

This means the electroweak symmetry breaking mechanism provides an elegant relation between the masses of the W and Z bosons via the electroweak interaction strengths. The W, Z and H boson masses are all proportional to the vacuum expectation value v.

The masses of fermions can now be introduced as well, although they do not arise from the covariant-derivative outlined above. Instead, to express the interactions between fermions and the scalar doublet, extra gauge-invariant 'Yukawa' terms are added in the SM Lagrangian density, of the form

$$\mathcal{L}_{\text{Yukawa}} = y_f \overline{\Psi}_L \Phi \psi_R. \tag{1.21}$$

Here y_f is the Yukawa coupling constant, $\overline{\Psi}_L$ represents a left-handed fermion $SU(2)$ doublet, and ψ_R a right-handed charged fermion $SU(2)$ singlet. In the unitary gauge (1.18), such a Yukawa term not only results in an interaction term between the fermion and the scalar field, but also in a fermion mass term with $m = vy_f/\sqrt{2}$. Note that this mass term is not gauge invariant anymore, as it is a product of a left-handed and right-handed field, that transform differently under electroweak gauge transformations.

Quark mixing and the Cabibbo-Kobayashi-Maskawa matrix

The Yukawa interactions introduce an interesting feature, subtle at first glance, but of major importance in the understanding of the quark (and lepton) flavor structure of the Standard Model. It involves the fact that the fermion mass eigenstates, i.e. the physical states of the fermions propagating through space, are not entirely the same as the weak-interaction (flavor) eigenstates that appear in the Yukawa couplings. Upon closer inspection, Yukawa interaction terms like (1.21) where only the fields of one generation appear at a time, are not the only allowed gauge-invariant terms. Focusing on the relevant terms for down-type quarks[9] as an example, the Yukawa coupling y_f can be extended to a complex 3×3 matrix in generation-space, with elements Y^d_{ij}:

$$\mathcal{L}^d_{\text{Yukawa}} = Y^d_{ij}\overline{\begin{pmatrix} u & d \end{pmatrix}_{Li}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_{Rj}, \tag{1.22}$$

where summations over the generation indices i and j are implied. After spontaneous symmetry breaking, we obtain the following mass terms for down-type quarks:

$$\mathcal{L}^d_{\text{Yukawa,mass}} = Y^d_{ij}\overline{d_{Li}}\frac{v}{\sqrt{2}}d_{Rj} = \overline{d_{Li}}M^d_{ij}d_{Rj}, \tag{1.23}$$

with M^d_{ij} the mass matrix of down-type quarks. A priori, this matrix does not have to be diagonal, but it can be diagonalized via a unitary matrix[10] V^d to obtain states with a definite mass. Equivalently, for up-type quarks a matrix M^u_{ij} and diagonalizing unitary matrix V^u can be obtained. When expressing the kinetic Lagrangian density of the weak interaction ($i\overline{\psi}\gamma^\mu D_\mu\psi$ using the covariant derivative in Equation (1.8)) in terms of the quark *mass* eigenstates instead of the *flavor* eigenstates, quark-mixing terms appear in the charged-current interaction[11] of the left-handed quarks:

$$\mathcal{L}_{\text{kinetic},W^\pm} = \frac{g}{\sqrt{2}}\overline{u_{Li}}(V^u V^{d\dagger})_{ij}\gamma^\mu W^-_\mu d_{Lj} + \frac{g}{\sqrt{2}}\overline{d_{Li}}(V^d V^{u\dagger})_{ij}\gamma^\mu W^+_\mu u_{Lj}. \tag{1.24}$$

In this equation, the d and u quarks now represent the physical mass eigenstates. The unitary 3×3 matrix $(V^d V^{u\dagger})_{ij}$ is commonly known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM}, and describes the mixing of quark mass eigenstates into weak-interaction eigenstates:

$$\begin{pmatrix} d_{\text{weak}} \\ s_{\text{weak}} \\ b_{\text{weak}} \end{pmatrix}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d_{\text{mass}} \\ s_{\text{mass}} \\ b_{\text{mass}} \end{pmatrix}_L. \tag{1.25}$$

Although a unitary $N \times N$ matrix has N^2 real parameters, not all of them are physically observable in the context of the CKM matrix. For instance, global complex

[9] The d, s and b are said to be down-type quarks because they are the 'lower' component of the $SU(2)_L$ doublet. They can be summarized by the notation d_i with i a generation index. Equivalently, one can write u_i for up-type quarks u, c and t.

[10] Unitary matrices have the property that $V^\dagger V = \mathbb{I}$.

[11] Interactions involving $W^\pm$ bosons (Z, γ and H bosons) are often named charged-current (neutral-current) interactions, referring to the electric charge transferred by the boson.

phases can be absorbed into the quark fields without changing physical properties. As it turns out, only $N(N-1)/2$ rotation angles and $(N-1)(N-2)/2$ complex phases are of physical interest. For the Standard Model with three generations, this corresponds to three quark-mixing angles and one complex phase. These parameters cannot be calculated from first principles but need to be measured.

Moreover, from Equation (1.24) we can deduce that the elements $|V_{qq'}|$ are related to the probability of a transition of a quark q to another quark q' via the weak interaction. The unitarity requirements can be summarized as:

$$\sum_k |V_{ik}|^2 = \sum_i |V_{ik}|^2 = 1 \tag{1.26}$$

$$\sum_k V_{ik} V_{jk}^* = 0 \quad (i \neq j). \tag{1.27}$$

The first relation expresses that the sum of squares of elements of a given row or a given column should be 1, consistent with the probabilistic view, while the second equation can be recognized as a constraint on three complex numbers (in the case of three generations). These numbers can be viewed as the sides of a triangle in the complex plane, the so-called *unitarity triangle*, whose area depends on the CKM phase. The existence of a non-zero complex CKM phase has an important consequence in the Standard Model, as it can be considered as a source of Charge Parity (CP) violation. A CP symmetry states that the laws of physics should stay the same when converting a particle to its antiparticle and mirroring the space coordinates. However, this symmetry appears to be violated in nature, as can be measured in studies of −for example− kaon-antikaon mixing.[12] Hence, the area of the unitarity triangle serves as a useful representation of the amount of CP violation predicted by the Standard Model.

Many experiments have been carried out to determine the absolute values of the CKM-matrix elements. These measurements include dedicated studies of meson (kaon, B and D meson) decays and meson-antimeson oscillations. The determination of $|V_{tb}|$ is possible from the measurement of the single-top-quark production cross section. The current world averages of direct measurements (hence not assuming unitarity of V_{CKM}) of the CKM-element magnitudes [4] are:

$$V_{\mathrm{CKM}} = \begin{pmatrix} 0.97425 \pm 0.00022 & 0.2252 \pm 0.0009 & (4.15 \pm 0.49) \times 10^{-3} \\ 0.230 \pm 0.011 & 1.006 \pm 0.023 & (40.9 \pm 1.1) \times 10^{-3} \\ (8.4 \pm 0.6) \times 10^{-3} & (42.9 \pm 2.6) \times 10^{-3} & 0.89 \pm 0.07 \end{pmatrix}. \tag{1.28}$$

The analogue of the CKM matrix in the leptonic sector is the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, describing the lepton mixing. The elements of the PMNS matrix are far more difficult to determine experimentally due to the difficulties of direct neutrino detection [13].

Perturbative quantum field theory and renormalization

Although the quantum field theory of the Standard Model is an elegant framework to describe in a relatively simple way the sub-atomic particles and their interactions, the

[12] A kaon is a meson containing a strange quark, while B and D mesons contain b and c quarks, respectively.

calculation of observable physical quantities is not guaranteed to be easy at all. The scattering matrix (S matrix) is a useful concept to predict quantities like production cross sections involving interactions of fields. This matrix contains the complete information about a certain collision process, and depends on the Lagrangian density of the fields involved. As long as the interaction coupling strength is not too large, the calculation of the probability amplitudes to go from one state to another can be calculated using a perturbative expansion of the S matrix. Leading-order (LO) terms in the expansion dominate over next-to-leading order (NLO) terms, and so on. The perturbative calculations can be simplified by considering a diagrammatic approach using Feynman diagrams. An example of such a pictorial representation of interactions of fields is given in Figure 1.1, where the leading-order ('tree-level') diagrams are shown for the production of a pair of top quarks via the strong interaction.

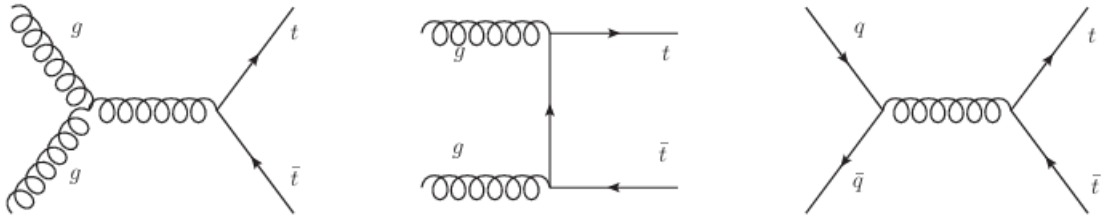

Figure 1.1: Leading-order Feynman diagrams for $t\bar{t}$ production via the strong interaction.

Quantum corrections to tree-level diagrams are then represented by higher-order Feynman diagrams, involving loops of virtual particles (i.e. particles that do not necessarily satisfy the mass-shell relation $E^2 - |\vec{p}|^2 = m^2$). Since extra loops correspond to extra interaction vertices, their contribution to physical observables is expected to be smaller as long as the coupling constant is small enough to allow for a meaningful perturbative expansion. However, in these perturbative calculations divergences often arise. For instance, when dealing with loops in diagrams one has to integrate over all possible combinations of energy and momentum of the virtual particle that could travel in the loop, sometimes resulting in divergent integrals. A particular example of an (ultraviolet) divergence is the vacuum polarization diagram shown in Figure 1.2. To treat the divergences in the Standard Model, a technique called 'renormalization' [14] is applied, where infinities are absorbed into unobservable *bare* quantities, like the bare electric charge or mass of a particle. This ensures that measurable quantities at some chosen renormalization scale are always finite.

1.1.3 Shortcomings of the Standard Model

The Standard Model is extremely successful in explaining a wide range of phenomenological observations, and has made profound predictions about the existence of new elementary particles. Nevertheless, cosmological evidence as well as a number of conceptual problems indicate that the Standard Model cannot be a complete theory of nature. The following open questions are some of the most puzzling in high-energy physics.

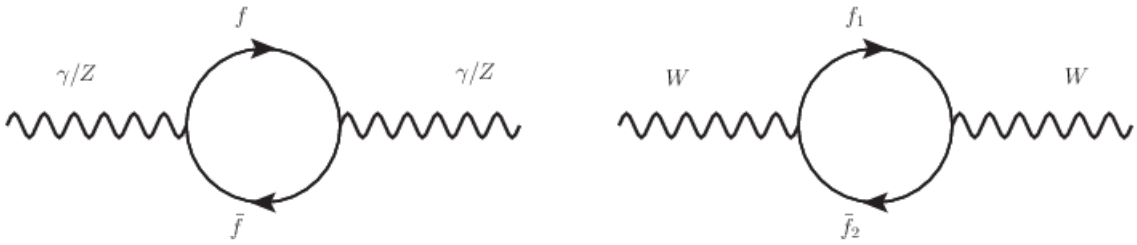

Figure 1.2: The vacuum polarization of the photon and Z boson (left) and W boson involves fermion-loop quantum corrections to the gauge fields.

- **Gravity.** The current widely accepted theory describing gravity is Einstein's theory of general relativity, attributing gravitational forces to the geometry of spacetime itself. However, it has been proven very difficult to reconcile the ideas of general relativity with quantum-physical principles, and a fully consistent 'quantum gravitational theory' has not yet been found.[13] At the currently probed energy scales of high-energy experiments, gravitational interactions between subatomic particles are assumed to be negligible. Nevertheless, in the early universe the energy density is thought to have been so high that a quantum-gravitational theory would be necessary to describe the fundamental particles and their interactions correctly.

- **Dark matter and dark energy.** Results from the Planck satellite [15] indicate that only a fraction, about 5%, of the energy density in the universe is made up by ordinary (baryonic) matter. Mysterious forms of energy, called dark matter and dark energy, represent as much as 26.8% and 68.3% of the universe, respectively. Little is known about the nature of the extremely weakly-interacting dark matter. However, because of its gravitational effects on visible matter it has a profound impact on the structure and the formation of galaxies and galaxy clusters. Dark energy is believed to be responsible for the acceleration of the expansion of the universe, but is even more enigmatic.

- **Naturalness of the Brout-Englert-Higgs boson mass.** The mass of the H boson has been measured by the CMS and ATLAS experiments to be about 125 GeV [5, 6]. The observable H-boson mass m_H is extremely sensitive to quantum corrections, mainly driven by top-quark loop effects due to the large Yukawa coupling between the H boson and the top quark. The dominating term of the quantum corrections to m_H from fermionic loops is quadratically divergent with the ultraviolet cut-off scale Λ_{UV} at which new-physics effects would become visible:

$$\Delta m_H^2 = -\frac{|y_f|^2}{8\pi^2}\left[\Lambda_{UV}^2 + \ldots\right] \tag{1.29}$$

  This is clearly problematic, because in case the Standard Model is valid up to

[13] In technical terms, general relativity is not a renormalizable theory, meaning one cannot get rid of infinities in predictions of physical observables. String theory provides a promising framework to unify all four fundamental forces of nature, as it predicts the existence of a spin-2 particle that would be the graviton, the quantum of the gravitational field.

the Planck scale (where gravity comes into play), Λ_{UV} would be about 16 orders of magnitude higher than the electroweak symmetry-breaking scale. Such a huge correction requires an incredible fine-tuning of the (unobservable) bare H-boson mass $m_{H,0}$, since $m_{H,\text{observable}} = m_{H,0} + \Delta m_H$. Several proposed extensions of the SM, like supersymmetry, provide new mechanisms or new particles at the TeV scale to cancel the fermionic corrections to the Higgs mass.

- **Disequilibrium between matter and antimatter.** While the Big Bang should have produced an equal amount of baryonic matter and antibaryonic matter, this is not measured in our observable universe. While most of the primordial matter and antimatter annihilated each other, a slight imbalance caused a fraction of matter to survive. Within the Standard Model, some asymmetry in the production rate of matter and antimatter could be induced by the CP-violating nature of the weak interaction. However, the amount of CP violation needed for the observed baryon asymmetry in the universe is about ten orders of magnitude larger than can be explained from measurements of the Standard Model. An additional puzzling observation, known as the *strong CP problem*, is that the strong interaction does not seem to violate the CP symmetry, while terms with a CP-violating phase (the QCD vacuum angle) are naturally allowed in the QCD Lagrangian.

- **Three generations of quarks and leptons.** The Standard Model seems to have a certain level of arbitrariness. If neutrinos are allowed to have a mass, the theory contains 26 parameters[14] whose value are not theoretically predicted, and have to be experimentally measured. Moreover, the origin of the number of fermion generations is unknown, because the theoretical framework of the SM would be fully consistent with extra generations of fermions.

Despite the shortcomings illustrated above, the Standard Model has been tremendously successful in explaining and predicting observations in high-energy experiments during the last decades. Therefore physicists obtained a high confidence that the quantum-field theory approach is sensible at energies that can be probed today. Adding new symmetries and fields, or extra gauge-invariant and renormalizable terms to the Lagrangian density while obeying the rules of quantum field theory, serves as the basis for numerous models beyond the SM that try to answer some of the open questions listed above.

1.2 A chiral fourth generation of quarks

When the muon, discovered in 1936 in cosmic-ray experiments, was identified as just a heavy short-lived copy of the electron, this came as a big surprise. Nobel laureate Isidor Isaac Rabi's famous reaction was "Who ordered that?", because the muon seemed not to play any role in nuclear physics. After the establishment of the three generations

[14] Six quark masses, six lepton masses, eight parameters related to the CKM and PMNS mixing matrices, three gauge coupling constants, the scalar self-interaction and vacuum expectation value, and the QCD vacuum angle.

of quarks and leptons throughout the 20$^{\text{th}}$ century, one can ask the question "Who didn't order more?". Not only is adding a fourth generation of fermions allowed in the Standard Model, there are several other well-motivated reasons to search for them in high-energy experiments, as will be outlined in Section 1.2.1. The actual way to extend the SM with a new generation is relatively trivial, and will be described in Section 1.2.2. Finally, some of the main indirect and direct constraints on the existence of a fourth generation are summarized in Section 1.2.3.

1.2.1 Motivations for a fourth generation

A chiral fourth generation could address some of the outstanding challenges in high-energy physics [16], of which some are listed below.

- **Possible explanation of matter-antimatter asymmetry in the universe.** As will be explained in more detail in Section 1.2.2, a chiral fourth generation would require the extension of the CKM quark-mixing matrix to a 4×4 matrix. It can be shown [17] that the amount of CP violation in the Standard Model depends on the area of the unitarity triangle as well as on the mass differences[15] between like-charged quarks. In a four-generation SM, the unitarity triangle equation (1.27) would involve an additional term, resulting in a larger unitarity-"triangle" area[16], enhancing the amount of CP violation [18]. Moreover, extra factors from the mass difference between heavy fourth-generation quarks and lighter-generation quarks could increase the amount of CP violation by many orders of magnitude. The combined effect might bridge the ten orders of magnitude needed to explain the baryon-antibaryon asymmetry in the universe.

- **Insights in fermion-mass hierarchy.** The scalar H field may provide a consistent mechanism to add fermion mass terms in the Standard Model, but the observed *hierarchy* of fermion masses remains a mystery. First-generation quarks and charged leptons are lighter than second-generation quarks and charged leptons, which are in turn lighter than their third-generation counterparts. This can be considered rather unnatural, suggesting a more *flavor-democratic* structure at higher energy scales, with some dynamical mechanism yielding the observed mass spectrum at low energy. In such a scenario, the actual fermion-Higgs Yukawa couplings of a given fermion type (e.g. down-type quarks) could be of the same order, resulting in a single eigenvalue of the mass matrix [19, 20]. Small perturbations to such a flavor-democratic mass matrix would then result in the observed masses of the first three generations. It turns out that the mass differences between the third-generation charged fermions would not allow this perturbative approach for a 3×3 mass matrix, while it would be valid in the context of a fourth generation with reasonably degenerate fermion masses. Moreover, light neutrino masses of the first three generations would be favored, consistent with observations.

[15] This might be intuitively understood from the fact that, if all quark masses would be equal, mass eigenstates and weak-interaction eigenstates would coincide, removing the need for a CKM mixing matrix.

[16] In fact, the triangle becomes a quadrangle.

- **New perspectives on electroweak symmetry breaking (*pre H-boson discovery argument*).** Before 2012, the existence of a new scalar H field related to electroweak symmetry breaking was not established. Since virtual loops from heavy fourth-generation quarks would influence the strengths of the coupling constants at different energy scales (referred to as the *running* of coupling constants via Renormalization Group Equations), the range of m_H allowed by electroweak precision measurements would get affected. Masses of the H boson as high as 600 GeV would become consistent with electroweak data [21], while the constraints within a three-generation Standard Model were much tighter and favored a light H boson. Electroweak symmetry breaking could even be realized by a condensate of fourth-generation quarks, without invoking the existence of an H field [22]. Of course, these arguments were much more powerful before the discovery of the Brout-Englert-Higgs boson with $m_H \approx 125$ GeV. Nevertheless, the interesting interplay between a fourth generation and the electroweak symmetry breaking mechanism did serve as a motivation to search for fourth-generation quarks in the early stages of data-taking at the Large Hadron Collider.

1.2.2 The four-generation Standard Model

When considering a sequential repetition of the generation structure of the SM, the fermion content and gauge transformation properties are well defined. The fourth-generation quarks and leptons should obey the chiral structure of the theory, or in other words, the left-handed components of the new quarks and leptons would transform as a $SU(2)$ doublet under the electroweak gauge group, while the right-handed components would transform as a $SU(2)$ singlet. Conventionally the new fourth-generation fermions are denoted by the symbols of the third-generation particles with a prime:

$$\text{fourth-generation quarks :} \qquad \begin{pmatrix} t' \\ b' \end{pmatrix}_L,\ t'_R,\ b'_R$$

$$\text{fourth-generation leptons :} \qquad \begin{pmatrix} \nu_{\tau'} \\ \tau' \end{pmatrix}_L,\ \nu'_{\tau,R},\ \tau'_R.$$

These particles have the same quantum numbers as their lower-generation cousins, hence the up-type fourth-generation quark t' (*t-prime*) has electric charge $+2/3$ and the down-type fourth-generation quark b' (*b-prime*) has charge $-1/3$.

As explained in Section 1.1.2, the unitary CKM matrix in generation space determines the mixing of the quark mass eigenstates to form the weak-interaction eigenstates. Promoting the CKM matrix from a 3×3 matrix to a 4×4 matrix results in:

$$\begin{pmatrix} d_{\text{weak}} \\ s_{\text{weak}} \\ b_{\text{weak}} \\ b'_{\text{weak}} \end{pmatrix}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} & V_{ub'} \\ V_{cd} & V_{cs} & V_{cb} & V_{cb'} \\ V_{td} & V_{ts} & V_{tb} & V_{tb'} \\ V_{t'd} & V_{t's} & V_{t'b} & V_{t'b'} \end{pmatrix} \begin{pmatrix} d_{\text{mass}} \\ s_{\text{mass}} \\ b_{\text{mass}} \\ b'_{\text{mass}} \end{pmatrix}_L . \tag{1.30}$$

Such an extended CKM matrix has three physically independent rotation angles and two CP-violating phases more than the SM 3×3 matrix. Analogous to the CKM

matrix, the PMNS lepton-mixing matrix would be extended with an additional row and column as well.

It is instructive to note that the presence of chiral fourth-generation quarks would imply the presence of chiral fourth-generation leptons and vice versa. The reason is the cancellation of gauge anomalies within one generation of fermions. A gauge anomaly is a quantum-mechanical effect that breaks renormalizability, and would appear because of the chiral nature of the weak interaction. Axial-vector couplings like in Equation (1.14) can lead to anamolies when calculating triangle Feynman diagrams in which a closed fermion loop couples to the gauge bosons. However, when grouping the quark and lepton doublets in one generation, a striking cancellation occurs (originating from the specific fermion hypercharges), leaving the Standard Model anomaly-free and renormalizable. To maintain these well-behaved gauge properties, a fourth generation would therefore include both quarks and leptons.

Production and decay of fourth-generation quarks

Focusing on the hypothetical quarks t' and b', we can extrapolate their possible production and decay properties principally from their third-generation counterparts. As will be explained in Section 1.2.3, the mass difference between the t' and b' quark is expected from global electroweak fits to be smaller than the W-boson mass. Hence, the decays $t' \to b'W$ (or $b' \to t'W$ if $m_{b'} > m_{t'}$) will be kinematically suppressed.[17] The Standard-Model CKM matrix in (1.28) exhibits a particular diagonal structure; the magnitudes $|V_{ud}|$, $|V_{cs}|$ and $|V_{tb}|$ are significantly larger than the off-diagonal elements, and the further from the diagonal, the smaller the elements become. It is therefore natural to assume a similar 'hierarchy': $|V_{t'b'}| > |V_{t'b}| > |V_{t's}| \gtrsim |V_{t'd}|$. Since these elements influence the strength of the decay of the t' and b' quarks via the weak interaction, and decays from one fourth-generation quark to another are kinematically suppressed, the fourth-generation quarks are expected to decay to third-generation quarks via the following dominating modes:

$$t' \to bW^+, \qquad \overline{t'} \to \bar{b}W^- \tag{1.31}$$

$$b' \to tW^-, \qquad \overline{b'} \to \bar{t}W^+. \tag{1.32}$$

The nature of the couplings of the sequential fourth-generation quarks is the same as in the Standard Model. Hence, the production of t' and b' quarks in proton-proton collisions at the Large Hadron Collider can happen through the strong interaction, completely analogous to the diagrams in Figure 1.1, where the top quarks are replaced by either t' or b' quarks. In that case, the production cross section and its dependence on the heavy-quark mass can be calculated using regular quantum chromodynamics, since the fourth-generation quarks have the same possible color charges as other quarks. The charged-current weak couplings of chiral fourth-generation quarks have a $V - A$ structure (see Equation (1.14)), similar to the weak couplings of Standard-Model quarks. For significant off-diagonal CKM magnitudes in the fourth row or column of (1.30), electroweak production modes become possible, and the relevant production

[17] There will be treshold effects, however, when the mass difference is close to the W-boson mass, and decay modes to off-shell W^* bosons could become important [23, 24].

cross sections will depend on the magnitudes of these CKM elements. For sizeable values of $|V_{t'b}|$, the possible diagrams resemble those of single-top production, with a t' instead of a t quark (see Figure 1.3).

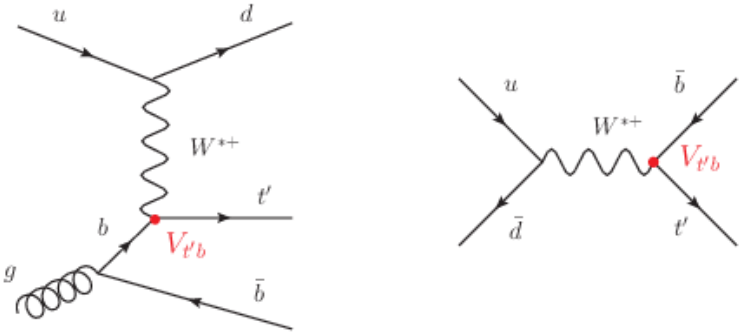

Figure 1.3: Feynman diagrams of single-t' quark production for the t-channel $2 \to 3$ (left) and s-channel (right) processes.

1.2.3 Experimental and theoretical constraints

Just like any model beyond the SM, a hypothetical fourth-generation model is subject to constraints from experimental data. This section outlines some of the main experimental and theoretical bounds to take into account, indicating the allowed parameter space.

CKM-matrix unitarity

As the magnitudes squared $|V_{ij}|^2$ of the CKM matrix elements in a given row or column should add up to one, this constrains the possible quark mixing from a fourth generation. Nevertheless, there is sufficient room for a fourth generation without violating unitarity. For instance, the sum $|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2$ from direct measurements is 0.9999 ± 0.0006 [4]. A value for $|V_{ub'}| = 0.026$ would be still allowed at the 68% confidence level, a value even larger than $|V_{ub}| \approx 0.004$. The mixing of the second generation is constrained more, since the sum of magnitudes squared in the second row in CKM matrix gives 1.067 ± 0.047, resulting in a bound on $|V_{cb'}|$ comparable to the measured value of $|V_{cb}| = (40.9 \pm 1.1) \times 10^{-3}$ [4]. This is still a reasonable restriction, given the small values of other off-diagonal CKM elements. Finally, the mixing of the fourth with the third generation is constraint the least, due to the large experimental uncertainty on $|V_{tb}|$. Without assuming unitarity of the 3×3 CKM matrix, the D0 experiment obtained a lower limit of $|V_{tb}| > 0.81$ at 95% CL from single-top production cross section measurements [25]. Therefore significant magnitudes of $V_{t'b}$ and $V_{tb'}$ are still allowed.

Electroweak precision data

The experiments at the Large Electron-Positron collider (LEP) at CERN measured the number of *light* neutrino types to be three (see Figure 1.4). As neutrinos cannot

be detected directly in collider experiments, this number was determined from the difference between the total Z-boson decay width and the decay widths to charged leptons and hadrons.[18] It is crucial to note that this measurement is only sensitive to the number of neutrino species lighter than half of the Z-boson mass, and stable fourth-generation Dirac (Majorana) neutrinos with a mass $m_{\nu'_\tau} > 45$ GeV (39.0 GeV) are still allowed [4].

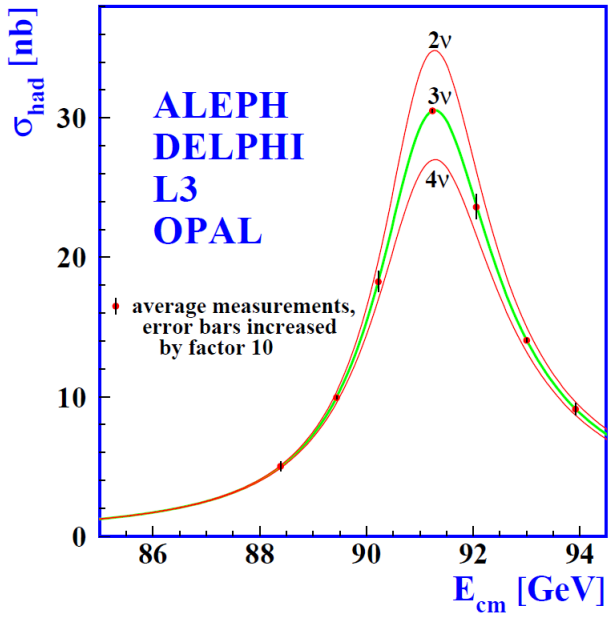

Figure 1.4: The hadron production cross section σ_{had} around the Z-boson resonance is very sensitive to the number of light neutrino species $N_{\nu,\text{light}}$ to which the Z boson can decay [26]. From this measurement, $N_{\nu,\text{light}}$ can be determined to be 2.92 ± 0.05, and a global fit to LEP data yields $N_{\nu,\text{light}} = 2.9840 \pm 0.0082$.

To parametrize potential new-physics contributions to electroweak precision observables, one can employ a set of quantities called the 'oblique parameters' S, T and U [27, 28]. They originate from new loop contributions to the vacuum-polarization functions, so they are sensitive to the effects that new particle-antiparticle pairs have on the *self-energy* of photons, Z and W bosons. It is assumed that the new physics is associated to a high mass scale and only contributes to electroweak precision observables via virtual loops. The oblique parameter S measures new-physics contributions to neutral-current processes, while T is sensitive to the difference between the contributions to neutral and charged current processes. The parameter U only affects the W-boson mass and width, and is usually small in new-physics models. The values of the oblique parameters can be obtained from a global fit to electroweak precision data [29] (taking as input a.o. measured values of masses of quarks, Z and W boson masses and widths, and the electromagnetic and strong coupling strengths). The confidence-level contours in the (S, T) plane resulting from this fit are shown in Figure 1.5, where the

[18] The decay width (or rate) Γ is defined as the probability per unit time that a given particle will decay. The total decay width of a particle is equal to the sum of the partial widths of all possible decay modes.

oblique parameters are calculated relative to a reference SM (fixing $m_H = 126$ GeV and $m_t = 173$ GeV). The oblique corrections δS, δT and δU from fourth-generation fermions are calculated in [30, 31], and the correction to S yields

$$\delta S = \frac{1}{3\pi}\left[2 + \ln\frac{m_{b'}m_{\nu'_\tau}}{m_{t'}m_{\tau'}}\right]. \tag{1.33}$$

An important observation is that δS is mainly sensitive to the mass differences within the $SU(2)_L$ doublets rather than to the masses itself. The first term in Equation (1.33) gives a large contribution of about 0.21 to S, and if the masses of all fourth-generation fermions would be degenerate, the global fit in Figure 1.5 would rule out a fourth generation when considering only the oblique parameter S. However, the constraints can be relaxed significantly by considering non-degenerate masses, and taking into account contributions to the T parameter as well.[19] From the constraints on the oblique

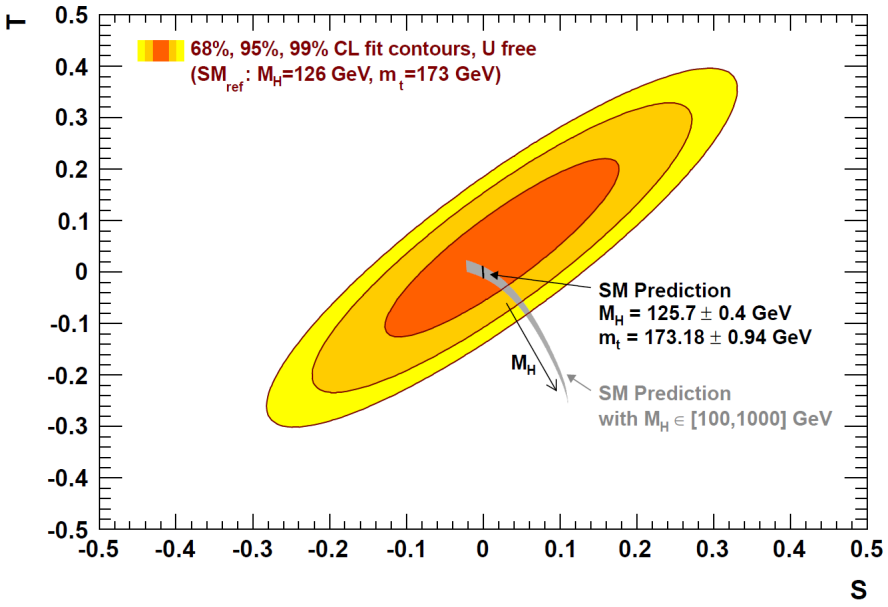

Figure 1.5: The 68%, 95% and 99% confidence level constraints (dark to light) on the S and T oblique parameters from a global fit to electroweak data, relative to a Standard-Model reference [29]. These constraints affect the possible parameter space of a fourth-generation model.

parameters, one can derive confidence level contours for the mass differences $m_{t'} - m_{b'}$ and $m_{\nu'_\tau} - m_{\tau'}$. Figure 1.6 illustrates how a relatively small mass difference $m_{t'} - m_{b'} < m_W$ is preferred, and the value of the mixing parameter $|V_{t'b}|$ has a visible influence on the constraints [24].

[19] Before the discovery of the H boson in 2012, m_H was still indetermined, hence constraints from the oblique parameters were somewhat different at the time the search presented in Chapter 4 was carried out.

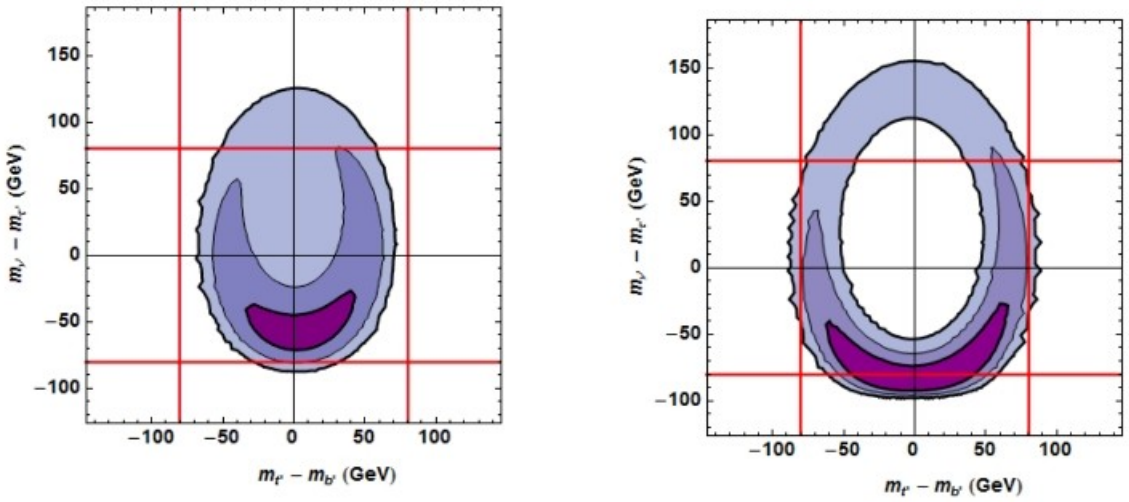

Figure 1.6: The 68%, 95% and 99% confidence level contours (dark to light) for the mass differences $m_{t'} - m_{b'}$ (x axis) and $m_{\nu'_\tau} - m_{\tau'}$ (y axis) [24]. The constraints are obtained with $|V_{t'b}| = |V_{tb'}| = 0.01$ (left) and $|V_{t'b}| = |V_{tb'}| = 0.1$ (right), no mixing between fourth-generation and first and second generation quarks, $m_H = 125$ GeV, $m'_t = 500$ GeV and $m_{\tau'} = 100$ GeV. The red lines indicate a mass difference equal to the W-boson mass.

Direct measurements

Fourth-generation fermion searches have been carried out at the LEP, Tevatron and LHC colliders. The CDF and D0 experiments at the Tevatron searched for new heavy down-type and up-type quarks in proton-antiproton collision data at a center-of-mass energy $\sqrt{s} = 1.96$ TeV, resulting in the bounds $m_{t'} > 335$ GeV [32] and $m_{b'} > 372$ GeV [33] at 95% CL. These searches assumed strong pair production of heavy quarks with subsequent decays $t' \to qW$ ($q = d, s, b$) and $b' \to tW$, respectively, and looked for an excess of events in the distributions of the reconstructed top-quark mass and/or distributions of H_T (the scalar sum of the transverse momentum of the final-state objects).

The most stringent limits on fourth-generation leptons come from the L3 experiment at LEP. Lower mass limits at 95% CL were derived for heavy charged leptons decaying to a neutrino and a W boson, resulting in $m_{\tau'} > 100.8$ GeV (102.6 GeV for a stable τ') [34]. The search for a heavy neutral lepton decaying to a charged lepton and a W boson yielded $m_{\nu'_\tau} > 90.3$ (80.5) for Dirac (Majorana) couplings. As mentioned before, from measurements at the Z-boson resonance, stable neutrinos with $m_{\nu'_\tau} \lesssim m_Z/2$ were excluded.

With the start-up of the Large Hadron Collider, a new energy frontier opened up, and the searches for fourth-generation quarks resulted in even higher limits. The CMS and ATLAS experiments searched extensively to fourth-generation quarks in $\sqrt{s} = 7$ and 8 TeV proton-proton collision data. As of February 2014, the highest limits on $t' \to bW$ and $b' \to tW$ are determined in the context of vector-like quark searches. Since the pair-production cross sections for vector-like and chiral fourth-generation

quarks are the same, limits on charged-current decay modes of vector-like quarks can be interpreted in sequential fourth-generation settings. By searching for an excess of events in the distribution of the reconstructed mass of the heavy quark, the ATLAS experiment excluded $m_{t'} < 740$ GeV at 95% CL, for a branching fraction BF($t' \to bW$) = 100% [35]. The CMS experiment searched for vector-like quarks in multilepton final states, and excluded $m_{b'} < 785$ GeV for BF($b' \to tW$) = 100% [36]. The different searches at the LHC for heavy quarks are summarized in Table 1.3. The combined search strategy for fourth-generation quarks presented in Chapter 4 resulted in the most stringent mass limits at the time.

Theoretical unitarity and perturbativity bounds

The direct searches pushing the lower mass limits in the regime of 500-800 GeV have important implications. For too large fermion masses, several theoretical issues arise. Firstly, there is the rough perturbative *unitarity bound* [37–39]. The scattering amplitudes in the S matrix (see Section 1.1.2) can be decomposed in amplitudes of partial waves. Each partial-wave amplitude must not violate unitarity (as this would contradict the probabilistic interpretation of quantum mechanics). However, for high energies, the calculation of the (hypothetical) tree-level $t'\overline{t'} \to t'\overline{t'}$ elastic scattering process can only satisfy unitarity if the t' mass is sufficiently low:

$$m_{t'} \lesssim v\sqrt{\frac{4\pi}{3}} \approx 500 \text{ GeV} \tag{1.34}$$

This assumes that the mass of the t' quark originates from a Yukawa coupling. The Renormalization-Group-Equations running of the Yukawa coupling is then responsible for the ill-defined behaviour of the scattering amplitude for high energies. The bound is in principle only indicating at which energy scale some new physics or strong dynamics should enter to save unitarity. Secondly, but somewhat connected to the first point, there is even a *perturbativity bound* itself. The Yukawa coupling for chiral fermions becomes non-perturbative at $y_f \gtrsim \sqrt{4\pi}$, and translated into a fermion mass bound, this would mean $m_f \lesssim 600$ GeV in order to maintain perturbativity. However, as mentioned before, the Yukawa coupling is not a fixed quantity, as it runs with the energy scale. In practice, this means Yukawa couplings for even lower fermion masses turn non-perturbative at some scale. The perturbative four-generation Standard Model would already break down at 1 TeV for t' and b' masses around 375 GeV [40].

Scalar boson searches

H bosons at the LHC are dominantly produced through gluon fusion. Since gluons do not couple to the H field, this production process requires a fermion loop, with the largest contribution coming from the top quark. New heavy fourth-generation quarks entering this loop would enhance the gluon-fusion H-production cross section $\sigma(gg \to H)$ by a factor between five and nine [41–43], depending on m_H, $m_{t'}$ and $m_{b'}$. As a result, probing the H-production cross section provides an indirect but strong test of the four-generation Standard Model. However, the relevant observable in the H-boson searches is $\sigma(pp \to H) \times \text{BF}(H \to X\overline{X})$, so an increased H-boson production

rate might be compensated by a smaller branching fraction to WW^*, ZZ^* and $\tau\tau$. A decrease of the branching fractions of these final states could exactly be caused by the invisible decay to fourth-generation neutrinos $H \to \nu'_\tau \nu'_\tau$, provided these neutrinos are close to their experimental bounds and about half the H-boson mass: $m_{\nu'_\tau} \sim 60$ GeV [44–46]. Moreover, because of an accidental cancellation between gauge-boson and fermion loops, the decay rate $H \to \gamma\gamma$ would be reduced in a four-generation SM and could compensate the enhancement of $\sigma(gg \to H)$ completely [47].

A global fit of the electroweak precision observables has been performed while taking into account the H-boson signal strengths in the decay modes to $\gamma\gamma$, WW^*, ZZ^*, $\tau\tau$ and $b\bar{b}$ measured by the LHC and Tevatron experiments [48]. The impact of the discovery of the H boson with $m_H \approx 125$ GeV and SM-like production and decay rates, is so profound that this fit excludes a sequential four-generation Standard Model at the 5σ level.

Clearly, the many experimental constraints outlined above do not favor the existence of a chiral fourth generation, or require the parameters to be fine-tuned to evade the existing bounds. Especially after the discovery of a light H boson in July 2012, and in combination with the extremely tight mass limits derived from indirect searches, the interest in SM-like fourth-generation quarks decreased. This did not signify the end of searches for new quarks, however, as the focus in the LHC experiments shifted more towards types of quarks that could evade many of the stringent indirect and direct constraints: vector-like quarks.

1.3 Vector-like quarks

Although the concept of vector-like quarks in quantum field theory has been studied since decades, the phenomenological interest in them increased significantly in the last few years. Vector-like quarks turn out to evade several extremely tight constraints on new heavy chiral quarks, and due to their decay properties, a potential signal could have been missed in preceding quark searches. Following a similar approach as in the previous Section, we will first motivate the search for vector-like quarks in Section 1.3.1. Next, the vector-like quark model relevant to this thesis will be described in Section 1.3.2. Finally, in Section 1.3.3 we will give an overview of the current experimental constraints on these types of quarks.

1.3.1 Motivations to search for vector-like quarks

Vector-like quarks emerge in many new-physics models, and fulfill diverse roles. Their mass eigenstates are expected to mix with those of SM quarks, possibly resulting in striking indirect and direct signatures [49, 50]. Because of their interesting phenomenological implications without invoking too much free parameters, they are very attractive for searches at the LHC. Some of the categories of models in which vector-like quarks appear are outlined below.

- **Little-Higgs models** [51, 52]. These models try to provide an explanation for the lightness of the H boson, by assuming that it originates from the breaking

of some larger symmetry group at $\mathcal{O}(10)$ TeV. Several new particles are then predicted at the TeV scale, one of which is a vector-like $T^{2/3}$ quark singlet. These new particles would couple to the H field and cancel the one-loop quadratically-divergent quantum corrections to the H-boson mass (see Equation (1.29)).

- **Composite Higgs models** [53]. Although an H boson has been observed at a mass of about 125 GeV, its nature and the complete picture of how electroweak symmetry breaking happens is not yet fully established. In composite Higgs models, the electroweak symmetry breaking might be driven by a condensate of the top quark and a vector-like partner.

- **Extra dimensions** [54]. Many models trying to unify gravity with the other fundamental forces of nature include extra warped spatial dimensions. When the SM-quark fields propagate in such a higher-dimensional space with compactified extra dimensions, a whole mass spectrum of excited states appears (a so-called Kaluza-Klein tower). The excited states of the right-handed SM-quark fields would appear to have exactly the same quantum properties as vector-like quarks.

- **Grand Unified Theories** (GUTs) [55]. These models try to unify the strong and electroweak forces into one big gauge group (like $SU(5)$, $SO(10)$ or E_6) corresponding to only one gauge coupling constant, instead of the three constants in the Standard Model. Both supersymmetric and non-supersymmetric versions of these GUTs predict the existence of vector-like states.

- **Models providing insights in the SM flavor structure** [56]. The flavor hierarchy in the SM seems unnatural, and some models postulate a higher flavor-symmetric gauge group that yields the observed fermion pattern when broken. To cancel anomalies in these gauge-group extensions, vector-like quarks can be added. Such flavor-symmetric gauge groups can also provide a solution for the strong CP problem.

In addition, vector-like quarks can explain certain anomalous observations.[20] One particular example is the $Z \to b\bar{b}$ forward-backward asymmetry observed at LEP, showing a difference of 2.8 standard deviations with respect to the SM expectation [26]. The mixing between the b quark and a vector-like partner $B^{1/3}$ can alter the $Zb\bar{b}$ vertex in such a way that the fit of electroweak precision observables actually improves [49, 57]. Moreover, just like a chiral fourth generation, new quark-mixing effects can provide a new source for CP violation to explain the matter-antimatter asymmetry in the universe.

[20] Although the Standard Model agrees very well with observations in general, a couple of three-sigma deviations from the expectation have been measured. With many hundreds of precision measurements performed in the last decades, some three-sigma deviations would be expected from a purely statistical point of view. Nevertheless, these deviating measurements are at least thought-provoking and are often used as guidance towards the development of models beyond the SM.

1.3.2 Vector-like quark model

Quarks accommodated in the Standard Model are chiral, meaning that left-handed and right-handed quark fields have different electroweak quantum numbers, or in other words, have different SM gauge-group transformation properties. One should keep in mind that the incorporation of this chiral structure was completely driven by experimental observations. Vector-like quarks are hypothetical spin-1/2 fermion fields for which the left- and right-handed components transform in the same way, and are allowed to be added to the Lagrangian density without breaking gauge invariance. They are still called *quarks* because they are assumed to have regular quark-color charges (i.e. they transform under the $SU(3)_c$ group as a triplet). The terminology of 'vector-like' quarks stems from their hypothetical *Vector* coupling (V) to the charged weak gauge bosons in Equation (1.13):

$$\begin{aligned} J^{\mu+} &= J_L^{\mu+} + J_R^{\mu+} \\ &= \bar{u}_L \gamma^\mu d_L + \bar{u}_R \gamma^\mu d_R = \bar{u} \gamma^\mu d, \end{aligned} \tag{1.35}$$

instead of a Vector minus Axial-vector structure ($V - A$) for chiral quarks. Because of the absence of axial-vector couplings, the Standard Model extended with vector-like quarks would still be anomaly free. In a chiral fourth generation, new leptons would be required to cancel gauge anomalies, but here, the presence of vector-like quarks does not necessarily imply the existence of vector-like leptons. Nevertheless, such exotic lepton types have been studied in the literature as well [58].

When vector-like quarks are required to mix with SM ones, this should happen via Yukawa couplings involving the H doublet field (this is also the case in regular SM-quark mixing). There are only a restricted amount of gauge-covariant vector-like quark multiplets associated to such renormalizable couplings [59]. These are $SU(2)_L$ singlets, doublets and triplets involving fields $U^{+2/3}$, $D^{-1/3}$, $X^{+5/3}$ and $Y^{-4/3}$, where the superscripts indicate the electric charge. The quarks with SM charges $+2/3$ and $-1/3$ can be referred to as up-type and down-type vector-like quarks, respectively. The possible weak-interaction eigenstate multiplets are

singlets:	$U_{L,R}, \quad D_{L,R}$
doublets:	$(X\ U)_{L,R}, \quad (U\ D)_{L,R}, \quad (D\ Y)_{L,R}$
triplets:	$(X\ U\ D)_{L,R}, \quad (U\ D\ Y)_{L,R}.$

Similar to the mechanism by which quark mixing occurs in the SM or in the hypothetical fourth-generation extension, the physical mass eigenstates of the SM and vector-like quarks of a certain type appear after diagonalizing the mass matrix. This time, however, the procedure is more complicated than obtaining just one 4×4 CKM matrix. Firstly, there is a dependence on the chosen multiplet scenario; for instance, in cases where up-type U and down-type D quarks don't exist simultaneously (like the singlet cases), a CKM-matrix element V_{UD} as analogue of $V_{t'b'}$ in a fourth-generation context would not make sense. Secondly, while in the SM mixing is only allowed between left-handed quarks because of the chiral nature of the weak interaction, mixing

of vector-like quarks with SM quarks can happen in the right-handed sector as well. It turns out that for new vector-like quark singlets and triplets the mixing angles in the right-handed sector are suppressed with respect to the left-handed ones by m_q/m_Q, while for new doublets the left-handed mixings are suppressed [60–62]. Thirdly, a peculiar feature of vector-like quarks is that they do not obtain their mass purely via the vacuum-expectation value of the SM H field. They are allowed to have a bare gauge-invariant mass term in the Lagrangian, because left-handed and right-handed fields transform covariantly under the gauge-group transformations:

$$\mathcal{L}_{m,Q} = -M_0 \overline{\Psi}\Psi, \tag{1.36}$$

where the form of Ψ depends on the multiplet representation. The bare masses M_0 of the components of a given multiplet are therefore the same, but the mixing with SM quarks induces relatively small mass splittings [49]. Note that from an effective point of view, the origin of this mass term is not relevant, but it could for example be generated by a Yukawa coupling to a scalar singlet that acquires a vacuum-expectation value much larger than $v = 246$ GeV.

The consequences of quark mixing are manifold. The SM couplings themselves are modified by the presence of vector-like quarks, which could result in deviations in electroweak precision observables. Also the oblique parameters get affected, but this would even be the case without mixing (since the heavy quarks would couple to gauge bosons anyway, and alter the vacuum polarization functions). The experimental constraints on the existence of vector-like quarks coming from the measurements of SM couplings will be discussed in more detail in Section 1.3.3. Furthermore, the vector-like quark mixing with SM-generation quarks induces the very interesting decay properties of these new heavy quarks, which will be outlined later in this Section. In scenarios where the up-type U and down-type D quarks mix dominantly with the third generation, they are often referred to as *top partners* or *bottom partners*, and usually denoted by T and B, respectively. In this thesis, we will mainly focus on vector-like quark mixing with the first generation (*up/down-quark partners*), with a search for down-type vector-like quarks presented in Chapter 5.

The relevant interaction terms in the Lagrangian density between first-generation quarks and up-type and down-type vector-like quarks can be written as [62, 63]:[21]

$$\begin{aligned}\mathcal{L}_{\text{interaction},U} = &\frac{g}{\sqrt{2}} W^-_\mu \kappa_{dU} \overline{d}_R \gamma^\mu U_R \\ &+ \frac{g}{2\cos\theta_W} Z_\mu \kappa_{uU} \overline{u}_R \gamma^\mu U_R \\ &- \frac{m_U}{v} H \kappa_{H,uU} \overline{u}_R U_L + \text{h.c.}\end{aligned} \tag{1.37}$$

$$\begin{aligned}\mathcal{L}_{\text{interaction},D} = &\frac{g}{\sqrt{2}} W^+_\mu \kappa_{uD} \overline{u}_R \gamma^\mu D_R \\ &+ \frac{g}{2\cos\theta_W} Z_\mu \kappa_{dD} \overline{d}_R \gamma^\mu D_R \\ &- \frac{m_D}{v} H \kappa_{H,dD} \overline{d}_R D_L + \text{h.c.}\ .\end{aligned} \tag{1.38}$$

[21] The interaction terms involving third-generation quarks would be completely analogous to (1.37) and (1.38).

Here only the terms for one helicity are written, but the others are analogous. The coupling parameters, denoted by κ, are model dependent, and originate from the mixing between SM quarks and vector-like quarks. The coupling κ_{qQ} can be reparametrized as

$$\kappa_{qQ} = \frac{yv}{\sqrt{2}m_Q}\left[1 + \mathcal{O}\left(\frac{v^2}{m_Q^2}\right)\right] = \frac{v}{\sqrt{2}m_Q}\tilde{\kappa}_{qQ}, \tag{1.39}$$

with y some model-dependent Yukawa parameter [63]. The new parameter $\tilde{\kappa}_{qQ}$ is naturally of order unity in a weakly coupled theory.

Production and decay of vector-like quarks

Just like for regular quarks, there could be various production modes for vector-like quarks. In case they are produced in pairs in proton-proton collisions, the production diagrams are again completely similar to Figure 1.1, with gluon fusion being the dominant contribution. The electroweak single production is a potentially important production mode as well, especially for heavy quarks coupling to first-generation quarks, considering the up and down valence quarks of the proton. For very large vector-like quark masses, single production can be expected to be the dominating production mode. The reasons are that it is kinematically more favorable to produce one heavy quark than two, and the parton distribution functions (that are discussed in more detail in Section 3.1.1) of the proton scale differently with energy for gluons then for quarks. To illustrate this, Figure 1.7 shows the cross section in proton-proton collisions, at a center-of-mass energy of $\sqrt{s} = 7$ TeV, for pair production and charged-current single production of heavy down-type vector-like quarks coupling to the first generation. Figure 1.8 represents example Feynman diagrams of charged and neutral current single production of a heavy vector-like quark Q.

Since the mass splitting between vector-like quarks in a given multiplet is typically $\mathcal{O}(1)$ GeV, the decay from one heavy quark into another is suppressed [49]. As a consequence, the possible decays of vector-like quarks are exclusively to Standard-Model quarks, depending on the amount of mixing with a particular SM generation, via W, Z or H bosons. Mixing with the third generation would result in the following decays of the top and bottom partners:

$$T \to\ bW^+ \,/\, tZ \,/\, tH, \qquad \overline{T} \to\ \bar{b}W^- \,/\, \bar{t}Z \,/\, \bar{t}H \tag{1.40}$$

$$B \to\ tW^- \,/\, bZ \,/\, bH, \qquad \overline{B} \to\ \bar{t}W^+ \,/\, \bar{b}Z \,/\, \bar{b}H. \tag{1.41}$$

When only mixing with the first generation is allowed, the possible decay modes are:

$$U \to\ dW^+ \,/\, uZ \,/\, uH, \qquad \overline{U} \to\ \bar{d}W^- \,/\, \bar{u}Z \,/\, \bar{u}H \tag{1.42}$$

$$D \to\ uW^- \,/\, dZ \,/\, dH, \qquad \overline{D} \to\ \bar{u}W^+ \,/\, \bar{d}Z \,/\, \bar{d}H. \tag{1.43}$$

From a phenomenological point of view, the decays of U and D are substantially different from the decays of T en B. The presence of top quarks, subsequently decaying to a b quark and a W boson, and of bottom quarks, will result in a high b-quark multiplicity and a potentially large amount of leptons in the final state. In a particle detector, the quarks will be detected as a collection of hadronized particles called jets,

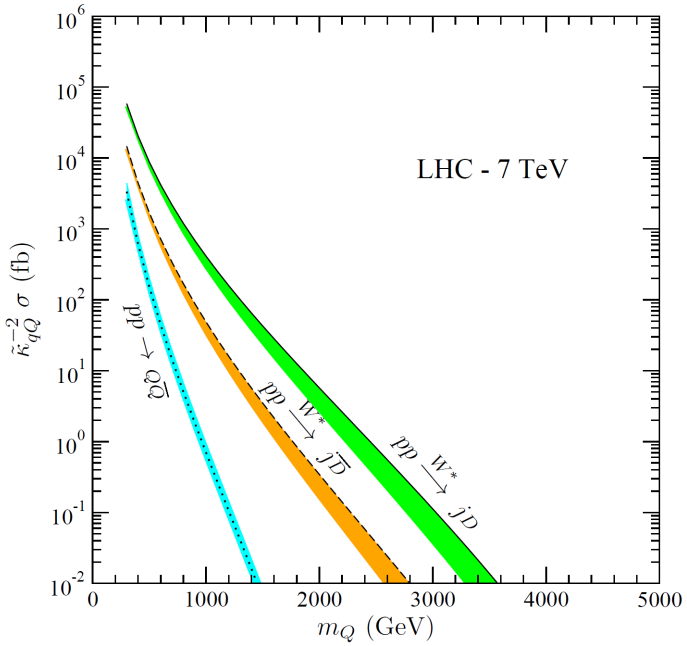

Figure 1.7: For high quark masses, the electroweak single-production mode of vector-like D quarks would dominate over the strong pair production [63]. Note that for the pair production, the y axis just represents the cross section σ, independent of $\tilde{\kappa}_{qQ}$.

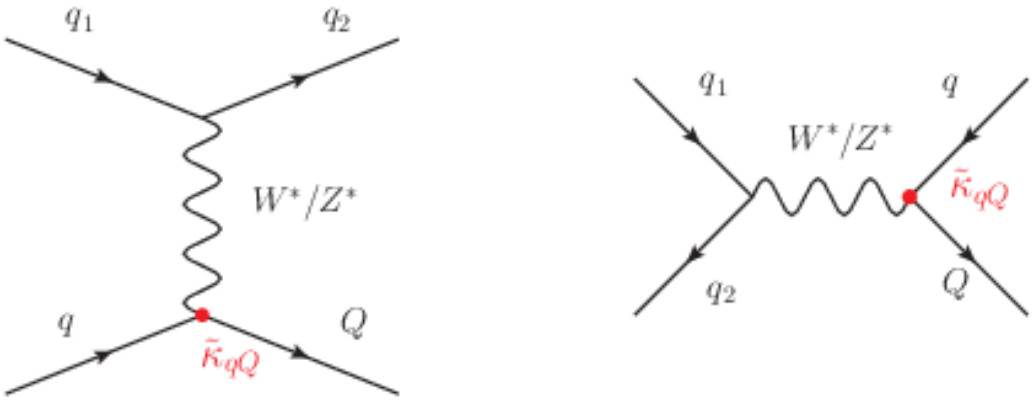

Figure 1.8: A vector-like quark Q coupling to quarks of the first generation can be produced singly through charged-current or neutral-current interactions, via t-channel (left) and s-channel (right) processes.

and a jet originating from a b quark can be identified via dedicated algorithms (see Section 3.3.5). Note that the only possible decay modes of the vector-like quarks with exotic charges are $X \to u_i W^+$ and $Y \to d_i W^-$, with u_i and d_i up-type and down-type SM quarks of generation i, respectively.

The branching ratios of the different decay modes would depend on the multiplet in which the vector-like quark resides [49]. In most models the neutral-current branching fractions $\mathrm{BF}(Q \to Zq)$ and $\mathrm{BF}(Q \to Hq)$ tend to be roughly of the same size, with $\mathrm{BF}(Q \to Wq)$ allowed to vary from 0 to 1 depending on the multiplet representation. Moreover, there is some dependency of the branching fractions on kinematic effects, because of the different masses of the decay products. This dependency tends to disappear for very high quark masses. The main restriction independent of the details of the exact model, is that the branching fractions should add up to one:

$$\mathrm{BF}(Q \to Wq) + \mathrm{BF}(Q \to Zq) + \mathrm{BF}(Q \to Hq) = 1. \tag{1.44}$$

1.3.3 Experimental constraints

Several of the indirect and direct constraints on chiral fourth-generation quarks are applicable to vector-like quarks as well. However, due to the different nature of their couplings to W and Z gauge bosons, the distinctive way of acquiring mass, and the broad decay possibilities, experimental bounds on vector-like quarks should be carefully evaluated.

CKM-matrix unitarity

A consequence of the mixing of vector-like quarks with SM quarks is that the measured 3×3 CKM matrix would not be unitary. The precise amount of deviation strongly depends on the multiplet representation of the new quarks, however. When considering a single vector-like quark representation, either the left-handed or the right-handed mixings will be suppressed by the ratio m_q/m_Q. Hence, in models with only sizeable right-handed couplings, the familiar CKM matrix, describing the mixing of left-handed fields, will not be severely affected. Nonetheless, if both U and D quarks exist, a CKM matrix will appear in the right-handed sector too.

It is important to note that in the Standard Model, the unitarity of the CKM matrix is required to suppress flavor-changing neutral current (FCNC) processes. At tree level those processes are forbidden because the Z boson couples to a linear combination of the electromagnetic current (which is flavor diagonal, i.e. where no mixing is apparent) and the current associated to the third component of the weak isospin, which is only flavor diagonal if the CKM matrix is unitary [64]. At higher orders FCNC processes are suppressed via the Glashow-Illiopoulos-Maiani (GIM) mechanism, which also requires a unitary CKM matrix [65]. Therefore vector-like quarks not only lead to tree-level FCNC couplings between SM quarks, but they might also spoil the GIM mechanism.

Electroweak precision data

The presence of new vector-like quarks induces modifiations of precisely measured observables at tree and loop level. Some qualitative descriptions of the main observables

that are sensitive to these new heavy quarks are summarized below [50].

- **Zq$\bar{\text{q}}$ couplings.** The experiments at the LEP collider performed precise measurements of the left-handed and right-handed $Zq\bar{q}$ couplings via studies of hadronic cross sections and asymmetries in $e^+e^- \to Z \to q\bar{q}$ processes [26]. The measured $Zc\bar{c}$ and $Zb\bar{b}$ couplings result in strong bounds on the tree-level mixing of vector-like quarks with the SM quarks. The contributions of vector-like quarks are only at the tree level if there is mixing with the SM quarks (e.g. $B-b$ mixing affecting $Zb\bar{b}$), and at the loop level in the absence of mixing (e.g. the effect of T on $Zb\bar{b}$).

  The most stringent bounds on couplings between vector-like quarks and first-generation quarks come from the measurement of the weak charge of the Cesium and Tallium atoms (Atomic Parity Violation) [66]. These measurements are very sensitive to $Zu\bar{u}$ and $Zd\bar{d}$ couplings, resulting in strong bounds on possible flavor-conserving neutral-current couplings. For instance, a stringent bound $\kappa_{uD} \approx \kappa_{uU} \lesssim 0.07$ can be obtained in a model with a single vector-like quark doublet $(U\ D)$ [63]. Measurements of meson mixing and decays constrain vector-like quark contributions as well. For instance, while the transition from a D^0 meson (formed by a charm and an up quark) to its antiparticle only occurs at loop level in the SM, tree-level $Zc\bar{u}$ couplings induced by vector-like quark mixing can contribute to the transition rates. Also additional loop effects in kaon-antikaon transitions might be generated in the presence of vector-like quarks.

- **Rare top-quark decays.** FCNC top-quark decays (e.g. $t \to Zq$) are extremely rare in the Standard Model and only happen via loop-induced diagrams, with $t \to Zq$ branching ratios of the order of 10^{-16} to 10^{-14}. Vector-like quarks mixing with the top quark would result in possible tree-level FCNC decays, with the rate depending on the amount of mixing, but able to reach $t \to Zq$ branching fractions up to 10^{-4} in a singlet scenario [67]. The current experimental reach of 5×10^{-4} for the $t \to Zq$ branching fraction [68] is competitive with such predictions. Vector-like quarks might also contribute at loop level to FCNC decays $t \to gq$, destroying the GIM suppression. Note that when the vector-like quarks would not mix with the top quark, these rare top-quark decays would not get affected.

- **Oblique parameters.** Loop-level constraints are much more model dependent than tree-level constraints, since cancellations can occur depending on the exact particle content and thus the actual representation and even the amount of vector-like quark multiplets that are present. Nevertheless, the new states would modify the vacuum-polarizations of electroweak gauge bosons and contribute to the S, T and U oblique parameters. Studies of the effect of vector-like singlets and doublets on the oblique parameters are performed for example in [69]. It is observed that the constraints on the oblique S parameter would still allow vector-like quarks easily up to a few TeV, while the measurement of the T parameter is more sensitive and puts constraints on both the mixing and mass.

In the literature as well as in experimental searches, vector-like quarks are often considered to mix mainly with the top and bottom quarks. This assumption is partly

driven by the hierarchical structure of the Yukawa couplings, but also by the more stringent indirect experimental constraints in the first and second generation sectors compared to the top-quark sector. From the low-energy precision measurements, a generic bound on the couplings κ in Equations (1.37) and (1.38) of the order of a few percent can be expected, without taking into account possible cancellations among contributions of different vector-like quarks [63]. However, such a bound can partially or completely disappear in specific models. For example, in some scenarios with two degenerate vector-like quark doublets (i.e. with the same mass and mixing to the first-generation quarks), the mixing with the SM quarks would not induce observable corrections to the SM quark couplings due to cancellations of vector-like quark contributions with opposite signs. The new quarks would still contribute to the oblique parameters, but only result in a bound $\kappa \lesssim \mathcal{O}(0.5)$, considerably weaker than what would be naively expected from precision measurements. Also partial cancellations in case of non-exact degeneracy can occur. Mixings between vector-like quarks and multiple generations at the same time should be considered as a possibility [62]. Therefore, sizable values of κ and hence $\tilde{\kappa}$ should still be taken into account in a model-independent search.

Brout-Englert-Higgs boson searches

The unitarity and perturbativity bounds originating from the Yukawa coupling to the H field are not directly applicable to vector-like quarks, since their heavy masses do not originate from such a coupling, but from the bare gauge-invariant mass term (1.36). This does not mean they do not couple to the Higgs field. Unlike chiral fourth-generation quarks, however, vector-like quarks are not expected to contribute significantly to Higgs-boson production and decay modes. One of the reasons is that vector-like quarks *decouple* from the H field when the mass parameter M_0 becomes large. When considering a T quark coupling to the top quark, there would be almost no modifications in the $gg \to H$ and $H \to \gamma\gamma$ rates due to cancellations between the T loop and the modified top-quark loop [49]. For vector-like quarks mixing with lighter quarks, these specific cancellations would not be present, but the decoupling effect still ensures that the modifications are relatively small. For the maximum allowed mixing from electroweak precision measurements, a calculation of the correction to the $gg \to H$ cross section due to a vector-like B quark yields an increase in the rate of about 6%, while the H-boson decay mode to $b\bar{b}$ would decrease with a similar fraction. The predicted deviations from the SM rates are still well below the current precision achievable by the CMS and ATLAS experiments.

Direct measurements

Most direct searches for new heavy vector-like quarks assume dominant mixing with third-generation quarks. Searches for T and B decaying via neutral-current interactions at the Tevatron and the LHC have been performed from early on, while results for charged-current decays could be reinterpreted from the regular chiral fourth-generation quark searches. However, these analyses often assumed for simplicity 100% branching ratios of one particular decay mode, which might not represent the actual expected

vector-like quark signals. Hence, large regions of phase space were not covered in these early searches. Since the last few years heavy-quark searches at the CMS and ATLAS experiments therefore try to be as model independent as possible, by scanning the three possible branching fractions in a so-called *branching-fraction triangle*. This makes sense because of the relation 1.44, and since the discovery of a SM-like H boson, the possibility of significant decay fractions to H bosons should be considered. Masses m_T up to 687 GeV and for some branching fractions even up to 800 GeV have been excluded at 95% CL for the up-type vector-like quarks coupling to third-generation quarks [35, 70]. For down-type vector-like quarks coupling to third-generation quarks, the lower bound on the mass is between 582 GeV and 785 GeV [36, 71].

Searches for vector-like quarks coupling to quarks of the first or second generation are complementary to these top-partner and bottom-partner searches. The most stringent limits on vector-like quarks in a light-quark partner scenario have been determined by the ATLAS experiment, excluding masses $m_U < 1080$ GeV for neutral-current single production and decay and $m_D < 1120$ GeV for charged-current single production and decay [72]. These limits are quoted for a coupling parameter $\tilde{\kappa} = 1$, but mass limits for smaller $\tilde{\kappa}$ can be deduced from these results as well. Still, in each of these cases, the branching fraction of one particular decay mode is assumed to be 100%. The constraints on the masses of fourth-generation and vector-like quarks by the CMS and ATLAS experiments at the LHC are summarized in Table 1.3.

Table 1.3: An overview of the searches for new heavy chiral (t' and b') or vector-like (T/U and B/D, and X) quarks at the LHC. Limits indicated with (†) are indicative, as they depend on the branching fractions to different bosons. The result indicated with ($\star$) is from the analysis presented in Chapter 4. It assumes deviations from degeneracy of the t' and b' quarks up to 25 GeV, and the limits depend on the value of the assumed extended CKM-matrix elements. Note that some of the results in this table are preliminary, others are published. Only the most stringent limits per process, channel and experiment are shown.

Process	Channel	Limit (95% CL)	Data
pair $t' \to (q = d, s, b)W$	2ℓ	350 GeV	ATLAS 1.04 fb^{-1}, 7 TeV [73]
pair $t' \to bW$	2ℓ	557 GeV	CMS 5.0 fb^{-1}, 7 TeV [74]
	ℓ+jets	570 GeV	CMS 5.0 fb^{-1}, 7 TeV [75]
	ℓ+jets	656 GeV	ATLAS 4.7 fb^{-1}, 7 TeV [76]
pair $b' \to tW$	ℓ+jets	675 GeV	CMS 5.0 fb^{-1}, 7 TeV [77]
	ℓ+jets	480 GeV	ATLAS 1.04 pb^{-1}, 7 TeV [78]
	$\ell^\pm\ell^\pm$ and 3ℓ	611 GeV	CMS 4.9 pb^{-1}, 7 TeV [79]
	$\ell^\pm\ell^\pm$	720 GeV	ATLAS 14.3 fb^{-1}, 8 TeV [80]
pair+single $t' \to bW$ pair+single $b' \to tW$	ℓ+jets, $\ell^\pm\ell^\pm$, 3ℓ	685^{+20}_{-20}-750 GeV ($\star$)	CMS 5.0 fb^{-1}, 7 TeV [81]
pair $T \to tZ$	ℓ+jets	625 GeV	CMS 5.0 fb^{-1}, 7 TeV [77]
	2ℓ	350-750 GeV (†)	ATLAS 14.3 fb^{-1}, 8 TeV [82]
	3ℓ	475 GeV	CMS 1.14 fb^{-1}, 7 TeV [83]
pair $T \to tH$	ℓ+jets	350-850 GeV (†)	ATLAS 14.3 fb^{-1}, 8 TeV [84]
pair $T \to tZ/tH$	$\ell^\pm\ell^\pm$	540 GeV (singlet)	ATLAS 14.3 fb^{-1}, 8 TeV [80]
pair $T \to tZ/tH/bW$	$\geq 1\ell$	687-782 GeV (†)	CMS 19.6 fb^{-1}, 8 TeV [70]
	ℓ+jets	600-800 GeV (†)	ATLAS 14.3 fb^{-1}, 8 TeV [35]
pair $B \to bZ$	2ℓ	550 GeV	CMS 5.0 fb^{-1}, 7 TeV [85]
	$2e$	400 GeV	ATLAS 2.0 fb^{-1}, 7 TeV [86]
pair $B(X) \to tW$	$\ell^\pm\ell^\pm$	670 GeV	ATLAS 4.7 fb^{-1}, 7 TeV [87]
pair $X \to tW$	$\ell^\pm\ell^\pm$	800 GeV	CMS 19.5 fb^{-1}, 8 TeV [88]
pair $B \to bZ/tW$	2ℓ	450-700 GeV	CMS 19.6 fb^{-1}, 8 TeV [89]
pair $B \to bZ/tW/bH$	ℓ+jets	582-732 GeV (†)	CMS 19.8 fb^{-1}, 8 TeV [71]
	$\geq 3\ell$	520-785 GeV (†)	CMS 19.5 fb^{-1}, 8 TeV [36]
	$\ell^\pm\ell^\pm$	590 GeV (singlet)	ATLAS 14.3 fb^{-1}, 8 TeV [80]
single $D(X) \to uW$	ℓ+jets	1120 (1420) GeV	ATLAS 4.66 fb^{-1}, 7 TeV [72]
single $U \to uZ$	2ℓ	1080 GeV	ATLAS 4.66 fb^{-1}, 7 TeV [72]

Chapter 2

The Large Hadron Collider and the CMS experiment

In order to explore the fundamental matter and interactions at high energies, physicists and engineers build powerful particle accelerators. Inside the accelerator, beams of particles are collided at a huge center-of-mass energy, from which new heavy particles may be created. Large particle detectors are built around the beam-collision points, to detect the particles that emerge from the collisions. The properties of these detected particles provide insights in the elementary particles and interactions that took part in the collision, and hence enable physicists to study the fundamental structure of matter in our universe.

The Large Hadron Collider (LHC) [90] at CERN (the European Organization for Nuclear Research) is the world's largest particle accelerator, and collides protons at the highest energies ever achieved in laboratory conditions. The design of the accelerator complex and the operation of the LHC is outlined in Section 2.1. One of the huge particle detectors that is built around a proton-beam collision point, is the Compact Muon Solenoid (CMS) detector. As described in Section 2.2, this detector is composed of several subdetectors, constructed in layers around the collision point. By measuring the hits and the energy deposits of particles traversing the detector material, individual particles can afterwards be reconstructed and identified, in order to obtain a complete as possible picture of a proton collision event.

2.1 The Large Hadron Collider

The LHC is part of an accelerator complex built at CERN, the world's largest particle physics laboratory, situated in Geneva, Switzerland. The complex consists of several interconnected linear and circular accelerators, as depicted in Figure 2.1, that gradually increase the energy of the protons before they are injected in the LHC ring. The consecutive accelerator rings increase in radius, to reduce the energy loss of the protons by synchrotron radiation. This radiation is proportional to the fourth power of the proton energy and inversely proportional to the bending radius. The acceleration chain starts with a linear accelerator (LINAC 2), which increases the energy of the protons up to 50 MeV before injecting them in the Proton Synchrotron Booster. Next, this

accelerator delivers the protons at an energy of 1.4 GeV to the Proton Synchrotron (PS), which accelerates the protons to 26 GeV. Finally, the Super Proton Synchrotron (SPS) injects two beams of protons with an energy of 450 GeV in opposite direction in the Large Hadron Collider. Here the protons are accelerated to 3.5 TeV (in 2010 and 2011) or 4 TeV (in 2012).

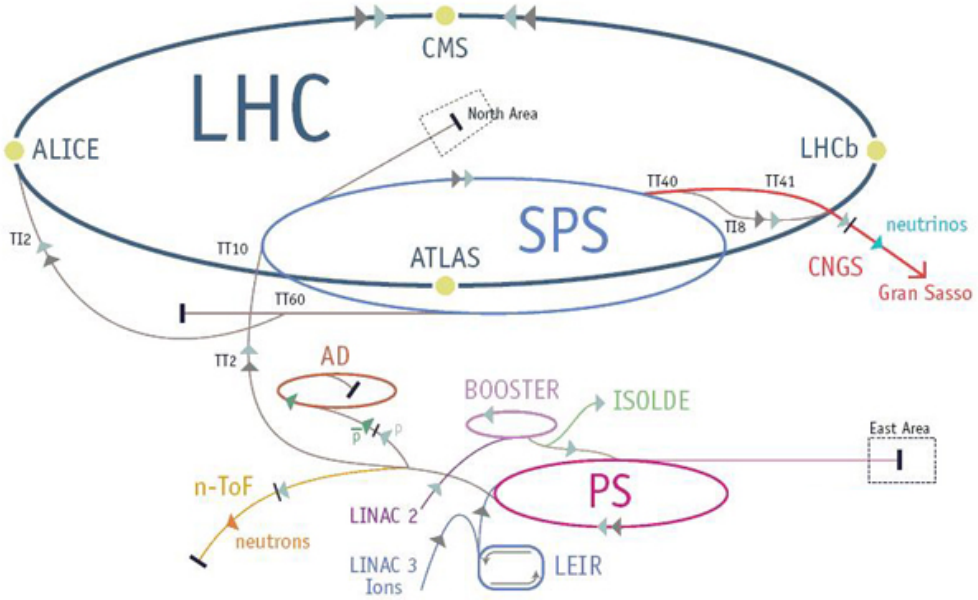

Figure 2.1: The CERN accelerator complex consists of a chain of linear and circular pre-accelerators that increase the energy of the protons step by step. The proton beams are finally injected in the LHC ring, and collided at the interaction points where particle detectors, such as the CMS detector, are located.

2.1.1 The LHC design and operation

The Large Hadron Collider is placed inside a circular tunnel with a circumference of about 27 km, at a depth ranging from 50 to 175 metres underground. This tunnel was home to the Large Electron Positron collider (LEP), which collided electrons and positrons (anti-electrons) from 1989 until 2000. The construction of the LHC was completed in 2008, and after a commissioning phase, the first proton collisions at a center-of-mass energy of $\sqrt{s} = 7$ TeV began in 2010. Before the start-up of the LHC, the previous collision energy record was held by the Tevatron collider at Fermilab, colliding protons with antiprotons at $\sqrt{s} = 1.96$ TeV.

The LHC consists of 1232 dipole magnets of about 15 m in length, that are designed to bend the protons into orbit. A schematic representation of the cross section of a dipole magnet is shown in Figure 2.2. The two proton beams circulate in opposite directions in two separate beam pipes inside the magnet. A strong electric current in the coils around the beampipe generate the magnetic fields to bend the protons

in the required orbits. By surrounding the magnet structure with a vessel filled with liquid Helium, the magnet is cooled down to only 1.9 K. At this temperature, the coil becomes superconducting and can generate, with the aid of an iron return yoke, a strong magnetic field of 8.3 T. In order to stabilize and focus the proton beams, over 8000 other higher-order multipole and corrector magnets are placed along the LHC tunnel.

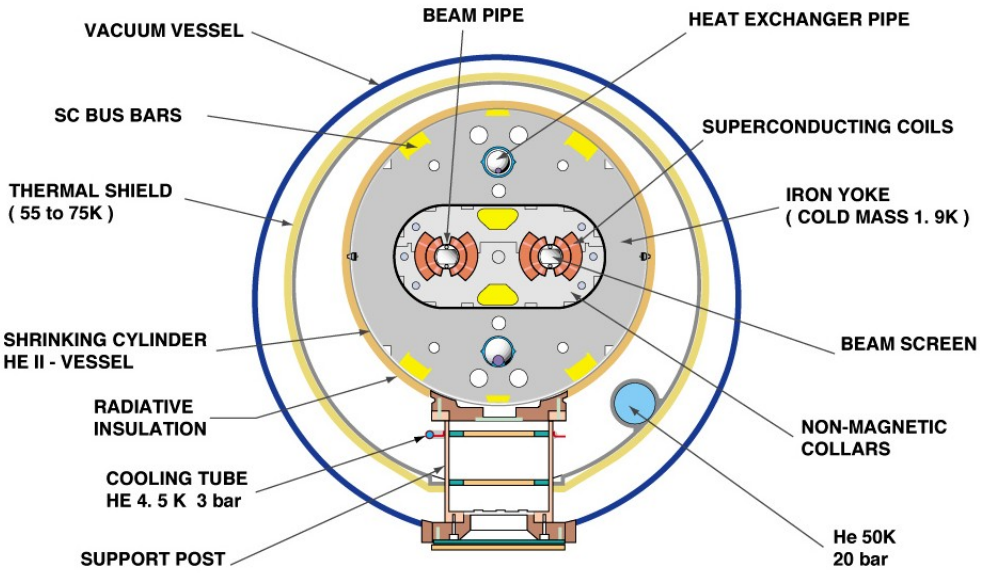

Figure 2.2: The cross section of a LHC dipole magnet shows the two beam pipes in which the proton beams circulate in opposite directions around the LHC ring. The superconducting coils generate a magnetic field of 8.3 T that steer the protons in the desired circular path.

The protons are accelerated in Radio-Frequency (RF) cavities, which generate an electromagnetic field that is tuned in such a way that the protons receive a 'kick' at each revolution in the accelerator ring. As a result, the energy of the protons increase, obtaining velocities very close to the light speed. Eventually, protons that are ideally timed and have exactly the desired energy, will feel a zero accelerating force from the RF cavity, while protons with slightly different energies that arrive earlier or later are decelerated or accelerated. As a result, the proton beams are divided in discrete 'bunches' of protons. The LHC has been designed to provide 2808 bunches per beam, with about 10^{11} protons per bunch.

Once stable beams are obtained, they are squeezed and collided at the four collision points where the particle detectors are located. An important parameter to characterize the performance of a collider is the *instantaneous luminosity* $\mathcal{L}$, a quantity influencing the rate of collisions. The instantaneous luminosity depends on the number of protons per beam, the transverse dimensions of the beams (the more the beams are squeezed,

the higher the probability that a proton-proton collision takes place), and the bunch-crossing frequency. The number of events N of a certain process with cross section σ produced per second can be expressed as

$$\frac{dN}{dt} = \mathcal{L}\ \sigma. \tag{2.1}$$

When the instantaneous luminosity is integrated over time, one obtains the *integrated luminosity* provided by the collider in a certain time range. The LHC has been designed to deliver an instantaneous luminosity of 10^{34} cm^{-1}s^{-1}, by colliding proton bunches every 25 ns. During a given bunch crossing, multiple proton-proton collisions can take place. These additional interactions are referred to as pile-up interactions. Since in these collisions particles are produced that are detected simultaneously with the particles from another potentially interesting collision event, pile-up interactions result in a serious challenge for physics experiments.

Throughout 2010, the LHC operated at a maximal instantaneous luminosity of 2×10^{32} cm^{-1}s^{-1} and delivered an integrated luminosity of about 45 pb^{-1} at 3.5 TeV per proton to the main experiments. In the next year 2011, the machine increased the instantaneous luminosity up to 4×10^{33} cm^{-1}s^{-1}, resulting in an integrated luminosity of about 6 fb^{-1}, over 100 times more than the amount of collisions obtained in 2010. In 2012, the energy per proton was raised to 4 TeV, and the instantanous luminosity was increased even further, such that an integrated luminosity of about 23 fb^{-1} was provided to the main experiments. The high instantaneous luminosities during 2011 and 2012 resulted in a lot of pile-up interactions per bunch crossing. For instance, on average 21 interactions per bunch crossing have been observed by the CMS experiment in 2012, as shown in Figure 2.3. Between the proton-collision runs, the LHC also collided lead ions at high energy.

2.1.2 The physics programme at the LHC

One of the main objectives of the LHC physics programme was to discover the Brout-Englert-Higgs boson, which was predicted as the cornerstone of the Standard Model. According to electroweak precision measurements, this boson was believed to have a mass at the electroweak energy scale, of the order of 100 GeV, but previous collider experiments did not find conclusive hints of its existence. Since the LHC was designed to collide protons at the TeV scale, the experiments were expected to shed more light on the electroweak symmetry breaking mechanism. In 2012, the two largest experiments at the LHC, the CMS and ATLAS experiments, announced the discovery of a new particle in the search for this elusive boson [5, 6]. The CMS [92] and ATLAS [93] detectors are designed as general-purpose experiments that can focus on a wide range of high-energy physics, including Standard-Model precision measurements and searches for physics beyond the Standard Model. The CMS detector is explained in more detail in Section 2.2.

Two other particle detectors are located at proton-beam collision points at the LHC. The LHCb experiment [94] focuses on precision experiments with bottom quarks, to ultimately shed light on the matter-antimatter asymmetry in the universe. The ALICE experiment [95] is designed to study heavy-ion collisions. In the collisions between lead

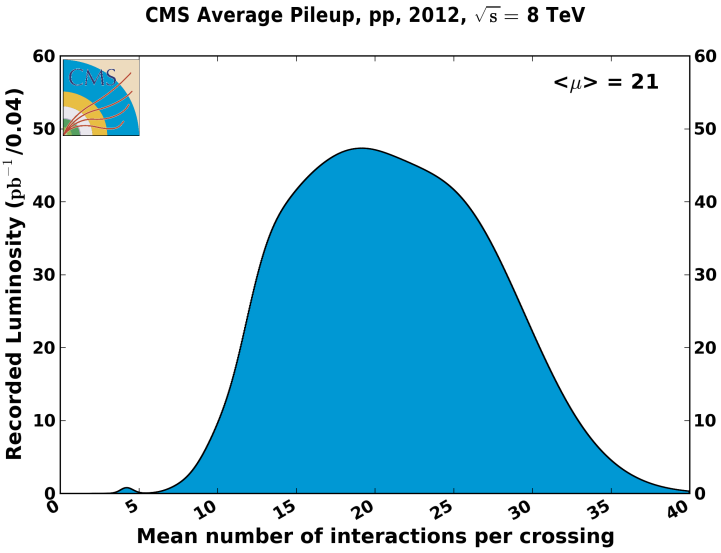

Figure 2.3: The distribution of the mean number of proton-proton collisions per bunch crossing shows that on average 21 collisions per crossing took place during the 2012 operation of the LHC [91].

ions, the energy density is so high that a new state of matter, called a gluon-quark plasma, can be formed that is believed to have been present in the universe right after the Big Bang.

In addition to these four main experiments, several smaller experiments are located at the LHC ring. These include the TOTEM [96] and LHCf [97] experiments, constructed with the purpose of studying particles emitted during proton collisions in the very forward region, close to the beam pipe.

2.2 The Compact Muon Solenoid detector

The CMS detector is one of the four main experiments located at the Large Hadron Collider. It is designed in layers of subdetectors, almost hermitically surrounding the nominal proton-proton collision point. The detector consists of a cylindrical-shaped *barrel* part and enclosed on the sides by *endcap* parts, as shown in Figure 2.4. The overall length of the CMS detector is nearly 22 m and its diameter about 15 m.

The geometry of the detector is described using the coordinate system defined in Section 2.2.1. The most inner subdetector layer consists of a dedictated tracking system, discussed in Section 2.2.2, to detect the hits and ultimately reconstruct the trajectories of charged particles emerging from the collisions. The next layers are the electromagnetic calorimeter and the hadronic calorimeter, described in Section 2.2.3 and 2.2.4, respectively, whose purpose it is to measure the energy deposited by electrons, photons and hadrons. The outermost layer consists of a dedicated muon system optimized for the detection of muons, as explained in Section 2.2.5. The tracker and the largest part of the calorimeters are enclosed within a superconducting solenoid,

generating a magnetic field of 3.8 T in order to bend charged particles and enable their momentum measurement. This strong magnetic field is obtained with the aid of an iron return yoke embedded within the muon chambers. The electronic signals from the hits and energy deposits of particles are read out and preprocessed by a data acquisition system as described in Section 2.2.6. Finally, the data from the most interesting collision events is processed and distributed for analyses, as outlined in Section 2.2.7.

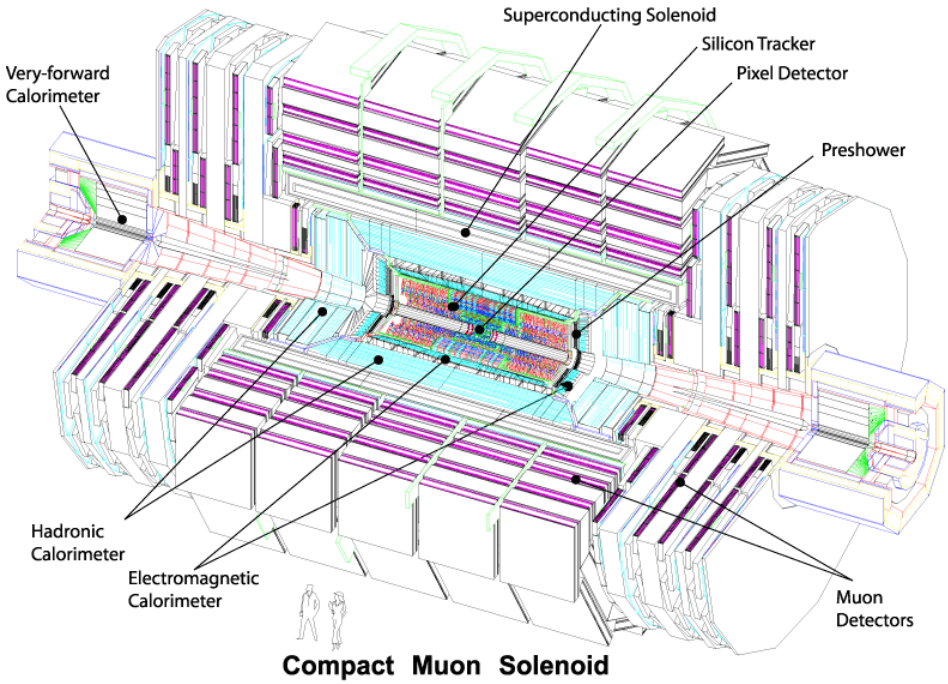

Figure 2.4: The CMS detector consists of layers of subdetectors, each with a specific purpose in the measurement of final-state particles in proton collisions. A strong superconducting magnet curves the trajectories of charged particles.

2.2.1 The CMS coordinate system

The right-handed coordinate system used in the CMS experiment has its origin at the nominal proton-proton interaction point. The x axis points from the origin towards the center of the LHC, while the y axis points up, perpendicular to the LHC plane. The z axis is pointing along the anticlockwise-beam direction. The azimuthal angle ϕ is measured in the (x, y) plane relative to the x axis, and the polar angle θ is the angle with respect to the z axis. For convenience, the polar angle is translated into another coordinate, the so-called pseudorapidity η:

$$\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.2}$$

Hence, the pseudorapidity is 0 in the direction perpendicular to the beam axis, and plus or minus infinity in the direction parallel to the beam axis. This quantity is more

practical as a measure of the 'angle' of a particle with respect to the z axis, because differences in pseudorapidity are invariant under Lorentz boosts along the z axis, for high enough particle energies. Since the boost of the center-of-mass system along the z direction in proton-proton collisions is not known on an event-by-event basis, this invariance is a useful property to describe the kinematics of final-state particles. The rapidity y of a particle with an energy E and a z component of the momentum, p_z, is defined as

$$y \equiv \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \tag{2.3}$$

Differences in rapidity are invariant under Lorentz boosts along the z axis, irrespective of the particle energy. The transverse momentum p_T of a particle is defined as the momentum component in the transverse (x, y) plane, and is often used in physics analyses. The reason is that, because of the conservation of momentum before and after the collision, the momenta of particles in the final state of a collision need to be balanced in the transverse plane. This information can be exploited when studying kinematic properties of produced particles.

2.2.2 The inner tracking system

The purpose of the inner tracker is to measure the curved trajectories of charged particles and to reconstruct vertices from which these particles may have originated. From the magnitude and the direction of the curvature, the momentum and the electric charge of the particles can be deduced. As schematically depicted in Figure 2.5, the tracking system consists of two major parts; a high-granularity silicon pixel detector close to the proton-proton interaction point and a silicon strip detector around the pixel detector. The tracker has a cylindrical shape, with an overall length of 5.8 m and a diameter of 2.5 m, and covers the pseudorapidity region up to $|\eta| < 2.5$.

The silicon pixel detector aims to measure the hits of charged particles with such a high precision, that even in the high-radiation environment close to the interaction point, separate charged-particle tracks can be disentangled and vertices can be reconstructed. This detector consists of three cylindrical pixel-cell layers located at a radius of 4.4 cm, 7.3 cm and 10.2 cm, and two endcap disks at each side of the interaction point, at a $|z|$ coordinate of 34.5 cm and 46.5 cm. The pixel detector contains in total 66 million pixel cells. All pixel cells have a surface of 100×150 μm^2, which allows for a single-hit spatial resolution of about 10 μm in the $r\phi$ dimension, and 15 to 20 μm in the z dimension [92].

Outside the pixel detector, the particle flux is reduced enough to use a less granular tracker. The silicon strip detector consists of layers of silicon-strip modules, where each module carries either one thin silicon strip detector (320 μm) or two thick strip detectors (500 μm). The strip detector is composed of four different subsystems, with a total of over 9 million silicon strip channels. These subsystems are the Tracker Inner Barrel (TIB), consisting of 4 layers covering up to $|z| < 65$ cm and up to a radius of 55 cm. At each side of the TIB, a Tracker Inner Disk (TID) is placed, with the strip modules oriented within three layers perpendicular to the beam line and covering up to $|z| = 100$ cm. The inter-strip distance in the TIB and TID ranges from 80 to 120 μm. Outside the TIB, six additional layers form the Tracker Outer Barrel (TOB), with

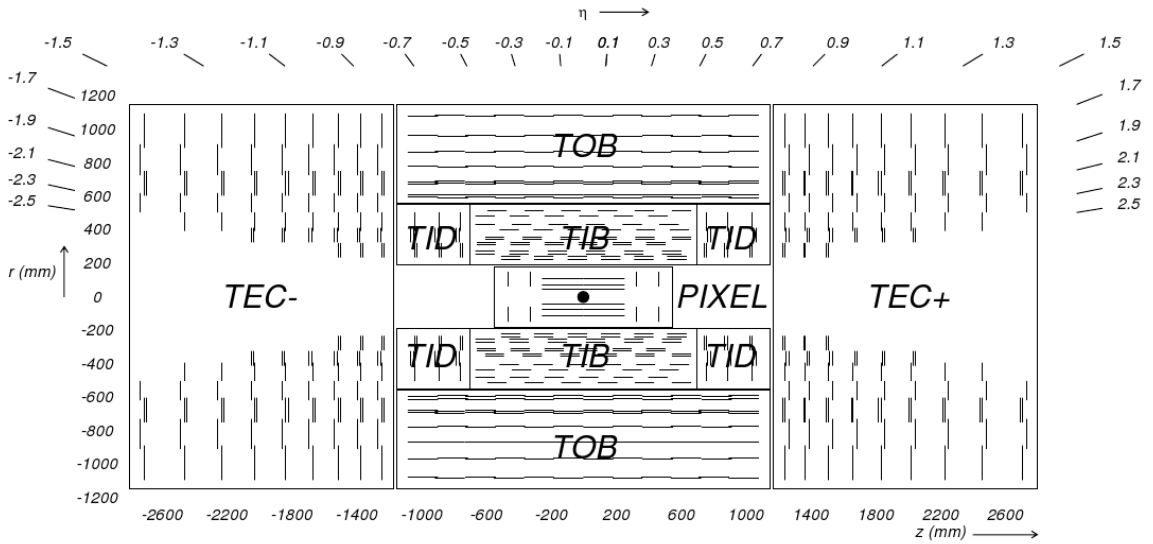

Figure 2.5: The inner tracking system of the CMS detector consists of two main components, the silicon pixel detector and the silicon strip detector. The latter is subdivided in several subcomponents, with a different number, arrangement and orientation of the silicon strip modules.

an outer radius of 116 cm. Finally, the Tracker Endcap (TEC) consists of 9 layers on each side of the tracker, and extends up to $|z| = 282$ cm. The single-hit resolution, in the $r\phi$ dimension, of the silicon strip tracker ranges from 23-35 μm in the TIB to 35-53 μm in the TOB. The resolution in the z direction is about 230 μm in the TIB and 530 μm in the TOB [92]. With this detector layout, at least nine points are measurable per charged-particle trajectory in the $|\eta|$ range up 2.4.

Track reconstruction

The reconstruction of charged-particle tracks in the inner tracking system is done with an iterative-tracking algorithm, where each iteration proceeds in four steps [92]; the track-seed generation, the pattern-recognition algorithm, the removal of track-hit ambiguities and a final track fit.

1. **Seed generation**. This first step consists of finding reconstructed hits that can be used as seeds for the subsequent track-finding algorithm. Seeds are identified either from a group of at least three reconstructed hits in the tracker, or from a pair of hits while requiring the origin of the track segment to be compatible with the nominal beam-collision point. Because of the higher granularity of the pixel detector, the seed-generation efficiency is higher in the pixel than in the strip detector, and the overall efficiency exceeds 99%.

2. **Pattern recognition**. The seeds are used as the starting point of a Kalman filter method [98]. This algorithm extrapolates the seed trajectory outward to the next tracker layer, taking into account the magnetic field and multiple-scattering

effects. If compatible hits in the next layer are found, the track parameters are updated (fitted) with this new information. This procedure continues until the outermost tracker layer is reached.

3. **Ambiguity removal**. The previous pattern-recognition step can result in multiple extrapolated tracks associated to the same seed, or different tracks sharing the same hits. Hence, a removal of the ambiguities is necessary. For this purpose, tracks sharing too many hits are removed from the list of track candidates, only keeping the tracks with the most hits. If the tracks contain the same number of hits, the track with the highest χ^2 of the track fit is discarded.

4. **Final track fit**. Finally, the track parameters are refitted with the Kalman filter method, taking into account all hits found in the pattern-recognition step. The fit is performed outwards from the beam line towards the calorimeters, and inwards from the outermost track hit to the beam line. Refitting twice improves the estimation of the track parameters.

The hits unambiguously associated to the final track are removed from the list of available hits, and the procedure is repeated, with slightly looser track-reconstruction criteria in each iteration. In this way, a high track-finding efficiency is obtained while keeping the rate of reconstructing fake tracks negligible. The global track reconstruction efficiency for muons exceeds 98% for most of the tracker acceptance, and is between 75% and 95% for charged hadrons. In general, the efficiency drops in the very high $|\eta|$ region due to lack of coverage of the two pixel disks. The transverse momentum resolution $\Delta p_T/p_T$ for muons with $p_T = 100$ GeV is about 1% to 2% up to $|\eta| < 1.6$ and becomes up to 7% at the boundary of the tracker acceptance. For smaller transverse momenta, the resolutions are below 2% in the whole pseudorapidity range [92].

Primary vertex reconstruction

The vertex reconstruction involves two steps, a vertex finding and a vertex fitting algorithm. Vertices can be found by grouping reconstructed tracks according to the z coordinate of their closest approach to the beam line. Next, a vertex fitting algorithm, such as the Adaptive Vertex fitter [99], is performed on these clusters of tracks to obtain the three-dimensional primary-vertex position. In this fit, tracks with a larger distance to the vertex are downweighted to reduce the contribution from long-lived hadron decays. Thereafter, the primary vertex corresponding to the highest sum of squared track transverse momenta is noted as the 'main' primary vertex. A vertex-position resolution of about 20 μm is obtained in the transverse dimensions, and about 30 μm in the z direction [100].

2.2.3 The electromagnetic calorimeter

The electromagnetic calorimeter system (ECAL) is designed to measure the energy of electrons and photons. It consists of 75,848 scintillating lead-tungstate ($PbWO_4$) crystals. The choice of lead tungstate as the scintillator material is motivated by its high density and small Molière radius. These properties guarantee a reasonably

small longitudinal and lateral spread of an electromagnetic shower from an electron or a photon entering the calorimeter. Moreover, this scintillating material has a fast response; about 80% of the light is emitted in 25 ns, the same order of magnitude as the LHC bunch crossing time.

The ECAL detector is built around the tracker in a cylindrical shape, as shown in Figure 2.6. The detector has three main parts, the barrel (EB), the endcap (EE) and the preshower (ES). The barrel part covers the pseudorapidity region $|\eta| < 1.479$, with the lead-tungstate crystals facing the nominal proton-proton collision point, but with a slight tilt to reduce the effect of particles crossing the small gaps between the crystals. The crystals have a length of 23 cm and a front-face surface of 22×22 mm^2. The endcap part extends from $|\eta| = 1.479$ to $|\eta| = 3.0$, with crystals with a front-face size of 28.6×28.6 mm^2 and a length of 22 cm. In front of the endcaps, from $|\eta| = 1.653$ to $|\eta| = 2.6$, a preshower is located. This subdetector contains layers of lead and silicon strips, and its main goal is to discriminate between photons and neutral pions.[1]

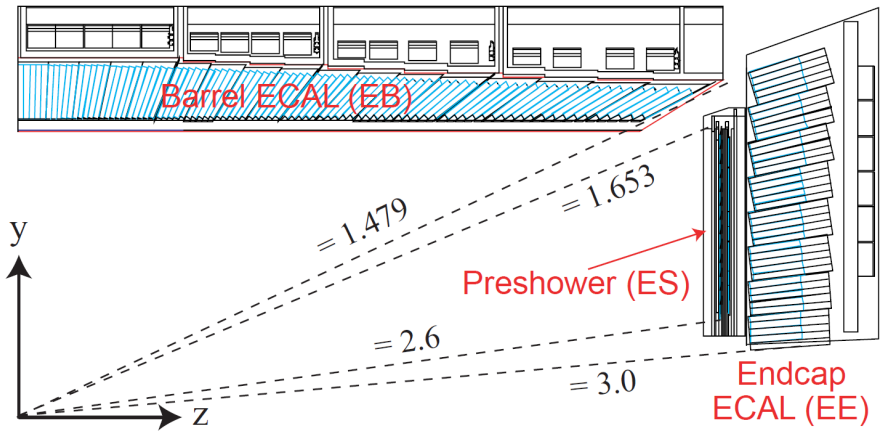

Figure 2.6: A schematic overview of one quarter of a longitudinal slice of the ECAL detector shows the main parts: the barrel (EB), the endcap (EE) and the preshower (ES) subdetectors. The scintillating crystals face towards the proton-proton interaction region.

For electrons from Z-boson decays, the electromagnetic calorimeter system provides an energy resolution better than 2% in the central region of the barrel ($|\eta| < 0.8$) and ranging from 2% to 5% elsewhere [101].

2.2.4 The hadronic calorimeter

The goal of the hadronic calorimeter (HCAL) is to measure the energy of hadrons emerging from the proton-proton collisions. Hence, it plays a crucial role in the reconstruction of the jets as well as the missing transverse energy in the event, as will be discussed in the next chapter. The HCAL detector is a sampling calorimeter made from

[1] A neutral pion is a superposition of an up−anti-up quark state and a down−anti-down quark state, and commonly decays to two photons.

brass absorber plates interleaved with plastic scintillator tiles to measure the deposited energy. The HCAL surrounds the ECAL detector, and, as illustrated in Figure 2.7, consists of four main parts. A barrel part (HB) and an endcap part (HE) are joint hermetically and are located within the region enclosed by the solenoid magnet. The barrel part extends to $|\eta| = 1.4$ and the endcap part covers the region $1.3 < |\eta| < 3.0$. Energetic hadrons may not deposit their full energy in the HB due to the limited length of the HB absorber material, such that additional scintillator layer rings are placed outside of the magnet to capture the additional energy of the hadronic shower. These rings compose the outer barrel hadronic calorimeter (HO). The absorber and scintillator plates of the HB and the HO parts form towers in (η, ϕ) space, with an area of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The forward calorimeter (HF) is designed to measure the energy of forward hadrons in the region $2.9 < |\eta| < 5.2$. The HF calorimeter, located at about 11.2 m from the interaction point, is built from steel absorbers embedded with quartz fibers. The tower front-face size in (η, ϕ) space decreases for higher $|\eta|$ in the HE and the HF subdetectors, to a maximal size of $\Delta\eta \times \Delta\phi = 0.350 \times 0.174$ in the HE subdetector and 0.302×0.348 in the HF subdetector.

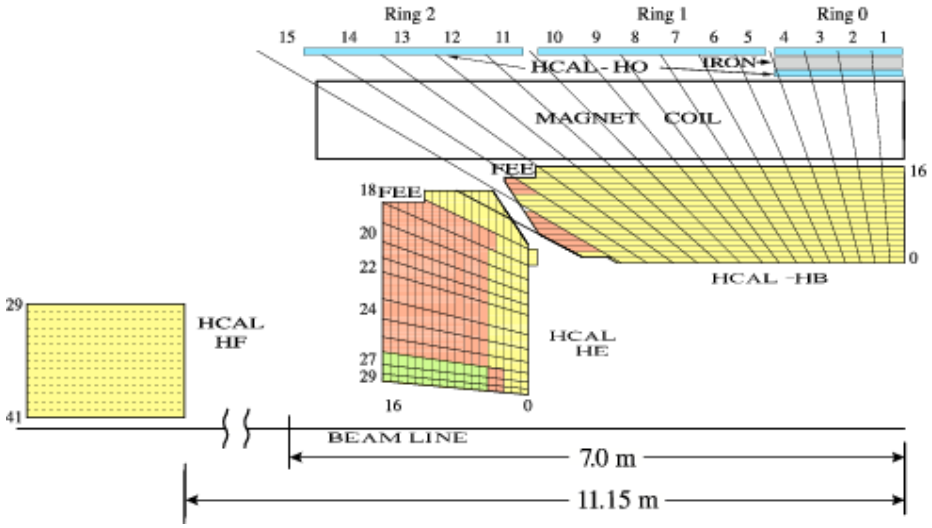

Figure 2.7: A schematic overview of one quarter of a longitudinal slice of the HCAL detector shows the different subcomponents: the barrel (HB), the outer barrel (HO), the endcap (HE) and the forward (HF) subdetectors. Towers of absorber and scintillating material face the proton-proton interaction region.

The resolution of pion energy measurements from test beams, using a prototype setup of a combined ECAL+HCAL system, was found to be between 15% and 20% below 50 GeV, and to improve down to 10% at 300 GeV [92].

2.2.5 The muon system

In order to precisely reconstruct muon candidates, the CMS detector is surrounded by a dedicated muon system. This system extends up to about 10.5 m from the nominal proton-proton interaction point along the z axis, and up to a radius of 7.5 m from

the beam line. As shown in Figure 2.8, the muon system consist of layers of muon chambers embedded in the iron yoke closing the magnetic field lines. In the barrel part of the CMS detector, covering $|\eta| < 1.2$, four layers of Drift Tubes (DT's) are installed. The drift tubes have a good single-hit spatial resolution of less than 100 μm, and an angular resolution of the order of 1 mrad [92]. Since the magnetic field as well as the muon flux increases at larger pseudorapidity, another technology is employed in the endcap region $0.9 < |\eta| < 2.4$. Here muon hits are recorded with Cathode Strip Chambers (CSC's), with typical spatial resolutions of less than 200 μm. The DT's and some of the CSC's are complemented by another type of muon chambers, Resistive Plate Chambers (RPC's), up to $|\eta| < 1.6$. These muon detectors are very fast, and improve the time resolution of muon-hit measurements down to 1 ns. Hence, the RPC's are useful in the triggering on muons. The purpose and design of the trigger system is explained in the next section.

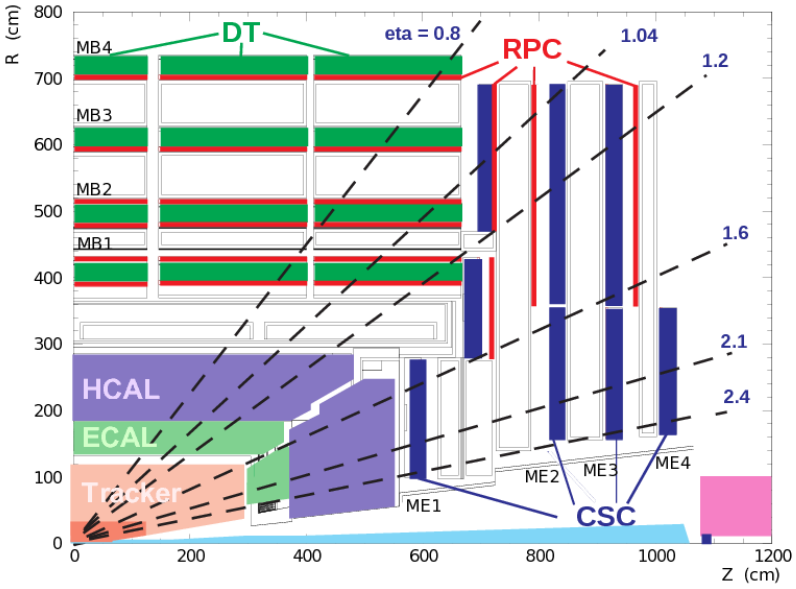

Figure 2.8: A schematic overview of one quarter of a longitudinal slice of the muon system shows the layers of muon chambers. In the barrel and endcap part, Drift Tubes (DT's) and Cathode Strip Chambers (CSC's) are used, respectively. These detectors are complemented by fast Resistive Plate Chambers (RPC) up to $|\eta| < 1.6$.

2.2.6 Data acquisition

The LHC is designed to provide proton bunch crossings at a rate of 40 MHz. Assuming about 20 inelastic proton collisions per bunch crossing, this corresponds to a collision-event rate of the order of 10^9 Hz. Processing and storing events at this rate and amount is beyond the present computational and technological capabilities. The data acquisition system of the CMS detector is therefore equipped with an online trigger system, to make a fast decision on whether or not the event is interesting enough to fully process and store on tape. The event rate needs to be reduced down to about 100

Hz in order for the processing to be feasible. The online trigger system consists of two trigger levels: the Level-1 (L1) trigger and the High Level Trigger (HLT).

The Level-1 trigger

The raw data from all CMS subdetectors are first buffered in front-end drivers, electronic hardware devices on the detector with a pipeline memory of 3.2 μs. Segmented information of the calorimeter and muon systems is sent to a control room in the cavern next to the CMS detector. Here, coarse L1 trigger algorithms are performed within only 1 μs, to decide to keep a particular event or not. The L1 trigger is able to reduce the event rate to the order of at most 100 kHz.

The High Level Trigger

Events passing the L1 trigger are processed by the HLT software, that uses more complicated reconstruction algorithms. Here the information of all subdetectors, including the tracker, can be used to perform algorithms on higher-level reconstructed objects like electrons, muons, photons, jets and the missing transverse energy. Typically, events from decays of heavy particles contain high-p_T decay products, hence these objects can be used to select the most interesting events. An HLT output rate of 100 Hz can be obtained, and these events are stored for further physics analysis.

2.2.7 The CMS computing model

The events that pass the online trigger requirements are stored, processed and distributed via an extensive computing infrastructure that is common to all LHC experiments: the Worldwide Large Hadron Collider GRID (WLCG) [102]. This computing model consists of several levels ('Tiers') to deal with the immense complexity and size of the data sharing network.

- **Tier-0**. A single Tier-0 centre is located at CERN. Here the raw data collected by the CMS detector is archived, and a first 'prompt' reconstruction of the data is performed. After this reconstruction step, the data is already in a file format that may be used in physics analyses. Next, the Tier-0 site distributes the data to the Tier-1 centres. The Tier-0 centre computing facilities can also be used to reprocess data, taking into account newly derived calibrations.

- **Tier-1**. A total of seven Tier-1 sites around the world are involved in the CMS experiment. In these centres, a large computing infrastructure is provided to carry out data reprocessing when needed. The Tier-1 sites may store real collision data received from the Tier-0 centre as well as simulated data, and they distribute the data further downstream to the different Tier-2 sites.

- **Tier-2**. Over 50 Tier-2 centres around the world receive data from the Tier-1 sites, and make this data accessible for physics analysis. The Tier-2 computing facilities are also widely used for the production of simulated data.

This distributing approach results in multiple copies of the same real or simulated data at different physical locations. Hence, these data can be accessed and processed in parallel by many physicists around the globe.

Chapter 3

Simulation and reconstruction of proton-proton collisions

In order to search for signatures of new particles in proton collisions recorded by the CMS detector at the LHC, one needs to be able to compare the observations with model predictions. For this purpose, collision events are simulated with Monte-Carlo event-generator and simulation software, from the hard parton collision up to the hadrons, photons and leptons that can be observed in the detector. The generation of proton collisions is discussed in Section 3.1. The final-state particles emerging from the collisions interact with the detector material, and this is simulated using dedicated detector simulation software, as described in Section 3.2. Finally, in Section 3.3 it is outlined how the physical objects that can be used in physics analyses are reconstructed from the energy deposits and the hits of charged particles in the detector. The algorithms used for the object reconstruction are applied to the observed and simulated events.

3.1 Collision event generation

When two protons collide, the probability for elastic, low-energetic scattering interactions is higher than for inelastic, high-energetic processes. However, the latter type of processes are generally considered to be the most interesting, as head-on collisions at a high center-of-mass energy increase the probability of producing heavy particles. The event generation chain is shown schematically in Figure 3.1 and can be summarized as follows [103].

- **Parton distribution functions of colliding protons.** The incoming protons consist of partons (quarks and gluons). The probability for a parton to carry a longitudinal momentum fraction x of the proton at a momentum transfer scale Q^2 is governed by the so-called parton distribution functions (PDFs) $f(x, Q^2)$. This is explained in more detail in Section 3.1.1.

- **Hard-scattering process.** The two interacting partons produce the actual process of interest, where particles are produced through the fundamental interactions. If these particles have a very short lifetime, like heavy quarks or

heavy bosons, the subsequent decay process is considered to be part of the hard interaction. The generation of the hard interaction is described in Section 3.1.1.

- **Parton shower.** The partons that collide are color charged. Hence, the hard scattering is accompanied by a cascade of radiation from QCD processes. If this radiation originates from the initial colliding partons, this is referred to as Initial State Radiation (ISR). The radiation from outgoing partons is called Final State Radiation (FSR). The generation of the resulting shower of gluon and quark radiation is outlined in Section 3.1.2.

- **Hadronization.** The emerging color-charged particles in the parton shower are connected with each other through the strong interaction in a highly complex way. Only color-neutral particles can be observed in the final-state, and the rearrangement of the partons into these color-neutral hadrons is described using non-perturbative phenomenological models, as explained in Section 3.1.3. The formed hadrons are often not stable, and can decay.

- **Underlying event.** The proton remnants that do not participate in the hard scattering are color-charged too, and will therefore involve additional QCD radiation and hadronization in the event. The modeling of this so-called underlying event is discussed in more detail in Section 3.1.4.

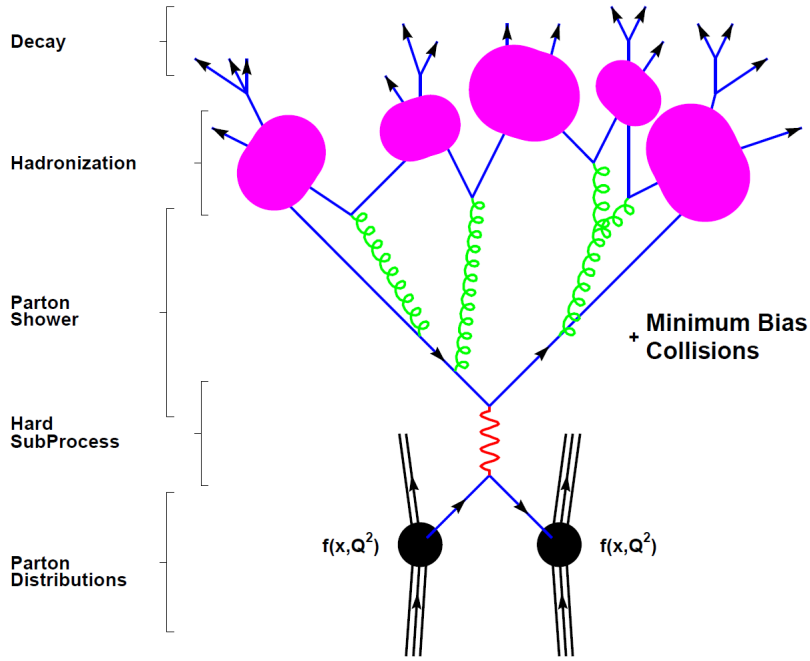

Figure 3.1: The collision event generation can be represented by a factorized chain of different steps [103], ordered from the lower part of the scheme towards the upper part.

3.1.1 Parton distribution functions and the hard interaction

When two protons collide, the hard interaction happens between two partons i and j that carry a fraction x_i and x_j of the momentum of the protons. Effectively, the center-of-mass energy $\sqrt{s}$ of the interaction gets reduced to $\sqrt{Q^2} \equiv \sqrt{\hat{s}} = \sqrt{x_i x_j s}$. The energy scale Q^2 is defined as the momentum transfer or the virtuality of the process. The hard fundamental interaction itself is dictated by the Lagrangian density of the considered model, where the possible interactions can be represented by Feynman diagrams. But since the proton is a composite particle, the differential cross section $d\sigma_{pp\to X}$ of a particular process can be written as a function of the differential partonic cross sections $d\hat{\sigma}_{ij\to X}$, weighted with the parton distribution functions [104]:

$$d\sigma_{pp\to X} = \sum_{i,j} \int\int dx_i\, dx_j\, f_i(x_i, Q^2)\, f_j(x_j, Q^2)\, d\hat{\sigma}_{ij\to X}. \tag{3.1}$$

Here the sum runs over the possible initial parton-flavor, spin and color configurations of the proton constituents. The parton distribution functions $f(x, Q^2)$, determining the probability to find a parton with momentum fraction x at an energy scale Q^2, depend on these parton configurations as well.

The parton distribution functions are obtained from global fits to experimental data, for instance from deep inelastic scattering experiments where the structure of the proton is probed with electrons or positrons. The simulated events in this thesis are generated using the PDF sets obtained by the CTEQ collaboration [105]. Figure 3.2 illustrates the dependency of the parton distribution functions on the momentum fraction x, for the CTEQ6L1 PDF set. In this figure, a scale $Q^2 = (350\ \text{GeV})^2$ is chosen, which is representative for the virtuality of top-quark pair production, since this implies a center-of-mass energy of the hard interaction of about twice the top-quark mass. It can be seen that for most values of the momentum fraction x, the gluon density dominates over the quark densities. Hence, it is easier to probe the gluons than the quarks. For x values close to 1, the parton densities of the up and down quarks (the valence quarks of the proton) dominate over the gluon density. The PDFs for anti-up and anti-down quarks are in general lower, since these sea quarks originate in the proton only from gluon splitting.

The differential partonic cross section $d\hat{\sigma}_{ij\to X}$ in Equation (3.1) involves a matrix-element amplitude of the hard interaction, derived from the Lagrangian density of the quantum-field theoretical description one considers. If the energy scale of the interaction is high enough, or equivalently, if the strong coupling constant $\alpha_s \equiv g_s^2/4\pi$ is small enough[1], the matrix element can be calculated up to a certain perturbative order in α_s. A certain renormalization scale μ_R needs to be chosen to deal with ultraviolet divergences, as mentioned in the last paragraph of Section 1.1.2. Since the strong coupling constant runs with the energy scale, α_s is evaluated at this energy scale μ_R.

[1]Including one-loop corrections, the running of the strong coupling constant is given by $\alpha_s(Q^2) = \alpha_s(\mu_R^2)/(1 + \beta\alpha_s(\mu_R^2)\ln\frac{Q^2}{\mu_R^2})$, with $\beta = (33 - 2n_f)/12\pi$, n_f the number of quark flavors and μ_R the renormalization scale. This means that α_s decreases with increasing energy, and when $\alpha_s \ll 1$, perturbative QCD calculations can be done. The phenomenon of decreasing strong-interaction strength with increasing energy is known as asymptotic freedom.

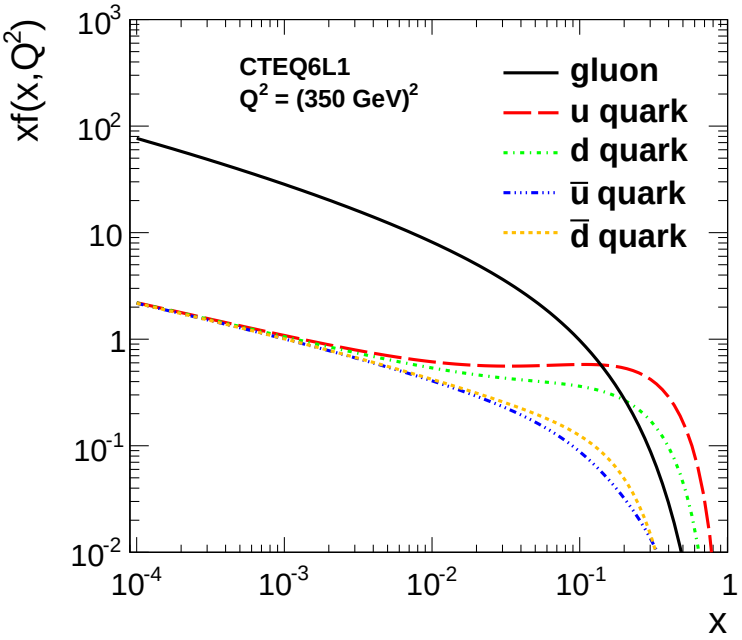

Figure 3.2: The parton distribution functions of the proton are shown at an energy scale $Q^2 = (350\ \text{GeV})^2$ for gluons, u and d quarks and $\bar{u}$ and $\bar{d}$ antiquarks [106].

Apart from ultraviolet divergences, also *infrared* and *collinear* divergences are encountered in the calculation of the matrix element. This is caused by the emission of soft (low-p_T) gluons, or gluons collinear to the partons. These phenomena correspond to terms in the calculation that spoil the perturbativity of the interaction. The validity of the equation given in (3.1) is therefore only true in the assumption that these non-perturbative effects can be factorized from the calculable perturbative effects. This is ensured by choosing an appropriate 'factorization scale' μ_F that defines the regime where the perturbative expansion in the matrix element calculation can be applied.

Event generators like MadGraph [107] can calculate matrix elements up to leading order (LO), but next-to-leading (NLO) generators like powheg [108] and aMC@NLO [109] are also available. Since virtual loops are not fully included in LO generators, and also real corrections are approximated, the quantity and properties of additional partons in the final state might not be described accurately. One can scale a LO cross section to the NLO level using a so-called k-factor, defined as the ratio of theoretical NLO and LO cross section calculations. Caution is needed, however, as such a k-factor may depend on the kinematic phase space and the probed energy scale.

The kinematic properties of the particles produced in the hard interation cannot be predicted on an event-by-event basis, due to the intrinsic quantum physical stochastic behaviour of the processes, as well as an incomplete knowledge of the initial state. Hence, once the matrix element is known, the kinematic phase space is sampled using Monte Carlo techniques in order to obtain a set of generated events.

The MadGraph/MadEvent generator

The MadGraph matrix element generator can be used to calculate tree-level Feynman diagrams for a wide variety of Standard-Model as well as hypothetical new-physics model processes. Using these matrix elements, the events are generated with the MadEvent generator and LO cross sections can be calculated. MadGraph/MadEvent is used extensively in the CMS collaboration to generate simulated event samples. Most SM background events relevant for the analyses presented in Chapters 4 and 5, and all chiral fourth-generation and vector-like quark signal events have been generated in this framework. The decay of produced particles can be generated in MadGraph as well. The events generated by MadGraph/MadEvent can be interfaced with general-purpose generators simulating also the parton-showers and the hadronization, such as pythia [110].

One can consider initial and final state radiation in the event generation, by requiring a certain number of additional partons in the final state of the matrix element. The additional partons correspond to extra 'legs' in the Feynman diagram. For the top-quark pair (W+jets and Z/γ^*+jets) production processes, up to three (four) additional partons are included in the event generation. It should be noted that this does not yield a complete NLO description of events, since MadGraph neglects loop corrections. Nevertheless, the higher-order corrections from multi-leg tree-level diagrams provide a good approximation when considering hard radiation.

The POWHEG generator

The simulated single top-quark event samples used in Chapters 4 and 5 are generated using the powheg generator. This program is able to generate events using NLO QCD computations, but for a relatively limited number of physics processes. Only one additional parton can be generated in the matrix element, which is a limiting factor as well, such that the description of additional QCD radiation should be provided later by a parton-shower program.

3.1.2 Parton showering

Incoming and outgoing partons from the hard interaction can radiate gluons and quarks. The perturbative evolution of the cascade of parton branchings to other partons can be characterized using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations [111–113]. In the parton shower approach, the infrared and collinear divergences are handled by introducing a so-called Sudakov form factor in the DGLAP equations. This factor handles the cancellation of real and virtual divergences and ensures conservation of parton branching probability. The DGLAP equations describe how the probability of an initial 'mother' parton (q or g) splitting at an energy scale Q^2 to other 'daughter' partons ($q \to qg$, $g \to gg$ or $g \to q\bar{q}$), evolves in time. At each splitting, the momentum of the mother parton is divided among the daughter partons, and these daughters can in turn branch to other partons at a lower Q^2 scale. This branching cascade continues down to a certain point, defined by the energy scale Λ_{QCD} at which α_S would be of order unity. Here the energy per parton would be so low that

perturbative QCD cannot be applied anymore. This cutoff scale is usually taken to be around 1 GeV. Below the Λ_{QCD} scale, hadronization models need to be invoked to describe the non-perturbative regime, as explained in Section 3.1.3.

The parton showering can be applied to radiation of partons in the final state, as well as in the initial state of the hard-scattering process. In the latter case, the shower evolution provided by the DGLAP equations are reversed towards the scale corresponding to the parton directly from the PDF. In the simulated event samples used in the physics analyses of this thesis, the parton showering implemented in the PYTHIA program is used, where a set of parameters determine the showering procedure and the amount of ISR and FSR. These parameters have been tuned to data from collider experiments to match the observations.

Matching matrix elements with parton showers

The description of additional radiation with matrix-element calculations has certain limitations, originating from the soft and collinear divergences one encounters, and the high computational cost. The parton shower approach is able to deal with these divergences and is computationally much faster, as DGLAP equations are a simplified way to treat parton branchings by avoiding extensive matrix-element calculations. Hence, the matrix-element and parton-shower approaches are complementary, and need to be interfaced properly.

The main issue that needs to be addressed when matching matrix elements with parton showers is the potential double counting of parton configurations. The reason is that an $(n+1)$-jet event can be obtained in two ways. Either a matrix-element event with $n+1$ hard partons is generated and showered to obtain $n+1$ jets, or a matrix element with n partons is generated, and an additional hard parton is emitted during the parton showering at a large enough angle with respect to the matrix-element partons. In the simulated event samples used in this thesis and generated with MadGraph/MadEvent, the MLM matching scheme [114] is used, which involves the following procedure.

Suppose one has generated a sample consisting of events with $n \leq N$ matrix-element partons. Consider the subsample of events with n matrix-element partons that survive basic acceptance cuts: transverse momentum $p_T > p_T^{min}$, pseudorapidity $|\eta| < |\eta^{max}|$ and $\Delta R > \Delta R^{min}$, where $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ represents the separation between partons in (η, ϕ) space. The pseudorapidity η is defined according to Equation (2.2). Imposing these thresholds ensures that the partons do not reside in phase-space regions with soft and collinear divergences. The events are passed to a parton-shower program like PYTHIA, providing the shower evolution down to lower Q^2 scales. The set of partons produced as such are then clustered with a generic clustering algorithm.[2] The obtained clusters of partons hence agree with some basic concept of jets, and they are required to have a transverse energy larger than some threshold (the matrix-element parton-shower matching threshold). Next, the initial matrix-element partons can be matched to the jets, starting from the hardest parton. A parton and a jet are matched if they have a small angular separation, namely $\Delta R < \Delta R^{match}$. Once a match is found, the jet is

[2]More details about jet clustering algorithms can be found in Section 3.3.4.

removed from the list of available jets to which the other partons are matched. This matching sequence continues until all matrix element partons have found a match. If $n < N$ and there are unmatched jets or partons left, the event is dropped. If $n = N$, meaning one considers the subsample of events with the maximal amount of additional partons produced in the matrix element, left-over jets are retained. In this way, all n-parton subsamples are exclusive, and can be combined in an inclusive sample with a high multiplicity of additional jets. The dedicated simulated event samples for the $t\bar{t}$+jets, W+jets and Z/γ^*+jets processes, used in the analyses presented in Chapters 4 and 5, have been created using this procedure.

In the event samples generated with the NLO POWHEG generator, the program used for the parton shower is also PYTHIA, but the matching procedure is somewhat different. The matrix elements in POWHEG samples contain up to one additional parton in the hard interaction at NLO accuracy. The clusters from the parton shower are ordered in p_T, and the highest-p_T cluster should then be matched with the emitted parton from the hard process. This results in events with up to one additional parton from the matrix element, and additional partons on top of the first one are generated by the parton shower program.

3.1.3 Hadronization and decay

In the parton-shower formalism the partons cascade down to a parton-splitting energy scale Λ_{QCD}. Since perturbative QCD calculations are not possible anymore below this scale, phenomenological models are used to describe how hadrons are formed from the partons produced in the shower.

In PYTHIA, this so-called hadronization is implemented according to the Lund string model [115], which is based on the idea that the strong-interaction potential $V(r) = kr$ is linear in the distance r between two quarks (this phenomenon originates from the self-interacting nature of gluons), with the proportional constant k of order 1 GeV/fm. Hence, when two quarks separate, the potential energy increases, and the gluon field connecting the quarks forms a 'color string' or tube. The potential energy stored in the color tube can be so high that a new quark-antiquark pair is created. The probability for this to happen decreases exponentially with the square of the quark mass, so this effect is only sizable for up, down or strange quarks. After the fragmentation of the color string in new quarks, the quarks are arranged in two quark-antiquark bound states. However, the quarks in a bound state can still have enough momenta to separate from each other, increasing the field strength and potentially creating a new quark-antiquark pair. These processes continue until only on-shell color-neutral hadrons remain. However, these particles are often not stable, and decay further to particles that are observed in the detector. The decay of the particles can be treated in PYTHIA, or in more specialized packages like TAUOLA [116].

Several parameters in the phenomenological hadronization model can be tuned to match the experimental data. The parameters in PYTHIA are for instance related to the functions describing the fragmentation probability for light-flavor (u, d and s) and heavy-flavor (c and b) quarks. It should be emphasized again that very heavy quarks, like the top quark (or hypothetical heavy fourth-generation or vector-like quarks), have such a short lifetime that they decay instead of hadronizing.

3.1.4 Underlying event

Since protons are composite particles, only a fraction of the total available energy is associated to the hard parton scattering in proton-proton collisions. The color charge carried away by the partons participating in the hard scattering interaction leaves the energetic proton remnants color charged as well. As a consequence, the remainder of the proton will hadronize as well, increasing the charged-particle multiplicity in the event. Moreover, the hadronization process of the proton remnants is not independent from the hadronization in the hard-scattering event, because color connections between these processes exist.

In addition, multiple parton collisions can occur in a given proton collision. The probability for this to happen is higher for central collisions than for peripheral collisions because of geometrical reasons, and the probability also increases for events involving a harder main scattering interaction. The beam-remnant hadronization and the multi-parton interaction phenomena are referred to as the *underlying event*. The underlying-event activity involves non-perturbative QCD effects, and is described using phenomenological models. In the event samples at a center-of-mass energy of 7 TeV used in Chapter 4, the parametrization of the underlying-event modeling is included in the so-called Z2 tune [117], using experimental data from the CDF experiment at the Tevatron accelerator, and including some updates for LHC experiments. The underlying-event tune in the simulated samples at a center-of-mass energy of 8 TeV used in Chapter 5, is an updated version of this Z2 tune using the collision data of LHC experiments, and is referred to as the Z2* tune. It has been checked that the data collected by the CMS experiment at 0.9 TeV and 7 TeV show a good overall agreement with the used underlying-event tuning incorporated in PYTHIA [118].

3.2 Detector simulation

After the event-generation chain outlined in Section 3.1, the particles that emerge from the collision pass through the detector, and therefore their interaction with the detector material needs to be simulated. This can be done using a full description of the CMS detector geometry and its subdetectors [92], using the GEANT4 toolkit [119]. This description provides a mapping of the magnetic field as well, which is crucial to simulate the curvatures of charged particles. In the GEANT4 program, the particle trajectories are tracked through the active detector regions (i.e. the sensitive layers of the detector that are designed to detect traversing particles) and the dead material regions (e.g. gaps between detector components, support structures and cables). The energy deposits in the various subdetectors are simulated as well. The simulated effects include for instance the Bremsstrahlung of charged particles, photon conversions to an electron-positron pair, the energy loss of charged particles by ionization of detector material, and the showering of electrons, photons and hadrons in the calorimeters.[3]

[3]The particle showering in the electromagnetic and hadronic calorimeters should not be confused with the parton showering explained in Section 3.1.2. The former effect is induced by the interaction with detector material, while the latter provides a description of the evolution of QCD radiation from a parton.

Afterwards, the electronic response from the hits in the active detector material is simulated. All SM background events in the analyses presented in this thesis are simulated using this full simulation of the CMS detector.

Fast simulation

The detailed simulation of the detector response to particles is computationally very intensive, taking over 100 seconds to simulate one event. To speed up the detector simulation, the CMS experiment has developed dedicated fast-simulation (FASTSIM) software [120], providing a simplified description of the detector geometry and response. Parametrized functions are obtained with the aid of full-simulation studies, and enable a faster determination of the particle trajectories and energy loss throughout the subdetector layers. This fast simulation reduces the event-simulation CPU time by about a factor of hundred compared to the full simulation based on GEANT4, while providing a comparable accuracy. The vector-like quark signal events used in the search presented in Chapter 5 are simulated using FASTSIM. The production of these samples involves many millions of events, and a full GEANT4-based simulation is rather impractical due to the high consumption of CPU resources.

Figure 3.3 gives a comparison between events obtained using the full simulation and the fast simulation for a pair-produced vector-like quark process where the heavy quarks decay to a W boson and a light-flavor quark. The generated mass of the vector-like quarks in these samples is 800 GeV. Apart from the used detector simulation framework, the generated events are identical. Basic kinematic properties are compared for the jet, muon, electron and missing transverse energy objects, whose reconstruction is discussed in more detail in Section 3.3. The same physics-object selections are applied on both samples; one isolated muon (with transverse momentum $p_T > 30$ GeV and pseudorapidity $|\eta| < 2.1$) and at least two jets (with $p_T > 30$ GeV and $|\eta| < 2.4$). For the plot of the electron p_T, an isolated electron ($p_T > 30$ GeV and $|\eta| < 2.5$) is required instead of a muon. In general, the agreement of these kinematic distributions is reasonable, with deviations smaller than 15%. The largest discrepancies are visible towards lower transverse momenta of the charged leptons. This indicates it is more difficult to parametrize the low-p_T behavior of particles traversing the detector. In the search for vector-like quarks presented in Chapter 5, a systematic uncertainty of 20% will be set on the fast-simulated signal process event yields, which covers the discrepancy between full and fast simulation.

Pile-up interactions

As mentioned in Section 2.1.1, additional proton-proton interactions can occur during the same (or preceding or following) proton-beam bunch crossing. These additional interactions usually involve softer final-state particles, but their energy deposits and detector hits still affect the recording of the event triggered by the main hard-scattering interaction. This pile-up effect is larger for higher instantaneous proton-beam luminosities. To simulate the effect of pile-up in the event samples used in this thesis, additional minimum-bias (i.e. soft) interactions are generated with PYTHIA, and the detector hits of the corresponding generated final-state particles are superimposed to

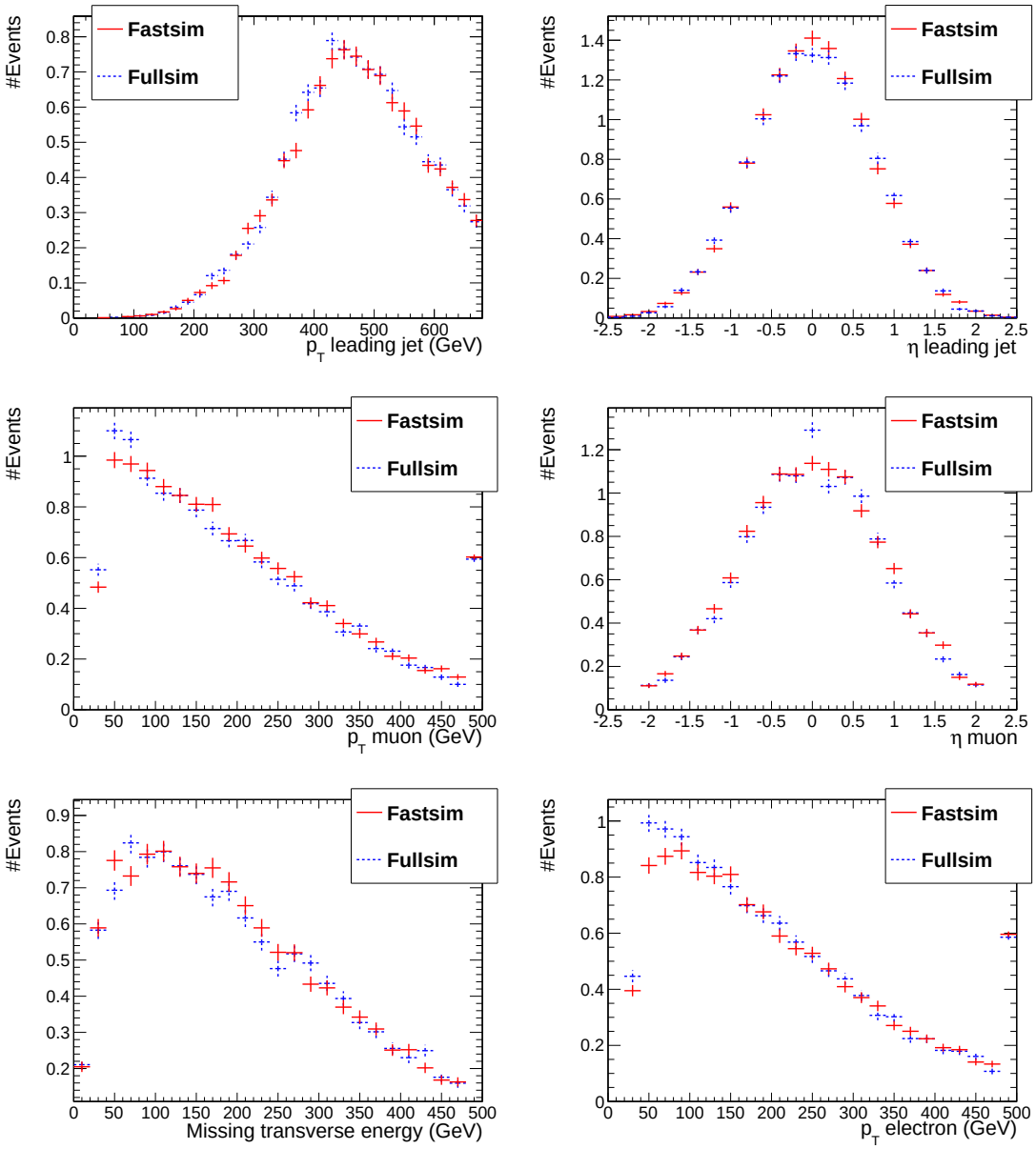

Figure 3.3: A comparison of full simulation (blue dashed) and fast simulation (solid red) for an example of simulated vector-like quark processes, namely $DD \to WqWq$ with a heavy-quark mass of 800 GeV. The spectra of the hardest-p_T ('leading') jet p_T and η (top, left and right), the muon p_T and η (middle, left and right), the missing transverse energy (bottom left) and the electron p_T (bottom right) deviate less than 15% in fast simulation with respect to full simulation.

the detector hits of the main interaction. The number of generated additional interactions follows an assumed distribution that is chosen to be reasonably representative for the high-luminosity environment at the LHC. In a physics analysis, the simulated events can then be reweighted such that the distribution of the observed number of pile-up interactions agrees with the distribution obtained from simulated events.

After the detector simulation step, the generated events are in the same format as real collision data events recorded by the CMS detector. Hence, the same algorithms and software can be employed to reconstruct the physics objects to be used in analyses, as discussed in the following section.

3.3 Physics-object reconstruction and identification

In both real and simulated collision events, the hits of particles with detector layers produce digital signals, but data in this raw format are impractical for most physics analyses. In many CMS analyses a specialized high-performance algorithm, called the particle-flow (PF) algorithm, is used with the aim of reconstructing a complete description of all stable particles traversing the detector in a given event. This PF algorithm is discussed in Section 3.3.1. The way muon and electron candidates are reconstructed is outlined in Sections 3.3.2 and 3.3.3, respectively. The reconstruction algorithm to cluster hadrons and other particles (photon and leptons) into jets is explained in Section 3.3.4. Jets are high-level physics objects, and the procedure to obtain a good estimate of their energy is also described in this section. This is followed in Section 3.3.5 by a discussion of the algorithms to identify jets originating from b quarks. Finally, the missing transverse energy reconstruction is described in Section 3.3.6.

3.3.1 The particle-flow event reconstruction method

The particle-flow algorithm [121, 122] aims to combine the information of the different CMS subdetectors, in order to reconstruct a full event picture by identifying all stable particles traversing the detector. The PF algorithm links clusters of calorimeter cells and charged-particle tracks into *blocks*, which are then interpreted in terms of electrons, muons, photons, and charged and neutral hadrons. The determination of PF-reconstructed particle momenta is more precise than when obtained from more traditional approaches, since the information of subdetectors is combined. The resulting list of particles is for instance used to build jets, to determine the missing transverse energy and to quantify charged-lepton isolation with respect to other particles.

The PF reconstruction algorithm starts by considering fundamental elements, namely charged-particle tracks and calorimeter clusters. The tracks are obtained from the iterative-tracking procedure described in Section 2.2.2, providing a high track-reconstruction efficiency and a low fake rate. The calorimeter clustering, performed in each subdetector separately, proceeds as follows. First, *cluster seeds* are identified as calorimeter-cell energy deposits above a given energy threshold. Next, 'topological clusters' are formed from the seeds by joining neighbouring cells if their corresponding energy is large enough. The energy thresholds are chosen to suppress contributions

from electronic noise. Each topological cluster can give rise to one or more 'particle-flow clusters', one for each seed. The energy of the topological cluster is shared among the PF clusters according to a cell-cluster distance algorithm.

In the following stage, the fundamental particle-flow elements are linked together, because in general, a particle traversing various subdetectors can give rise to several distinct elements. A group of linked elements is called a 'block'. For instance, tracker tracks can be linked to calorimeter clusters by extrapolating the track into the ECAL or the HCAL. The quality of the link is decided based on the distance in (η, ϕ) space between the extrapolated track position and the cluster position. Links can also be formed between two calorimeter clusters (either from the ECAL and the HCAL, or from the ECAL and the preshower subdetectors) as well as between a charged-particle track in the tracker and a muon track in the muon system. Hence, from the charged-particle tracks and the PF clusters, muon and electron candidates can be reconstructed, which is explained in more detail in the dedicated Sections 3.3.2 and 3.3.3.

Once muons and electrons are identified, the corresponding linked elements are removed from the list of blocks. From the remaining charged-particle tracks and PF clusters, charged hadrons, neutral hadrons and photons are reconstructed as follows. For tracks linked to PF clusters, the track momentum is compared to the PF-cluster momentum. If both momenta are compatible within uncertainties, a charged PF hadron is identified. If the track momentum is much smaller than the PF-cluster momenta, this can be interpreted as an energy deposit on top of a charged-hadron deposit, hence a charged PF hadron together with a neutral hadron or photon are identified. A track connected to a PF cluster with a significantly smaller momentum is not consistent with a charged-hadron hypothesis, and therefore additional PF muons with loose selection criteria are searched for. After the identification of the charged PF hadrons, the remaining PF clusters in the HCAL and ECAL are identified as neutral PF hadrons and PF photons, respectively.

Additional higher-level algorithms, for various purposes, can be applied on the list of PF particles in an event. One particular useful feature is that charged hadrons can be removed from the event if they are matched to other primary vertices than the main primary vertex. As mentioned in Section 2.2.2, the main primary vertex is identified as the interaction vertex corresponding to the highest sum of squared track transverse momenta. The other primary vertices are assumed to correspond to pile-up interactions, hence removing charged particles originating from these vertices is an effective method to reduce the influence of pile-up activity in the event. This method is referred to as *charged-hadron subtraction*, and is used in the analyses presented in this thesis.

3.3.2 Muon reconstruction and identification

At least one isolated muon is expected in many of the final-state topologies of fourth-generation or vector-like quark production, as well as important SM background processes like top quark pair production. To study these events and ultimately search for new-physics signals in the collected data, an excellent reconstruction and identification of muons in the CMS detector is crucial. Muons do not in general experience severe energy losses in the inner parts of the detector, such that they can easily reach the

dedicated outer muon system. The reconstruction of muons with the CMS detector can be done in different ways [92, 123].

Muon reconstruction

In a so-called local-reconstruction step, hits in the drift-tubes (DT), the cathode strips chambers (CSC) and the resistive plate chambers (RPC) are first combined to form track segments. Next, the track segments in the innermost muon chambers are condidered as seeds, where a Kalman-filter technique is used to propagate the track inside out towards the outermost chamber. This technique incorporates the effects of energy losses and multiple Coulomb scatterings while the muon traverses the detector layers. At each layer, the track fit is updated, and the final fitted track is refitted outside in using the available track segments, in order to improve the determination of the final track parameters. Muons reconstructed like this are referred to as *standalone muons*, because their reconstruction solely makes use of the muon system, and not of the central tracker.

A second type of muon is obtained by matching a standalone muon track to a track in the inner silicon tracker. When a match is found, another Kalman filter is applied starting from the muon system towards the central tracking system. Since several inner tracks might be matched to one muon-system track, only the muon candidate corresponding to the lowest acceptable χ^2 results in a so-called *global muon*.

The third type of muon candidates are obtained by extrapolating all tracker tracks with $p_T > 0.5$ GeV and total momentum $p > 2$ GeV inside out to the muon system, taking into account the magnetic field, the average expected energy loss, and multiple Coulomb scattering in the detector material. If a matching muon track segment is found (based on geometrical coordinates), the associated track in the tracker is said to correspond to a *tracker muon*. Tracker-muon reconstruction has been found to be more efficient than global-muon reconstruction for very low-p_T muons ($p_T \lesssim 5$ GeV), since a global muon typically requires hits in multiple muon stations [123]. Hence tracker muon candidates are used in the reconstruction of soft muons that can appear from a meson decay in a jet originating from a b or c quark.

In the search for fourth-generation quarks presented in Chapter 4, a part of the analysis strategy focuses on the search for an excess of events with two leptons with the same electric charge. This means a correct determination of the charge of muons is needed. The charge of a muon candidate is determined from the direction of the curvature of its track in the magnetic field. A soft charged particle will bend more strongly, hence the charge identification is more precise for lower transverse momentum. Using cosmic-ray muon data recorded in 2008, the CMS experiment measured the charge misidentification probability of muons [124], presented in Figure 3.4. This probability is found to be below or at the promille level for a global muon with $p_T \lesssim 300$ GeV, which corresponds to the bulk of the spectrum of SM background events.

Muon identification and isolation

In the searches for new quarks presented in Chapter 4 and 5, PF muons are used with slightly different identification and isolation requirements, as they correspond to

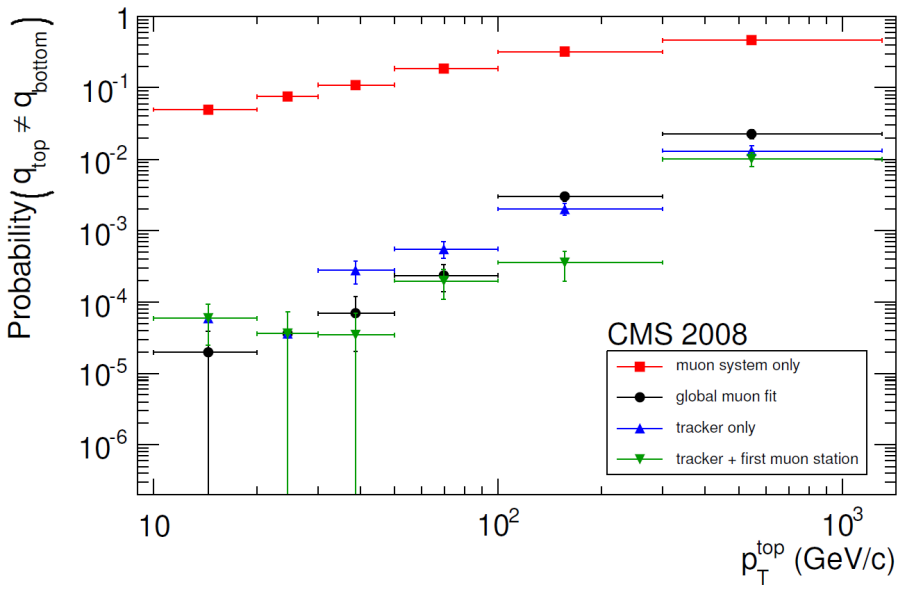

Figure 3.4: The charge misidentification probability for muons has been measured with the CMS detector using cosmic-ray events [124]. The probability that the charge corresponding to a cosmic muon track in the top and bottom half of the detector is measured to be different ($q_{\text{top}} \neq q_{\text{bottom}}$), is shown as a function of the measured muon p_T corresponding to the track in the top half of the detector (p_T^{top}). Various muon-reconstruction algorithms result in a different performance, but whenever the information of the tracker is used, the charge misassignment remains well below 0.1% up to $p_T \approx 100$ GeV, and becomes only about 1% at $p_T \approx 500$ GeV.

$\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV datasets, respectively.

The muon identification criteria [123, 125] that are applied in the analyses are summarized in Table 3.1. In the 7 TeV analysis, the muon candidates are required to be global muons, with a normalized χ^2 (i.e. χ^2/ndf, the χ^2 divided by the number of degrees of freedom) of the muon trajectory fit smaller than 10. There should be at least one hit in the muon system ($n_{MuonHits} > 0$), and more than one muon station should match the global muon track ($n_{MuonStations} > 1$). The number of valid hits $n_{TrackerHits}$ in the inner tracker system needs to be larger than 10, with at least one hit in the pixel detector ($n_{PixelHits} > 0$). The transverse impact parameter d_0 and the longitudinal impact parameter $|\Delta z|$ of the muon with respect to the main primary vertex should be smaller than 0.02 cm and 1 cm, respectively. These impact-parameter requirements ensure a good rejection of cosmic muons and muons originating from additional proton collisions in the event. Finally, the muon candidate should be well separated from reconstructed jets (see Section 3.3.4), $\Delta R(\mu, \text{jet}) > 0.3$, with $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ the separation in (η, ϕ) space. In the 8 TeV analysis, mostly the same identification requirements are applied on muon candidates. However, the criterion on the number of inner tracker hits is replaced by a requirement on the number of layers of the tracker associated to a valid hit, $n_{TrackerLayers} > 5$. The longitudinal impact parameter cut is tightened to $|\Delta z| < 0.5$ cm, to reduce the effect of the increased amount of pile-up interactions in the 8 TeV data set.

Table 3.1: The set of identification criteria applied on muon candidates in the 7 TeV analysis of Chapter 4 differs slightly from the 8 TeV analysis of Chapter 5.

	$\sqrt{s} = 7$ TeV	$\sqrt{s} = 8$ TeV
Global muon	yes	yes
χ^2/ndf	< 10	< 10
$n_{MuonHits}$	> 0	> 0
$n_{MuonStations}$	> 1	> 1
$n_{TrackerHits}$	> 10	$-$
$n_{TrackerLayers}$	$-$	> 5
$n_{PixelHits}$	> 0	> 0
$\lvert d_0 \rvert$	< 0.02 cm	< 0.02 cm
$\lvert \Delta z \rvert$	< 1 cm	< 0.5 cm
$\Delta R(\mu, \text{jet})$	> 0.3	> 0.3

In addition to the identification requirements outlined above, some isolation criteria are applied on muon candidates. Since the PF algorithm reconstructs all charged and neutral hadrons, as well as photons, the relative isolation of a muon with transverse momentum p_T^μ can be quantified as

$$I_{rel} = \frac{\sum p_T^{\text{charged hadron}} + \sum p_T^{\text{neutral hadron}} + \sum p_T^{\text{photon}}}{p_T^\mu}, \tag{3.2}$$

where $p_T^{\text{charged hadron}}$ and $p_T^{\text{neutral hadron}}$ are the transverse momenta of charged and neutral hadrons, respectively, and p_T^{photon} is the transverse momentum of a photon. The sums run over all corresponding identified PF particles within an isolation cone $\Delta R = 0.4$ around the muon.

Equation (3.2) represents the relative isolation definition used in the 7 TeV analysis. However, in the 8 TeV collision run, the relative isolation has been redefined to cope with the higher pile-up environment. In practice, an extra term is introduced in the definition, with the aim of subtracting the effect of neutral particles originating from pile-up interactions:

$$I_{rel}^{\Delta\beta} = \frac{\sum p_T^{\text{charged hadron}} + \max\left(0, \sum p_T^{\text{neutral hadron}} + \sum p_T^{\text{photon}} - 0.5 \sum p_T^{\text{PU}}\right)}{p_T^\mu}. \tag{3.3}$$

This is referred to as a $\Delta\beta$-corrected relative isolation. The sum $\sum p_T^{\text{PU}}$ runs over all PF particles found within a cone of $\Delta R = 0.4$ around the muon by the charged-hadron subtraction algorithm to originate from pile-up interactions. The factor 0.5 is motivated by the observation that jets contain on average about half as many neutral PF particles than charged PF particles [126].

3.3.3 Electron reconstruction and identification

In the search for new heavy quarks, electrons fulfill an equally important role as muons. Due to the lower mass compared to muons by two orders of magnitude, electrons suffer

more from energy loss by Bremsstrahlung. As a result, they will not reach the outer muon detector layers, but instead deposit most of their energy in the electromagnetic calorimeter. In general, the reconstruction of electrons is more complicated compared to muons.

Electron reconstruction

Two methods are used to reconstruct electrons in this thesis, a particle-flow based electron reconstruction [122, 127] and an ECAL-driven reconstruction [128, 129]. The former is used in the 7 TeV analysis presented in Chapter 4, while the latter is used in Chapter 5 since the ECAL-driven type of reconstruction proved to retain a higher identification efficiency in the high-pile-up environment of the 8 TeV proton collision run.

The ECAL-driven electron reconstruction starts by clustering energy deposits in the ECAL subdetectors into a *supercluster*, which is extended more in the ϕ direction than in the η direction. The reason for this is to account for the fact that electrons, while traversing the tracker material, bend in the ϕ direction by the strong magnetic field of the solenoid. Hence, the Bremsstrahlung they emitted results in a large spread of the energy deposits in the ϕ direction. From the energy distribution in the ECAL crystals associated to the supercluster, the electron impact point can be estimated and propagated through the magnetic field towards the inner pixel detector. Compatible hits are then identified and used as seeds for the electron track reconstruction.[4]

The track fit is performed with a Gaussian-sum filter (GSF) [130], which is an extension of the Kalman filter. The latter algorithm assumes gaussian probability density distributions of the energy loss and is effective for the description of multiple scattering, but this is not adequate in the presence of significant Bremsstrahlung effects. The GSF algorithm models the Bremsstrahlung energy loss distribution by a weighted sum of Gaussian distributions, and proves to improve the momentum resolution of electrons compared to the standard Kalman filter technique. The GSF electron track fit is performed from the track seed, and propagated layer by layer by identifying compatible hits. This fitting procedure ends when the outermost tracker layer is reached, resulting in the final reconstructed electron track.

Alternatively, in the PF-based electron reconstruction, the seeding for the GSF track fitting starts from a PF cluster rather than an ECAL supercluster. Electron candidate tracks are linked with PF clusters in the same way as described in Section 3.3.1, but additional clusters originating from electron Bremsstrahlung photons are recovered in the linking process. This is done by extrapolating tangents to the GSF track to the ECAL surface from the intersection points between the track and each of the tracker layers. In addition, the PF algorithm reconstruction incorporates a low-level identification procedure, by considering a Boosted Decision Tree (BDT) multivariate discriminator based on various tracking variables as well as variables related to the matching between the track and the ECAL clusters. The purpose of this discriminator is to distinguish electrons from charged hadrons. Particle-flow electrons are then

[4] Another reconstruction method with *tracker-driven* seeding is used in the CMS collaboration, which is more efficient for low-p_T electron reconstruction, while the ECAL-driven seeding performs better for high-p_T isolated electrons.

Table 3.2: The set of cut-based identification criteria applied on electron candidates in the 7-TeV analysis of Chapter 4 depends on whether the electron is associated to a supercluster in the ECAL barrel or in the endcap.

	Barrel	Endcap
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
H/E	< 0.04	< 0.025
$\Delta\phi$(supercluster,track)	< 0.06	< 0.03
$\Delta\eta$(supercluster,track)	< 0.004	< 0.007

defined as electrons with a BDT discriminator value exceeding -0.1. For instance, imposing this threshold retains roughly 65% of non-isolated electrons from a B-hadron decay in jets, while rejecting about 99% of pions [131]. The efficiency for isolated electrons is about 99% for electrons in a $Z \to e^+e^-$ data sample.

The electric charge identification for electrons, just like for muons, is derived from the direction of the curvature of the reconstructed track. The charge misidentification probability for electrons has been measured in data from $Z/\gamma^* \to e^+e^-$ events to be within 0.1 to 0.4%, increasing with the pseudorapidity of the electron [132].

Electron identification and isolation

The collection of PF electrons is used in the 7 TeV analysis of Chapter 4 [129]. Just like for muons, the transverse and longitudinal impact parameters with respect to the primary vertex are required to be $|d_0| < 0.02$ cm and $|\Delta z| < 1$ cm. The electron should also be separated from any reconstructed jet in the event such that $\Delta R > 0.3$. For the electron identification a set of criteria is defined, corresponding to an electron reconstruction efficiency of about 80%. These requirements involve four variables, for which the thresholds imposed are summarized in Table 3.2. The cuts differ for electrons associated to superclusters in the barrel ($|\eta_{supercluster}| < 1.4442$) and the endcap ($|\eta_{supercluster}| > 1.5560$) of the ECAL subdetector. The variable $\sigma_{i\eta i\eta}$ represents the width of the electron shower in the η direction as calculated from a 5×5 array of crystals around the electron energy deposit in the ECAL. The H/E variable is calculated as the ratio between the energy deposited in the HCAL and the energy deposited in the ECAL, and should be low for electrons. The ϕ and η variables indicate the spatial matching distance between the electron GSF track and the ECAL supercluster. Here, the position of the supercluster corresponds to an energy-weighted mean position.

The electron candidates are required not to originate from a photon conversion to an electron-positron pair. Electron candidates are rejected if the most inner expected hit of the reconstructed track is missing, or if a second track is found close to the electron track. The latter requirement is quantified as $\Delta \tan\theta < 0.02$, calculated as the difference between the $\tan\theta$ of the tracks, and $|Dist| < 0.02$ cm, where $|Dist|$ measures the distance between the two tracks in the transverse plane at the point where the tracks are parallel.

The relative isolation of a PF electron with transverse momentum p_T^e can be ex-

pressed in the same manner as Equation (3.2) for PF muons:

$$I_{rel} = \frac{\sum p_T^{\text{charged hadron}} + \sum p_T^{\text{neutral hadron}} + \sum p_T^{\text{photon}}}{p_T^e}. \tag{3.4}$$

The isolation is calculated using the PF reconstructed charged hadrons, neutral hadrons and photon within a cone of $\Delta R = 0.4$ around the electron.

As mentioned before, in the 8 TeV analysis of Chapter 5 we do not use the collection of PF electrons, due to a loss of identification efficiency in the high pile-up scenario of the 2012 proton collision run. Moreover, the identification definition is altered [133]. The transverse and longitudinal impact parameter requirements are the same, as well as the requirement that electrons should not have a missing hit in the innermost tracker layer compatible with the GSF track. However, rather than using the cut-based identification criteria from Table 3.2, electron candidates are identified using a multivariate (MVA) discriminator, referred to as the $mvaId$ variable. The MVA technique combines different observable into one discriminator that separates well-identified electrons from fake electrons. The used variables include for instance tracking variables like the χ^2 of the GSF track fit, ECAL-based variables like the shower width or the ratio of the energy deposited in a 1×5 array of crystals divided by the energy in a 5×5 array, the tracker-ECAL matching variables $\Delta\phi$ and $\Delta\eta$, and the H/E variable. The resulting discriminator tends to have higher (lower) values for good-quality (bad-quality) electrons.

The photon conversion rejection for electrons in the 8 TeV data set is performed somewhat differently than for the 7 TeV analysis. Now pairs of charged-particle tracks are used to fit a conversion vertex. If the fit corresponds to a probability higher than 10^{-6}, the transverse decay length is greater than 2 cm, and the associated tracks have no hits before the position of the fitted vertex, a conversion vertex is said to be found. In that case, the electron is rejected.

In order to cope with the average high pile-up activity in event during the 8 TeV collision run, the relative electron isolation has been modified, using a similar idea of subtracting the expected pile-up effect as done for muons in Equation (3.3). This time, however, an *effective-area* corrected relative isolation is defined:

$$I_{rel}^{EA} = \frac{\sum p_T^{\text{charged hadron}} + \max\left(0, \sum p_T^{\text{neutral hadron}} + \sum p_T^{\text{photon}} - \rho \cdot A_{eff}\right)}{p_T^e}, \tag{3.5}$$

where the sums run over the PF particles[5] in a cone of $\Delta R = 0.3$ around the electron. The variable ρ is defined as the 'transverse-momentum density' of the event, and A_{eff} is the effective area, a measure of the extension of the object that one considers (in this context a cone around the electron) in which one is affected to a certain extend by pile-up effects. The ρ and A_{eff} variables also appear later in the context of jet energy corrections in Section 3.3.4.

[5]Note that while the PF electrons are not used in the 8 TeV analysis, the other particles in the event are identified using the PF algorithm.

3.3.4 Jet reconstruction

In Section 3.1.2 and 3.1.3 we explained that quarks and gluons are not observed directly in the detector. Instead, after a fragmentation and hadronization process that can be phenomenologically modeled, these partons result in a collection of observable hadrons. These particles can be clustered into jets, and in order to obtain a good estimation of the jet energy, one has to correct for various sources affecting the reconstructed energy scale. Furthermore, the energy resolution of simulated jets, that quantifies to which extend the parton energy or transverse momentum can be resolved, needs to be calibrated to the resolution observed in data. Finally, the identification criteria applied on jets used in the analysis will be outlined at the end of this section.

Jet clustering

Various jet-clustering algorithms exist, with a wide range of theoretical and topological properties. They all start from a set of particles or calorimeter deposits and cluster them together in jets. Two important properties are required for any theoretically well-behaved jet-clustering algorithm, namely the insensitivity to additional soft radiation ('infrared safety' of the algorithm) and the insensitivity to the collinear splitting of a hard parton ('collinear safety').

The most common algorithm used within the CMS collaboration is the so-called anti-k_T algorithm [134], which is both infrared and collinear safe. In order to decide whether or not two objects i and j should be clustered together, an abstract 'distance' d_{ij} between the objects is defined. In addition, a distance d_{iB} between object i and the beam is introduced. The k_T-like algorithms all use a similar definition of these distances:

$$d_{ij} = \min\left(k_{Ti}^{2p}, k_{Tj}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2}, \tag{3.6}$$

$$d_{iB} = k_{Ti}^{2p}, \tag{3.7}$$

where k_{Ti} and k_{Tj} are the transverse momenta of objects i and j, respectively. The distance $\Delta R_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}$ between the two objects is calculated in the (y, ϕ) space, with y the rapidity as defined in Equation (2.3), and the dimensionless parameter R is referred to as the radius or distance parameter of the jet-clustering algorithm. The parameter p regulates the relative importance of the momenta and spatial distance scales, and is set to $p = -1$ in the anti-k_T algorithm. An interesing property of the anti-k_T algorithm is that it produces cone-shaped jets in (y, ϕ) space that are robust with respect to the addition of soft particles, such that R can indeed be interpreted as a cone radius parameter. The algorithm first considers the minimum distance among all d_{ij} and d_{iB} in the event. If this minimum corresponds to d_{ij} and is smaller than d_{iB}, the momenta of the objects i and j are summed, resulting in a new recombined object that replaces objects i and j. If the minimum is d_{iB} instead, this indicates that the object i is well separated from all other objects j, and is therefore considered as a jet. The objects constituting this jet are removed from the list, and the clustering procedure is continued, until the list of input objects is exhausted.

Particle-flow jets are clustered from the reconstructed PF particles in the event. The clustering algorithm can also be applied to the list of particles generated by PYTHIA, resulting in so-called generator-level jets (GenJets). The energy of the jets used in this thesis is obtained from the so-called E recombination scheme, which dictates that the four-momenta of the jet constituents should be added to form the four momentum of the jet. However, a reliable determination of the jet energy is not so straightforward, since many effects can distort the energy estimation. These effects include the calorimeter response, the limited particle reconstruction efficiency, the fact that low-p_T charged particles can bend out of the jet cone due to the strong magnetic field, and the underlying event and pile-up interaction activity. This necessitates the careful calibration of the jet energy, as described in the following paragraph.

Jet energy corrections

The CMS collaboration has developed a factorized approach to correct the jet energy scale from simulation and data toward a more performant estimate [135–137]. The jet-energy corrections can be performed in the following sequential steps:

- **Pile-up correction (L1).** The first level of jet energy correction aims to correct for the effect of pile-up interactions. This can be studied in the simulation by comparing the energy of generator-level jets to the energy of reconstructed raw (i.e. uncorrected) PF jets, since particles originating from pile-up interactions are not considered in generator-level jets (but only entered at the detector-simulation level, as mentioned before). The L1 pileup correction used in CMS [138] is based on the idea that the additional energy from pile-up interactions can be subtracted using the concept of *jet areas* [139, 140]. The L1 corrected transverse momentum $p_{T,j}^{L1}$ of a PF jet j with uncorrected transverse momentum $p_{T,j}^{raw}$ and pseudorapidity η_j can be written as

$$p_{T,j}^{L1} = p_{T,j}^{raw} - \text{Offset}\left(\rho, A_j, \eta_j, p_{T,j}^{raw}\right). \tag{3.8}$$

  Here, the offset term depends on the pseudorapidity of the jet, the average p_T density ρ in the event, and the effective jet area A_j. Only in the 8 TeV analysis, the offset calibration is also considered to be a function of the transverse momentum $p_{T,j}^{raw}$. The effective jet area A_j is determined by injecting a large number of very soft particles in the event prior to the jet clustering. The spread in (y, ϕ) space of these soft particles in each jet defines the jet area. The p_T density ρ is calculated as the median of the $p_{T,j}/A_j$ ratios of all jets in the event. To correct for differences in the offset in observed data and simulation, an additional residual correction is applied on jets in data. The exact parametrization of the offset term in Equation (3.8) is different in the 7 TeV analysis compared to the 8 TeV analysis, but the calibration results in both cases in a correction that takes into account energy from pile-up on an event-by-event and on a jet-by-jet basis. The resulting jet energy response is mostly independent of the number of pile-up interactions in the event.

- **Relative η and absolute p_T correction (L2L3).** After the L1 jet energy correction, the jet energy response $\mathcal{R} = p_{T,j}^{L1}/p_{T,GenJet}^{L1}$ is still non-unity and depends

on the pseudorapidity of the jets due to the non-uniformity of the calorimeter response in η. The corresponding calibration factors are determined from QCD multijet simulated events by measuring the jet energy response $\mathcal{R}$ as a function of the L1-corrected PF-jets η and p_T. This results in the following L2L3-corrected jet transverse momentum:

$$p_{T,j}^{L1L2L3} = \frac{p_{T,j}^{L1}}{\langle \mathcal{R}(p_{T,j}^{L1}, \eta_j) \rangle}. \tag{3.9}$$

- **Residual η and p_T corrections (L2L3Residual).** Since the correction given in Equation 3.9 is obtained from simulation, additional residual corrections are applied on the jets observed in data [141]. First, the L2Residual correction aims to remove the η dependence of the measured jet energy response in the data. Correction factors can be obtained by requiring a transverse-momentum balance of jets in di-jet events. Here, the jets in the barrel ($|\eta| < 1.3$) are used as a reference. The L3Residual correction can be determined from data by studying Z/γ^*+jet events where the jet is back-to-back with a $Z \to \ell^+\ell^-$ boson or photon. Since the momentum of muons, electrons and photons is measured relatively precisely, the p_T balance of the reconstructed Z or γ^* boson and the jet can be exploited.

Additional jet energy correction steps have been developed in the CMS collaboration, to correct for jet-flavor dependencies of the energy response, the underlying event activity and the energy of the jet at parton level (L5-L7). These optional corrections are not used in this thesis, as they generally involve smaller corrections compared to the recommended L1, L2 and L3 corrections.

The jet energy scale corrections outlined above are associated to uncertainties that can be propagated as systematic uncertainties in physics analyses. These uncertainties include for instance observed differences of the p_T pile-up offset term in data and simulation, uncertainties in the jet energy resolution, and in the energy scale of the photons and leptons used to determined residual corrections.

Jet energy resolution

The energy resolution of PF jets has been measured at 7 TeV from QCD di-jet and γ+jet events using the p_T balance the corresponding final-state objects [135]. The jet resolution measured in data has been observed to be 5% in the barrel to 29% worse in the endcap compared to the resolution obtained from simulation. Hence an η-dependent smearing of the energy resolution of simulated PF jets is applied, such that the resolution describes the one observed in data. The uncertainty in the scale factors applied to the jet resolution can be propagated as a systematic uncertainty in a physics analysis.

Jet identification

Sets of identification criteria are imposed on the PF jets used in the analyses presented in Chapter 4 and 5 to retain good-quality jets. A jet is required to consist of at least

two particles. The fractions of the jet energy corresponding to either neutral hadrons or photons should not exceed 99%. In addition, for jets within the tracker acceptance ($|\eta| < 2.4$), the fraction of energy attributed to electrons should also be less than 99%, while there should at least be some energy deposited in the HCAL that corresponds to a charged hadron.

3.3.5 Identification of jets from b quarks

An essential ingredient in many searches for new physics beyond the Standard Model is the identification of jets originating from b quarks, via so-called b-tagging algorithms. Since chiral fourth-generation quarks are assumed to decay to third-generation quarks, the final state of such topologies are expected to contain at least one b jet. Even in the search for vector-like quarks decaying to light-flavor quarks, b-tagging is an important aspect, since it can be applied for instance to *reject* events with b jets to retain mostly the light-flavor jet final states.

b-jet identification algorithms

Several b-tagging algorithms have been developed in the CMS collaboration [142]. The main idea behind all of these algorithms exploits the fact that the hadronization of bottom quarks involves a B meson, a bound state of a bottom quark and an up, down, strange or charm quark. These mesons have a short lifetime, but live long enough to travel a measurable distance in the inner pixel detector before decaying to other hadrons or leptons. In this section, we will only outline the most relevant ones for this thesis, which are based on the impact parameter of charged-particle tracks with respect to the primary vertex, and on the reconstruction of a secondary vertex that would correspond to the meson decay vertex. This is illustrated in Figure 3.5, depicting three jets of which one contains displaced tracks, originating from a secondary-vertex decay that is itself displaced with respect to the primary vertex. The track impact parameter (IP) with respect to the primary vertex is defined as the distance between the vertex and the tangent to the track extrapolated from the point of closest approach between the track and the jet axis. This variable tends to be larger for displaced tracks than tracks originating from the primary vertex. The measurement of the impact parameter has a corresponding uncertainty σ_{IP}, such that a more practical observable is the so-called track *impact parameter significance* IP/σ_{IP}. When a particle is produced along the jet axis starting from the interaction point, one can attach a *sign* to this variable that is positive, while for tracks that have a point of closest approach to the vertex in the opposite direction of the jet axis, this sign is negative. The impact parameter can be measured in the three-dimensional space, or in the two dimensions of the transverse plane.

- **Track counting b-tagging algorithms.** The impact parameter significance can be used as a discriminating variable between b jets and light-flavor jets. The tracks associated to the jet are first ordered according to decreasing impact parameter significance. When the IP significance of the second-highest IP

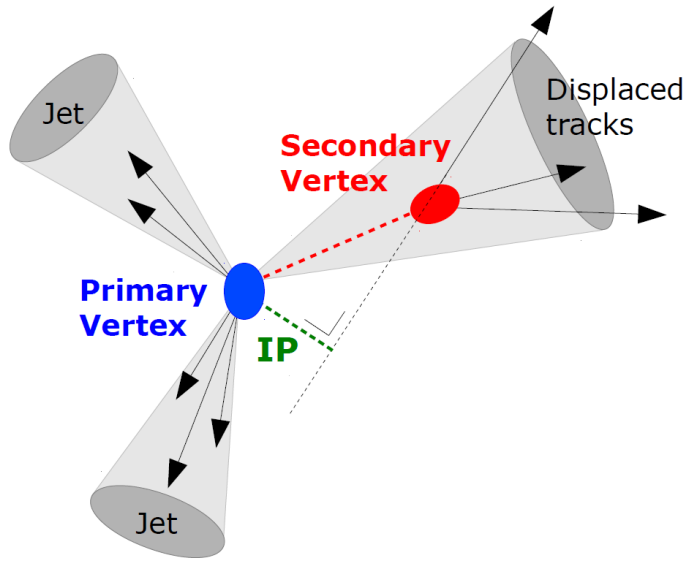

Figure 3.5: A jet originating from a b quark often involves a displaced secondary vertex, corresponding to a B-meson decay. Tracks originating from this secondary vertex hence appear to be displaced with respect to the primary vertex. The secondary vertex can not always be reconstructed, but the information of the displaced tracks themselves, quantified by the track impact parameter IP, can still be used to identify such a jet as a b jet.

significant track is used as discriminating variable, this is referred to as the *Track-Counting High-Efficiency* (TCHE) discriminator. When the IP significance of the third-highest IP significant track is used, one obtains the *Track-Counting High-Purity* (TCHP) discriminator. Since not all B-meson decays involve three well-reconstructed displaced charged-particle tracks, the efficiency of identifying b jets with the TCHP algorithm is decreased with respect to the TCHE algorithm, but the purity is increased, as the probability for three tracks to be displaced in non-b jets is very low. These b-tagging algorithms are used in the fourth-generation quark search presented in Chapter 4.

- **Secondary vertex based algorithms.** The secondary vertex originating from a B-hadron decay can be reconstructed using specialized algorithms like the Adaptive Vertex fitter [99]. The so-called *flight distance significance* D/σ_D, defined as the distance between the reconstructed secondary vertex and the primary vertex, normalized by its uncertainty, is used in the Simple Secondary Vertex (SSV) algorithm to distinguish between b-quark jets and non-b-quark jets. Just like the track-based algorithms, two variants are defined, dedicated to obtain a high b-jet efficiency (Simple Secondary Vertex High Efficiency, SSVHE) or a high b-jet purity (Simple Secondary Vertex High Purity, SSVHP). These algorithms differ in the amount of tracks that are required to be associated to the vertex, namely two for the SSVHE algorithm, and three for the SSVHP algorithm.

- **Combined Secondary Vertex algorithm.** A particularly powerful b-tagging algorithm is obtained if one combines the information of track-based variables with the properties of a reconstructed secondary vertex in a jet. This is the aim of the Combined Secondary Vertex (CSV) algorithm, that uses a multivariate analysis technique to combine multiple sensitive variables into one discriminator.

  First, the Adaptive Vertex fitter algorithm tries to reconstruct secondary vertices using good-quality tracks of the jet. If such a vertex is reconstructed, the tracks associated to that vertex are removed from the list of tracks to be considered and the fitting procedure is repeated. In case a vertex reconstruction fails and there are at least two reconstructed tracks in the jet, an attempt is made to reconstruct a 'pseudo-vertex' using the tracks with a signed impact parameter significance in the transverse plane exceeding 2. Such a pseudo-vertex can be considered as a looser definition of an actual reconstructed vertex and without performing a vertex fit. Hence, one can divide the jets in three vertex categories; jets with a reconstructed vertex, a pseudo-vertex and without a vertex. The latter contains all jets that do not fall in the former two categories.

  Next, discriminating variables are defined in these three categories. These include:

  - The flight distance significance in the transverse plane.
  - The mass of the secondary vertex, calculated as the invariant mass of the charged particle tracks associated to the secondary vertex.
  - The number of tracks associated to the secondary vertex. This track multiplicity can be large for a B meson decay.
  - The pseudo-rapidity, relative to the jet axis, of the secondary-vertex tracks.
  - The vertex energy fraction, defined as the ratio of the energy of the summed four-momenta of all secondary-vertex tracks and the energy of the summed four-momenta of all tracks associated to the jet.
  - The signed transverse impact parameter significance of the track that raises the invariant mass of the set of considered secondary-vertex tracks (sorted in signed transverse impact parameter significance) above the charm-quark mass.
  - The number of tracks associated to the jet.
  - The three-dimensional impact parameter significance for all tracks associated to the jet.

  The last two types of variables are the only ones available if no reconstructed or pseudo-vertex is found. Finally, these variables are combined in a likelihood-ratio, one for each vertex category:

$$\text{Likelihood ratio} = \frac{\Pi_i \; p_i^b(x_i)}{\Pi_i \; p_i^b(x_i) + \Pi_i \; p_i^{c/udsg}(x_i)}. \tag{3.10}$$

  Here $p_i^b(x_i)$ corresponds to the probability density function of variable x_i for b jets, and $p_i^{c/udsg}(x_i)$ the probability density function for c or u, d, s, g jets.

These functions are obtained from simulated jets. The products run over all input variables listed above that are relevant for the considered vertex category. Afterwards, the likelihood-ratio outputs (considering either c or $udsg$ background jets) related to each of the three vertex categories are combined into another likelihood ratio. The final CSV discriminator is obtained by summing the charm and light-flavor discriminators with a weight of 25% and 75%, respectively.[6]

The advantage of this multivariate approach is that even when one fails to reconstruct a secondary vertex, which happens in about 65% of the cases, the information of the impact parameter significances still provides discriminating power between b and non-b jets. This CSV b-tagging algorithm is used in the vector-like quark search presented in Chapter 5.

Recently we developed within the CMS collaboration a new CSV algorithm (combining discriminating variables in a neural network) that shows significant improvements with respect to the standard CSV algorithm. When including a novel vertex reconstruction algorithm, the so-called Inclusive Vertex Finder (IVF) [143], the performance is improved even further. The IVF algorithm reconstructs all secondary vertices in an event, independent of the jet reconstruction. This technique improves the vertex reconstruction efficiency significantly, resulting in a higher amount of jets in the sensitive reconstructed-vertex category of the CSV algorithm. The new neural-network based CSV algorithm (CSVV2) will be released for the LHC Run 2, starting in 2015, colliding protons at a center-of-mass energy of $\sqrt{s} = 13$ TeV.

Performance of b-jet identification algorithms

Given a b-tagging algorithm, one can impose a threshold on the corresponding discriminating variable for a jet. When a jet has a discriminator value exceeding the threshold, the jet is said to be b-tagged, as this corresponds to the b-jet hypothesis (with a certain probability).

The performance of b-tagging algorithms can be defined as the ability to correctly identify jets originating from b quarks while suppressing the probability to misidentify jets from non-b quarks. To illustrate this, Figure 3.6 shows, for various algorithms used in the CMS collaboration, the probability that non-b jets are misidentified as b jets as a function of the efficiency of b jet identification. These performance curves are obtained from simulated $t\bar{t}$ events (at 8 TeV), by varying the threshold on the discriminator values of a given algorithm, and calculating, for each jet flavor, the efficiencies of jets passing this threshold. A performance curve shifted more towards the lower right, corresponds in general to a better performance. From performance curves like in Figure 3.6, b-tagging *operating points* (alternatively called *working points*) are defined, corresponding to particular minimal thresholds on the discriminator. These generally involve a 'loose', 'medium' and 'tight' operating point, defined such that the efficiency to tag light-parton jets is close to 10%, 1% and 0.1%, respectively for jets with an average p_T of about 80 GeV. However, these percentages strongly depend on the

[6]These fractions are chosen to match the approximate non-b jet flavor composition in top quark pair events.

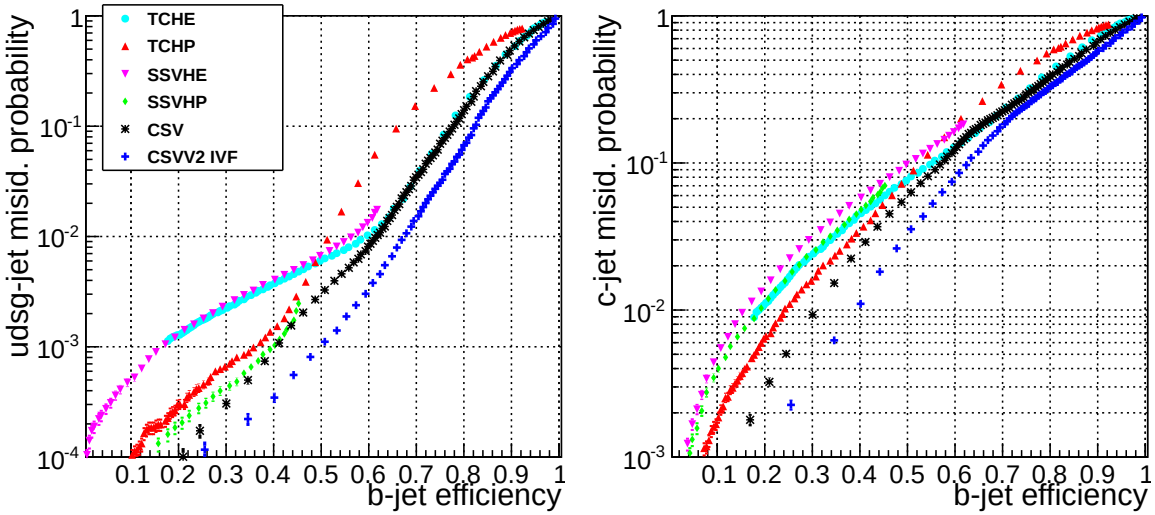

Figure 3.6: The b-tag misidentification probability for light-flavor parton jets (left) and c-quark jets (right) as a function of the b-tag efficiency of b-quark jets can be used as a measure to evaluate the performance of the b-tagging algorithms from simulation [142]. The TCHE and TCHP discriminators result in a particularly good performance towards higher b-jet efficiency and lower non-b-jet efficiency, respectively. The CSV discriminator proves to have a good performance over the whole efficiency spectrum, clearly improving the performance compared to track-based (TCHE and TCHP) and secondary vertex based (SSVHE and SSVHP) algorithms. I contributed to the development of a new CSV algorithm based on a neural network and using the Inclusive Vertex Finder algorithm (CSVV2 IVF), drastically improving the performance with respect to the other algorithms.

kinematic phase space of the jets. For PF jets with $p_T > 30$ GeV in 7 TeV simulated proton collisions [144], the b-tag efficiencies correspond to about 77%, 64% and 50% for the loose, medium and tight working points of the TCHP algorithm, respectively, while the combined c+$udsg$ misidentification probabilities are about 22%, 4% and 1%. For PF jets with $p_T > 30$ GeV at 8 TeV, the loose, medium and tight operating points of the CSV discriminator result in a simulated b-tag efficiency of about 83%, 70% and 52%, respectively. The corresponding total misidentification probabilities are about 15%, 3% and 1%. All these efficiencies can be measured in data as well [142, 144, 145], and jet p_T and η dependent scale factors have been determined to correct the b-tag and 'mistag' (i.e. non-b tag) efficiencies in simulation to those observed in data.

3.3.6 Missing transverse energy reconstruction

Neutrinos are commonly produced as final-state particles in proton collisions, for instance originating from the leptonic decay of a W boson. Hence, many SM background and hypothetical new heavy quark processes correspond to topologies with at least one neutrino. However, the neutrino only interacts very weakly, and escapes from the de-

tector without interacting with the detector material. Nevertheless, the presence of neutrinos can be inferred from the momentum balance in the transverse plane, since the net balance before the collision was zero. This is referred to as the missing transverse momentum, or in a looser but more commonly used terminology, the missing transverse energy. A momentum balance along the beam axis cannot be used, since the longitudinal momentum of the center-of-mass system of the colliding partons is not known on an event-by-event basis.

The missing transverse energy vector [146] in the event is calculated from the vector sum of the transverse momenta $\vec{p}_T$ of all reconstructed particles:

$$\vec{\not{E}}_T^{raw} = -\sum \vec{p}_T. \tag{3.11}$$

Since the energy of jets are corrected, as explained in Section 3.3.4, these corrections are propagated to the missing transverse energy. The propagation of L2L3 jet energy scale corrections to the $\vec{\not{E}}_T$ yields:

$$\vec{\not{E}}_T = \vec{\not{E}}_T^{raw} - \sum_{PFjets} \left(\vec{p}_T^{\,L1L2L3} - \vec{p}_T^{\,L1}\right), \tag{3.12}$$

where the sum runs only over PF jets with a fully-corrected transverse momentum $\vec{p}_T^{\,L1L2L3}$, hence exceeding 10 GeV, since jet energy corrections below this threshold are not reliable. The magnitude of the missing transverse energy vector is referred to as the missing transverse energy $\not{E}_T$, but whether this symbol indicates the magnitude or the vector itself, should be clear from the context.

Chapter 4

Search for the quarks of a sequential fourth generation

The existence of three generations of fermions has been firmly established experimentally. Before 2012, the possibility of a fourth chiral generation of fermions had not been excluded, although it was already strongly constrained by precision measurements of electroweak observables. Previous direct searches at hadron colliders have considered pair production of only *one* of the fourth-generation quarks, and single production only in the context of vector-like quarks. At the time of the search presented in this Chapter, the most stringent limits excluded at 95% CL the existence of a down-type (up-type) fourth-generation quark with a mass below 611 (570) GeV [75, 79]. If a fourth generation of fermions would exist, both the down-type and the up-type fourth-generation quarks should be able to be produced via strong and electroweak processes in proton collisions. Hence, to increase the sensitivity and to use a consistent approach while searching for a new generation of quarks, we have developed a simultaneous search for the up-type and down-type fourth-generation quarks, based on both the electroweak and strong production mechanisms. Since the observed proton-proton collision data at $\sqrt{s} = 7$ TeV recorded by the CMS experiment is found to agree with the expectations from the Standard Model, we set lower limits on the mass of the fourth-generation quarks.

In Section 4.1, we will describe our assumptions on the considered fourth-generation model, which are directly related to the applied analysis strategy. The studied collision data as well as the simulated Standard-Model background and fourth-generation signal processes will be discussed in Section 4.2. The basic selection criteria applied on the reconstructed physics objects are outlined in Section 4.3, with additional optimized selection criteria for candidate fourth-generation signal events described in Section 4.4. Next, the observables sensitive to the presence of a possible fourth-generation signal are constructed in Section 4.5. The method to perform the combined search is outlined in Section 4.6, where it is crucial to take into account the relevant systematic uncertainties in the background and signal modeling. Finally, the results of this combined search for fourth-generation quarks are presented and discussed in Section 4.7.

4.1 Model assumptions

A new generation of chiral fermions requires not only the existence of two additional quarks and two additional leptons, but also an extension of the quark-mixing (CKM) and lepton-mixing (PMNS) matrices. As illustrated in Section 1.2.3, new mixing-matrix elements are constrained by the requirement of consistency with electroweak precision measurements. However, the electroweak corrections to the oblique parameters are mainly influenced by the mass differences between the fourth-generation leptons or quarks. In particular, scenarios with a mass difference between the fourth-generation quarks smaller than the mass of the W boson are preferred and even fourth-generation quarks with degenerate masses are allowed.

If fourth-generation t' and b' quarks exists, their electroweak production cross sections and decay branching fractions will be governed by an extended 4×4 CKM matrix, $V_{\mathrm{CKM}}^{4\times4}$. For simplicity, we assume a model with one free parameter, A, where $0 \leq A \leq 1$:

$$V_{\mathrm{CKM}}^{4\times4} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} & V_{ub'} \\ V_{cd} & V_{cs} & V_{cb} & V_{cb'} \\ V_{td} & V_{ts} & V_{tb} & V_{tb'} \\ V_{t'd} & V_{t's} & V_{t'b} & V_{t'b'} \end{pmatrix} = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(0) & \mathcal{O}(0) & 0 \\ \mathcal{O}(0) & \mathcal{O}(1) & \mathcal{O}(0) & 0 \\ \mathcal{O}(0) & \mathcal{O}(0) & \sqrt{A} & \sqrt{1-A} \\ 0 & 0 & -\sqrt{1-A} & \sqrt{A} \end{pmatrix}, \tag{4.1}$$

where the complex phases are not shown for clarity. Note that the extra CP violation resulting from additional phases is not relevant in the context of the presented search, as this will only affect dedicated precision measurements. Our fourth-generation model assumptions can be categorized into three (intertwined) classes that are discussed below: the quark mixing, the considered fourth-generation processes and the allowed masses and decays of the new quarks.

4.1.1 Quark mixing

Within the CKM model (4.1) presented above, mixing is allowed only between the third and the fourth generations. This is a reasonable assumption since it is observed from the 3×3 CKM matrix (1.28) that the mixing between the third and the first two generations is small. Nevertheless, the exclusion limits presented in this analysis could be relaxed if there is a fourth generation that mixes only with the first two generations, or when the size of the mixing with the third generation is about the same as the mixing with the first two generations. The implications of the assumption of small mixing with lower generations is not straight-forward, for instance due to the requirement in our analysis of jets identified as b jets. The effect of a different model of the CKM matrix should therefore be carefully examined.

With this search, we will set limits on the masses of the fourth-generation quarks as a function of the $V_{\mathrm{CKM}}^{4\times4}$ parameter A. Since $\sqrt{A} = |V_{tb}|$, the lower limit of $|V_{tb}| > 0.81$ from single-top production cross-section measurements [25] translates into a lower limit on the mixing between the third- and fourth-generation quarks in our model of $A > 0.66$. In case A *exactly* equals 1, hence assuming no mixing of the third generation with the fourth generation, the new quarks would be stable. Stable or quasi-stable fourth-generation quarks would lead to different signals in particle detectors than

usually considered in direct searches. Although searches for long-lived new quarks are worthwile doing, these are outside of the scope of this thesis. Nevertheless, our results will be valid for values of A extremely close to unity. The distance between the primary vertex and the decay vertex of the fourth-generation quarks is less than 1 mm for $1 - A > 2 \times 10^{-14}$, a number obtained using the leading-order formula for the decay width of the top quark in which the top-quark mass is replaced with a fourth-generation quark mass of 600 GeV.[1]

4.1.2 Fourth-generation processes

We search for fourth-generation quarks that are produced in pairs through the strong interaction, namely $b'\overline{b'}$ and $t'\overline{t'}$, or through electroweak production, in particular $t'b$, tb', and $t'b'$. The charges are omitted in the notation of the latter processes, but we do consider the production of particles and anti-particles coherently. While the cross sections of the pair-production processes (analogous to Figure 1.1) do not depend on the value of A, the production cross sections of the $t'b$ and tb' processes depend linearly on $(1 - A)$ (see Figure 1.3), and the SM single-top and $t'b'$ cross sections on A. We do not include the $t't$, $t't'$, $b'b$ and $b'b'$ electroweak production channels. The Feynman diagrams of these subleading processes involve multiple vertices associated to the CKM-matrix elements, and can be neglected. The last three production mechanisms are especially suppressed compared to the other ones. The $t't'$ process requires two heavy quarks to be produced through the electroweak coupling, and the $b'b$ and $b'b'$ processes involve heavily-suppressed gluon splitting to a top-quark pair. Furthermore, for none of these production channels NLO cross sections are available in the literature. Considering LO cross sections would result in relatively smaller cross sections for these processes compared to the other processes. Omitting them will result in a worse sensitivity to the fourth-generation quark processes and hence in more conservative limits on the masses of the new quarks and on the value of A.

4.1.3 Masses and decay topologies

Motivated by the allowed parameter space from precision measurements (see Figure 1.6), we assume the t' and b' quark masses to be degenerate within 25 GeV. In the case they are degenerate, and A is not *exactly* 1, they will decay in 100% of the cases to the third-generation quarks, since the decay of one fourth-generation quark to the other via a massive boson is kinematically not allowed. Even for non-zero mass differences, the branching fractions of the $t' \to bW$ and the $b' \to tW \to (bW)W$ decays are close to 100%, provided that the mass difference is small [23]. For instance, for a mass splitting of 25 GeV, and for $V_{t'b} = 0.005$ (which would correspond to $A = 0.99975$ in our model), less than 5% of the decays will be $t' \to b'W^*$ (in the case $m_{t'} > m_{b'}$) or $b' \to t'W^*$ (in the case $m_{t'} < m_{b'}$). For larger values of $V_{t'b}$, the branching fractions of $t' \to b'W^*$ (or $b' \to t'W^*$) decrease even further. Therefore, the decay chains remain

[1] The LO formula for the decay width of the top quark is $\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} \times |V_{tb}|^2$, with $G_F = 1.17 \times 10^{-5}$ GeV^{-2} the Fermi coupling constant.

unchanged as long as the mass splitting is relatively small. We expect the following decay topologies to dominate:

- $t'b \to bWb$;
- $t'\overline{t'} \to bWbW$;
- $b't \to tWbW \to bWWbW$;
- $b'\overline{t'} \to tWbW \to bWWbW$;
- $b'\overline{b'} \to tWtW \to bWWbWW$.

These decay chains imply that two jets from b quarks and one to four W bosons are expected in the decay chains for fourth-generation quarks produced both through the electroweak and the strong interaction. The W bosons subsequently decay to either hadronic ($W \to q\bar{q}$) or leptonic ($W \to \ell\bar{\nu}_\ell$) final states. Events with either one isolated lepton (muon or electron) or two same-sign leptons or three leptons will be selected. The events will be classified according to the number of observed W bosons, in order to optimize the sensitivity to the signal processes listed above.

4.2 Data and simulation

The search for the fourth-generation quarks is performed using $\sqrt{s} = 7$ TeV proton-proton collisions recorded by the CMS experiment at the LHC. We have analyzed the full dataset collected in 2011 corresponding to an integrated luminosity of (5.0 ± 0.1) fb^{-1}. This data has been certified to ensure the good operational status of all subdetectors at the time of data taking. In addition, several filters have been applied to reject electronic noise as well as backgrounds from protons hitting the edge of the beam pipe, as these effects might resemble energy deposits from genuine proton-proton collisions in the detector. Various Standard-Model background processes can mimic the expected signatures from fourth-generation quarks. In Sections 4.2.1 and 4.2.2, we will discuss the simulation of these SM-background and the fourth-generation quark processes, respectively. In Section 4.2.3 we briefly describe the trigger requirements applied when recording the data.

4.2.1 Standard-Model background simulation

The observed data are compared to simulated data generated with powheg 301 for the single-top process, pythia 6.4.22 for the diboson processes, and MadGraph 5.1.1 for the other Standard-Model processes. The powheg and MadGraph generators are interfaced with pythia for the decay of the particles as well as for the hadronization using a CMS custom underlying-event tune (tune Z2). The matching of the matrix-element partons to the parton showers is obtained using the MLM matching algorithm. The cteq6L1 leading-order parton distribution functions are used in the event generation. The generated events are passed through the full CMS detector

Table 4.1: Overview of the cross sections σ (and uncertainty) for the relevant Standard-Model background processes. The uncertainties on them are discussed in Section 4.6.2. The equivalent luminosity $\mathcal{L}_{eq}$ is defined as the number of simulated events normalized by the cross section.

	Generator	σ (pb) at 7 TeV	$\mathcal{L}_{eq}$ (fb^{-1})
$t\bar{t}$+jets	MADGRAPH	154±18 (12.0%)	23.6
t (t channel)	POWHEG	41.92 (NNLL)	92.8
$\bar{t}$ (t channel)	POWHEG	22.65 (NNLL)	88.6
t (tW channel)	POWHEG	7.87 (NNLL)	103.2
$\bar{t}$ (tW channel)	POWHEG	7.87 (NNLL)	102.6
t (s channel)	POWHEG	3.19 (NNLL)	81.3
$\bar{t}$ (s channel)	POWHEG	1.44 (NNLL)	95.5
W+1jet $\to \ell\nu$	MADGRAPH	5014.6	13.4
W+2jets $\to \ell\nu$	MADGRAPH	1610.3	15.6
W+3jets $\to \ell\nu$	MADGRAPH	342.6	19.4
W+4jets $\to \ell\nu$	MADGRAPH	193.6	66.7
Z/γ^*+jets $\to \ell\ell$ ($m_{\ell\ell} > 50$ GeV)	MADGRAPH	3078±139 (5%)	11.7
WW	PYTHIA	43.0±15.1 (35%) (NLO)	76.4
WZ	PYTHIA	18.2±7.6 (42%) (NLO)	250.8
ZZ	PYTHIA	5.9±1.6 (27%) (NLO)	1102.3
$t\bar{t}$+W	MADGRAPH	0.170±0.032 (19%) (NLO)	329.9
$t\bar{t}$+Z	MADGRAPH	0.143±0.040 (28%) (NLO)	334.8
$W^{\pm}W^{\pm}$+2jets	MADGRAPH	0.159±0.080 (49%) (NLO)	323.0

simulation based on GEANT4, and then processed by the same reconstruction software as used for the collision data.

The background processes taken into account, together with their cross sections σ are listed in Table 4.1. The equivalent luminosity $\mathcal{L}_{eq}$ is defined as the number of simulated events normalized by the cross section, and is therefore a measure of the size of the simulated sample. The cross sections of the top-quark pair and W and Z boson productions (in association with jets) correspond to the values measured by the CMS experiment [147, 148]. We use the theoretically predicted cross-section values for the single-top, $t\bar{t} + W$, $t\bar{t} + Z$, and same-sign WW processes [149–152]. The cross-section values for the diboson production are obtained with the MCFM generator [153, 154]. The simulated sample for the W-boson production is splitted in bins of the number of jets produced in association with the W boson. The cross sections for these *exclusive* processes are obtained by multiplying the *inclusive* cross section measured by the CMS experiment with the ratio of the LO cross sections of the exclusive and inclusive W-production processes. These ratios are determined from calculations with the MADGRAPH generator.

4.2.2 Fourth-generation signal simulation

The MadGraph 5.1.1 generator is used to produce the signal-process samples, again interfaced with pythia for the decay of the particles and for the hadronization. The subsequent matching, detector simulation and reconstruction are completely similar to the simulated SM samples. For the pair-production of the fourth-generation quarks we use the approximate next-to-next-to-leading-order cross-section values from the Hathor package [155], shown in Table 4.2 for various fourth-generation quark masses. We simulate samples for a whole range of heavy-quark masses since the mass greatly influences the kinematics of the final state objects (leptons, jets, missing transverse energy). For the electroweak production processes mentioned in Section 4.1.2, we rescale the NLO cross sections at $\sqrt{s} = 14$ TeV [156] to 7 TeV using a scale factor defined as the ratio of the LO cross section at 7 TeV and the LO cross section at 14 TeV as obtained by the MadGraph event generator. The resulting *maximal* production cross sections (assuming $|V_{tb'}| = |V_{t'b}| = |V_{t'b'}| = 1$) are listed in Table 4.3, and need be scaled according to the value of A. It is clear that the electroweak production cross section of the b' process is much lower than that of the t' process. The reason is that the b' quark needs to be produced in association with a top quark, which is much heavier than the b quark associated with the t' quark in the electroweak t' production. A similar reasoning leads to the understanding of the very small cross section of $t'b'$. As this process involves the production of two heavy particles, the cross section rapidly drops with increasing fourth-generation quark mass.

4.2.3 Trigger requirements in data and simulation

Events are triggered by requiring an isolated muon or electron, where the latter is accompanied by at least one jet identified as a b jet. To cope with the increasing instantaneous luminosity throughout the data taking, the transverse-momentum, p_T, requirement of the trigger-level muons was raised at several moments in time. The muon-p_T trigger threshold ranged from 17 GeV at the start-up to 30 GeV, with an additional pseudo-rapidity requirement of $|\eta| < 2.1$, during the second half of data taking. The electron trigger requirement consists of an isolated electron with $p_T > 25$ GeV and at least one central jet with $p_T > 30$ GeV and $|\eta| < 2.4$, being identified as a b jet by the Track-Counting High-Efficiency (TCHE) algorithm (described in Section 3.3.5). The selection criteria of the triggers are also applied on the simulation for consistency. Both single-muon and single-electron trigger efficiencies are found to be similar in data and simulation, hence no scale factors were applied to correct for potential differences in trigger efficiency.

4.3 Baseline event selection

In order to select potential fourth-generation events while reducing as much as possible the contamination of SM background events mimicking these new-physics processes, a dedicated procedure needs to be set up. Basic selection criteria on the missing transverse energy, leptons and jets are outlined in Sections 4.3.1 to 4.3.3. These criteria

Table 4.2: Overview of the $Q'\overline{Q'}$ production cross sections σ (approximate NNLO) for different assumed t' and b' quark masses [155], ranging from 400 GeV to 775 GeV in steps of 25 GeV. As these are QCD calculations, they do not depend on the $V_{\mathrm{CKM}}^{4\times4}$ parameter A. The equivalent luminosity $\mathcal{L}_{eq}$ is defined as the number of simulated events normalized by the cross section.

$m_{Q'}$ (GeV)	σ (pb) at 7 TeV	$\mathcal{L}_{eq}$ (fb^{-1}) $t'\overline{t'}$	$\mathcal{L}_{eq}$ (fb^{-1}) $b'\overline{b'}$
400	1.406	77	77
425	0.957	172	123
450	0.662	97	98
475	0.464	358	252
500	0.330	323	314
525	0.237	710	466
550	0.171	154	600
575	0.125	1324	924
600	0.0923	1124	1137
625	0.0685	2379	1581
650	0.0511	642	2122
675	0.0384	4187	4345
700	0.0290	5582	5808
725	0.0220	7443	5154
750	0.0168	9518	10250
775	0.0129	12879	8990

Table 4.3: Overview of the cross sections σ at 7 TeV, for the t' (b') electroweak single production in the t-channel for the different assumed t' and b' masses. The listed cross sections are maximal, hence for the case where $V_{t'b} = 1$, $V_{tb'} = 1$ or $V_{t'b'} = 1$. Charges are omitted in the notation. In the $t'b'$ process, the masses of the heavy quarks are assumed to be equal. The equivalent luminosity $\mathcal{L}_{eq}$ is defined as the number of simulated events normalized by the cross section.

	$t'b$		tb'		$t'b'$	
$m_{Q'}$ (GeV)	σ (pb)	$\mathcal{L}_{eq}$ (fb^{-1})	σ (pb)	$\mathcal{L}_{eq}$ (fb^{-1})	σ (pb)	$\mathcal{L}_{eq}$ (fb^{-1})
400	18.3	33	1.4	414	0.38	1572
425	15.8	38	1.3	470	0.30	2022
450	13.7	44	1.1	534	0.23	2601
475	11.9	50	1.0	594	0.18	3325
500	10.3	58	0.99	687	0.14	4253
525	8.9	63	0.87	766	0.11	5474
550	7.7	77	0.77	880	0.085	7013
575	6.7	89	0.60	1000	0.066	9050
600	5.8	103	0.53	1140	0.052	11623
625	5.0	119	0.46	1285	0.040	14859
650	4.3	138	0.41	1454	0.031	18994
675	3.8	158	0.36	1652	0.024	24607
700	3.3	178	0.32	1888	0.019	31213
725	2.8	212	0.28	2138	0.015	40340
750	2.4	245	0.25	2419	0.012	52057
775	2.1	283	0.22	2297	0.009	66289

will not yet be optimized for fourth-generation signal events, but can be considered as the common starting point for more optimized selections, splitting the event samples into the search categories defined in Section 4.4 later on. We will consider both single-lepton channels and multilepton channels, to increase the sensitivity to fourth-generation quarks.

The physics objects in the final state are reconstructed using the Particle-Flow (PF) algorithm developed by the CMS collaboration, as introduced in Section 3.3.1. The events are required to have a good-quality main primary vertex (see Section 2.2.2) within a cylinder of radius 2 cm and length 24 cm centered around the nominal interaction point. A PF charged-hadron subtraction algorithm is applied to the different object collections, to ensure the removal of charged hadrons that cannot be matched with the main primary vertex. In addition, the simulated events are reweighted to match the observed distribution of the number of simultaneous proton interactions, so-called pile-up. For the full dataset collected in 2011, we observe on average about nine interactions in each event.

4.3.1 Lepton selection criteria

The basic selection criteria on the muon and electron candidates, reconstructed as explained in Section 3.3, are summarized in Table 4.4. For the muons, we define a *tight* collection as containing 'global' or 'tracker' muons with $p_T > 20$ GeV and $|\eta| < 2.1$. In addition, the PF-based relative isolation I_{rel} from Equation (3.2), calculated from the deposited energies of other PF particles within a cone of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.4$ around the muon, is required to be smaller than 0.125. To veto additional muons in the single-lepton channels, and for the data-driven background estimation in the same-sign dilepton channel, we consider a collection of *loose* muons. The loose muons are defined as global muons with $p_T > 10$ GeV, $|\eta| < 2.5$ and $I_{rel} < 0.2$, also calculated within a cone of $\Delta R < 0.4$ around the muon candidate.

For the same reasons, we define also two collections of electron candidates. The tight collection contains the electrons with $p_T > 20$ GeV and $|\eta| < 2.5$, that are identified using the identification criteria outlined in Table 3.2. We exclude electron candidates that correspond to a supercluster in the transition region between the barrel and endcap ($1.4442 < |\eta| < 1.5660$) of the electromagnetic calorimeter, since the reconstruction of an electron object in this region is not optimal. The relative isolation, as calcualted from Equation (3.4), is required to be smaller than 0.1 in a cone of $\Delta R < 0.4$ around the electron candidate. This isolation requirement is tighter than for muons because the backgrounds for electrons are higher than for muons. Loose electrons should have $p_T > 15$ GeV, the same $|\eta|$ requirement as tight electrons, and a more relaxed isolation requirement, namely $I_{rel} < 0.2$ within a cone of $\Delta R < 0.4$ around the electron candidate. Moreover, the identification criteria from Table 3.2 are not applied on these loose electrons.

In order to validate the modeling of the lepton identification and selection, we compare kinematic distributions between observed data and SM simulation. Figure 4.1 shows the transverse momentum and the pseudorapitidy of the two tight muons in selected opposite-sign dimuon events. In these types of events we require the highest-p_T ('leading') muon to have a transverse momentum exceeding 40 GeV, while the second-

Table 4.4: An overview of the collections of leptons used in the analysis with the basic selection requirements defining them. Note that the tight leptons are a subset of the loose leptons.

Muons		Electrons	
Tight	Loose	Tight	Loose
$p_T > 20$ GeV	$p_T > 10$ GeV	$p_T > 20$ GeV	$p_T > 15$ GeV
$\|\eta\| < 2.1$	$\|\eta\| < 2.5$	$\|\eta\| < 2.5$	$\|\eta\| < 2.5$
$I_{rel} < 0.125$	$I_{rel} < 0.2$	$I_{rel} < 0.1$	$I_{rel} < 0.2$

highest-p_T ('subleading') muon should have $p_T > 20$ GeV. Events with additional loose muons and electrons are vetoed. Similar distributions for electrons in the opposite-sign dielectron events are shown in Figure 4.2. In these control plots no requirement on the missing transverse energy is made, but at least one jet with $p_T > 30$ GeV is required that is identified as a b jet by the Track-Counting High-Purity (TCHP) algorithm (see Section 3.3.5). We observe a reasonable agreement between the data and the SM background simulation, mainly consisting of Z+jets and $t\bar{t}$ processes. This is expected from the decay of the Z boson into two opposite-sign leptons, and leptonically decaying W bosons from top-quark decays.

In the analysis, we select events with at least one high-quality isolated (i.e. tight) muon or electron with $p_T > 40$ GeV. To suppress Z+jets background events, we exclude events where the invariant mass of the four-vector sum of this muon (electron) with a loose muon (electron) is inside a 10 GeV window around the Z-boson mass.

4.3.2 Jet selection criteria

In Section 3.3.4, we described the jet-clustering algorithm as well as the identification criteria and jet-energy corrections that are applied to these jets, to make sure their properties are well-behaved and the energy response in the detector is realistically described. The jets in this analysis are required to have a corrected transverse momentum exceeding 30 GeV and $|\eta| < 2.4$.

After the jet-energy corrections are applied, we smear the jet energies in the simulation to match the resolution measured with data [135]. Next, at least one of the jets within the tracker acceptance ($|\eta| < 2.4$) needs to be identified as a b jet, using the TCHP b-tagging algorithm. Here we require the signed impact-parameter significance of the third track in the jet (sorted by decreasing significance) to be larger than a value chosen such that the probability for a light-quark jet to be misidentified as a b jet is about 1%. For top-quark pair events this results in a b-tag efficiency of about 60%. The choice of this algorithm is recommended, as it can be considered as a tighter selection than the b-tag requirement used in the electron-trigger preselection. We apply p_T and jet-flavor dependent scale factors measured from data to the simulated events to take into account the different b-jet efficiency and the different probability that a light-flavor quark or gluon is identified as a b jet in data and simulation [157]. In case the jet is originating from a c quark, the same scale factor is applied as for the b jets, but the

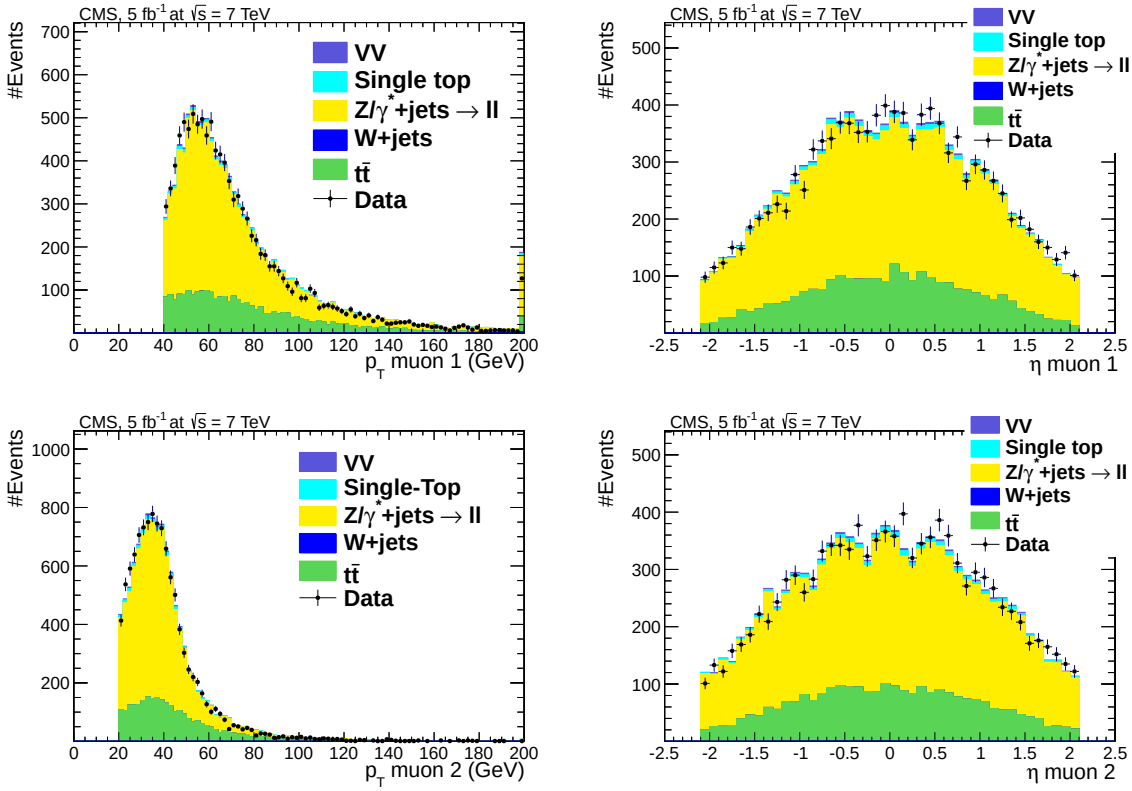

Figure 4.1: Control plots for opposite-sign dimuon events; the p_T (left) and η (right) of the highest-p_T muon (top), and the p_T and η of the second-highest-p_T muon (bottom). The total simulated background is normalized to the observed number of data events. The shapes of the data and simulation distributions agree reasonably well.

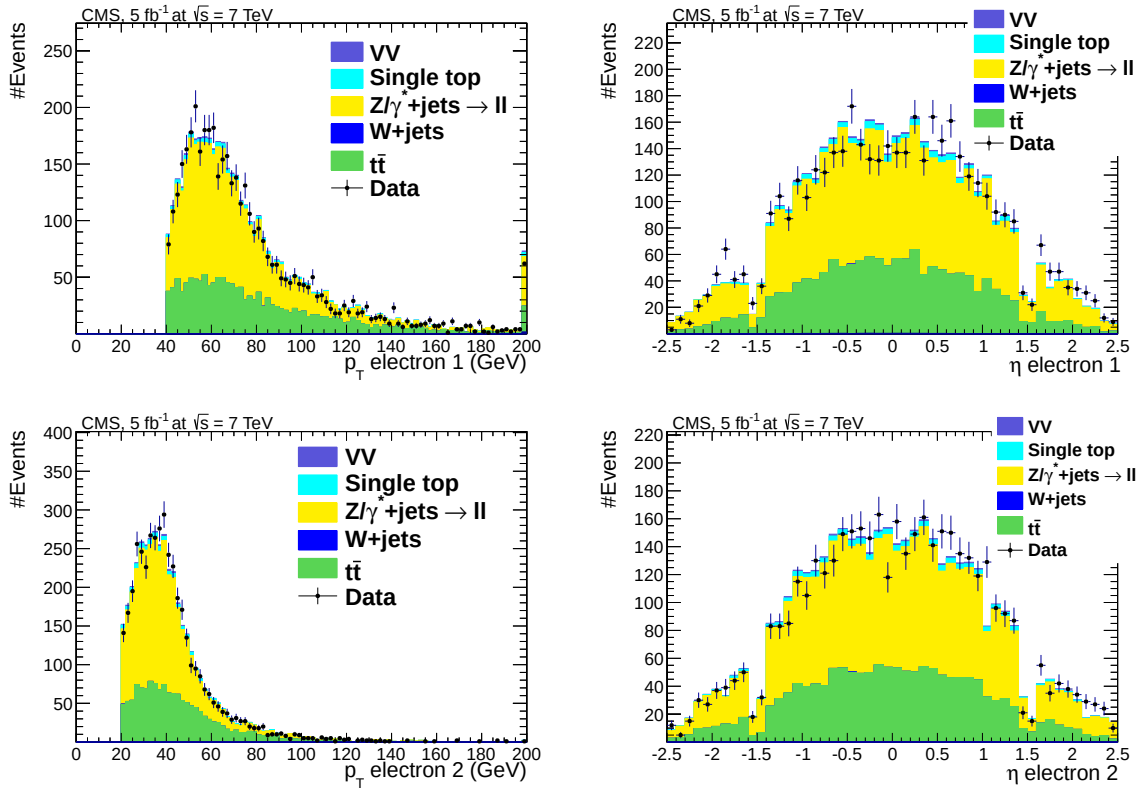

Figure 4.2: Control plots for opposite-sign dielectron events; the p_T (left) and η (right) of the highest-p_T electron (top), and the p_T and η of the second-highest-p_T electron (bottom). The total simulated background is normalized to the observed number of data events. The shapes of the observed and simulated distributions agree reasonably well.

uncertainty is taken twice as large.

A control region to validate the jet spectra can be defined by requiring the missing transverse energy to be between 30 GeV and 40 GeV, and asking for one tight electron with $p_T > 40$ GeV (while vetoing for additional loose electrons and muons). Here we also require at least four selected jets, at least one of which is identified as a b jet. Figure 4.3 shows the agreement between data and simulation for the kinematic distributions of the first until the fourth jet, ordered according to transverse momentum. Due to the applied selection criteria in this control region, the SM background is dominated by the semi-leptonic $t\bar{t}$ process, where one of the top quarks decays to a leptonically decaying W boson, and the other top quark to a hadronically decaying W boson.

4.3.3 Missing transverse energy selection criteria

The missing transverse energy $\not\!\!E_T$ is calculated by summing the transverse momenta of all the PF reconstructed objects and requiring a balance of the momentum in the transverse plane, as explained in Section 3.3.6. We require the missing transverse energy in this analysis, that has been corrected for L2L3 jet energy corrections according to Equations (3.9) and (3.12), to exceed 40 GeV. This requirement increases the purity of fourth-generation events by suppressing QCD multijet background processes. This type of background processes has a huge production cross section, but have usually *softer* spectra for the kinematics of the final-state objects. For instance, in case a neutrino from a semi-leptonic B-hadron decay is responsible for the missing transverse energy in a QCD multijet event, this neutrino will have a relatively low transverse momentum. In other cases where there is no neutrino (and the triggered lepton is for example a charged-hadron track misidentified as a lepton, the missing transverse energy can come purely from the non-perfect measurement of energies of objects, distorting the momentum balance. Due to these reasons, multijet events tend to have a low missing transverse energy. Moreover, the final state of all considered fourth-generation processes will contain at least one relatively high-p_T neutrino, hence such a tight requirement on the missing transverse energy will not reject too much signal events.

4.4 Event classification

Starting from the baseline physics-object selection outlined in the previous section, we define different channels according to the number of W bosons in the final state. Given that the t'-quark decay mode is the same as the top-quark decay mode, the $t'b$ and $t'\overline{t'}$ processes will yield signatures that are very similar to respectively the single-top and $t\bar{t}$ processes in the Standard Model. We select these processes through the single-lepton decay channels. In the signal final states that contain a b' quark, we expect three or four W bosons. If two or more of the W bosons decay to leptons, we may have events with two leptons of the same charge or with three charged leptons. Although the branching fraction of these decays is small compared to that of other decay channels, these final states are very interesting because of the low background that is expected from SM

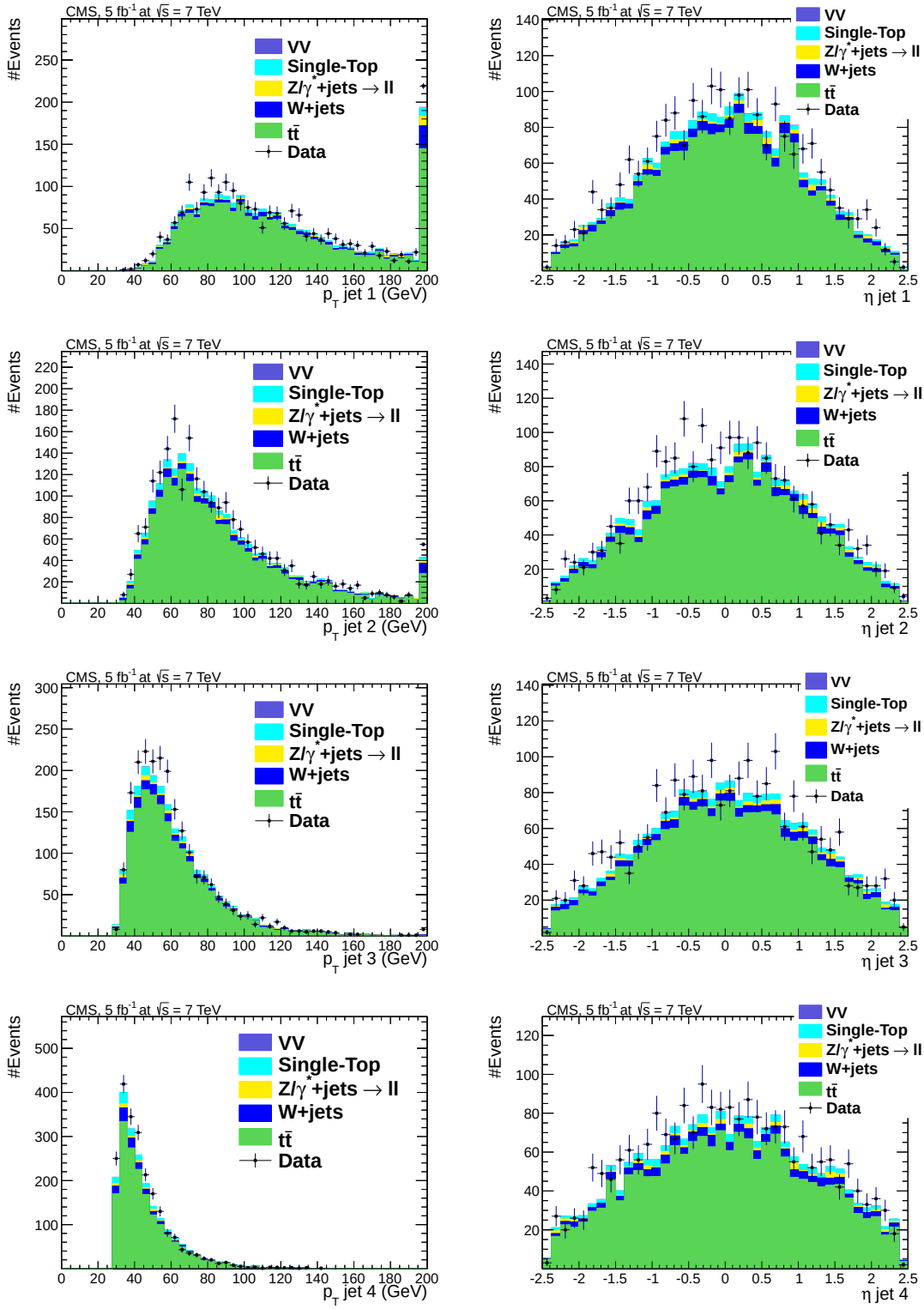

Figure 4.3: Control plots for the jets in single-electron events: the p_T (left) and η (right) of the first (top) to fourth (bottom) jet in the event, sorted by transverse momentum. The last bin includes the overflow bin. The total simulated background, normalized to the integrated luminosity of the collected data, agrees reasonably well with the observed data.

processes.

In Section 4.4.1 we will discuss the procedure to count the number of W bosons in the event, and list the constructed subcategories of single-lepton events to optimize the search for a fourth-generation signal. The channels with two same-sign leptons and three leptons in the final state are described in Section 4.4.2. For same-sign dilepton events, we estimate the SM background from lepton-charge misidentification and from jets wrongly identified as leptons in a data-driven way.

4.4.1 The single-muon and single-electron decay channels

In the single-muon (single-electron) channel, we require exactly one isolated tight muon (electron), and veto for additional loose leptons as defined in Table 4.4. Since we will divide the selected single-lepton events in different subsamples according to the number of W-boson candidates, first a dedicated W-boson counting procedure is set up.

W-boson candidate counting procedure

Each event is assumed to have at least one W boson decaying to leptons, consistent with the requirements of an isolated lepton and a large missing transverse momentum from the neutrino that escapes detection. The decay of W bosons to $q\bar{q}$ final states are reconstructed with the following procedure. For each event, we have a collection of selected jets used as input for the reconstruction of the W-boson candidates. As the one or two b-tagged jets with the highest value of the b-tag TCHP discriminator are assumed to originate from a b quark, these jets are not considered. W-boson candidates are constructed from all possible pairs of the remaining jets in the collection. We use both the expected mass, $m_W^{\text{fit}} = 84.3$ GeV, and the width, $\sigma_{m_W}^{\text{fit}} = 9.6$ GeV from a Gaussian fit to the reconstructed-mass distribution of jet pairs matched to partons from the decay of a W boson in simulated $t\bar{t}$ events, as illustrated in Figure 4.4. The W-boson candidate with a mass that matches the value of m_W^{fit} best, is chosen as a W boson if its mass is within a $\pm 1\sigma_{m_W}^{\text{fit}}$ window around m_W^{fit}. The jet pair that provided the hadronically decaying W boson is removed from the collection and the procedure is repeated until no more candidates are found for W bosons decaying to jets. Figure 4.5 shows the reconstructed W-boson candidate mass from all jet pairs in single-lepton events. The agreement between the observed data and simulated distributions, peaking at the W-boson mass, is reasonable.

The key element in the identification procedure for hadronically decaying W bosons is to have two jets matching the expected W-boson reconstructed mass value. Hence with this procedure we will identify correct W bosons but also fake W bosons. A fake W boson is defined as a dijet system that does not originate from a true W boson, but nevertheless matches with the W boson mass. This can occur for both Standard-Model and fourth-generation signal events. This would cause a 'migration' of events from one subsample (later on defined according to the number of b jets and W bosons) to another. When this migration is too large, the sensitivity of the method to differentiate between heavy quarks produced via the electroweak or via the strong interaction (and thus to the $V_{\text{CKM}}^{4\times4}$ parameter A), would be reduced. For simulated semileptonically decaying $t\bar{t}$ events, the hadronically decaying W boson is not reconstructed in 65% of the events.

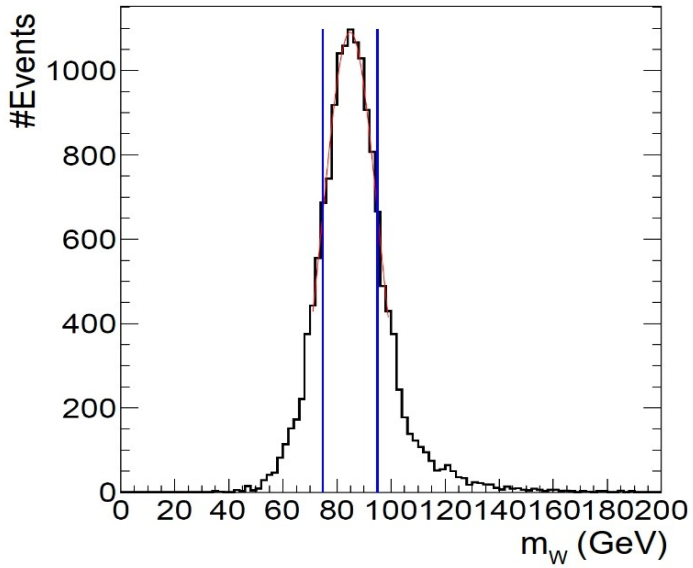

Figure 4.4: From the distribution of the mass of the hadronically decaying W boson, reconstructed from the jets that are matched to the partons in muon + jets $t\bar{t}$ events, the mean $m_W^{\text{fit}} = 84.3$ GeV and standard deviation $\sigma_{m_W}^{\text{fit}} = 9.6$ GeV are extracted by a Gaussian fit. The two vertical lines illustrate the window that is taken to decide whether or not a hadronically decaying W boson is found.

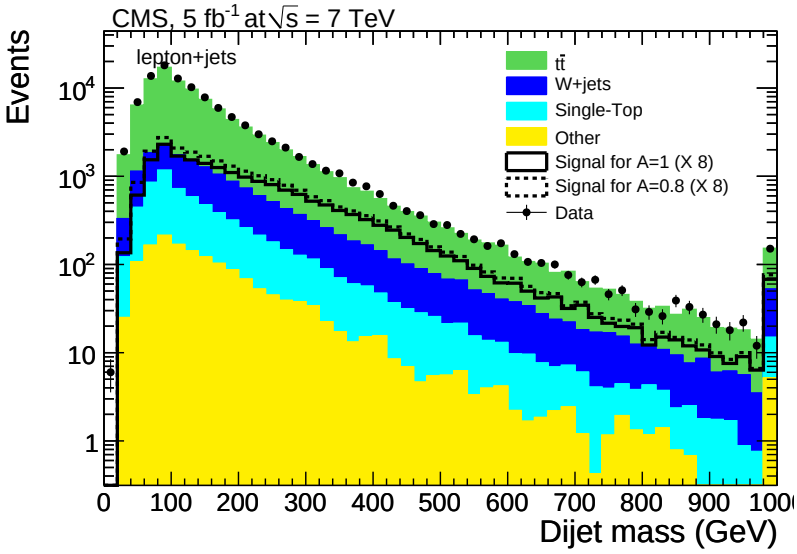

Figure 4.5: The observed data and the simulation agree well for the distribution of the invariant mass for all jet pairs in single-lepton events.

For the remaining events, a hadronically decaying W boson is reconstructed, but in about half of those events the W boson is fake. However, the most important aspect of the W-counting procedure is that the simulation agrees with the observed data. Figure 4.6 shows that this is indeed the case, since the shapes of the reconstructed W-boson mass from the chosen jet pair in the event for data and simulation agree. Therefore, the only impact of the fake rate intrinsically present in the method, is that some sensitivity might be lost when events are not categorized in the correct subsample.

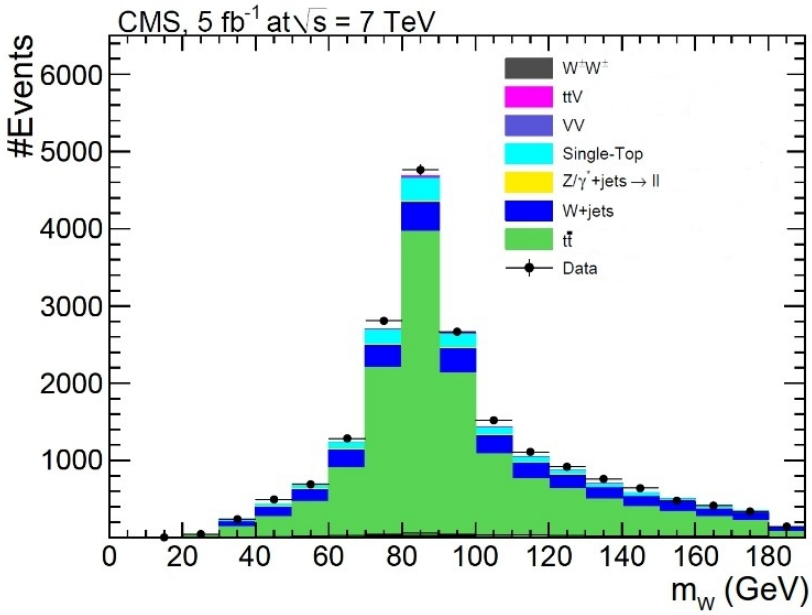

Figure 4.6: Reconstructed W-boson mass of the jet pair with the closest invariant mass to $\sigma^{\text{fit}}_{m_W}$ in the event.

Single-lepton subsamples

In the single-lepton channels, we define different exclusive subsamples according to the number of b jets (exactly one or at least two) and the number of W-boson candidates (one, two, three, and at least four). For the muon and electron channels, there are seven subsamples each, because we do not consider the subsample with only one b jet and one W boson, as it proved to add little sensitivity in the final result. The different single-lepton subsamples are named as NbMW, where N represents the number of required b jets, and M the number of assumed W bosons. Since we require one lepton in the final state, the number of W bosons that are considered to decay to two jets is M $-$ 1.

- **2b1W category.** The subsample with two b jets and one W boson ($W \to \ell\nu$ with $\ell = e, \mu$) is dominated by t' events produced through the electroweak interaction. In this subsample, we apply a veto for additional jets with a transverse momentum exceeding 30 GeV. Furthermore, we suppress the background from $b\bar{b}$ by imposing the requirement $\Delta\phi(j_1, j_2) < \frac{\pi}{2} + \pi(|p_T^{j1} - p_T^{j2}|)/(p_T^{j1} + p_T^{j2})$ on the

two b jets j_1 and j_2, since this background tends to have jets that are produced back-to-back with balanced p_T.

- **1b2W and 2b2W categories.** At least three selected jets are required in addition to the b-tagged jet from the baseline selection, resulting in a topology with at least four jets. In the case of the 2b2W category, one of the additional jets is also required to be identified as a b jet. The hadronically decaying W boson is found from the remaining jets, as described in the W-boson counting procedure. These categories target semi-leptonic $t'\bar{t}'$ production. Here the background is dominated by $t\bar{t}$ production, due to the similarity in decay topology.

- **1b3W and 2b3W categories.** The events in these categories need to have at least five additional selected jets on top of the b-tagged jet from the baseline selection. In the case of the 2b3W category, one of the additional jets is required to be tagged as a b jet. Among the non-b-tagged jets, two independent jet pairs form two reconstructed hadronically decaying W bosons. Because of the high number of required W bosons, these subsamples are designed to be sensitive to electroweak and pair production of b' quarks.

- **1b4W and 2b4W categories.** These subsamples are analogous to the 1b3W and 2b3W categories, but now at least seven additional selected jets on top of the b-tagged jet from the baseline selection are required. Independent jet pairs form at least three reconstructed hadronically decaying W bosons.

Table 4.5 summarizes the requirements that define the different single-lepton decay subsamples, after the criteria on the missing transverse energy, and the lepton and jet p_T and η are applied. The observed and predicted event yields in the different

Table 4.5: Overview of the event-selection requirements defining the seven different subsamples in the single-lepton decay channel, according to the number of b jets and the number of W-boson candidates. The indicated number of jets include the total number of required jets ($p_T > 30$ GeV, $|\eta| < 2.4$), regardless of their (non-)identification as b jets. The number of hadronically decaying W-boson candidates ($W \to q\bar{q}$) are found from the W-boson counting procedure described in Section 4.4.1.

Single-lepton decay channel			
1 W	2 W	3 W	4 W
$= 2$ jets	≥ 4 jets	≥ 6 jets	≥ 8 jets
$= 2$ b jets	either $= 1$ or ≥ 2 b jets		
$\Delta\phi(j_1, j_2)$ requirement	1 $W \to q\bar{q}$	2 $W \to q\bar{q}$	3 $W \to q\bar{q}$

subsamples of the single-lepton channel are shown in Table 4.6. After the selection criteria, the dominant background contributions result from the production of top quark pairs, W+jets, and single top. Other processes with very small contributions to the total background are Z+jets and diboson production, and also top-quark pairs produced in association with a W or Z boson. The combined event yield of these

processes is about 1% of the total SM contribution. The QCD multijet background is found to be negligible in each of the subsamples. The reason is the requirements of an isolated muon or electron with $p_T > 40$ GeV, a missing transverse energy larger than 40 GeV and at least one jet identified as a b jet. Data and simulation are found to agree within the combined statistic and systematic uncertainties.

Table 4.6: Event yields in the single-lepton (e or μ) channel. Uncertainties reflect the combined statistical and systematic uncertainties. The prediction for the signal is shown for two different values of A and for a fourth-generation-quark mass $m_{Q'} = 550$ GeV.

	1b2W	1b3W	1b4W	2b1W	2b2W	2b3W	2b4W
$t\bar{t}$+jets	5630 ± 410	230^{+29}_{-26}	$3.0^{+1.9}_{-1.3}$	819^{+59}_{-62}	2810 ± 240	85^{+12}_{-10}	$0.6^{+0.8}_{-0.5}$
W+jets	490 ± 180	$8.0^{+3.1}_{-3.0}$	$0.3^{+0.9}_{-0.3}$	150^{+47}_{-46}	37 ± 12	$1.1^{+1.0}_{-0.4}$	$0.0^{+0.8}_{-0.0}$
Z+jets	36^{+5}_{-6}	$1.0^{+0.2}_{-0.1}$	0	$7.1^{+1.0}_{-0.6}$	$2.8^{+1.0}_{-0.3}$	0	0
Single top	346 ± 64	$6.5^{+1.6}_{-1.5}$	$0.2^{+0.3}_{-0.2}$	200 ± 34	110 ± 19	$2.5^{+0.7}_{-0.5}$	$0.0^{+0.1}_{-0.0}$
VV	15 ± 2	$0.4^{+0.3}_{-0.1}$	$0.0^{+0.1}_{-0.0}$	15 ± 2	1.8 ± 0.3	$0.0^{+0.1}_{-0.0}$	$0.0^{+0.1}_{-0.0}$
$t\bar{t}V$	28 ± 3	3.4 ± 0.5	0.1 ± 0.0	0.7 ± 0.2	15 ± 5	$1.5^{+0.3}_{-0.2}$	0
Total background	6550 ± 450	249^{+29}_{-26}	$3.6^{+2.1}_{-1.3}$	1190^{+83}_{-85}	2970 ± 240	91^{+12}_{-10}	$0.6^{+1.2}_{-0.5}$
Observed	7003	242	8	1357	3043	91	4
Signal ($A = 1$)	55 ± 1	12 ± 1	0.9 ± 0.2	$1.0^{+0.2}_{-0.3}$	49 ± 2	8.1 ± 0.4	0.5 ± 0.2
Signal ($A = 0.8$)	85 ± 2	14 ± 1	1.0 ± 0.2	69 ± 3	66 ± 2	9.2 ± 0.4	0.5 ± 0.2

4.4.2 The same-sign dilepton and trilepton decay channels

Since the probability for fourth-generation $b'\overline{b'}$ or $b't$ and $b't'$ production to result in final states with same-sign dileptons or trileptons is relatively high, it is worth considering these topologies. The transverse momentum of at least one of the tight leptons in the multilepton channel is required to be larger than 40 GeV, while the threshold is reduced to 20 GeV for additional tight leptons. Events with two muons or electrons with a mass within 10 GeV of the Z-boson mass are rejected to reduce the SM background with Z bosons in the final state. We require at least four jets, of which one is required to be a b jet, for the same-sign dilepton events. These events can consist of two muons, two electrons, or a muon and an electron. In the case of the trilepton events, consisting of either three muons, two muons and an electron, a muon and two electrons, or three electrons, the minimum number of required jets is reduced to two, of which one is required to be a b jet. Table 4.7 summarizes the event selection requirements defining the same-sign dilepton and trilepton decay channels that are applied on top of the other requirements on the missing transverse energy, and lepton and jet p_T and η from the baseline selection.

There are several contributions to the total SM background for the same-sign dilepton events. One of these contributions comes from events for which the charge of

Table 4.7: Overview of the event selection requirements specific to the same-sign dilepton and trilepton decay channels. The indicated number of jets include the total number of required jets ($p_T > 30$ GeV, $|\eta| < 2.4$), regardless of their (non-)identification as b jets.

Same-sign dilepton	Trilepton
= 2 isolated leptons with same charge	= 3 isolated leptons
≥ 4 jets	≥ 2 jets
≥ 1 b jet	≥ 1 b jet

one of the leptons is misreconstructed, for instance in $t\bar{t}$ events with two W bosons decaying into leptons. Secondly, there are events with one prompt lepton and one non-prompt lepton passing the isolation and identification criteria. We obtain from the data the predicted number of background events for these two contributions to the total background in the same-sign dilepton subsample. Finally, there is an irreducible contribution from SM processes with two prompt leptons of the same sign; e.g. $W^{\pm}W^{\pm}$, WZ, ZZ, $t\bar{t}+W$ and $t\bar{t}+Z$. Except for $W^{\pm}W^{\pm}$, these processes are also the main contributions to the total background for the trilepton subsample. The event yields for the irreducible component of the background for the same-sign dilepton channel and the total background in the case of the trilepton subsample are taken from the simulation.

Charge misidentification estimation in the same-sign dilepton channel

For the same-sign dilepton events with at least one electron, the background is estimated from control samples. The charge of a lepton in the CMS detector can be determined from the direction of the curvature in the magnetic field, and this curvature is in general more precisely measured for muons than for electrons. For muons, the charge misidentification rate is negligible, and for electrons we determine this rate using a double-isolated-electron trigger applied on a Z+jets enriched data sample. The reason we want to select such a sample is that in these types of events we can exploit the prior knowledge that the charge of two leptons from a Z-boson decay should be opposite. To achieve this, two isolated electrons in data are required with an invariant mass within 10 GeV of the Z-boson mass. We select events with $\not\!\!E_T < 20$ GeV and a transverse mass formed by the lepton and the missing transverse energy

$$M_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\ell}\not\!\!E_T}[1 - \cos(\Delta\phi(\ell, \not\!\!E_T))] \tag{4.2}$$

less than 25 GeV to suppress background from top-quark and W+jets events. Figure 4.7 shows that after the cuts on $\not\!\!E_T$ and M_{T}, the dominating process is Z+jets, as desired. The total simulated distributions are normalized to the observed number of data events, as a prescaled trigger was used for which we did not calculate the collected integrated luminosity. The shape from data and simulation agrees reasonably well, except for low invariant mass, since only events with $m_{\ell\ell} > 50$ GeV were generated in

the simulated Z+jets sample. At these low invariant masses, also QCD multijet events contribute to the spectrum.

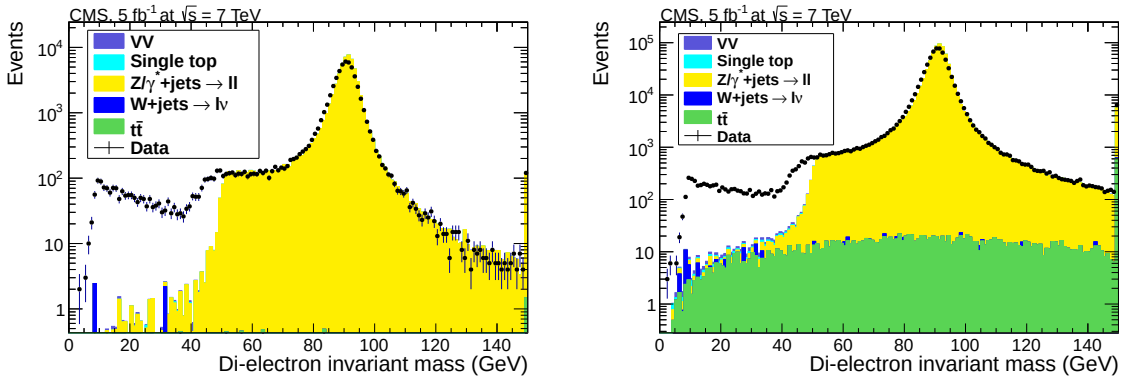

Figure 4.7: These distributions show the invariant mass of the dilepton pair in the events selected for the estimation of the charge mis-identification probability. The left plot represents the invariant mass after all the selection cuts are applied, while in the right plot the cuts on the missing transverse energy and the transverse mass are not applied. In both cases the Z-boson mass peak around 91 GeV is clearly visible.

In this Z+jets enriched control sample, the number of events N_{SS} with two electrons of the same charge can be written as a function of the number of events N_{OS} with two electrons of opposite charge as

$$\begin{aligned} N_{SS} &= N_{e^+e^+} + N_{e^-e^-} = (N_{e^+e^-} + N_{e^+e^-}) \times R \\ &= N_{OS} \times 2R, \end{aligned} \tag{4.3}$$

with R defined as the charge misidentification ratio R. Hence, R is calculated as $R = N_{SS}/2N_{OS}$. We obtain $R_B = (0.140 \pm 0.015)\%$ and $R_E = (1.39 \pm 0.018)\%$ for ECAL barrel and endcap electron candidates, respectively. Next, when applying the full same-sign dilepton signal event selection (left column of Table 4.7), but with the charge requirement reversed, we obtain a number of selected data events with two electrons and with an electron and a muon in the final state. The background with two electrons or with an electron and a muon with the same sign is obtained by taking the number of opposite-sign events and scaling it with R. This procedure is summarized in Table 4.8.

The p_T spectrum of the electrons in the control sample and the signal region is similar. Therefore, no correction is applied for the p_T dependency of the charge misidentification ratio. We also verified that changing the window around the Z-boson mass does not alter the charge misidentification ratio significantly. We verified that the same-sign fourth-generation signal-event contamination in the considered Z-boson window is negligible and does not affect the charge misidentification rate. Only 0.08 signal events are expected to end up in this dilepton invariant-mass region, to be compared to about 80000 Z+jets events for which 80 events have two leptons with the same charge.

Table 4.8: Estimated number of background events in the same-sign dilepton signal category obtained from the data-driven charge misidentification rates of barrel and endcap electrons.

	Di-electron			Electron+Muon	
	2 in barrel	barrel+endcap	2 in endcap	barrel	endcap
Observed data	69	18	0	220	29
Scale factor	$\times 2R_B$	$\times(R_B + R_E)$	$\times 2R_E$	$\times R_B$	$\times R_E$
Estimated background	0.19 ± 0.03	0.28 ± 0.07	0.0±0.0	0.31 ± 0.03	0.40 ± 0.05

Fake-lepton estimation in the same-sign dilepton channel

Another important background contribution to the same-sign dilepton channel originates from jets being misidentified as an electron or a muon (non-prompt or "fake" leptons). For the background estimation due to fake electrons (muons), we apply a single-isolated-electron (single-muon) trigger on a QCD multijet enriched data sample. The p_T thresholds in these prescaled triggers are low, namely 8 GeV. The reason we want to select a multijet enriched sample, is that we expect no prompt leptons in these types of events, only non-prompt leptons from hadron decays, or charged-hadron tracks being misidentified as leptons. Hence, these events provide a handle on the fake leptons we want to study. When these fake leptons happen to be isolated by coincidence, they can still pass the tight isolated-lepton criteria.

We require at least one loose electron or muon. Additionally, we require $\not\!\!E_T < 20$ GeV and $M_T < 25$ GeV to suppress background from top-quark and W+jets events. Moreover, we veto events with leptons of the same flavor that have a dilepton mass within 20 GeV of the Z-boson mass. We count the number of loose and tight leptons with a p_T below 35 GeV. Leptons produced in jets have typically a soft p_T spectrum, so the threshold on the p_T is required to suppress contamination from W+jets events that would bias the estimation. The probability that a loose (L) lepton passes the tight (T) selection criteria (defined in Table 4.4) is then given by the ratio $\epsilon_{\mathrm{TL}} = N_{\mathrm{T}}/N_{\mathrm{L}}$. Consequently, the probability that a loose lepton does *not* pass the tight selection criteria, is $(1 - \epsilon_{\mathrm{TL}})$. Next, we count the number of events in data that pass the event selection criteria with one lepton passing the tight selection criteria and a second lepton passing the loose, but *not* the tight, criteria. If we denote this number of events by $N_{\mathrm{T,L-T}}$, we can express the number of estimated events $N_{\mathrm{T,T}}$ with two tight leptons as

$$\begin{aligned} N_{\mathrm{T,T}} &= N_{\mathrm{T,L}} \times \epsilon_{\mathrm{TL}} \\ &= \frac{N_{\mathrm{T,L-T}}}{(1 - \epsilon_{\mathrm{TL}})} \times \epsilon_{\mathrm{TL}} \end{aligned} \tag{4.4}$$

Therefore, we need to apply the factor $R_f^\ell = \epsilon_{\mathrm{TL}}^\ell/(1 - \epsilon_{\mathrm{TL}}^\ell)$, depending on the lepton type ℓ, to the number $N_{\mathrm{T,L-T}}$ obtained from data to estimate the number of events

Table 4.9: Estimated number of background events in the same-sign dilepton channel obtained from the loose-tight method.

	T muon + L muon	T electron + L electron	T muon + L electron	T electron + L muon
Observed data	2	4	4	5
Scale factor	R_f^μ	R_f^e	R_f^e	R_f^μ
Estimated background	0.07 ± 0.05	0.36 ± 0.18	0.36 ± 0.18	0.17 ± 0.08

with a non-prompt lepton in the analysis.[2] We obtain $\epsilon^\mu_{\mathrm{TL}} = 3.31 \pm 0.19\%$ for muons and $\epsilon^e_{\mathrm{TL}} = 8.21 \pm 0.71\%$ for electrons. The estimated number of background events in the same-sign dilepton channel from fake leptons are summarized in Table 4.9. The statistical uncertainty in the estimated number of events is large because only a few events are selected with one tight and one loose, but not tight, lepton.

As a closure test, we repeated the procedure on the simulation. The probability that a loose lepton passes the tight cuts is obtained from QCD multijet events, and yields $\epsilon^\mu_{\mathrm{T,L}} = 3.45\%$ for muons and $\epsilon^e_{\mathrm{T,L}} = 8.73\%$ for electrons. We also determined these probabilities with all the simulated samples and obtain similar results. This demonstrates that the applied cuts reject efficiently the contributions from W+jets and top-quark pair events. These probabilities are then applied to the number of events in the simulation with one loose, but not tight, and one tight lepton. From this closure test, we estimate a total of 0.15 events. This number is in agreement with the prediction of 0.23 ± 0.23 dilepton events from top-quark pair events that decay semi-leptonically. It is found from the simulation that there are no contributions from other processes.

The total number of expected background events for the same-sign dilepton and trilepton channels are summarized in Table 4.10. We observe no significant excess in data from a fourth-generation signal. Note that in the trilepton subsample, we do not impose a lepton-charge requirement, hence an evaluation of the charge misidentification rate is not relevant. The contribution from fake leptons is expected to be negligible in this case. The reason is that the efficiency for a loose lepton to pass the tight selection criteria has been found similar in data and simulation, as mentioned above. At the same time, no simulated non-prompt trilepton events pass the trilepton event selection criteria, hence this indicates that the fake-lepton contribution is negligible in data as well.

[2]This formula and the corresponding numbers are corrected with respect to the formula given in the published paper [81], where the ϵ_{TL} and $(1 - \epsilon_{\mathrm{TL}})$ factors were multiplied instead of divided. The resulting yields differ only slightly within the uncertainties, and we checked this has no impact on the derived exclusion limits.

Table 4.10: The prediction for the total number of background events compared with the number of observed events in the same-sign dilepton and the trilepton subsamples. The numbers of expected signal events for a fourth-generation-quark mass $m_{Q'} = 550$ GeV are also shown for two possible scenarios.

	2 muons	2 electrons	electron+muon	trilepton
Irreducible background	0.77 ± 0.08	0.59 ± 0.08	1.10 ± 0.11	0.96 ± 0.12
Charge misid background	–	0.47 ± 0.08	0.71 ± 0.06	–
Fake-lepton background	0.07 ± 0.05	0.36 ± 0.18	0.53 ± 0.20	–
Total background	0.84 ± 0.09	1.42 ± 0.21	2.34 ± 0.24	0.96 ± 0.12
Observed	2	2	2	1
Signal ($A = 1$)	3.31 ± 0.15	2.03 ± 0.36	5.29 ± 0.19	3.37 ± 0.16
Signal ($A = 0.8$)	3.79 ± 0.15	2.29 ± 0.36	6.00 ± 0.19	3.65 ± 0.16

4.5 The search variables

We have defined different subsamples according to the reconstructed final state. In this section, we reconstruct observables that are sensitive to the presence of the fourth-generation quarks. These observables will be used in Section 4.6 as input to a fit of the combined distributions for the SM (background-only) hypothesis and the signal-plus-background hypothesis.

In the subsamples with two leptons of the same sign, the trilepton subsample, and the two single-lepton subsamples with four W-boson candidates, the expected number of events is small. Therefore, the event counts in each of these subsamples, given in Tables 4.6 and 4.10, are used as the 'observable'. In the single-lepton subsamples with one or three W bosons, we use S_T, the scalar sum of the transverse momenta of the reconstructed objects in the final state, as the discriminating observable. The S_T variable is discussed in more detail in Section 4.5.1. To increase the sensitivity in the subsamples with two W bosons, we reconstruct also m_{bW}, the mass of the hadronic bW system. The definition and the reconstruction procedure of this observable are described in Section 4.5.2.

An overview of the observables used in each subsample is presented in Table 4.11. The way to consistently combine the information of these variables in the search for fourth-generation quarks will be discussed in Section 4.6.

4.5.1 The S_T variable

We define the S_T variable in this analysis as the scalar sum of the transverse momenta of the reconstructed objects in the final state;

$$S_T = \not{E}_T + p_T^{\ell} + p_T^{b} + p_T^{j} + \sum_{i=1}^{N \leq 3} p_T^{W^i_{q\bar{q}}}, \tag{4.5}$$

where the sum runs over the number of reconstructed hadronically decaying W bosons. Here p_T^{ℓ} is the transverse momentum of the lepton, p_T^{b} the transverse momentum of the

Table 4.11: Overview of the observables to be used in the limit calculation.

Subsample	Observable
single-lepton $1W$	S_T
single-lepton $2W$	S_T and m_{bW}
single-lepton $3W$	S_T
single-lepton $4W$	event yield
same-sign dilepton	event yield
trilepton	event yield

b-jet, p_T^j the transverse momentum of the second b jet or, in case there is no additional jet identified as a b jet, the p_T of the jet with the highest transverse momentum in the event not used in the hadronically decaying W-boson reconstruction, and $p_T^{W^i_{q\bar{q}}}$ the transverse momentum of the i^{th} W boson decaying to jets.

The distributions of the S_T variable in each of the exclusive subsamples defined in Section 4.4.1 is shown in Figure 4.8 for the single-muon channel and in Figure 4.9 for the single-electron channel. The red dashed bands in these plots represent the total systematic uncertainty in the simulation, taking into account all the systematic effects described in Section 4.6.2. These include for example uncertainties in the used cross sections of the SM background processes, in the jet energy scale, and in the b-tag scale factors applied on the simulation. We do not show the S_T distributions for the categories with four reconstructed hadronically decaying W bosons, because in these cases the number of events are used as search observable. The overlayed signal distributions are shown only for the degenerate fourth-generation quark masses of 550 GeV. The electroweak t', b', and single-top production cross sections are taken to be maximal for illustrational purposes, and all signal cross sections are scaled with a factor of five for visibility. Due to the high masses of the fourth generation quarks, their decay products tend to be very energetic. This causes the signal distributions to shift to higher S_T values compared to the SM background distributions, providing the desired discriminating power. From Figures 4.8 and 4.9 we see that the designed subsamples are indeed sensitive to the appropriate fourth-generation processes. For example, the event yield of the electroweak t' process is the largest in the 2b1W category, while the subsamples with three W bosons are mainly sensitive to b' production processes where multiple W bosons are expected in the final state.

4.5.2 The m_{bW} variable

The single-lepton subsamples with two W bosons are designed to be sensitive to t'-quark pair production. Since the main difference with SM top-quark pair production is the mass of the heavy quark, it is sensible to reconstruct the mass of the t' quark and the top quark and to use this as a discriminating variable. In practice, we will consider m_{bW}, the invariant mass of the system formed by a b jet and a hadronically decaying W boson. In the SM $t\bar{t}$ process, this would indeed correspond to the reconstructed mass of a top quark decaying to a W boson that subsequently decays to two jets.

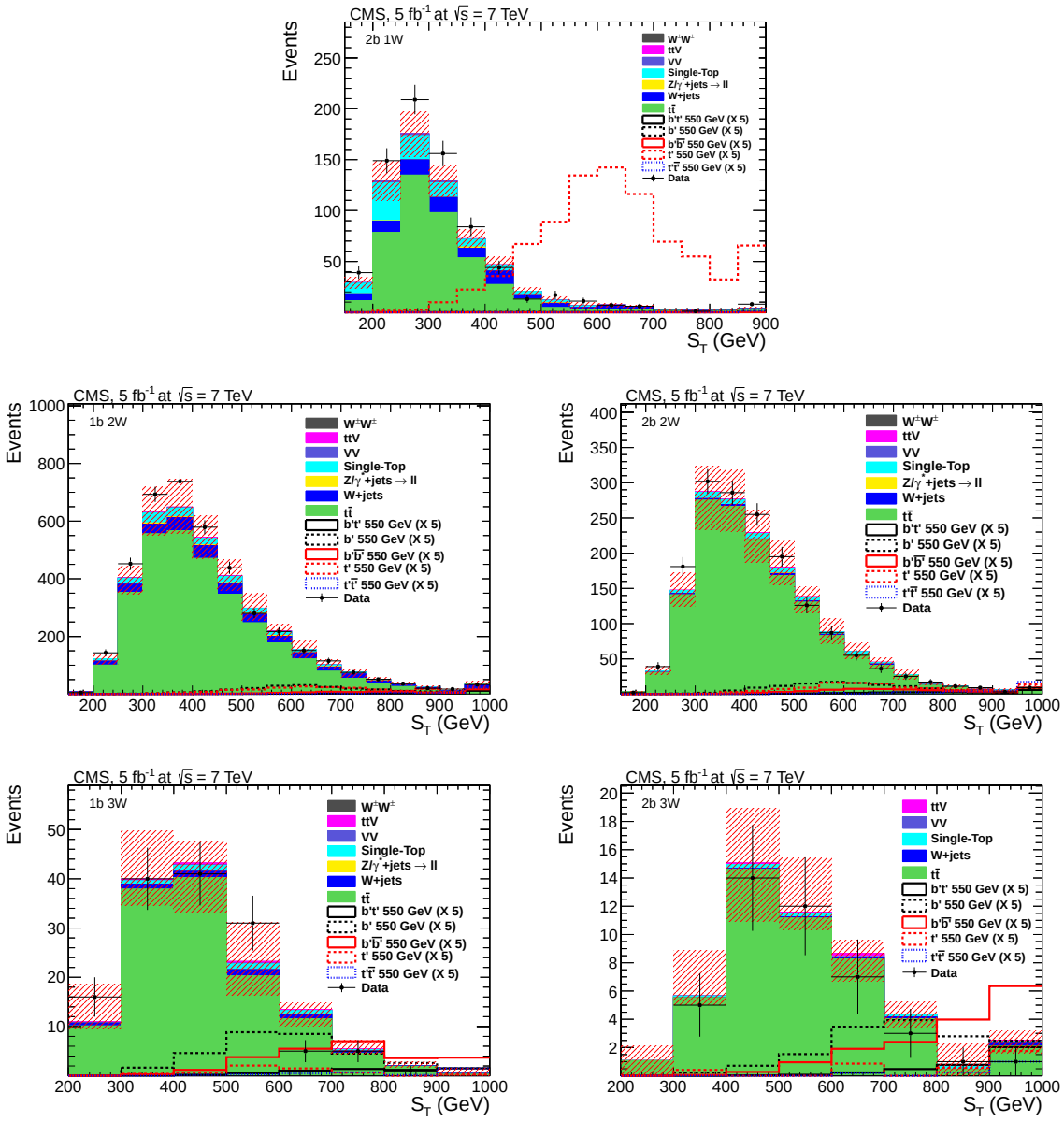

Figure 4.8: The S_T distributions for the different subsamples in the single-muon channel; the subsample with two b jets and one W boson (top), with two W bosons (one b jet middle left, two b jets middle right) and with three W bosons (one b jet bottom left, two b jets bottom right). Maximal cross sections are assumed for the signal processes, and all overlayed signal processes are scaled with a factor of five for visibility. The last bin includes the events in the overflow bin.

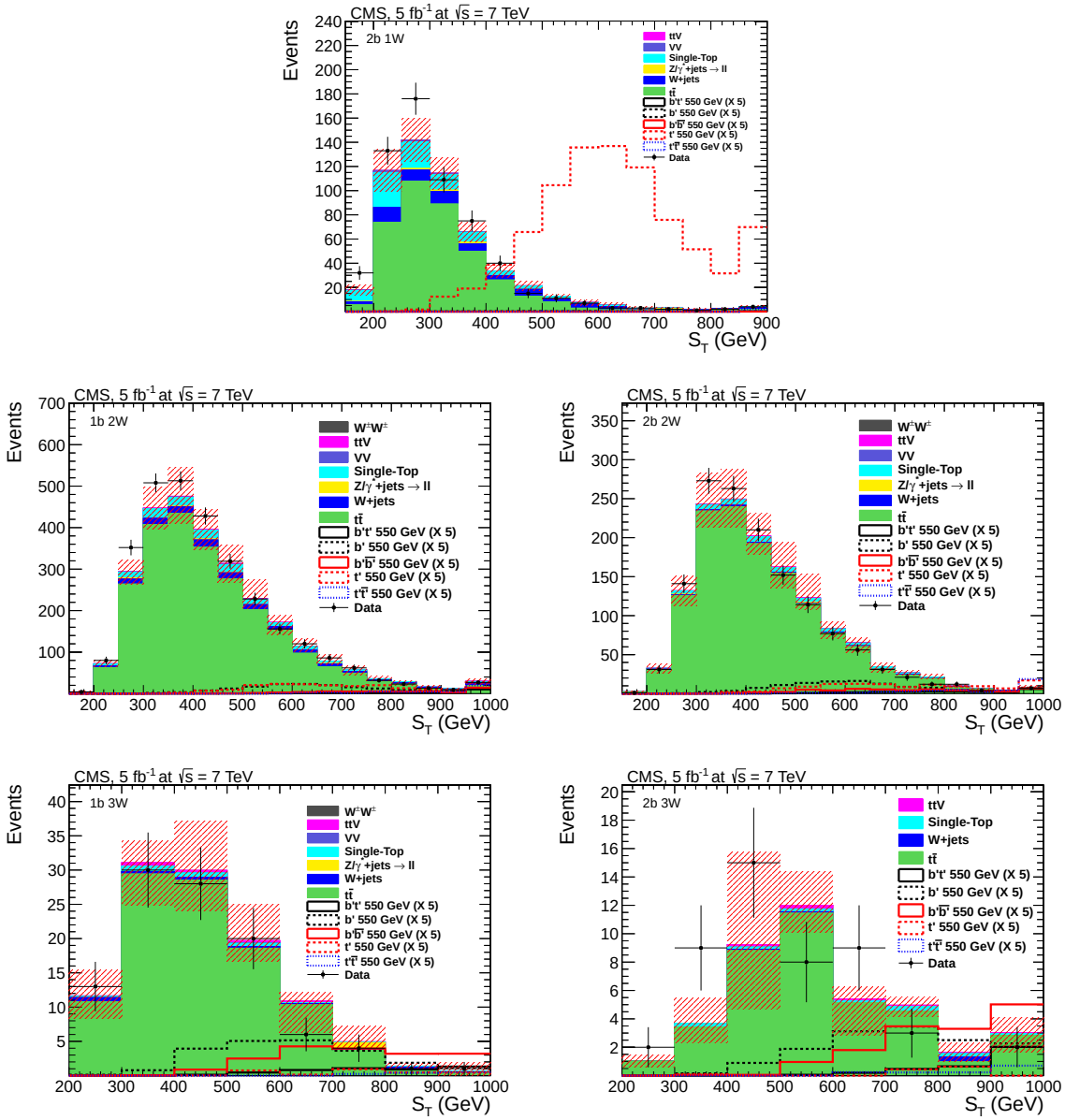

Figure 4.9: The S_T distributions for the different subsamples in the single-electron channel; the subsample with two b jets and one W boson (top), with two W bosons (one b jet middle left, two b jets middle right) and with three W bosons (one b jet bottom left, two b jets bottom right). Maximal cross sections are assumed for the signal processes, and all overlayed signal processes are scaled with a factor of five for visibility. The last bin includes the events in the overflow bin.

However, in semileptonic top-quark pair topologies, as depicted in Figure 4.10, there is an ambiguity to match the jets to the partons from which they originate. If the jets are wrongly matched, the reconstructed mass distributions will be less peaked and the sensitivity to the fourth-generation signal would be reduced. Hence, to reconstruct the mass of the hadronic bW system, a procedure to choose the proper jet combination is applied.

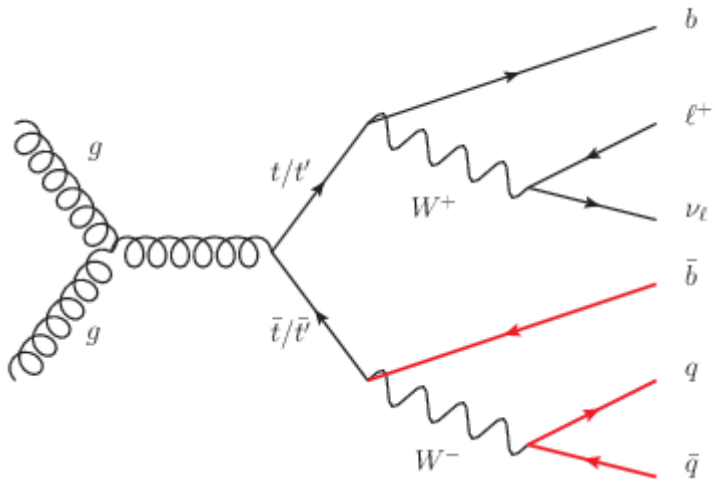

Figure 4.10: In semileptonic top-quark (or t'-quark) pair topologies, the assignment of jets to the underlying partons is ambiguous. When a correct jet combination is found, the invariant mass of the bW system (the 'hadronic side' of the event, indicated in red), will be representative for the mass of the top (or t') quark.

Choosing the best jet combination

As explained in section 4.4.1, the event selection in the 2b2W category requires at least two b jets, and a pair of jets satisfying the W-boson mass constraint. Hence, these four jets are chosen to reconstruct the event topology $t'\bar{t}' \to WbWb \to q\bar{q}b\ell\nu_\ell b$, but the assignment of each jet to a specific parton from this topology will be done entirely via a multi-variate analysis (MVA) technique. Without the use of b-tagging there would be $4! = 24$ possible combinations to decide which of the four jets originates from which hard parton. The jets assigned to the W boson can be interchanged without any effect on the mass of the W boson. Therefore, the problem is reduced to a 12-fold ambiguity. Each choice of jet combination results in a different event topology with different kinematic properties for the underlying top quarks and W bosons. Several variables are identified that differentiate between a correct jet combination and a wrong jet combination. For this purpose only selected simulated events in which a correct jet combination exists are studied. In a correct jet combination it is assumed that each of the three partons on the hadronic side matches a reconstructed jet with an angular criterion of $\Delta R(\text{jet}, \text{parton}) < 0.3$, where the ΔR definition uses the angular

metric $\Delta R = \sqrt{\Delta\theta^2 + \Delta\eta^2}$.[3] In [158] diverse observables are studied for this purpose as well as various multi-variate methods to combine the statistical information of these variables. Those variables that are strongly correlated with the top-quark mass (mass of the three-jet system associated to the hadronically decaying top quark) are rejected from the list, since we do not want to bias the mass reconstruction. Six variables are selected with a strong differentiation power between correct and wrong jet combinations in events where a correct combination exists. The following six variables are used:

- The sum of the TCHP b-tag discriminator values of the two jets assumed to originate from the b quarks in the jet combination, divided by the sum of the b-tag discriminator values of the four jets in the topology.
- The ratio of the transverse momentum of the hadronic top quark over the sum of the transverse momenta of all jet combinations that make a top-quark candidate in the event.
- The reconstructed W-boson mass in the $W \to q\bar{q}$ decay.
- The space angle $\Delta\Omega = \sqrt{\Delta\theta^2 + \Delta\phi^2}$ between the hadronic top quark and the lepton.
- The space angle $\Delta\Omega$ between the hadronic top quark and the b jet associated to the leptonically decaying top quark.
- The space angle $\Delta\Omega$ between the lepton and the b jet associated to the leptonically decaying top quark.

The distributions of these six variables for correct and wrong combinations in semimuonic $t\bar{t}$ events where a correct combination exists are shown in Figure 4.11. The correlation between the variables is small in general, with the largest linear correlation factor smaller than 40%. Given the small correlation, the six observables are employed in a likelihood-ratio technique to obtain one single MVA likelihood-ratio variable:

$$\text{Likelihood ratio} = \frac{S}{S+B} = \frac{\Pi_{i=1}^{6} f_i^S(x_i)}{\Pi_{i=1}^{6} f_i^S(x_i) + \Pi_{i=1}^{6} f_i^B(x_i)}, \tag{4.6}$$

where $f_i^B(x_i)$ and $f_i^S(x_i)$ are the probability density functions for variable x_i for wrong jet combinations ('background') and correct jet combinations ('signal'), respectively. When these probability density functions are determined from simulation, the multivariate analysis technique is said to be *trained*.

The jet combination which has the largest likelihood-ratio value is selected as the best combination in the event, as it is assumed to be most compatible with a semileptonic $t\bar{t}$-like topology. The procedure to choose the jet combination for the hadronic

[3] This means that sometimes, even in semileptonic $t\bar{t}$ events, jet-parton matching can fail. Final-state radiation can alter the direction of the final-state parton, such that the clustered jet has a larger spatial angular separation from its associated parton than the matching algorithm allows. Another situation where the algorithm can fail is when one of the selected jets actually originates from the initial or final state radiation itself.

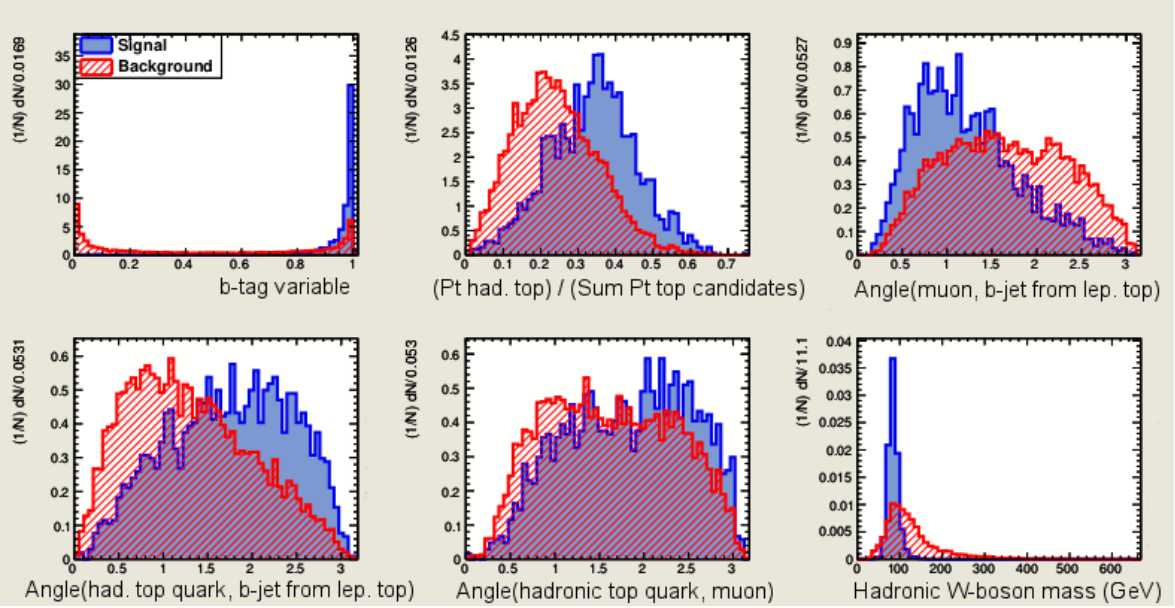

Figure 4.11: The correct (blue) and wrong (red) jet combination distributions of the six input variables used in the multi-variate analysis technique to choose the jet combination most compatible with a semileptonic $t\bar{t}$-like topology.

bW-mass reconstruction in the 1b2W category is treated analogously to the 2b2W category. The only difference is the list of four input jets to start with. Both for the 1b2W and 2b2W subsamples, an event is required to have at least four jets. In the 1b2W category there is only one b-tagged jet, and therefore the highest p_T jet among the remaining selected jets in the event is taken as the fourth input jet for the MVA jet-combination procedure. Figure 4.12 shows the distribution of the highest MVA value among all combinations in the event, for the selected data sample and the simulated event samples.

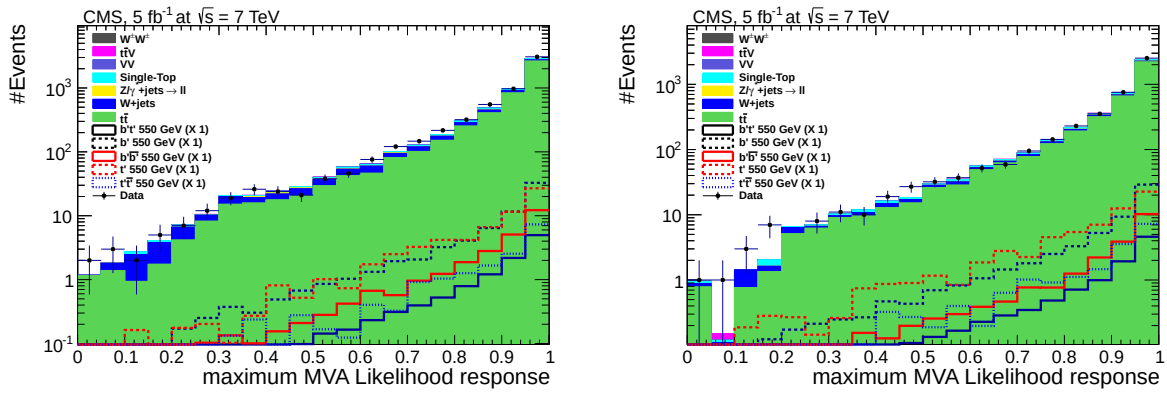

Figure 4.12: Maximum likelihood-ratio value among all jet combinations in the event, for the single-muon channel (left) and the single-electron channel (right), with the 1b2W and 2b2W subsamples combined.

For $t\bar{t}$ events in which a correct combination exists the efficiency of this choice is about 60% for the 1b2W and 2b2W subsamples together and in case the MVA method is trained on simulated $t\bar{t}$ events. This number should be compared to 8% (=1/12) in case a random choice is made. For t' pair production with a t'-quark mass of 550 GeV this efficiency drops to about 30%. When the MVA method is trained on $t'\bar{t}'$ signal events instead of $t\bar{t}$ events, the efficiency for taking the correct jet combinaton increases for the signal. However, at the same time the efficiency decreases for the $t\bar{t}$ events, in a way that a larger tail of the top-quark mass distribution (caused by the wrong reconstruction of the top-quark mass) tends to contaminate the $t'\bar{t}'$ signal region. This is illustrated in Figure 4.13. The most important aspect of the m_{bW} variable is not necessarily that it reconstructs the t'-quark mass as good as possible, but rather to construct a variable that has a good separation power between SM processes and the fourth-generation quark signal. Hence, we train on $t\bar{t}$ events in order to increase the signal over background ratio in the high-mass tail of the distribution, and in this way we have a fixed training to be used for all heavy-quark mass hypotheses.

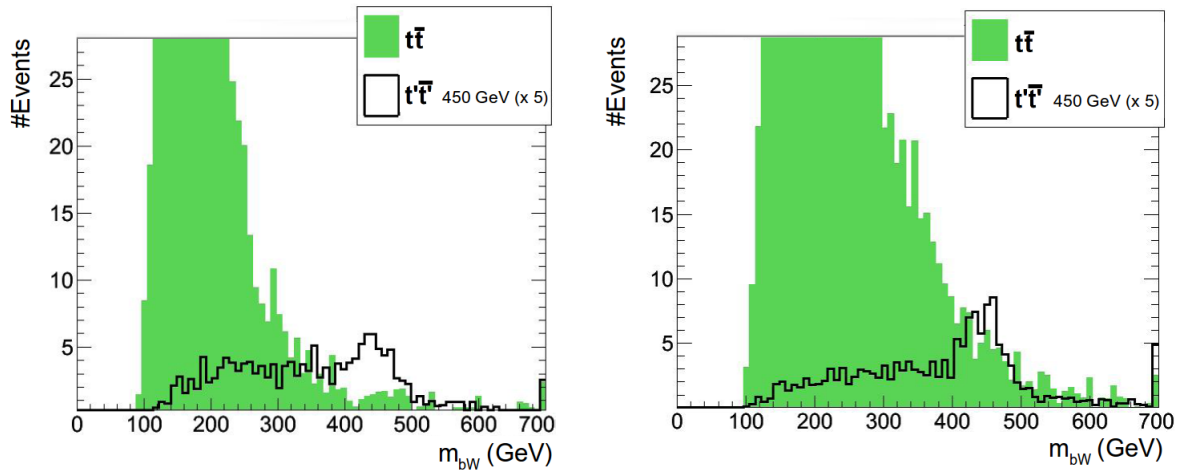

Figure 4.13: The distributions of m_{bW} for the $t\bar{t}$ background and an example $t'\bar{t}'$ signal of $m_{t'} = 450$ GeV, for the case where the training of the multivariate analysis technique was performed on simulated $t\bar{t}$ events (left) and on simulated $t'\bar{t}'$ events of mass $m_{t'} = 500$ GeV (right).

Reconstructing the m_{bW} variable

Once the best jet combination is chosen, the mass of the hadronic bW system is reconstructed. As mentioned before, the best combination is not necessarily the correct one, and the effect of correct and wrong combinations on the m_{bW} distribution for semimuonic $t\bar{t}$ is illustrated in Figure 4.14. Here the distributions are shown for the highest likelihood-ratio and for the hadronic bW mass, both for for correct and wrong jet combinations. It is clear that correct jet combinations tend to lead to a higher maximum likelihood-ratio value, as well as a more peaked mass spectrum around the generated top-quark mass of 172.5 GeV.

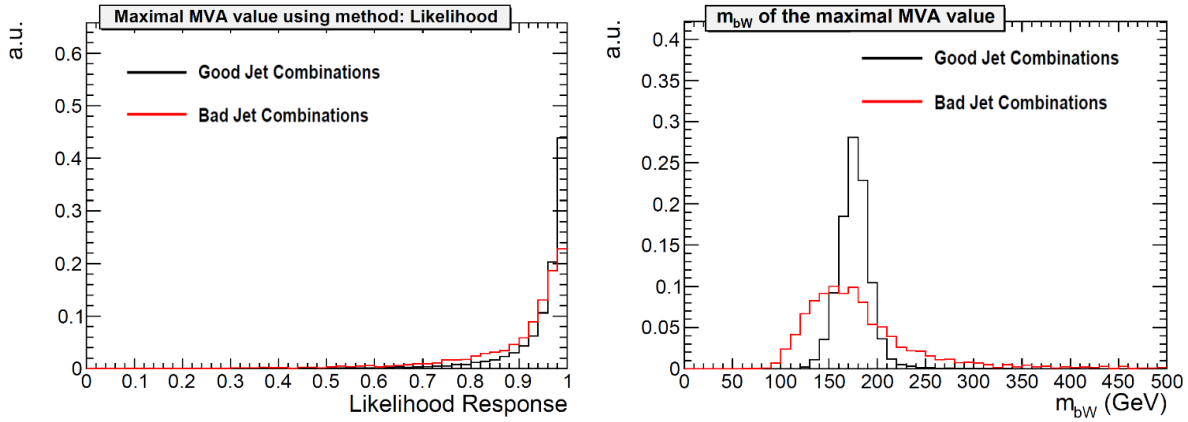

Figure 4.14: Left: the maximum likelihood-ratio values of simulated semimuonic top-quark pair events for which the associated jet combination is correct (black) or wrong (red). Right: the reconstructed mass of the hadronic bW system for simulated semimuonic top-quark pair events where the maximum likelihood-ratio corresponds to a correct or wrong jet combination.

The mass distributions for data and simulation in the subsamples with two W bosons are shown in Figure 4.15 for the single-muon channel, and in Figure 4.16 for the single-electron channel. A good agreement between data and simulation is observed. Comparing with the middle-row S_T distributions in Figures 4.8 and 4.9, the m_{bW} distributions exhibit a more peaked structure in the $t\bar{t}$ background, resulting in more discriminating power between the $t\bar{t}$ background and the $t'\bar{t}'$ signal. Moreover, a mass peak around the generated fourth-generation quark mass of 550 GeV is noticable, especially in the subsample with two b jets.

4.6 Combined search

We described the observables sensitive to the presence of fourth-generation quarks in the different subsamples in the previous section. These observables are defined in distinct exclusive subsamples, namely the seven single-lepton categories (catalogued according to the number of W bosons and b jets) and the multilepton final states. The purpose of this subdivision is to be sensitive to diverse topologies when considering electroweak and strong production of up-type and down-type fourth-generation quarks. In Section 4.6.1, we discuss how the information of the different topologies and the distributions of the relevant search variables are combined. The systematic sources affecting the modeling of the background and signal are outlined in Section 4.6.2, and the calculation of exclusion limits on the parameters of our simplified fourth-generation model is described in Section 4.6.3.

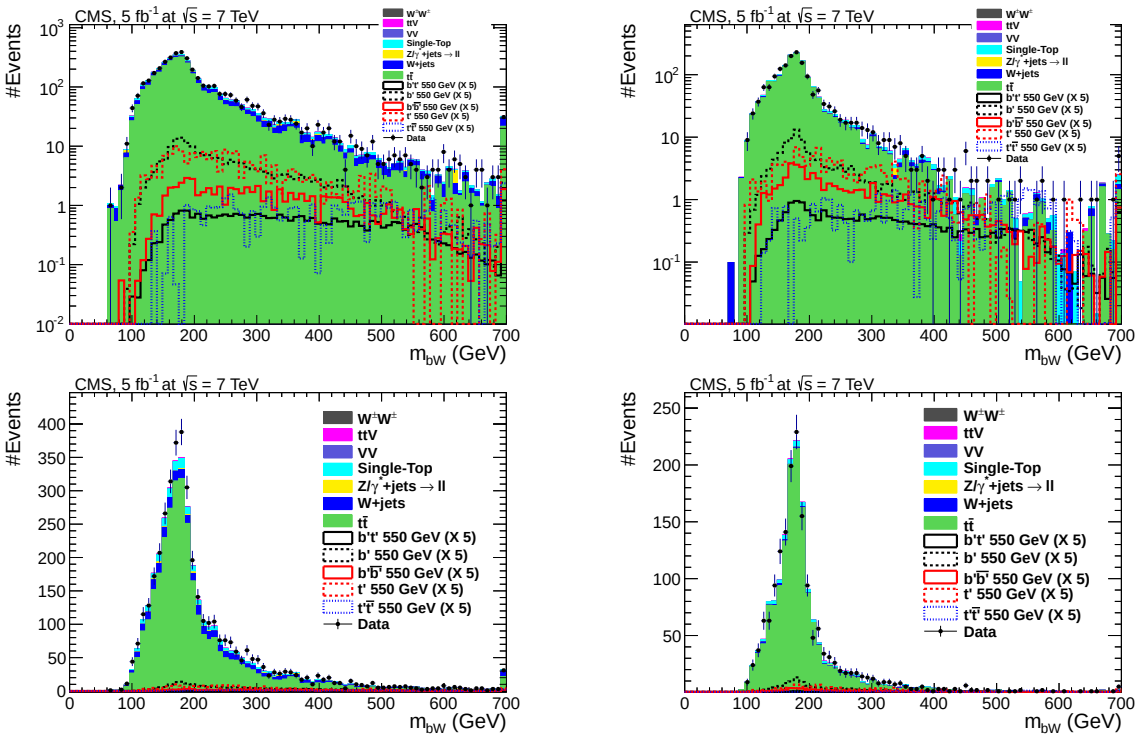

Figure 4.15: The observed and simulated distributions of the mass of the hadronic bW system for the 1b2W subsample (left) and the 2b2W subsample (right) in the single-muon channel and for fourth-generation quarks with mass 550 GeV (scaled by a factor of five for visibility). In the top plots the y axis is in logarithmic scale while in the bottom plots it is in linear scale.

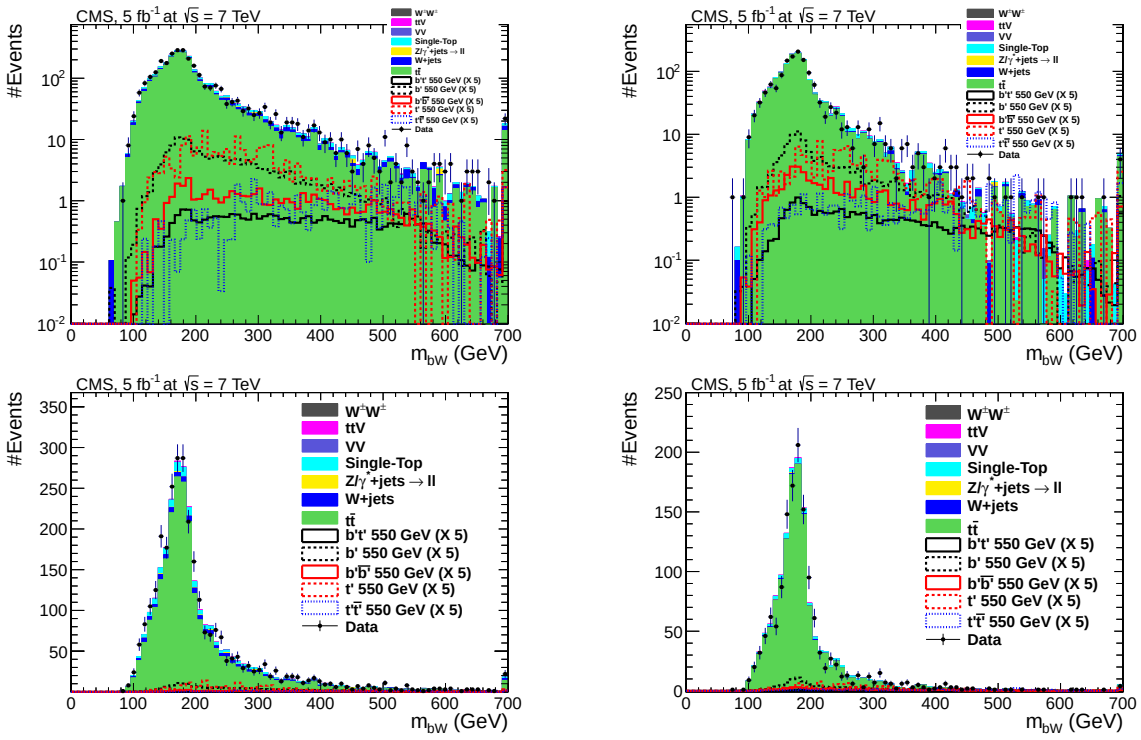

Figure 4.16: The observed and simulated distributions of the mass of the hadronic bW system for the 1b2W subsample (left) and the 2b2W subsample (right) in the single-electron channel and for fourth-generation quarks with mass 550 GeV (scaled by a factor of five for visibility). In the top plots the y axis is in logarithmic scale while in the bottom plots it is in linear scale.

4.6.1 Construction of template distributions

We build three different one-dimensional distributions, referred to as *templates*, corresponding to three channels. The division into channels is mainly for technical purposes, as the separate contribution of the channels to the exclusion limits can be easily evaluated in this way. The distributions of subsamples can be combined in one histogram because the subsamples were defined to be exclusive, in other words, they do not have data events in common and are therefore statistically independent.

- **The single-muon (single-electron) template.** This distribution is built from the observables of the different single-muon (single-electron) subsamples: the S_T variable, the mass of the hadronic bW system, and event counts. The actual construction of this template will be discussed below.

- **The multilepton template.** For the same-sign dilepton and trilepton final states only event counts are considered. Therefore a template with four bins is constructed; three bins containing the number of same-sign dilepton (ee, $e\mu$ and $\mu\mu$) events, and another bin with the trilepton events.

Template binning procedure in subsamples with two W bosons

As explained in Section 4.5, two kinematic variables are defined for the 1b2W and 2b2W subsamples of the single-lepton topology: the S_T variable, and the mass of the reconstructed hadronic bW system. These two observables can be combined in a two-dimensional distribution, with a binning chosen such that no bins are depleted in SM background events. The reason bins without SM background events should be avoided is because this is required by the likelihood-ratio hypothesis-test statistic that we will employ in the limit setting procedure in Section 4.6.3. Correlations between the S_T and m_{bW} variables are naturally taken into account, because later on a fit of these two-dimensional distributions will be performed.

To construct the two-dimensional binning, first the binning in the dimension of the hadronic bW mass variable is defined in such a way that the selected SM top-quark pair events are uniformely distributed over N bins. In a second stage, in each of these hadronic bW mass bins a binning is chosen in the dimension of the S_T variable to acquire also in this dimension a uniform distribution of the top-quark pair events in M bins. The values of N and M are chosen to be 18 and 10, respectively, for the 1b2W category, and 12 and 6 for the 2b2W category. Hence, a total of 180 (72) bins are obtained in the 1b2W (2b2W) category, each with the same expected yield for selected top quark pair events. The number of bins is chosen such that the two-dimensional distributions have about 15 to 20 expected top-quark pair events per bin. By decreasing the background expectation in each bin one can expect the signal-to-background ratio to increase in some bins. However, if the background expectation becomes too small, one becomes too sensitive to statistical fluctuations in the background template, due to the finite amount of simulated events available. The two-dimensional distributions with the bin boundaries obtained from the described binning procedure are shown in Figure 4.17 in the 1b2W and 2b2W categories in the single-muon channel, for $t\bar{t}$

background and $t'\bar{t}'$ signal events seperately. This illustrates the different population of these processes in the two-dimensional (S_T, m_{bW}) space.

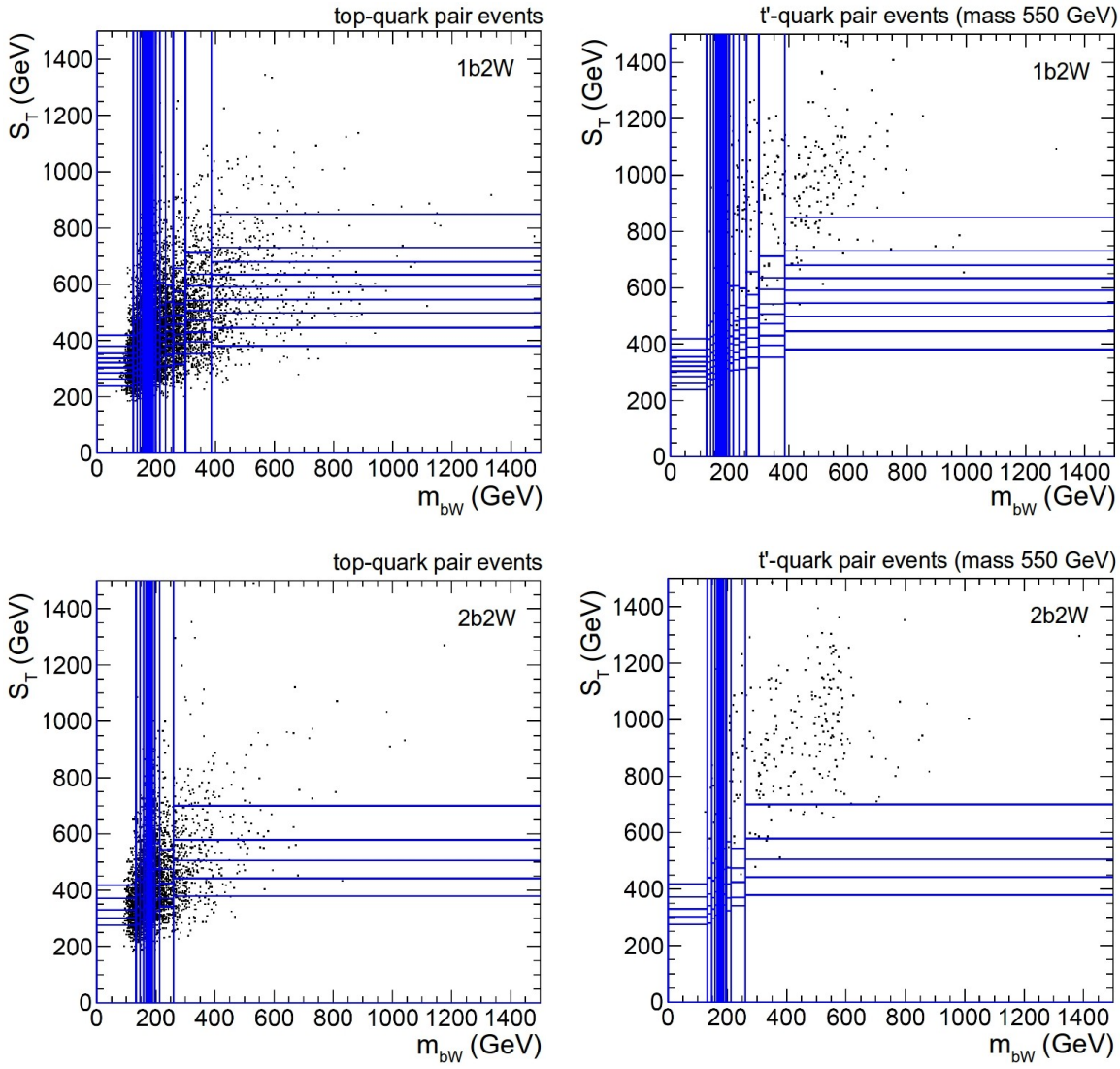

Figure 4.17: The two-dimensional hadronic bW mass versus S_T distributions for top-quark pair events (left) and signal t' pair events with a t'-quark mass of 550 GeV (right), in the 1b2W subsample (top) and 2b2W subsample (bottom) in the single-muon channel. The binning, indicated with the blue lines, is chosen to obtain a two-dimensional template with a uniformely distributed background from top-quark pair events.

Next, we *unroll* the two-dimensional distributions to a one-dimensinal distribution by joining the N different S_T distributions together. This is merely for representational purposes (to visualize the agreement between simulation and data) and technical reasons (because the limit calculation tool is based on one-dimensional distributions as input). The unrolled distributions are shown in Figure 4.18 for the single-muon channel, and Figure 4.19 for the single-electron channel. One can see that the distributions

are indeed flat in top-quark pair background, by construction. Events of non-$t\bar{t}$ SM background processes like the W+jets process tend to accumulate mostly in the bins corresponding to low and high m_{bW} values (low and high bin numbers as displayed in the unrolled distributions). This is an artifact of the m_{bW} mass reconstruction and the binning procedure because of the following. The m_{bW} distribution is peaked around the generated top-quark mass for $t\bar{t}$ events, but tends to be broader for the W+jets process since no 'correct' jet combinations (as defined in Section 4.5.2) exist in these topologies. Hence, when flattening the top-quark pair distribution, a finer binning will be taken around the top-quark mass peak (see Figure 4.17) than in the tail(s) of the distribution. This results in a larger relative accumulation of non-$t\bar{t}$ SM background events towards the right (and the left) of the unrolled distributions compared to the center.

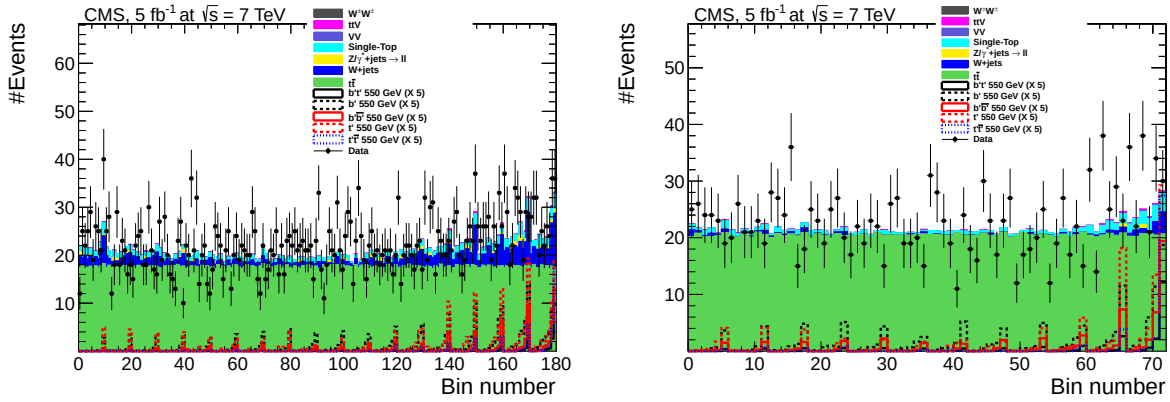

Figure 4.18: The unrolled two-dimensional distribution of the S_T variable and the mass of the hadronic bW system, for the 1b2W subsample (left) and 2b2W subsample (right) in the single-muon channel. In the left distribution, the first 10 bins represent the S_T distribution in the first bin of the m_{bW} distribution, bin 11 to 20 represent the S_T distribution in the second bin of the m_{bW} distribution, and so on.

Construction of combined single-lepton templates

In the distributions corresponding to the seven subsamples of the single-lepton topology, each bin has an expected amount of events for SM-processes and for fourth-generation quark processes depending on the model parameters, $m_{t'}$, $m_{b'}$ and the $V_{\mathrm{CKM}}^{4\times4}$ matrix element A. All bins in all of the seven subsamples are treated independently from each other. The correlation between the bins is introduced by the systematic effects that can modify the expected number of events in a bin. For example, if the jet energy scale is altered, events might migrate from one bin to another, and even from one subsample to another. In practice this means that it is allowed to combine the S_T distributions of the seven subsamples 2b1W, 1b3W, 2b3W, 1b4W, 2b4W and the unrolled two-dimensional distributions of the two subsamples 1b2W and 2b2W into one larger histogram. The amount of bins of this histogram, shown in Figure 4.20 for the single-muon channel, equals the sum of the number of bins in each of these independent seven subsamples.

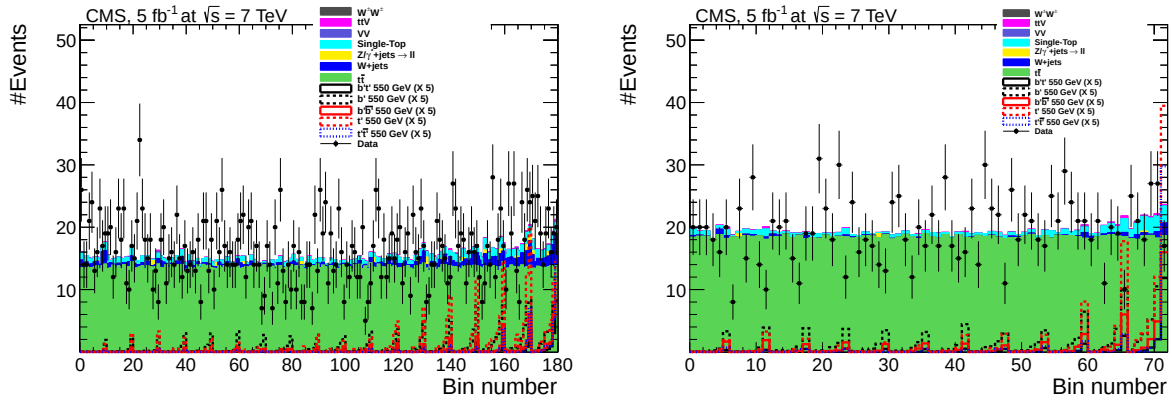

Figure 4.19: The unrolled two-dimensional distribution of the S_T variable and the mass of the hadronic bW system, for the 1b2W subsample (left) and 2b2W subsample (right) in the single-electron channel.

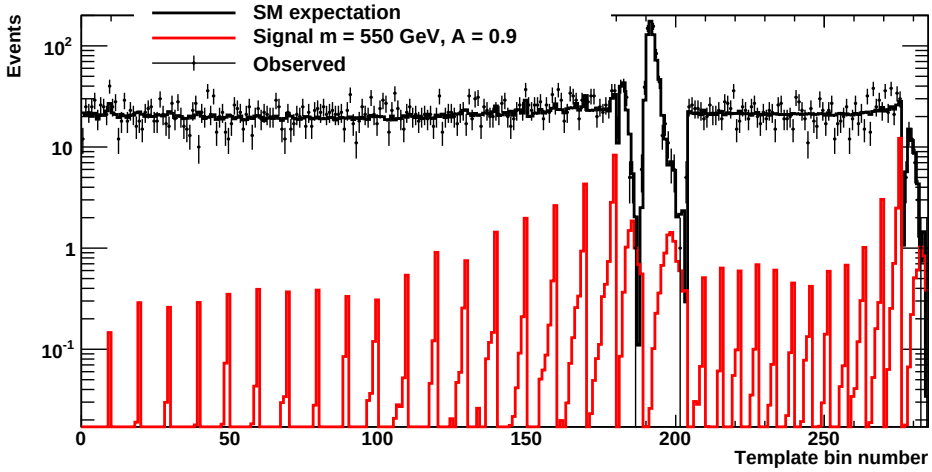

Figure 4.20: The combined one-dimensional histogram of the search variables in the seven subsamples of the single-muon channel. The total Standard-Model contribution is indicated in black, the total fourth-generation signal (with $A = 0.9$ and for a $m_{t'} = m_{b'} = 550$ GeV, and *without* extra scaling for visibility) in red, and the observed data is superimposed with black crosses. The order of the subsamples in the plot is the following (from left to right): 1b2W (the part where the SM contribution is more or less uniform), 1b3W, 1b4W, 2b1W (the higher peak in the middle of the plot), 2b2W (SM contribution mostly uniform), 2b3W, 2b4W.

4.6.2 Systematic uncertainties

Various systematic sources can affect both the shape and the normalization of the distributions of the observables. The effect can be different in each of the subsamples because a different composition of SM processes make up the background in each of them. The following sources are studied and the systematic effects corresponding to an uncertainty of one standard deviation ($\pm 1\sigma$) are considered as the systematic uncertainty that will be used in the statistical fit procedure afterwards.

- **Integrated luminosity.** The integrated luminosity of the collected data is measured with a precision of 2.2% [159], and has the same normalization effect on all the templates.

- **Muon and electron identification, selection and trigger efficiency.** In this analysis we use the isolated-muon and electron+jet triggers. The trigger efficiency for the isolated-muon triggers has been estimated from data using a tag-and-probe technique with leptonically decaying Z bosons, and compared to their expectation from the simulation. A very good agreement to the percent level has been observed for the trigger efficiency, as well as for the muon identification. We use a total of 3% (5%) as a 1σ effect on the total muon (electron) efficiency.

- **Cross section of the different background processes.** The dominant background processes in the seven subsamples of the single-lepton topology are the top-quark pair, W+jets and single-top production. For the simulated top-quark pair and W+jets processes we use the cross section values measured by the CMS experiment, and the uncertainties in them are taken as the corresponding systematic uncertainties. The most important contributions that affect the normalization of the templates are the 12% (30%) uncertainty for the top-quark pair [147] (single-top) production cross section and a 50% uncertainty for the W production cross section because of the large fraction of selected events with jets from heavy-flavor quarks.[4] Relevant for the multilepton channel, we include the uncertainties in the production cross sections of Z+jets (5% [148]), WW (35%), WZ (42%), ZZ (27%), $t\bar{t}+W$ (19%), $t\bar{t}+Z$ (28%) and $W^{\pm}W^{\pm}$ (49%). The uncertainties in the normalization of diboson and top-quark pair production in association with a boson are taken from a comparison of the NLO and the LO predictions.

- **Jet energy scale.** The event selection and the search variables in particular are sensitive to the energy scale of the jets as well as the missing transverse energy. Therefore a correct estimate of the jet energy scale of the observed jets is crucial for this analysis. We apply the standard CMS recipe to alter the energy scale of the reconstructed jets according to the estimated p_T and η dependent

[4] In this context, the heavy-flavor quarks that are produced in association with W (or Z) bosons are b or c quarks, and in general the cross section of the corresponding processes is not as precisely known as the cross section for W (or Z) bosons being produced in association with light-flavor quarks u, d and s. The difficulty to model the production cross section of a W boson in association with a b jet is illustrated by the measurement of the CDF experiment [160] of a deviation of about three standard deviations with respect to NLO predictions.

uncertainties [135]. The missing transverse energy is altered by this recipe as well, since the momenta of the final-state objects should stay balanced in the transverse plane. When the jet energy scale is modified to a lower (higher) value, this results in a shift of the S_T values to lower (higher) values. For this purpose, two new templates are determined for each physics process which reflect the 1σ upper and lower scaling of the jet energy scale.

- **Jet energy resolution.** The overall mismatch in the jet energy resolution between the simulation and the collision data has already been included in the nominal templates. This was done by smearing the jet resolution in the simulation in order to match the resolution obtained from data. The uncertainty in the estimation of the jet energy resolution in data depends on the pseudorapidity of the jets. Templates are made with a $\pm1\sigma$ effect of the jet energy resolution. The resolution of the missing transverse energy is also altered with this procedure.

- **Scale factor for b-tag efficiency.** Since we make extensive use of b-tagging in this analysis, the corresponding systematic uncertainties have a clear impact on the overall scale of the templates of the discriminating observables. Scale factors to account for differences in b-tagging efficiency in simulation and data are estimated from the collision data and applied in the nominal templates. These scale factors and their uncertainties depend on the transverse momentum and the pseudorapidity of the jets. The -1σ ($+1\sigma$) templates for the diverse processes for this systematic uncertainty are obtained using scale factors which are shifted down (up) by the uncertainty in the measured scale factor [157].

- **Scale factor for non-b-tag efficiency ('mistag rate').** From the collision data, scale factors are measured that accommodate the difference in efficiency of tagging a light-flavor quark or gluon as a b jet in simulation and data. The nominal template and the $\pm1\sigma$ templates for the physics processes for this systematic uncertainty are obtained by weighting each simulated event according to the number of b-tagged light-quark or gluon jets.

- **Pile-up reweighting.** The distribution of the primary vertices in the simulated samples are shifted up and down with -5% and $+5\%$, resulting in different expected templates. This uncertainty originates from the uncertainty in the inelastic proton-proton collision cross section.

- **Data-driven background estimation.** The uncertainties in the data-driven background estimation for same-sign dilepton events where one of the leptons has a misidentified charge or is a fake lepton are listed in Table 4.8. Although these uncertainties are essentially statistical, they are propagated as systematic uncertainties to the statistical fit procedure.

Figure 4.21 shows the single-muon SM template distribution for the nominal and for the $\pm1\sigma$ systematically shifted expectation of the jet energy scale. This is the dominating systematic source that can alter the shape of the distributions rather than merely the normalization.

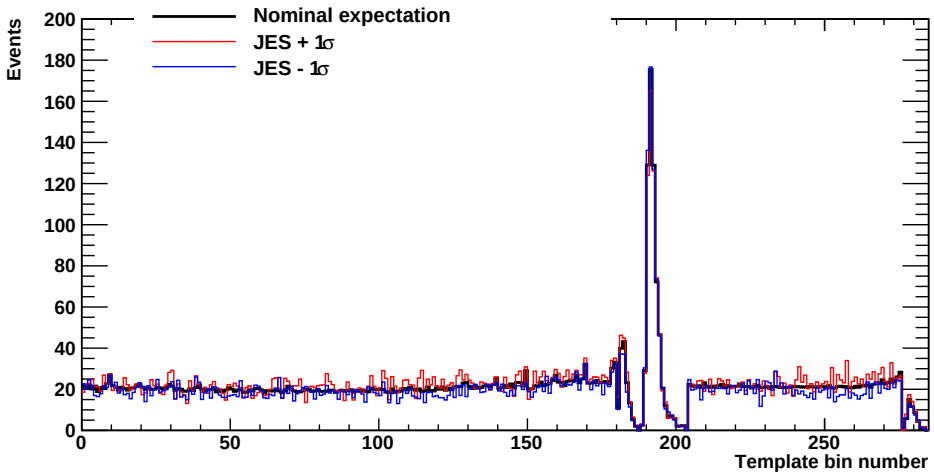

Figure 4.21: The combined one-dimensional histogram of the search variables in the seven subsamples of the single-muon channel, for the nominal SM expectation and the templates shifted by the $\pm 1\sigma$ variations on the jet energy scale.

4.6.3 Limit setting procedure

We use the predicted templates constructed in Section 4.6.1 and 4.6.2 to test for the compatibility of the observed data with either the Standard-Model background-only hypothesis or the background plus fourth-generation signal hypothesis. Rather than making explicit cuts on these templates to isolate potential signal events, we use a profile likelihood ratio method to make use of the information of the shapes of all nominal and systematically shifted distributions in a more optimal way.

The probability model

Our statistical model depends on several parameters that are unkown and that can alter the shape and normalization of the defined discriminating variables. We make a distinction between the *parameter of interest* μ (commonly referred to as the signal strength) and the *nuisance parameters* $\boldsymbol{\alpha}$:

- **Parameter of interest.** In our case, the parameter of interest is the strength of the combined fourth-generation signal, which can be interpreted as some assumed cross section of the combined signal processes normalized by the expected combined signal cross section (from theoretical calculations). In this way, $\mu = 0$ indicates the absence of signal, and $\mu = 1$ the presence of signal with the yield predicted by the theory. In this analysis, the signal is composed of multiple processes, and depends implicitly on the masses of the fourth-generation quarks and the $V_{\mathrm{CKM}}^{4\times4}$ parameter A. Clearly, the parameter A governs the cross section of the electroweak production of fourth-generation quarks, and the masses $m_{t'}$ and $m_{b'}$ affect the cross sections of the strong and electroweak processes. Since the relation between the signal strength and these underlying theoretical parameters is

non-trivial, we will need to scan over A and $m_{Q'}$, and in each probed scan point, the total signal distribution is the sum of all individual signal distributions.

- **Nuisance parameters.** In our context, these parameters $\boldsymbol{\alpha}$, which can be viewed as a set $\{\alpha_p\}$, are corresponding to the systematic uncertainties; these include theoretical uncertainties in cross sections of simulated processes, and experimental uncertainties in the response and calibration of the detector (e.g. jet energy scale, b-tag uncertainties, and so on). Even though we are not directly interested in the determination of these unknown nuisance parameters, they still affect the predicted template distributions and can therefore influence the statistical inference of the parameter of interest. In practice a nuisance parameter can be considered as a parameter interpolating between the nominal and systematically shifted distributions. For example, a nuisance parameter value $\alpha_{t\bar{t}} = 0$ corresponds to the nominal measured value of the $t\bar{t}$ production cross section, and $\alpha_{t\bar{t}} = \pm 1$ indicates a shift in the $t\bar{t}$ cross section of one standard deviation up or down, $\pm 1\sigma$. Values of $|\alpha_{t\bar{t}}|$ between 0 and 1 indicate a systematic shift less than one standard deviation, and are interpreted as a linear scaling between the nominal and systematically shifted templates. In case the systematic uncertainty alters the shape of the templates (e.g. the jet energy scale), template morphing [161, 162] is used to interpolate linearly on a bin-by-bin basis between the nominal templates and the systematically shifted ones.

We use the implementation of the statistical model in the HistFactory tool [163]. Schematically, the total joint probability model f_{tot}, incorporating systematic uncertainties, for the observed data set $\mathcal{D}$ in several channels with n_c events is a product of probability density functions:

$$f_{\text{tot}}(\mathcal{D}, \boldsymbol{a} \mid \mu, \boldsymbol{\alpha}) = \prod_{c \in \text{channels}} \left[\text{Pois}(n_c | \nu_c(\mu, \boldsymbol{\alpha})) \prod_{e=1}^{n_c} f_c(x_e | \mu, \boldsymbol{\alpha}) \right] \cdot \prod_p f_p(a_p | \alpha_p). \quad (4.7)$$

Here $f_c(x_e|\mu, \boldsymbol{\alpha})$ represents the probability density function (for channel c) of a discriminating variable evaluated in the measured values x_e, while $f_p(a_p|\alpha_p)$ are so-called constraint terms. The observable a_p represents the maximum-likelihood estimator of the underlying unknown nuisance parameter α_p, and has corresponding $\pm 1\sigma$ measurement uncertainties. For instance, it could describe a measurement of the jet energy scale that only *estimates* the true unknown value of the jet energy scale. The constraint terms are then idealized probability density functions describing such an *auxiliary* measurement a_p that constrains the nuisance parameter α_p. These are often taken to be Gaussian, Poisson or log-normal distributions, but in this analyis we choose Gaussian constraints. We write $\boldsymbol{a} = \{a_p\}$ as the set of estimators of nuisance parameters, corresponding to the set $\boldsymbol{\alpha} = \{\alpha_p\}$ of nuisance parameters themselves. The product of the constraint terms is taken over the set of all systematic sources. The factor $\text{Pois}(n_c|\nu_c)$ in Equation (4.7) determines the probability to observe n_c events in channel c, given an expecation of ν_c, which in principle is also a function of the model parameters μ and $\boldsymbol{\alpha}$. In the single-lepton channels, the discriminating variable x rather refers to the combination of discriminating variables in the histogram template.

The profile likelihood ratio test statistic

In order to perform hypothesis tests, a test statistic needs to be defined. This is a single real-valued quantity that describes the outcome of a whole experiment. In searches for new physics, the test statistic should be defined in such a way that it is discriminating between the background-only and the signal-plus-background hypothesis.

From the probability model (4.7), we can construct the likelihood function, which has the same expression, but is rather considered as an explicit function of the parameter of interest μ and the nuisance parameters $\boldsymbol{\alpha}$, for a given data set $\mathcal{D}$ and estimators $\boldsymbol{a}$ from auxiliary measurements:

$$L(\mu, \boldsymbol{\alpha}; \mathcal{D}, \boldsymbol{a}) = f_{\text{tot}}(\mathcal{D}, \boldsymbol{a} \mid \mu, \boldsymbol{\alpha}). \tag{4.8}$$

The maximum-likelihood estimates $\hat{\mu}$ and $\hat{\boldsymbol{\alpha}}$ are the values of the parameters that maximize this likelihood function. Note that these estimates depend on the observed data, and in practice they will originate from a fit of templates to the data. On the other hand, the conditional maximum-likelihood estimates $\hat{\hat{\boldsymbol{\alpha}}}(\mu)$ are the values of the nuisance parameters that maximize the likelihood function with the signal strength μ fixed. The procedure of choosing specific values of the nuisance parameters is called 'profiling'.

Given these definitions, the profile likelihood ratio $\lambda(\mu)$ can be defined as the ratio of likelihood functions:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\alpha}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\alpha}})}. \tag{4.9}$$

The profile likelihood ratio $\lambda(\mu)$ only depends explicitly on the parameter of interest (implicitly on the data $\mathcal{D}$ and auxiliary measurements $\boldsymbol{a}$), while it is independent of the true nuisance parameters $\boldsymbol{\alpha}$ since they are profiled ('fitted to the data'), either with μ fixed or not fixed. Small values of the profile likelihood ratio generally indicate that the data agrees poorly with the hypothesized signal strength μ. For setting a one-sided upper limit on the signal strength, we use the test statistic

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{4.10}$$

which is set to zero for values of the signal strength μ smaller than the best-fitted value $\hat{\mu}$. Note that the test statistic for the observed data is evaluated at an assumed signal strength μ that is hypothesized and on which one aims to set a limit.

Distributions of the test statistic can be obtained for both the background-only hypothesis and the signal-plus-background hypothesis, for instance by generating pseudo-experiments using Monte Carlo techniques. While the profile likelihood ratio itself is independent of the nuisance parameters, the *distributions* of the test statistic might still depend on $\boldsymbol{\alpha}$. For a large enough data set, asymptotic formulae for the test statistic distributions have been derived [164]. They are based on the identity $-2 \ln \lambda(\mu) = (\mu - \hat{\mu})^2/\sigma^2 + \mathcal{O}(1/\sqrt{N})$ [165], where N represents the data sample size and σ the variance of a Gaussian distribution of $\hat{\mu}$. This variance can be calculated from the covariance matrix of the estimators of the parameter of interest and the nuisance parameters. As a consequence, if the sample size is large enough, asymptotic test statistic distributions can be derived without performing pseudo-experiments.

The CLs method

We use the CLs method with an asymptotic Gaussian approximation of the test statistic distributions [164, 166] to extract one-sided confidence intervals on the signal strengths. Given a background-only distribution and a signal-plus-background distribution of the profile likelihood ratio test statistic (4.10), as well as the observed test statistic from data $q_{\mu,\text{obs}}$, one can calculate for both hypotheses the p-value of the data. The p-value assuming a particular hypothesis $(\mu, \boldsymbol{\alpha})$ is the probability to observe an equal or more outlying outcome than observed. Hence, small values mean a low compatibility of the data with that hypothesis. The CLs value is defined from the p-values under signal-plus-background and background-only hypotheses as:

$$\text{CLs} = p_{s+b}/(1 - p_b). \tag{4.11}$$

The signal-plus-background hypothesis will be rejected at a confidence level of 95% if the observed CLs value is equal or below 0.05. The CLs method is designed to prevent downward fluctuations in the background that would exclude models for which the search has little sensitivity.[5] An *expected* exclusion limit with $\pm 1\sigma$ and $\pm 2\sigma$ 'uncertainty' bands can be calculated as well. This is done by considering the CLs value of the median of the background-only test-statistic distribution and the CLs values of the 68% and 95% quantiles of the background-only distribution.

The procedure of finding the signal strength for which CLs = 0.05 and that therefore can be excluded, requires scanning over different values for the signal strength. In this analysis, we probe each point in the fourth-generation model parameter space $(A, m_{Q'})$. If the excluded signal strength is smaller than one (which is the signal strength set equal to the expected one from theoretical predictions of the combined fourth-generation model), we conclude that the model point $(A, m_{Q'})$ is excluded at the 95% confidence level.

4.7 Results and discussion

Using the full set of combined templates for observed data, and Standard-Model and fourth-generation expectations including all systematic uncertainties, we can construct the full probability model as described in Section 4.6.3. When considering the profile likelihood ratio test statistic, we perform fits (maximum-likelihood estimations) of the nuisance parameters corresponding to the systematic uncertainties, both seperately and simultaneously with the combined fourth-generation signal strength. The results of the fits for the background-only model are discussed in Section 4.7.1. In Section 4.7.2 and Section 4.7.3 we derive the final limits on the parameters of our fourth-generation quark model for degenerate t' and b' quark masses and non-degenerate masses, respectively. In Section 4.7.4 we perform a signal injection test that serves as a cross check that our search would indeed be sensitive to the presence of a fourth-generation quark signal.

[5] When the background-only and signal-plus-background hypothesis have almost completely overlapping test statistic distributions, one should not expect to be able to exclude the signal-plus-background hypothesis. The CLs method indeed prevents this, because for such overlapping distributions, p_{s+b} will be similar to $(1 - p_b)$, causing the CLs value to be CLs $\approx 1 > 0.05$.

We test the sensitivity provided by each defined subsample as well. To conclude, Section 4.7.5 summarizes the analysis approach and the main results of this combined search for chiral fourth-generation quarks.

4.7.1 Fitted background model

To quantify the level of agreement of the data with the background-only hypothesis, we can fit the nuisance parameters, corresponding to the systematic uncertainties, to the data. Note that most systematic uncertainties are cross-correlated over the single-muon, single-electron and multilepton channels. For instance, the jet-energy scale is not expected to differ from channel to channel, so the same nuisance parameter is describing this effect throughout all channels. Other systematic effects can affect only one channel, however, like the fake-lepton and charge misidentification uncertainties that are only relevant to the multilepton channel. The fitted nuisance parameters $\hat{\hat{\boldsymbol{\alpha}}}(\mu = 0)$ we obtain from the fit of the background-only hypothesis to the data are quoted in Table 4.12. A $V_{CKM}^{4\times4}$ parameter value of $A = 1$ has been used reflecting the maximal production cross section for the single-top process as expected from the Standard Model ($|V_{tb}| \approx 1$). For some nuisance parameters, the values can be interpreted as the fitted effect on the physics quantity when possible. Systematic uncertainties affecting also the shape of the templates are less straightforward to interpret, as they have been obtained using template morphing and do not only result in a normalization effect.

To match the data in the best possible way given the background-only model and the nuisance parameters described, we see from Table 4.12 that the top-quark pair cross section is scaled down by 4%, the W+jets cross section is scaled down by 9% and the single-top cross section scaled up by 75%. These numbers should not be taken as measurements of the respective production cross sections of these processes because the uncertainty in these fitted nuisance parameters is rather large, and the event selection is not optimized for their measurements. The maximum likelihood estimations of the nuisance parameters are a multidimensional fit with many degrees of freedom. Hence, the interplay between effects is non-trivial. Furthermore, some simplifying assumptions on the constraint terms $f_p(a_p|\alpha_p)$ are made, since they are idealized terms like Gaussian distributions and do not necessarily provide the most realistic description of the auxiliary measurements. The most reliable estimation of the underlying nuisance parameters still comes from the auxiliary measurement a_p themselves, since these have been determined in dedicated optimized analyses. In addition, the discriminating variables used in the fit in this analysis have the purpose to differentiate between the fourth-generation processes and the SM processes. They have less separation power to differentiate between the SM processes themselves.

The expected S_T and m_{bW} distributions of a background-only model can be compared to data after including the systematic shifts given by the fitted nuisance parameters $\hat{\hat{\alpha}}(\mu = 0)$. These distributions are shown in Figure 4.22 and 4.24 for the single-muon channel and Figure 4.23 and 4.24 for the single-electron channel. Figures 4.25 and 4.26 show the distributions for the single-muon and single-electron channels combined after inclusion of the fitted nuisance parameters, where the red dashed bands indicate the total systematic uncertainty in the unfitted background modeling. In general, the fitted background expectation and the observed data agree well. The signal in these

Table 4.12: Overview of the nuisance parameters (describing the effects of the systematic uncertainties) fitted to data with a background-only model. For convenience, the fitted values of systematic sources affecting the normalization of the templates are transformed in the third column into relative shifts on the relevant physics parameter.

Nuisance parameter	Fitted $\hat{\alpha}(\mu = 0)$	Fitted effect
integrated luminosity	/	+0.5%
$\alpha_{muon\ id}$	+0.37 σ	+1.11%
$\alpha_{electron\ id}$	-0.18 σ	-0.9%
$\alpha_{t\bar{t}}$	-0.33 σ	-4.0%
α_{W}	-0.18 σ	-9%
α_{Z}	+0.003 σ	0.02%
α_{t}	+2.5 σ	75%
α_{WW}	-0.03 σ	-1.1%
α_{WZ}	+0.03 σ	1.3%
α_{ZZ}	-0.002 σ	-0.05%
$\alpha_{t\bar{t}W}$	+0.12 σ	2.3%
$\alpha_{t\bar{t}Z}$	-0.04 σ	-1.1%
$\alpha_{W^{\pm}W^{\pm}}$	+0.13 σ	6.37%
α_{JES}	+0.44 σ	/
α_{btag}	+0.35 σ	/
α_{mistag}	-0.85 σ	/
α_{JER}	-0.22 σ	/
α_{PU}	+0.79 σ	/
$\alpha_{charge\ misid\ e}$	+0.04 σ	/
$\alpha_{fake\ lepton\ e}$	+0.07 σ	/
$\alpha_{fake\ lepton\ \mu}$	+0.02 σ	/

distributions, scaled by a factor eight for visibility, corresponds to the combined signal processes for $m_{t'} = m_{b'} = 550$ GeV, for a $V_{CKM}^{4\times4}$ parameter A of 0.8 and 1, as indicated in the legends.

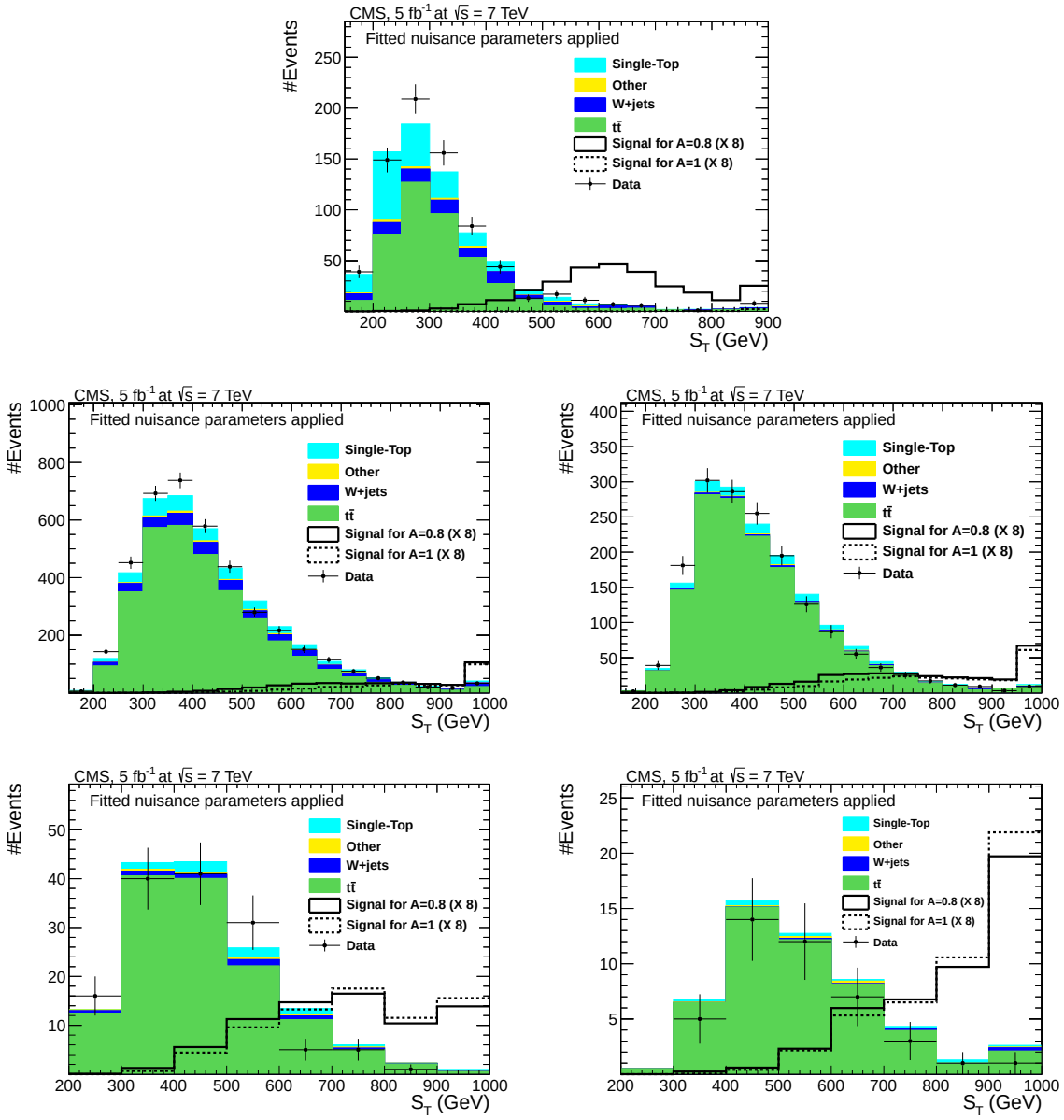

Figure 4.22: The S_T distributions in the different subsamples of number of b-tagged jets and number of hadronically decaying W bosons after the inclusion of the fitted nuisance parameters from Table 4.12, for the single-muon channel. The first row shows the S_T distribution in the 2b1W category, the middle row corresponds to the 1b2W (left) and the 2b2W (right) categories, and the bottom row distributions are for the 1b3W (left) and the 2b3W (right) categories.

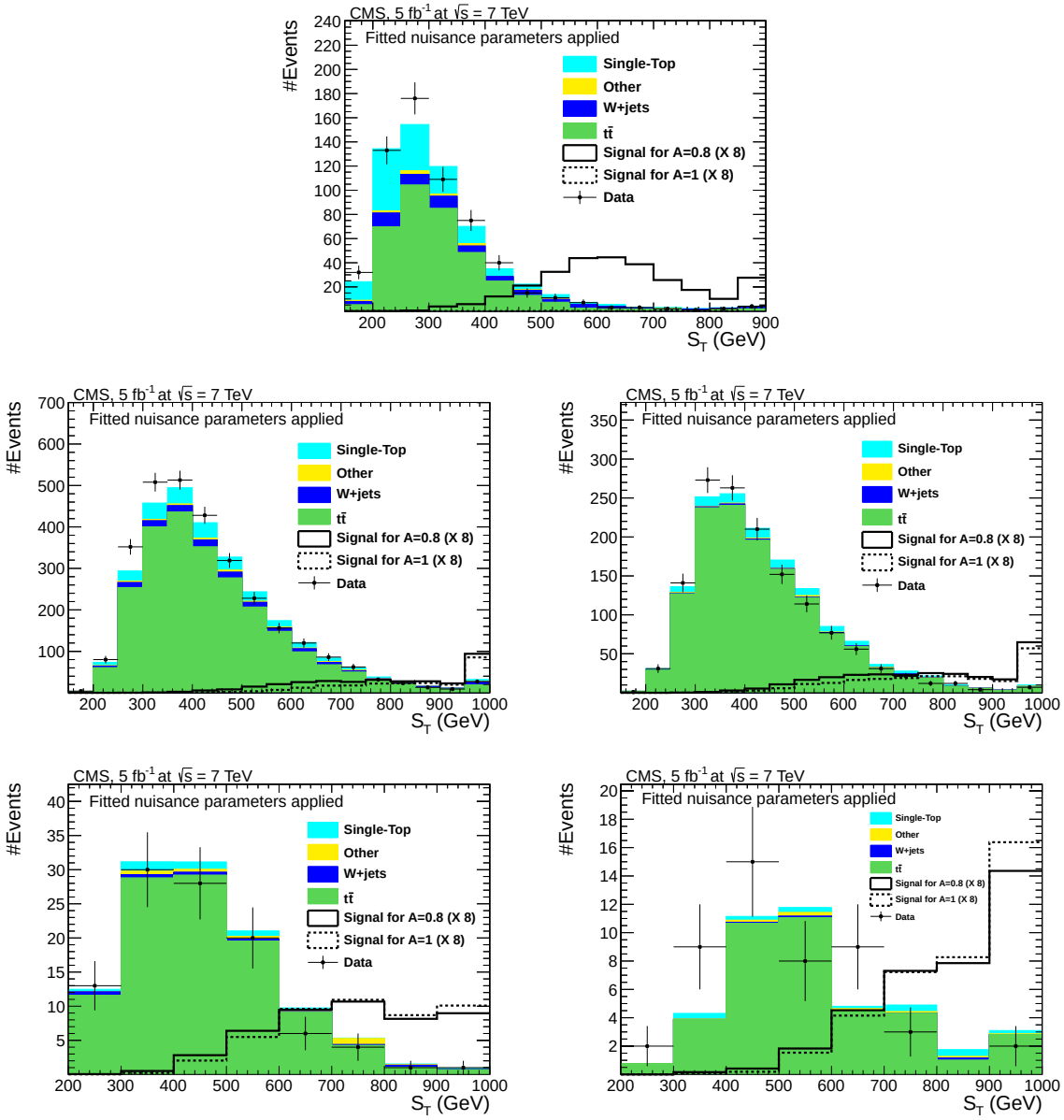

Figure 4.23: The S_T distributions in the different subsamples of number of b-tagged jets and number of hadronically decaying W bosons after the inclusion of the fitted nuisance parameters from Table 4.12, for the single-electron channel. The first row shows the S_T distribution in the 2b1W category, the middle row corresponds to the 1b2W (left) and the 2b2W (right) categories, and the bottom row distributions are for the 1b3W (left) and the 2b3W (right) categories.

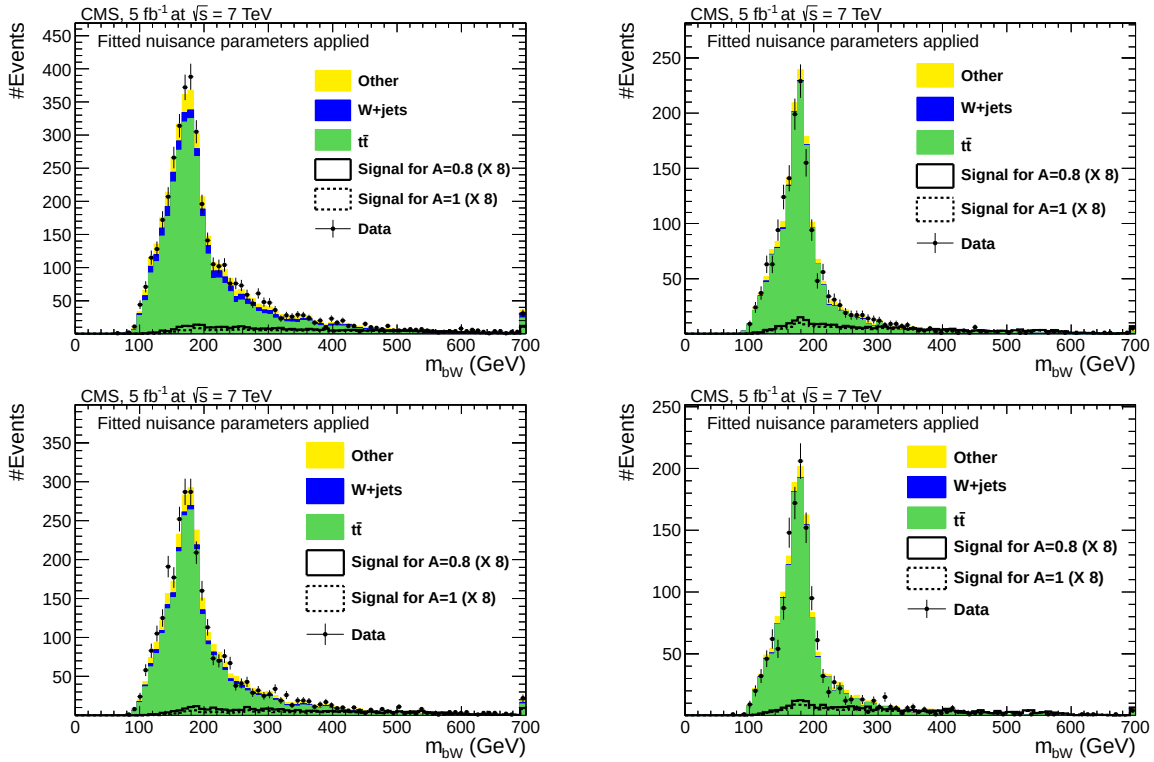

Figure 4.24: The reconstructed m_{bW} distributions in the 1b2W (left) and 2b2W (right) categories in the single-muon (top) and single-electron (bottom) channel, after the inclusion of the fitted nuisance parameters from Table 4.12.

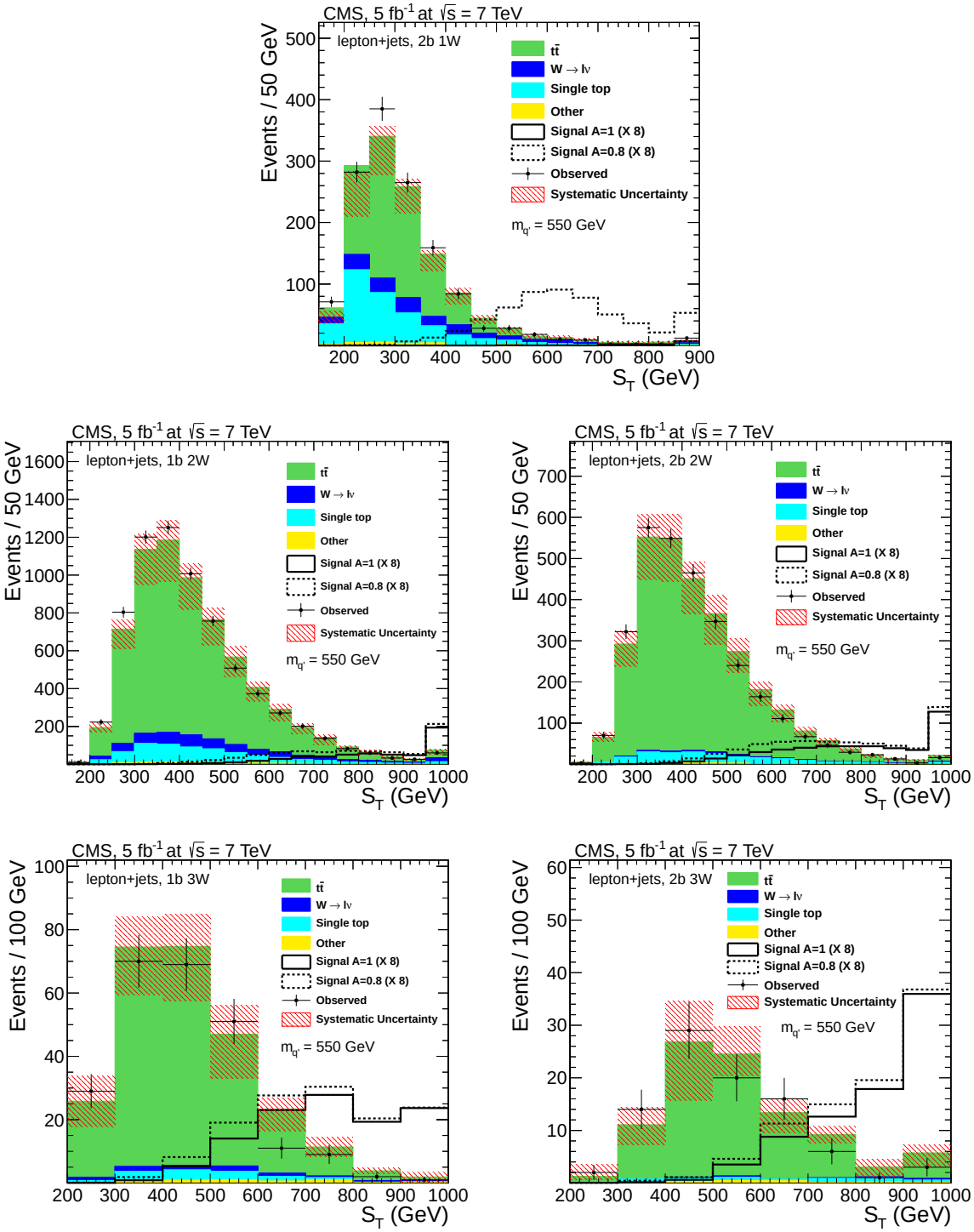

Figure 4.25: The observed and expected distributions of S_T in the relevant combined single-muon and single-electron subsamples agree well, after the inclusion of the fitted nuisance parameters. The distribution for the signal is shown for two different values of the $V_{CKM}^{4\times4}$ parameter A and for b' and t' masses of 550 GeV. In the case that $A = 1$, the $t'b$ and tb' production cross sections are zero and the $t'b'$ cross section is maximal. For decreasing values of A, the $t'b$ and $t'b'$ production cross sections start increasing, while the production cross section of $t'b'$ decreases. The cross section of the signal processes in these distributions is scaled with a factor eight for visibility.

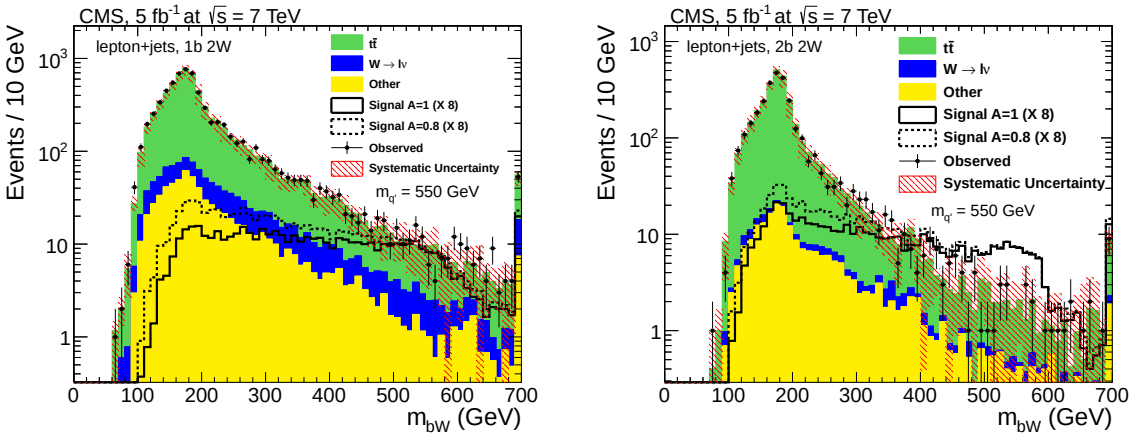

Figure 4.26: The observed and expected distributions agree well for the m_{bW} variable in the subsamples with two W bosons in the combined single-muon and single-electron channel, after the inclusion of the fitted nuisance parameters.

4.7.2 Exclusion limits for degenerate fourth-generation quark masses

Using the probability model (4.7) with the single-muon, single-electron and multilepton channnels, the observed data is compared to the expectation with and without the fourth-generation signal. We use the CLs procedure on the profile likelihood ratio test statistic distributions to extract the 95% CL limits on the combined signal strength μ, for assumed parameters A and $m_{t'} = m_{b'}$. From this, we extract limits on these fourth-generation model parameters themselves.

Figure 4.27 at the left (right) shows the limit obtained using only the single-muon (single-electron) channel seperately. The existence of fourth-generation quarks with degenerate masses is excluded for all parameter values below the black solid line using the assumed model of the $V_{\mathrm{CKM}}^{4\times4}$ matrix. For $A \sim 1$, we obtain an observed (expected) exclusion limit $m_{t'} = m_{b'} > 645$ (605) GeV at the 95% CL using the single-muon channel alone. For $A \sim 1$, the observed (expected) limit obtained in the single-electron channel is 635 (600) GeV at 95% CL. The limit calculation for the same-sign leptons and trilepton channel, using only the event counts is shown in Figure 4.28. For $A \sim 1$, the observed (expected) limit on the fourth generation quarks obtained in these multilepton channels is 595 (625) GeV at 95% CL.

Figure 4.29 shows the exclusion limit obtained when combining the single-muon, single-electron and multilepton channels. Fourth-generation quarks with a degenerate mass below 685 (670) GeV are observed (expected) to be excluded at the 95% CL for a parameter value of $A \sim 1$. It is worth noting that no limits can be set for A exactly equal to unity ($A = 1$), because in this special case the fourth-generation quarks would be stable in the assumed model. Nevertheless, as mentioned in Section 4.1.1, the analysis is valid for values of A extremely close to unity. When the value of the $V_{\mathrm{CKM}}^{4\times4}$ parameter A approaches unity, the SM single-top and the $t'b'$ processes reach their maximal values for the production cross section. When the value of A decreases, the cross section of these processes decreases linearly with A. At the same time the expected

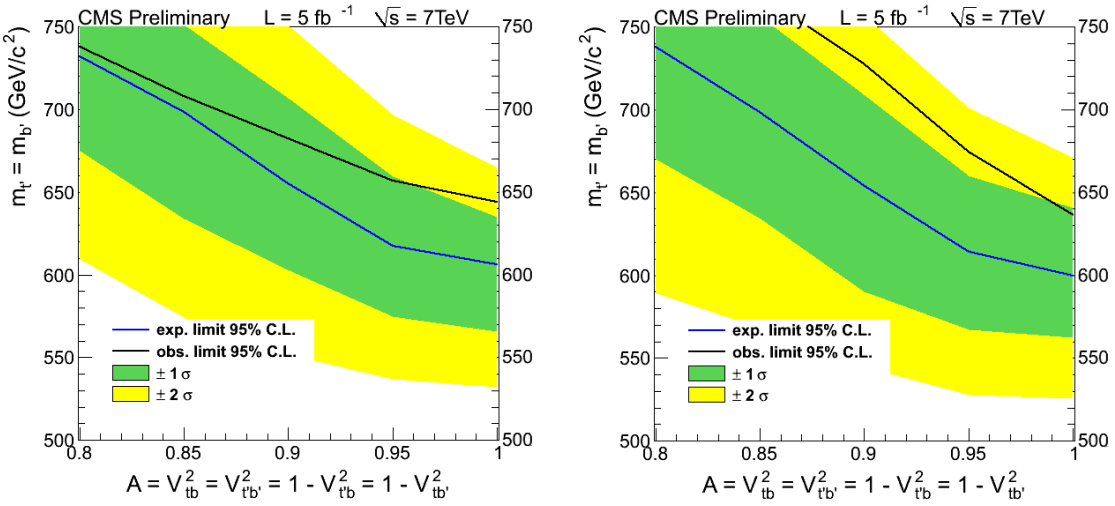

Figure 4.27: Exclusion limit on $m_{t'} = m_{b'}$ as a function of the $V_{CKM}^{4\times4}$ parameter A for the single-muon (left) and single-electron (right) channel. The inner (outer) band indicates the 68% (95%) confidence interval around the expected limit.

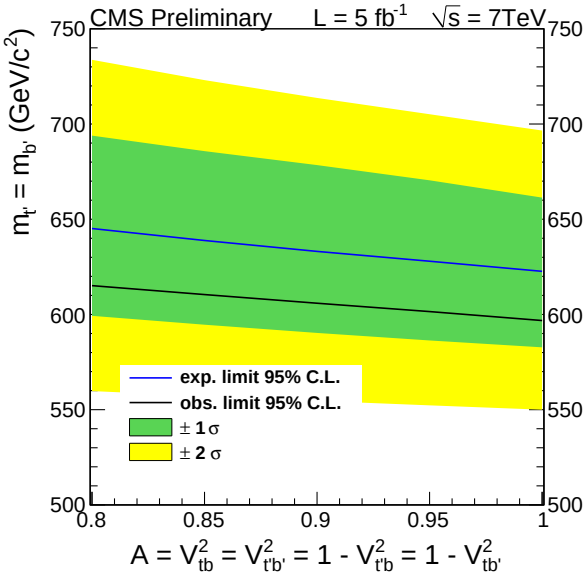

Figure 4.28: Exclusion limit on $m_{t'} = m_{b'}$ as a function of the $V_{CKM}^{4\times4}$ parameter A for the same-sign dilepton and trilepton channel.

cross section of the $t'b$ and tb' processes increases with $(1 - A)$ and is equal to zero for $A = 1$. Therefore, the $t'b$ and tb' processes are expected to enhance the sensitivity for fourth-generation quarks when the parameter A decreases. This tendency is visible in all limit plots where both the expected and observed limits on $m_{Q'}$ are more stringent for smaller values of A. For instance, combining all channels, the limit on the fourth-generation quark masses increases by 70 GeV for $A = 0.9$ compared to the value of the limit for $A \sim 1$. While the $t'b$ and tb' processes do not contribute for $A \sim 1$, the inclusion of the $t'b'$ process results in a more stringent limit (a difference of about 30 GeV) compared to when this process is not taken into account.

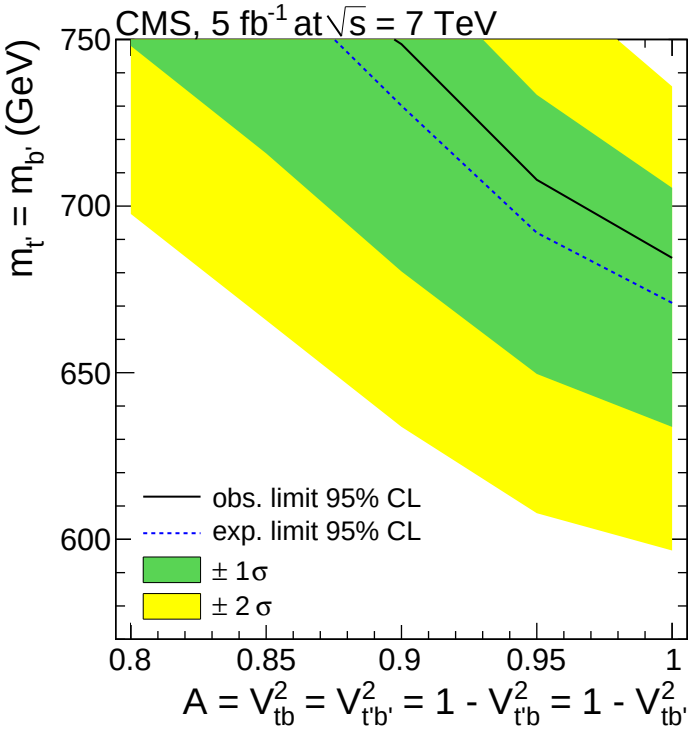

Figure 4.29: Exclusion limit on $m_{t'} = m_{b'}$ as a function of the $V_{\mathrm{CKM}}^{4\times4}$ parameter A. This limit is obtained using all channels together. The parameter values below the solid line are excluded at 95% CL. The inner (outer) band indicates the 68% (95%) confidence interval around the expected limit. The slope indicates the sensitivity of the analysis to the $t'b$ and tb' processes.

4.7.3 Exclusion limits for non-degenerate fourth-generation quark masses

Up to now, the masses of the fourth-generation quarks were assumed to be degenerate. However, if a fourth generation of chiral quarks exists, this is not necessarily the case. As noted in Sections 1.2.3 and 4.1.3, the mass splitting is expected to be smaller than the W-boson mass. Requiring the rate of decays of fourth-generation quarks to each other to be negligible, mass splittings up to 25 GeV can be probed using the same decay chains as for degenerate masses. The sensitivity of the analysis increases or decreases

depending on the specific values of the masses and hence the production cross sections of the fourth-generation quarks. The effect of the mass difference between the fourth-generation quarks on the combined exclusion limit is shown in Figure 4.30 for a $V_{\mathrm{CKM}}^{4\times4}$ parameter $A \sim 1$. For instance in case $m_{t'} = m_{b'} + 25$ GeV ($m_{t'} = m_{b'} - 25$ GeV), the limit on $m_{t'}$ increases about +20 (−20) GeV with respect to the degenerate-mass case. To obtain this limit, we do not take into account the electroweak $t'b'$ process, resulting in more conservative exclusion limits. This is the reason why for $m_{t'} - m_{b'} = 0$ and $A \sim 1$, the limit on $m_{t'}$ is weaker compared to the one shown in Figure 4.29.

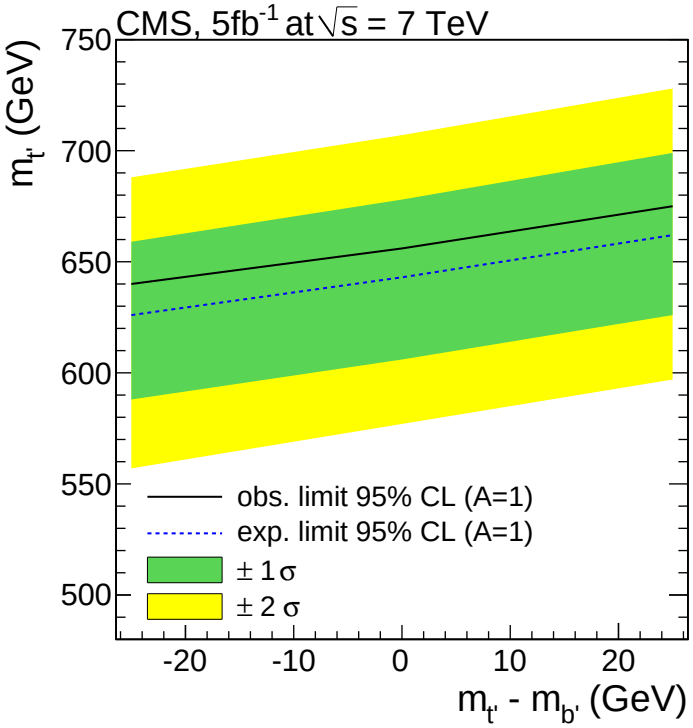

Figure 4.30: For a $V_{\mathrm{CKM}}^{4\times4}$ parameter value $A \sim 1$, the exclusion limit on $m_{t'}$ versus $m_{t'} - m_{b'}$ is shown. The exclusion limit is calculated for mass differences up to 25 GeV. The existence of up-type fourth-generation quarks with mass values below the observed limit are excluded at the 95% CL.

4.7.4 Additional sensitivity tests

As described before, we use two-dimensional distributions of S_T and m_{bW} in the single-lepton channels. The sensitivity when only S_T is used has been checked and we observe that the expected combined exclusion limit at $A \sim 1$ is about 30 GeV lower when using only the S_T variable. This clearly motivates the use of the reconstructed mass of the hadronic bW system as a discriminating variable. We perform several other studies to test the sensitivity of our search. One particularly interesting study is to check the outcome of the statistical procedure when injecting the combined total simulated background and the signal as pseudodata. Another cross check is to test the sensitivity added by each subsample of the single-lepton categories.

Signal injection test

We create pseudo-data constructed from the sum of SM background simulation and a simulated signal process of fourth-generation quarks with a mass of 625 GeV for an A value of 0.8. The input cross section of this signal is set to 1 pb, so numerically it corresponds to the signal strength μ. The signal strength $\hat{\mu}$ we obtain from the fit (i.e. the value of μ that maximizes $L(\hat{\mu}, \hat{\alpha})$ is 1.00 ± 0.20; hence we are able to extract the input cross section. The exclusion limit corresponding to this signal injection test is shown in Figure 4.31. It is clear that the presence of signal in the pseudo-data causes the pseudo-observed exclusion limits to be less stringent.

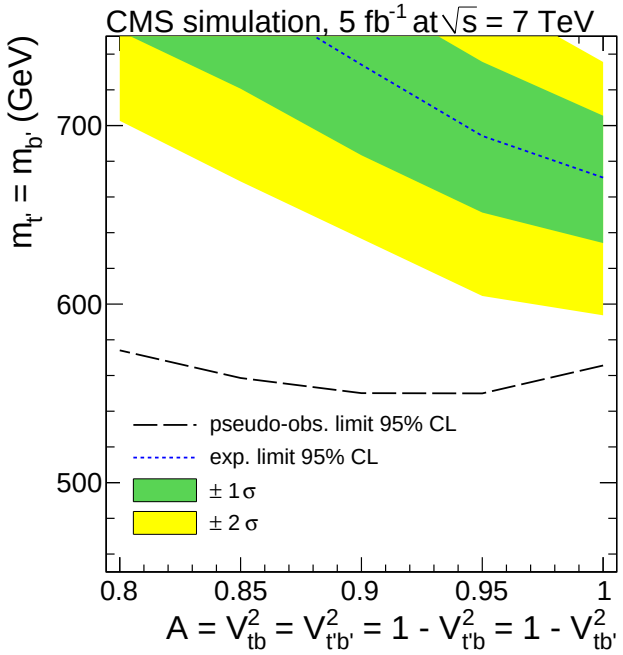

Figure 4.31: The exclusion limits for the test where the SM simulation and an injected fourth-generation signal with a mass of 625 GeV and $A = 0.8$ is treated as pseudo-observed data. The limit corresponding to this pseudo-observed data deviates a lot from the expected exclusion limit, clearly indicating the sensitivity of our search to this signal.

As a second more quantitative test, we perform pseudo-experiments rather than a single injection of signal. We generate these *toy* experiments with a signal of mass of 550 GeV and $A = 0.8$, with a cross section corresponding to 1 pb. Each toy experiment is generated using random numbers drawn from Gaussian distributions for the systematic uncertainties. The statistical effects are taken into account as well, by replacing each bin content with a random number from a Poisson distribution with the bin content as mean. The resulting distribution of the fitted output cross section has a mean of 1 pb, providing more confidence in the fitting procedure, as shown in Figure 4.32. The pull for each toy experiment with output cross section σ_{toy} and statistical uncertainty s_{toy}, is defined as $(\sigma_{toy} - \sigma_{input})/s_{toy}$, where σ_{input} is the input cross section of 1 pb. The resulting pull distribution is centered around zero and has a width close to 1, indicating

that the uncertainties in the fitted cross section are well estimated.

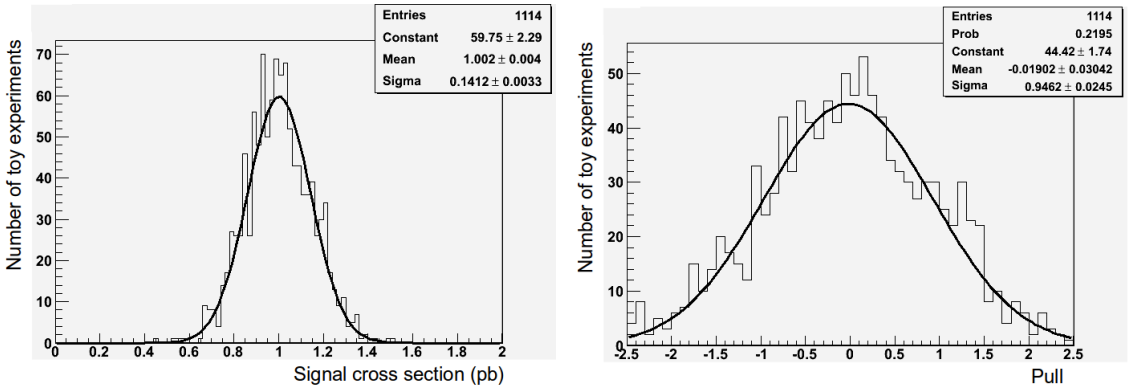

Figure 4.32: The mean of the distribution of the fitted output cross section for 1000 toy experiments agrees with the signal cross section of 1 pb used in the toy experiments (left). The pull distribution, defined in the text, proves the uncertainties in the fitted cross section are well estimated (right).

Sensitivity per single-lepton subsample

We study the effect of the different subsamples on the combined single-muon and single-electron expected exclusion limits, by calculating the limits, each time omitting a different subsample. The exclusion limit used as a reference to which the other limits are compared, considers all subsamples (2b1W, 1b2W, 2b2W, 1b3W, 2b3W, 1b4W and 2b4W). For $A \sim 1$ the expected exclusion limit for degenerate fourth-generation quark masses is 639 GeV, and for $A = 0.9$ this is 710 GeV. The difference of these values with the limits obtained when omitting one specific subsample can be interpreted as the added sensitivity of this subsample. The results of the sensitivity of the subsamples for two values of the A parameter ($A \sim 1$ and $A = 0.9$ are summarized in Table 4.13. The 2b1W is very important for smaller A values, since it is constructed to be sensitive to electroweak single t' and b' production. The 2b2W subsample is most important for the case $A \sim 1$, as it is designed to be particularly sensitive for t' pair production by considering the two-dimensional S_T and m_{bW} distribution.

Table 4.13: The added sensitivity for each subsample, obtained by omitting the subsamples one by one and calculating the full expected single-lepton limits. Both three and four W-boson categories contain a relatively low number of events, so their corresponding subsamples with one and two b jets are taken together.

	2b1W	1b2W	2b2W	1b3W +2b3W	1b4W +2b4W
$A \sim 1$	+5 GeV	+14 GeV	+31 GeV	+15 GeV	+3 GeV
$A \sim 0.9$	+58 GeV	+8 GeV	+10 GeV	+8 GeV	+1 GeV

4.7.5 Summary

We have performed a search for a sequential fourth generation of quarks using the full $\sqrt{s} = 7$ TeV data set collected by the CMS experiment in 2011, corresponding to an integrated luminosity of 5 fb^{-1}. A simple model for a unitary CKM matrix has been defined based on a single parameter $A = |V_{tb}|^2 = |V_{t'b'}|^2$. First degenerate masses have been assumed for the fourth-generation quarks, hence $m_{t'} = m_{b'}$. The information is combined from different subsamples corresponding to different final states with at least one electron or muon. Observables have been constructed in each of the subsamples and are used to differentiate between the Standard-Model background and the processes with fourth-generation quarks. With this strategy the search for electroweakly and strongly produced t' and b' quarks has been combined in a coherent way into a single analysis. Model-dependent limits are derived on the mass of the quarks and the $V_{\mathrm{CKM}}^{4\times4}$ matrix element A. The existence of fourth-generation quarks with masses below 685 GeV is excluded at 95% confidence level for minimal off-diagonal mixing between the third- and the fourth-generation quarks. A non-zero cross section for the single fourth-generation quark production processes, corresponding to a value of the $V_{\mathrm{CKM}}^{4\times4}$ parameter $A < 1$ gives rise to a more stringent limit. When a mass difference of 25 GeV is assumed between t' and b' quarks, the limit on $m_{t'}$ shifts by about +20 (−20) GeV for $m_{b'} = m_{b'} + 25$ GeV ($m_{t'} = m_{b'} - 25$ GeV). These results significantly reduce the allowed parameter space for a fourth generation of fermions and raise the lower limits on the masses of the fourth generation quarks beyond the theoretical unitarity and perturbativity bounds.

Chapter 5

Search for vector-like quarks decaying to light quarks

A chiral fourth generation of quarks has been strongly constrained by direct searches such as the one presented in Chapter 4 as well as by the discovery of the Standard-Model-like Brout-Englert-Higgs boson at the Large Hadron Collider in 2012. Vector-like quarks can evade several of these very tight constraints. They couple differently to the W, Z and H bosons compared to chiral quarks, resulting in a potentially rich phenomenology. Vector-like quarks appear in many new-physics models that try to address open questions in high-energy physics. As they are expected to have masses around the TeV scale, they are well-motivated candidates for the next new particles to be discovered at the LHC.

In this analysis, we search for vector-like quarks of charge $-1/3$ that couple to quarks of the first generation in proton-proton collisions at a center-of-mass energy of 8 TeV collected by the CMS experiment in 2012. The relative branching fractions of vector-like quarks decaying to heavy Standard-Model bosons appear to be different in many models involving these new particles. Therefore an inclusive analysis is set up that is sensitive to many possible decay modes of the vector-like quarks. We consider both single and pair production of these hypothetical quarks, and search for them in final-state topologies with at least one isolated muon or electron. We observe no excess of events in the collected data with respect to the Standard-Model background expectation, hence we set exclusion limits on the parameters of the considered vector-like quark model.

In Section 5.1, we outline the basic considerations on the vector-like quark model in the performed search. The analyzed $\sqrt{s} = 8$ TeV proton collision data set and the relevant simulated Standard-Model background and vector-like quark signal processes are described in Section 5.2. The reconstructed physics objects and the basic selection criteria applied on them are outlined in Section 5.3. Next, in Section 5.4, the data is splitted in categories that are optimized to be sensitive to specific vector-like quark production and decay modes, while in Section 5.5, several data-driven estimations are performed to model the background expectation in these subsamples. The observables that are sensitive to the presence of vector-like quarks are discussed in Section 5.6, and the simultaneous search strategy for the different final states is outlined in Section 5.7. The results of this search are presented and discussed in Section 5.8.

5.1 Vector-like quark model considerations

In this analysis, we start from the generic vector-like quark extension discussed in Section 1.3.2, where the new quarks are assumed to couple to first-generation quarks. Even though simplifying assumptions in an experimental analysis are unavoidable for practical and computational reasons, we aim to be as much as possible model-independent, while maintaining a good sensitivity to the presence of vector-like quarks. Several remarks regarding the free parameters of the considered model are made in Section 5.1.1. The studied production processes are described in Section 5.1.2, and the resulting signal topologies are summarized in Section 5.1.3.

5.1.1 Coupling parameters

From the Lagrangian density describing the interaction between down-type vector-like quarks (with electric charge $-1/3$) and first-generation quarks, Equation (1.38), it is clear that several unknown model parameters appear. For a given quark mass, there are three free parameters, corresponding to the couplings to the three bosons: $\tilde{\kappa}_{uD}$, $\tilde{\kappa}_{dD}$ and $\tilde{\kappa}_{H,dD}$. We will use an equivalent shorthand notation $\tilde{\kappa}_{uD} \equiv \tilde{\kappa}_W$, $\tilde{\kappa}_{dD} \equiv \tilde{\kappa}_Z$ and $\tilde{\kappa}_{H,dD} \equiv \tilde{\kappa}_H$, with the boson subscript indicating the boson involved. We do not know a priori what values these parameters should have, but they may be of order unity. As mentioned in Section 1.3.3, experiments might constrain $\kappa = \frac{v}{\sqrt{2}m_Q}\tilde{\kappa}$ to the percent level, but this still allows for relatively large values of $\tilde{\kappa}$ for high vector-like quark masses. Moreover, cancellations among multiple new quark contributions can relax these constraints, such that values $\tilde{\kappa} = \mathcal{O}(1)$ cannot be incontrovertibly excluded by the available indirect measurements.

We can represent the information from the three $\tilde{\kappa}$ parameters in another equivalent set of parameters that are easier to relate to phenomenological observables:

$$\tilde{\kappa}_W, \quad \mathrm{BF}_W, \quad \mathrm{BF}_Z, \tag{5.1}$$

with $\mathrm{BF}_W = \mathrm{BF}(Q \to qW)$ the branching fraction for the decay of the heavy quark to a W boson, and $\mathrm{BF}_Z = \mathrm{BF}(Q \to qZ)$ the branching fraction for the decay of the heavy quark to a Z boson. For given non-zero values of these three parameters, both $\mathrm{BF}_H = \mathrm{BF}(Q \to qH)$ (or $\tilde{\kappa}_H$) and $\tilde{\kappa}_Z$ are uniquely determined. The first just follows from the fact that the branching fractions should add up to 1 (see Equation (1.44)). For the second statement, we use the calculations of the partial widths from [62] to derive an analytical relation. When setting the SM-quark masses to zero, and this is a good approximation for first-generation quarks, it can be calculated that the partial widths $\Gamma(Q \to qV)$, with V either a W, Z or H boson, are directly proportional to the 'kinematic functions' Γ_V^0:

$$\Gamma(Q \to qV) \approx \kappa_V^2 \frac{M_Q^3 g^2}{64\pi m_W^2} \Gamma_V^0 \tag{5.2}$$

where the kinematic functions are given by

$$\Gamma^0_W \approx 1 + \mathcal{O}(M_Q^{-4}) \tag{5.3}$$

$$\Gamma^0_Z \approx \frac{1}{2} + \mathcal{O}(M_Q^{-4}) \tag{5.4}$$

$$\Gamma^0_H \approx \frac{1}{2} - \frac{m_H^2}{M_Q^2} + \mathcal{O}(M_Q^{-4}). \tag{5.5}$$

The branching fractions of decays to a boson V can then be calculated as

$$\mathrm{BF}(Q \to qV) = \frac{\tilde{\kappa}_V^2 \Gamma^0_V}{\sum_{V'=W,Z,H} \tilde{\kappa}_{V'}^2 \Gamma^0_{V'}}, \tag{5.6}$$

when considering only decays to the first generation. Ignoring terms of $\mathcal{O}(M_Q^{-4})$, the following relations are obtained:

$$\mathrm{BF}_W = \frac{\tilde{\kappa}_W^2}{\tilde{\kappa}_W^2 + \frac{1}{2}\tilde{\kappa}_Z^2 + (\frac{1}{2} - \frac{m_H^2}{M_Q^2})\tilde{\kappa}_H^2} \tag{5.7}$$

$$\mathrm{BF}_Z = \frac{\frac{1}{2}\tilde{\kappa}_Z^2}{\tilde{\kappa}_W^2 + \frac{1}{2}\tilde{\kappa}_Z^2 + (\frac{1}{2} - \frac{m_H^2}{M_Q^2})\tilde{\kappa}_H^2}, \tag{5.8}$$

and consequently, if $\tilde{\kappa}_W \neq 0$:

$$\tilde{\kappa}_Z = \sqrt{2\frac{\mathrm{BF}_Z}{\mathrm{BF}_W}}\tilde{\kappa}_W. \tag{5.9}$$

The relation (5.9) means that, for an assumed ratio $\frac{\mathrm{BF}_Z}{\mathrm{BF}_W}$ of branching fractions, and assuming a value of the strength of the single vector-like quark production via the charged current interaction (proportional to $\tilde{\kappa}_W^2$), the value of $\tilde{\kappa}_Z$ and hence the single-production signal strength via the neutral-current interaction can be calculated analytically. Performing a three-dimensional scan over the parameters (5.1) is therefore sufficient to probe the coupling parameters in a model-independent way, as long as $\tilde{\kappa}_W \neq 0$.

Caution is needed for the extreme cases, e.g. when $\tilde{\kappa}_W = 0$. This implies $\mathrm{BF}_W = 0$, and Equation (5.9) is not applicable. Similarly, $\mathrm{BF}_W = 0$ would imply $\tilde{\kappa}_W = 0$. In such a case, the coupling parameter space is essentially reduced to two dimensions, and one should probe the parameters $\tilde{\kappa}_Z$ and BF_Z in a two-dimensional scan.

For simplicity, we assume only right-handed couplings of vector-like quarks to first-generation quarks in this analysis. As mentioned before, it is often expected that one of the two chiral couplings is suppressed by a factor m_q/M_Q with respect to the other coupling [60]. We do not explicitly exploit angular correlations between final-state objects, hence this assumption should not be a limiting case, but should still be taken with care.

5.1.2 Production processes

We consider both electroweak single production and strong pair production of vector-like D quarks, without assuming a specific underlying $SU(2)$ multiplet structure to

which this hypothetical quark may belong. This means the analysis is not optimized for a combined search for all quarks in a given multiplet, contrary to the analysis presented in Chapter 4. Therefore exclusion limits in this analyis are expected to be more conservative than could be obtained in a dedicated model-dependent search combining the signal from D and U quarks. On the other hand, this allows for a more model-independent interpretation.

The Feynman diagrams for the single production modes we consider have been shown in Figure 1.8. If the vector-like quark with a given mass is produced via a charged-current or neutral-current interaction, the cross section of this process scales with $\tilde{\kappa}_W^2$ or $\tilde{\kappa}_Z^2$, respectively. We assume the production mechanisms involving the exchange of a H boson to be negligible (even for potentially sizable $\tilde{\kappa}_H^2$), since these are suppressed by the Yukawa couplings of light quarks, which are very small. Since quantum-mechanical interference effects of the charged-current and neutral-current production process amplitudes are expected to be small [62], the total single-production cross section σ^Q of a vector-like quark Q would therefore be the sum of the cross sections of both possible production mechanisms:

$$\sigma^Q \equiv \sigma(pp \to Qq) = \sigma(pp \xrightarrow{W^*} Qq) + \sigma(pp \xrightarrow{Z^*} Qq) \tag{5.10}$$

$$\equiv \sigma^Q_{CC} + \sigma^Q_{NC} \tag{5.11}$$

These cross sections can even be large when the $\tilde{\kappa}$ coupling strengths are small, because the partons in the initial state are valence quarks of the proton. Anti-D-quark production is considered as well.

For the pair production we only consider the model-independent strong production, where the cross sections $\sigma^{Q\overline{Q}} \equiv \sigma(pp \to Q\overline{Q})$ are governed by well-known perturbative quantum chromodynamics. It is important to note that in principle, pair production of vector-like quarks coupling to first-generation quarks can also occur through the t-channel exchange of a H boson. The corresponding Feynman diagram would contain two vertices related to $\tilde{\kappa}_H^2$. Ignoring this process with respect to QCD pair production assumes relatively small values of $\tilde{\kappa}_H$. Nevertheless, the most important consequence of ignoring this H-boson exchange production process is that the derived exclusion limits on the vector-like quark parameters will only be more conservative.

Previous searches for vector-like quarks coupling to first-generation quarks only considered electroweak production modes [72, 167, 168]. Hence, exclusion limits on the masses can be avoided if one allows for small enough values of the $\tilde{\kappa}$ coupling parameters. By considering the strong pair production, one would be able to exclude vector-like quark masses independent of $\tilde{\kappa}$. Clearly, a search involving the *combination* of single and pair production processes, which is the main aim of this analysis, will still result in exclusion limits depending on $\tilde{\kappa}$, as long as the contribution of single production is not negligible.

5.1.3 Signal topologies

Because of their mixing properties with SM quarks, vector-like quarks could decay in various modes. These decays to W, Z and H bosons have been summarized for vector-like D quarks in Equation (1.43). Given the pair and single production modes,

Table 5.1: Due to the pair and single production modes that should be present if vector-like quarks Q exist, and the various possible decay modes, the hypothetical vector-like quark processes involve many distinct final-state topologies. The processes indicated by (†) are not considered in the analysis, as the probability for them to result in at least one isolated lepton is low.

Production	Topology
Single (electroweak)	$Qq \to Wqq$
	$Qq \to Zqq$
	$Qq \to Hqq$ (†)
Pair (strong)	$Q\overline{Q} \to WqWq$
	$Q\overline{Q} \to WqZq$
	$Q\overline{Q} \to WqHq$
	$Q\overline{Q} \to ZqZq$
	$Q\overline{Q} \to ZqHq$
	$Q\overline{Q} \to HqHq$ (†)

we expect the signal topologies that are listed in Table 5.1. The W and Z bosons can either decay to jets or leptons. The H boson has many possible decay modes, with the decay to two b quarks having the largest branching fraction (about 58% for $m_H = 125$ GeV [169]). We will require at least one lepton in the final state, and since the branching fraction of the H boson to leptons is relatively low, we omit the decay of a singly produced quark $Qq \to Hqq$ and the decay of pair-produced quarks $Q\overline{Q} \to HqHq$ in the analysis. Exclusive event categories will be reconstructed according to the number of observed isolated leptons as well as some additional selection criteria. With this strategy, we exploit the topology and kinematic properties of the different final-state modes and optimize the sensitivity to the presence of vector-like quarks. Contrary to chiral fourth-generation quarks where jets originating from b quarks are expected in the final state, the topologies of vector-like quarks coupling to first-generation quarks do not in general contain b jets. Only in the case of a significant decay branching fraction for $Q \to Hq$, the final state will contain b jets originating from the decay of the H boson.

5.2 Data and simulation

In this search for vector-like quarks we use the proton-proton collisions at $\sqrt{s} = 8$ TeV collected by the CMS experiment during the 2012 LHC run. The full data set corresponds to a total integrated luminosity of (19.6 ± 0.5) fb^{-1}. The data has been certified to ensure the good quality for physics analysis, and filters are applied to reject events where electronic noise or proton-beam backgrounds mimic energy deposits in the detector. The Standard-Model processes that constitute the main backgrounds in this search for new quarks are discussed in Sections 5.2.1 and the simulated vector-like quark processes in Section 5.2.2. Events for the background and signal processes are

generated with the same generators as in Chapter 4, but at a different center-of-mass energy. The high-level trigger requirements that are applied to collect the data are specified in Section 5.2.3.

5.2.1 Standard-Model background simulation

Many of the physics processes that constitute the SM background in this analysis are similar to those in the combined search for fourth-generation quarks presented in Chapter 4. However, due to the applied event selection criteria their relative contributions will differ.

The POWHEG 1.0 generator is used to generate the single-top processes, PYTHIA 6.4.24 for the diboson WW, WZ and ZZ processes, and MadGraph for the other SM processes. The CTEQ6M parton distribution functions (PDFs) are used in POWHEG, and for the other generators the CTEQ6L1 PDF set is used [105]. The hadronization and parton showering are simulated with PYTHIA with the underlying-event tune Z2*. Next, the MLM matching algorithm is used for the matching of matrix-element partons to the parton showers. Additional minimum-bias events generated with PYTHIA are superimposed to all generated hard-scattering events to simulate the pileup effect of multiple collisions in the same bunch crossing. The generated SM-background events are finally processed through the GEANT4 CMS detector simulation.

All SM processes with the corresponding cross sections σ and equivalent integrated luminosities $\mathcal{L}_{eq}$ of the generated samples are shown in Table 5.2. We use generated samples in jet bins for the production processes of a W or Z boson in association with jets. For these processes, the exclusive cross section in each bin of the jet multiplicities are obtained by scaling the inclusive NNLO cross-section, calculated with the FEWZ generator [170], by the ratio of the LO cross section of the exclusive and inclusive production processes. These cross sections are only used as a starting point, however, because the actual normalizations will be derived from the observed data. The cross section of the top-quark pair production corresponds to the measurement by the CMS experiment [171], multiplied by the measured branching fractions of the top quark. The single-top production cross section values are obtained from approximate NNLO calculations [172]. For the WZ and ZZ diboson processes and $t\bar{t}$ in association with a W or Z boson, the NLO cross sections are calculated with the MCFM generator. For the WW diboson process, the CMS measured cross section value is used. The triboson NLO production cross sections are obtained with the aMC@NLO generator. In the generated triboson samples the off-shell photon (γ^*) process is not generated, so we will consider a large uncertainty on these cross sections. Note that the triboson cross sections are extremely small, of the order of tens of femtobarns. Finally, the LO cross section of the small background coming from two same-sign W bosons is obtained by the MadGraph generator.

5.2.2 Vector-like quark signal simulation

All vector-like quark processes considered in the analysis are generated with MadGraph 5.1.5, with the implemented model Lagrangian density of [63] and using the CTEQ6L1 PDF set. Vector-like quark events are simulated for charged-current and

Table 5.2: Overview of the cross sections σ for the relevant Standard-Model background processes in the vector-like quark search.

Process	Generator	σ (pb) at 8 TeV	$\mathcal{L}_{eq}$ (fb^{-1})
W+1jet $\to \ell\nu$	MadGraph	6520.0 (scaled to NLO)	3.5
W+2jets $\to \ell\nu$	MadGraph	2113.0 (scaled to NLO)	16.1
W+3jets $\to \ell\nu$	MadGraph	626.6 (scaled to NLO)	24.7
W+4jets $\to \ell\nu$	MadGraph	258.4 (scaled to NLO)	51.3
Z/γ^*+1jet $\to \ell\ell$ ($m_{\ell\ell} > 50$ GeV)	MadGraph	671.7 (scaled to NLO)	35.7
Z/γ^*+2jets $\to \ell\ell$ ($m_{\ell\ell} > 50$ GeV)	MadGraph	216.7 (scaled to NLO)	10.6
Z/γ^*+3jets $\to \ell\ell$ ($m_{\ell\ell} > 50$ GeV)	MadGraph	61.2 (scaled to NLO)	173.6
Z/γ^*+4jets $\to \ell\ell$ ($m_{\ell\ell} > 50$ GeV)	MadGraph	27.6 (scaled to NLO)	226.3
$t\bar{t}$+jets (semi-lept.)	MadGraph	100.1	168.8
$t\bar{t}$+jets (full-lept.)	MadGraph	24.0	396.7
$t\bar{t}$+jets (full-hadr.)	MadGraph	104.4	157.6
t (t channel)	Powheg	56.4 (approx. NNLO)	66.0
$\bar{t}$ (t channel)	Powheg	30.7 (approx. NNLO)	62.1
t (tW channel)	Powheg	11.1 (approx. NNLO)	44.5
$\bar{t}$ (tW channel)	Powheg	11.1 (approx. NNLO)	44.5
t (s channel)	Powheg	3.79 (approx. NNLO)	68.6
$\bar{t}$ (s channel)	Powheg	1.76 (approx. NNLO)	79.5
WW	Pythia	69.9	143.1
WZ	Pythia	33.2 (NLO)	301.1
ZZ	Pythia	17.7 (NLO)	548.4
$t\bar{t}$+W	MadGraph	0.23 (NLO)	845.0
$t\bar{t}$+Z	MadGraph	0.21 (NLO)	1021.7
W^+W^+	MadGraph	0.25 (LO)	403.2
W^-W^-	MadGraph	0.089 (LO)	1084.3
WWW	MadGraph	0.081 (NLO)	2737.0
WWZ	MadGraph	0.058 (NLO)	3834.9
WZZ	MadGraph	0.020 (NLO)	11170.5
ZZZ	MadGraph	0.0055 (NLO)	40691.9

neutral-current electroweak single production, and strong pair production. In all these cases, the vector-like quarks are decayed in MadGraph to a W, Z or H boson in association with a first-generation quark. The bosons are subsequently decayed in pythia via all possible decay modes, after which the hadronization and parton showering is done. The H boson is generated with a mass of 125 GeV, and with the corresponding theoretically predicted decay modes [169]. The widths of the new quarks are set to 1% of the mass, which is an approximation, but due to the limited detector resolution we do not expect to be sensitive to this. The CMS detector response of the generated vector-like quark signal events is simulated using the CMS FastSim framework (see Section 3.2), that uses parametrizations and simplifications in order to speed up the event simulation.

Table 5.3: Overview of the $Q\overline{Q}$ production cross section (approximate NNLO) for different assumed heavy quark masses [155].

m_Q (GeV)	$\sigma^{Q\overline{Q}}$ (fb) at 8 TeV
500	570
550	306
600	170
650	97.1
700	56.9
750	34.1
800	20.8
850	12.9
900	8.1
950	5.1
1000	3.3

Table 5.3 shows the approximate NNLO cross section of the pair production of heavy quarks [155].[1] As these are QCD calculations, they do not depend on the model parameters $\tilde{\kappa}$. Separate samples have been generated for every pair-production topology in Table 5.1 (except $Q\overline{Q} \to HqHq$). The equivalent integrated luminosity of these samples, defined as the number of generated events divided by the cross section, is at least 170 fb^{-1}, corresponding to about 9 times the integrated luminosity of the collected data at 8 TeV. For quark masses above 1 TeV, the cross section drops down to the femtobarn level, making the pair-produced signal practically impossible to observe with about 20 fb^{-1} of data. Therefore, samples above 1 TeV have not been generated.

The LO charged-current and neutral-current single-production cross sections obtained with MadGraph are listed in Table 5.4 for vector-like D quarks.[2] For each production mode in this table, separate samples where generated for the decay modes $Q \to Wq$ and $Q \to Zq$. The equivalent luminosity of each sample corresponds to at least 100 fb^{-1}, which is about 5 times the integrated luminosity of the collected data at 8 TeV. The charged-current production cross section is larger than the neutral-current production cross section. This is due to the fact that for the charged-current (neutral-current) single production of down-type vector-like quarks, an up quark (down quark) is required in the initial state. Since the colliding protons consist of two valence up quarks and only one valence down quark, the probability to produce a D quark via W-boson exchange is higher than via Z-boson exchange. Another reason for the difference in size of the cross sections is that the Z-boson mass is heavier compared to the

[1] It should be noted that the cross section for strongly produced fourth-generation and vector-like quarks is the same assuming a certain quark mass.

[2] Note that the maximal single-production cross section of these vector-like quarks is larger than that of sequential fourth-generation quarks. In the former case, only first-generation quarks are required in the initial state, while in the latter case at least a third-generation quark should be present in the initial state.

Table 5.4: Overview of the LO cross section for the charged-current (σ^D_{CC}) and neutral-current (σ^D_{NC}) electroweak single production for the different assumed vector-like D quark masses. The listed cross sections are for $\tilde{\kappa}_W = 1$ or $\tilde{\kappa}_Z = 1$, and should be scaled with $\tilde{\kappa}^2$. All cross sections are the sum of the cross sections for the heavy particle and the anti-particle.

m_Q (GeV)	σ^D_{CC} (pb) at 8 TeV	σ^D_{NC} (pb) at 8 TeV
500	12.38	3.611
550	8.401	2.445
600	5.851	1.711
650	4.17	1.209
700	3.025	0.8648
750	2.205	0.6325
800	1.67	0.47
850	1.248	0.3545
900	0.9555	0.2676
950	0.7369	0.2049
1000	0.5708	0.159
1050	0.4459	0.1232
1100	0.3515	0.09749
1150	0.2789	0.07642
1200	0.224	0.06055
1250	0.1781	0.04832
1300	0.1432	0.03872
1350	0.1158	0.03101
1400	0.09397	0.0252
1450	0.07643	0.0202
1500	0.06247	0.01654
1550	0.05124	0.01342
1600	0.04194	0.01095
1650	0.03427	0.008979
1700	0.02855	0.007398
1750	0.02347	0.006036
1800	0.01952	0.005025

W-boson mass. Note that the antiparticles of the heavy quarks require an anti-quark in the initial state. In a proton these are only present as sea quarks from gluon splitting, not as valence quarks, hence the corresponding cross sections are smaller.

5.2.3 Trigger requirements in data and simulation

The trigger preselection applied to the data and simulated events require the presence of at least one isolated muon or electron. The trigger-level muon is required to have a

transverse momentum exceeding 24 GeV and an absolute value of the pseudorapidity smaller than 2.1. The p_T threshold on the trigger-level electron is 27 GeV. Since the trigger efficiencies in data and simulation have been observed to differ at the percent level, we apply p_T and η dependent scale factors to the simulated events. These scale factors have been determined with tag-and-probe techniques [123], where a $Z \to \ell\ell$ enriched data sample is used to probe the trigger efficiency of one of the two leptons (i.e. the probe) given a triggered lepton (the tag).

5.3 Baseline physics-object selection

Various SM background processes can mimic the signatures of vector-like quarks. A set of optimized selection criteria on the physics objects in the final state is crucial to suppress this background and enhance the purity of potential signal events. In Sections 5.3.1 and 5.3.2, the basic selection criteria imposed on the lepton and jet candidates in the event are outlined. The refined selection in Section 5.4 aim to increase the sensitivity to the presence of vector-like quarks and are based on these initial lepton and jet selection criteria. The missing transverse energy $\not{E}_T$ is reconstructed and corrected similarly as described in Section 3.3.6, but we do not require a minimal threshold on the missing transverse energy in the baseline selection. The reason is that some of the vector-like quark signatures that we search for do not result in large missing transverse energy.

The physics objects in the final state are reconstructed using the Particle-Flow (PF) algorithm described in Section 3.3.1, except the electron candidates. An identification efficiency loss for PF electrons has been observed with respect to an ECAL-driven electron reconstruction, due to the high pile-up activity in the 8 TeV proton collision run. Therefore, the ECAL-driven reconstruction as outlined in Section 3.3.3 has been used to reconstruct electron candidates in this analysis. A good main primary vertex is required in the event, and the contribution from charged hadrons not matched with the main primary vertex is subtracted from the physics objects. The simulated events are reweighted to match the observed distribution of the number of simultaneous proton interactions. For the full dataset collected in 2012, we observe on average about 21 interactions in each event.

5.3.1 Lepton selection criteria

We define tight and loose charged-lepton collections, with the requirements defining them summarized in Table 5.5. The tight-lepton collection is a subset of the loose collection. In the analysis, events with at least one tight muon or electron are selected, and the loose-lepton collections are constructed with the aim of vetoing additional leptons in the event and to estimate from the observed data the fake-lepton contribution to events with three isolated leptons.

The tight-muon collection is defined as containing muons with $p_T > 20$ GeV, $|\eta| < 2.1$ and $\Delta\beta$-corrected relative isolation $I_{rel}^{\Delta\beta} < 0.12$ (calculated in a cone of $\Delta R = 0.4$ around the muon), while muon candidates in the loose-muon collection are defined by the looser requirements $p_T > 10$ GeV, $|\eta| < 2.5$ and relative isolation $I_{rel}^{\Delta\beta} < 0.2$.

Table 5.5: An overview of the basic selection requirements defining the collections of leptons used in the vector-like quark search.

Muons		Electrons	
Tight	Loose	Tight	Loose
$p_T > 20$ GeV	$p_T > 10$ GeV	$p_T > 20$ GeV	$p_T > 15$ GeV
$\|\eta\| < 2.1$	$\|\eta\| < 2.5$	$\|\eta\| < 2.5$	$\|\eta\| < 2.5$
$I_{rel}^{\Delta\beta} < 0.12$	$I_{rel}^{\Delta\beta} < 0.2$	$I_{rel}^{EA} < 0.1$	$I_{rel}^{EA} < 0.2$
		$mvaId > 0.5$	$mvaId > 0$

Here the relative isolation is calculated according to Equation (3.3). Tight electrons are required to have $p_T > 20$ GeV, $|\eta| < 2.5$ and effective-area corrected relative isolation $I_{rel}^{EA} < 0.1$ (calculated in a cone of $\Delta R = 0.3$ around the electron). The $mvaId$ variable, a quantity related to the electron identification as explained in the identification paragraph of Section 3.3.3, is required to be larger than 0.5, indicating a well identified electron. Loose electrons are defined as electron candidates satisfying $p_T > 15$ GeV, $|\eta| < 2.5$ and relative isolation $I_{rel}^{EA} < 0.2$. The $mvaId$ variable for loose electrons is relaxed to exceed 0. Electron candidates in the transition region between the barrel and endcap ($1.4442 < |\eta| < 1.566$) of the electromagnetic calorimeter are excluded. The relative isolation of all lepton candidates are calculated within a cone of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.4$ around the lepton.

Scale factors are applied to the simulated events to account for differences in lepton trigger, identification and isolation efficiency in the simulation with respect to the observed data. In events with only one isolated lepton, this scale factor is the multiplication of the trigger, identification and isolation efficiency scale factors, and in multilepton events we apply for each additional isolated lepton the scale factor accounting for the identification and isolation efficiency. These factors depend on the transverse momentum and the pseudorapidity of the leptons.

Figure 5.1 and 5.2 show example distributions of kinematic properties of selected muons and electrons, respectively. The transverse momentum and pseudorapidity is compared between the observed data and the simulation for events with two opposite-sign leptons. Here the leading lepton is required to have a transverse momentum above 30 GeV, and the subleading lepton should have a transverse momentum exceeding 20 GeV. Events with additional loose muons and electrons are vetoed. In addition, the invariant mass of the two leptons is required to exceed 50 GeV, since the simulated Z+jets events are generated with this threshold. We also require the presence of at least four selected jets. These selections are merely intended to select events in order to validate the modeling of the kinematic properties of leptons, and are not part of the baseline selection for the rest of the analysis. The QCD multijet background in these plots is estimated from data, as described in Section 5.5, but the uncertainty on this estimation is large.

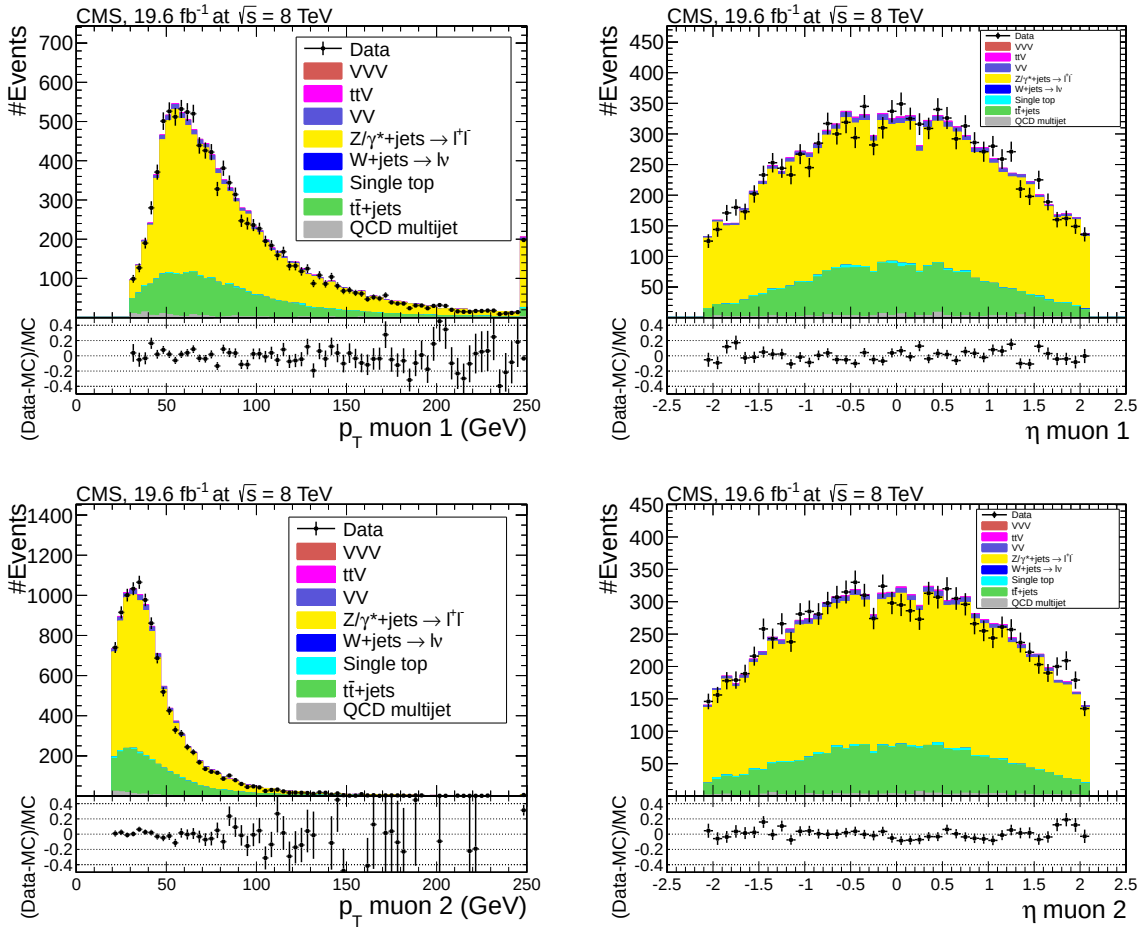

Figure 5.1: Control plots for dimuon events; the p_T (left) and η (right) of the highest-p_T muon (top), and the p_T and η of the second-highest-p_T muon (bottom). The total simulation is normalized to the observed number of data events. The shapes of the data and simulation distributions agree reasonably well.

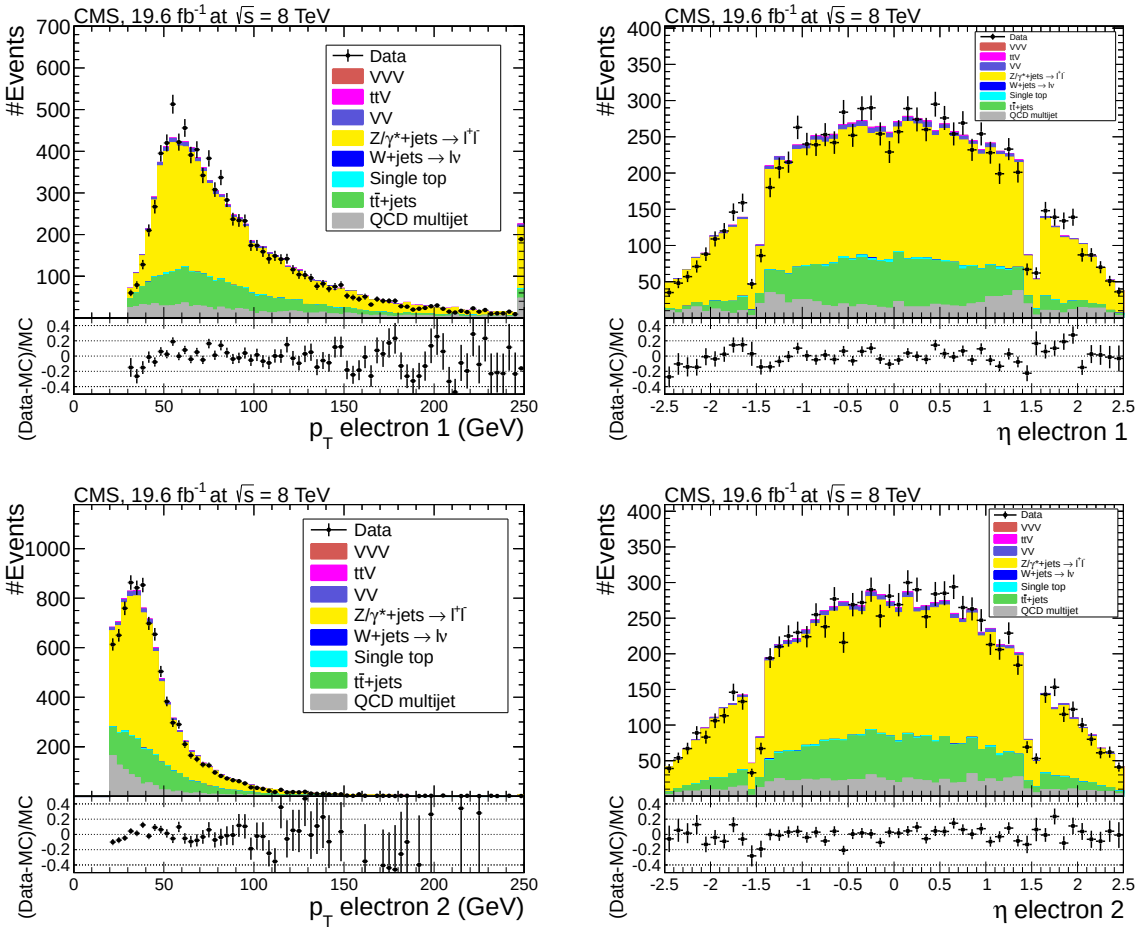

Figure 5.2: Control plots for dielectron events; the p_T (left) and η (right) of the highest-p_T electron (top), and the p_T and η of the second-highest-p_T electron (bottom). The shapes of the data and simulation distributions agree reasonably well.

5.3.2 Jet selection criteria

Two collections of jets are defined in this analysis. For both jet collections, the jets need to pass the jet-identification criteria and have a transverse momentum exceeding 30 GeV. The first collection of jets consists of all jets with $|\eta| < 2.4$ and will be later on referred to as the *selected jets*, while the second collection contains the jets with $2.4 < |\eta| < 5.0$ and will be called *selected forward jets*.

Jet energy corrections are applied to the jets as well as a smearing of the energy resolution of simulated jets to match the resolution in data [137], as described in Section 3.3.4. To validate the modeling of the jet kinematics in simulated events, we compare data and simulation in a $t\bar{t}$ enriched region. Here we select events with exactly one isolated muon with $p_T > 30$ GeV and with $|\eta| < 2.1$, and veto for additional loose muons and electrons. We require the presence of at least four selected jets, and a missing transverse energy exceeding 60 GeV. The transverse momentum and the pseudorapidity of the leading and subleading jets are shown in Figure 5.3. The QCD multijet background is estimated from data, as described in Section 5.5. In these plots and in the rest of the analysis, we apply a reweighting of the simulated spectrum of the transverse momentum of the top quark, because in dedicated studies the spectrum has been observed to be harder in the MadGraph simulation than in the data [173].[3]

Identification of b jets

As mentioned before, we do not expect in general many b jets in the final state of the considered vector-like quark signal topologies. Nevertheless, we will use a b jet identification algorithm, the Combined Secondary Vertex (CSV) algorithm, to distinguish b jets from non-b jets. As explained in Section 3.3.5, the CSV b-tagging algorithm uses a likelihood-ratio technique to combine the information of properties of reconstructed secondary vertices and tracks with a high impact parameter with respect to the main primary vertex. Higher values of the likelihood-ratio are more consistent with jets originating from b quarks. Hence, by imposing a threshold on the resulting CSV discriminator, which has a value between 0 and 1, one can identify b jets with a certain efficiency.

The CSV b-tagging discrimator of the leading jet is shown in Figure 5.4. It can be seen that the CSV discriminator exhibits a shape difference between the observed and expected distributions. This feature is also observed in [145] and is the reason we apply scale factors to simulated events to account for differences in efficiency when imposing a threshold on the CSV discriminator. The procedure of the application of these calibration scale factors is explained in more detail below. The discrepancy between the shape of the observed and simulated CSV distribution can be partly attributed to the different calibration of the alignment of the tracker modules used in data and simulation. Since the CSV algorithm exploits various properties of individual tracks,

[3] The empirically determined weight applied on simulated top-quark pair events is $\sqrt{\exp(0.156 - 0.00137 \times p_T^{\text{top}}) \times \exp(0.156 - 0.00137 \times p_T^{\text{antitop}})}$, with the transverse momenta expressed in GeV.

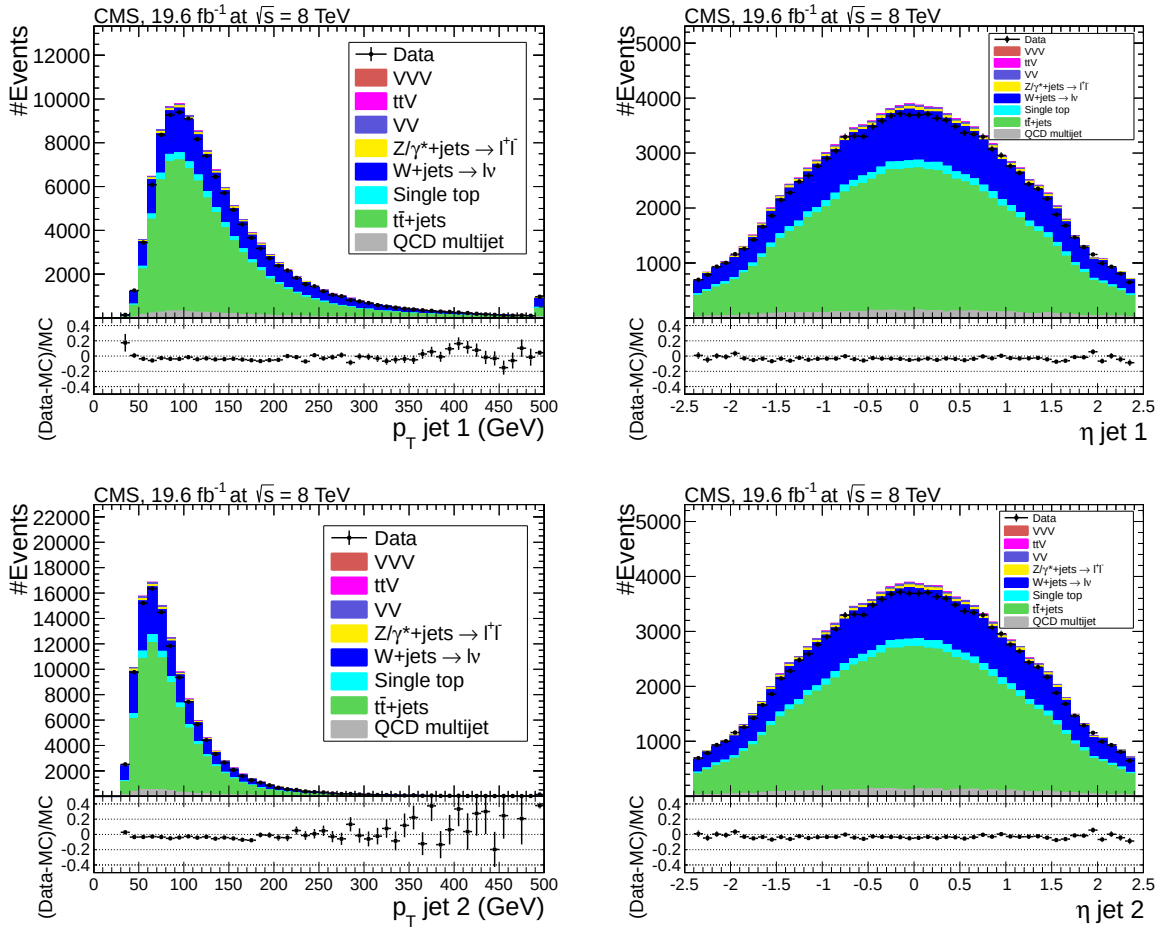

Figure 5.3: Control plots for the jets in single-muon events: the p_T (left) and η (right) of the leading (top) to subleading (bottom) jet in the event. The observed data and simulation agree reasonably well.

this algorithm is expected to be particularly sensitive to these differences in alignment.

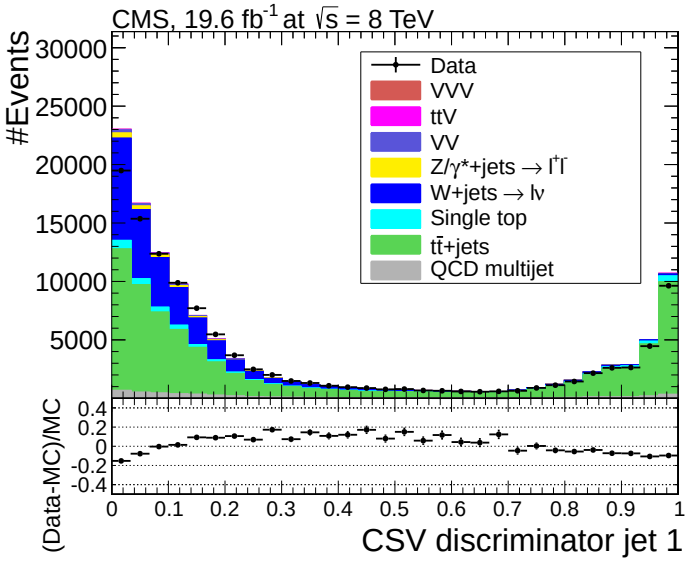

Figure 5.4: The shape of the distributions in the Combined Secondary Vertex discriminator for observed data and simulation tends to be more peaked to the extremes in simulation. For this reason we apply scale factors to simulated events with a certain configuration of b-tagged and anti-b-tagged jets.

In most of the event categories defined in Section 5.4, we impose a veto for b jets, a procedure that could be referred to as *anti-b-tagging* of jets. To get an efficient rejection of b jets, we require the CSV discriminator to be *smaller* than a certain loose CSV threshold. In the event categories in which we aim to be sensitive to H-boson decays, we require at least one b jet, hence for these jets we require the CSV discriminator to be *larger* than a reasonably tight CSV threshold. More concretely we impose the following b-jet identification criteria in the analysis:

$$\text{anti-}b\text{-tagged jet:} \quad 0 \leq \text{CSV discriminator} < 0.244 \tag{5.12}$$

$$b\text{-tagged jet:} \quad 0.679 < \text{CSV discriminator.} \tag{5.13}$$

The specific values of the thresholds are chosen according to the operating points that are recommended and deployed in the CMS collaboration. With an upper threshold of 0.244, corresponding to the 'loose operating point' of the CSV algorithm, we ensure that the efficiency of anti-b-tagging light-flavor quarks is about 90%. The probability for jets originating from b quarks to pass the anti-b-tagging requirement is less than 20%. With the lower CSV threshold of 0.679 for b-tagged jets, referred to as the 'medium operating point' of the CSV algorithm, a tagging efficiency of b-quark jets of about 70% can be achieved, while rejecting jets from light-flavor quarks to the percent level. The actual values of the efficiencies depend on the jet kinematics, as shown below.

Application of b-tag efficiency scale factors

Dedicated studies showed differences in the b-tag and non-b-tag efficiency between data and simulation, and scale factors have been derived to account for these effects [145]. Some of the subsamples defined in the previous section have a jet configuration with both b-tagged jets and anti-b-tagged jets, which somewhat complicates the application of these scale factors to the simulated events.

From simulated $t\bar{t}$ events we obtain the probability that a jet originating from a b quark is tagged as a b jet. For b-tagging we use a different threshold of the CSV discriminator than for anti-b-tagging, hence we derive the b-tagging efficiencies ϵ^M (for the medium CSV operating point) and ϵ^L (for the loose CSV operating point). These efficiencies depend not only on the flavor of the underlying quark, but also on the jet p_T and η, as shown in Figure 5.5. The dependence on the jet kinematics originates from the fact that the track and the secondary-vertex reconstruction tends to be less efficient in the forward region of the tracker, or in case the jet has a high transverse momentum. The probability P_{sim} to obtain a jet configuration in a simulated event with n jets b-tagged using the medium CSV operating point (CSV > 0.679)), m jets anti-b-tagged using the loose CSV operating point ($0 \leq$ CSV < 0.244)) and q jets not b-tagged using the medium CSV operating point but also not anti-b-tagged using the loose operating point ($0.244 <$ CSV < 0.679)) is then given by:

$$P_{sim} = \prod_{i=1}^{n} \epsilon_i^M \times \prod_{j=1}^{m} (1 - \epsilon_j^L) \times \prod_{k=1}^{q} (\epsilon_k^L - \epsilon_k^M) \tag{5.14}$$

The probability P_{obs} to observe such a jet configuration in the data is different due to the differences in b-tag efficiencies. The efficiency in data can be obtained by multiplying the derived scale factors SF (which in general also depend on the flavor and the p_T and η of the jet) with the efficiencies obtained in simulation, hence the probability P_{obs} can be written as:

$$P_{obs} = \prod_{i=1}^{n} \mathrm{SF}_i^M \epsilon_i^M \times \prod_{j=1}^{m} (1 - \mathrm{SF}_j^L \epsilon_j^L) \times \prod_{k=1}^{q} (\mathrm{SF}_k^L \epsilon_k^L - \mathrm{SF}_k^M \epsilon_k^M). \tag{5.15}$$

To account for the differences in b-tag efficiency in data and simulation, we then weight each simulated event by the ratio of these probabilities:

$$\text{event weight} = \frac{P_{obs}}{P_{sim}}. \tag{5.16}$$

5.4 Subsamples optimized for vector-like quark processes

A priori, the branching fractions of the vector-like quarks to W, Z or H bosons, as well as the strength of the electroweak vector-like quark production, are unknown. This results in a large parameter space that should be considered in order to maximize the

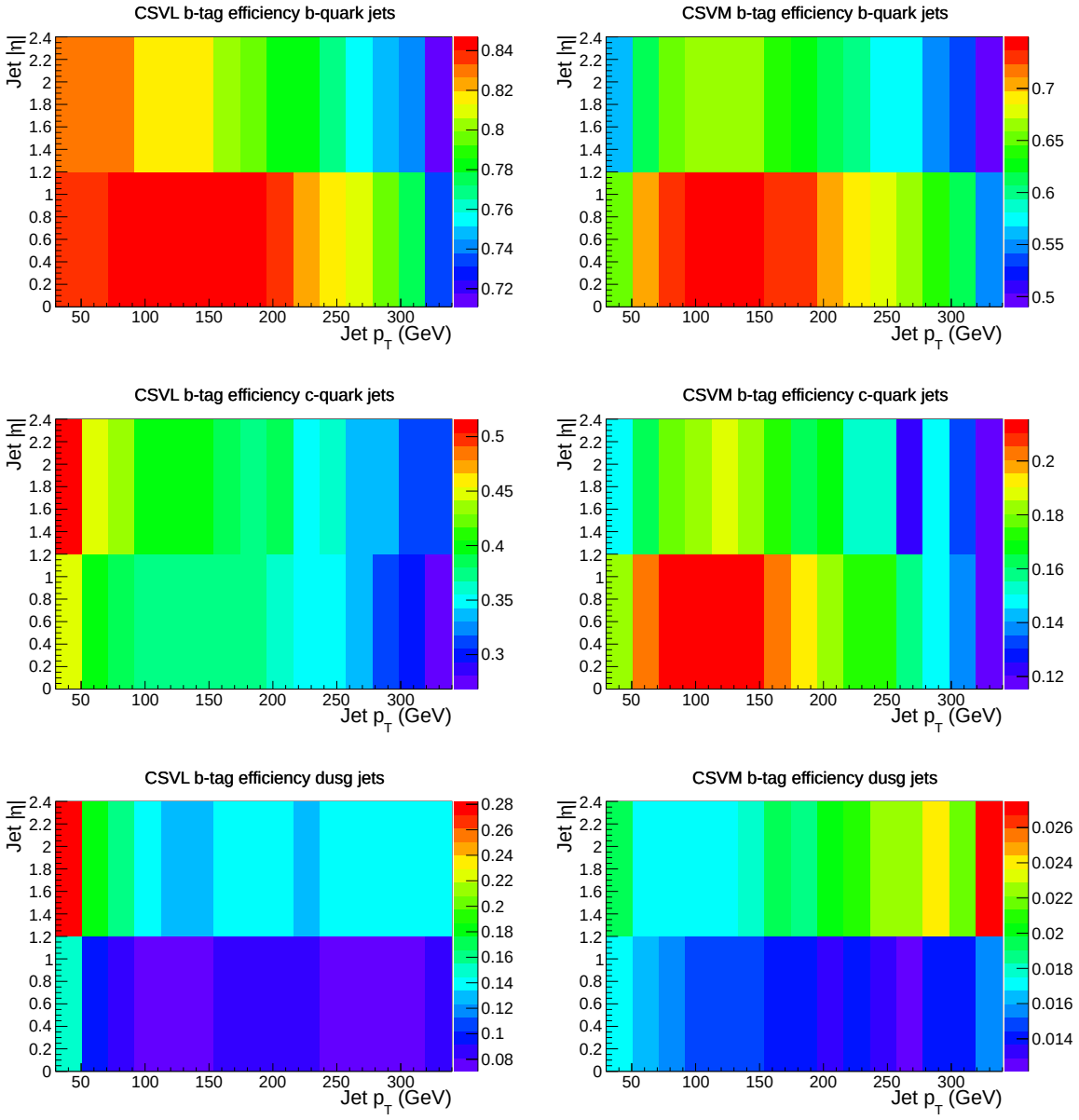

Figure 5.5: The b-tag efficiency for simulated jets originating from b quarks (top), c quarks (middle) and d, u, s quarks and gluons (bottom) are determined in bins of jet p_T and η, and for the loose (left) and the medium (right) Combined Secondary Vertex operating points. These efficiencies, obtained from $t\bar{t}$ simulated events, are used in the determination of the simulated-event weights to account for differences in b-tag and anti-b-tag efficiencies in data and simulation.

probability of observing the potential signatures of these new quarks. We define an inclusive strategy, where seven subsamples are created by applying selection criteria on the physics objects in the events. The selection within each subsample is optimized for specific vector-like quark production and decay modes that are most promising for a potential discovery of these new particles.

We categorize the events according to the number of isolated leptons and additional selection criteria on the jets and the missing transverse energy. Each of the exclusive subsamples is aimed to be particularly sensitive to one or more of the topologies in Table 5.1. This is reflected in the names used to designate the subsamples. Because of the complexity of the selection, the definition of each subsample is summarized in Tables 5.6 and 5.7. It is worth to mention that vector-like quark signal topologies involving tau leptons from W or Z boson decays are possible, but the defined subsamples are not optimized for these cases.

- **W^+qq and W^-qq categories.** When a single vector-like quark is produced via the electroweak interaction in the t-channel, it is associated with a forward jet. When the heavy vector-like quark decays to a W boson and a light-flavor jet, these decay products will tend to have a large boost. When the W boson subsequently decays to a charged lepton and a neutrino, the final state will consist of a high-p_T charged lepton, high missing transverse energy, a very high-p_T jet and a high-p_T forward jet. Hence we require exactly one isolated tight muon or electron with $p_T > 30$ GeV, and veto for additional loose leptons. One or two anti-b-tagged selected jets, where the leading jet should have $p_T > 200$ GeV, and exactly one selected forward jet are required. We ask for the missing transverse energy to exceed 60 GeV, and the transverse mass M_T of the lepton and the missing transverse energy (as defined in Equation (4.2)) to be larger than 40 GeV. The selection requirements on $\not\!\!E_T$ and M_T are imposed to suppress the QCD multijet background. Since the W boson is expected to be boosted, we require the transverse momentum of the W boson candidate, constructed from the muon or electron and the reconstructed neutrino, to be larger than 150 GeV. The procedure to reconstruct the neutrino is explained in Section 5.6. In addition we require an angular separation ΔR between the lepton and the reconstructed neutrino to be smaller than 1.5, since for boosted W bosons these leptons are expected to be close together. The left side of Figure 5.6 illustrates with simulated single-muon events that the reconstructed W-boson candidate is indeed more boosted for a reference vector-like quark D signal (mass $m_Q = 1100$ GeV, $\tilde{\kappa}_W = 1$, $\mathrm{BF}_W = 0.5$ and $\mathrm{BF}_Z = 0.25$) than for SM background processes. The right side of Figure 5.6 shows the ΔR between the lepton and the reconstructed neutrino is on average lower than for the SM background processes. In these figures, the QCD multijet estimation and the normalization of the W+jets and Z+jets processes are determined from data as described in Section 5.5. After all selection criteria described above, the main SM background originates from W+jets processes.

  We can further split the obtained event sample, which can be referred to as the Wqq category, in two exclusive subsamples according to the charge of the lepton, the W^+qq and the W^-qq subsample, where the charge of the lepton is assumed to correspond to the charge of the W boson. The motivation of this additional

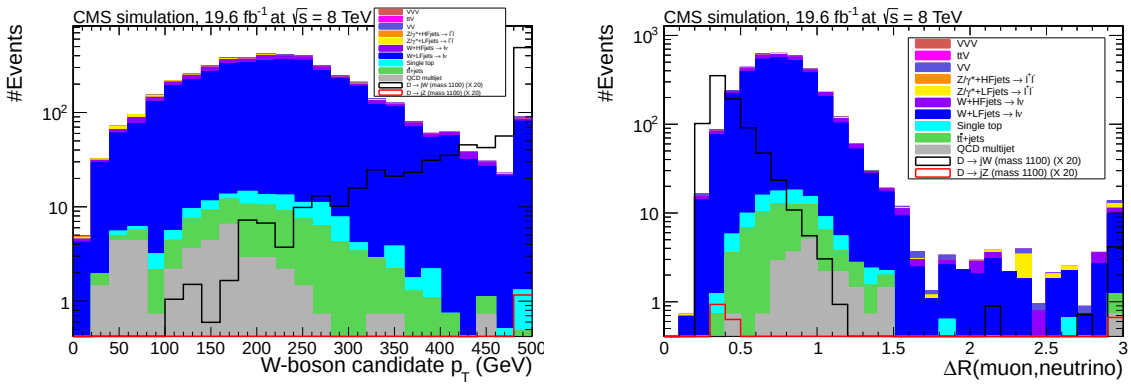

Figure 5.6: The transverse momentum of the reconstructed W-boson candidate tends to be higher for vector-like quarks than for SM background processes (left). The boost of the W boson also results in a more collimated lepton-neutrino system in vector-like quark processes (right). In the analysis, the requirements $p_T(W \to \ell\nu) > 150$ GeV and $\Delta R(\ell, \nu) < 1.5$ are imposed in the Wqq event category.

subdivision is that because of the parton density function of the proton, the production rate of a D quark would be higher than that of a $\overline{D}$ quark. At the same time, the Standard-Model W-boson production is also known to be charge-asymmetric due to the proton parton structure, with more W^+ bosons produced than W^- bosons. In the actual search for vector-like D quarks, we will only consider the $\mathrm{W^-qq}$ subsample, since using this subsample results in a significant enhancement of the signal-to-background ratio compared to using the Wqq subsample. It should be noted that in a U-quark scenario, the main signal would contribute to the $\mathrm{W^+qq}$ subsample instead of the $\mathrm{W^-qq}$ subsample, because of its electric charge of $+2/3$.

- **Zqq category.** This category is designed to be sensitive to the single production of a vector-like quark decaying to a Z boson and a light-flavor jet. Considering a leptonic decay of the Z boson, we expect two opposite-sign same-flavor leptons in addition to a very high-p_T jet and a forward jet. We select events with two tight muons or two tight electrons of opposite charge and within 7.5 GeV around the expected Z-boson mass. The leading lepton is required to have $p_T > 30$ GeV and the subleading lepton $p_T > 20$ GeV. Additional loose leptons are vetoed. The jet requirements are the same as in the Wqq categories. We reconstruct the Z-boson candidate from the two tight leptons and require the transverse momentum of this Z-boson candidate to be larger than 150 GeV. We do not expect neutrinos in this topology, hence we do not apply a threshold on the missing transverse energy or on M_T. The SM background in this subsample is dominated by Z+jets processes.

- **WWqq category.** This subsample is optimized for the pair production of vector-like quarks, both decaying to a W boson and a very high-p_T jet. Considering the leptonic decay of both W bosons, we select two tight muons, two tight electrons or a tight muon and a tight electron, while vetoing for additional loose

leptons. The two leptons are required to be of opposite charge, since the heavy-quark pair is produced as a quark and anti-quark. When the charged leptons are of the same flavor, they are required to have an invariant mass more than 7.5 GeV away from Z-boson mass. One can expect significant missing transverse energy in these topologies, hence we require $\not\!\!E_T > 60$ GeV. We require at least two selected jets in the event, and they should all pass the anti-b-tagging requirement. Contrary to the single heavy-quark production, the final state is likely to contain two very high-p_T jets. Therefore we require the leading and the subleading jet to have a transverse momentum larger than 200 GeV and 100 GeV, respectively. Because of the ambiguity in reconstructing the neutrinos, we do not reconstruct W-boson candidates in this subsample. The SM processes that survive these selections are mainly the $t\bar{t}$ and Z+jets processes.

- **WHqq category.** When the branching fractions of vector-like quarks to a W boson and to a H boson are sizeable, some of the pair-produced heavy quarks will result in mixed decay topologies. Here one of the quarks would decay to a W boson and the other to a H boson, where both bosons are accompanied by a very high-p_T jet. We consider the leptonic decay of the W boson, and the decay of the H boson to a $b\bar{b}$ quark pair. Hence, we select these events by requiring exactly one tight muon or electron with $p_T > 30$ GeV, while vetoing for other loose leptons. The missing transverse energy should be larger than 60 GeV. We require at least three selected jets, with two of them anti-b-tagged. The transverse momentum of the leading and subleading anti-b-tagged jet are required to exceed 200 GeV and 100 GeV, respectively. In addition, we require the presence of at least one b jet. We outline in Section 5.6 how the H-boson candidate is reconstructed from the selected jets. In case a H-boson candidate is found, we require its transverse momentum to be above 75 GeV. We reconstruct neither the neutrino nor the W-boson candidate. Because of the application of b-tagging, the main SM background in this event category is top-quark pair production, but also processes involving the production of a W boson in association with heavy-flavor jets.

- **ZHqq category.** The selection criteria in this subsample are similar to the WHqq subsample, except that it is optimized for the case where one of the pair-produced vector-like quarks decays to a Z boson, instead of a W boson. We require two opposite-sign muons or electrons within the Z-boson mass window of 7.5 GeV around the expected mass, and the resulting Z-boson candidate should have a transverse momentum larger than 150 GeV. Additional loose leptons are vetoed. Contrary to the WHqq category, no H-boson candidate is reconstructed and hence no requirements are imposed on such a candidate. The main background in this subsample is Z-boson production in association with heavy-flavor jets.

- **VZqq semi-leptonic category.** In this subsample we are sensitive to the pair production of vector-like quarks in two particular situations. The first is the topology where one heavy quark decays to a leptonically decaying Z boson, while the other one decays to a W boson, that subsequently decays to jets. The second

situation that we consider is where both vector-like quarks decay to Z bosons, but one decays to two leptons and the other one to two jets. In both scenarios there are two leptons from a Z boson decay. Hence, we require the presence of two tight opposite-sign same-flavor leptons, the leading lepton with $p_T > 30$ GeV and the subleading one with $p_T > 20$ GeV, and we veto for additional loose leptons. The invariant mass of the leptons must be inside the mass window of 7.5 GeV around the Z-boson mass, and the p_T of the Z candidate formed by the leptons should be larger than 150 GeV. We require at least four selected jets and impose a b-jet veto in the event. At least two anti-b-tagged jets should be present, one with $p_T > 200$ GeV and the other with $p_T > 100$ GeV. The background in this subsample is dominated by Z+jets production.

- **VZqq full-leptonic category.** This category is again defined to be sensitive to the pair production of vector-like quarks where one of the quarks decays to a W boson and the other to a Z boson, or where both quarks decay to a Z boson. This time, however, we consider the case where all bosons decay leptonically, resulting in final states of three or four leptons. We require at least one tight lepton with $p_T > 30$ GeV, and the p_T of the other tight leptons must be above 20 GeV. We only consider events for which the lepton configuration is compatible with the assumed topology, for instance, in events with two muons and one electron, the two muons are automatically the only option to reconstruct the Z boson. However, they still need to satisfy the opposite-charge and the invariant-mass requirements in order to be considered as a Z boson candidate. When such a candidate is found, it transverse momentum is required to exceed 150 GeV. We do not reconstruct any W-boson candidate in this subsample. An additional threshold on the ΔR angular separation between other-flavor leptons is imposed, as motivated in the last paragraph of Section 5.5. At least two anti-b-tagged jets should be present, one with $p_T > 200$ GeV and the other with $p_T > 100$ GeV. In events with three tight leptons, a missing transverse energy above 60 GeV is required. Here the selected event yield is low, and the background consists of two potential sources. The first one involves processes with at least three prompt leptons (WZ, ZZ and to a lesser extend the triboson processes). This is referred to as the irreducible background. The second component originates from fake leptons in processes with two prompt leptons like the Z+jets process.

The event yields in these subsamples from the observed data as well as the SM expectation are shown in Table 5.8 for the muon channel and Table 5.9 for the electron channel. Two expected event yields for two reference vector-like D quark signals are shown, one for $m_Q = 1100$ GeV and $\tilde{\kappa}_W = 1$ ('single-production reference signal'), and the other for $m_Q = 600$ GeV and $\tilde{\kappa}_W = 0.1$ ('pair-production reference signal'). The choices of the branching fractions are in both cases $\mathrm{BF}_W = 0.5$ and $\mathrm{BF}_Z = \mathrm{BF}_H = 0.25$. The QCD multijet estimation and the normalization of the W+jets and Z+jets processes in this table are determined from data as described in Section 5.5.

Table 5.6: The subsamples in this table are optimized for the single production of vector-like quarks. They are defined according to the number of tight muons or electrons and additional criteria optimized for specific vector-like quark topologies. In each subsample, we veto for additional loose leptons. The W^+qq subsample is not directly used to search for D quarks, but would be used in a search for U quarks.

Subsample	Tight leptons (μ,e)	Additional selection criteria
(W^+qq) W^-qq	1 with $p_T > 30$ GeV	1 or 2 selected anti-b-tagged jets leading $p_T > 200$ GeV 1 selected forward jet $p_T(W \to \ell\nu) > 150$ GeV $\Delta R(\ell, \nu) < 1.5$ $\not{E}_T > 60$ GeV, $M_T > 40$ GeV
Zqq	2 opposite-sign same-flavor leading $p_T > 30$ GeV subleading $p_T > 20$ GeV	1 or 2 selected anti-b-tagged jets leading $p_T > 200$ GeV 1 selected forward jet $\lvert m_{\ell\ell} - m_Z \rvert < 7.5$ GeV $p_T(Z \to \ell\ell) > 150$ GeV

5.5 Data-driven background estimations

In some of the subsamples defined in Section 5.4 we estimate the SM background from the observed data. For the W+jets and Z+jets processes, we derive normalization factors in a data-driven way. These normalization factors are afterwards propagated into the estimation of the QCD multijet background, for which we derive both the shape and the normalization of the corresponding distributions from data. In the event category with more than three leptons, we apply the loose-to-tight method to estimate from data the contribution of fake leptons. A dedicated data-driven estimation of the combined W+jets and QCD multijet processes in the Wqq subsample is performed later in Section 5.6.1, since it makes use of the actual reconstructed search variable in this event subsample.

Normalization of the W+jets and Z+jets processes

The SM processes involving the production of a W or Z boson in association with jets are important backgrounds in our search for vector-like quarks, especially in the Wqq and Zqq subsamples that are sensitive to the single-production mode where the heavy quark decays to a W or Z boson. Although the inclusive cross sections of these processes have been measured by the CMS experiment with a precision of less than 4% [174], they are not necessarily well known in the kinematic region we study in this analysis. Therefore, as described below, we fit the normalization of the simulated Z+jets process to the data in a control region, and do the same for the W+jets background. The purpose of fitting the normalization of these backgrounds is mainly for later use in the determination of the QCD multijet background.

In the definition of the control region for the Z+jets background, we employ similar

Table 5.7: The subsamples in this table are optimized for the pair production of vector-like quarks. They are defined according to the number of tight muons or electrons and additional criteria to optimized for specific vector-like quark topologies. In each subsample, we veto for additional loose leptons.

Subsample	Tight leptons (μ,e)	Additional selection criteria
WWqq	2 opposite-sign leading $p_T > 30$ GeV subleading $p_T > 20$ GeV	≥ 2 selected jets all jets anti-b-tagged leading $p_T > 200$ GeV subleading $p_T > 100$ GeV $\lvert m_{\ell\ell} - m_Z \rvert > 7.5$ GeV (same flavor) $\not\!\!E_T > 60$ GeV
WHqq	1 with $p_T > 30$ GeV	≥ 3 selected jets ≥ 2 anti-b-tagged jets leading $p_T > 200$ GeV subleading $p_T > 100$ GeV ≥ 1 b-tagged jet $p_T(H \to b\bar{b}) > 75$ GeV $\not\!\!E_T > 60$ GeV
ZHqq	2 opposite-sign same-flavor leading $p_T > 30$ GeV subleading $p_T > 20$ GeV	≥ 3 selected jets ≥ 2 anti-b-tagged jets leading $p_T > 200$ GeV subleading $p_T > 100$ GeV ≥ 1 b-tagged jet $\lvert m_{\ell\ell} - m_Z \rvert < 7.5$ GeV $p_T(Z \to \ell\ell) > 150$ GeV
VZqq semi-leptonic	2 opposite-sign same-flavor leading $p_T > 30$ GeV subleading $p_T > 20$ GeV	≥ 4 selected jets ≥ 2 anti-b-tagged jets leading $p_T > 200$ GeV subleading $p_T > 100$ GeV veto b-tagged jets $\lvert m_{\ell\ell} - m_Z \rvert < 7.5$ GeV $p_T(Z \to \ell\ell) > 150$ GeV
VZqq full-leptonic	3 or 4 leading $p_T > 30$ GeV others $p_T > 20$ GeV	≥ 2 selected jets all jets anti-b-tagged leading $p_T > 200$ GeV subleading $p_T > 100$ GeV $\lvert m_{\ell\ell} - m_Z \rvert < 7.5$ GeV $p_T(Z \to \ell\ell) > 150$ GeV $\not\!\!E_T > 60$ GeV (3 leptons) $\Delta R(\ell, \ell) > 0.05$ (other flavor)

Table 5.8: Event yields in the muon channel, for the subsamples with one or two isolated leptons. The indicated uncertainties are only statistical. The prediction for the signal is shown for two different reference signals, as defined in the text.

$\mathcal{L} = 19.6$ fb^{-1}	W^+qq	W^-qq	Zqq	WWqq	WHqq	ZHqq	VZqq semi-lep.
$t\bar{t}$+jets	$26.0^{+5.5}_{-4.6}$	$28.2^{+5.7}_{-4.8}$	$0.2^{+1.2}_{-0.2}$	$60.5^{+8.0}_{-7.1}$	$825.1^{+9.8}_{-10.5}$	$1.8^{+1.9}_{-0.9}$	$0.2^{+1.2}_{-0.2}$
W+jets	$2088.4^{+28.7}_{-28.9}$	$1173.5^{+28.0}_{-27.7}$	$0.0^{+1.0}_{-0.0}$	$0.8^{+1.5}_{-0.5}$	$199.6^{+14.2}_{-13.3}$	$0.0^{+1.0}_{-0.0}$	$0.0^{+1.0}_{-0.0}$
Z+jets	$17.3^{+4.6}_{-3.7}$	$22.8^{+5.2}_{-4.3}$	$537.5^{+14.9}_{-15.1}$	$76.2^{+8.8}_{-7.9}$	$13.4^{+4.2}_{-3.2}$	$36.1^{+6.4}_{-5.4}$	$217.1^{+13.2}_{-12.6}$
Single top	$19.8^{+4.4}_{-3.7}$	$10.3^{+3.5}_{-2.7}$	$0.0^{+1.0}_{-0.0}$	$4.4^{+2.6}_{-1.6}$	$79.5^{+4.7}_{-5.1}$	$0.0^{+1.0}_{-0.0}$	$0.0^{+1.0}_{-0.0}$
VV	$29.1^{+4.6}_{-4.2}$	$31.3^{+4.7}_{-4.4}$	$10.3^{+3.4}_{-2.7}$	$8.9^{+3.2}_{-2.5}$	$6.0^{+2.8}_{-2.0}$	$1.1^{+1.6}_{-0.7}$	$3.8^{+2.4}_{-1.5}$
$t\bar{t}V$	$0.1^{+1.0}_{-0.1}$	$0.1^{+1.0}_{-0.1}$	$0.0^{+0.9}_{-0.0}$	$0.6^{+1.2}_{-0.4}$	$9.1^{+1.0}_{-1.6}$	$1.1^{+1.4}_{-0.7}$	$0.2^{+1.1}_{-0.2}$
VVV	$0.3^{+0.9}_{-0.2}$	$0.3^{+0.9}_{-0.2}$	$0.1^{+0.8}_{-0.1}$	$1.2^{+0.9}_{-0.7}$	$0.7^{+0.9}_{-0.5}$	$0.2^{+0.8}_{-0.2}$	$0.4^{+0.9}_{-0.3}$
QCD multijet	$12.2^{+3.8}_{-2.9}$	$8.6^{+3.3}_{-2.4}$	$0.0^{+1.0}_{-0.0}$	$43.2^{+5.7}_{-5.3}$	$79.2^{+5.9}_{-6.0}$	$0.7^{+1.5}_{-0.5}$	$0.0^{+1.0}_{-0.0}$
Total background	$2193.2^{+36.8}_{-36.6}$	$1275.2^{+31.7}_{-31.2}$	$548.1^{+22.7}_{-21.9}$	$195.8^{+14.2}_{-13.3}$	$1212.6^{+31.2}_{-30.6}$	$41.0^{+6.9}_{-5.9}$	$221.7^{+15.1}_{-14.1}$
Observed	2082	1112	527	178	1172	54	249
Pair ref. signal	$1.9^{+1.9}_{-1.0}$	$4.3^{+2.4}_{-1.6}$	$1.5^{+1.7}_{-0.8}$	$12.9^{+3.3}_{-2.8}$	$15.4^{+3.5}_{-3.1}$	$4.0^{+2.4}_{-1.5}$	$11.0^{+3.2}_{-2.7}$
Single ref. signal	$6.6^{+2.8}_{-2.0}$	$35.0^{+3.3}_{-3.6}$	$10.5^{+3.2}_{-2.6}$	$0.2^{+1.1}_{-0.2}$	$0.7^{+1.4}_{-0.5}$	$0.2^{+1.1}_{-0.1}$	$1.0^{+1.6}_{-0.6}$

Table 5.9: Event yields in the electron channel, for the subsamples with one or two isolated leptons. The indicated uncertainties are only statistical. The prediction for the signal is shown for two different reference points, as defined in the text.

$\mathcal{L} = 19.6$ fb^{-1}	W^+qq	W^-qq	Zqq	WWqq	WHqq	ZHqq	VZqq semi-lep.
$t\bar{t}$+jets	$22.3^{+5.1}_{-4.2}$	$22.5^{+5.2}_{-4.2}$	$0.0^{+1.0}_{-0.0}$	$21.5^{+5.1}_{-4.1}$	$765.9^{+7.7}_{-8.5}$	$1.4^{+1.8}_{-0.8}$	$0.5^{+1.3}_{-0.4}$
W+jets	$1930.0^{+27.5}_{-27.7}$	$1079.2^{+26.9}_{-26.6}$	$0.0^{+1.0}_{-0.0}$	$0.4^{+1.3}_{-0.3}$	$183.1^{+13.6}_{-12.7}$	$0.0^{+1.0}_{-0.0}$	$0.0^{+1.0}_{-0.0}$
Z+jets	$10.0^{+3.7}_{-2.7}$	$8.1^{+3.4}_{-2.4}$	$440.5^{+13.5}_{-13.6}$	$62.4^{+8.0}_{-7.2}$	$15.4^{+4.4}_{-3.4}$	$28.8^{+5.7}_{-4.8}$	$189.4^{+12.2}_{-11.7}$
Single top	$19.8^{+4.4}_{-3.7}$	$11.7^{+3.7}_{-2.9}$	$0.0^{+1.0}_{-0.0}$	$1.9^{+1.9}_{-1.0}$	$82.9^{+4.8}_{-5.2}$	$0.0^{+1.0}_{-0.0}$	$0.0^{+1.0}_{-0.0}$
VV	$27.9^{+4.4}_{-4.1}$	$31.5^{+4.5}_{-4.3}$	$8.6^{+3.2}_{-2.4}$	$3.9^{+2.4}_{-1.5}$	$5.4^{+2.7}_{-1.8}$	$0.7^{+1.4}_{-0.5}$	$3.8^{+2.4}_{-1.5}$
$t\bar{t}V$	$0.0^{+0.9}_{-0.0}$	$0.0^{+0.9}_{-0.0}$	$0.0^{+0.9}_{-0.0}$	$0.2^{+1.0}_{-0.2}$	$7.8^{+0.7}_{-1.4}$	$0.7^{+1.2}_{-0.5}$	$0.2^{+1.1}_{-0.2}$
VVV	$0.4^{+0.8}_{-0.3}$	$0.3^{+0.8}_{-0.2}$	$0.1^{+0.8}_{-0.1}$	$0.5^{+0.8}_{-0.4}$	$0.7^{+0.8}_{-0.5}$	$0.1^{+0.8}_{-0.1}$	$0.4^{+0.8}_{-0.3}$
QCD multijet	$13.4^{+3.8}_{-3.0}$	$14.3^{+3.8}_{-3.1}$	$0.4^{+1.3}_{-0.3}$	$14.7^{+4.1}_{-3.3}$	$54.6^{+6.6}_{-6.1}$	$0.0^{+1.0}_{-0.0}$	$0.0^{+1.0}_{-0.0}$
Total background	$2023.7^{+35.0}_{-34.8}$	$1167.6^{+30.3}_{-29.7}$	$449.7^{+20.7}_{-19.9}$	$105.5^{+10.7}_{-9.7}$	$1116.0^{+29.9}_{-29.4}$	$31.7^{+6.1}_{-5.1}$	$194.2^{+14.1}_{-13.2}$
Observed	1862	1039	440	105	1055	49	208
Pair ref. signal	$1.4^{+1.7}_{-0.8}$	$3.8^{+2.3}_{-1.5}$	$1.5^{+1.7}_{-0.8}$	$5.3^{+2.5}_{-1.8}$	$14.2^{+3.1}_{-2.8}$	$3.8^{+2.3}_{-1.5}$	$8.6^{+2.8}_{-2.3}$
Single ref. signal	$5.7^{+2.6}_{-1.9}$	$30.7^{+3.1}_{-3.4}$	$9.5^{+3.0}_{-2.4}$	$0.1^{+1.1}_{-0.1}$	$0.4^{+1.3}_{-0.3}$	$0.1^{+1.1}_{-0.1}$	$1.0^{+1.6}_{-0.6}$

selection criteria as in the Zqq subsample (see Table 5.6) but veto on forward jets and require the leading jet p_T to be between 150 GeV and 200 GeV. Also the requirement on the transverse momentum of the Z-boson candidate is dropped. In this way the selected events have reasonably similar properties to the events in the Zqq subsample, but a significant signal-event contamination in the control region is avoided. We use a maximum-likelihood technique to fit the leading-jet p_T distributions of the Z+jets, $t\bar{t}$ and all other backgrounds combined, to the observed data in the muon channel. In this fit, the normalization of the Z+jets background is unconstrained. The normalization of the other background processes are constrained within the largest uncertainty of the dominating diboson components (taken to be 17% from the theoretical uncertainty in the WZ production). Keeping the $t\bar{t}$ normalization fixed or constrained within its uncertainty from the CMS measurement (14%), does not make a difference in the fit, as the $t\bar{t}$ contribution is very small. The fitted normalization factors are 0.86 for Z+jets, and compatible with unity for the other backgrounds.

Similarly, in the muon-channel control region in which we estimate the normalization factor of the W+jets component, a veto for forward jets is applied and the leading jet p_T is required to be between 150 GeV and 200 GeV. The other selections are the same as in the Wqq event category (see Table 5.6), except that the requirements on the p_T of the W-boson candidate and the ΔR between the lepton and neutrino are dropped. Now the W+jets background normalization is unconstrained in the fit, and the $t\bar{t}$ normalization is constrained to its uncertainty from the CMS measurement. The remaining background is dominated by diboson and Z+jets processes. The normalization of the latter background is fixed to the one derived in the previous fit. It has been checked that allowing the non-$t\bar{t}$ backgrounds to vary within a conservative 20% in the fit, does not affect the fitted normalization factor of W+jets significantly. The background of QCD multijet events is assumed to be negligible. The fitted normalization factor for the W+jets background is found to be 0.86, hence of the same magnitude as the normalization factor for Z+jets. The fitted normalization factor of the $t\bar{t}$ background is again compatible with unity.

The resulting fitted distributions of the transverse momentum of the leading-jet p_T in the defined control regions are shown in Figure 5.7. It should be noted that all fitted normalizations are still allowed to be modified in the combined fit of the systematic nuisance parameters that will be performed in the statistical limit setting procedure, as described in Section 5.7.3. Hence, the fitted Z+jets and W+jets normalizations are only a reference with respect to which the normalization can still be varied afterwards. As the relative contribution of vector bosons in association with light-flavor (LF) jets and with heavy-flavor (HF) jets is rather uncertain, the normalizations of these two contributions will be allowed to vary separately in the limit setting procedure. Moreover, in the Wqq event category, another data-driven approach will be used to estimate the shape and normalization of the background distribution of the corresponding search variable. This will be explained in more detail in Section 5.6.1.

QCD multijet estimation

In the Wqq subsamples we apply an event selection of one isolated lepton, one or two central non-b jets and one forward jet. When a non-prompt lepton originating from a

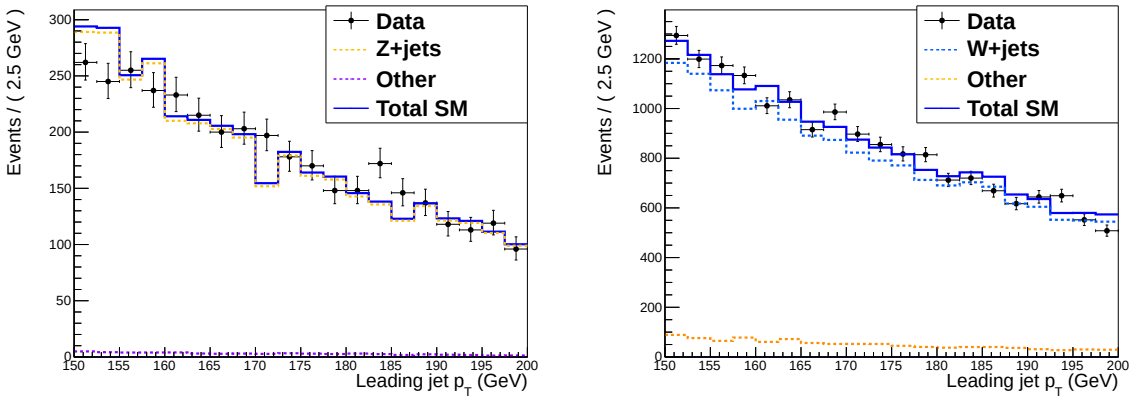

Figure 5.7: The spectra of the transverse momentum of the leading jet in regions with selection criteria similar but not overlapping with the Zqq subsample selections (left) and the Wqq sample selections (right) are fitted to the data. The dashed lines represent separate contributions, while the solid blue line is the total fitted spectrum. Normalization factors for the Z+jets and W+jets background processes are derived from these fits.

jet happens to survive the lepton and isolation selection criteria, or an isolated track is misidentified as a lepton, QCD multijet events can contribute to the background. The probability of such a fake lepton passing the tight-lepton criteria is low. However, since the cross section of multijet events is very high, this background can be significant, especially in the electron channel because the fake rate for electrons is higher than for muons. Moreover, the large QCD multijet cross section makes it difficult to generate a large enough sample of simulated events that is adequate for a reliable comparison between data and simulation. Therefore we estimate this background in a data-driven way. In the Wqq subsample, the QCD multijet estimation is determined simultaneously with the W+jets component from data, as described later in Section 5.6.1. Nevertheless, we still estimate this QCD multijet contribution separately to be able to assess its relative contribution to the background in this subsample, as well as to extrapolate its estimation to the other subsamples in the analysis.

We select a QCD multijet enriched data sample by applying the event selection criteria on the observed data but inverting the lepton isolation criterion. For the muon channel, the muon requirement is $I_{rel} > 0.2$. In the electron channel, a QCD multijet enriched sample can be obtained by requiring the electron isolation to be $I_{rel} > 0.2$, and inverting the $mvaId$ identification variable as well, $mvaId < 0.5$. The normalization of the QCD multijet background is then determined with single-lepton events with a lepton with $p_T > 30$ GeV and a veto on additional loose leptons. In addition, we require 1 or 2 anti-b tagged jets, with the leading jet p_T between 150 GeV and 200 GeV and veto on forward selected jets. We do not impose a threshold on the missing transverse energy or transverse mass M_T. No additional cuts on neutrino and lepton systems have been applied, since neither a neutrino or a W-boson candidate is reconstructed. We perform a maximum-likelihood fit of the spectrum of the missing transverse energy to data, with the normalizations of the W+jets and Z+jets fixed as determined by the

previous fits. The QCD multijet normalization is left unconstrained in the fit. The resulting fitted distribution of the missing transverse energy in the muon channel and in the electron channel are shown in Figure 5.8. In the electron channel the QCD multijet background is much higher than in the muon channel, due to the higher probability of a non-prompt electron to pass the trigger, identification and isolation selection criteria.

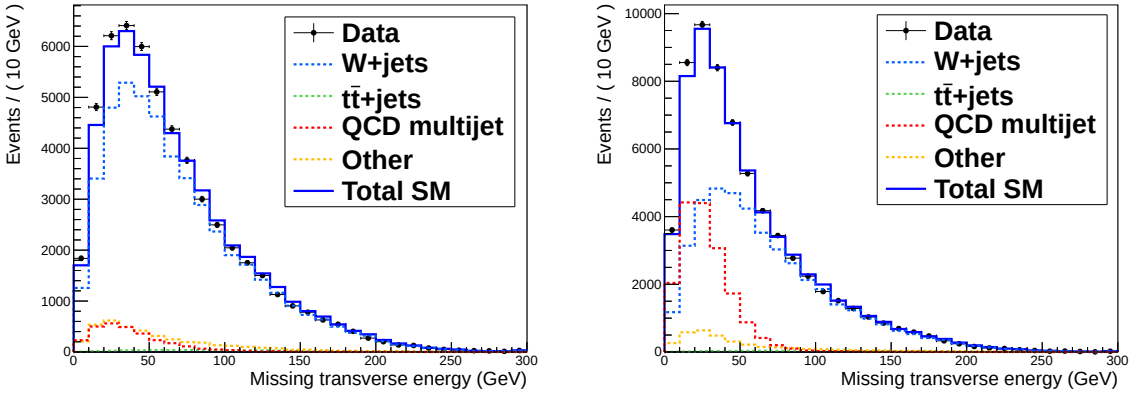

Figure 5.8: The spectra of the missing transverse energy in regions with selection criteria similar but not overlapping with the Wqq sample selection are fitted to the data in the muon channel (left) and in the electron channel (right). The dashed lines represent separate contributions, while the solid blue line is the total fitted spectrum. Normalization factors for the QCD multijet component of the SM background are derived in these fits.

Fake-lepton estimation in trilepton events

Processes resulting in final-state topologies with three or four leptons are very rare in the Standard Model. The main contributions in our analysis originate from diboson (WZ and ZZ) and triboson processes (WWW, WWZ, WZZ and ZZZ). In the full-leptonic VZqq event category, introduced in Section 5.4, we require all selected jets to be anti-b-tagged, which removes another potential source of prompt multilepton processes, namely top-quark pair production in association with a W or a Z boson. The diboson and triboson processes comprise the so-called irreducible background, as they involve three or four prompt charged leptons.

Another potential background source in the full-leptonic VZqq event category involves dilepton processes with a non-prompt lepton. As a Z-boson candidate is required, this would mainly consist of Z+jets events where a non-prompt lepton originates from a meson decay inside a jet. Since such a lepton is not a prompt lepton, it is referred to as a fake lepton. In simulated Z+jets events passing all full-leptonic VZqq selection criteria as listed in Table 5.7, not considering the ΔR cut on the leptons, we identify another source of fake leptons. Since a boosted Z candidate is required, at least one of the selected tight muons has a very high transverse momentum, easily above 100 GeV. When these very energetic muons deposit a significant amount of energy in the electromagnetic calorimeter, this may wrongly result in the identification

of an electron. In addition, high-p_T muons can radiate high-energetic photons through Bremsstrahlung, and if these photons convert to an electron-positron pair, the electron or the positron might be wrongly reconstructed as a prompt electron, with an associated track very close to the radiating muon. Both effects result in events with a reconstructed muon-electron pair very close in ΔR. Table 5.10 shows the six simulated Z+jets events passing the full-leptonic VZqq selection requirements, omitting a cut on the ΔR between the leptons. This motivates an additional angular requirement on charged leptons of different flavor. Therefore we impose a lower threshold of 0.05 for the ΔR between a muon and an electron in the full-leptonic VZqq subsample to reject Z+jets events with a fake electron very close to a prompt muon. It is checked that this requirement removes almost no vector-like quark signal events. We do not apply this cut on same-flavor leptons, since same-flavor leptons from a highly-boosted Z boson are expected to be reasonably close in ΔR.

Table 5.10: All simulated Z+jets events passing the full-leptonic VZqq selection criteria with the exception of the ΔR criterion on leptons, are events with two muons from the decay of the Z boson, and a non-prompt electron very close in ΔR to the leading muon. The electrons originate from the misidentification of a muon calorimeter energy deposit, or from the conversion of a photon radiated by the high-energetic muon. These types of events can easily be rejected in the analysis.

Z+jets event	Tight leptons	Leading-muon p_T	ΔR(leading muon, electron)
1	$\mu\mu e$	352.7 GeV	6.6×10^{-5}
2	$\mu\mu e$	296.5 GeV	1.7×10^{-4}
3	$\mu\mu e$	178.5 GeV	3.4×10^{-4}
4	$\mu\mu e$	236.8 GeV	2.1×10^{-4}
5	$\mu\mu e$	272.9 GeV	1.4×10^{-4}
6	$\mu\mu e$	288.3 GeV	2.4×10^{-4}

Although no simulated events from non-irreducible background processes are observed to pass the applied lepton ΔR criterion, we still estimate in a data-driven way the contribution from fake leptons passing the tight-lepton selection criteria. For this, we use the loose-to-tight method, which has already been used in the context of same-sign lepton events in Section 4.4.2 of the fourth-generation quark search.

First we estimate the probability ϵ_{TL} that a loose (L) lepton passes the tight (T) selection criteria. To apply the method on muons, we consider a single-muon data set with a prescaled single-muon trigger with a muon-p_T trigger requirement of 8 GeV. A double-electron data set with a prescaled single-electron trigger with a electron-p_T trigger requirement of 8 GeV is used for the electron fake-rate study. In these data sets a QCD multijet enriched data sample is selected by requiring at least one loose muon or electron, $\not\!\!E_T < 20$ GeV and the transverse mass of the lepton and missing transverse energy $M_T < 25$ GeV. We veto events with same-flavor leptons with an invariant mass within 20 GeV of the Z-boson mass to suppress Z+jets production processes. The number of tight and loose leptons with a p_T below 35 GeV are counted. The ratio ϵ_{TL} of these numbers can thus be calculated, and we obtain $(5.40 \pm 0.57)\%$ for the probability

that a loose muon passes the tight-muon selection criteria, and $(6.32 \pm 0.82)\%$ for the probability that a loose electron passes the tight-electron selection criteria.

Next, we count the number of trilepton events $N_{\mathrm{T,T,L-T}}$ in the analysis with the full selection of the full-leptonic VZqq subsample, except that only two of the leptons are tight leptons, and the third one is a loose-not-tight lepton. Analogous to formula (4.4), that was used for the fake-lepton contribution in same-sign dilepton events, we can write a similar equation to estimate the number of events $N_{\mathrm{T,T,T}}$ with three tight leptons where one lepton originates from a fake lepton:

$$N_{\mathrm{T,T,T}} = N_{\mathrm{T,T,L-T}} \times \frac{\epsilon_{\mathrm{TL}}}{(1 - \epsilon_{\mathrm{TL}})}. \tag{5.17}$$

We observe only one data event with this selection. This event has a tight muon, a tight electron and a loose-not-tight muon. Hence, when applying the factor $R_f^\mu = \epsilon_{\mathrm{TL}}^\mu/(1-\epsilon_{\mathrm{TL}}^\mu)$, we obtain a fake-lepton estimation of 0.06 ± 0.06. This result is compatible with the results from the SM background simulation, where only events from irreducible processes pass the full selection requirements of the full-leptonic VZqq subsample.

5.6 Search variables

In each of the exclusive subsamples defined in Section 5.4, an observable is constructed with a high discriminating power between SM background and vector-like quark processes. In several of the subsamples we reconstruct the mass of the vector-like quark, as discussed in Section 5.6.1. In other subsamples where the reconstruction of the heavy-quark mass is more ambiguous, or where the event yield is too low, we use a simpler observable like the S_T variable or the event counts, as described in Section 5.6.2.

5.6.1 Reconstruction of the mass of the heavy quark

The way the mass of the vector-like quark is reconstructed differs from subsample to subsample, as the corresponding final-state topologies are substantially different. Any mass reconstruction requires some assumption on the correspondence between the reconstructed physics objects and the presumed decay products of the initial heavy particle, and in some of the subsamples the possible ambiguity is higher than in others. Nevertheless, as soon as one can achieve more discriminating power with a mass variable than with a simpler observable such as the S_T variable, the sensitivity to the presence of the heavy quarks is increased.

The mass variable in the $\mathrm{W^-qq}$ category

In the $\mathrm{W^-qq}$ subsample, sensitive to the single production of a vector-like D quark decaying to a W boson and a high-p_T jet, we aim to reconstruct the mass of the vector-like quark by first reconstructing the neutrino from the W-boson decay. The left diagram of Figure 5.9 illustrates that afterwards, from this neutrino, the charged lepton and the leading jet in the event, the four-vector of the vector-like quark and hence its mass can be reconstructed.

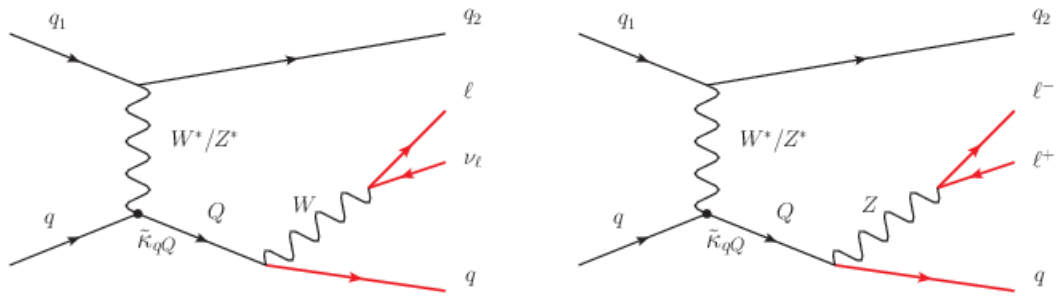

Figure 5.9: In the single-production modes, the mass of the vector-like quarks is reconstructed from the charged lepton, the neutrino and the leading jet in case the heavy quark decays to a W boson and a high-p_T jet (left), or from the charged-lepton pair and the leading jet when the vector-like quark decays to a Z boson and a high-p_T jet (right).

The neutrino momentum vector in the transverse plane (i.e. the x and y components of the momentum) is assumed to correspond to the missing-transverse energy vector. One can obtain the z component of the neutrino momentum by requiring the mass of the four-vector sum of the neutrino and the charged lepton to be equal to the W-boson mass. More concretely, the equation that needs to be solved is

$$(E_\nu + E_\ell)^2 - ((\vec{p}_\nu + \vec{p}_\ell) \cdot (\vec{p}_\nu + \vec{p}_\ell)) = m_W^2, \tag{5.18}$$

with $(E_\nu, \vec{p}_\nu)$ and $(E_\ell, \vec{p}_\ell)$ the energy-momentum four-vectors of the neutrino and the charged lepton, respectively. However, since the momentum balance along the direction of the beam pipe is not known on an event-by-event basis, the z coordinate of the neutrino momentum $p_{\nu,z}$ is ambiguous, which is reflected in the quadratic equation one obtains for the z component of the neutrino momentum. If the solution is complex, the real part is taken as the z component of the neutrino momentum. An imaginary part in the solution can result from the fact that the W-boson mass constraint is strict, while the lepton and missing transverse energy objects have a certain resolution. About 27% of the simulated vector-like quark $D \to Wq$ events of mass 1100 GeV have such a complex solution for $p_{\nu,z}$. In case the solution is real, two possibilities exist to solve the quadratic equation, and we take the solution for which the total reconstructed neutrino momentum has the largest difference in pseudorapidity with the leading jet in the event. Figure 5.10 shows the distribution of the reconstructed mass of a vector-like quark of mass 1100 GeV, when $p_{\nu,z}$ is chosen as the one for which the $\Delta\eta$ between the neutrino and the leading jet is the largest or the smallest, in case of real solutions of the W-mass constraint equation (5.18). The resolution of the distribution of the reconstructed vector-like quark mass is better (worse) for the choice of $p_{\nu,z}$ associated to the largest (smallest) pseudorapidity difference between the reconstructed neutrino and the leading jet in the event. The motivation is that in the rest frame of the heavy quark, the W boson and the high-p_T jet would decay back-to-back, hence have a large angular separation. Since the mass of the vector-like quark is expected to be very heavy, the boost of the heavy quark itself will be reasonably low, such that its direct decay products will have a large angular separation in the laboratory rest frame too.

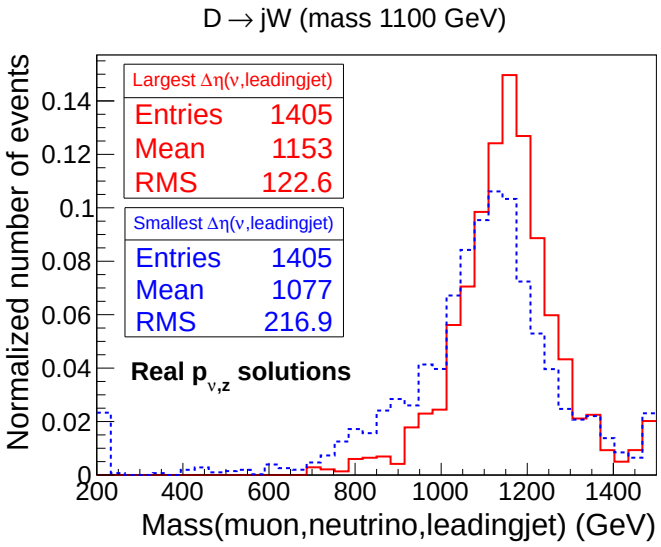

Figure 5.10: The W-boson mass constraint results in two possible choices of the z component of the neutrino momentum $p_{\nu,z}$, in case the solution is not complex. The resolution of the normalized distribution of the reconstructed vector-like quark mass is better (worse) for the choice of $p_{\nu,z}$ associated to the largest (smallest) pseudorapidity difference between the reconstructed neutrino and the leading jet in the event.

Once the neutrino is reconstructed, the W-boson candidate can be reconstructed, and consequently the vector-like-quark candidate.

The distributions of the mass of the vector-like quark candidate (the neutrino, charged lepton and leading-jet system) are shown in Figure 5.11, comparing the observed data with the SM expectation in the muon and electron channel in the Wqq category. The SM background is dominated by the W+jets process. The vector-like D quark signal processes for a mass of 1100 GeV (single-production reference signal) is shown, where the yield is multiplied by a factor 20 for visibility. It is clear that the singly-produced vector-like quark decaying to a W boson is indeed the largest signal contribution in this subsample, and the distribution of the reconstructed heavy-quark mass peaks at the generated mass of 1100 GeV. As mentioned before, we split the Wqq category in two subsamples, according to the charge of the lepton. From Figure 5.12, the charge asymmetry of the W+jets background and especially in the vector-like quark signal is clear. As mentioned before, we will not use the background simulation directly to compare to the observed data, but instead perform a data-driven estimation of the reconstructed mass distribution of the combined W+jets and QCD multijet processes, as described in the next paragraph.

Data-driven determination of the reconstructed mass distribution in the W^-qq category

Since the vector-like D-quark signal is most prominent in the W^-qq subsample and rather small in W^+qq subsample, we only keep the W^-qq subsample in the actual search for D quarks, and estimate the corresponding background from a W^+qq-like data sample. Not only is this method more reliable than putting confidence in the simulation

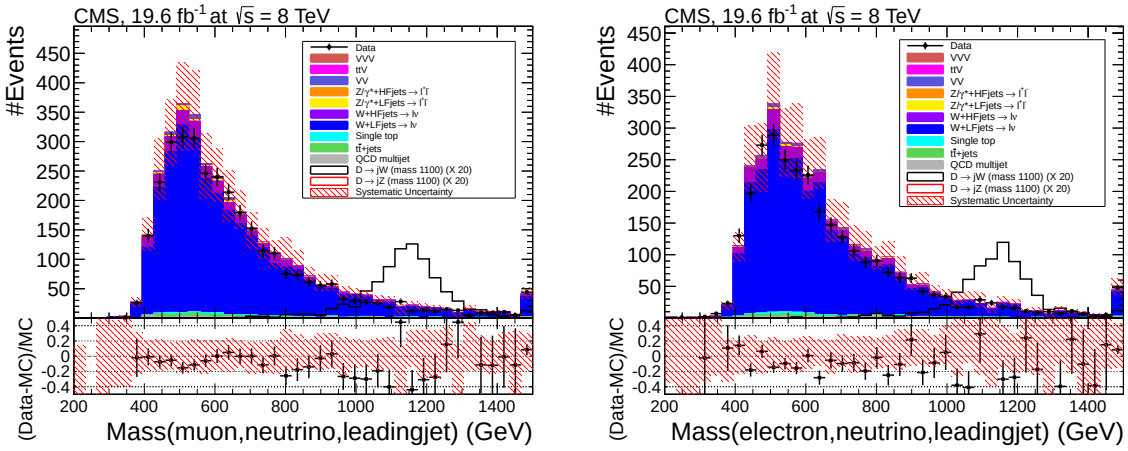

Figure 5.11: The reconstructed mass of the vector-like quark in the Wqq event category, in the muon channel (left) and the electron channel (right). The signal is scaled up by a factor 20 for visibility.

while incorporating many systematic uncertainties, one is also less affected by statistical fluctuations in the background expectation that is present due to the relatively low number of simulated W+jets events. For this purpose, we consider a control sample, defined as a W^+qq-like subsample, but with a forward-jet veto instead of a forward-jet requirement. By vetoing for forward jets, the signal contamination is reduced significantly. For a vector-like D mass of 1000 GeV, with $\tilde{\kappa}_W = 1$ and the highest-possible decay branching fraction to a W boson, the relative signal contamination in this control region is found from simulation to be only about 0.1%. This can be compared with the relative amount of signal events in the W^-qq subsample (with forward-jet requirement), which is about 7%. It should be noted that the 95% CL lower limit derived by the ATLAS experiment on a D quark with this choice of parameters is already 1120 GeV [72].

In the defined control subsample, also dominated by the W+jets background, we reconstruct the mass of the vector-like quark with the same procedure as described. Next, the shape of the reconstructed-mass distribution in the control subsample is compared with the distribution in the W^-qq signal subsample, for the W+jets background. These distributions are shown in Figure 5.13 for the muon channel. Since we notice a slightly harder spectrum in the control sample with respect to the distribution in the W^-qq signal region, we correct for this shape difference as explained in the following.

First we subtract from the data histogram in the control region the expectation from all simulated SM backgrounds other than W+jets and QCD multijets. In this way, we are left with a data sample that can be expected from the combined W+jets and QCD multijet processes, since we verified the vector-like quark signal contamination to be negligible. Next, this data distribution is normalized by the ratio of the number of expected W+jets events in the W^-qq subsample and the number of expected W+jets events in the W^+qq-like control sample, where both numbers are taken from simulation. Afterwards we correct for the aforementioned shape difference observed in the W+jets simulation, by applying the difference of the distributions in Figure 5.13 to this

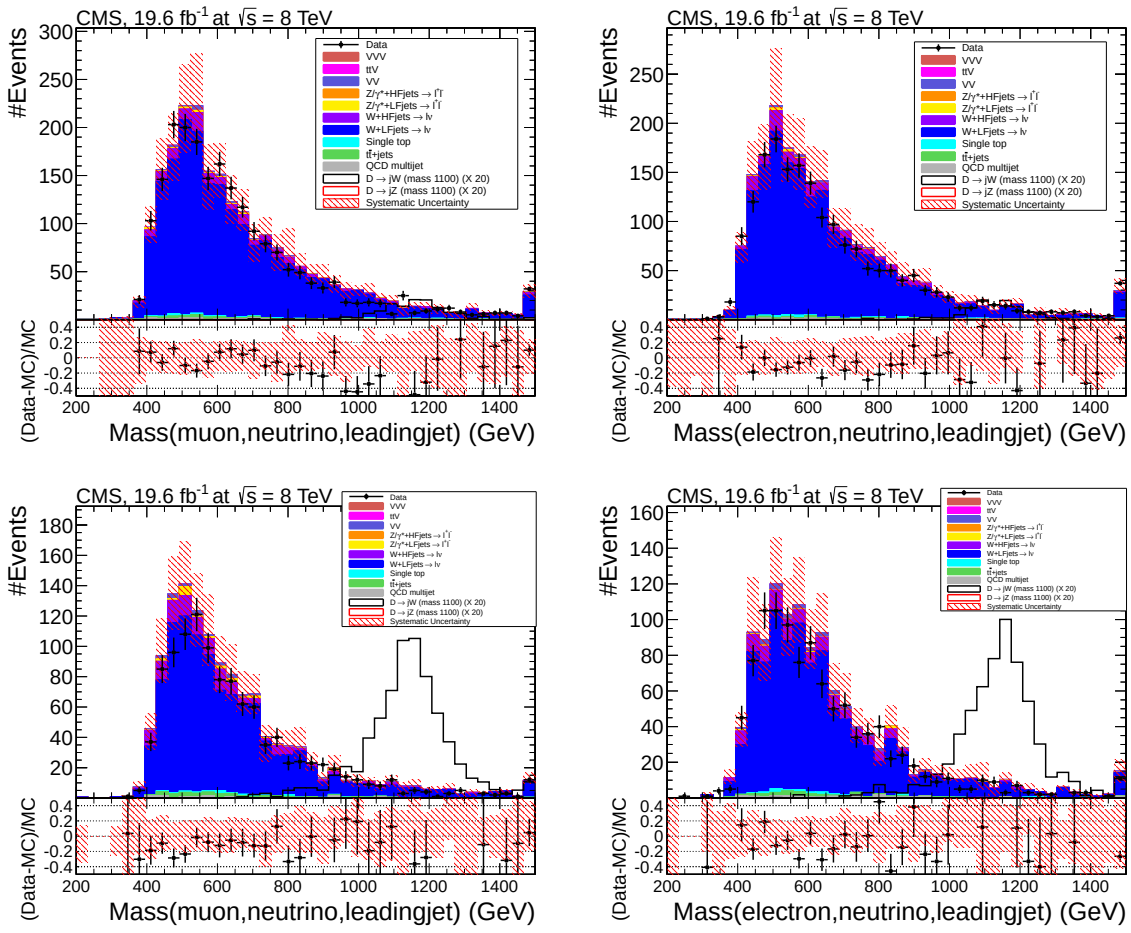

Figure 5.12: The reconstructed mass of the vector-like quark in the W^+qq subsample (top) and the W^-qq subsample (bottom), in the muon channel (left) and the electron channel (right). The enhanced relative vector-like D quark signal contamination in the W^-qq subsample with respect to the W^-qq subsample is clear. The single-production reference signal is scaled up by a factor 20 for visibility.

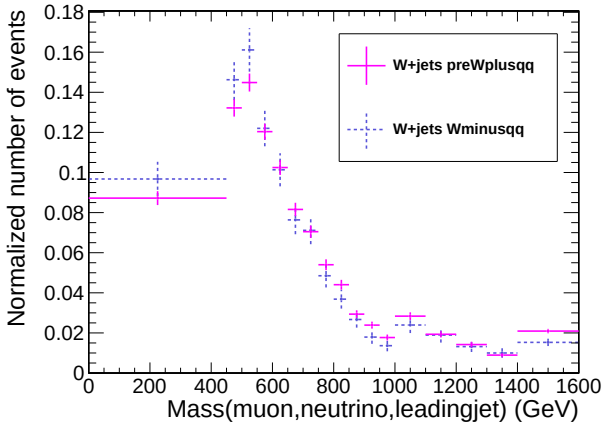

Figure 5.13: For the dominating W+jets background, the spectrum of the reconstructed mass of the vector-like quark in a W^+qq-like control sample (solid line, same selections as applied in the W^+qq category but with a forward-jet veto) is slightly harder than the spectrum in the W^-qq signal subsample (dashed line). Hence, a shape correction is applied when extrapolating the data distribution from the control to the signal subsample.

normalized control data sample. Finally, we add up the simulated SM background processes other than the W+jets and QCD multijet processes that pass the W^-qq signal category selection. The result is a shape and normalization corrected data-driven background distribution of the reconstructed vector-like quark mass. Figure 5.14 shows the resulting data-driven background distribution for the muon and electron channels, and compares this with the nominal distribution obtained from simulation and the data-driven distribution if one were not to correct for the shape difference in the control and signal subsample. Note that for the data-driven estimation of the mass distribution of a U quark, the roles of the subsamples are reversed, such that the distribution in the sensitive W^+qq subsample is estimated from a control W^-qq-like sample with a forward-jet veto.

The mass variable in the Zqq category

The Zqq subsample is defined to be sensitive to the presence of singly-produced vector-like quarks that decay to a Z boson. The Z-boson is reconstructed from the two oppositely-charged same-flavor leptons if they are within the Z-boson mass window. The right side of Figure 5.9 show that combining this Z-boson candidate with the leading jet in the event, we can reconstruct the vector-like quark and its mass. The reconstructed mass variable is shown for data and expectation in Figure 5.15 for the muon and the electron channel. The SM background is completely dominated by the Z+jets process, and the reconstructed mass for the singly-produced vector-like quark decaying to a Z boson has a sharp peak around the generated heavy-quark mass of 1100 GeV.

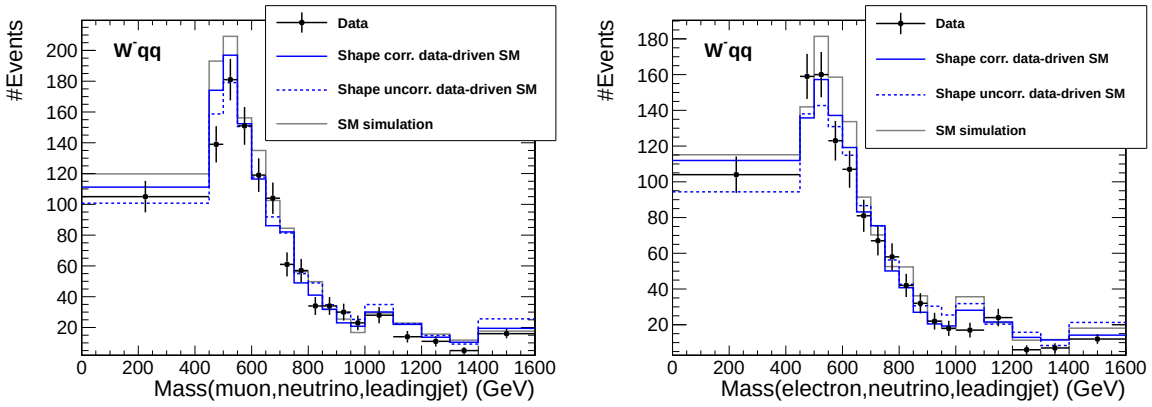

Figure 5.14: The data-driven background distribution (solid blue) of the reconstructed heavy-quark mass shows a reasonably good agreement with the observed data, in muon channel (left) and electron channel (right). The agreement is in general better than the distribution taken straight from simulation (grey), especially towards lower mass values, and comparable to the data-driven distribution without shape correction (blue dashed).

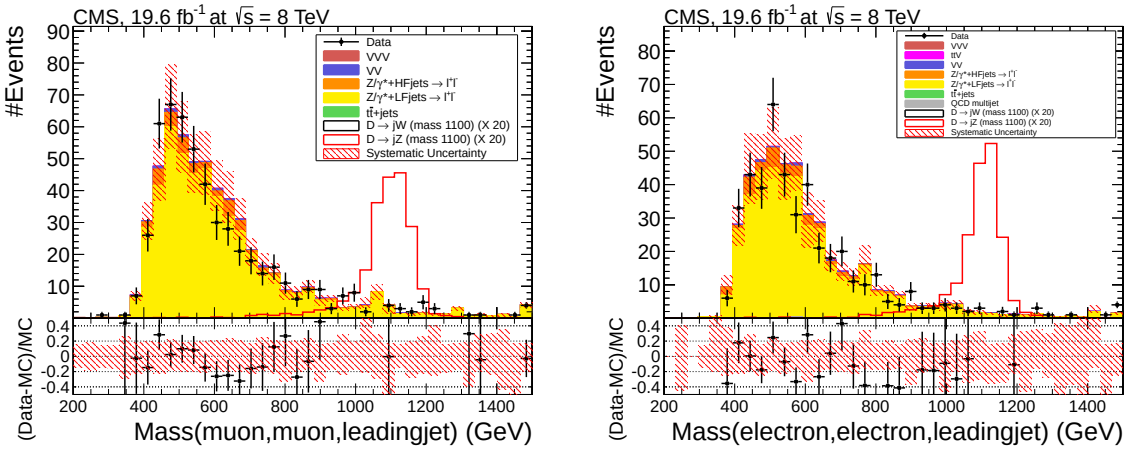

Figure 5.15: The reconstructed mass of the vector-like quark in the Zqq subsample, in the muon channel (left) and the electron channel (right). The single-production reference signal is scaled by a factor 20 for visibility.

The mass variable in the WHqq category

The selection criteria in the WHqq subsample are optimized for the pair production of vector-like quarks where one of the quarks decays to a W boson, and the other to a H boson. In the Wqq subsamples the neutrino is reconstructed by choosing the solution of the z component of the neutrino momentum such that the pseudorapidity difference of the neutrino with the leading jet is maximal. The final state in the WHqq subsample, however, involves *two* high-p_T jets, which increases the ambiguity in the neutrino reconstruction. Therefore we consider the other 'side' of the diagram for the mass reconstruction, as shown in Figure 5.16 at the left, where the H boson decays to two b quarks.

In the reconstruction of the H-boson candidate, we exclude the two leading anti-b-tagged jets from the list of available jets. There are three possible scenarios where we construct the H-boson candidate. In the first case, if there are at least two b-tagged jets, the two b jets with the smallest $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ are combined to form a H-boson candidate. In the second case, if there is only one b jet but other jets are available, the jet closest in ΔR to the b jet is chosen in the H-boson reconstruction. The mass of the H-boson candidate in these scenarios is shown for data and expectation in the left side of Figure 5.17. The signal corresponds to the pair production of vector-like quarks of mass 600 GeV (pair-production reference signal) where one decays to a H boson and the other to a W boson. If the H-boson candidate is found from two jets j_1 and j_2, the mass of the candidate is required to be in a 30 GeV window around the assumed H-boson mass of 125 GeV, $|m_{j_1 j_2} - m_H| < 30$ GeV. In the third possible case, if there are only three jets in the event (namely the two anti-b-tagged jets and the b jet), the b jet itself is considered as the H-boson candidate. This is motivated by the possibility that the H boson might be very boosted, with the decay products collimated in one jet. In this case we do not impose a requirement on $|m_j - m_H|$.

From the H-boson candidate and the anti-b-tagged high-p_T jet with the largest ΔR to the H-boson candidate, the vector-like quark candidate and its mass is reconstructed. From the right side of Figure 5.17, it can be seen that the distribution for the reconstructed mass of vector-like quarks of mass 600 GeV is more peaked with this choice of the anti-b-tagged high-p_T jet compared to the choice of the anti-b-tagged high-p_T jet with the *smallest* ΔR to the H-boson candidate. Figure 5.18 shows the resulting vector-like quark mass variable distributions in the muon channel and in the electron channel. The main SM backgrounds in this subsample are the $t\bar{t}$ and the W-boson in association with heavy-flavor jets processes.

The mass variable in the ZHqq category

In the ZHqq subsample, the Z boson originating from a vector-like quark decay is considered to decay to two oppositely-charged same-flavor leptons. Hence, these charged leptons are used to reconstruct the Z boson candidate, if they satisfy the invariant-mass constraint $|m_{\ell\ell} - m_Z| < 7.5$ GeV. The produced heavy quarks both decay to a high-p_T light-flavor jet, as depicted at the right side of Figure 5.16, and we choose the high-p_T anti-b-tagged jet with the largest ΔR separation from the Z-boson candidate in the vector-like quark mass reconstruction. The resulting mass variable is shown in

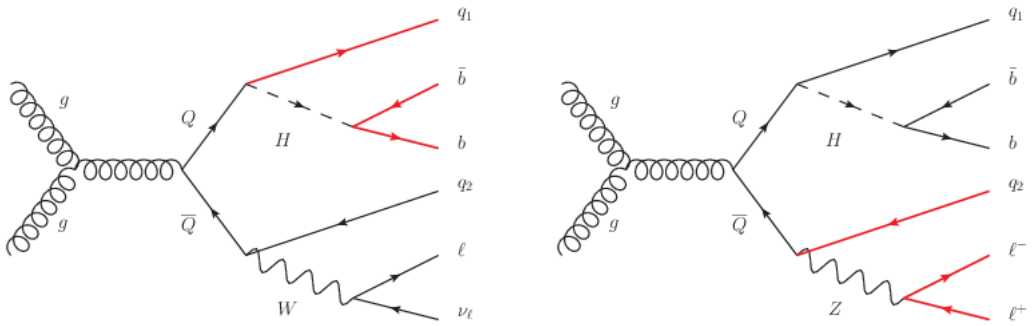

Figure 5.16: In the pair-production modes where one of the heavy quarks decays to a H boson, the mass of the vector-like quark is either reconstructed from the jet(s) originating from b quarks, and a high-p_T jet, in case the other heavy quark decays to a W boson (left), or from the charged-lepton pair and the leading jet when the other heavy quark decays to a Z boson (right).

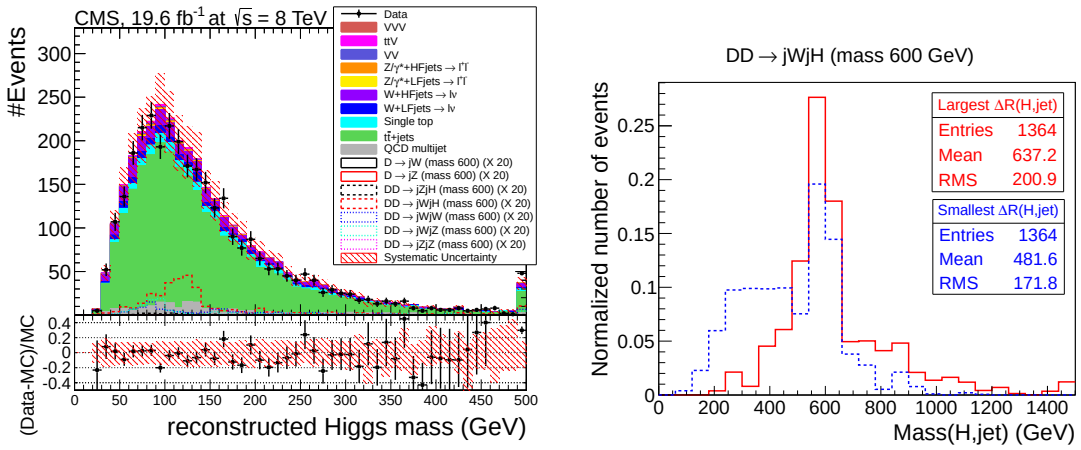

Figure 5.17: Left: The mass of the H boson is reconstructed in the WHqq subsample from two b-tagged jets, or a b-tagged jet and the jet closest in ΔR. A peak around the generated H-boson mass of 125 GeV is visible in the simulated pair-production signal of a vector-like quark decaying to a W boson and the other one decaying to a H boson. The pair-production reference signal is scaled up by a factor 20 for visibility. Right: Choosing the anti-b-tagged high-p_T jet with the largest ΔR separation to the H-boson candidate results in a distribution of the reconstructed heavy-quark mass that is peaked at the generated quark mass. Choosing the jet according to the smallest ΔR separation, broadens the distribution towards lower reconstructed mass values and would result in less discriminating power of the search observable in the WHqq subsample.

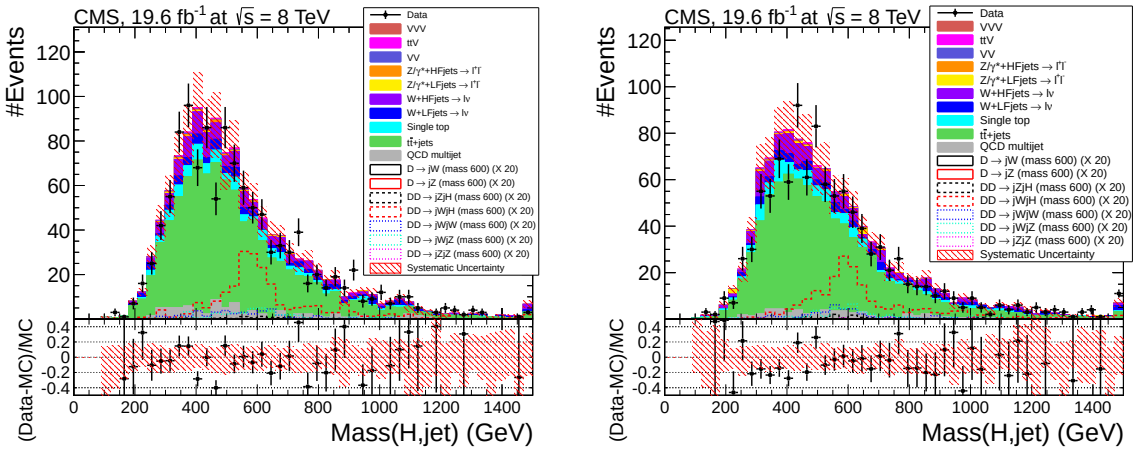

Figure 5.18: The reconstructed mass of the vector-like quark in the WHqq subsample, in the muon channel (left) and the electron channel (right). The pair-production reference signal is scaled up by a factor 20 for visibility.

Figure 5.19 for the muon channel and for the electron channel. The background consist mainly of Z+jets, with a large contribution from the component with associated heavy-flavor jets.

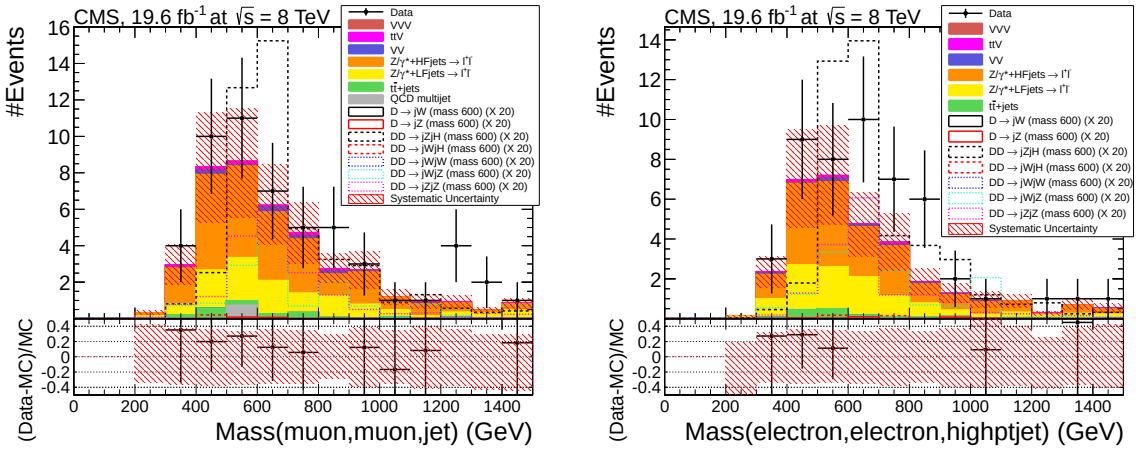

Figure 5.19: The reconstructed mass of the vector-like quark candidate in the ZHqq subsample, in the muon channel (left) and the electron channel (right). The pair-production reference signal is scaled up by a factor 20 for visibility.

The mass variable in the semi-leptonic VZqq category

The last subsample in which a mass variable is reconstructed is the subsample optimized for pair-produced heavy quarks where a Z boson originating from a heavy quark decays to two charged leptons, and a W or a Z boson decays to two jets. This is illustrated in Figure 5.20. The construction of the Z-boson candidate and the choice of the high-p_T anti-b-tagged jet is completely similar to the procedure in the ZHqq subsample. The

distribution of the mass variable reconstructed as such is shown in Figure 5.21 for the muon and electron channel.

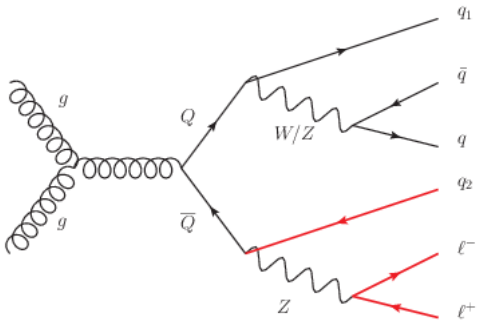

Figure 5.20: In the pair-production modes where one of the heavy quarks decays to a Z boson and the other one to a W or a Z boson that decays to two jets, the mass of the vector-like quarks is reconstructed from the charged-lepton pair and a high-p_T jet.

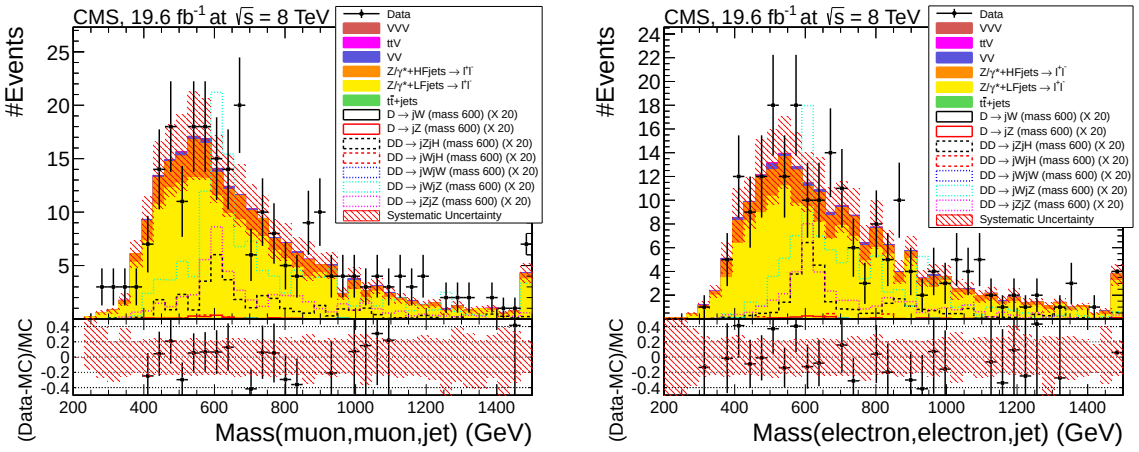

Figure 5.21: The reconstructed mass of the vector-like quark candidate in the semi-leptonic VZqq subsample, in the muon channel (left) and the electron channel (right). The pair-production reference signal is scaled up by a factor 20 for visibility.

5.6.2 The S_T variable and event counts

In the WWqq subsample we use the S_T variable, which is defined in this analysis as the scalar sum of the reconstructed physics objects in the final-state:

$$S_T = \not{E}_T + \sum p_T^{\ell} + \sum p_T^{jet}. \tag{5.19}$$

Because of the two neutrinos assumed in the topology of the WWqq subsample, as illustrated in the left diagram of Figure 5.22, the reconstruction of a mass variable

is highly ambiguous. Hence we use the S_T variable to discriminate between SM and vector-like quark processes. The data and SM expectation is compared in Figure 5.23 for the muon and electron channel.

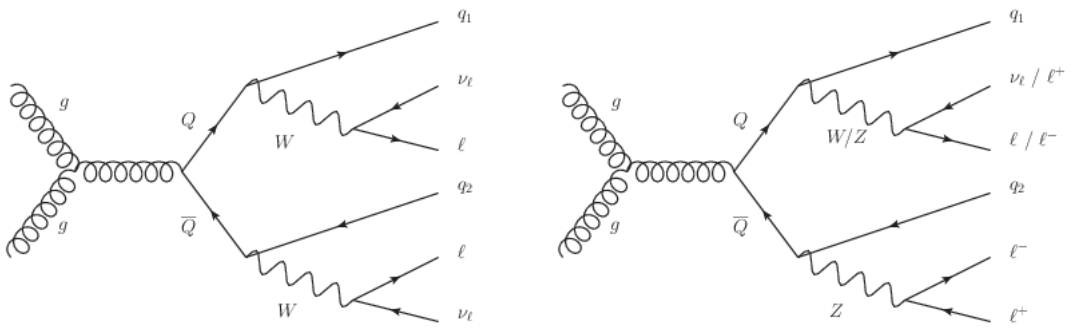

Figure 5.22: In the pair-production modes where both heavy quarks decay to a W boson that subsequently decays leptonically, the considered discriminating observable is the S_T variable, calculated as the scalar sum of the transverse momenta of the final-state objects (left). When one of the heavy quarks decays to a W or a Z boson, and the other to a Z boson, and only leptonic decays are considered, we just use the event counts as the observable, since the event yield is very low (right).

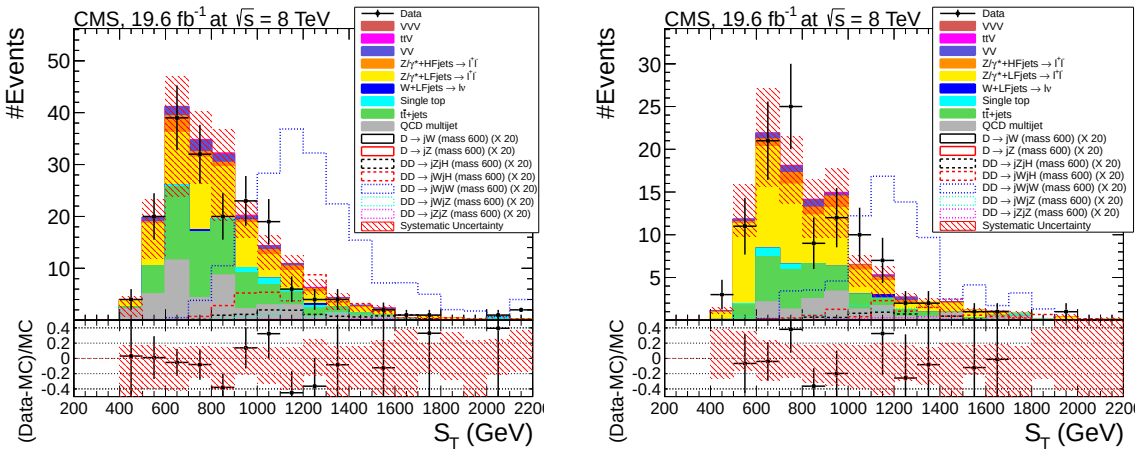

Figure 5.23: The S_T variable in the WWqq subsample, in the muon channel (left) and in the electron channel (right). The pair-production reference signal is scaled up by a factor 20 for visibility.

In the full-leptonic VZqq subsample the selected number of events is too low to obtain a sensible distribution of events, so we use the event counts as the discriminating observable. The assumed topology in this subsample is illustrated in the right side of Figure 5.22. The expected number of isolated charged leptons is three or four, and the number of events observed and expected are summarized in Table 5.11. The main SM background originates from irreducible diboson and triboson processes with three prompt charged leptons. A smaller contribution comes from fake leptons, as estimated

in Section 5.5. Note that the four-lepton event yield is even low for the pair-production reference signal, but this yield can increase for a larger decay branching fraction of the vector-like quark to a Z boson. For instance, for a branching fraction $\mathrm{BF}_Z = 1$, the four-lepton event yield would increase with a factor of 16 with respect to the reference scenario $\mathrm{BF}_Z = 0.25$ shown in Table 5.11.

Table 5.11: The prediction for the total number of background events compared to the number of observed events in the full-leptonic VZqq subsample. The numbers of expected signal events for a pair-production reference signal of vector-like quark mass 600 GeV is shown (corresponding to $\mathrm{BF}_W = 0.5$, $\mathrm{BF}_Z = 0.25$). The indicated uncertainties are statistical.

	3 leptons	4 leptons
Irreducible background	0.4 ± 0.2	0
Fake-lepton background	0.06 ± 0.06	–
Total background	0.5 ± 0.2	0
Observed	2	0
Pair ref. signal	2.6 ± 0.8	$0.1^{+0.7}_{-0.1}$

5.7 Inclusive search

We use the observables constructed in the previous section to search in an inclusive way for down-type vector-like quarks coupling to first-generation quarks. In Section 5.7.1, the combination of the distributions of the discriminating observables into template distributions is described. The systematic sources affecting these template distributions are discussed in Section 5.7.2, and a brief review of the limit setting procedure is given in Section 5.7.3.

5.7.1 Construction of template distributions

For technical purposes, we build two different one-dimensional distributions (templates); one for the muon channel and another one for the electron channel. These channels contain both the single-lepton and the dilepton subsamples. In the subsample with three and four leptons, we use event counts, resulting in a third 'template' of one bin. In each subsample, the distribution of the associated observable is binned such that the SM background expectation in each bin predicts at least about 10 events. This choice ensures that the asymptotic formulae for the test statistic distribution in the limit setting can be applied safely. It has been checked in various CMS analyses [75, 175] that the use of these asymptotic approximations give compatible results with full Monte Carlo pseudoexperiment techniques, even with bins with a low number of expected events.

A specific vector-like quark signal depends not only on the three free parameters from Equation (5.1), namely the strength of the coupling to the W boson, and the

branching fractions of the charged-current and neutral-current decay modes, but also on the mass of the vector-like quark. For given values of these four parameters, the yield and the shape of the variable distributions of a signal process are determined. Hence, the total signal template for a specific model point is obtained from the weighted sum of the separate possible processes. Here the charged-current and neutral-current single-production process cross sections are weighted by $\tilde{\kappa}_W^2$ or $\tilde{\kappa}_Z^2$, respectively. The reason is that an interaction vertex in the matrix-element Feynman diagram involves the $\tilde{\kappa}$ coupling parameter, and the cross section is obtained from the modulus squared of the matrix element. The pair production cross section remains unaffected by the choice of these strengths. Note that if $\tilde{\kappa}_W \neq 0$, the value of $\tilde{\kappa}_Z$ is determined from the assumed value of $\tilde{\kappa}_W$, and the branching fractions to a W and a Z boson, via Equation (5.9). Next, the branching fractions to a W, Z or H boson are used to weight the simulated signal samples to obtain the full assumed signal model. Hence, the total signal can be schematically represented as:

$$\begin{aligned}
\text{Total signal} = {} & \tilde{\kappa}_W^2 \ \mathrm{BF}_W \times (pp \xrightarrow{W^*} Qq \to Wqq) \\
& + \tilde{\kappa}_Z^2 \ \mathrm{BF}_W \times (pp \xrightarrow{Z^*} Qq \to Wqq) \\
& + \tilde{\kappa}_W^2 \ \mathrm{BF}_Z \times (pp \xrightarrow{W^*} Qq \to Zqq) \\
& + \tilde{\kappa}_Z^2 \ \mathrm{BF}_Z \times (pp \xrightarrow{Z^*} Qq \to Zqq) \\
& + \mathrm{BF}_W^2 \times (pp \to QQ \to WqWq) \\
& + 2 \ \mathrm{BF}_W \ \mathrm{BF}_Z \times (pp \to QQ \to WqZq) \\
& + 2 \ \mathrm{BF}_W \ \mathrm{BF}_H \times (pp \to QQ \to WqHq) \\
& + \mathrm{BF}_Z^2 \times (pp \to QQ \to ZqZq) \\
& + 2 \ \mathrm{BF}_Z \ \mathrm{BF}_H \times (pp \to QQ \to ZqHq)
\end{aligned} \tag{5.20}$$

Figure 5.24 illustrates the template one obtains by joining together the single-lepton and dilepton subsamples. Here the observed data is compared with the total SM expectation, and the single-production (m_Q = 1100 GeV) and pair-production (m_Q = 600 GeV) reference signal for vector-like D quarks is overlayed.

5.7.2 Systematic uncertainties

The sources of systematic effects that potentially alter the template distributions are outlined below. Most of them have already been introduced in Section 4.6.2.

- **Integrated luminosity.** The uncertainty in the pixel-based luminosity measurement of the integrated luminosity of the collected data is 2.6% [176]. This systematic source has the same normalization effect on all template distributions.
- **Normalization of the different background processes.** The dominant background processes in the single-lepton and dilepton subsamples are the W+jets, Z+jets and top-quark pair production. The uncertainties in the light-flavor component of the W+jets, the Z+jets production processes and the data-driven

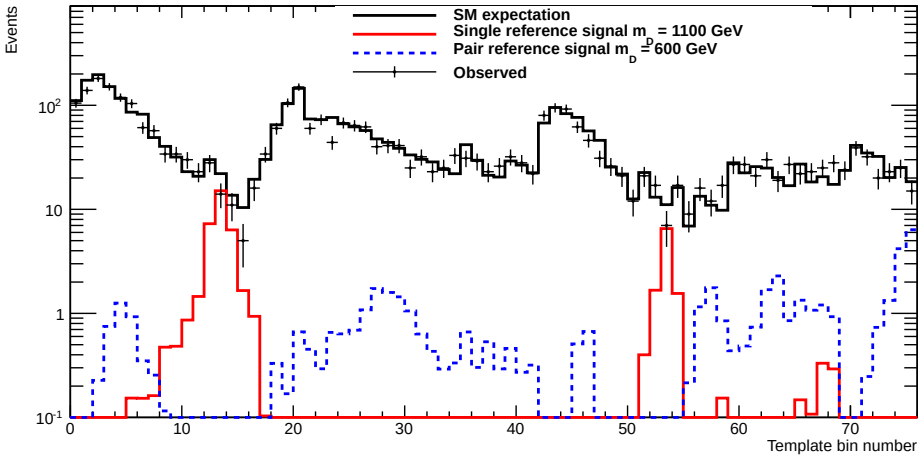

Figure 5.24: The combined one-dimensional histogram of the search variables in the six single-lepton and dilepton subsamples of the muon channel. The total Standard-Model contribution is indicated in black, the observed data is superimposed, and the expectations for two reference D-quark signal scenarios are overlayed. From left to right, the following subsamples can be distinguished: W^-qq (bin numbers 0 to 16), WHqq (bin numbers 17 to 41), Zqq (bin numbers 42 to 54), ZHqq (bin numbers 55 to 58), semi-leptonic VZqq (bin numbers 59 to 68) and WWqq (bin numbers 69 to 75).

background in the Wqq subsamples, are taken as the difference in the fitted cross section value (see Section 5.5) and the NLO theoretical production cross section. This results in a normalization uncertainty on 14% for these simulated or estimated processes. The uncertainty in the normalization of the heavy-flavor component of W+jets and Z+jets is taken to be 50%. The uncertainty in the measured value of the top-quark pair production cross section from the CMS experiment is 14%. The single-top cross section uncertainty from NLO calculations is 3.4%. The uncertainty in the measured WW production cross section by CMS is taken to be 10%, and for both WZ and ZZ diboson processes, the uncertainties from the theoretical NLO calculations are found to be 4%. The same-sign $W^{\pm}W^{\pm}$ background contribution in the analysis is small, and is merged in the statistical procedure with the WW process. The NLO calculations of the $t\bar{t}+W$ and $t\bar{t}+Z$ cross sections result in uncertainties of 32% and 12%, respectively. The conservative uncertainty in the fitted normalization of the QCD multijet background process is taken to be 100%.

- **Jet energy scale.** The energy scale of the jets and the missing transverse energy affect the event selection and the reconstruction of the search variables in the defined subsamples. The energy scale of the reconstructed jets is varied to a lower and higher value according to the estimated p_T and η dependent uncertainties [137]. In this procedure, the missing transverse energy is altered as well. This results in two new template distributions for each physics process, one corresponding to the upper scaling of the jet energy scale by one standard

deviation, and the other to the lower scaling.

- **Jet energy resolution.** In the nominal template distributions, the jet resolution in the simulation was smeared in order to match the resolution obtained from the observed data. The estimation in data of the jet energy resolution and its uncertainty depends on the pseudorapidity of the jets. We construct templates with a $\pm 1\sigma$ variation of the jet energy resolution. The resolution of the missing transverse energy is also altered with this procedure.

- **Unclustered missing transverse energy.** The missing transverse energy in the event is reconstructed from the balance of the energy deposits in the detector. Most of the energy deposited is clustered in jets or in charged-lepton objects, but some fraction is not clustered. This unclustered missing transverse energy component can be obtained by subtracting the reconstructed jets and leptons in the event from the missing transverse energy. The remaining energy is varied up and down by $\pm 10\%$ to evaluate the $\pm 1\sigma$ variations of the unclustered energy. The resulting effect on the search variable distribution has been found to be negligible, inducing a global decrease or increase of the selected event yield much below 1%. As a consequence, we do not propagate this systematic uncertainty in the statistical limit setting procedure.

- **Scale factor for non-b-tag efficiency (mistag rate).** Since we use anti-b-tagging in all subsamples, as well as b-tagging in some of them, this analysis is sensitive to the difference in efficiency of b-tagging jets originating from light-flavor jets in data and simulation. We vary the scale factors that account for these differences with one standard deviation up and down, to obtain the corresponding systematically shifted templates.

- **Scale factor for b-tag efficiency.** The application of b-tagging and anti-b-tagging in the various subsamples results in a source of systematic uncertainty related to the probability that a jet originating from a b quark is identified as a b jet. Templates corresponding to $\pm 1\sigma$ variations of the b-tag efficiency scale factors are obtained by shifting the scale factors by their measured uncertainties.

- **Pile-up reweighting.** The distribution of the number of proton-proton interactions in the simulated events has been reweighted to match the distribution derived from the data. This measurement depends on the instantaneous luminosity durig data taking and on the assumed total inelastic cross section of proton-proton collisions. The mean number of primary vertices in the simulated samples are varied by $\pm 6\%$ to evaluate the systematic effect originating from the imperfect knowledge of these quantities. The level of the distributions is found to shift with less than 1%, and since other larger systematic uncertainties (e.g. in the luminosity and the used cross section values) easily cover such normalization effects of the templates, the systematic uncertainty on the pile-up reweighting is not evaluated during the limit setting.

- **Muon and electron trigger, identification and isolation efficiency.** The scale factors that are applied as weights to the simulated events in order to ac-

count for differences in muon and electron trigger, identification and isolation efficiency in data and simulation are varied by their $\pm 1\sigma$ uncertainties. These scale factors and uncertainties depend on the transverse momentum and the pseudorapidity of the leptons, and as a result both the shape and the normalization of the templates are affected.

- **Shape correction of data-driven reconstructed mass distribution in the W^-qq event category.** As explained in Section 5.6.1, the reconstructed mass distribution in the W^-qq category is estimated from data. However, the estimation is still partly driven by simulation, since a shape-correction is applied to extrapolate the data distribution from a W^+qq-like control region to the W^-qq signal region. This 'nominal' correction, which has been determined from simulated W+jets events, may potentially be mismodeled. Hence, we construct two systematically shifted template distributions for the data-driven reconstructed mass distribution in the W^-qq subsample; a distribution without any shape correction performed, and another one for which the effect of the nominal shape correction is doubled. A nuisance parameter is assigned to this systematic uncertainty, such that the size of the correction is fitted in the limit setting procedure.

- **Fake lepton estimation.** The uncertainty in the data-driven background estimation for trilepton events where one of the leptons is a fake lepton is listed in Table 5.5. This results in an uncertainty in the normalization of the prediction in the full-leptonic VZqq subsample.

- **Signal yield.** As mentioned in Section 3.2, we observe deviations less than 15% when comparing kinematic distributions of physics objects for fast-simulated vector-like quark events and full-simulated events. In addition, for the single production processes we use LO cross sections, which may differ from more accurate higher-order calculations (that are not available in the literature). To cover the potential discrepancies, a systematic uncertainty of 20% is set on the normalization of the combined signal yield.

When using template distributions constructed from a limited number of simulated events, one can be affected by statistical fluctuations that are not representative for the systematic source being evaluated. This could be especially the case in the subsamples dominated by processes for which simulated event samples are available with an integrated luminosity equivalent to or lower than the integrated data luminosity. From Table 5.2 and considering the background composition in the subsamples, statistical effects on the systematic template distributions are expected to be the largest for the Z+jets process in the Zqq subsample, but also for the $t\bar{t}$ background in the WHqq subsample. For the jet energy scale as well as the jet energy resolution for these processes (Z+jets in the Zqq subsample and $t\bar{t}$ in the WHqq subsample), we consider the ratio of the nominal search variable distributions with the systematically shifted distributions. These uncertainties result mainly in a global difference of selected events, as the acceptance of the selected events increases or decreases according to the chosen jet energy scale and resolution. This can be seen in Figure 5.25, where a uniform distribution is used to fit the ratio of the nominal over the systematically shifted distributions. Note

that the resulting normalized χ^2 of these fits are low because of statistical correlations between bins[4], but the general tendency is indeed compatible with a uniform distribution. The fitted effects are used to evaluate the corresponding systematic uncertainties in the statistical procedure.

For all background contributions other than Z+jets in the Zqq subsample and $t\bar{t}$ in the WHqq subsample, the systematic uncertainties in the jet energy scale and in the jet energy resolution are evaluated using the systematically shifted template distribtuions, without performing uniform fits. Furthermore, for the data-driven W+jets and QCD multijet background estimations in the Wqq category, evaluating systematic uncertainties that aim to describe the differences between data and simulation is not needed, because this background is determined in a data-driven way.

5.7.3 Limit setting procedure

In this analysis, we follow a similar statistical procedure as for the chiral fourth-generation search described in Section 4.6.3. The template distributions constructed in Sections 5.7.1 and 5.7.2 are used to test the compatibility of the collected data with the Standard-Model background-only hypothesis and the background plus vector-like quarks signal hypothesis.

The probability model has the same form as in Equation (4.7). The parameter of interest in the statistical model is the total signal strength μ, depending on some assumed set of model parameters (the single-production strengths, the branching fractions and the mass of the vector-like quark). We use the profile likelihood ratio test statistic to derive one-sided upper limits on the signal strength. In this procedure, the nuisance parameters are fitted to the data, both simultaneously with the signal strength (resulting in fitted values $\hat{\boldsymbol{\alpha}}$ and $\hat{\mu}$) and separately (resulting in fitted values $\hat{\hat{\boldsymbol{\alpha}}}$ for a fixed assumed value of μ). Via asymptotic approximations of the distributions of the profile likelihood ratio test statistic for the background-only hypothesis and the background-plus-signal hypothesis, a 95% CL exclusion limit on the signal strength can be derived via the CLs method. This limit can then be translated into an exclusion limit on the specific probed set of model parameters. If the excluded combined signal strength is smaller than the theoretically predicted strength, the model point is said to be excluded at the 95% CL.

5.8 Results and discussion

To compare the observed data to the SM background-only model and the background-plus-signal model, we use the templates of the observed data as well as the nominal and systematically shifted SM background and vector-like quark signal simulation. Since we do not observe a significant data excess with respect to the SM expectation, we set 95% CL exclusion limits on vector-like quark model parameters. Using the profile likelihood

[4]When evaluating the systematic uncertainties on the jet energy scale and resolution, the same simulated events are used in the nominal and the systematically shifted distributions. Hence, the fit points are not distributed truly randomly from one bin to another, such that they agree too well with the fitted hypothesis, resulting in a low χ^2/ndf of the fit.

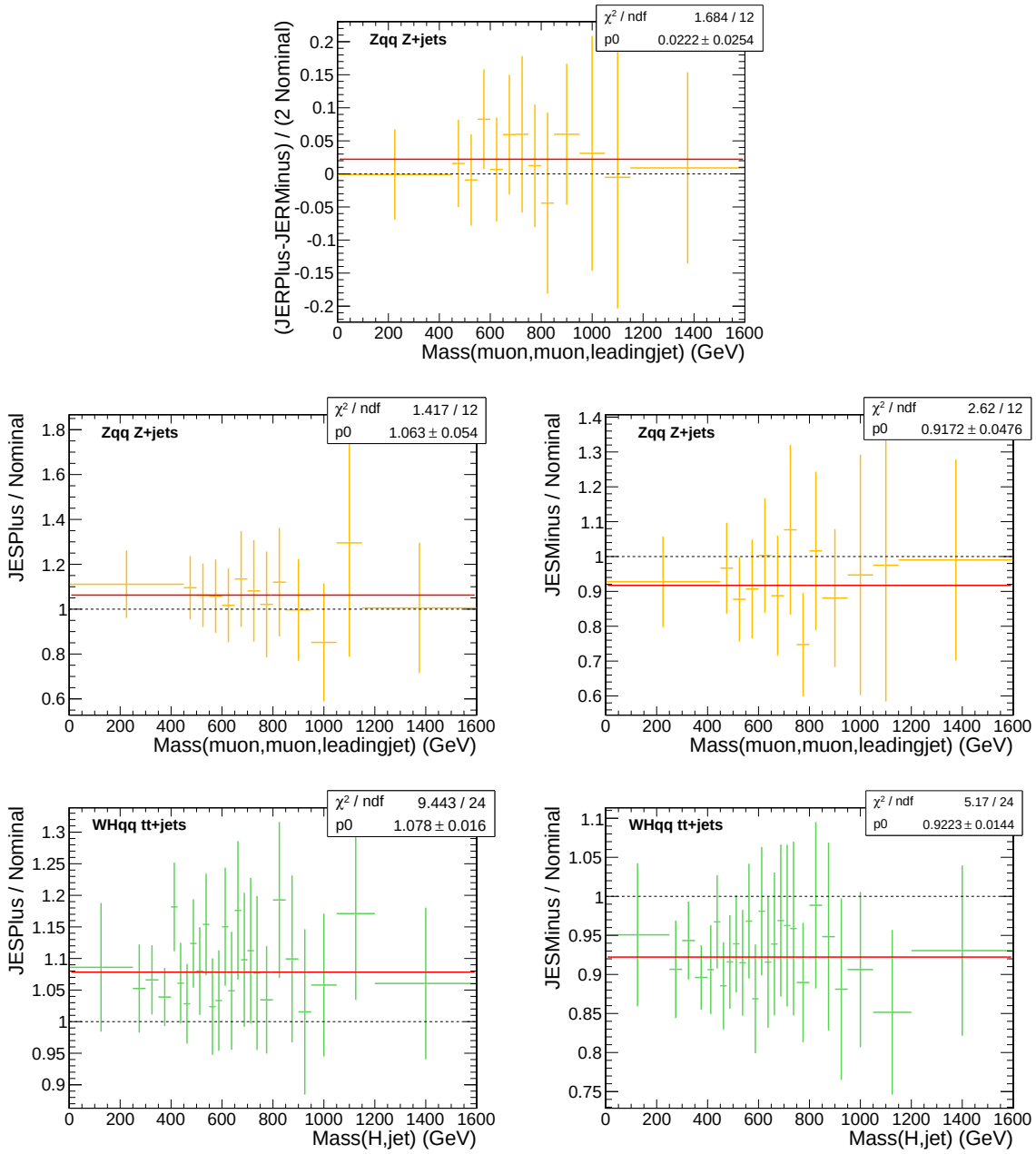

Figure 5.25: The background templates corresponding to the systematically shifted jet energy scale and resolution variations are affected by stastistical fluctuations. Therefore uniform fits are performed on the ratio of the relevant systematically shifted template and the nominal template. This is shown for the average jet energy resolution variation (top) and jet energy scale up and down variation (middle left and right, respectively) for the Z+jets process in the Zqq subsample. Fits on the up and down variation of the jet energy scale on the $t\bar{t}$ background in the WHqq subsample are shown as well (bottom left and right). These fits are used later on to describe the $\pm 1\sigma$ effect of these systematic sources in the statistical limit setting procedure.

ratio test statistic, the nuisance parameters related to the systematic uncertainties in the analysis are fitted to the data, and the results of a background-only fit are presented in Section 5.8.1. We discuss the exclusion limits on down-type vector-like quarks in Section 5.8.2. These limits are compiled in so-called branching fraction triangle plots, for different values of the charged-current single-production strength $\tilde{\kappa}_W$. Finally, in Section 5.8.3 we briefly summarize the analysis approach and the main results.

5.8.1 Fitted background model

By performing a background-only fit of the nuisance parameters, one can study the overall agreement of the data with the SM expectation. In practice, this can be obtained by maximizing the likelihood function (4.9) while fixing the signal strength μ to 0. The resulting conditional maximum-likelihood estimates for the nuisance parameters $\boldsymbol{\alpha}$ are denoted by $\hat{\hat{\boldsymbol{\alpha}}}(\mu = 0)$. Table 5.12 summarizes these fitted nuisance parameters, which are expressed in number of standard deviations σ, corresponding to the various systematic uncertainties considered in this analysis. Note that some of the systematic sources (e.g. the jet energy scale and the normalization of the Z+jets process) are relevant to multiple subsamples, while other sources (e.g. the fake-lepton estimation, and the uncertainty in the shape of the data-driven mass distribution in the $\mathrm{W^-qq}$ subsample) only affect one subsample. Where possible, the fitted effect on the normalization of the corresponding nominal template is indicated.

The largest fitted systematic effect ($+1.2\ \sigma$) corresponds to the heavy-flavor component of the Z+jets process. This reflects the moderate data excess in the ZHqq subsample in Figure 5.19. However, it should be emphasized again that the numbers in Table 5.12 should not be interpreted as measurements of the corresponding cross sections or calibrations, since the used selection in the analysis does not aim to create optimized phase-space regions for such measurements. Moreover, the simultaneous fit of the whole set of nuisance parameters may cause a non-trivial interplay between the systematic sources. For example, a specific value of a nuisance parameter can be preferred because it improves the overall agreement between the observed data and the background, but this does not necessarily imply that the fitted value is optimal for one exclusive subsample.

5.8.2 Exclusion limits for down-type vector-like quarks decaying to light quarks

As mentioned in Section 5.7.1, one can obtain the total expected vector-like quark signal via Equation (5.20) for a specific choice of model parameters: the charged-current single-production strength ($\tilde{\kappa}_W$), and the branching ratios to a W boson (BF_W), and to a Z boson (BF_Z). By scanning the possible mass values of the vector-like quark and computing for each mass the signal strength that is excluded at 95% CL, we can infer a lower limit at 95% CL on the heavy quark mass.

Table 5.12: Overview of the nuisance parameters (describing the effects of the systematic uncertainties) fitted to data with a background-only model. For convenience, the fitted values of systematic sources affecting the normalization of the templates are transformed in the third column into relative shifts on the relevant physics parameter. The nuisance parameters α_{WLF} and α_{ZLF} denote the nuisance parameters corresponding to the light-flavor component of the W+jets and Z+jets process, respectively, while α_{WHF} and α_{ZHF} correspond to the heavy-flavor components.

Nuisance parameter	Fitted $\hat{\hat{\alpha}}(\mu = 0)$	Fitted effect
integrated luminosity	/	-0.1%
$\alpha_{t\bar{t}}$	$+0.003\ \sigma$	-0.04%
α_{WLF}	$-0.30\ \sigma$	-4.2%
α_{WHF}	$-0.19\ \sigma$	-9.5%
α_{ZLF}	$-0.59\ \sigma$	-8.3%
α_{ZHF}	$+1.2\ \sigma$	$+60\%$
α_{t}	$-0.04\ \sigma$	-0.1%
α_{WW}	$+0.09\ \sigma$	$+0.9\%$
α_{WZ}	$+0.04\ \sigma$	$+0.2\%$
α_{ZZ}	$+0.009\ \sigma$	$+0.04\%$
$\alpha_{t\bar{t}W}$	$+0.002\ \sigma$	$+0.06\%$
$\alpha_{t\bar{t}Z}$	$+0.01\ \sigma$	$+0.12\%$
α_{VVV}	$+0.20\ \sigma$	$+10\%$
α_{QCD}	$+0.00\ \sigma$	$+0\%$
α_{JES}	$+0.33\ \sigma$	/
α_{JER}	$+0.39\ \sigma$	/
α_{btag}	$-0.58\ \sigma$	/
α_{mistag}	$+0.41\ \sigma$	/
$\alpha_{muon\ id}$	$-0.02\ \sigma$	/
$\alpha_{electron\ id}$	$-0.06\ \sigma$	/
$\alpha_{fake\ lepton}$	$+0.15\ \sigma$	$+15\%$
$\alpha_{Wqq\ shape}$	$-0.31\ \sigma$	/

Results for $\tilde{\kappa}_W = 1$

The inference of a mass limit from a limit on the signal strength is illustrated in Figure 5.26, for the reference choice of model parameters $\tilde{\kappa}_W = 1$, $\mathrm{BF}_W = 0.5$ and $\mathrm{BF}_Z = 0.25$ at the left side, and for the case where only a vector-like quark coupling to the W boson is allowed, $\tilde{\kappa}_W = 1$ and $\mathrm{BF}_W = 1$, at the right side. The median, $\pm 1\sigma$ and $\pm 2\sigma$ expected exclusion limits are indicated with a dashed line, and green and yellow bands, respectively, while the observed limit is represented by black points connected by a solid line. A signal strength of 1 corresponds to the expected yield of the combined signal processes, using the theoretical cross-section calculations. Hence, when the 95% CL exclusion limit on the signal strength is lower than 1, the corresponding mass is excluded at 95% CL. A heavy vector-like D quark with a mass below 1585 GeV (1440 GeV) is observed (expected) to be excluded at 95% CL with the reference choice of model parameters: $\tilde{\kappa}_W = 1$, $\mathrm{BF}_W = 0.5$ and $\mathrm{BF}_Z = 0.25$. In case the vector-like quark only couples to the W boson, a D quark below mass 1740 GeV (1565 GeV) is observed (expected) to be excluded at 95% CL. The latter limits are more stringent than those derived by the ATLAS experiment at 7 TeV, which excluded a down-type vector-like quark decaying to a W boson and a light quark below 1120 GeV (1160 GeV expected) [72].

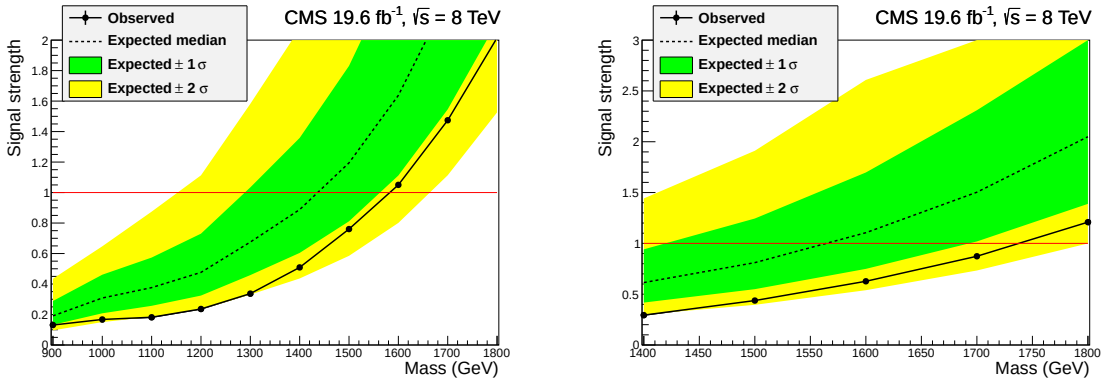

Figure 5.26: The 95% CL exclusion limits on the signal strength corresponding to different sets of model parameters ($\tilde{\kappa}_W = 1$, $\mathrm{BF}_W = 0.5$, $\mathrm{BF}_Z = 0.25$ at the left, and $\tilde{\kappa}_W = 1$, $\mathrm{BF}_W = 1$ at the right) as a function of the hypothetical D-quark mass. The median expected and the observed exclusion limits are indicated with a dashed and a solid line, respectively. The $\pm 1\sigma$ and $\pm 2\sigma$ expected exclusion limits are represented by the inner (green) and outer (yellow) bands around the median.

Next, one can apply this procedure to each model point while scanning the branching fractions to the different bosons in steps of 0.1. The resulting 95% CL exclusion limits on the heavy quark mass can be summarized in a branching fraction triangle. Each point in such a triangle corresponds to a unique set of branching fractions (BF_W, BF_Z and BF_H). In the corner points of the triangle, the associated branching fraction is 1, while the others are 0. The expected (observed) limit contours are shown in the top (bottom) plot of Figure 5.27, for $\tilde{\kappa}_W = 1$. Because $\tilde{\kappa}_W$ is assumed to be equal to 1 in these figures, the probed signal is dominated by single-production processes. The

higher the branching fraction to a W boson, the higher the limit on the mass. This is due to the signal event yield that is increased in the Wqq subsample, designed to be sensitive to the single production of heavy quarks decaying to a W boson and a light quark. In a similar way, a higher branching fraction to a Z boson increases the mass limits, because the signal events yield in the Zqq subsample is increased. The increase in sensitivity is higher towards the $\mathrm{BF}_Z = 1$ corner than towards the $\mathrm{BF}_W = 1$ corner, because the mass reconstruction in the Zqq subsample is more sharply peaked for the signal than in the Wqq subsample (see Figures 5.12 and 5.15). Hence, this provides more discriminating power between the SM background and the vector-like quark signal. For an increased decay rate to H bosons, the sensitvity and the lower mass limits decrease drastically. The reason for this is that we did not construct a subsample that would be particularly sensitive to the single production of vector-like quarks decaying to a H boson. This is due to the fact that the branching fraction of a H boson to leptonic final states is very low. The white shaded region in the observed triangle indicates the mass limit cannot be evaluated, as it is lower than 500 GeV and no simulated signal samples are produced for this mass.

The black shaded region below $\mathrm{BF}_W \approx 0.1$ in the branching fraction triangles indicates the region where one should be careful with the interpretation. In this region BF_W approaches 0, but since $\tilde{\kappa}_W$ is fixed to 1, this implies via Equation (5.9) that the neutral-current single-production strength parameter $\tilde{\kappa}_Z$ diverges, and when BF_W is exactly equal to 0, Equation (5.9) is not applicable anymore.

Results for $\tilde{\kappa}_W = 0.7$ and $\tilde{\kappa}_W = 0.4$

The same limit-setting procedure can be applied for different choices of the $\tilde{\kappa}_W$ parameter. For a choice of $\tilde{\kappa}_W = 0.7$ and $\tilde{\kappa}_W = 0.4$, the charged-current single-production signal cross section, proportional to $\tilde{\kappa}_W^2$, decreases by approximately a factor 2 and a factor 6, respectively. As a result, the total signal yield as well as the lower limits on the mass decrease. This is shown in the expected (top) and observed (bottom) triangle plots in Figure 5.28 for $\tilde{\kappa}_W = 0.7$ and Figure 5.29 for $\tilde{\kappa}_W = 0.4$. The tendency of increasing limits towards the corners corresponding to $\mathrm{BF}_W = 1$ and $\mathrm{BF}_Z = 1$ is still present. It should be noted that pair-produced vector-like quarks do not contribute to mass limits above 1000 GeV, because the signal yield of pair-production processes for masses above 1000 GeV is totally negligible. The effect of pair production is only expected to become relevant towards the right bottom corner of these triangles, at least if the branching fraction to a W or Z boson is still high enough to contribute to the signal event yield in the WHqq and ZHqq subsamples, respectively. Due to fluctuations in the observed data, the observed exclusion limit contours are not expected to be as smooth as the expected limit contours. The different model points in the branching fraction triangles are all sensitive to another composition of signal processes. Therefore the relevant subsamples and background composition differs, resulting in a different set of fitted nuisance parameters and fitted signal strength. In general the observed limits are within the 2σ band around the median expected limits, except in a few occasions over the 55 branching-fraction scan points per triangle.

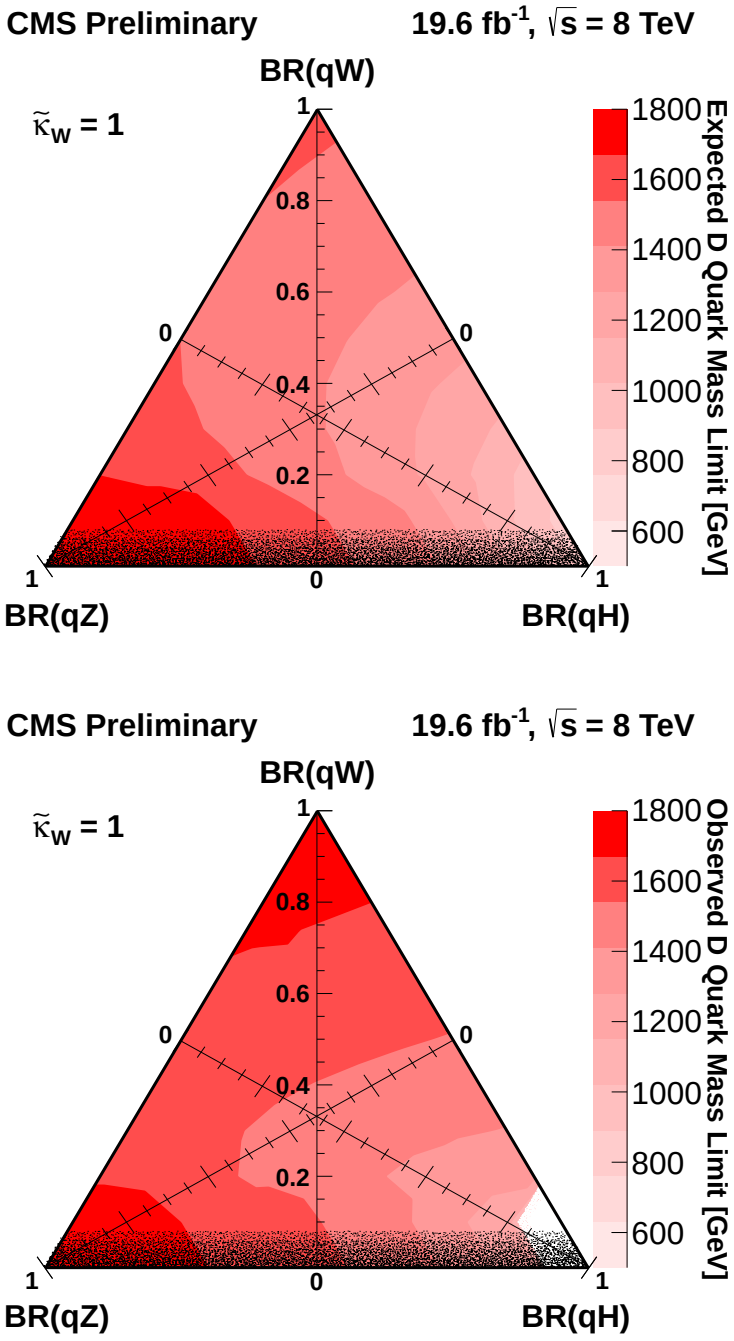

Figure 5.27: The expected (top) and observed (bottom) lower mass limits can be represented in triangles, where each point in the triangle corresponds to a given set of branching fractions ($BF_W \equiv$ BR(qW), $BF_Z \equiv$ BR(qZ), $BF_H \equiv$ BR(qH)). The shown limit contours correspond to a choice of $\tilde{\kappa}_W = 1$, such that the relevant signal is dominated by electroweak single production.

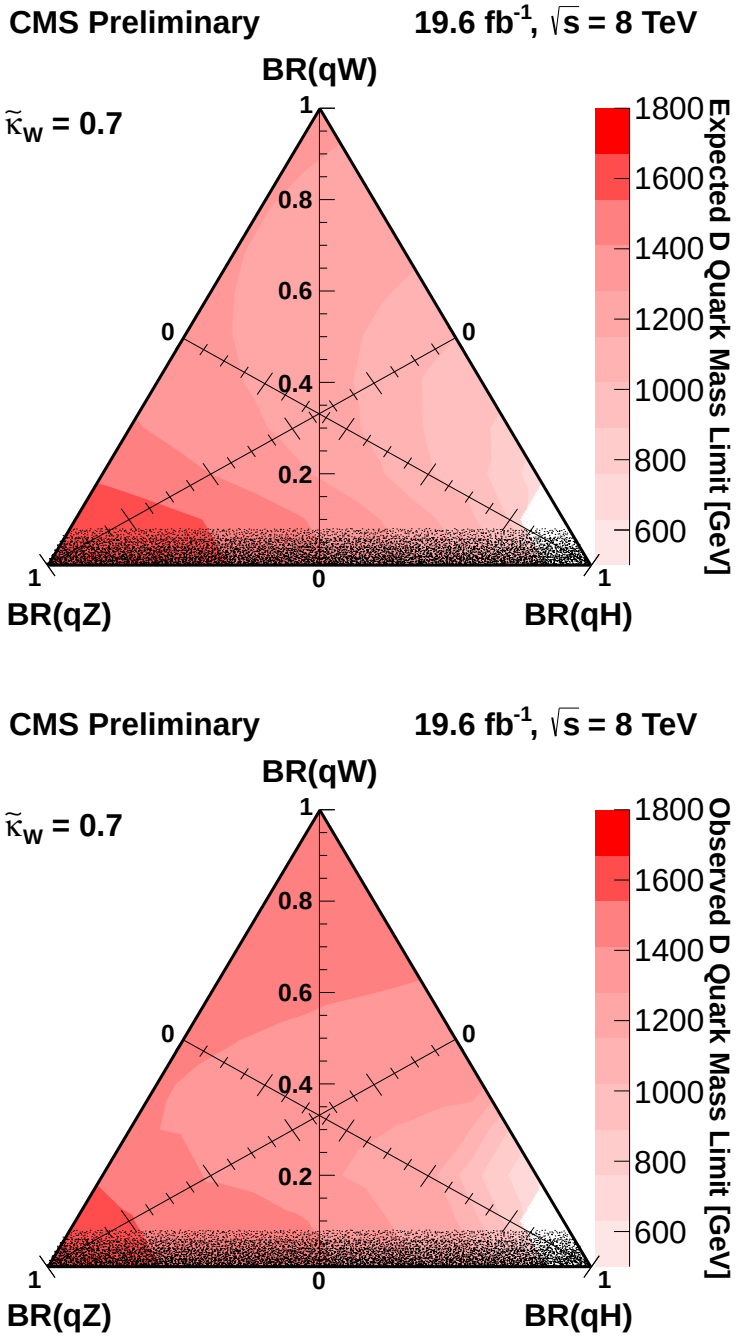

Figure 5.28: The expected (top) and observed (bottom) lower mass limits can be represented in triangles, where each point in the triangle corresponds to a given set of branching fractions ($BF_W \equiv BR(qW)$, $BF_Z \equiv BR(qZ)$, $BF_H \equiv BR(qH)$). The shown limit contours correspond to a choice of $\tilde{\kappa}_W = 0.7$, such that the relevant signal is dominated by electroweak single production in most of the triangle parameter space.

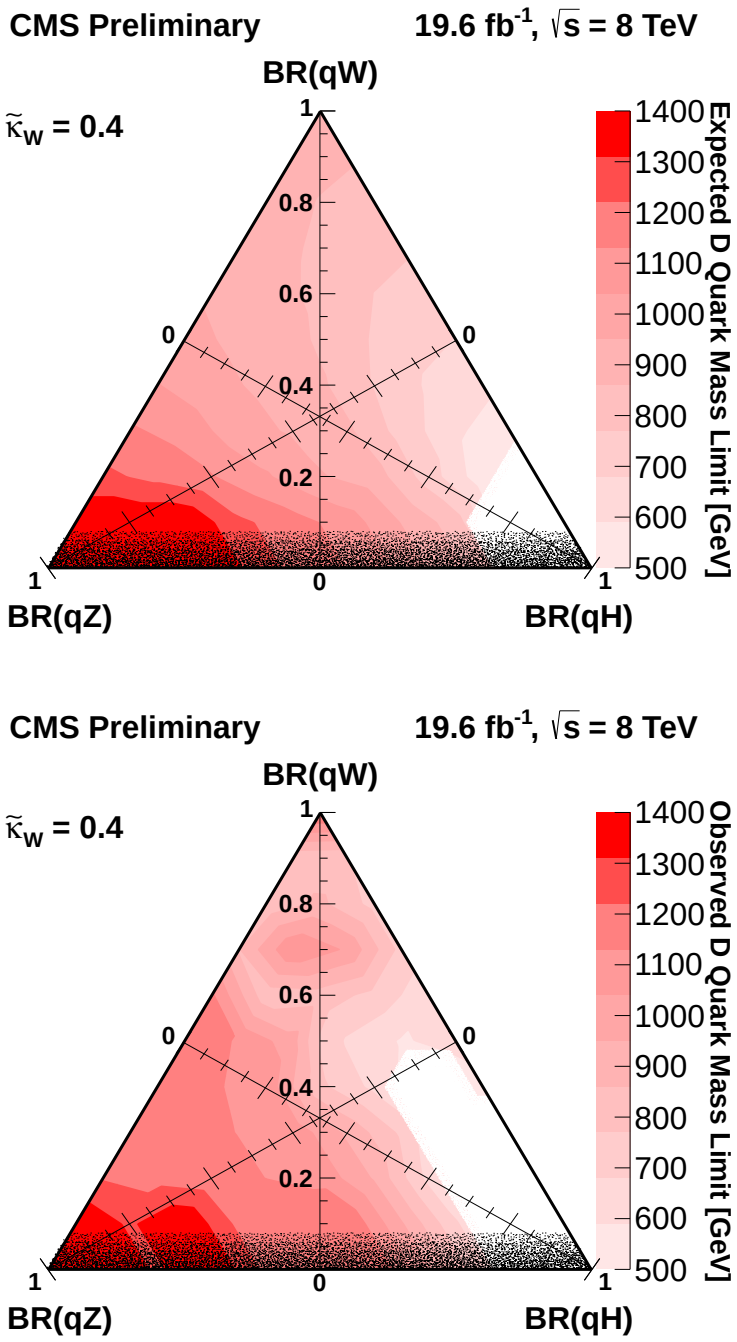

Figure 5.29: The expected (top) and observed (bottom) lower mass limits can be represented in triangles, where each point in the triangle corresponds to a given set of branching fractions ($BF_W \equiv$ BR(qW), $BF_Z \equiv$ BR(qZ), $BF_H \equiv$ BR(qH)). The shown limit contours correspond to a choice of $\tilde{\kappa}_W = 0.4$, such that the relevant signal is dominated by electroweak single production in most of the triangle parameter space, but the relative importance of the pair-produced signal is increased.

Results for $\tilde{\kappa}_W = 0.1$

For the choice of $\tilde{\kappa}_W = 0.1$, the strength of the charged-current single-production process of vector-like quarks is reduced by a factor 100. As a consequence, the pair-production modes, which are independent of the size of the $\tilde{\kappa}_W$ parameter, would contribute significantly to the total signal yield. The 95% CL limits on the signal strength as a function of the vector-like quark mass is shown in Figure 5.30 on the left side for the choice of model parameters $\tilde{\kappa}_W = 0.1$, $\mathrm{BF}_W = 0.5$ and $\mathrm{BF}_Z = 0.25$. For this set of parameters, we exclude at 95% CL vector-like D quarks with a mass below 530 GeV, while the expected exclusion limit is 640 GeV. On the right side of Figure 5.30, the exclusion limits are shown for $\tilde{\kappa}_W = 0.1$ and $\mathrm{BF}_W = 1$. In this case we exclude at 95% CL vector-like D quarks with a mass below 710 GeV, with an expected lower limit of 720 GeV. Note that in the latter scenario, the only signal processes that contribute, correspond to the pair-production mode $QQ \to WqWq$ and the single-production mode $Qq \to Wqq$. Hence, only the WWqq and Wqq subsamples are expected to be sensitive to the presence of vector-like quarks. Indeed, from the simulation of D-quark processes with a generated quark mass of 700 GeV, and for $\tilde{\kappa}_W = 0.1$ and $\mathrm{BF}_W = 1$, we expect in the muon channel about 14 signal events in the WWqq subsample and about 5 events in the Wqq subsample. About 3 events of the $QQ \to WqWq$ process end up in the WHqq category, due to the misidentification of b jets.

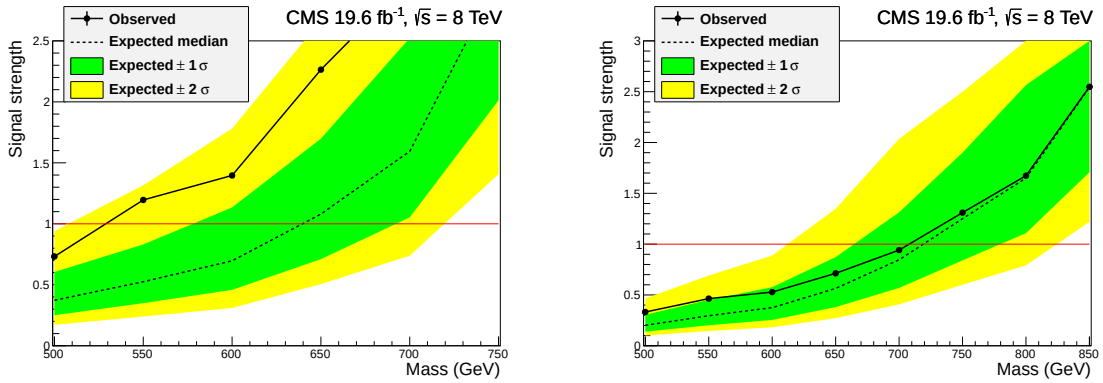

Figure 5.30: The 95% CL exclusion limits on the signal strength corresponding to different sets of model parameters ($\tilde{\kappa}_W = 0.1$, $\mathrm{BF}_W = 0.5$, $\mathrm{BF}_Z = 0.25$ at the left, and $\tilde{\kappa}_W = 0.1$, $\mathrm{BF}_W = 1$ at the right) as a function of the hypothetical D-quark mass. The median expected and the observed exclusion limits are indicated with a dashed and a solid line, respectively. The $\pm 1\sigma$ and $\pm 2\sigma$ expected exclusion limits are represented by the inner (green) and outer (yellow) bands around the median.

The result of the scan over the branching fractions for $\tilde{\kappa}_W = 0.1$ is shown in Figure 5.31 for the expected lower limits on the mass (top) and the observed limits (bottom). A slight data excess in the subsamples sensitive to pair production signal processes (in particular the ZHqq subsample and the full-leptonic and semi-leptonic VZqq subsamples, see for instance Tables 5.8, 5.9 and 5.11) cause the limits in a large part of the triangle to be less stringent than expected. Nevertheless, the deviations remain within the 2σ band around the median expected limits.

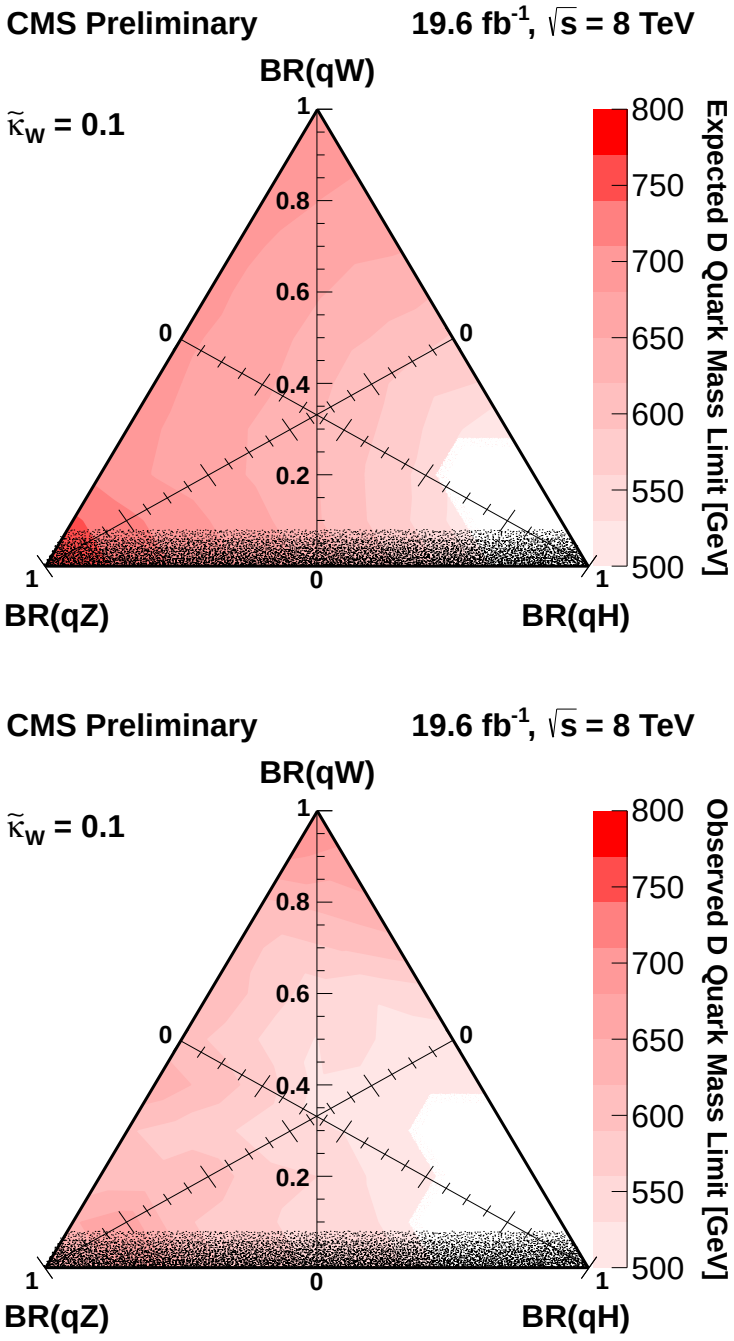

Figure 5.31: The expected (top) and observed (bottom) lower mass limits can be represented in triangles, where each point in the triangle corresponds to a given set of branching fractions ($BF_W \equiv BR(qW)$, $BF_Z \equiv BR(qZ)$, $BF_H \equiv BR(qH)$). The shown limit contours correspond to a choice of $\tilde{\kappa}_W = 0.1$, such that the relevant signal is dominated by strong pair production in most of the triangle parameter space.

5.8.3 Summary

We have performed a search for vector-like quarks with electric charge $-1/3$ decaying to light-flavor quarks using $\sqrt{s} = 8$ TeV proton-proton collisions recorded by the CMS experiment. The full data set collected in 2012, corresponding to an integrated luminosity of 19.6 fb^{-1}, has been analyzed. The production and decay modes of these hypothetical new quarks result in a rich phenomenology, since they may couple to W, Z and/or H bosons. In order to be sensitive to a wide range of decay channels, an inclusive search strategy has been employed, considering both the electroweak single production and the strong pair production of these new quarks.

We have defined various sets of selections on the reconstructed physics objects, subdividing the data set in different event categories. This division has been based on the number of isolated leptons and some additional selection criteria on leptons, jets and the missing transverse energy. The selection in each category has been optimized for specific production and decay modes. Some of the components of the SM background have been estimated from data, while others are taken from simulation. In each category, a sensitive observable has been constructed, to discriminate between the SM background and the vector-like quark signal. Whenever feasible, this observable has been defined as the reconstructed mass of the hypothetical heavy quark.

No significant excess of data with respect to the SM expectation has been observed, hence exclusion limits at 95% CL on the signal model parameters have been derived. The excluded masses range from 500 GeV to 1800 GeV, depending on the vector-like quark branching fractions to W, Z or H bosons and the electroweak single-production strength $\tilde{\kappa}_W$. The limits are evaluated for scenarios in which the single vector-like quark production would dominate over the pair production, and vice versa. Since the applied novel search strategy does not assume any particular assumption on the branching fractions, this analysis is the first to set limits on vector-like quarks decaying to light quarks in such a wide model parameter space. For a down-type vector-like quark only coupling to a W boson and with a single-production coupling parameter equal to unity, a mass below 1740 GeV (1565 GeV) is observed (expected) to be excluded at 95% CL. This tightens the previous constraint significantly, derived by the ATLAS experiment at $\sqrt{s} = 7$ TeV.

Chapter 6

Conclusions and outlook

Throughout the 20$^{\text{th}}$ century, the Standard Model of particle physics has been formulated, which describes and predicts the interactions of fundamental matter particles with an astonishing precision. The Standard Model incorporates three sequential generations of fermions, each containing an up-type and a down-type quark with charges $+2/3$ and $-1/3$ respectively, and a charged and neutral lepton. The generations appear to be identical except for the mass of the corresponding member particles. The heaviest known elementary particle to date, the top quark with a mass of 173 GeV, belongs to the third generation of fermions, and was only discovered in 1995 at the Tevatron accelerator at Fermilab. Since then, many searches have been performed for new heavy quarks, because no conclusive theoretical mechanism has been proposed to prevent the existence of quarks beyond the known three generations. Moreover, the existence of new quarks may address many unsolved questions in high-energy physics.

The production of heavy particles requires a large center-of-mass energy, and with the Large Hadron Collider at CERN, a new era in high energy physics started by colliding protons at the highest energies ever. Huge state-of-the-art particle detectors, such as the Compact Muon Solenoid (CMS) detector, are placed around the proton interaction points, and detect the particles emerging from the collisions. During 2011 and 2012, proton-collision data sets have been collected by the CMS experiment at a center-of-mass energy of $\sqrt{s} = 7$ TeV and 8 TeV, respectively. In this thesis, these data sets have been analyzed to search for the signatures of new heavy quarks. Using the 7 TeV data, a search has been performed for up-type and down-type quarks of a sequential fourth generation. This search is summarized in Chapter 4, and involved a novel inclusive search strategy for single and pair production of heavy quarks. The philosophy of the inclusive search proved successful in excluding previously unprobed parameter space, and this has guided the development of a search at 8 TeV for a more exotic type of quarks, so-called vector-like quarks, as presented in Chapter 5. This analysis specifically focused on down-type vector-like quarks decaying to quarks of the first generation. By considering multiple production modes and many final states of vector-like quark decays simultaneously, very high masses as well as small coupling parameters have been probed. In both analyses, no significant excess of data has been observed with respect to the SM expectation, and lower mass limits on the mass of new heavy quarks have been set at 95% confidence level.

6.1 Combined search for the quarks of a sequential fourth generation

In the search for chiral fourth-generation quarks, we assumed a simple model for an extended unitary CKM matrix, parametrized by a single parameter $A = |V_{tb}|^2 = |V_{t'b'}|^2$. We searched for the presence of up-type (t') and down-type (b') fourth-generation quarks simultaneously, since both quarks should be present in a consistent chiral fourth-generation model. Moreover, global fits of electroweak precision observables have shown that a relatively small mass splitting between t' and b' quarks would be favored. The fourth-generation quarks have been assumed to decay to third generation quarks, as motivated by theoretical considerations. In order to be sensitive to many possible final-state topologies originating from strong pair produced and electroweak single produced fourth-generation quarks, we subdivided the data in event categories according to the number of b jets and W-boson candidates. We required at least one isolated muon or electron in the final state. In these event categories, discriminating observables were defined, such as the scalar sum of the transverse momenta of final-state objects and the reconstructed t'-quark mass.

Using about 5 fb^{-1} of data at $\sqrt{s} = 7$ TeV, we exclude at 95% CL the existence of mass-degenerate fourth-generation quarks with masses below 685 GeV, assuming a minimal mixing between the third- and the fourth-generation quarks. A non-zero cross section for the single fourth-generation quark production processes, corresponding to a value of the extended CKM-matrix parameter $A < 1$, results in a more stringent limit. When a mass difference of 25 GeV is assumed between the t' and b' quarks, the lower limits on the t'-quark mass shifts by about +20 (−20) GeV for $m_{b'} = m_{b'} + 25$ GeV ($m_{t'} = m_{b'} - 25$ GeV).

Before the above analysis results were made public, the most stringent limits excluded at 95% CL the existence of a down-type (up-type) fourth-generation quark with a mass below 611 (570) GeV [75, 79]. Hence, the published results of our analysis [81] reduced the allowed parameter space for a fourth generation of fermions significantly. However, the assumptions on the mass degeneracy within 25 GeV, although motivated by global fits of electroweak observables, and the assumptions on the extended CKM matrix, should be taken with care. For instance, since we made extensive use of b tagging, the lower bounds on the mass may be relaxed for significant decay rates of fourth-generation quarks to first or second generation quarks. Also, in a scenario in which fourth-generation quarks would have an extremely small mixing with any lower generations, the quarks might have such a long lifetime that their signature in a particle detector would be drastically different.

Perspectives

This analysis pushed the lower limits on the mass of fourth-generation quarks well into the mass regime where theoretical problems appear. These involve on the one hand the so-called unitarity bound, which indicates that the mass of chiral fourth-generation quarks should not exceed about 500 GeV, in order not to contradict the probabilistic interpretation of quantum mechanical calculations. On the other hand, a

perturbativity bound of about 600 GeV marks the regime where the Yukawa coupling becomes so large that perturbative calculations make little sense.

Furthermore, since the discovery of the Brout-Englert-Higgs boson in 2012 with a mass of about 126 GeV, and production and decay rates that are consistent with a three-generation Standard Model, a sequential chiral fourth generation of quarks has been regarded as ruled out [48]. From global fits of electroweak precision observables, the existence of a relatively light H boson is found to be highly inconsistent with a chiral fourth generation.

Taking the above considerations into account, the existence of chiral fourth-generation quarks has become very unlikely. A high level of parameter fine-tuning would be needed to avoid all constraints. As a consequence, the general focus on the search for new quarks shifted towards vector-like quarks, that do evade many of the experimental and theoretical constraints. Vector-like quarks are defined as quarks with the same weak-interaction transformation properties for the left-handed and right-handed chirality components. This is in contrast with the concept of chiral quarks, whose left-handed and right-handed chiralities transform differently. Vector-like quarks can evade the tight bounds from the H-boson discovery, and since they do not obtain their mass from a Yukawa coupling to the Standard-Model H field, their mass is not limited by theoretical bounds. The electric charge of these quarks may be $-1/3$ (D), $+2/3$ (U), $+5/3$ (X) or $-4/3$ (Y). Within the theoretical formulation of these exotic objects, vector-like quarks are allowed to mix with all three SM generations. In the 8 TeV analysis presented in this thesis, we focused on mixing with the first generation.

6.2 Search for vector-like quarks decaying to light quarks

Similar to regular SM quarks, vector-like quarks can be produced in pairs via the strong interaction and singly via the electroweak interaction. In the particular case of heavy quarks coupling to first-generation quarks, a large single-production rate is expected in proton collisions, since the partons required in the initial state are valence quarks of the proton. Vector-like quarks may couple to W, Z and H bosons, and a priori, a wide variety of choices of the branching fractions to these bosons is possible. This motivates an inclusive search strategy, where we defined and optimized event categories in order to be sensitive to singly or pair produced vector-like quarks, with specific decays to the different bosons. Since the theory allows for only one type of vector-like quark to exist, the analysis strategy was developed searching for one vector-like quark type at a time, in contrast with the combined search for quarks of a sequential fourth generation. We required at least one isolated muon or electron in the final state. In the event categories, sensitive observables have been reconstructed, like the reconstructed mass of the vector-like quark or the sum of the transverse momenta of final-state objects.

Assuming a specific value of the charged-current single production strength parameter $\tilde{\kappa}_W$, we scanned the decay branching fractions BF_W, BF_Z and BF_H of the heavy quarks to W, Z and H bosons, respectively. Using about 20 fb^{-1} of data collected at

$\sqrt{s} = 8$ TeV, we derived lower limits on vector-like D quark masses between 500 GeV and 1800 GeV, depending on the set of branching fractions. For a particular choice of parameters $\tilde{\kappa}_W = 1$, $\mathrm{BF}_W = 0.5$ and $\mathrm{BF}_Z = 0.25$, we exclude at 95% CL D quarks with a mass below 1585 GeV. When decreasing the charged-current single-production strength by a factor of 100 by setting $\tilde{\kappa}_W = 0.1$, we set a lower limit at 95% CL of 530 GeV on the heavy quark mass. The bounds for $\tilde{\kappa}_W = 0.1$ are much less stringent than those for $\tilde{\kappa}_W = 1$, and the signal yield becomes dominated by pair production processes.

In a scenario where the vector-like D quark only couples to W bosons, we exclude masses below 1740 GeV at 95% CL for $\tilde{\kappa}_W = 1$. This significantly improves the previous constraint on these quarks, obtained by the ATLAS experiment [72]. Moreover, since we did not assume a specific branching fraction to W, Z or H bosons, but derived limits for all possibilites, a vast amount of previously unexplored parameter space for vector-like quarks has been probed in this analysis.

Perspectives

The presented search for vector-like quarks decaying to light quarks may be improved or extended by taking into account the following considerations.

- **Improved signal event samples.** The detector response for the vector-like quark events used in this thesis has been simulated with CMS fast-simulation sofware, which uses fast but simplified techniques to describe the response of generated particles traversing the detector. For kinematic distributions of reconstructed physics objects, we observed deviations with the more detailed CMS full-simulation software smaller than 15%, but a full simulation is expected to be more accurate. In addition, if higher-order calculations of the cross section of singly-produced vector-like quarks would become available, the uncertainty on the signal yield, which has now been set to 20%, may be decreased. Finally, event samples for quark masses below 500 GeV could be generated to evaluate the sensitivity towards very high branching fractions to H bosons.

- **Exploiting properties of boosted bosons decaying to jets.** When a very massive particle decays to a W, a Z or a H boson, this boson in general has a large transverse momentum. In the case where the boson decays to two jets, the boost of the boson may result in a very small angular separation between the jets, even smaller than the typical jet cone. Hence, the jets would be overlapping or completely merged, resulting in so-called fat jets. Via advanced jet substructure techniques, one can identify such boosted bosons, and improve the sensitivity in searches for new heavy particles [177–179].

- **Increasing the sensitivity to the H-boson channels.** Since its discovery at the LHC, the H boson is considered to be more and more interesting as a probe to search for exotic physics. We considered vector-like quark decays to H bosons only in pair production processes, and even in these cases, the branching fraction to a W or a Z boson should still be non-zero to retain some sensitivity. One could use non-lepton trigger requirements, such that the single-production

process $pp \xrightarrow{V^*} Qq \rightarrow Hqq$ could be included in the inclusive search. Other production modes are also promising at future LHC runs, such as vector boson H fusion and associated production of a vector-like quark with a H boson [180].

- **Probing higher masses in LHC Run 2.** The Large Hadron Collider is scheduled to resume operations in 2015, colliding protons at a center-of-mass energy of $\sqrt{s} = 13$ TeV [181]. The instantaneous luminosity will increase as well, such that the LHC will deliver about 40 fb^{-1} of integrated luminosity per year. The increased heavy quark production cross section at this high collision energy, together with the high amount of data that is expected to be collected, will open up a new mass domain to probe.

With the above improvements we should be able to probe vector-like quark at the TeV level regardless of the production mode and the branching fractions, and masses up to 2 TeV for the most sensitive channels. The same analysis strategy can also be applied to up-type vector-like quarks (U) or quarks with an exotic charge (X and Y) that decay to quarks of the first generation. The main differences between U quarks and D quarks are the single-production cross sections, and the sign of the lepton charge in the selection optimized for the single-production of the vector-like quarks decaying to a W boson.

Bibliography

[1] S. L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961) 579–588.

[2] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[3] S. Weinberg, "Elementary Particle Physics: Relativistic Groups and Analyticity", (1968).

[4] J. B. et al. (Particle Data Group), "The Review of Particle Physics", *Phys. Rev. D* **86** (2012) 010001.

[5] 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:hep-ex/1207.7235.

[6] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, arXiv:hep-ex/1207.7214.

[7] CMS Collaboration, "Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV", CMS Physics Analysis Summary HIG-13-005, (2013).

[8] M. Peskin and D. Schroeder, "An Introduction to Quantum Field Theory",. USA: Addison-Wesley (1995) 842 p.

[9] C.S. Wu et al., "Experimental Test of Parity Conservation in Beta Decay", *Phys. Rev.* **105 (4)** (1957) 14131415.

[10] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–322.

[11] P. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Rev. Lett.* **13** (1964) 508–509.

[12] G. Guralnik, C. Hagen, and T. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587.

[13] T2K Collaboration, "Evidence of Electron Neutrino Appearance in a Muon Neutrino Beam", *Phys. Rev. D* **88** (2013) 032002, doi:10.1103/PhysRevD.88.032002, arXiv:hep-ex/1304.0841.

[14] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields", *Nucl. Phys. B* **44** (1972) 189–213.

[15] Planck Collaboration, "Planck 2013 results. I. Overview of products and scientific results", (2013).

[16] B. Holdom et al., "Four Statements about the Fourth Generation", *PMC Phys. A* (2009) 4, arXiv:hep-ph/0904.4698v2.

[17] C. Jarlskog, "Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation", *Phys. Rev. Lett.* **55** (1985) 10391042.

[18] G. W. Hou, "Source of CP Violation for the Baryon Asymmetry of the Universe", *Int. J. Mod. Phys. D* **20** (2011) 1521–1532, doi:10.1142/S0218271811019694, arXiv:hep-ph/1101.2161.

[19] H. Fritzsch, "Hierarchical chiral symmetries and the quark mass matrix", *Phys. Lett. B* **289** (1992) 92.

[20] A. Datta, "Flavour Democracy Calls for the Fourth Generation", *Pramana* **40** (1993) L503–L509, arXiv:hep-ph/9207248v1.

[21] The Gfitter group, "Updated Status of the Global Electroweak Fit and Constraints on New Physics", *Eur. Phys. J. C* **72** (2012) 2003.

[22] G. W. Hou, "A brief (p)review on a possible fourth generation world to come", *PoS ICHEP2010* **244** (2010) arXiv:hep-ph/1101.2158v1.

[23] Y. Chao et al., "Threshold Effects in the Decay of Heavy b' and t' Quarks", *Phys. Rev. D* **84** (2011) 014029, arXiv:hep-ph/1101.0592.

[24] M. Buchkremer, J.-M. Gérard, and F. Maltoni, "Closing in on a perturbative fourth generation", *JHEP* **06** (2012) 135, doi:10.1007/JHEP06(2012)135, arXiv:hep-ph/1204.5403v2.

[25] D. Collaboration, "An Improved determination of the width of the top quark", *Phys. Rev. D* **85** (2012) 091104, arXiv:hep-ex/1201.4156v1.

[26] The ALEPH, DELPHI, L3, OPAL and SLD Collaborations, The LEP Electroweak Working Group, The SLD Electroweak and Heavy Flavour Groups, "Precision Electroweak Measurements on the Z Resonance", *Phys. Rept.* **427** (2006) 257–454, arXiv:hep-ex/0509008v3.

[27] M. Peskin and T. Takeuchi, "New Constraint on a Strongly Interacting Higgs Sector", *Phys. Rev. Lett.* **65 (8)** (1990) 964967.

[28] M. Peskin and T. Takeuchi, "Estimation of Oblique Electroweak corrections", *Phys. Rev. D* **46** (1992) 381409.

[29] The Gfitter group, "The electroweak fit of the Standard Model after the Discovery of a New Boson at the LHC", *Eur. Phys. J. C* **72** (2012) 2205.

[30] H.-J. He, N. Polonsky, and S. Su, "Extra Families, Higgs Spectrum and Oblique Corrections", *Phys. Rev. D* **64** (2001) 053004.

[31] O. Eberhardt, A. Lenz, and J. Rohrwild, "Less space for a new family of fermions", *Phys. Rev. D* **82** (2010) 095006.

[32] CDF Collaboration, "Search for Heavy Top $t' \to Wq$ in Lepton Plus Jets Events in 4.6 fb^{-1}", CDF public conference note CDF/PUB/TOP/PUBLIC/10110, (2010).

[33] T. C. Collaboration, "Search for heavy bottom-like quarks decaying to an electron or muon and jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **106** (2011) 141803.

[34] L. Collaboration, "Search for Heavy Neutral and Charged Leptons in e+ e- Annihilation at LEP", *Phys. Lett. B* **517** (2001) 75–85, `arXiv:hep-ex/0107015v1`.

[35] ATLAS Collaboration, "Search for pair production of a heavy top-like quark decaying to a high-p_T W boson and a b quark in the lepton plus jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", ATLAS Note ATLAS-CONF-2013-060, (2013).

[36] CMS Collaboration, "Search for Vector-Like b' Pair Production with Multilepton Final States in pp collisions at $\sqrt{s} = 8$ TeV", CMS Pysics Analysis Summary B2G-13-003, (2013).

[37] M. Chanowitz, M. Furman, and I. Hinchliffe, "Weak Interactions of Ultraheavy Fermions (II)", *Nucl. Phys. B* **153** (1979) 402, `arXiv:hep-ph/1003.3211v3`.

[38] J. Erler and P. Langacker, "Precision Constraints on Extra Fermion Generations", *Phys. Rev. Lett.* **105** (2010) 031801, `arXiv:hep-ph/1003.3211v3`.

[39] W. J. Marciano, G. Valencia, and S. Willenbrock, "Precision Constraints on Extra Fermion Generations", *Phys. Rev. D* **40** (1989) 1725, `arXiv:hep-ph/1003.3211v3`.

[40] A. Wingerter, "Implications of the Stability and Triviality Bounds on the Standard Model with Three and Four Chiral Generations", *Phys Rev D* **84** (2011) 095012, `arXiv:hep-ph/1109.5140v1`.

[41] J. Gunion, D. W. McKay, and H. Pois, "Four Generations and Higgs Physics", *Phys. Rev. D* **53** (1996) 1616–1647, `arXiv:hep-ph/9507323v2`.

[42] G. Kribs et al., "Four Generations and Higgs Physics", *Phys. Rev. D* **76** (2007) 075016, `arXiv:hep-ph/0706.3718`.

[43] N. B. Schmidt et al., "The Fourth Standard Model Family and the Competition in Standard Model Higgs Boson Search at Tevatron and LHC", *Eur. Phys. J. C* **66 (1)** (2009) 119–126.

[44] A. N. Rozanov and M. I. Vysotsky, "Tevatron constraints on the Higgs boson mass in the fourth-generation fermion models revisited", *Phys. Lett. B* **700** (2011) 313–315, arXiv:hep-ph/1012.1483v2.

[45] S. A. Cetin et al., "Impact of the relatively light fourth family neutrino on the Higgs boson search", *Phys. Lett. B* **710 (2)** (2011) 328, doi:10.1016/j.physletb.2012.02.088, arXiv:hep-ph/1108.4071v2.

[46] S. A. etin et al., "Status of the Fourth Generation - A Brief Summary of B3SM-III Workshop in Four Parts", (2011). arXiv:hep-ph/1112.2907v1.

[47] A. Denner et al., "Higgs production and decay with a fourth Standard-Model-like fermion generation", *Eur. Phys. J. C* **72** (2012) 1992, doi:10.1140/epjc/s10052-012-1992-3, arXiv:hep-ph/1111.6395.

[48] O. Eberhardt et al., "Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations", *Phys.Rev.Lett.* **109** (2012) 241802, arXiv:hep-ph/1209.1101v2.

[49] J. Aguilar-Saavedra et al., "A handbook of vector-like quarks: mixing and single production", *Phys. Rev. D* **88** (2013) 112003, arXiv:hep-ph/1306.0572v3.

[50] Y. Okada and L. Panizzi, "LHC signatures of vector-like quarks", *Adv. High Energy Phys.* **2013** (2013) 364936, arXiv:hep-ph/1207.5607v3.

[51] N. Arkani-Hamed et al., "The Littlest Higgs", *JHEP* **0207** (2002) 034, arXiv:arXiv:hep-ph/0206021v2.

[52] M. Schmaltz and D. Tucker-Smith, "Little Higgs Review", *Ann. Rev. Nucl. Part. Sci* **55** (2005) 229–270, arXiv:arXiv:hep-ph/.

[53] D. Marzocca, M. Serone, and J. Shu, "General Composite Higgs Models", *JHEP* **08** (2012) 013, doi:10.1007/JHEP08(2012)013, arXiv:hep-ph/1205.0770v3.

[54] H.-C. Cheng, B. Dobrescu, and C. Hill, "Electroweak Symmetry Breaking and Extra Dimensions", *Nucl. Phys. B* **589** (2000) 249–268, arXiv:hep-ph/9912343v3.

[55] J. Kang, P. Langacker, and B. Nelson, "Theory and Phenomenology of Exotic Isosinglet Quarks and Squarks", *Phys. Rev. D* **77** (2008) 035003, arXiv:hep-ph/0708.2701v2.

[56] D. Guadagnoli, R. Mohapatra, and I. Sung, "Gauged Flavor Group with Left-Right Symmetry", *JHEP* **1104** (2011) 093, arXiv:hep-ph/1103.4170v3.

[57] D. Choudhury, T. Tait, and C. Wagner, "Beautiful Mirrors and Precision Electroweak Data", *Phys. Rev. D* **65** (2002) 053002, arXiv:hep-ph/0109097v2.

[58] K. Ishiwata and M. Wise, "Phenomenology of heavy vectorlike leptons", *Phys. Rev. D* **88** (2013) 055009, arXiv:hep-ph/1307.1112v2.

[59] F. del Aguila, M. Perez-Victoria, and J. Santiago, "Observable contributions of new exotic quarks to quark mixing", *JHEP* **0009** (2000) 011, arXiv:hep-ph/0007316v2.

[60] G. Cacciapaglia et al., "Bounds and Decays of New Heavy Vector-like Top Partners", *JHEP* **1011** (2010) 159, doi:10.1007/JHEP11(2010)159, arXiv:hep-ph/1007.2933v2.

[61] G. Cacciapaglia et al., "Heavy Vector-like Top Partners at the LHC and flavour constraints", *JHEP* **1203** (2012) 070, arXiv:hep-ph/1108.6329v2.

[62] M. Buchkremer et al., "Model Independent Framework for Searches of Top Partners", *Nucl.Phys. B* **876** (2013) 376–417, arXiv:hep-ph/1305.4172v3.

[63] A. Atre et al., "Model-Independent Searches for New Quarks at the LHC", *JHEP* **1108** (2011) 080, arXiv:hep-ph/1102.1987v1.

[64] L. Maiani, "The GIM Mechanism: origin, predictions and recent uses", (2013). arXiv:hep-ph/1303.6154v1.

[65] S. Glashow, J. Iliopoulos, and L. Maiani, "Weak Interactions with LeptonHadron Symmetry", *Phys. Rev. D* **2** (1970) 1285.

[66] A. Deandrea, "Atomic parity violation in cesium and implications for new physics", *Phys. Lett. B* **409** (1997) 277–282, doi:10.1016/S0370-2693(97)00843-5, arXiv:hep-ph/9705435v3.

[67] J. A. Aguilar-Saavedra, "Top flavour-changing neutral interactions: theoretical expectations and experimental detection", *Acta Phys. Polon. B* **35** (2004) 2695–2710, arXiv:hep-ph/0409342v4.

[68] CMS Collaboration, "Search for flavor-changing neutral currents in top-quark decays t to Zq in pp collisions at sqrt(s)= 8 TeV", *Submitted to Phys. Rev. Lett,* (2013) arXiv:hep-ex/1312.4194.

[69] G. Cynolter and E. Lendvai, "Electroweak Precision Constraints on Vector-like Fermions", *Eur. Phys. J. C* **58** (2008) 463–469, arXiv:hep-ph/0804.4080.

[70] CMS Collaboration, "Inclusive search for a vector-like T quark with charge 2/3 in pp collisions at sqrt(s) = 8 TeV", *Phys. Lett. B* **729** (2014) 149, arXiv:hep-ph/1311.7667v2.

[71] CMS Collaboration, "Search for pair-produced vector-like quarks of charge -1/3 in lepton+jets final state in pp collisions at $\sqrt{s} = 8$ TeV", CMS Pysics Analysis Summary B2G-12-019, (2013).

[72] ATLAS Collaboration, "Search for Single Production of Vector-like Quarks Coupling to Light Generations in 4.64 fb^{-1} of ATLAS Data at sqrt(s) = 7 TeV", ATLAS Note ATLAS-CONF-2012-137, (2012).

[73] ATLAS Collaboration, "Search for pair-produced heavy quarks decaying to Wq in the two-lepton channel at sqrt(s) = 7 TeV with the ATLAS detector", *Phys. Rev. D* **86** (2012) 012007, arXiv:hep-ex/1202.3389.

[74] CMS Collaboration, "Search for heavy, top-like quark pair production in the dilepton final state in pp collisions at sqrt(s) = 7 TeV", *Phys. Lett. B* **716** (2012) 103, arXiv:hep-ex/1203.5410.

[75] CMS Collaboration, "Search for pair produced fourth-generation up-type quarks in pp collisions at sqrt(s) = 7 TeV with a lepton in the final state", *Phys. Lett. B* **718** (2012) 307, arXiv:hep-ex/1209.0471v2.

[76] ATLAS Collaboration, "Search for pair production of heavy top-like quarks decaying to a high-pT W boson and a b quark in the lepton plus jets final state at sqrt(s)=7 TeV with the ATLAS detector", *Phys. Lett. B* **718** (2013) 1284–1302, arXiv:hep-ex/1210.5468.

[77] CMS Collaboration, "Search for heavy quarks decaying into a top quark and a W or Z boson using lepton + jets events in pp collisions at sqrt(s) = 7 TeV", *JHEP* **01** (2013) 154.

[78] ATLAS Collaboration, "Search for down-type fourth generation quarks with the ATLAS detector in events with one lepton and hadronically decaying W bosons", *Phys. Rev. Lett.* **109** (2012) 032001, arXiv:hep-ex/1202.6540v2.

[79] CMS Collaboration, "Search for heavy bottom-like quarks in 4.9 inverse femtobarns of pp collisions at sqrt(s) = 7 TeV", *JHEP* **05** (2012) 123, arXiv:hep-ex/1204.1088v2.

[80] ATLAS Collaboration, "Search for anomalous production of events with same-sign dileptons and b jets in 14.3 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", ATLAS Note ATLAS-CONF-2013-051, (2013).

[81] CMS Collaboration, "Combined search for the quarks of a sequential fourth generation", *Phys. Rev. D* **86** (2012) 112003, arXiv:hep-ex/1209.1062v2.

[82] ATLAS Collaboration, "Search for pair production of new heavy quarks that decay to a Z boson and a third generation quark in pp collisions at $\sqrt{8} = 8$ TeV with the ATLAS detector", ATLAS Note ATLAS-CONF-2013-056, (2013).

[83] CMS Collaboration, "Search for a Vector-like Quark with Charge 2/3 in t + Z Events from pp Collisions at sqrt(s) = 7 TeV", *Phys. Rev. Lett.* **107** (2011) 271802, arXiv:hep-ex/1109.4985v1.

[84] ATLAS Collaboration, "Search for heavy top-like quarks decaying to a Higgs boson and a top quark in the lepton plus jets final state in pp collisions at sqrt(s) = 8 TeV with the ATLAS detector", ATLAS Note ATLAS-CONF-2013-018, (2013).

[85] CMS Collaboration, "Search for a vector-like quark of charge -1/3 and decaying to bZ in pp collisions at $\sqrt{s}$ =7 TeV", CMS Pysics Analysis Summary EXO-11-066, (2012).

[86] 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:hep-ex/1204.1265v2.

[87] ATLAS Collaboration, "Search for exotic same-sign dilepton signatures (b' quark, T5/3 and four top quarks production) in 4.7 fb-1 of pp collisions at sqrt(s)=7 TeV with the ATLAS detector", ATLAS Note ATLAS-CONF-2012-130, (2012).

[88] CMS Collaboration, "Search for top-quark partners with charge 5/3 in the same-sign dilepton final state", *Submitted to Phys. Rev. Lett.* (2013) arXiv:hep-ph/1312.2391v1.

[89] CMS Collaboration, "Search for anomalous production of events with same-sign dileptons and b jets in 14.3 fb^{-1} of pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector", CMS Physics Analysis Summary B2G-12-021, (2013).

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

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

[92] CMS Collaboration, "CMS Physics : Technical Design Report Volume 1: Detector Performance and Software", Technical Design Report CMS CERN-LHCC-2006-001 CMS-TDR-8-1, (2006).

[93] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[94] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[95] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[96] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007, doi:10.1088/1748-0221/3/08/S08007.

[97] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006, doi:10.1088/1748-0221/3/08/S08006.

[98] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *J. Fluids Eng.* **82(1)** (1960) 35, doi:10.1115/1.3662552.

[99] CMS Collaboration, "Adaptive Vertex Reconstruction", CMS Note CMS-NOTE-08-033, (2008).

[100] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", CMS Physics Analysis Summary TRK-10-005, (2010).

[101] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at sqrt(s) = 7 TeV", *JINST* **8** (2013) P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:hep-ex/1306.2016.

[102] CMS Collaboration, "LHC computing Grid : Technical Design Report", Scientific Committee Paper CERN-LHCC-2005-024, (2005).

[103] M. Dobbs et al., "Les Houches Guidebook to Monte Carlo Generators for Hadron Collider Physics", (2004). arXiv:hep-ph/0403045.

[104] J. C. Collins, D. E. Soper, and G. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1988) 1–91, arXiv:hep-ph/0409313.

[105] J. Pumplin et al., "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis", *JHEP* **0207** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195v3.

[106] "Online PDF plotting and calculation",. http://hepdata.cedar.ac.uk/pdf/pdf3.html.

[107] J. Alwall et al., "MadGraph 5: going beyond", *HEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:hep-ph/1106.0522.

[108] S. Frixione and C. O. P. Nason, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **0711** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:hep-ph/0709.2092.

[109] S. Frixione and B. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **0206** (2002) 029, arXiv:hep-ph/0204244.

[110] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[111] V. Gribov and L. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438.

[112] Y. 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.

[113] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298.

[114] M. Mangano et al., "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **0701** (2007) 013, arXiv:hep-ph/0611129.

[115] B. Andersson, "The Lund Model". Cambridge University Press, 1998.

[116] S. Jadach, "TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons", *Computer Physics Communications* **64 (2)** (1991) 275, doi:10.1016/0010-4655(91)90038-M.

[117] R. Field, "Early LHC Underlying Event Data - Findings and Surprises", (2010). arXiv:hep-ph/1010.3558.

[118] CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC with sqrt(s)= 7 TeV and Comparison with sqrt(s) = 0.9 TeV", *JHEP* **9** (2011) 109, arXiv:hep-ph/1107.0330.

[119] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[120] D. Orbaker on behalf of the CMS collaboration, "Fast simulation of the CMS detector", *J. Phys.: Conf. Ser.* **219** (2010) 032053, doi:10.1088/1742-6596/219/3/032053.

[121] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary PFT-09-001, (2009).

[122] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV", CMS Physics Analysis Summary PFT-10-003, (2010).

[123] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV", *JINST* **7** (2012) P10002, arXiv:hep-ex/1206.4071.

[124] CMS Collaboration, "Performance of CMS muon reconstruction in cosmic-ray events", *JINST* **5** (2010) T03022, doi:10.1088/1748-0221/5/03/T03022, arXiv:hep-ex/0911.4994.

[125] CMS Collaboration, "Single Muon efficiencies in 2012 Data", CMS Detector Performance Summary DP-2013-009, (2013).

[126] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary PFT-10-002, (2010).

[127] CMS Collaboration, "Electron reconstruction within the Particle Flow Algorithm", CMS Analysis Note AN-10-034, (2010).

[128] S. Baffioni et al., "Electron reconstruction in CMS", *Eur. Phys. J. C* **49** (2007) 1099–1116, doi:10.1140/epjc/s10052-006-0175-5.

[129] CMS Collaboration, "Electron reconstruction and identification at sqrt(s) = 7 TeV", CMS Physics Analysis Summary EGM-10-004, (2010).

[130] W. Adam et al., "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *J. Phys. G: Nucl. Part. Phys* **31 N9** (2005) doi:10.1088/0954-3899/31/9/N01, arXiv:physics/0306087.

[131] F. Beaudette and D. Benedetti and P. Janot and M. Pioppi, "Electron Reconstruction within the Particle Flow Algorithm", CMS Analysis Note AN-10-034, (2010).

[132] cms Collaboration, "Measurement of the lepton charge asymmetry in inclusive W production in pp collisions at sqrt(s) = 7 TeV", *JHEP* **1104** (2011) 050, doi:0.1007/JHEP04(2011)050, arXiv:hep-ex/1103.3470.

[133] CMS Collaboration, "Electron performance with 19.6 fb[1] of data collected at $\sqrt{s} = 8$ TeV with the CMS detector.", CMS Detector Performance Summary DP-2013-003, (2013).

[134] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **0804** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:hep-ph/0802.1189.

[135] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **06** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:hep-ex/1107.4277.

[136] CMS Collaboration, "Jet Energy Scale performance in 2011", CMS Detector Performance Summary DP-2012-006, (2012).

[137] CMS Collaboration, "Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11 fb^{-1} of data in CMS", CMS Detector Performance Summaries CMS-DP-2013-011, (2013).

[138] A. Perloff on behalf of the CMS collaboration, "Pileup measurement and mitigation techniques in CMS", *J. Phys. Conf. Ser.* **404** (2012) 012045, doi:10.1088/1742-6596/404/1/012045.

[139] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **0804** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:hep-ph/0802.1188.

[140] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119, doi:0.1016/j.physletb.2007.09.077, arXiv:hep-ph/0707.1378.

[141] H. Kirschenmann on behalf of the CMS collaboration, "Determination of the Jet Energy Scale in CMS", *J. Phys.: Conf. Ser.* **404** (2012) 012013, doi:10.1088/1742-6596.

[142] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/6/09/P09001, arXiv:hep-ex/1211.4462.

[143] CMS Collaboration, "Measurement of B anti-B Angular Correlations based on Secondary Vertex Reconstruction at sqrt(s)=7 TeV", *JHEP* **1103** (2011) 136, doi:10.1007/JHEP03(2011)136, arXiv:hep-ph/1102.3194.

[144] M. Maes, "Measurement of the top quark pair production cross section at the LHC with the CMS experiment". PhD thesis, Vrije Universiteit Brussel, Brussels, 2013. Presented on October 2013.

[145] CMS Collaboration, "Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events", CMS Physics Analysis Summary BTV-13-001, (2013).

[146] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) 09001, doi:10.1088/1748-0221/6/09/P09001, arXiv:hep-ex/1106.5048.

[147] CMS Collaboration, "Measurement of the $t\bar{t}$ Production Cross Section in pp Collisions at 7 TeV in Lepton + Jets Events Using b-quark Jet Identification", *Phys. Rev. D* **84** (2011) 092004, arXiv:hep-ex/1108.3773.

[148] CMS Collaboration, "Measurement of the inclusive W and Z production cross sections in pp collisions at sqrt(s) = 7 TeV with the CMS experiment", *JHEP* **10** (2011) 132, arXiv:hep-ex/1107.4789.

[149] N. Kidonakis, "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production", *Phys. Rev. D* **83** (2011) 091503, arXiv:hep-ph/1103.2792.

[150] N. Kidonakis, "NNLL resummation for s-channel single top quark production", *Phys. Rev. D* **81** (2010) 054028, arXiv:hep-ph/1001.5034.

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

[152] V. Hirschi et al., "Automation of one-loop QCD corrections", *JHEP* **05** (2011) 044, arXiv:hep-ph/1103.0621.

[153] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10, arXiv:hep-ph/1007.3492.

[154] J. M. Campbell and R. K. Ellis, "An update on vector boson pair production at hadron colliders", *Phys. Rev. D* **60** (1999) 113006, arXiv:hep-ph/9905386.

[155] M. Aliev et al., "HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034, arXiv:hep-ph/1007.1327.

[156] J. Campbell et al., "NLO predictions for t-channel production of single top and fourth generation quarks at hadron colliders", *JHEP* **10** (2009) 042, arXiv:hep-ph/0907.3933.

[157] CMS Collaboration, "b-Jet Identification in the CMS Experiment", CMS Physics Analysis Summary BTV-11-004, (2011).

[158] P. Van Mulders, "Calibration of the jet energy scale using top quark events at the LHC. oai:cds.cern.ch:1308729". PhD thesis, Vrije Universiteit Brussel and Antwerp University, Brussels and Antwerp, 2010. Presented on June 2010.

[159] C. Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update", technical report, (2012).

[160] CDF Collaboration, "First Measurement of the b-jet Cross Section in Events with a W Boson in p-pbar Collisions at sqrts = 1.96 TeV", *Phys. Rev. Lett.* **104** (2010) 131801, doi:10.1103/PhysRevLett.104.131801, arXiv:hep-ex/0909.1505v5.

[161] A. Read, "Linear interpolation of histograms", *Nucl. Instrum. Meth. A* **425** (1999) 357, doi:10.1016/S0168-9002(98)01347-3.

[162] J. Conway, "Nuisance Parameters in Likelihoods for Multisource Spectra", in *Proceedings of PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, H. B. Propser and L. Lyons, eds., p. 115. CERN, 2011.

[163] K. Cranmer et al., "HistFactory: A tool for creating statistical models for use with RooFit and RooStats", (2012).

[164] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0., arXiv:1007.1727v3.

[165] A. Wald, "Tests of Statistical hypotheses concerning several parameters when the number of observations is large", *Transactions of the American Mathematical Society* **54** (1943) 426–482.

[166] A. L. Read, "Presentation of search results: the CL(s) technique", *J. Phys. G* **28** (2002) 2693.

[167] D0 Collaboration, "Search for single vector-like quarks in ppbar collisions at sqrt(s) = 1.96 TeV", *Phys. Rev. Lett.* **106** (2011) 081801, doi:10.1103/PhysRevLett.106.081801, arXiv:hep-ex/1010.1466v3.

[168] 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) 22–39, arXiv:hep-ex/1112.5755v2.

[169] The LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", (2013). arXiv:hep-ph/1307.1347v2.

[170] R. Gavin et al., "W physics at the LHC with FEWZ 2.1", *Comput.Phys.Commun.* **184** (2013) 208–214, doi:10.1016/j.cpc.2012.09.005, arXiv:hep-ph/1201.5896v1.

[171] CMS Collaboration, "Top pair cross section in e/mu+jets at 8 TeV", CMS Physics Analysis Summary TOP-12-006, (2012).

[172] N. Kidonakis, "Differential and total cross sections for top pair and single top production",. arXiv:hep-ph/1205.3453. Presented at International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS2012), Bonn, Germany, March 26-30, 2012.

[173] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in pp collisions at sqrt(s) = 7 TeV", *Eur. Phys. J. C* **73** (2013) 2339, arXiv:hep-ex/1211.2220.

[174] CMS Collaboration, "Measurement of inclusive W and Z boson production cross sections in pp collisions at sqrt(s) = 8 TeV", *Submitted to Phys. Rev. Lett* (2014) arXiv:hep-ex/1402.0923v1.

[175] CMS Collaboration, "Search for standard model four top quark production at 8 TeV in the lepton + jets channel", CMS Physics Analysis Summary TOP-13-012, (2013).

[176] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analysis Summary LUM-13-001, (2013).

[177] J. M. Butterworth et al., "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:hep-ph/0802.2470.

[178] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Techniques for improved heavy particle searches with jet substructure", *Phys. Rev. D* **80** (2009) 051501, doi:10.1103/PhysRevD.80.051501, arXiv:hep-ph/0903.5081.

[179] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", *Phys. Rev. D* **81** (2010) 094023, doi:10.1103/PhysRevD.81.094023, arXiv:hep-ph/0912.0033.

[180] A. Atre, M. Chala, and J. Santiago, "Searches for New Vector Like Quarks: Higgs Channels", *JHEP* **1305** (2013) 099, doi:10.1007/JHEP05(2013)099, arXiv:hep-ph/1302.0270.

[181] "CERN press release: The first LHC protons run ends with new milestone",. http://press.web.cern.ch/press-releases/2012/12/first-lhc-protons-run-ends-new-milestone.

Summary

The biggest quest in elementary particle physics is to uncover which fundamental building blocks constitute our universe, and how they interact with each other. Throughout the 20^{th} century, physicists have formulated the Standard Model of particle physics, a mathematically elegant theory that describes subatomic particles and their interactions at the most fundamental level. The Standard Model has been remarkably successful in explaining countless observations in high-energy experiments. Moreover, from this theory profound predictions were made, for instance about the existence of the famous Brout-Englert-Higgs (H) boson, which has been discovered in 2012 at the CERN laboratory.

It appears that in this theory the fundamental matter particles, the so-called quarks and leptons, can be categorized into three families or 'generations'. The first generation consists of the up and down quark (that make up protons and neutrons), and the electron and the electron neutrino. The second generation of matter particles contains copies of these quarks and leptons, but with a higher mass. The third generation contains even more massive particles, with the top (t) quark —the heaviest elementary particle known to date— and the bottom (b) quark. Interestingly, no theoretical restrictions exist on the number of additional generations that might exist. Quarks of a fourth generation would need to be very massive, as they have not yet been observed in high-energy experiments.

In order to search for new heavy quarks beyond the Standard Model, the Large Hadron Collider at CERN accelerates and collides protons at huge energies. Heavy quarks that might be produced in these collisions would decay to Standard-Model quarks and leptons, that can be detected with enormous particle detectors, such as the Compact Muon Solenoid (CMS) detector. By measuring the properties of these decay products, the underlying fundamental interactions and the heavy particles produced in the collision can be studied. In this thesis two searches for new heavy quarks beyond the Standard Model are described.

The first analysis, using proton-proton collision events collected by the CMS experiment in 2011, involves a simultaneous search for fourth generation up-type (t') and down-type (b') quarks, which are considered to be heavier copies of the top and the bottom quark, respectively. Motivated by the observed decay patterns of heavy Standard-Model quarks, the hypothetical t' quark is assumed to decay to a b quark and a W boson, the electrically charged force-carrier of the weak interaction. The b' quark is assumed to decay to a t quark and a W boson. Since top quarks and W bosons subsequently decay as well, one may obtain interesting final-state signatures in the detector. We exploit the fact that, in a consistent fourth-generation model, both

t' and b' quarks are expected to be produced in pairs via the strong interaction and singly via the electroweak interaction. By defining a set of optimized selection criteria, the collision events are categorized in subsamples that are sensitive to particular fourth-generation final-state topologies with at least one electrically charged lepton. As we do not observe any excess of collision events over the Standard-Model expectation, we set lower limits on the mass of t' and b' quarks. We exclude the existence of mass-degenerate fourth-generation quarks below a mass of 685 GeV/c^2 at the 95% confidence level. When the mass difference is assumed to be 25 GeV, the lower mass limits change by about 20 GeV. These results significantly reduce the allowed parameter space for a fourth generation of quarks.

In 2012 the ATLAS and CMS experiments discovered a new boson with a mass of 125 GeV/c^2 compatible with the H boson, which meant a breakthrough in particle physics. The existence of the H boson with the observed mass and production and decay properties is highly incompatible with a sequential fourth-generation model. However, many theoretical extensions of the Standard Model allow for the existence of a more exotic type of quarks, so-called vector-like quarks, that can evade the extremely stringent constraints from the H boson discovery. These quarks are expected to have different production and decay properties compared to Standard-Model quarks. For instance, they can interact with Standard-Model quarks not only via W bosons, but also via a Z boson, the electrically neutral force-carrier of the weak interaction, or a H boson. This potentially leads to a very rich phenomenology in proton-proton collisions.

The second analysis described in this thesis presents a search for vector-like quarks decaying to first-generation quarks, using the proton-proton collision events recorded by the CMS experiment in 2012 at even higher collision energies than in 2011. We consider the strong pair production and the electroweak single production of down-type vector-like (D) quarks. Since the decay of the vector-like quarks to W, Z and H bosons is considered, without assuming particular relative decay probabilities, this leads to a wide variety of possible final-state signatures. An inclusive search strategy is set up in order to increase the sensitivity to the production of these vector-like quarks decaying to first-generation quarks. Event categories are defined according to the number of charged leptons in the final state and additional selection criteria, such that each category is sensitive to a particular production and decay mode of the vector-like quarks. No excess of collision events over the Standard-Model expectation is observed. By scanning over the relative decay probabilities to W, Z and H bosons, as well as an electroweak single-production strength parameter, lower limits on the mass of the vector-like D quarks are set. We derived lower mass limits between 500 GeV and 1800 GeV at the 95% confidence level, depending on the set of scanned parameters.

In this thesis, we searched for new heavy quarks in the data collected by the CMS experiment in 2011 and 2012, but found no evidence for the existence of new quarks. Nevertheless, the presented analyses excluded a large amount of parameter space, guiding particle physicists in new directions to improve the understanding of the laws of nature and the elementary particles. The search for new quarks is not over, and with the startup of the Large Hadron Collider in 2015 at almost twice the collision energy of 2011, exciting times lie ahead.

Samenvatting

Zoektocht naar nieuwe zware quarks met de CMS detector bij de Large Hadron Collider

Het ultieme doel in elementaire deeltjesfysica is het opstellen van een theorie die alle fundamentele bouwstenen van ons universum en hun onderlinge interacties beschrijft. Doorheen de twintigste eeuw hebben fysici het Standaard Model van de deeltjesfysica geformuleerd, een wiskundig elegante theorie die subatomaire deeltjes en hun onderlinge interacties beschrijft op het meest fundamentele niveau. Het Standaard Model is uitzonderlijk succesvol gebleken bij het verklaren van talrijke observaties in hoge-energie experimenten. Bovendien werden vanuit deze theorie diepgaande voorspellingen gedaan, zoals over het bestaan van het intussen beroemde Brout-Englert-Higgs (H) boson, hetgeen experimenteel werd aangetoond in 2012 in het CERN laboratorium.

Het blijkt dat in deze theorie de fundamentele materiedeeltjes, de zogenaamde quarks en leptonen, gecategoriseerd kunnen worden in drie families of 'generaties'. De eerste generatie bestaat uit de up en down quark (waaruit protonen en neutronen zijn opgebouwd), en het elektron en elektron-neutrino. De tweede generatie van materiedeeltjes bevat kopieën van deze quarks en leptonen, maar met een hogere massa. De derde generatie bevat nog massievere deeltjes, met de top (t) quark —het tot nog toe zwaarst gekende elementaire deeltje— en de bottom (b) quark. Opvallend genoeg is er geen theoretische beperking op het aantal bijkomende generaties die zouden bestaan. Quarks van een vierde generatie zouden een zeer grote massa moeten bezitten, aangezien ze nog nooit in hoge-energie experimenten zijn waargenomen.

Om te zoeken naar nieuwe zware quarks, versnelt en botst de Large Hadron Collider bij het CERN protonen met een ontzagwekkende energie. Zware quarks die mogelijk in deze botsingen geproduceerd worden, zouden vervallen naar Standaard-Model quarks en leptonen, die gedetecteerd kunnen worden in enorme deeltjesdetectoren, zoals de Compact Muon Solenoid (CMS) detector. Door het meten van de eigenschappen van deze vervalproducten, kan men de onderliggende fundamentele interacties en de zware deeltjes die in de botsing geproduceerd worden onderzoeken. In deze thesis worden twee zoektochten naar nieuwe zware quarks beschreven.

In de eerste analyse bestuderen we proton-proton botsingen die geregistreerd zijn door het CMS experiment in 2011. In deze data zoeken we simultaan naar up-type (t') en down-type (b') quarks van de vierde generatie, die beschouwd worden als zwaardere kopieën van respectievelijk de t en de b quarks. Gemotiveerd door het geobserveerde vervalpatroon van Standaard-Model quarks, nemen we aan dat de hypothetische t' quark vervalt naar een b quark en een W boson, het elektrisch geladen krachtdragend

deeltje geassocieerd met de zwakke wisselwerking. Voor de b' quark wordt aangenomen dat het vervalt naar een t quark en een W boson. Aangezien t quarks en W bosonen op hun beurt ook vervallen, zou men dus interessante karakteristieken waarnemen in de detector. We maken eveneens gebruik van het feit dat men, in een consistent vierde-generatie model, zowel t' en b' quarks kan verwachten, waarbij de quark en antiquark samen geproduceerd worden via de sterke wisselwerking en enkelvoudig via de zwakke wisselwerking. Door geoptimaliseerde selectiecriteria te definiëren kan de totale verzameling van botsingsgebeurtenissen ingedeeld worden in kleinere verzamelingen die gevoelig zijn aan specifieke vierde-generatie topologieën met tenminste één elektrisch geladen lepton. We observeren geen overschot aan botsingsgebeurtenissen ten opzichte van wat we verwachten voor Standaard-Model processen, en daarom bepalen we een ondergrens op de massa van t' en b' quarks. Op deze manier sluiten we het bestaan uit van vierde-generatie quarks met een massa kleiner dan 685 GeV/c^2, met 95% betrouwbaarheid, indien t' en b' quarks een gelijke massa hebben. Wanneer een massaverschil tussen t' en b' quarks van 25 GeV verondersteld wordt, veranderen de ondergrenzen op de massa met ongeveer 20 GeV. Deze resultaten reduceren sterk de toegelaten parameterruimte voor een vierde generatie van quarks.

In 2012 ontdekten de ATLAS en CMS experimenten een nieuw boson met een massa van 125 GeV/c^2 compatibel met een H boson, wat een doorbraak betekende in de deeltjesfysica. Het bestaan van het H boson met de geobserveerde massa en productie- en vervaleigenschappen is incompatibel met het daarnet beschouwde vierde-generatie model. Vele theoretische uitbreidingen van het Standaard Model laten echter het bestaan toe van een meer exotisch type van quarks, zogenaamde *vector-like* quarks, die de bijzonder strikte beperkingen van de ontdekking van het H-boson kunnen omzeilen. Deze quarks zouden verschillende productie- en vervaleigenschappen bezitten vergeleken met Standaard-Model quarks. Zo kunnen vector-like quarks bijvoorbeeld niet alleen interageren met Standaard-Model quarks via de W bosonen, maar ook via een Z boson, het elektrisch neutrale krachtdragend deeltje geassocieerd met de zwakke wisselwerking, of een H boson. Dit leidt tot een potentieel zeer rijke fenomenologie in proton-proton botsingen.

De tweede analyse in deze thesis beschrijft een zoektocht naar vector-like quarks die vervallen naar eerste-generatie quarks, gebruik makende van proton-proton botsingen geregistreerd door het CMS experiment in 2012, bij een nog hogere botsingsenergie dan in 2011. We beschouwen zowel paarproductie van down-type vector-like (D) quarks via de sterke wisselwerking, als enkelvoudige productie via de zwakke wisselwerking. Aangezien het verval van de vector-like quarks naar W, Z en H bosonen beschouwd wordt zonder specifieke relatieve vervalwaarschijnlijkheden aan te nemen, geven deze quarks aanleiding tot een ruime verscheidenheid aan mogelijke configuraties en eigenschappen van de gedetecteerde vervalproducten. We stellen een inclusieve zoekstrategie op om de gevoeligheid te verhogen voor de productie van deze vector-like quarks die vervallen naar eerste-generatie quarks. Botsingsgebeurtenissen worden ingedeeld in deelverzamelingen volgens het aantal geobserveerde leptonen en bijkomende selectiecriteria, zodanig dat elke verzameling gevoelig is aan een bepaalde productie- en vervalmodus van de vector-like quarks. We observeren geen overschot aan gebeurtenissen ten opzichte van de Standaard-Model verwachting. Door te scannen over de relatieve

vervalwaarschijnlijkheden naar W, Z en H bosonen, en over een parameter die de sterkte van enkelvoudige D quark productie bepaalt, kunnen we ondergrens afleiden op de massa van vector-like D quarks. We bepalen ondergrens op de massa tussen 500 GeV en 1800 GeV met 95% betrouwbaarheid, afhankelijk van de waarden van de gescande parameters.

In deze thesis zochten we naar nieuwe zware quarks in proton-proton botsingen geregistreerd met de CMS detector in 2011 en 2012. We vonden geen significante aanwijzingen voor het bestaan van nieuwe quarks. Met de beschreven analyses sluiten we echter een groot deel van de parameterruimte uit, wat bijdraagt tot een verbeterd begrip van de natuurwetten en de elementaire deeltjes. De zoektocht naar nieuwe quarks is nog niet ten einde, en met de heropstart van de Large Hadron Collider in 2015 met bijna twee keer de botsingsenergie bereikt in 2011, wachten spannende tijden.

Acknowledgements

The fascinating journey through particle physics is one that cannot be made alone. Throughout the four years as a PhD student, I could count on the valuable expertise and help of many people. The enthusiasm of my promotor Jorgen D'Hondt, who introduced me to particle physics research when I was still a Bachelor student, has always been extremely inspiring. Thank you for guiding me through all crucial moments during the PhD research, and for being a true leader instead of a boss. Already from the beginning of my research, I worked closely together with my copromotor Petra Van Mulders. Your LHC data Survival Kit has proven to be very useful, as it seems I more or less survived! I cannot describe how grateful I am for the countless discussions we had about physics, but also about all kind of random things, which made the research so much more fun. I also want to thank the members of the jury, as I really appreciate the efforts they did to read my manuscript in such a short time. I think this thesis greatly benefited from your feedback.

During my PhD research I was very lucky to have worked with amazing colleagues, with whom I could talk about both work-related and non-work-related things. Working in versatile and multicultural labs like the Interuniversity Institute for High Energies and CERN has been extremely enriching on a personal and professional level. I started in a small, relatively dark office, and ended up in a large office that was in fact so bright that I had to buy a parasol. I shared the first office with Michael, Stijn and Alexis. I really want to thank you for all the great help you provided to a less experienced researcher, and for the many discussions during lunch, summer schools, conferences and at CERN. It was a real pleasure to meet amazing postdoctoral researchers like Rebeca, Grégory, Nadjieh, James, Taejeong, Natalie and Nadir. I also thank my colleague PhD students and office coworkers Annik, Quentin, Isis, Lieselotte, Kevin, Isabelle and Shimaa, and Master student Bert. I think each one of you will deliver a great thesis and will one day have the honor to write acknowledgements. Thank you all for being such enjoyable company, making the atmosphere in the office unique!

I want to express my gratitude to the phenomenologists I worked with, Kentarou, Karen, Bettina, Mathieu, Fabio and Anupama, with whom we had many fruitful and interesting discussions, and who aided in the implementation of the new-physics models in the event generators. Also special thanks go to the computing wizards Olivier, Shkelzen and Stéphane for their persistent technical support. And of course, for all administrative and logistic support I could always count on Marleen.

A thesis cannot be realized without the unconditional support of family and friends. Firstly, I want to thank my parents for believing in me and letting me pursue a career in science, rather than forcing me to become a professional footballer like so many

parents do to their child. Your support during often intense times has been crucial in the success of this research! My brother Stijn and my sister Ines are of course the most fun brother and sister one can imagine, although I might be able to imagine even better brothers and sisters if I try really hard. I should also thank all my friends, who supported me and made sure I was not totally consumed by work. Thank you Raf, Leen, Kristine, Joke, Gert, Sophie, Laurent, Ruben, Lynx, Karen, Robbe, Eva, Eline, Kaat, Jeroen, Tinne, Tom, Ayla, Vanessa, Elke, Jonas, Bavo, Robert, Emanuele and all my other concert-band, tennis and CERN friends for the nice times and your support! A final word of gratitude goes to my cats, who fed me all these years and let me sleep in their house.