Vrije Universiteit Brussel

VRIJE
UNIVERSITEIT
BRUSSEL

Faculteit Wetenschappen en Bio-ingenieurswetenschappen
Departement Natuurkunde

A search for flavour changing neutral currents involving a top quark and a Higgs boson with the CMS experiment

Kevin Deroover

Promotor: Prof. Dr. Jorgen D'Hondt
Co-promotor: Dr. Kirill Skovpen
Proefschrift ingediend met het oog op het behalen van
de academische graad Doctor in de Wetenschappen

September 2017

Doctoral examination commission:

Prof. Dr. Nick van Eijndhoven (Vrije Universiteit Brussel, chair)
Prof. Dr. Freya Blekman (Vrije Universiteit Brussel, secretary)
Prof. Dr. Jorgen D'Hondt (Vrije Universiteit Brussel, promotor)
Dr. Kirill Skovpen (Vrije Universtiteit Brussel, co-promotor)
Prof. Dr. Nuno Castro (LIP, Minho)
Prof. Dr. Alberto Mariotti (Vrije Universiteit Brussel)
Prof. Dr. Pascal Vanlaer (Université Libre de Bruxelles)
Prof. Dr. Nathalie Vermeulen (Vrije Universiteit Brussel)

Cover illustration: Visualisation of the reconstructed particles in the CMS detector originating from a proton-proton collision at a center of mass energy of 13 TeV, where a pair of top quarks decay semi-muonic. The back illustration shows Feynman diagrams of flavour-changing neutral currents involving a top quark and a Higgs boson.

Contents

Introduction

Nature has always bewildered the curious mind. It are such curious minds that set on the adventure of uncovering the workings of the universe. By means of experimental discoveries, theoretical predictions, through trial and (many) error(s), a vast field of exact sciences arose as a result of that, rendering humankind able of grasping the subtleties of nature. The current knowledge of the most fundamental building blocks of nature and their interactions is described in the Standard Model of elementary particle physics. The Standard Model has proven to be a very accurate description at electroweak energy scales, making countless predictions which have been experimentally verified with breathtaking precision. The discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments, is the most recent demonstration of the predictive power of the Standard Model.

However, many phenomena are yet unexplained by the Standard Model, such as the existence of dark matter and dark energy. All in all, the Standard Model is able to account for only 4% of the universe's energy content. A whole series of not yet refuted theories, like Supersymmetric extensions to the Standard Model or including extra dimensions, can account for phenomena like dark matter, though have no experimental evidence so far. The search for such new physics in current experiments is not straightforward, as its characteristics are often hard to disentangle from those of the Standard Model. However, processes where new physics could be significantly enhanced with respect to the Standard Model are so-called Flavour-Changing Neutral Currents (FCNC). These processes change the flavour of quarks, by means of a neutral boson, without changing their charge and are highly suppressed in the Standard Model, far beyond the sensitivity of current experiments. The observation of such processes would be clear evidence of new physics, which makes the search for FCNC vital in the quest for solving the mysteries of the universe.

The high suppression of FCNC in the Standard Model was first experimentally observed by measuring the decay rates of kaons in the 1960s. Subsequently, Glashow, Iliopoulos and Maiani [1] provided an explanation for this observed suppression in 1970 by introducing a new elementary particle to the Standard Model, called the charm quark. The direct observation of the charm quark in the J/Ψ meson decay [2, 3] in 1974 was the final confirmation for the validity of the FCNC suppressing mechanism developed by Glashow, Iliopoulos and Maiani. A first observation of FCNC involving bottom quarks, matching the predictions from the Standard Model, was made in

2005 at the CDF experiment [4]. Another precise observation of rare FCNC involving bottom quarks was made in a combination of data from the CMS and LHCb experiments [5] in 2015. All these observations agree with Standard Model predictions and therefore put stringent constraints on new-physics models. However, a recent reinterpretation [6] in 2013 of experimental results on FCNC measurements involving bottom quarks shows some first discrepancies with Standard Model predictions that could indicate the existence of a particle with a mass of about 3 TeV/c^2. To date, searches for FCNC involving top quarks at the ATLAS and CMS experiments [7–11] are still far from being sensitive to Standard Model predictions. This leaves a large window of unexplored phase space where new physics could be hiding, making the search for FCNC involving top quarks extremely interesting.

In this thesis, FCNC are investigated in high-energy proton-proton collisions at a center-of-mass energy of 13 TeV, produced by the Large Hadron Collider and recorded by the CMS experiment at CERN in 2016 with an integrated luminosity of 36 fb^{-1}. More specifically the coupling of top quarks to up (or charm) quarks and a Higgs boson, decaying into a pair of bottom quarks, is investigated. The Standard Model of elementary particles is shortly introduced in the first chapter of this thesis, with a focus on the flavour regulating mechanisms and the current limits on FCNC involving top quarks. Chapter 2 provides a description of the Large Hadron Collider and CMS experiment, where the proton collision data is recorded as used in this thesis. In order to understand the actual collision data, simulations of known collision processes are used as a comparison. The mechanisms on which simulations are based, together with the reconstruction and identification of detector signatures, is explained in Chapter 3. In Chapter 4, a phenomenological study of FCNC involving top quarks is performed to estimate the sensitivity of actual searches in proton collision events. The analysis strategy developed in this thesis for the search for FCNC with the CMS experiment, is described in Chapters 5 and 6. The first of those two chapters details the selection requirements on collision events and the technique developed to reconstruct FCNC event signatures. The second analysis chapter describes how the search discriminates FCNC events from known Standard Model processes and the obtained limits on FCNC processes. In the final chapter, the search is concluded with a reflection on the obtained results, a comparison with similar measurements and an outlook to the future of this analysis.

The author's personal work is represented in almost all parts of the entire Chapters 5 and 6. This work is, at the time of writing this thesis, going through the final steps of an internal review by the CMS collaboration. It is expected that the results in this thesis will be made public by the CMS collaboration shortly after in the form of a publication as written in [12]. Furthermore, in Chapter 4, the phenomenological study of top-FCNC in the H → b$\bar{\text{b}}$ channel has been conducted by the author. Besides his work on FCNC, the author has also contributed to the development and maintenance of the most performing and most recent b-tagging algorithms (see Section 3.3.5) as experimental work beneficial for numerous results published by the CMS collaboration.

Parts of this work are documented in [13], which, at the time of writing this thesis, has passed the final steps of an internal review by a dedicated CMS analysis review committee and will soon be submitted to a journal.

Chapter 1

Flavours in and beyond the Standard Model

It has always been apparent that invisible mechanisms are responsible for the workings of nature. At first, supernatural entities such as god(s) were attributed to those mechanisms where humankind lacked the appropriate knowledge. However, as human knowledge culminated through the ages, science arose as a more relevant explanation. Currently, the best knowledge about the fundamental constituents of nature is encapsulated in the Standard Model (SM) of elementary particle physics, which will be introduced in Section 1.1. Though the SM describes phenomena observed in dedicated experiments with extraordinary accuracy, some observed phenomena lie beyond the scope of the SM. This drives physicists to further develop the current understanding of nature by expanding on the SM. Many extensions of the SM predict the occurence at a higher rate than in the SM of so-called flavour-changing neutral currents (FCNC), which are discussed in Section 1.2.

1.1 The Standard Model of particle physics

The fundamental building blocks of nature and their interactions are described in the theory of the Standard Model. The SM succeeds in combining the weak, strong and electromagnetic forces into a consistent quantum field theory [14, 15]. A quantum field theory is a theoretical framework in which elementary particles and their interactions are described by treating them as fields. In such a framework the interactions between elementary particles are propagated by force fields, of which the associated particles are called force carriers. The particle content of the SM is summarised in Section 1.1.1, followed by a brief introduction to the underlying theoretical framework of the quantum field theory in Section 1.1.2. Finally, some of the shortcomings of the SM are discussed in Section 1.1.3.

1.1.1   Particle content and interactions

All visible matter in the universe is made up of matter particles, called fermions, which interact via force carriers, called bosons. Fermions carry half-integer spin, whereas bosons have integer spin. Every fermion has an anti-fermion, which has the same mass and is oppositely charged. Furthermore, fermions can be classified into particles carrying colour charge[1], which are called quarks, and those who do not, which are called leptons. A distinction is made between leptons that have charge $\pm 1e$, where e represents the electron charge, and neutral leptons, referred to as *neutrinos*. Fermions are categorized into three generations, each identical to the other except for their mass and flavour quantum number. The *up* and *down* quark, together with the leptons *electron* and *electron-neutrino*, form the first generation. The second generation consists of the *charm* quark, *strange* quark, *muon* and *muon-neutrino*. Finally, the *top* and *bottom*[2] quark make up the third generation, together with the *tau* and *tau-neutrino*. The different types of quarks and leptons are also characterized as *flavours*.

     The gauge bosons that are the force carriers of the strong interactions, are called gluons g, which are electrically neutral, massless and carry colour charge. Another neutral and massless boson is the photon γ which mediates the electromagnetic force. The third force described by the SM, the weak force, gives rise to three massive bosons: the charged $W^\pm$ and neutral Z^0. The SM is completed by a scalar boson, commonly referred to as the *Higgs boson*, experimentally confirmed on July 4 2012 [16]. The symmetry breaking mechanism is held responsible for giving mass to the elementary particles, as theorized by R. Brout, F. Englert [17] and P. Higgs [18]. The particle content of the SM, together with the particles' properties, is pictorially represented in Figure 1.1.

     The quarks of the SM are bound through gluons into mesons and baryons. Mesons (e.g.: π^0, K^0) are composed of a quark and an anti-quark. Baryons, such as protons and neutrons, are composed of three quarks. Almost all mesons and baryons are short-lived and unstable particles and decay via $W^\pm$ and Z^0 bosons, usually associated with a charged lepton, neutrino and quarks. The only stable baryon is the proton, consisting of two up quarks and one down quark.

1.1.2   Theoretical Framework

The theoretical framework of the SM is a quantum field theory. The dynamics and kinematics of particle and force fields in the theory are controlled by a Lagrangian $\mathcal{L}$. The Lagrangian of the SM is constructed such that it contains a certain set of symmetries which represent physics conservation laws. Conservation of energy, momentum, angular momentum and charge are just a few examples of conservation laws coming from symmetries which need to be represented in the SM Lagrangian. In order

---

[1]The colour charge indicates the ability of a particle to interact via the strong force.
[2]The bottom quark is also referred to as the beauty quark.
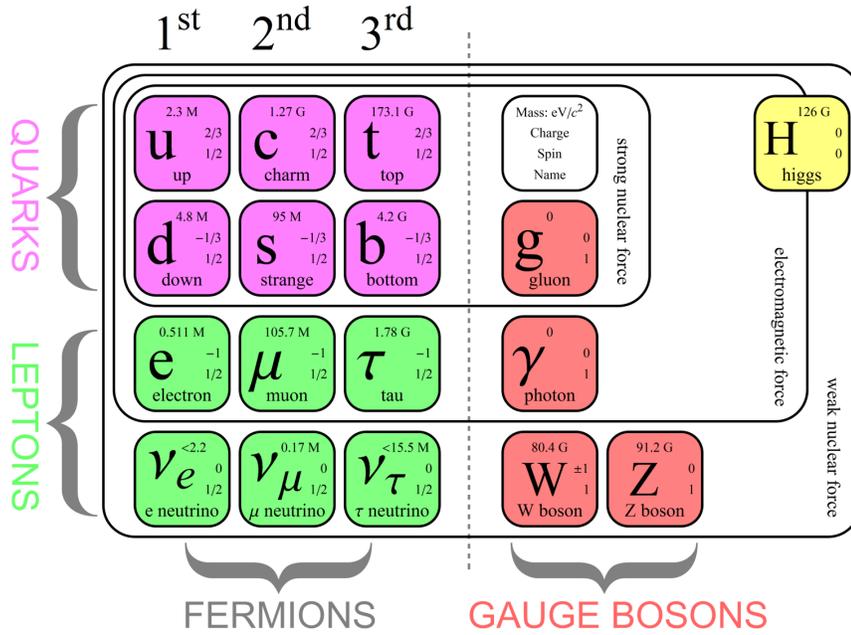
Figure 1.1: The fermions and gauge bosons of the Standard Model, with their experimentally confirmed properties as indicated [19].

to sustain such symmetries under local group transformations, *gauge invariance* is imposed.

Gauge invariance

Consider for example the wave function of a spin-1/2 fermion, with mass m and four-coordinates x in space-time, to be a Dirac spinor $\psi(x)$, for which the Lagrangian in natural units ($\hbar = c = 1$) is

$$\mathcal{L}_{\mathrm{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi, \tag{1.1}$$

with γ^μ the Dirac matrices[3] and $\bar{\psi} = \psi^\dagger\gamma^0$ with $\psi^\dagger$ the hermitian conjugate of ψ. This Lagrangian is invariant under global phase transformations ($\psi(\mathrm{x}) \to \psi'(\mathrm{x}) = \mathrm{e}^{ig\alpha}\,\psi(\mathrm{x})$, with g and α constant). Now consider a *local* phase transformation represented by the unitary matrix $\mathrm{U}(\alpha)$ of a group acting on the fermion field as

$$\psi(x) \to \psi'(x) = U(\alpha(x))\psi(x) = e^{ig\alpha(x)^\nu T_\nu}\psi(x), \tag{1.2}$$

with α^ν representing the transformation parameters, T_ν the generators of the group and g a real constant. Summation over ν is implied in this notation, where ν runs

---

[3] The Dirac matrices are 4×4 matrices, defined by $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ with $\eta^{\mu\nu} = \mathrm{diag}(+\ -\ -\ -)$ the Minkowski metric. The notation $\{a,b\}$ denotes the anti-commutator, i.e. $ab + ba$.

from 1 to the dimension of the group (i.e. the number of generators). Inserting (1.2) in (1.1) one gets the following Lagrangian:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - g\bar{\psi}(x)\gamma^\mu\psi(x)\partial_\mu(\alpha^\nu(x)T_\nu). \qquad (1.3)$$

Consequently the Dirac Lagrangian as defined in (1.1) is not invariant under local transformations. To restore the internal symmetry, the partial derivative is replaced with a covariant derivative, defined by

$$\partial_\mu \to \mathcal{D}_\mu = \partial_\mu - igT_\nu A^\nu_\mu, \qquad (1.4)$$

with A^ν_μ introduced as vectorial gauge fields. The transformation properties of the vectorial gauge field are determined such that they ensure invariance under local transformation. This whole process is referred to as imposing gauge invariance on a Lagrangian.

Substituting (1.4) into (1.1) gives

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^\mu T_\nu A^\nu_\mu\psi. \qquad (1.5)$$

This Dirac Lagrangian led to the development of quantum electrodynamics (QED), in which the transformation of the Lagrangian is considered under the Abelian group[4] U(1) and is the starting point of the SM. The corresponding interpretation of the gauge field $A_\mu(x)$ has lead to the description of the photon. The Dirac Lagrangian describes the propagation of free leptons (first two terms) and their interactions through photons (third term). The interaction strength between fermions and gauge bosons is represented by 'g', called the *coupling constant*.

Electroweak theory

The electroweak theory (EW) is described by requiring gauge invariance under the $SU(2)_L \otimes U(1)_Y$ group. The non-Abelian group $SU(2)_L$ has three generators, introducing three gauge fields W^α_μ ($\alpha \in$ {1,2,3}). The subindex L indicates that the gauge fields only couple to left-handed[5] fermions, as required by the observed parity-violating nature of the weak force. The left-handed fermions are grouped into *doublets* (e.g. the doublet (e_L, ν_{eL})) in the SM, while the right-handed fermions remain isolated in *singlets*. The Abelian group $U(1)_Y$ has one generator, namely the hypercharge Y, leading to only one gauge field B_μ. Two coupling constants are introduced in this gauge symmetry group, namely g_1 for $U(1)_Y$ and g_2 for $SU(2)_L$.

The physically observable gauge bosons of the electroweak theory, like the photon field A_μ and the Z^0_μ- and $W^\pm_\mu$-fields, are superpositions of the four gauge fields of $SU(2)_L \otimes U(1)_Y$:

$$A_\mu = \sin\theta_W W^3_\mu + \cos\theta_W B_\mu, \qquad (1.6)$$

---

[4] In an Abelian group the generators T_ν commute, whereas in non-Abelian groups the commutator of generators is not zero, e.g. in SU(2) $[T_\nu, T_\mu] = i\sum_\gamma^3 \epsilon^{\mu\nu\gamma}T_\gamma$ with $\epsilon^{\mu\nu\gamma}$ the Levi-Civita symbol.

[5] The handedness or *chirality* of a Dirac fermion field is defined as $\psi_L = \frac{1}{2}(1 - \gamma_5)\psi$ for left-handed fermions and $\psi_R = \frac{1}{2}(1 + \gamma_5)\psi$ for right-handed fermions, with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^4$

$$Z_\mu = \cos\theta_\mathrm{W} \mathrm{W}_\mu^3 - \sin\theta_\mathrm{W} \mathrm{B}_\mu, \qquad (1.7)$$

$$W_\mu^\pm = \sqrt{\frac{1}{2}}(W_\mu^1 \mp iW_\mu^2), \qquad (1.8)$$

with θ_W the weak mixing angle, defined as

$$\tan\theta_\mathrm{W} = \frac{\mathrm{g_1}}{\mathrm{g_2}}. \qquad (1.9)$$

Quantum Chromodynamics (QCD)

Quantum chromodynamics is represented in the Standard Model by the non-Abelian group $SU(3)_c$. This group has eight generators and thereby eight associated gauge boson fields G_μ^α ($\alpha \in \{1,2,..,8\}$), which are massless and known as the physical gluon fields. The subindex c denotes the fact that QCD only describes the interactions of coloured particles, namely gluons and quarks. The conventional names for the colour charges are *red*, *green* and *blue* and their anti-colours. The coupling constant representing the strength of QCD interactions is denoted as α_s.

An important characteristic of QCD is asymptotic freedom, which states that the strong coupling constant becomes weaker as the energy with which the interaction between strongly interacting particles is probed increases, and stronger as the distance between such particles increases. This phenomenon results in *colour confinement*, meaning quarks and gluons can not exist on their own and are not observed as individual particles. Instead, all quarks and gluons are bound in colour-neutral states called hadrons. The main consequence of colour confinement is the hadronisation, i.e. the formation of hadrons, of individual quarks and gluons.

Electroweak symmetry breaking

The SM does not contain mass terms for the elementary particles, hence it can not explain the observed particle masses. Such mass terms can not be introduced explicitely, because this would break gauge invariance. In order to accomodate mass for the massive bosons of the SM in a gauge invariant way, a complex scalar doublet ϕ is introduced, with a non-zero vacuum expectation value (VEV) v. The introduction of the scalar doublet induces an interaction between ϕ and the massive boson fields, giving rise to mass terms in the Lagrangian, which are gauge invariant. In a similar way, mass is given to fermions by adding interaction terms between ϕ and the fermionic fields in the Lagrangian. The coupling strength between ϕ and the fermion fields is quantified by the Yukawa couplings. The Lagrangian for the scalar boson is

$$\mathcal{L}_\phi = (D^\mu\phi)^\dagger (D_\mu\phi) - V(\phi), \qquad (1.10)$$

with V the potential of the scalar doublet. The scalar potential V is defined as

$$V(\phi) = -\frac{1}{2}\mu^2\phi^\dagger\phi + \frac{1}{4}\lambda^2(\phi^\dagger\phi)^2, \tag{1.11}$$

where μ is a real constant with units of mass and λ is a dimensionless parameter that accounts for the quartic self interactions of the scalar field. The potential V has an infinite set of minima (ground states) and by expanding the field around an arbitrary choice of ground state, the electroweak symmetry is broken. This spontaneous symmetry breaking leaves gauge invariance of the SM intact and gives mass to the $W^\pm$ and Z bosons as:

$$m_W = \frac{1}{2}\,v\,|g_2| \qquad\qquad m_Z = \frac{v}{2}\sqrt{g_1^2 + g_2^2}. \tag{1.12}$$

Out of the four degrees of freedom from the complex scalar doublet, one gives rise to a physically observable particle, called the Brout-Englert-Higgs boson. The remaining degrees of freedom couple to the gauge fields and mix with the W^+, W^- and Z boson.

In order to be in agreement with experimental results and to get a renormalisable theory, meaning that the physical predictions in terms of a finite number of free parameters remain finite, the SM combines three symmetry groups into one gauge symmetry group

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y. \tag{1.13}$$

With the introduction of all gauge fields and coupling constants, the total covariant derivative (1.4) for the SM becomes

$$D_\mu = \partial_\mu + ig_1\frac{Y}{2}B_\mu + ig_2\frac{\tau^a}{2}W_\mu^a + i\alpha_s\frac{\lambda^a}{2}G_\mu^a, \tag{1.14}$$

with Y, τ^a and λ^a being respectively the hypercharge, Pauli matrices and the Gell-Mann matrices representing the generators of respectively $U(1)_Y$, $SU(2)_L$ and $SU(3)_c$. The full Lagrangian describing the SM arises by writing down the Dirac terms for the fermions (1.1) and replacing the partial derivative by the covariant derivative defined by the gauge symmetry group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ (1.14). Also the electroweak symmetry breaking sector is accounted for by adding kinetic and potential terms for the scalar doublet, together with the Yukawa coupling terms for fermion fields. Finally, kinetic terms for the gauge fields are included to provide a description of the propagation of free gauge bosons.

Quantum field theory and observable quantities

The quantum field theory description of the SM by means of a Lagrangian is a fundamental representation of the interactions between elementary particles. Making the

connection to physically observable quantities is, however, less straightforward. One way of calculating observable quantities of elementary particle interactions is through the Feynman rules. The Feynman rules are developed such that one can construct the so-called *matrix element* $\mathcal{M}$ of a process based on a pictorial representation of the process, which is referred to as a Feynman diagram. The matrix element represents the core of the interaction, containing information about the four-momenta and masses of the involved particles, as well as the interaction vertices derived from the Lagrangian. For example, the cross section of an interaction process, which is an observable measure for the occurence of the process, is proportional to $|\mathcal{M}|^2$. Figure 1.2 represents a Feynman diagram of the interaction of a quark and anti-quark producing a Z boson or virtual photon[6], which decays into a charged lepton pair. This process is commonly referred to as the Drell-Yan process [20].
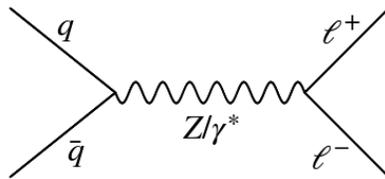
Figure 1.2: A Feynman diagram representation of the Drell-Yan process.

The Feynman diagram of a process which involves the minimal amount of particles is called the tree-level Feynman diagram and is also referred to as the Leading-Order process. For every interaction vertex added to the diagram, the order of the process increases. To obtain the highest precision of an observable quantity, such as the cross section, the process should be considered to an infinite amount of orders. However, due to computational limitations, one typically is limited to Next-to-Leading-Order (NLO) calculations depending on the process.

1.1.3 Drawbacks of the Standard Model

The SM is not a complete theory and does not provide a full description of nature. First of all the theory only successfully combines three out of four fundamental forces. By not accounting for the gravitational force, the SM does not describe phenomena where quantum and gravitational effects are equally important, such as at the Big Bang or the center of black holes.

An uncharming aspect of the SM, known as the *hierarchy problem* [21], is its inability to explain the large gap between the electroweak symmetry breaking scale ($\sim 10^2$ GeV/c^2) and the Planck scale ($m_{Planck} \approx 10^{19}$ GeV/c^2), beyond which quantum and graviational effects are not reconcilable. The hierarchy problem becomes apparent

---

[6]A virtual particle is an off-shell particle that does not satisfy $E^2 = p^2 + m^2$, but has the same quantum numbers as its real counterpart.

when calculating the Higgs boson mass to all orders of perturbation:

$$m_H^2 = m_{H,0}^2 + \Delta m_H^2, \qquad (1.15)$$

with m$_H$ the real physical mass of the Higgs boson, $m_{H,0}$ its bare mass and Δm$_H$ the quantum loop corrections. The quantum loop correction to the Higgs boson mass due to fermions in the loop expand as [22]

$$(\Delta m_H^2)_f = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + ..., \qquad (1.16)$$

with λ_f the Yukawa coupling of the fermion and Λ_{UV} the ultraviolet cut-off[7]. The most dominant contribution comes from the top quark, which has the largest Yukawa coupling to the scalar field, caused by its large mass. If the SM is the only physics theory up to the Planck scale, the cut-off would be $\Lambda_{UV} =$ m$_{Planck}$. In that case, in order for the theorised prediction to agree with the experimentally measured mass (m$_H$ ≈ 125 GeV/c^2), a fine-tuning cancellation between the bare mass and the quantum loop corrections of over 30 orders of magnitude is needed, which would need to be accounted for by the new-physics arising at the Planck scale.

Another issue arises in the scaling of the three coupling constants g_1, g_2 and α_s to the energy scale with which infinities are regulated, referred to as the renormalisation scale. This phenomenon is also known as the running of coupling constants. Grand Unified Theories (GUTs) [23] aim to unify all forces at a certain energy scale, such that there is only one coupling constant at that energy scale, the GUT scale. In order to accomplish the idea of unification, these three SM coupling constants should be equal at a certain value for the renormalisation scale. The SM however does not predict such a unification, but instead predicts that the coupling constants do not converge at the Planck scale.

An issue which is also not addressed by the SM is the existence of non-baryonic dark matter and dark energy, making up respectively 23% and 72% of the universe's energy density. Cold dark matter is presumably only sensitive to the weak and gravitational force. Since the only known SM particles with these characteristics are neutrinos, being too light to accomodate this great amount of dark matter, other models are needed to accomodate such Weakly Interacting Massive Particles (WIMPs).

Some other issues the SM also struggles with are: why are there exactly three generations? How do neutrinos obtain mass? Is it possible to predict the values of

---

[7] Λ_{UV} is a cut-off parameter introduced to regulate divergent integrals arising from quantum loops at higher order calculations. The value of Λ_{UV} is usually chosen as the scale at which new physics becomes apparent.

all 19 free parameters[8] in the theory? Why is the observed CP violation[9] so strong? All these unsolved riddles make the SM an incomplete theory to describe all observed phenomena and indicate that it is expected to break down at energies between the electroweak symmetry breaking scale and the Planck scale.

1.2 Flavour-changing neutral currents (FCNC)

By classifying fermions according to flavour, the interplay between flavours through the weak interaction drove the experimental and theoretical study of the SM to new discoveries. Flavour-changing currents in the SM are discussed in Section 1.2.1. Flavour-changing neutral currents (FCNC) are highly suppressed in the SM. Many beyond the SM theories however, predict the occurence of FCNC at much higher rates than predicted in the SM. By means of an effective field theory description, FCNC can be studied in a model-independent way. This effective approach is described in Section 1.2.2. Since the aim of this thesis is the study of FCNC in the top quark sector, the focus in this chapter is shifted towards top-FCNC, summarizing the status of searches for top FCNC in Section 1.2.3.

1.2.1 Flavour-changing currents in the Standard Model

Of the four gauge bosons in the electroweak theory, only the charged W boson changes the flavour of quarks. When the first description of such flavour-changing charged currents was introduced in 1963 by Nicola Cabibbo [24], the only known quarks were the up u, down d and strange s quark. Cabibbo then postulated the charged (weak) current as a coupling between the up quark and $d_{weak} = [\cos\theta_c d + \sin\theta_c s]$, a linear combination of the down d and strange s quarks. The linear combination is a direct consequence of the chosen rotation

$$\begin{pmatrix} d_{weak} \\ s_{weak} \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}, \qquad (1.17)$$

where the rotation angle θ_c is known as the Cabibbo angle. The charged weak current between u and d quarks is then defined as

$$J_\mu = \overline{u}\gamma_\mu(1 + \gamma_5)d_{weak}. \qquad (1.18)$$

The introduction of Cabibbo's rotation matrix leaves the strange-quark interaction field $s_{weak} = [-\sin\theta_c d + \cos\theta_c s]$ uncoupled. Consequently, Glashow, Iliopoulos and

---

[8] These parameters are: 3 coupling constants (g_1, g_2 and α_s), 2 parameters from the Higgs sector, 9 masses of the quarks and leptons, 4 parameters related to flavour mixing in the CKM matrix and 1 related to QCD. The number of free parameters can even be increased when taking the neutrino mass giving mechanism into account.

[9] Charge conjugation Parity symmetry or CP symmetry dictates that the laws of physics are invariant if a particle is interchanged with its anti-particle. The violation of CP symmetry is strongly correlated to the fact that matter is more abundantly present in the universe than anti-matter.

Maiani (GIM) in 1970 [1] required the existence of a fourth quark with charge 2/3, the charm quark, to couple to s_{weak}. They defined the charged weak current as

$$J_\mu = \overline{U}\gamma_\mu(1 + \gamma_5)\mathcal{R}D, \tag{1.19}$$

with $\mathcal{R}$ the rotation matrix introduced in (1.17) and the matrices U and D being the column matrices of the up-type quarks (u and c) and down-type quarks (d and s) respectively. The existence of the charm quark was first observed in the J/Ψ meson decay [2, 3] in 1974. The neutral weak current, defined as

$$J_3 = \overline{U}\gamma_\mu(1 + \gamma_5)[\mathcal{R}, \mathcal{R}^\dagger]D, \tag{1.20}$$

is then diagonal in flavour space, meaning no FCNC occur at tree-level Feynman diagrams.

The introduction of a third generation of quarks to the SM leads to a generalisation of the Cabibbo rotation matrix by Kobayashi and Maskawa [25]. The result is a 3×3 unitary matrix, called the CKM matrix, controlling the mixing of the weak interaction states of down-type quarks as

$$\begin{pmatrix} d_{weak} \\ s_{weak} \\ b_{weak} \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.21}$$

The squared values of the matrix elements $V_{qq'}$ represent the transition probability of a quark q going to a quark q'. The unitarity of the matrix $(U^\dagger U = I)$ implies that the sum of the squared elements in one row (column) equals unity. The experimentally determined values for $V_{qq'}$ are currently [26]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97425 \pm 0.00022 & 0.2253 \pm 0.0008 & (4.13 \pm 0.49) \times 10^{-3} \\ 0.225 \pm 0.008 & 0.986 \pm 0.016 & (41.1 \pm 1.3) \times 10^{-3} \\ (8.4 \pm 0.6) \times 10^{-3} & (40.0 \pm 2.7) \times 10^{-3} & 1.021 \pm 0.032 \end{pmatrix}.$$

The CKM matrix indicates that top quarks predominantly transition via charged weak currents to bottom quarks, with a probability consistent with unity. FCNC in the top quark sector of the SM are still far beyond the sensitivity of all current experiments. Higher-order loop Feynman diagrams are the only way for FCNC to occur in the SM, but are consequently highly suppressed. Figure 1.3 shows an example of one-loop contributions to a FCNC process, where a top quark decays to a charm quark with the associated production of a photon. The expected transition probabilities (or branching ratio BR) of a top quark going to an up or charm quark through FCNC in the SM is summarised in Table 1.1.

1.2.2   An effective approach beyond the Standard Model

Since top-FCNC are highly suppressed in the SM, far beyond the sensitivity of current experiments, the observation of such processes would be clear evidence of new physics.
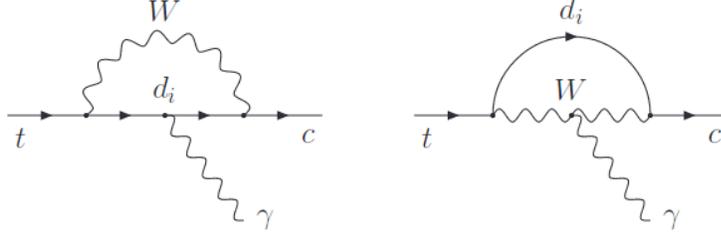
Figure 1.3: One-loop contribution to FCNC involving a top and charm quark in association with a photon [27].

A multitude of existing new-physics models predict strong enhancements of top-FCNC branching ratios, of which some are summarised in Table 1.1. The observation of top-FCNC that can be matched to these new-physics model predictions would be a first indication of the validity of those models. However, if observed top-FCNC don't match existing new-physics predictions, brand-new models would need to be developed.

Table 1.1: Predicted branching ratios, for top-FCNC in the SM, quark singlet (QS) models, a generic two-Higgs doublet model (2HDM) and the minimal supersymmetric extension to the SM (MSSM) [27].

Process	SM	QS	2HDM	MSSM
$t \to uZ$	8×10^{-17}	$\leq 1.1 \times 10^{-4}$	-	$\leq 2 \times 10^{-6}$
$t \to u\gamma$	4×10^{-16}	$\leq 7.5 \times 10^{-9}$	-	$\leq 2 \times 10^{-6}$
$t \to ug$	4×10^{-14}	$\leq 1.5 \times 10^{-7}$	-	$\leq 8 \times 10^{-5}$
$t \to uH$	2×10^{-17}	$\leq 4.1 \times 10^{-5}$	$\leq 5.5 \times 10^{-6}$	$\leq 10^{-5}$
$t \to cZ$	1×10^{-14}	$\leq 1.1 \times 10^{-4}$	$\leq 10^{-7}$	$\leq 2 \times 10^{-6}$
$t \to c\gamma$	5×10^{-14}	$\leq 7.5 \times 10^{-9}$	$\leq 10^{-6}$	$\leq 2 \times 10^{-6}$
$t \to cg$	5×10^{-12}	$\leq 1.5 \times 10^{-7}$	$\leq 10^{-4}$	$\leq 8 \times 10^{-5}$
$t \to cH$	3×10^{-15}	$\leq 4.1 \times 10^{-5}$	$\leq 10^{-3}$	$\leq 10^{-5}$

The FCNC enhancement is often caused by the introduction of new heavy particles in the model, which contribute via loops to the FCNC. In order to investigate FCNC enhancements in a model independent way, an effective Lagrangian is defined. The effective Lagrangian is part of an effective field theory (EFT) approach, which modifies the SM as

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i, \tag{1.22}$$

where Λ is the scale at which new physics is considered, c_i the coupling constants of the SM fields to the new-physics fields and $\mathcal{O}_i$ the higher-order effective operators. Such an EFT describes the SM as a low-energy theory to which higher-dimensional operator effects become apparent at higher energies. For simplicity, it is assumed that the

leading new-physics effects are solely described by dimension-six operators, the energy scale Λ being large, such that all effective higher-order contributions are suppressed by a power in Λ greater than 2 and can be neglected. The set of operators is reduced by removing redundant operators as discussed in [28]. After the spontaneous breaking of the electroweak symmetry, the operators generate flavour-changing neutral interactions of the top quark that are not present in the SM. The effective Lagrangian for top-FCNC interactions can be written as

$$
\begin{aligned}
\mathcal{L}_{eff} = \frac{1}{\sqrt{2}} \sum_{q=u,c} \Bigg[& \frac{g_1}{\Lambda} \kappa_{\gamma qt} A_{\mu\nu} \bar{t} \sigma^{\mu\nu} (f_{\gamma q}^L P_L + f_{\gamma q}^R P_R) q \\
& + \frac{g_2 \kappa_{Zqt}}{2 \cos\theta_w \Lambda} Z_{\mu\nu} \bar{t} \sigma^{\mu\nu} (f_{Zq}^L P_L + f_{Zq}^R P_R) q \\
& + \frac{\sqrt{2} g_2 \zeta_{Zqt}}{4 \cos\theta_w} \bar{t} \gamma^\mu (f_q^L P_L + f_q^R P_R) q Z_\mu \\
& + \frac{\alpha_s \kappa_{gqt}}{\Lambda} \bar{t} \sigma^{\mu\nu} T_a (f_{gq}^L P_L + f_{gq}^R P_R) q G_{\mu\nu}^a \\
& + \kappa_{Hqt} \bar{t} (f_{hq}^L P_L + f_{hq}^R P_R) q H + \text{h.c.} \Bigg],
\end{aligned}
\tag{1.23}
$$

following the naming conventions introduced in Section 1.1, with κ and ζ representing the coupling strengths of the respective interactions. Two coupling constants arise for the Zqt-vertex, which is a residue of the electroweak symmetry breaking[10] of $\mathcal{L}_{EFT}$. The left-handed and right-handed chirality projectors are represented by P_L and P_R, which can be modified by the chiral parameters f^L and f^R respectively. The normalisation of the chiral parameters is imposed as $|f^L|^2 + |f^R|^2 = 1$. The field strength tensors of the photon A_μ, Z boson Z_μ and gluon G_μ are introduced as

$$
\begin{aligned}
A_{\mu\nu} =& \partial_\mu A_\nu - \partial_\nu A_\mu \quad , \qquad Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu \quad , \\
& G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + \alpha_s f_{bc}^a G_\mu^b G_\nu^c ,
\end{aligned}
\tag{1.24}
$$

with f_{bc}^a being the structure constants of the SU(3) group.

The effective Lagrangian in (1.23) provides a description in terms of the coupling strengths κ_{tqX}, with X being a neutral gauge boson or Higgs boson. The connection to a branching ratio, which allows an easier interpretation across different EFT's, is made as

$$
BR(t \to qX) = \frac{\delta_{Xqt}^2 \Gamma_{t \to qX}}{\Gamma_t},
\tag{1.25}
$$

with $\Gamma_{t \to qX}$ the FCNC decay width[11] for coupling strength $\delta_{Xqt} = 1$ and Γ_t the full SM decay width of the top quark. Supposing a top-quark mass of 172.5 GeV/c^2, Γ_t becomes 1.32 GeV/c^2 [29].

---

[10]The dimension-six gauge-invariant operators $\mathcal{O}_i$ represent all possible anomalous couplings between SM fields. This includes that the massive Z boson appears both in the Z_μ-field (Eq. 1.7), as in the covariant derivative of the scalar field. The latter results in an extra Zqt-vertex.

[11]The decay width is a measure for the probability per unit time that a given particle will decay. The total decay width Γ_{total} is inversely proportional to the particle's lifetime.

For every top-FCNC vertex appearing in the Feynman diagram of a process, the matrix element $\mathcal{M}$ multiplies by a factor κ_{tqX}. Since the cross section of a process is quadratically proportional to $\mathcal{M}$, processes with one top-FCNC vertex scale quadratically to κ_{tqX}, or, equivalently, linearly to BR(t→qX).

1.2.3 Overview of experimental limits

The expected branching ratios for top-FCNC in the SM are of the order 10^{-12}-10^{-17} and thereby far beyond the reach of current experiments. Table 1.1 predicts enhanced top-FCNC which are possibly within the reach of current experiments. Many experimental searches for top-FCNC have been performed, based on an EFT approach. Such searches focus on one of the FCNC vertices described in the effective Lagrangian by putting all other FCNC coupling strengths to zero. Depending on the experimental setup and the investigated FCNC vertex, a search is usually sensitive to one of two distinct search modes: single-top production or $t\bar{t}$ decay. In single-top production, the FCNC vertex is responsible for the production of a single top quark, whereas in $t\bar{t}$ decay one looks at $t\bar{t}$ events where one of the top quarks decays via the FCNC vertex. These two cases are demonstrated in Figure 1.4 for the anomalous tqH-vertex, where the Higgs boson decays into two b-quarks and (one of the) top quark(s) decays leptonically.
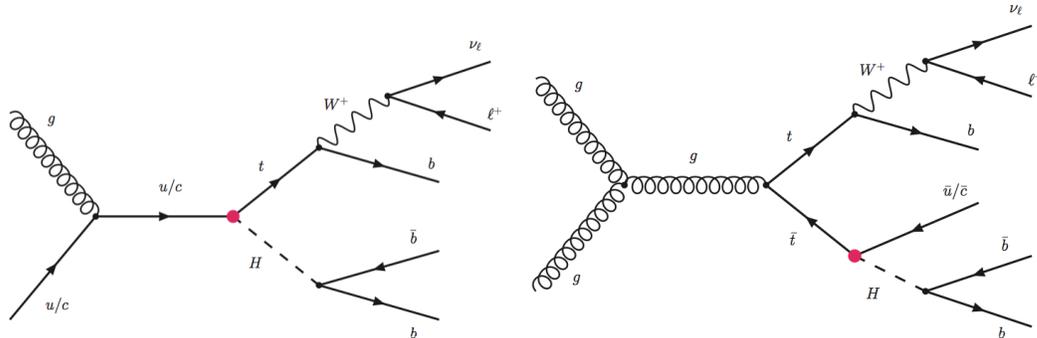
Figure 1.4: Feynman diagrams for FCNC tqH processes: associated production of top quark with Higgs boson (left) and FCNC decay of top quark in $t\bar{t}$ (right).

As the observation of top-FCNC eluded all searches so far, experimental upper bounds on the branching ratios have been determined. An overview of the current best experimental limits on top-FCNC is given in Table 1.2. A comparison of Tables 1.1 and 1.2 indicates that the current limits for top-FCNC involving a photon are still far away from any new-physics models' prediction. However, searches for top-FCNC involving a Higgs boson are narrowing in on some two-Higgs doublet model predictions and searches involving a Z boson or gluon become respectively sensitive to QS and MSSM models.

Table 1.2: Overview of the most stringent (observed and expected) experimental limits on top-FCNC branching ratios $\mathcal{B}$ at 95% confidence level.

Process	Search mode	Observed $\mathcal{B}$	Expected $\mathcal{B}$	Experiment	Ref.
$t \to uZ$	$t\bar{t}$ decay	2.2×10^{-4}	2.7×10^{-4}	CMS	[7]
$t \to u\gamma$	single-top production	1.3×10^{-4}	1.9×10^{-4}	CMS	[8]
$t \to ug$	single-top production	4×10^{-5}	3×10^{-5}	ATLAS	[9]
$t \to uH$	$t\bar{t}$ decay	4.5×10^{-3}	2.9×10^{-3}	ATLAS	[10]
$t \to cZ$	$t\bar{t}$ decay	4.9×10^{-4}	12×10^{-4}	CMS	[7]
$t \to c\gamma$	single-top production	1.7×10^{-3}	2.0×10^{-3}	CMS	[8]
$t \to cg$	single-top production	2×10^{-4}	1.5×10^{-4}	ATLAS	[9]
$t \to cH$	$t\bar{t}$ decay	4×10^{-3}	4.3×10^{-3}	CMS	[11]

Chapter 2

The CMS experiment at the Large Hadron Collider

One of the great successes of the SM is its predicitive power and the fact that it has been experimentally confirmed with great accuracy. In order to achieve the accuracy necessary to bolster theoretical claims from the SM and to investigate where it falls short, one needs state-of-the-art experimental instruments. Particle accelerators have been a very successful tool so far to experimentally probe the SM and beyond. From a collision of relativistic particles, a vast array of secondary particles are produced, of which the presence is recorded by particle detectors. From the recorded particles and their measured properties, the collision event can be reconstructed in order to provide insights in the interactions taking place at the collision.

The European Organization for Nuclear Research (CERN) at Geneva built the most powerful accelerator to date, called the Large Hadron Collider (LHC) [30], of which the main purpose is to accelerate and collide protons[1]. The main physics motivation for the LHC was twofold: the discovery (or exclusion) of the SM Higgs boson on the one hand and the search for physics beyond the SM on the other hand. The design of the LHC was mainly driven by the mass range in which the Higgs boson is predicted. This design is discussed in Section 2.1 together with the LHC operation.

Around the proton interaction points along the LHC, several particle detectors are built. One of them is the Compact Muon Solenoid (CMS), which is a multi-purpose particle detector. The collision data recorded by CMS is used in this thesis and will therefore be discussed in more detail in in Section 2.2.

2.1   The Large Hadron Collider

The LHC is the final acceleration stage of CERN's accelerator complex, which will be discussed in more detail in Section 2.1.1. Currently colliding beams of protons at a

---

[1]Apart from protons, lead-ions are also accelerated and collided in the LHC, but are not considered in this thesis.

center-of-mass energy of 13 TeV and with a design collision energy of 14 TeV, the LHC is the most powerful man-made particle accelerator. Four main detectors, along with a couple of smaller experiments, are built around the four different beam-interaction points (see Section 2.1.1). The design and operation of the LHC are explained in Section 2.1.2.

2.1.1   The CERN accelerator complex and LHC experiments

CERN has a rich history of collision physics, dating all the way back to its foundation in 1954. Several particle accelerators have come and gone over the past six decades at CERN, each and every one of them breaking grounds in terms of accelerator design as well as physics research. This has led CERN to become the leading laboratory in elementary (and nuclear) particle research, providing a vast accelerator complex. The resulting complex is a sequence of accelerators, each one boosting the energy of a particle to the next level. The LHC is the latest (and largest) addition to this complex.

The acceleration process, all the way up to the injection in the LHC, starts from the acquisition of protons. Protons are obtained from a hydrogen gas of which the electrons are stripped by an electric field. Once stripped from electrons, the gas enters a linear accelerator, called Linac 2, which acccelerates protons to an energy of 50 MeV. The resulting beam of protons is then boosted in the Proton Synchrotron Booster (PSB) to an energy of 1.4 GeV. A further acceleration to 25 GeV is provided by the Proton Synchrotron (PS), followed by another acceleration to 450 GeV by the Super Proton Synchrotron (SPS). From there on, the protons are injected in the LHC into two separate beam pipes, flowing in opposite directions, which ultimately accelerates protons to 6.5 TeV. The protons are injected in compressed bunches, containing 10^{11} protons each. The whole LHC injection chain is schematically presented in Figure 2.1.
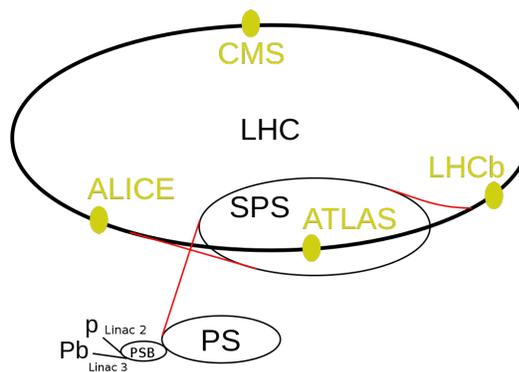
Figure 2.1: Schematic overview of the LHC injection chain and main experiments.

Along the LHC, four beam-interaction points are introduced (see Figure 2.1), around which different particle detectors are constructed. In this way, the LHC houses four

main experiments at those interaction points: ALICE [31], ATLAS [32], CMS [33] and LHCb [34]. The ATLAS and CMS experiments are two general purpose detectors at the LHC. Both experiments measure collision data (primarily from proton collisions) via their respective detectors in order to study both SM and beyond the SM physics. The complementarity of the ATLAS and CMS physics programs is set into place to have an independent confirmation of an experimental result through different detection techniques. Their greatest success was the discovery of the Higgs boson announced in July 2012. The ALICE experiment records lead collisions in order to study quark-gluon plasma[2]. The LHCb experiment performs precise measurements of CP-violation and rare decays, relying on the excellent identification of bottom quarks in proton collisons. Two smaller experiments reside at the LHC, next to the CMS and ATLAS detectors. Close to the CMS detector, the TOTEM experiment [35] measures elastic scatterings, diffractive processes and total cross sections of proton collisions. The other smaller experiment, LHCf [36], is located next to the ATLAS detector and measures forwardly produced neutral particles.

2.1.2   The LHC design

The LHC is built in a circular tunnel of 27 km circumference at a depth ranging between 50 and 175 m. The tunnel originally hosted the Large Electron Positron collider (LEP) between 1989 and 2000, which collided electrons with positrons at a center-of-mass energy of 208 GeV. Despite the many successes of LEP[3], it was not succesful at discovering the Higgs boson nor any beyond the SM physics. The desire to continue such searches and discover the Higgs boson drove the design of the LHC in the LEP tunnel, such that more energetic collisions could be produced at much higher collision frequencies.

   The collision-energy reach of LEP was mainly limited due to synchrotron radiation. Charged particles being accelerated to ultrarelativistic speeds in a circular orbit are subject to synchrotron radiation. Synchrotron radiation causes orbiting particles to lose energy at each revolution as

$$\Delta E \propto \frac{E^4}{m^4 R},$$

with E and m the energy and mass of the particle and R the radius of the orbit. Building a larger circular electron collider is very costly and it was therefore decided to collide protons instead of electrons[4] in the LEP-tunnel. As protons are 2000 times

---

[2]Quark-gluon plasma is a state of matter which emerges at extremely high temperatures and densities. This state is alleged to be present shortly after the Big Bang.

[3]The biggest accomplishments of LEP were the precise determination of the W- and Z-boson masses. The accurate determination of the Z-boson mass led to the conclusion that there are exactly three neutrino generations in the SM.

[4]No elementary particles besides the electron lives long enough to be appropriately accelerated. Therefore hadrons are the only alternative.

heavier than electrons, they are much less subject to synchrotron radiation. The LHC is designed to reach center-of-mass collision energies of 14 TeV, accelerating particles in each beam to 7 TeV.

The rate at which particles collide is represented by the instantaneous luminosity $\mathcal{L}$. In collider experiments, where bunches of particles are fired at each other in opposing directions, the instantaneous luminosity is calculated as

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y}, \tag{2.1}$$

with f the bunch crossing frequency, n_1 and n_2 the number of particles in the respective colliding bunches and σ_x and σ_y the widths of the bunches in the beam. The occurence rate N of a particular physics process scales linearly to the instantaneous luminosity and its cross section as

$$N = \mathcal{L} \sigma_{process}. \tag{2.2}$$

The design instantaneous luminosity of the LHC is $10^{34} \mathrm{cm}^{-2}\mathrm{s}^{-1}$, which is obtained by colliding bunches of protons, containing 10^{11} protons each and with $\sigma_x = \sigma_y = 17 \ \mu$m, every 25 ns. A direct consequence of squeezing so many protons into such a small bunch is the occurence of multiple collisions per bunch crossing. This feature is commonly referred to as pileup interactions and provides an added experimental challenge when analysing collision data. Additional pileup is caused due to the fact that bunches cross every 25 ns, during which the remnants of a previous crossing have only traversed about 7 meters of detector material. This effect is commonly referred to as out-of-time pileup.

The major change in design of the LHC with respect to LEP lies in the magnetic field configuration used to bend the opposing beams in a circular motion. As the colliding particles are no longer oppositely charged[5], as was the case at LEP, it is not possible to bend both beams with the same single magnetic field configuration. Given that the dimensions of the LEP tunnel are too small[6] to host two separate rings of beampipes, the LHC makes use of dipole magnets [37] for bending the opposing proton beams. Dipole magnets are a two-in-one magnetic field, designed such that it creates oppositely oriented magnetic fields in the two adjacent beampipes of the LHC. The LHC accommodates 1232 such dipole magnets, measuring 14 m and weighing around 35 tonnes. The production of these dipole magnets, costing about 0.5 million CHF each, was the most costly aspect of the LHC. Each dipole magnet can provide 8.4 T magnetic field strength at a current of 11.7 kA, which is achieved under superconducting conditions at 1.9 K. The cooling to 1.9 K is achieved by a closed liquid helium circuit. A schematic representation of a dipole magnet at the LHC is given in Figure 2.2. In order to focus the bunches to a width of 17 μm, 858 quadrupole magnets are implemented along the accelerator ring.

---

[5]It was no option to collide protons with anti-protons, as it is too difficult to produce the amount of anti-protons necessary to achieve the design instantaneous luminosity.

[6]The diameter of the LEP tunnel is about 4 m.
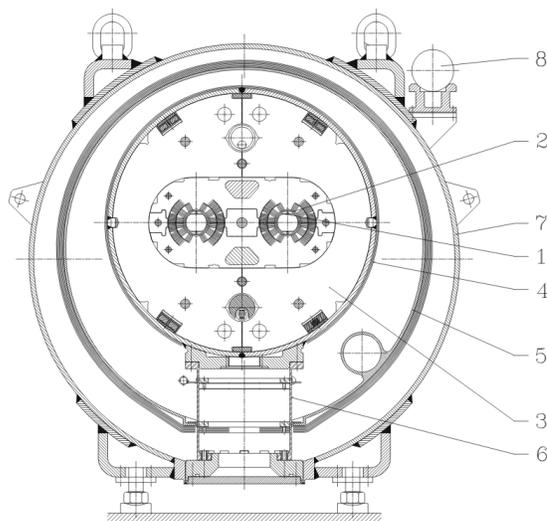
Figure 2.2: Schematic representation of a cross section of a dipole magnet at the LHC [38]. The diameter of the tube is about 1 m. 1: Beam screen, 2: Magnetic coil windings, 3: Cold mass at 1.9 K, 4: Radiative insulation, 5: Thermal shield (55 to 75 K), 6: Support post, 7: Vacuum vessel, 8: Alignment target.

To achieve a bunch crossing rate of 40 MHz, the ring gets filled with 2808 bunches of protons, which takes about 4 minutes. The LHC acceleration of the injected protons from 450 GeV to 7 TeV is controlled by 16 radiofrequency (RF) cavities along the ring. These cavities are metallic chambers that contain electromagnetic fields which oscillate at a frequency of 400 MHz. This final acceleration of the protons takes about 20 minutes, after which collisions can be recorded by the respective experiments at the design collision energy.

In this thesis, the collision data recorded during 2016 is used. During that year, the LHC operated at a maximum instantaneous luminosity of $1.4 \times 10^{34} \mathrm{cm}^{-2}\mathrm{s}^{-1}$ (40% above the design value), colliding protons at a center-of-mass energy of 13 TeV. The evolution of the total recorded integrated luminosity over the course of a year is shown in Figure 2.3 for 2011, 2012, 2015 and 2016. This graph shows 2016 was a very productive collision year and exceeded its projected luminosity, due to an increase in the number of proton bunches.

2.2   The Compact Muon Solenoid detector

The Compact Muon Solenoid (CMS) detector [40] is one of the four detectors along the LHC. Being a general purpose detector, CMS is designed in order to support a wide-range physics program, covering both SM precision measurements as well as searches for new physics phenomena. To achieve this, the detector consists as a whole of four main detector systems, each one designed for the detection of specific particle
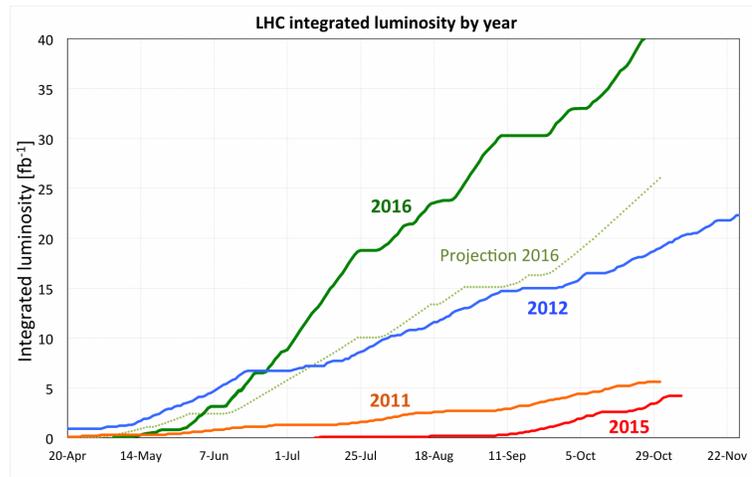
Figure 2.3: Integrated luminosity of the LHC in 2011, 2012, 2015 and 2016 and the predicted performance for 2016 [39].

signatures. The four detector systems are: the inner tracking system, electromagnetic and hadronic calorimeters and the muon detection system. These detectors are placed almost hermetically around the nominal interaction point and are designed with high granularity. As depicted in Figure 2.4, the detector systems are placed in a layered fashion around the beampipe, creating a cylindrical structure. The central part of the cylindrical structure is labelled as the *barrel*, which is closed of at the sides by two *endcaps*. The CMS detector spans a total length of 21.6 m, has a diameter of 14.6 m and weighs approximately 12 500 tonnes. The inner tracking system is placed as close as possible to the beampipe and consists of a high-resolution pixel detector enclosed by the silicon tracker system. It serves to detect charged particle hits in order to reconstruct their trajectories. The design and functionality of the inner tracking system is discussed in Section 2.2.1. The electromagnetic and hadronic calorimeters, which measure energy deposits, surround the inner tracker system and are summarised in Sections 2.2.2 and 2.2.3. The central feature of CMS is its superconducting solenoid, coiling cylindrically around the hadronic calorimeter in the barrel region. It produces a magnetic field of 3.8 T that bends charged particles such that their momentum can be determined. The magnetic field is confined by a steel return yoke in which the muon detection system is embedded (see Section 2.2.4). The electronic signals of each detector are processed to reconstruct the overall collision signature. As collisions happen at a rate of 40 MHz, not all collision events can be processed due to the limitations of cumputational resources. Therefore, only events with an interesting collision signature are triggered to be stored for later analysis. This trigger system is discussed in Section 2.2.5 together with the computing infrastructure that processes and distributes recorded collision data for analysis.

A right-handed coordinate system is defined at the CMS detector, with its origin
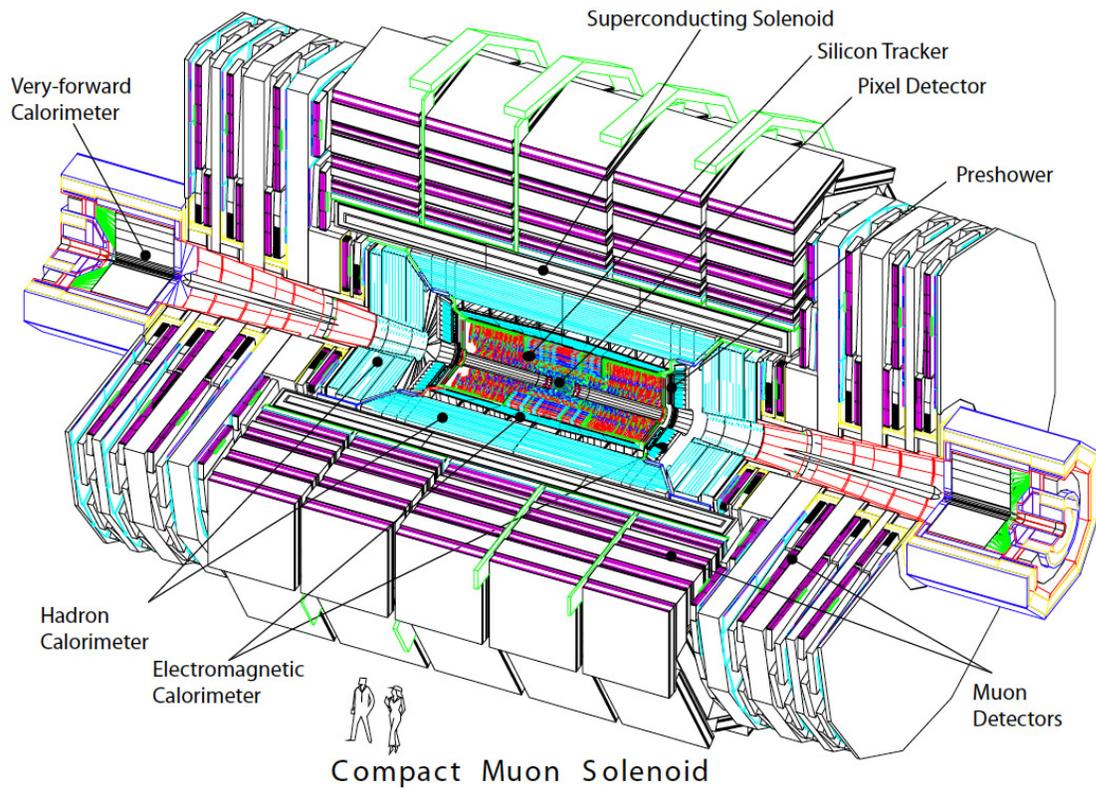
Figure 2.4: Perspective view of the CMS detector [41]. The coloured regions represent the different detectors. The blank areas in between the muon detectors represent the steel return yoke of the magnetic field.

established at the nominal interaction point. The x-axis points towards the center of the LHC ring, the y-axis points upwards and the z-axis is oriented along the anti-clockwise direction of the beam. The (x,y) plane is commonly referred to as the transverse plane. The transverse momentum p_T of a particle is accordingly defined as the momentum in the transverse plane. In spherical coordinates, the polar angle θ refers to the angle with respect to the z-axis and the azimuthal angle ϕ is defined in the (x,y) plane with respect to the x-axis. The *pseudorapidity* η is used instead of the polar angle and relates to it as

$$\eta = -ln\left(\tan\frac{\theta}{2}\right). \tag{2.3}$$

This choice is motivated by the fact that differences in pseudorapidity, unlike the polar angle, are invariant under Lorentz boosts. As the center-of-mass of each collision event has an unknown boost along the z-axis, variables that are invariant under Lorentz boost are more practical.

2.2.1   Detection and reconstruction of charged-particle tracks

A solid understanding of a collision event starts from the accurate reconstruction of its primary interaction vertex. This requires a high-resolution reconstruction of the paths of charged particles emerging from the primary interaction vertex, a feat that is mainly achieved by the high-granularity silicon pixel detector. The pixel detector is installed as close as possible around the beam pipe, making it subject to a very high flux of particles at about 1 MHz per mm². Consisting of 66 million pixel cells in total, each with a surface area of 100×150 μm², the pixel detector is able to cope with the high-radiation environment whilst disentangling separate hits from charged particles. Three layers of pixel cells are placed at radii of 4.4 cm, 7.3 cm and 10.2 cm and two more endcap layers at each side of the nominal interaction point, at a $|z|$ coordinate of 34.5 cm and 46.5 cm.

The flux of particles reduces to about 60 kHz per mm² at 20 cm distance from the beampipe. From this point on, a silicon strip detector is used for detecting charged particle crossings. The silicon strip detector is divided in four subsystems, each with a different configuration. Surrounding the pixel detector, the Tracker Inner Barrel (TIB) is placed, which consists of four layers of silicon strips at radii 25.5, 33.9, 41.9 and 49.8 cm and extend up to $|z|$ = 70 cm. At $|z|$ between 70 and 100 cm, the Tracker Inner Disks (TID) are located. The TID contain three disks of silicon strips, oriented perpendicular to the beam pipe, extending to a radius of 55 cm. Finally, in the barrel region, the Tracker Outer Barrel (TOB) configuration consists of six silicon strip layers, ranging between a radius of 55 cm to 116 cm. The two Tracker Endcaps (TEC) are composed of nine layers of silicon strips, placed in disks oriented perpendicular to the beam pipe. These disks range up to $|z|$ = 284 cm, rendering a total pseudorapidity coverage of the inner tracking system up to $|\eta|$ = 2.5. The total configuration of the inner tracking system is schematically represented in Figure 2.5.
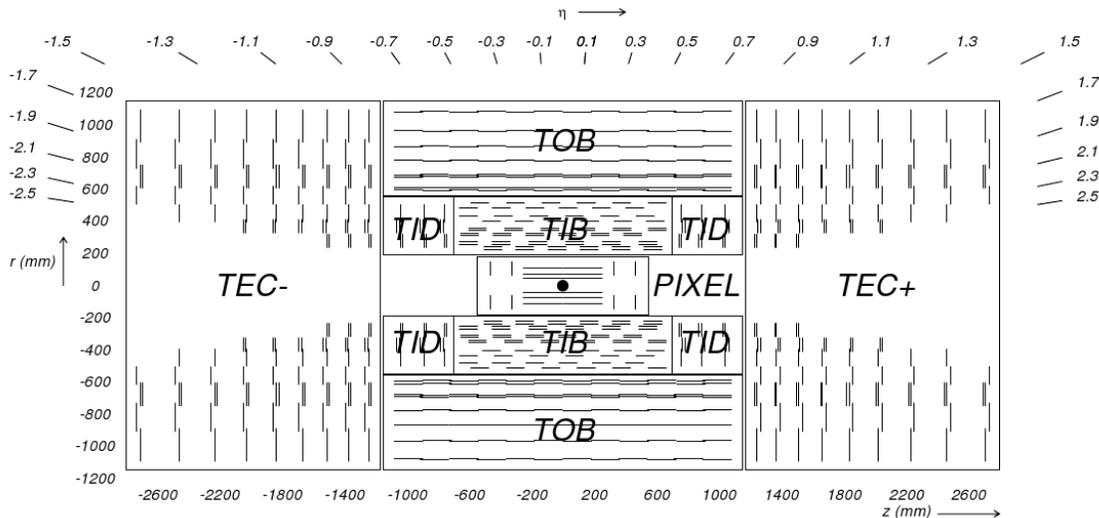
Figure 2.5: Schematic view of the inner tracking system configuration [42].

As charged particles traverse the detector, specific detector hits are recorded where the particle interacts with the silicon material. The path of the charged particle is reconstructed by combining the detector hits into a track. The tracks will have a helicoidal structure throughout the largest part of the detector volume, due to the magnetic field being constant. Based on this knowledge, an iterative track reconstruction procedure is developed. The iterative procedure is decomposed in four steps [43, 44]: seed generation, pattern recognition (or trajectory building), ambiguity resolution and final track fit.

1. **Seed generation.** As a starting point for the track reconstruction, a track seed is defined. In order to have a minimal description of the helix, at least three points in space are required. This is achieved by either grouping three detector hits or associating a pair of pixel hits to the nominal beam spot. From these hits, an initial estimation of the track parameters is made.

2. **Pattern recognition (or trajectory building).** The trajectory of the seed is extrapolated to the next detector layer, based on the estimated track parameters and their uncertainty. For every compatible detector hit, a new trajectory candidate is created, fitting the helix with a Kalman Filter technique [45]. Additionally, a trajectory candidate is added where no measured detector hit is found, to account for particles that didn't leave a hit in that detector layer. Such a hit is called an *invalid hit*. This procedure is repeated until the final detector layer is reached.

3. **Ambiguity resolution.** Ambiguities arise due to the fact that one seed can be associated to several trajectory candidates. To avoid double counting of tracks

associated to the passage of one charged particle, these ambiguities must be resolved. This is achieved by requiring that for each pair of trajectories, the fraction of shared detector hits is less than 50%. If a pair of tracks does not satisfy this condition, the track with the least number of hits is removed. If both tracks have an equal amount of hits, the track with the highest χ^2 of the track fit is discarded.

4. **Final track fit.** The final track is obtained by refitting it twice at the end of the process, using all associated detector hits at once. Firstly, the refitting is performed going from the beam spot outwards and after that, the track is refit again going from the most outer detector hit towards the beamspot.

Once a track has been reconstructed, the corresponding detector hits are removed and the procedure is repeated to find a next track. The iterative procedure allows a looser definition of the reconstruction criteria at each step. This ensures the minimisation of the reconstruction of fake tracks. The track reconstruction efficiency depends on the transverse momentum of the particle and its pseudorapidity [46]. Particles with p_T below 1 GeV have a very low reconstrucion efficiency, due to the fact that such particles may become trapped in the volume of the tracker as they spiral in the magnetic field. This produces many hits, which complicates the track fitting procedure. High-p_T tracks (typically above 100 GeV) are also more subject to misreconstruction, due to the fact that such particles are more likely to be produced in a more collimated beam of particles. Track reconstruction in such a high-density environment is more prone to misreconstruction. The track reconstruction efficiency dependence on pseudorapidity is a bit more ambiguous, dropping where the transition from TIB to TID happens and at very high $|\eta|$. At $|\eta| < 1$, the track reconstruction efficiency lies between 80% and 95%, whereas it varies between 75% and 90% for $1 < |\eta| < 2.1$. The track reconstruction efficiency of muons is significantly higher than for electrons and charged hadrons, varying between 95% and 99% for almost the full tracker acceptance. This is due to the fact that muons are little affected by bremsstrahlung and nuclear interactions. The relative transverse momentum resolution of the muon is outstanding, staying below 2% for muons with a $p_T \approx 100$ GeV and $|\eta| < 1.5$.

Primary vertex

From the collection of reconstructed tracks the primary interaction vertex can be reconstructed per collision event. The primary vertex reconstruction algorithm [44] starts by grouping tracks based on the longitudinal distance to the beam spot. Applying an adaptive vertex fitting algorithm, which will be discussed further on in Section 3.3.5, the position of the primary vertex is established. Tracks from long-lived hadrons may distort the primary vertex position. In order to reduce this distortion effect, tracks that are displaced with respect to the vertex get a smaller weight in the fitting process. The primary vertices are determined with a resolution of 20 μm in the transverse plane and 30 μm in the longitudinal direction.

Due to pileup interactions, several primary vertices are reconstructed per bunch crossing. The main primary vertex is identified by selecting the primary vertex with the largest value of summed physics-object[7] p_T^2. In the collision data recorded at CMS in 2016, on average 23 collisions per bunch crossing happened, with extrema of up to 50 interactions per bunch crossing.

2.2.2 Electromagnetic calorimeter

Scintillation detectors are excellent calorimeters for measuring the energy of electrons and photons. At CMS, the electromagnetic calorimeter (ECAL) consists of 75,848 scintillating crystals made up of lead tungstate ($PbWO_4$). The crystals are connected to photon multipliers, which transform the scintillation light into an electronic signal. The choice for lead tungstate is motivated by its excellent response time (80% of the scintillation light is emitted in 25 ns) and its short radiation length of about 0.85 cm. The response time of 25 ns is identical to the time between two bunch crossings, making it capable of dealing with the high-luminosity environment of the LHC. The small radiation length implies that the electromagnetic shower will be contained in a relatively small volume, which allows CMS to obtain a high granularity in the ECAL.

The layout of the ECAL is schematically presented in Figure 2.6. Three substructures make up the total ECAL: the Barrel ECAL (EB), the endcap ECAL (EE) and the preshower (ES). The crystals in the EB are oriented towards the nominal interaction point, are 23 cm long and having a surface area of 22×22 mm^2. The pseudorapidity covered by the EB is up to $|\eta| = 1.479$. From $|\eta| = 1.479$ to $|\eta| = 3.0$ the EE crystals are placed. These crystals have a length of 22 cm and a surface area of 28.6×28.6 mm^2. In front of the EE, the ES is placed covering a pseudorapidity going from $|\eta| = 1.653$ to $|\eta| = 2.6$. The preshower consists of two layers of lead absorbers and silicon strip sensors. Its purpose is to distinguish promptly produced photons from colinear diphotons produced in the decay of neutral pions.

The ECAL performs outstandingly, achieving an energy resolution for electrons from Z-boson decays better than 2% for $|\eta| < 0.8$ and varying between 2-5% elsewhere.

2.2.3 Hadronic calorimeter

Hadrons are hardly affected by the scintillator detectors of the ECAL and will only lose a small portion of their energy in the ECAL. Therefore, CMS employs a sampling calorimeter as hadronic calorimeter (HCAL) surrounding the ECAL. The sampling is procured by alternating layers of brass absorbers and plastic scintillator tiles. By interacting with the dense absorbers, the hadrons cascade and produce electromagnetic showers, which are recorded by the scintillator tiles.

---

[7]The physics objects, defined using charged-particle tracks only, are the jets, remaining unclustered tracks, and missing transverse momentum associated with the vertex.
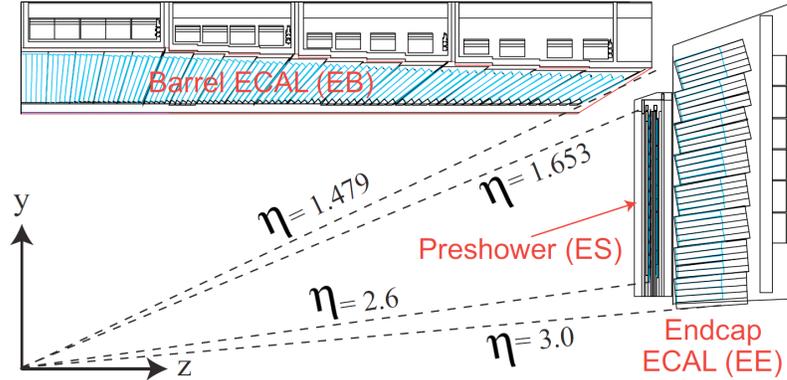
Figure 2.6: A pictorial representation of a quarter view of the ECAL detector in the (y,z) plane [40].

In the barrel region, the HCAL is divided into two parts: the barrel part (HB) and the outer barrel part (HO). The HB is placed inside the volume enclosed by the solenoid, consisting of 17 layers of scintillator tiles interspersed with brass absorber layers. The individual tiles in the HB cover an area of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The HB is placed at radii between 177.7 cm and 287.6 cm, filling up the remaining space between the ECAL and the solenoid. Due to the limited space available inside the solenoid volume, an extension of the HB is installed outside the solenoid, being labelled as HO. This extension only exists of scintillator detectors, in order to detect remnants of the hadronic cascade shower outside the solenoid. In total, the HB and HO cover up to $|\eta| = 1.4$. The endcap part of the HCAL (HE) covers the region $1.3 < |\eta| < 3.0$, creating a small overlap in pseudorapidity coverage with the barrel region. The scintillator tiles in the HE cover an area of $\Delta\eta \times \Delta\phi = 0.350 \times 0.174$, composing in total 19 scintillator layers interspersed with brass absorbers. The pseudorapidity coverage of the HCAL is further extended by the forward calorimeters (HF). These calorimeters are placed 11.5 m away from the nominal interaction point, covering $3.0 < |\eta| < 5.0$. As these forward regions are much more subject to particle radiation, different materials are used; namely steel absorbers with quartz fibers as scintillator detectors. The layout of the HCAL is schematically represented in Figure 2.7.

The energy resolution for hadrons in the barrel region, measured by a combination of HCAL and ECAL energy deposits, goes as

$$\frac{\sigma[E(GeV)]}{E(GeV)} = \left(\frac{A[\sqrt{E(GeV)}]}{\sqrt{E(GeV)}} \right)^2 + B^2, \tag{2.4}$$

where A represents a stochastic term and B a constant noise factor. During a test beam phase, these terms were determined as A = 0.847 $\pm$ 0.016 and B = 0.074 $\pm$ 0.008. For a hadron with an energy of 100 GeV, the energy resolution corresponds to about 15%.
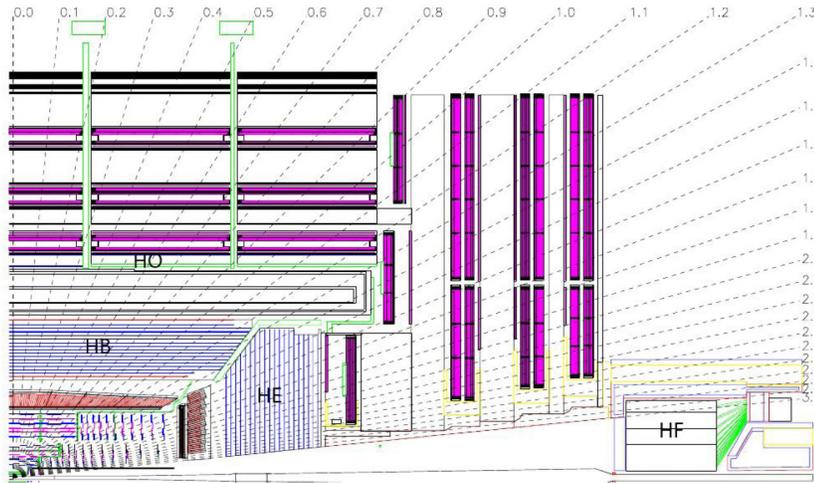
Figure 2.7: A quarter view on the CMS detector in the (y,z) plane, highlighting the HCAL parts: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF) [47].

2.2.4   Muon detection system

Muons are hardly affected by any type of material they traverse. Therefore, like neutrinos, they will traverse the CMS detector without being absorbed, leaving almost nothing but a trajectory in the tracker. In order to identify muons originating from a collision, the CMS detector has a special muon detection system placed outside of the solenoid, embedded in the steel return yoke of the magnet.

The CMS muon detection system consists of several layers of Drift Tubes (DT) in the barrel and Cathode Strip Chambers (CSC) in the endcap. These two types of muon detectors are gaseous detectors recording the passage of charged particles with a high spatial resolution, being 100 μm for the DT and 200 μm for the CSC. The angular resolution in the ϕ-direction is 1 mrad for DT and 10 mrad for CSC. The DTs cover the region $|\eta| < 1.2$ and the CSCs cover $|\eta| \in [1, 2.4]$. The real succes of the CMS muon detection system lies in the use of Resistive Plate Chambers (RPC), which have a timing resolution of 1 ns. Due to the excellent time resolution, the RPC are able to associate detected muons to individual bunch crossings. RPCs are placed in both barrel and endcaps, always in combination with a DT or CSC, as illustrated in Figure 2.8.

2.2.5   Processing and filtering collision events

At the intended instantaneous luminosity of 10^{34}cm^{-2}s^{-1}, CMS has to deal with a collision rate of the order of 10^9 Hz, coming from bunches crossing over at a rate of 40 MHz. Most of these collisions are not interesting for further physics analysis, because they are low energy hadronic processes and elastic collisions. As there is not enough
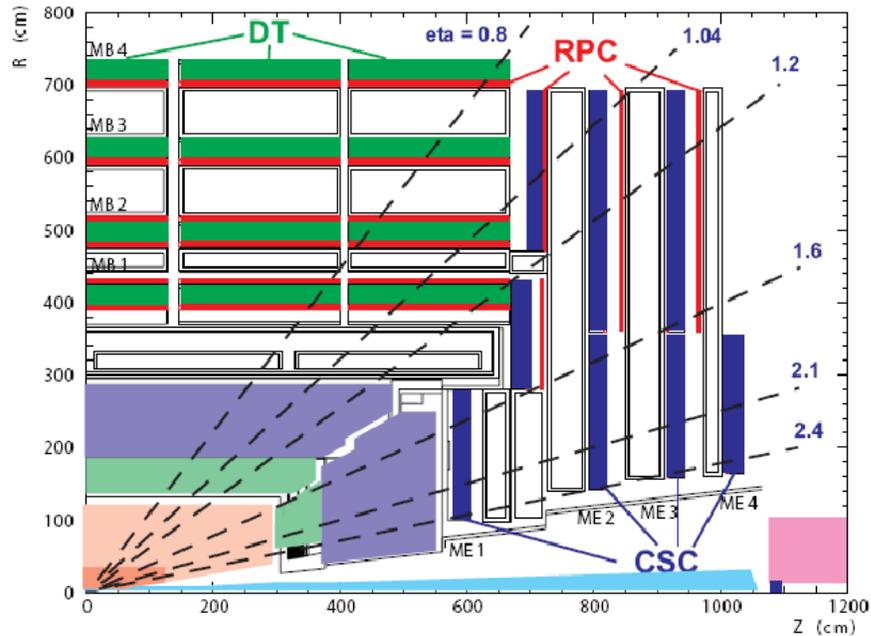
Figure 2.8: Quarter view of the CMS muon detection system. The non-coloured zones in the outer layers represent the return yoke of the magnet [48].

data storage room for all events, online triggers are installed to reject uninteresting events before they are stored for physics analysis. CMS relies on two triggers to do this: the Level-1 (L1) trigger, a hardware trigger located near the detector, and a High Level Trigger (HLT), using computer farms at the surface to run the software.

1. **Level-1 trigger.** The electronic signals from the detector are converted into digital information using digitizers. This raw data is sent to front-end pipelines, storing the information for 3.2 μs while L1 decides, based on fast algorithms, whether to further process the event or not. These algorithms need to be very fast and therefore only use information from the muon detection system and the calorimeters. The event rate after the L1 is reduced to the order of 100 kHz.

2. **High Level Trigger.** After the L1 has passed an event, the information of the event is read out from a buffer. Switching networks then combine all the information from the parts of the detector to reconstruct the event. The event information now contains reconstructed tracks and more detailed information from the calorimeters. The HLT uses this to make a more refined selection in the data in order to reduce the event rate to 1 kHz.

The whole trigger process is schematically represented in Figure 2.9.

Triggered events from actual recorded collision data are stored offline at two grand computing infrastructures in CERN and Budapest. This is the first level of data processing (and storage) of the extensive computing infrastructure of all LHC experiments,
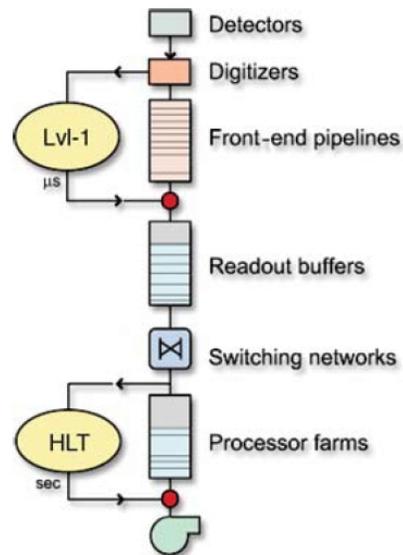
Figure 2.9: Schematic representation of the CMS trigger system [49].

also referred to as the Worldwide Large Hadron Collider Grid (WLCG) [50]. Combining the computing resources of many of the institutes participating to LHC experiments around the world, the WLCG provides a stable and reliable access to both simulated and actual collision data. Labelling the first level as *Tier-0*, two more levels are introduced as *Tier-1* and *Tier-2*. All LHC experiments combined have thirteen Tier-1 sites distributed around the globe, which are used for reprocessing of data and storage of both real collision data and simulated samples. These samples of real and simulated data are distributed further downstream to about 160 Tier-2 centers. From these Tier-2 centers, physicists around the world can access and analyse the data. A pictorial representation of the hierarchal WLCG structure is given in Figure 2.10. The analysis tools necessary for analysing, processing and simulating data are developed and maintained by physicists around the world, centralized in the common CMS Software framework (CMSSW) [51].
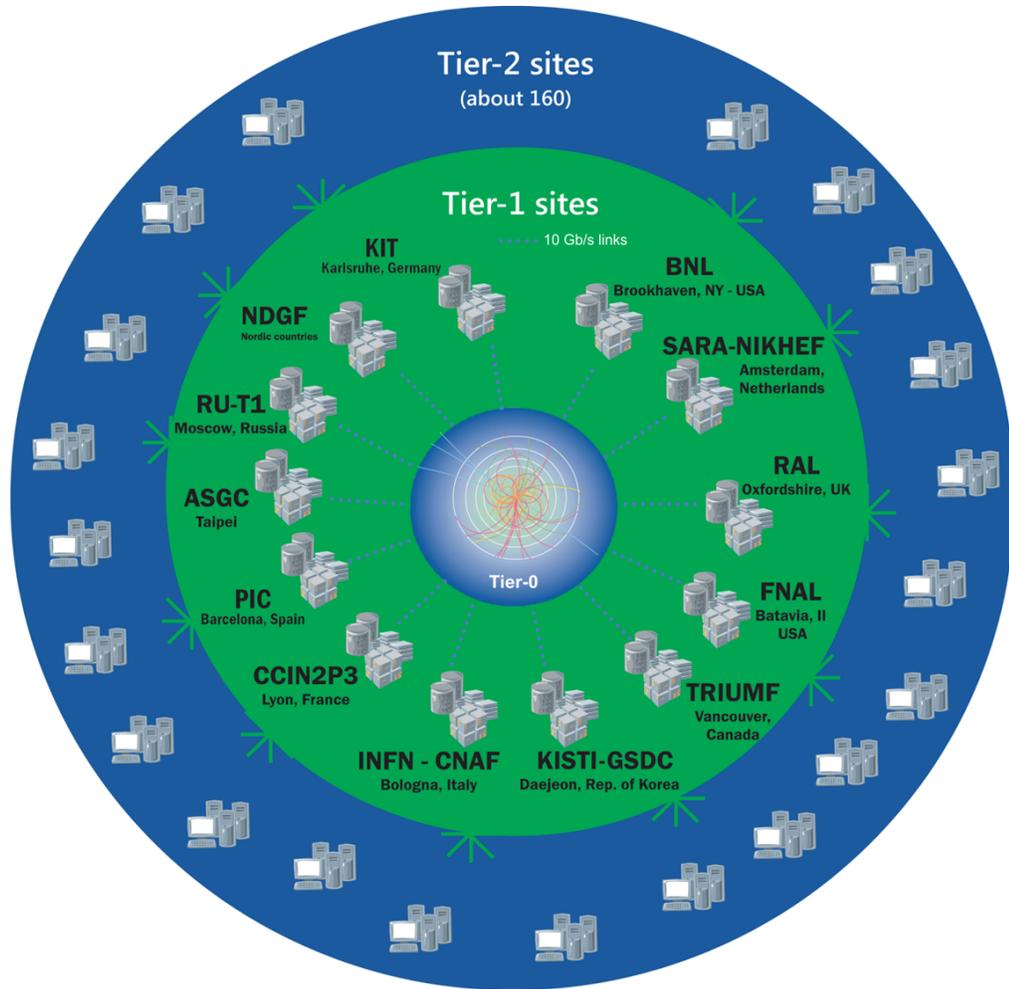
Figure 2.10: Pictorial representation of the WLCG tiered structure [52].

Chapter 3

Simulation and reconstruction of proton-proton collisions

As CMS records data from proton-proton collisions (pp-collisions), a comprehensive description of the fundamental aspects of these collisions is required. With this description, physics models are developed for simulating the hard collisions, including known SM processes as well as processes beyond the SM (e.g. Supersymmetry, Dark Matter). These generated physics processes cause specific detector signals to arise, which are in turn simulated based on the known geometry of the CMS detector and its materials. The raw detector signals are then translated into physics objects, such as electrons and jets, using reconstruction techniques designed for the CMS experiment. These techniques are developed to be identical for simulated and real collision events. The entire simulation process ensures recorded data can be compared to simulated events and conclusions can be made on specific phenomena.

The model for simulating pp-collisions is introduced in Section 3.1. The specific software used in the CMS collaboration to simulate the physics processes of interest to this thesis, is described in Section 3.2. Section 3.3 summarises the reconstruction of physics objects resulting from a collision, as done at the CMS experiment.

3.1 Fundaments of a proton-proton collision

Simulations of pp-collisions in the CMS experiment build on the generation of the hard process followed by showering and hadronisation models [53]. The typical event chain of these generators is schematically represented in Figure 3.1 in chronological order going from bottom to top.

The evolution of the event chain can be summarised as:

- A constituent (or so-called parton) from each proton "emerges" with a certain probability density $f(x, Q^2)$, determined by the momentum fraction x carried by the constituent in the proton and the momentum transfer Q^2. The parton
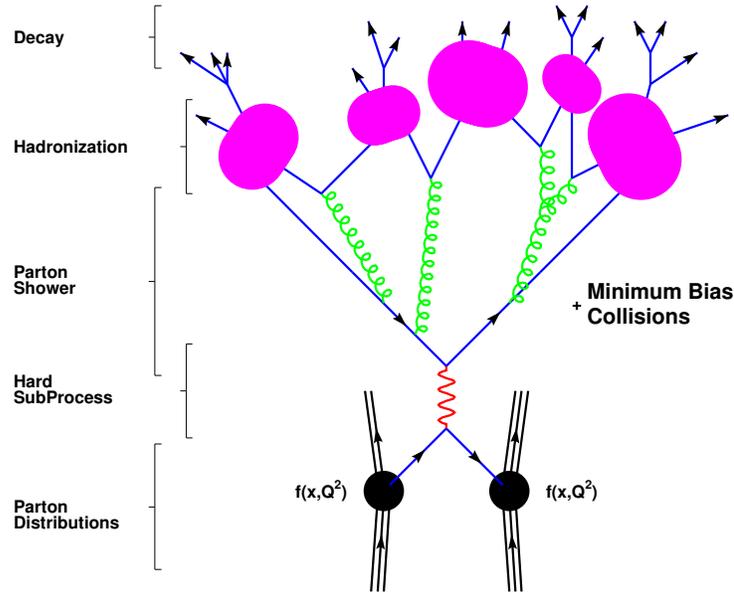
Figure 3.1: Schematic representation of an event chain originating from a pp-collision, starting with the resolution of partons from the proton (bottom) and ending with the decay (top) [53].

model of protons resolves these density functions, fitted on data, as explained in Section 3.1.1.

- The interaction of the two emerging partons is labelled as the hard (sub)process (see Section 3.1.1). The hard process encapsulates both the interaction between the emerging partons as well as the decay of any resonances, such as $W^{\pm}$ bosons decaying to leptons, arising from the interaction.

- Partons may branch into other partons and so on, creating a shower of partons. The process of parton showering is summarised in Section 3.1.2.

- Due to colour confinement, partons are prohibited to exist on their own. This forces partons to fragment and group into hadrons. This step is called hadronisation and is dictated by the non-perturbative regime, using phenomenological models as introduced in Section 3.1.3. The resulting hadrons are then decayed further.

- The remnants of the two protons, having lost their colour neutrality due to the emerging partons, will fragment and hadronise. These fragmentations causes the underlying event to arise (see Section 3.1.4).

3.1.1   Parton model and hard scattering

Quarks that determine the electric charge and baryon number of a hadron or meson are called valence quarks. In the case of the proton, the valence quarks are two up and one down quark (uud). Valence quarks are usually embedded in a broiling sea of gluons and (anti)quarks. This sea of partons is an everchanging sea, where partons continuously emit and radiate gluons and gluons split into quark-antiquark pairs. The flavour of the (anti)quarks in the sea depends on the scale Q at which the proton is probed, allowing only flavours for which the mass $m_q \ll Q$. This model is referred to as the parton model [54, 55]. When protons collide, the actual collision process, labelled as the hard process, is described by the interaction between two partons resolving from each proton. The probability at which a parton i, carrying a certain fraction x of the proton's total momentum, resolves from the proton is given by parton density functions (PDFs) $f_i(x, Q^2)$. The PDFs of hadrons are determined by global fits to data from deep inelastic scattering, Drell-Yan and jet processes. Collaborations such as CTEQ [56], MSTW [57] and NNPDF [58] perform such global fits for PDFs of protons and update them regularly with new data or adapted theoretical models. Examples of PDFs scaled with x for protons as provided by MSTW are given in Figure 3.2 at scales $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2, with their associated uncertainty bands at 68% confidence level. The relative uncertainties on PDFs become smaller for lower fractions x when probing the proton at higher energies, as one is able to resolve deeper in the proton. Gluons are dominantly dense in protons, especially at low x, which leads to a much higher interaction rate between gluons than quarks[1]. Furthermore, there's an asymmetry in the density of up and down quarks versus their anti-particles, which is not there for heavier quarks.

Parton density functions are measured at one specific scale and can be extrapolated by perturbative QCD to other scales, using the so called DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations [60–62].

Based on the parton model of protons, the cross section of pp-collisions producing a final state X, denoted as $\sigma_{pp \to X}$, can be formulated as the convolution of the *partonic* cross section $\hat{\sigma}_{ab \to X}$, representing the hard interaction, and the PDFs of the involved partons [63]

$$\sigma_{pp \to X} = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_a(x_a, Q^2) f_b(x_b, Q^2) \hat{\sigma}_{ab \to X}(x_a, x_b, \alpha_s(Q^2)), \qquad (3.1)$$

where the sum runs over all possible partons (taking into account spin, flavour, ...), integrating out the momentum fractions x_a and x_b of the partons. The differential partonic cross section is given by

$$d\hat{\sigma}_{ab \to X} = \frac{|\mathcal{M}_{ab \to X}|^2}{64\pi^2 s_{real}} dcos\theta d\phi, \qquad (3.2)$$

---

[1]This feature of proton PDFs is also one of the reasons why the LHC is sometimes quoted as a top quark factory in the context of top quark physics, as the main production mode of top quark pairs is via the interaction of two gluons.
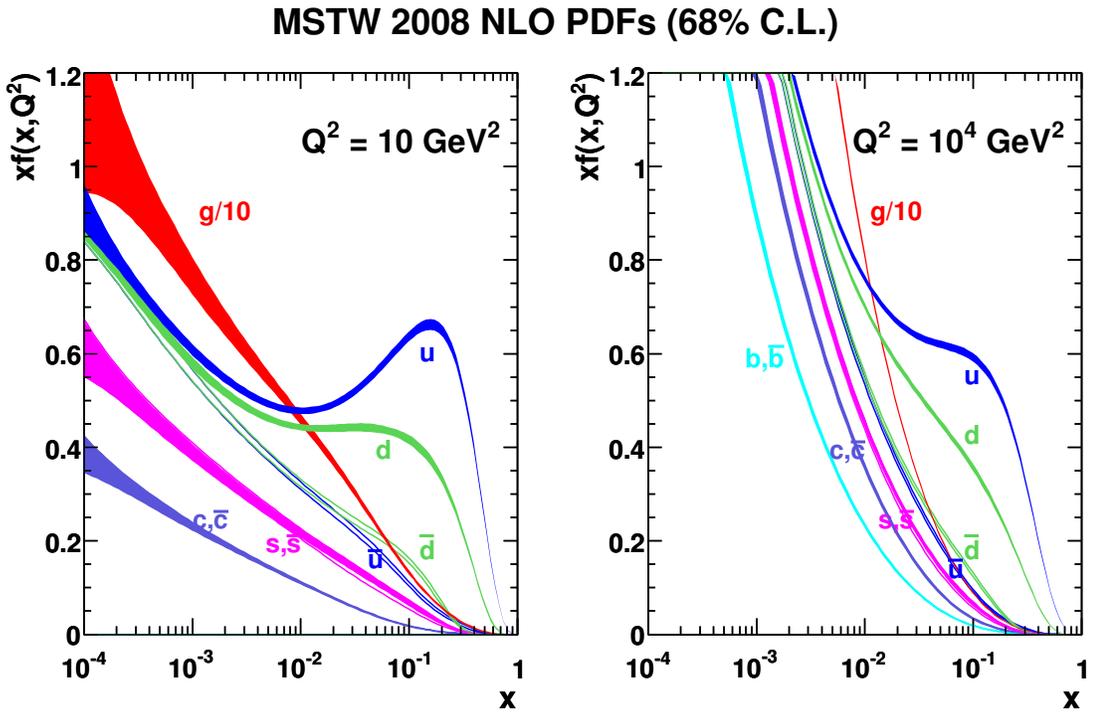
Figure 3.2: Parton density functions of partons in a proton at scales $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right) from the MSTW 2008 [59]. The gluon PDFs are downscaled by a factor 10 to make them visible.

with s_{real} the real center-of-mass energy, θ the polar angle, ϕ the azimuthal angle, and $\mathcal{M}$ the matrix element of the process, which can be derived from the Lagrangian density of the quantum field theory under consideration.

As the matrix element depends on the strong coupling constant α_s, it may be calculated to a certain perturbative order, given α_s is small enough (or equivalently, the interaction energy scale is high enough).

The renormalisation scale μ_R (see Section 1.1.3), at which α_s is evaluated, needs to be chosen such that it deals with ultraviolet divergences. Another energy scale, the factorisation scale μ_F, is introduced to regulate infrared and collinear divergences. Infrared divergences are caused by the emission of soft (low-energy) gluons, whereas collinear divergences arise from gluons that are emitted collinear to the parton. These two divergences appear in the calculation of the matrix element. However, such divergences can be factorized from perturbativity by introducing the factorisation scale. The renormalisation scale is limited to the calculation of the matrix element only and appears solely in the partonic cross section. Note that if the total hadronic cross section (3.1) is calculated to all orders in perturbation theory, it should be independent of the factorisation and renormalisation scales. In reality however, these calculations need to be restricted to a low order of perturbation such that renormalisation and factorisation scales need to be chosen.

3.1.2   Parton showering

Partons can radiate soft gluons or split into two collinear partons, resulting in a shower of partons. Depending on whether this happens to partons going into the hard interaction or coming out of it, the process is labelled as respectively Initial State Radiation (ISR) or Final State Radiation (FSR). Parton showering techniques are approximate perturbative treatments of QCD to all orders, describing the shower evolution at short distances until the perturbative description breaks down at a lower scale Λ_{QCD} (where α_s approaches unity). The approximate perturbative treatment is found in the DGLAP formalism. Considering a parton may branch into two daughter partons, the DGLAP formalism describes its branching probability at an energy scale Q^2. The three branching possibilities of a parton (quark q or gluon g) into daughter partons are $q \to qg$, $g \to gg$ and $g \to q\bar{q}$. Possible infrared and collinear divergences may result in branching probabilities above unity. In order to avoid such nonsensical parton branching probabilities, infrared and collinear divergences are cancelled by introducing Sudakov form factors[2] into the DGLAP formalism. The branching of partons is progressed until Λ_{QCD} is reached, which is usually chosen around 1 GeV. Once Λ_{QCD} is reached,

---

[2]Sudakov form factors take virtual (quantum loop) effects that are of the same order of real parton emissions into account, wich are neglected in the DGLAP evolution, as probabilities of not splitting a parton during the scale evolution.

hadronisation models are used to further describe the non-perturbative regime, except for top quarks which decay before they can hadronise. In ISR, the parton showering process is reversed until the energy scale of the parton, emerging from the proton, is reached.

3.1.3   Hadronisation

As the parton splitting energy goes down to Λ_{QCD}, perturbative QCD breaks down and cannot be used for a further description of the parton evolution. At this point, due to colour confinement, the partons start to form colour neutral hadrons. The hadronisation of partons is theoretically not yet fully understood, imposing the use of phenomenological models. Currently there are mainly two phenomenological approaches, one based on string fragmentation and the other based on the cluster model.

A main proponent of string fragmentation is the Lund string model [64], which is based on the idea of a string connecting a quark q and antiquark $\bar{q}$. As the two quarks move apart, the string gets stretched and potential energy builds up. The potential energy is assumed to increase linearly with distance r as κr, with $\kappa \approx 1$ GeV/fm. When the potential energy becomes large enough to produce a new pair of quarks $q'\bar{q}'$ with a mass m, the string can break with a probability proportional to

$$\exp(\frac{-\pi(m^2 + p_T^2)}{\kappa}),\tag{3.3}$$

with p_T the transverse momentum of the quarks in the pair. The quark pair $q\bar{q}$ is now split into two new pairs $q\bar{q}'$ and $q'\bar{q}$, lowering the available energy for a next splitting. The splitting process continues until only colour-neutral hadrons with an on-shell mass remain. Heavy quarks (charm, bottom and top) are ignored in this model as the probability to create a heavy quark pair is heavily suppressed[3]. Gluons are pictured in the string model as kinks on the string between two quarks, assigning the sum of the color and anticolor of the involved quarks to the gluon. As such, the Lund string model is infrared and collinear safe [65], because soft and collinear gluon emissions are represented as vanishingly small kinks on the string.

3.1.4   Underlying event

Though the hard interaction is the process of interest when studying pp-collisions, additional activity takes place in these collision events, labelled as the *underlying event* (UE) [66]. As the parton in the hard interaction is resolved from the proton, the remainder of the proton is no longer colour neutral. Colour confinement forces

---

[3]The probability ratios, supposing the same transverse momentum, for the production of u:d:s:c $\approx$ 1:1:0.3:10^{-11}.

the remainder of the proton, commonly referred to as the beam remnant, to hadronise as well, resulting in a higher charged-particle multiplicity. The description of the evolution of beam remnants into hadrons is more ambiguous than the process described in Section 3.1.3. This is due to the unknown distribution of the energy in the beam remnant and its colour connection to the parton participating in the hard interaction.

Additional to beam remnants, there is also the possibility of multiple parton ($2 \rightarrow 2$) interactions contributing to the UE. The probability of an extra hard interaction with high transverse momentum to arise from extra parton interactions is very small. Typically the activity coming from multiple parton interactions is much less energetic than the hard interaction, producing mainly low energetic hadrons. The differential cross section of multiple parton interactions diverges as p_T^{-4}, introducing the need to implement a phenomenological cutoff p_{T_0} such that

$$p_T^{-4} \rightarrow (p_T + p_{T_0})^{-4}. \tag{3.4}$$

This transformation ensures a good description of the perturbative result for large p_T, leaving a finite result as p_T approaches zero. The cutoff p_{T_0} depends on the center-of-mass collision energy $\sqrt{s}$ and is regulated as [67]

$$p_{T_0}(\sqrt{s}) = p_{T0}^{ref} \frac{\sqrt{s}}{\sqrt{s_0}}, \tag{3.5}$$

where p_{T0}^{ref} is the reference cutoff at $\sqrt{s_0}$. Consequently, the simulation of the UE is tuned by, among others, the choice of p_{T0}^{ref}, $\sqrt{s_0}$, the PDF and α_s.

3.2 Event and detector simulation

To simulate a specific physics process, such as the production of a top quark pair, several methods exist which emulate the various aspects of a pp-collision as described in Section 3.1. In the CMS collaboration, three types of software tools are interfaced with one another, each one representing a specific part of the collision process:

- Matrix element generators, e.g. `MadGraph/MadEvent` [68, 69] and `POWHEG` [70–73], simulate the hard process. Additional partons may be added to consider ISR and FSR.

- `PYTHIA` [74, 75] manages parton showering and hadronisation as described in Sections 3.1.2 and 3.1.3.

- `GEANT4` [76] simulates the detector response.

The UE of the simulated physics processes used in this thesis, is controlled in `PYTHIA` by the CUETP8M2T4 tune [77] for $t\bar{t}$ events and single top events in the t channel. All other proccesses are tuned with the CUETP8M1 tune [66].

3.2.1   Event simulation

The `MadGraph/MadEvent` matrix element generator

`MadGraph` generates matrix elements based on tree-level Feynman diagrams and calculates the corresponding LO cross section. Interfaced with `MadGraph`, the hard process events are generated by `MadEvent` based on the generated matrix elements. Extra partons may be added to the hard process as extra legs in the Feynman diagram to get an approximate description of ISR and FSR. This LO matrix element generator is very widely used for the simulation of both known SM backgrounds as well as beyond the SM physics processes. In this thesis, `MadGraph/MadEvent` is used for the simulation of a couple of background processes (e.g. $t\bar{t} + V$, W+jets) and the FCNC processes. For the FCNC signal processes specifically, up to two additional partons are added in the initial hard process at LO for the FCNC top quark pair production mode. No additional partons are included in generation of events for the single top production process. The Lagrangian terms from (1.23) are implemented in `MadGraph` by means of the FeynRules package [78] and the Universal FeynRules Output (UFO) format [79]. These are convoluted with the leading-order CTEQ6 set of parton density functions [80].

The merging of `MadGraph/MadEvent` generated parton configurations with `PYTHIA` parton showers is regulated by merging schemes. Several such schemes exist, but for `MadGraph/MadEvent` generated processes in this thesis, the MLM merging scheme [81] is applied. The MLM merging goes as follows. When N partons are considered in the hard processes, $n \leq N$ partons are generated in separate samples, constrained by phase-space cuts $p_T > p_T^{min}$, $|\eta| < |\eta^{max}|$. An extra angular separation between two separate partons i and j is required in the (η, ϕ) plane as $\Delta R_{i,j} = \sqrt{(\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2} > \Delta R^{min}$. The n partons are showered with `PYTHIA` which are then clustered based on a clustering algorithm (see Section 3.3.4). The clusters are required to have a minimum transverse energy, defined by the matrix-element parton-shower threshold. The generated partons from the n parton collection are matched to a cluster if $\Delta R_{part,clus} < \Delta R^{match}$, which is a fixed parameter. If a match is found for the hardest parton, the cluster is removed from the list of clusters and a match is sought for the second hardest parton and so on. For $n < N$ parton configurations, events are rejected when partons are not matched to a cluster or when there are more clusters than partons. For the $n = N$ parton configuration, events are only rejected if there are unmatched partons and/or there are unmatched clusters which are harder than the softest parton. In this thesis, the generation of a single boson in association with extra jets is one example of a physics process generated with `MadGraph/MadEvent`.

NLO matrix elements can be calculated as well in `MADGRAPH5_aMC@NLO` [68]. The merging with `PYTHIA` parton showers is more ambiguous than the MLM merging scheme, due to the fact that the matrix element includes up to $N + 1$ partons. The

final $N + 1$ parton state can be produced from either an N parton matrix element, where an extra parton showering was generated in the showering process, or from an $N + 1$ parton matrix element. The MC@NLO [82] program calculates how an N parton matrix element showering populates the $N + 1$ parton phase space. This calculation is subtracted from the $N + 1$ parton matrix element showering, representing the exclusive $N + 1$ parton final state. Having defined the exclusive $N + 1$ parton final state as such and the N parton final state as the subtracted part, double counting is avoided by applying showering separately to these two cases. A couple of less dominant background processes are generated with `MADGRAPH5_aMC@NLO`, such as $t\bar{t}$ in association with a gluon and double Z boson production.

The `POWHEG` matrix element generator

`POWHEG` simulates the hard collision event based on NLO calculations for a select set of physics processes. Due to the NLO precision of the hard interaction, up to one additional parton is generated in the hard process. The matching between `POWHEG` and `PYTHIA` is based on the ordering of the parton emissions in p_T, matching the highest p_T showers to the hard interaction. In this thesis `POWHEG` is used to simulate top quark pair production, $t\bar{t}+H$, single top production in t-channel and single top production in association with a W boson.

3.2.2   Detector simulation

The interaction of particles emerging from the collision with the detector is simulated by the `GEANT4` software [76]. The `GEANT4` toolkit is used to develop a very detailed representation of the full CMS detector, including active detector layers specifically designed to interact with the particles from the collisions and dead zones that are created by, among others, support structures, the magnet and cables. Based on a solid understanding of the underlying interaction mechanisms, the energy losses of the traversing particles is simulated as well as their trajectories in the detector. The description of the trajectory of charged particles through the detector relies on the excellent description of the magnetic field. Next to the interaction of the traversing particles with the detector material, the detector response and its conversion into electric signals is also simulated by the `GEANT4` toolkit. Due to the multitude of different detector materials, millions of readout channels and the complexity of the magnetic field, the detector response simulation is very CPU intensive and can take (depending on the process) several minutes for one event.

Pileup interactions are simulated by generating extra (soft) pp-interactions and processing the corresponding detector hits with `GEANT4`. The number of simulated pileup interactions is chosen such that it approximates the distribution of number of expected additional soft pp-collisions. As the number of simulated pileup interactions follows

an estimated distribution, this number will be reweighted once the true distribution of number of pileup interactions per collision event is measured in actual data.

3.3   Reconstruction of physics objects

Both measured data and simulated events result in a collection of electronic signals. As the electronic signals are read out, they are translated back to their corresponding detector hits and from these hits the event is reconstructed by associating physics objects to specific detector signatures. At CMS, particle-flow (PF) reconstruction [83, 84] is applied to identify objects in a collision event.

The elements and workings of the PF reconstruction will be discussed in Section 3.3.1. The PF reconstruction of muons and electrons is discussed in Sections 3.3.2 and 3.3.3 respectively. Particles (hadrons and charged leptons) that end up in close proximity of each other in the detector, are clustered into a jet as explained in Section 3.3.4 aiming to reconstruct hadronised partons. An important feature of this thesis is the identification of jets originating from b quark hadronisation, which is summarised in Section 3.3.5. The evaluation of missing transverse energy is described in Section 3.3.6. Next to the identification of particles emerging from the hard collision, the PF reconstruction provides the possibility to further scrutinize the event for particle contributions not arising from the hard collision or UE (see Section 3.3.7).

3.3.1   Particle-flow reconstruction

The PF reconstruction combines information from all CMS detector systems to reconstruct and identify specific detector signatures with known particles. This results in a more precise determination of the particles' momenta than using detector information separately. The elements that serve as input for the particle-flow reconstruction are reconstructed tracks (cfr. Section 2.2.1), muon tracks (cfr. Section 2.2.4) and calorimeter clusters [4]. PF elements are linked into a block, which then gets identified as a muon, electron, photon, charged- or neutral hadron. For example, electrons can be identified by linking an extrapolated track to a calorimeter cluster in the ECAL. Due to the high granularity of the CMS detector, the quality of the link can be determined with high precision in the (η,ϕ) plane. The commissioning of the PF reconstruction [85] confirms its excellent performance. Once a block is found, the corresponding elements are removed from the event and a next link is sought for, and so on until no more elements are left.

The PF reconstruction algorithm starts by extrapolating all charged-particle tracks throughout the detector. At first it looks for corresponding PF elements in the muon

---

[4]Calorimeter clusters are formed in each calorimeter detector separately, starting from a calorimeter-cell seed. Energy deposits above a certain threshold, defined to suppress electronic noise, from adjacent cells are added to the cluster until no more such cells are found. The calorimeter-cell seeds are selected as cells with an energy deposit above a certain threshold.

tracker and the ECAL, aiming to reconstruct muons and electrons first. Once all muons and electrons in the event are reconstructed, the identification and reconstruction of charged hadrons goes as follows. The remaining links between charged-particle tracks and calorimeter clusters (ECAL or HCAL) are evaluated based on their respective momenta. When the momentum of the charged-particle track and the calorimeter cluster are compatible (within uncertainties), the link is established as a charged hadron. If the momentum of the charged-particle track is much lower than the cluster momentum, the particle gets identified as a charged hadron with an additional neutral hadron or photon energy deposit, depending on whether the calorimeter excess is measured in the HCAL or ECAL respectively. A charged hadron can not be identified with charged-particle tracks that have momenta much higher than the calorimeter cluster energy deposit. In this case the algorithm tries to assign an additional muon with loose selection criteria to it. Once all charged hadrons in the event are found, the remaining calorimeter clusters in the HCAL and ECAL get reconstructed as neutral hadrons and photons respectively.

3.3.2   Muon object reconstruction

Muon identification is based on selection criteria applied to so-called *global* and *tracker* muon qualities.

- A global muon starts from the reconstruction of hits in the muon detector systems, i.e. drift-tube (DT), cathode-strip-chamber (CSC) and resistive-plate-chamber (RPC) hits (cfr. Section 2.2.4). DT and CSC hits are matched to form track segments. Once all tracks segments are collected, a track fit of the segments is performed, reconstructing a so-called *standalone muon track*. A match of the standalone muon track to a track in the CMS tracker system is found by propagating both to a common surface. The global muon is then reconstructed by fitting the combined hits from the track in the CMS tracker system and the standalone muon track, based on a Kalman-filter technique [86].

- A tracker muon is constructed by considering all tracks in the CMS tracker system with a $p_T > 0.5$ GeV and total momentum $p > 2.5$ GeV as muon candidates. These candidates are extrapolated to the muon spectrometer, trying to find a match[5] with at least one muon track segment. The extrapolation takes into account the effects of Coulomb scatterings, magnetic field and expected energy losses in detector material. If such a match exists, the track is identified as a tracker muon.

For low momenta, the tracker muon definition is more efficient at identifying muons due to the fact that only one track segment is required. Tracker muons are therefore

---

[5]A match is considered valid if the distance between the extrapolated track and the track segment is less than 3 cm.

more capable of identifying muons coming from the hadronisation of b or c quarks. For reconstructing isolated global muons, it is required that in a cone with size $\Delta R = 0.3$ around the muon, the sum of p_T of additional tracker tracks and energy deposits in the calorimeters is less than 10% of the muon p_T.

3.3.3   Electron object reconstruction

The traditional approach for identifying electrons is founded in the so-called *ECAL-driven approach*. In this approach an energetic ECAL cluster ($E_T > 4$ GeV) is used as a seed for the electron reconstruction. The ECAL cluster is enlarged in the ϕ direction, the direction along which the magnetic field bends the electron, to include all energy deposits due to radiations (Bremsstrahlung) of the electron. This enlarged cluster is commonly referred to as the super-cluster. The super-cluster is linked to tracker seeds by backwardly extrapolating the energy-weighted position of the super-cluster into the tracker. Those tracker seeds are redefined as electron seeds.

An issue in the ECAL-driven approach is the identification of electrons in the hadronically enriched regions. Due to the overlap of particle contributions from hadrons in the ECAL, the energy-weighted position of the super-cluster is biased, causing large linking inefficiencies. On top of that, the presence of possibly charged hadrons in that part of phase-space may lead to a backward extrapolation being compatible with many tracker seeds. To incorporate these electrons into the PF reconstruction, the *tracker-driven approach* is adopted.

The tracker-driven approach consists of considering all tracks from iterative tracking (cfr. Section 2.2.1) with a $p_T > 2$ GeV as electron track seeds. A pre-identification is applied considering two separate cases, in order to reduce fake identification of charged hadrons as electrons. The first case considers electrons with limited radiation, in which case the track seed generally has hits in all tracker layers. This allows to easily propagate the track to the ECAL inner surface and match it to an ECAL energy cluster, pre-identifying it as an electron if the cluster energy and the track momentum are compatible. The second case considers electrons that have large radiation effects due to the interaction with the thick tracker material. Tracks coming from high-radiating electrons typically have less hits in the tracker layers and a high χ^2 of the Kalman-filter track fit. An alternative track fit is applied to such tracks using a Gaussian-sum filter (GSf) instead of the Kalman-filter. Using a sum of Gaussian distributions in the fitting procedure to estimate the effects of energy loss due to Bremsstrahlung and interactions with tracker material, is way more effective than the single-Gaussian distribution in the Kalman-filter to describe energy losses in the tracker. These GSf tracks are then propagated to the ECAL surface through a multivariate analysis using GSf track- and Kalman filter properties combined with the extrapolated distance to the closest ECAL cluster.

The seeds from the ECAL- and tracker-driven approach are collected into a list, which are then all passed through the GSf tracking procedure. The GSf track seeds, together with the linked ECAL clusters and other associated tracks from the PF block,

are submitted for further PF electron identification via a multivariate analysis.

3.3.4   Jet object reconstruction

An experimentally unfavourable feature in identifying and reconstructing partons, is the fragmentation and hadronisation of these particles. As opposed to electrons or muons, the parton can not be associated to one of the PF-level particles. The reconstruction of partons becomes a matter of correctly clustering PF-level particles into a *jet*. A jet serves then as the high-level physics object representation of the parton in the final event reconstruction. In clustering hadrons into jets, several sources may affect the reconstructed energy scale of the jet, which needs to be corrected for. Apart from the scale, the energy resolution of jets in simulated samples differs from the one measured in the data, introducing the need to calibrate the simulated jet energy resolution.

The jet anti-k_t clustering algorithm

Two general distinctions are made in the area of jet clustering algorithms: cone and sequential recombination algorithms [87]. The main difference between the two lies in the starting point of the algorithms. Cone algorithms start from a seed particle and cluster particles comparing their spatial separation (with respect to the seed particle) to a predefined cone-size. Sequential recombination algorithms start from the full list of particles in the event, basing the clustering decision on the comparison of inter-particle distance to the particle-beam distance. Due to the nature of the two types of algorithms, their computing times significantly differ. Some first rough estimates for specific cone algorithms gave a computing time proportional to $N2^N$, with N being the number of particles, whereas some sequential recombination algorithms timings were proportional to N^3. This proportionality to computing time is a first indication of sequential recombination algorithms being preferrable in high-luminosity environments. However different these algorithms may be, the common goal is to reconstruct jets that are infrared safe (i.e. insensitive to soft radiations) and collinear safe (i.e. insensitive to collinear splitting).

Jets that are used in this thesis, are clustered according to the anti-k_t algorithm [88], which is an example of a sequential recombination algorithm. In general for sequential recombination algorithms, one considers the following distance definitions for particles i, j and the beam B:

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{3.6a}$$

$$d_{iB} = k_{T,i}^{2p}, \tag{3.6b}$$

where k_T represents the transverse momentum of the particle and $\Delta_{ij}^2 = \Delta y_{ij}^2 + \Delta \phi_{ij}^2$, with y the rapidity. The parameter R is referred to as the radius parameter and p a parameter to regulate the relative power of the transverse momenta with respect to

the geometrical difference Δ_{ij}^2. For the anti-k_T algorithm, the parameter p assumes a value of -1. This choice for p ensures soft particles are more likely to be associated to a closeby hard particle. For every pair of particles i and j, the distance d_{ij} is compared to the minimum of d_{iB} and d_{jB}. If the distance d_{ij} is smaller, the two particles are combined into a *pseudojet* by adding their four-momenta. The two particles are then removed from the list and replaced by the pseudojet. This process is repeated over again until d_{ij} is larger than d_{iB}, in which case the pseudojet i becomes a reconstructed jet. The radius parameter is chosen as 0.4 in the PF event reconstruction at CMS.

Jet energy scale corrections

A complication in reconstructing jets arises when comparing the generated parton energy to the jet energy. A fraction of the parton energy will not be registered due to dead zones in the detector or because of low momentum charged particles being trapped in the strong magnetic field. Other factors complicating the jet energy reconstruction are, among others, the non-linearity of the detector response and the presence of pileup particles. To correct the jet energy scale for these effects, CMS has adopted a factorized approach [89, 90]. At each level the jet four-momentum is corrected and fed to the next level of the factorized approach. The applied sequential levels are:

- **L1 Pileup.** The effects of pileup on the jet energy scale are estimated from simulations. In simulations, pileup effects are only introduced when simulating detector response (Section 3.2.2). This means one can disentangle the effects from pileup contributions by comparing the fully reconstructed jet to *generator jets* (which are reconstructed jets before detector interactions are simulated). The contribution from pileup interactions to the jet energy is estimated based on the concept of *hybrid jet areas*[6]. By subtracting the estimated pileup contribution, the transverse momentum of the jet gets corrected as $p_{T,L1} = p_{T,uncorr}$ - κ, where κ represents the pileup induced offset.

- **L2L3 Monte Carlo correction.** After correcting for pileup effects, the generator-jet transverse momentum p_T^{gen} still differs from the L1 corrected jet transverse momentum. The transverse momentum gets corrected for this difference as

$$p_{T,L2L3} = \left(\langle \frac{p_{T,L1}}{p_T^{gen}} \rangle [\eta, p_{T,L1}] \right)^{-1} p_{T,L1}, \tag{3.7}$$

  where $\langle \rangle$ represents the average for pre-defined bins in $[\eta, p_{T,L1}]$. These binned correction factors are determined from a multijet simulation sample.

- **L2L3 Residuals.** Some small residual differences remain in the jet energy response for simulation and data. These residuals are corrected for jets observed in

---

[6]Hybrid jet areas are defined by multiplying the effective jet area with the average energy density in the event. In this definition, the jet area represents the softness of jets in the event.

data only. The L2 residual correction is measured in dijet events where the two jets are back-to-back in azimuth. One of the jets is required to lie in the central region of the detector $|\eta| < 1.3$, where the detector is uniform. By requiring a balance in the transverse momentum between the two jets, the difference in jet energy between the two jets is used to correct the jet energy dependence of η. The L3 residual is obtained from Drell-Yan events in association with jets. Jets that are produced back-to-back with the leptonically decaying Z or γ can be used for jet calibration. The resolution of the transverse momentum of the lepton pair is relatively high compared to that of the jet. Therefore it can be used to estimate the transverse momentum residual of the jet.

More jet energy calibrations are available, correcting for example underlying event activity. However, these corrections are not applied in this thesis, as they have minimal effects.

Jet energy resolution correction

The resolution of the jet energy is corrected by the scaling method as documented in [89]. This method alters the jet p_T by rescaling it with

$$c_{JER} = 1 + (s_{JER} - 1)\frac{p_T - p_T^{ptcl}}{p_T}, \tag{3.8}$$

where p_T is the reconstructed transverse momentum, p_T^{ptcl} is the transverse momentum of the corresponding jet clustered from generator-level particles, and s_{JER} is the resolution scale factor. The resolution scale factor is measured in bins of η and is determined as given in Table 3.1 [91].

Table 3.1: Jet energy resolution scale factors in bins of η with uncertainty.

$|\eta|$	Scale factor	Uncertainty ($\pm 1\sigma$)
0-0.5	1.109	0.008
0.5-0.8	1.138	0.013
0.8-1.1	1.114	0.013
1.1-1.3	1.123	0.024
1.3-1.7	1.084	0.011
1.7-1.9	1.082	0.035
1.9-2.1	1.140	0.047
2.1-2.3	1.067	0.053
2.3-2.5	1.177	0.041

A comprehensive demonstration of the application of jet energy calibrations to the jet energy scale and resolution is given in Figure 3.3. This figure shows the normalized

distribution of the reconstructed mass of a hadronically decaying W boson in semi-leptonic $t\bar{t}$ events. The mass gets reconstructed by taking the invariant mass of the two jets that are matched with generator partons coming from the W boson. The calibration of those two jets results in a slightly broader mass resolution and a shift of the mass peak by $\sim +1$ GeV.
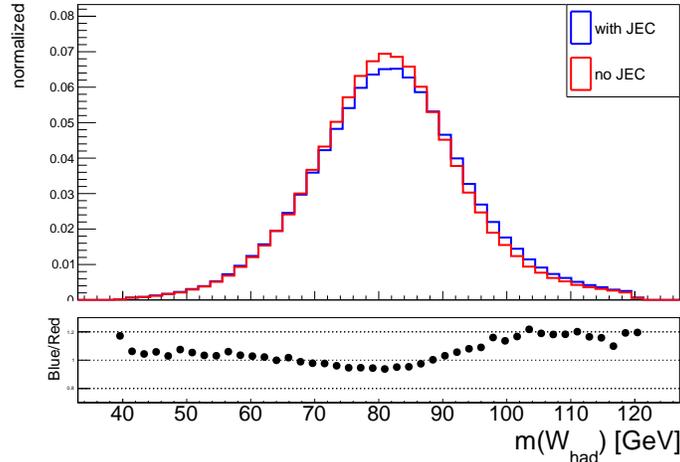
Figure 3.3: Reconstructed mass of a hadronically decaying W boson in semi-leptonic $t\bar{t}$ events before (red) and after (blue) applying jet energy calibrations. The ratio of the two distributions is indicated by the black dots.

3.3.5   Identification of b quark jets

The hadronisation and decay characteristics of light quarks (u, d and s) and gluons are almost identical. Therefore, jets originating from these partons have practically the same characteristics such that the identification of the original parton flavour of these jets is nearly impossible. Heavier quarks show distinct decay and hadronisation properties, resulting in observable differences with respect to other partons. As mentioned before, top quarks decay before hadronisation and are therefore not subject to jet reconstruction techniques. Bottom and charm quarks however may hadronise into heavy hadrons, which result into recognizable detector signatures. The CMS collaboration has developed several algorithms for the identification of jets originating from b and c quarks [92]. The identification of b quark jets is a crucial part of this thesis and will therefore be discussed in more detail.

The techniques for identifying c quark jets [93] are very similar to those used for b quark jet identification. The c quark identification algorithm, being developed in 2016, is a novelty within the CMS collaboration, and is considered as a very promising addendum to explore uncovered areas in new-physics searches.

Identification algorithms

Over the years, many different b quark jet identification algorithms have been developed by the CMS collaboration. All these algorithms are based on the same characteristic hadronisation process of b quarks, namely the production of a B meson, which is a bound state of a b quark and a light or c quark. As illustrated in Figure 3.4, the B meson has a lifetime which allows it to travel a measurable distance. The decayed B meson products are mostly charged particles, which produce hits in the tracker. The associated reconstructed tracks point to the same origin, constituting a displaced secondary vertex with respect to the primary interaction point. When extrapolating tangents to these tracks back to the primary vertex, the impact parameter (IP) can be defined as the extrapolated-track distance with respect to the primary vertex.
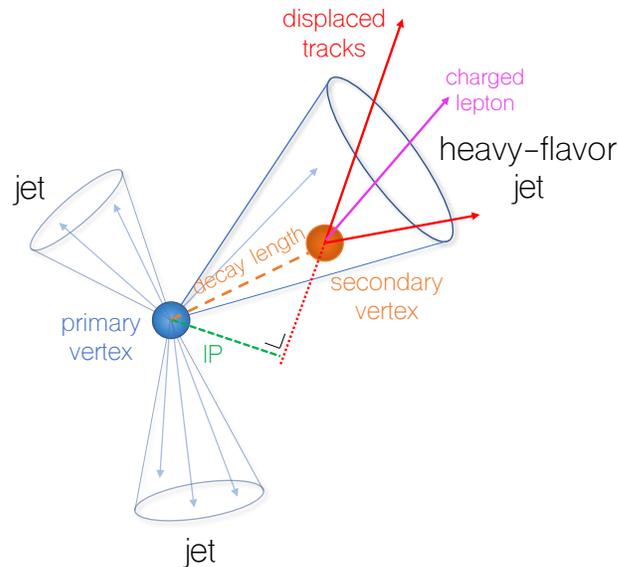
Figure 3.4: Illustration of a b quark jet with a secondary vertex from the decay of the B meson resulting in charged particle tracks (including possibly a soft lepton) that are displaced with respect to the primary interaction vertex, and hence with a large impact parameter (IP) value [13].

Tracks coming from B meson decays are displaced with respect to the primary vertex and therefore tend to have larger IP values than reconstructed tracks in light- and gluon jets. The IP significance defined as IP/σ_{IP}, with σ_{IP} the IP uncertainty, renders a more appropriate base for defining the degree of track displacement. The IP significance of the track with the highest IP is shown in Figure 3.5 for a selection enriched in b-flavoured jets. This illustrates that b-flavoured jets tend to contain tracks with higher IP significance. A mismatch between data and simulation exists for this (and other) variable(s), which contributes to an overall mismatch between data and simulation for b jet identification algorithms. Corrections factors are introduced to correct such behaviour, as will be discussed further in this Section. A sign is also

attributed to the IP, giving a positive sign to tracks that originate along the jet direction and a negative sign for the other tracks. In about 20% of B meson decays, a muon or electron is produced. These leptons are generally softer than promptly produced and isolated leptons.
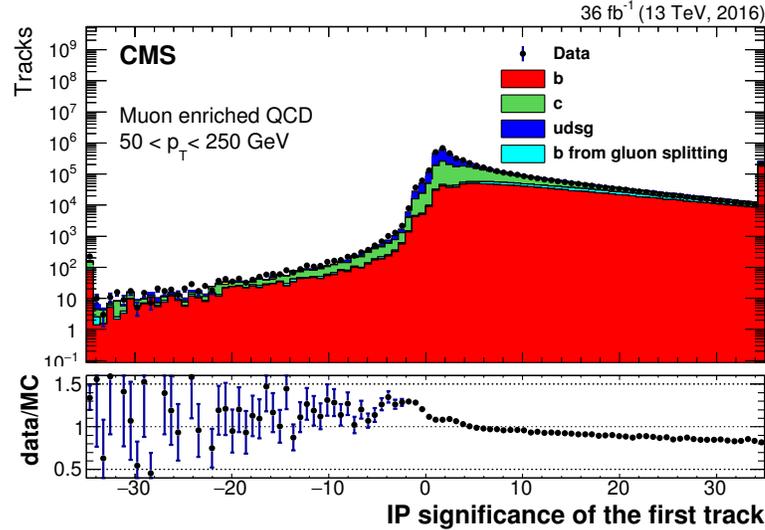
Figure 3.5: Data to simulation agreement for the IP significance of the track with the highest IP significance in a jet. The selection is enriched in multijets containing soft muons.

Only jets with well-reconstructed tracks are allowed for b quark identification algorithms. A number of quality requirements are imposed on the reconstructed tracks to achieve this. Jets that don't fulfill the quality requirements, are assigned a default (negative) b quark identification value.

For analysing pp-collisions recorded during LHC Run 2, the CMS experiment has developed (and maintained) the following various b quark jet identification algorithms:

- **Jet probability (JP) algorithms.** The probability for the jet to originate from the primary vertex can be calculated. This probability is calculated based on the individual track probabilities to originate from the primary vertex. The probability calculation builds on the track IP significance in comparison to the determined IP resolution[7]. A variant to the JP algorithm, called the jet B-probability (JBP), is considered where only the four tracks with the highest IP are used for the jet probability calculation.

- **Combined Secondary Vertex algorithms.** Multivariate techniques can be used to combine information from displaced tracks and secondary vertices. The

---

[7]The IP resolution is constructed from tracks with a negative IP. Such tracks are more likely to constitute jets coming from light-quarks or gluons.

algorithms developed by the CMS collaboration based on such techniques are called Combined Secondary Vertex (CSV) algorithms [94]. For LHC Run 1, a likelihood ratio [95] was used as multivariate technique, whereas for Run 2 a second version using neural networks [96] is introduced as CSVv2. One of the novel CSV developments in 2016 is the *DeepCSV* algorithm, which performs a deep learning neural network training.

The CSVv2 algorithm is trained on multijet events in three independent vertex categories and in bins of jet p_T and $|\eta|$. The three vertex categories are labelled as *reco-vertex*, *pseudo-vertex* and *no-vertex*. Jets ending up in the reco-vertex category have at least one well reconstructed secondary vertex. When no secondary vertices are found, a pseudo-vertex is considered for tracks with a signed impact parameter significance exceeding 2. If no such pseudo-vertex is found, the jets end up in the no-vertex category.

There are two ways in which one can try to reconstruct secondary vertices: *Adaptive Vertex Reconstruction (AVR)* [97] or *Inclusive Vertex finding (IVF)*. The AVR fits tracks to a vertex, using only good-quality tracks from the jet in its procedure. Once a vertex is found in the jet, the associated tracks are removed and a next vertex is sought for with the remaining tracks. The IVF technique starts from the collection of all tracks in the event, with a transverse momentum exceeding 0.8 GeV and a longitudinal IP smaller than 0.3 cm. From this collection, track seeds are selected based on their IP and IP significance. Tracks are then clustered based on the compatibility with the track-seed. The cluster of tracks is then passed through the adaptive vertex fitting. If at least 70% of the tracks between two IVF secondary vertices are shared and the distance significance between the two vertices is less than 2, one of the vertices is removed.

The efficiency to reconstruct a secondary vertex for b quark jets using the IVF algorithm is about 10% higher compared to the efficiency to reconstruct a secondary vertex with the AVR algorithm. For light quark/gluon jets the probability to find a secondary vertex increases by about 8%. This means that the impact of using the IVF vertices instead of AVR vertices in the CSVv2 algorithm is nontrivial.

From the reconstructed vertices and the collection of tracks, a set of discriminating variables is chosen to train the neural network in the three vertex categories. Some examples of the discriminating vertex-related variables are: the mass of the first reconstructed secondary vertex[8], its flight distance significance in the transverse plane and the number of secondary vertices. Considered track-related discriminating variables are, among others, the number of tracks in the jet, their 3D impact parameters and the signed IP of the track that raises the mass of the secondary vertex above the c quark mass. The training in each of the three vertex categories and $[p_T, |\eta|]$-bins is done considering c quark jets

---

[8]The secondary vertices are sorted according to increasing uncertainty on the flight distance.

and light quark/gluon jets separately. The resulting neural network outputs are then combined, weighting each vertex category training according to the relative appearances in a $t\bar{t}$+jets sample. The final CSVv2 discriminator attained by summing the c quark and light-quark/gluon training results with a weight of 25% and 75% respectively.

- **Soft electron (SE) and soft muon (SM) algorithms.** For B meson decays that contain a charged leptonic decay, soft electron and muon algorithms are developed. To identify the jet flavour, discriminating variables derived from the soft lepton track(s) and their relation to the jet-axis, are used in a Boosted Decision Tree (BDT) to construct the SE and SM discriminator.

- **Combined MVA (cMVAv2) algorithm.** The combined MVA uses a BDT to construct a single b quark identifying discriminator, based on six of the aforementioned algorithm discriminators. The six discriminators are: JP, JBP, SE, SM, CSVv2 with AVR vertices and CSVv2 with IVF vertices.

The performance of these algorithms is generally evaluated with respect to two different scenarios: discriminating b quark jets from light-quark/gluon jets on the one hand, and from c quark jets on the other. By defining a threshold in the algorithm's discriminator, jets exceeding (failing) the threshold are labelled as b-tagged (non b-tagged). The efficiency for jets exceeding the threshold is then used as a figure of merit for comparison purposes. Figure 3.6 shows the performance comparison of the CSVv2 (IVF), DeepCSV and cMVAv2 algorithms. As the purpose of the algorithms is to correctly b-tag jets originating from b quarks whilst keeping the misidentification of non-b quark jets as low as possible, performance curves that approach the lower-right corner in Figure 3.6 are desirable. Depending on the needs of an analysis, one can define working points in the discriminator space. Typically, three working points (loose, medium and tight) are defined, corresponding to a misidentifiation probability for light-quark/gluon jets of respectively 10%, 1% and 0.1%. It must be noted that these performance curves strongly depend on the sample on which they are validated, which is a $t\bar{t}$ sample in this case.

Data to simulation scale factors

Simulated physics samples provide a decent, well-understood description of the measured data for all variables used in b quark identification algorithms. However, in the final discrimination variables some small deviations in the data to simulation agreement exist, as shown in Figure 3.7. These deviations can result in different tagging performances between data and simulated samples. In order to correct for these deviations, scale factors are measured to correct simulated events. The scale factors are measured depending on η, p_T and discriminator value for jet flavours f, as

$$SF_f(\eta, p_T, \text{disc.}) = \frac{\epsilon_f^{data}(\eta, p_T, \text{disc.})}{\epsilon_f^{sim}(\eta, p_T, \text{disc.})}, \tag{3.9}$$
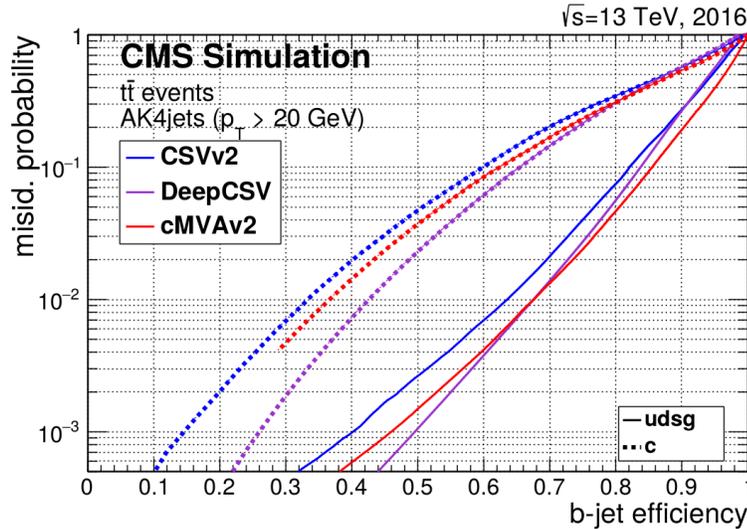
Figure 3.6: Performance of the CSVv2, cMVAv2 and DeepCSV algorithms demonstrating the probability for non-b quark jets to be misidentified as b quark jets as a function of the efficiency to correctly identify b quark jets. The curves are obtained from simulated $t\bar{t}$ events using jets with transverse momentum above 20 GeV. The DeepCSV algorithm outperforms all other taggers for both c quark jets as well as light-quark/gluon jets [13].

with ϵ the b-tagging efficiency. The flavour of jets in simulations is determined from the matched generated hadrons. In data the efficiencies are measured by selecting a sample enriched in jets with flavour f.

Depending on the use of b-tagging in an analysis, the scale factors are applied (and measured) in different ways. When an analysis solely uses b-tagging as a way of selecting events with a number of (non-)b-tagged jets, the scale factors are applied to the b-tagging efficiencies and mistag rates according to the chosen b-tagging discriminator working point [13]. If, however, an analysis depends on the distribution of the b-tagging discriminator, scale factors are applied to calibrate the data to simulation agreement. This method is referred to as the IterativeFit [98].

The b-tagging efficiencies in data for the IterativeFit are measured in events with two oppositely charged leptons (electron or muon) and exactly two jets. To enrich the sample in b-flavoured jets, a control region[9] is defined to select $t\bar{t}$-like events. In these events, the two jets are expected to originate mostly from b quarks. By applying a medium b-tag requirement on one of those jets, the remaining jet is used for the b-

---

[9] The control region is implemented by requiring the invariant mass of the two leptons to not coincide with the mass of the Z-boson (i.e. $|M_{\ell\ell} - M_Z| > 10$ GeV). The missing transverse energy is required to exceed 30 GeV.
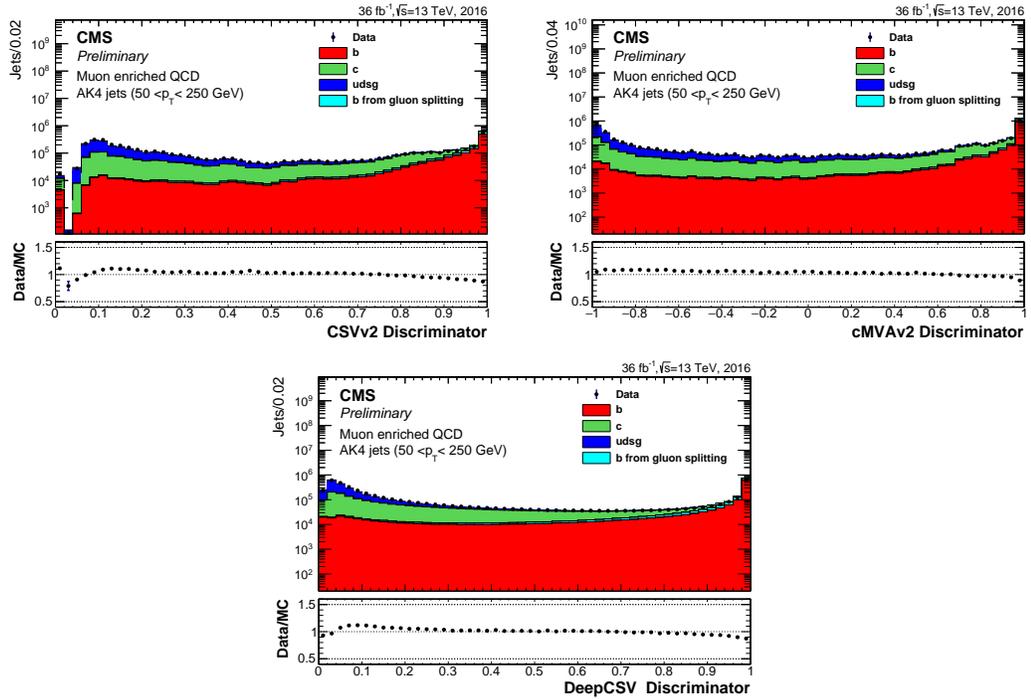
Figure 3.7: Data to simulation agreement for the CSVv2 (upper left), cMVAv2 (upper right) and DeepCSV (bottom) discriminators in a multijet sample, enriched in soft muons [13].

flavour scale factor measurement. The contamination of non-b quark jets is estimated from simulation and is subtracted from the measurement. To enrich the sample in light-flavour jets, a control region is defined to select Z+jets events, which is achieved by inverting the $t\bar{t}$ control region selection requirements. Requiring one of the two jets to not be loosely b-tagged, the remaining jet is expected to originate from a light quark or gluon and can therefore be used to measure the light-flavour scale factors. The contamination of b- and c quark jets is estimated from simulation and is subtracted from the measurement.

As the b-flavour scale factor measurement impacts the light-flavour scale factors (and vice versa), the scale factors are determined iteratively (hence IterativeFit). The iterative procedure first measures the scale factors when no scale factors are applied to the simulations used for the contamination estimation. The scale factor measurement is iterated, applying the scale factors from the previous iteration to the simulation. This procedure is iterated until the scale factors are stable with respect to the previous iteration, which is achieved after three iterations. The scale factors for c-flavour jets is set to unity during the whole procedure. In the end, scale factors are obtained in bins of p_T, η and CSVv2 discriminant value are determined for the different flavours as

$$
\begin{aligned}
SF^{\text{b flavour}}(p_T, \eta, \text{disc.}) &= \frac{N_{data}^{\text{b flavour}} - N_{sim}^{\text{b flavour}}}{N_{sim}^{\text{b flavour}}} \\
SF^{\text{light flavour}}(p_T, \eta, \text{disc.}) &= \frac{N_{data}^{\text{light flavour}} - N_{sim}^{\text{light flavour}}}{N_{sim}^{\text{light flavour}}} \\
SF^{\text{c flavour}}(p_T, \eta, \text{disc.}) &= 1,
\end{aligned}
\tag{3.10}
$$

with N the number of selected jets in the considered (p_T, η, discriminant) bin.

3.3.6   Missing transverse energy

Events where non-detectable particles such as neutrinos and potential dark matter particles are produced, give rise to an energy inbalance in the transverse plane. The energy inbalance, commonly referred to as missing transverse energy, in the PF event reconstruction is calculated as

$$
\vec{\slashed{E}}_T = - \sum_{i=1}^{N_{particles}} \vec{p}_{T,i}.
\tag{3.11}
$$

In order to take the jet energy corrections into account, the $\vec{\slashed{E}}_T$ is corrected such that the uncorrected transverse momentum of each PF jet with $p_T > 10$ GeV gets replaced by the corrected transverse momentum $\vec{p}_{T,j}^{\,corr}$ as

$$
\vec{\slashed{E}}_T = - \sum_{i=1}^{N_{particles}} \vec{p}_{T,i} - \sum_{j=1}^{N_{PFjets}} (\vec{p}_{T,j}^{\,corr} - \vec{p}_{T,j}).
\tag{3.12}
$$

Note that the missing transverse energy should not be solely interpreted as unde-tectable particles in the event. Resolution effects as well as detectable particles passing through dead zones of the detector contribute to the $\vec{\not{E}}_T$ as well.

3.3.7   Event post-processing

In the PF event reconstruction misidentifications and misreconstructions have a non-zero probability of occurance. Usually this has a very small impact on the event, e.g. photons escaping the ECAL calorimeter and being misidentified as neutral hadrons. Some misreconstructions or misidentifiations however can have a large impact on the $\vec{\not{E}}_T$ and/or the jet energy. Muons from cosmic rays can artificially and substantially increase the $\not{E}_T$ when they coincide with a bunch crossing. Cosmic muons are removed from the event by requiring the distance of the muon track to the beam pipe to be larger than 1 cm and the removal resulting in a reduced $\not{E}_T$.

Misreconstruction of the muon momentum may happen as well, as there might be a difference between the momentum reconstructed in the tracker and the muon system. This difference can arise from interactions in the iron yoke or synchroton radiation. As PF reconstruction may not always make the best choice for the muon momentum, the muon momentum is corrected to the choice rendering the smallest $\not{E}_T$.

Another cause of artificially high $\not{E}_T$ is the misidentifiation of charged hadrons punching through to the muon system as muons. In this case the charged hadron will be double counted as being a muon (due to the reconstruction of muon tracks) and as a neutral hadron (due to the energy deposits in the calorimeters). These cases are corrected for by replacing the wrongly identified muon and neutral hadron by a charged hadron.

Finally the PF event reconstruction provides the possibility to remove charged hadrons from the event that are associated to a primary vertex other than the main primary vertex. Charged hadrons associated to other primary vertices are assumed to come from pileup interactions. Removing these hadrons from reconstructed jets is referred to as *charged hadron subtraction*.

Chapter 4

A phenomenological sensitivity projection for top-FCNC

Many searches for various top-FCNC have been performed, but no deviations from SM predictions have been found so far. All top-FCNC searches resulted only in upper limits on their branching ratio, as summarised in Section 1.2.3. The current best limits have all been obtained by the CMS and ATLAS collaborations, using pp-collisions recorded during LHC Run 1 in 2012 at a collision energy of 8 TeV and about 20 fb^{-1} of collision data. During Run 2 of the LHC however, protons are collided at 13 TeV and a much higher luminosity of collision data will be recorded. With more data available at higher collision energies, experimental searches are more sensitive to top-FCNC than during Run 1. In order to estimate the sensitivity to several top-FCNC topologies under LHC Run 2 collision conditions, a phenomenological study has been performed. The sensitivity estimation is done for a benchmark scenario that can be achieved by the LHC during Run 2, namely an integrated luminosity of 100 fb^{-1} of recorded pp-collisions at a center-of-mass energy of 13 TeV.

An account of the top-FCNC analysis spectrum considered in this phenomenological study is given in Section 4.1. In order to evaluate the sensitivity at a benchmark scenario, a simulation of all involved SM and top-FCNC processes is performed. The detector signatures of these processes are emulated by means of a DELPHES [99] simulated LHC detector as described in Section 4.2. Finally, a short overview of possible analysis stategies is given in Section 4.3, together with their projected sensitivity at the chosen benchmark scenario.

This (unpublished) phenomenological study was performed in collaboration between three research groups from the *Vrije Universiteit Brussel*, the *Institut Pluridisciplinaire Hubert CURIEN* at Strasbourg and the *Chonbuk National University* in South-Korea. The author was one of the 10 researchers participating in this study, mainly contributing to the Hqt-FCNC (H $\rightarrow$ b$\bar{\text{b}}$) analysis (see Section 4.3). Therefore, relatively more details about that specific analysis are given in this chapter compared to the other analyses.

4.1   The spectrum of top-FCNC

The FCNC interactions, as described by the effective Lagrangian in Equation (1.23), constitute new-physics signals. For each anomalous coupling in the Lagrangian, a new-physics signal is produced by only considering that coupling as non-vanishing. The complex chiral parameters are arbitrarily chosen and all fixed such that $f_{Xq}^{L} = 0$ and $f_{Xq}^{R} = 1$. In this way five new-physics signals can be constructed, corresponding to the couplings κ_{gqt}, $\kappa_{\gamma qt}$, κ_{Zqt}, ζ_{Zqt} and κ_{Hqt}. Each of the new-physics signals is composed of two components; the SM production of a top quark pair followed by the FCNC decay of one of the top quarks and the SM decay of the other top quark, and the anomalous FCNC production of a single top quark decaying in a SM way. The cross section for the $t\bar{t}$ component is derived from the SM $t\bar{t}$ cross section by considering the decay as: $t\bar{t} \to (bW^{\pm})\,(Xqt)$. Taking the branching ratio for the SM top decay as $\mathcal{B}(t \to bW^{\pm})$ = 1 and knowing the FCNC branching ratio $\mathcal{B}((t \to Xq)$, the calculation of this cross section is straightforward. The branching ratio for each of the five anomalous decays is obtained from its partial decay width (see Equation (1.25)), which is calculated by the `MADGRAPH5_aMC@NLO` event generator and given in Table 4.1.

Table 4.1: Leading-order partial decay widths related to the anomalous decay modes of the top quark. The new-physics scale Λ has to be given in GeV. Small differences at the permille level in decay widths exist for q = u and q = c, but are not shown here for simplicity.

Coupling	Partial decay width [GeV]
κ_{gqt}	$3.67 \times 10^{5}\,(\kappa_{gqt}/\Lambda)^{2}$
$\kappa_{\gamma qt}$	$1.99 \times 10^{4}\,(\kappa_{\gamma qt}/\Lambda)^{2}$
κ_{Zqt}	$1.64 \times 10^{4}\,(\kappa_{Zqt}/\Lambda)^{2}$
ζ_{Zqt}	$0.169\,\zeta_{Zqt}^{2}$
κ_{Hqt}	$0.190\,\kappa_{Hqt}^{2}$

The anomalous production of a single top quark occurs when a gluon from one proton interacts with an up (or charm) quark from the opposing proton. In the case where only the coupling κ_{gqt} is considered, this leads to a single top quark in the final state at LO, as depicted in Figure 4.1. For the $\kappa_{\gamma qt}$, κ_{Zqt}, ζ_{Zqt} and κ_{Hqt} couplings, a new up (or charm) quark is created which then anomalously produces a single top quark in association with a photon, Z or H boson respectively, as illustrated in Figure 4.1. The LO cross sections are calculated by the event generator and summarised in Table 4.2. Due to the parton composition of protons, the anomalous single-top production cross sections involving an up quark are roughly ten times higher than when a charm quark is involved.
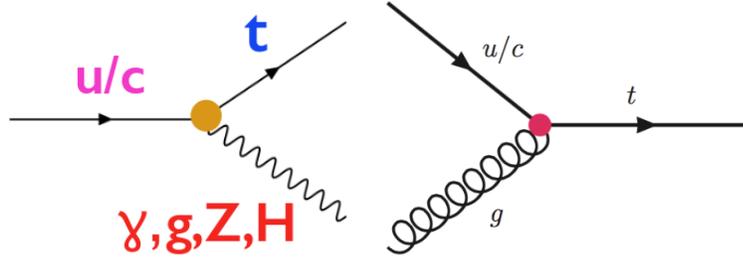
Figure 4.1: LO Feynman diagrams for the anomalous single top quark production in association with a photon, gluon, Z or H boson (left). The coupling κ_{gqt} can also lead to the anomalous production of a single top quark (right), without any additional particle.

Table 4.2: Leading-order single top production cross sections for the anomalous top interactions, when the new physics scale Λ is given in GeV.

Process	Coupling	Cross section [pb]
pp→ t $\oplus$ $\bar{t}$	κ_{gut}	1.15×10^{12} $(\kappa_{gut}/\Lambda)^2$
	κ_{gct}	2.01×10^{11} $(\kappa_{gct}/\Lambda)^2$
pp→ tγ $\oplus$ $\bar{t}\gamma$	$\kappa_{\gamma ut}$	1.41×10^{7} $(\kappa_{\gamma ut}/\Lambda)^2$
	$\kappa_{\gamma ct}$	2.01×10^{6} $(\kappa_{\gamma ct}/\Lambda)^2$
pp→ tZ $\oplus$ $\bar{t}$Z	κ_{Zut}	1.68×10^{7} $(\kappa_{Zut}/\Lambda)^2$
	κ_{Zct}	2.24×10^{6} $(\kappa_{Zct}/\Lambda)^2$
	ζ_{Zut}	72.8 ζ_{Zut}^2
	ζ_{Zct}	11.0 ζ_{Zct}^2
pp→ tH $\oplus$ $\bar{t}$H	κ_{Hut}	72.6 κ_{Hut}^2
	κ_{Hct}	9.85 κ_{Hct}^2

In this phenomenological study all relevant coupling strengths are constrained to

$$\frac{\kappa_{Xqt}}{\Lambda} = 0.1 \text{ GeV}^{-1} \quad \text{or} \quad \zeta_{Zqt} = 0.1 \quad \text{or} \quad \kappa_{Hqt} = 0.1 \ , \tag{4.1}$$

with X either the gluon, Z boson or photon. Each of the five new-physics signals can be classified into different final state topologies, depending on the number of charged leptons, photons and jets that are produced in the process. For practical reasons, the $\kappa_{\gamma qt}$ coupling is not considered in this phenomenological study. The ζ_{Zqt} coupling is not considered as well, due to the fact that its cross sections are much lower than the κ_{Zqt} case, which has an identical final state topology. With these simplifications the topology classification is considered for three new-physics cases: *gqt-FCNC*, *Zqt-FCNC* and *Hqt-FCNC*.

gqt-FCNC

The $t\bar{t}$ component of the gqt-FCNC signal results in topologies with high hadronic activity due to the hadronisation of the gluon. As such hadronic activity does not provide any discriminatory power from SM $t\bar{t}$ topologies, this component is not considered.

The leading-order single top production component however, has an advantageous characteristic final state topology with respect to SM processes. Single top quark production in the SM is always associated with either an extra quark or W boson, resulting in a final state with at least one extra jet compared to the LO gqt-FCNC signal. The decay of the single top quark is considered only for the leptonic case ($t \to bW \to b\ell\nu$), since the hadronic decay ($t \to bW \to bqq\prime$) is overwhelmed by QCD multijet events. In terms of identifiable physics objects, the final state topology of interest consists of one isolated electron or muon[1] , missing transverse energy from the neutrino and one jet coming from a bottom quark.

Zqt-FCNC

The presence of a leptonically decaying Z boson in the signal process provides for an experimentally advantageous event signature[2]. When requiring the top quark in the signal process to decay leptonically as well, the final state topology has a very distinct signature. With the single top component of the signal decaying as $tZ \to (b\ell\nu)(\ell^+\ell^-)$ and the $t\bar{t}$ component as $t\bar{t} \to (bW)(qZ) \to (b\ell\nu)(q\ell^+\ell^-)$, a final state topology with exactly three isolated leptons (electrons and/or muons) with at least one jet coming from a b quark is achieved.

The hadronic decay of the Z boson is not considered for a combination of reasons. This final state topology would involve only one isolated lepton from the SM top decay,

---

[1]Tau leptons are not considered, as the branching ratio of tau leptons to hadrons is about 65%[26]. The remaining 35% represents a leptonic decay ($\tau \to \nu_\tau e\nu_e$ or $\tau \to \nu_\tau \mu\nu_\mu$).

[2]Two charged leptons from a Z boson decay typically have a sharp invariant mass peak around the Z boson mass.

opening up the analysis to contributions from large SM background processes, such as W + jets and $t\bar{t}$ production. One could increase the signal to background ratio by b jet identification algorithms in order to aim for the approximately 15% Z boson decay rate to two bottom quarks. However, assuming a 65% b jet identification efficiency (see Figure 3.6), one would only select about 40% of those Z $\rightarrow$ b$\bar{\text{b}}$ decays. In this scenario, b jet identification could make the analysis sensitive to about 6% of the total Z boson decay rate, which is of the same level as the 6.7% electron and muon decay rate of the Z boson. Taking into account the higher level of background events and the worse mass resolution for dijet systems, it is much more effective to search for Zqt-FCNC in a trileptonic environment.

Hqt-FCNC

The final state topologies for the Hqt-FCNC signal process are more diverse than in the previous cases. This is due to the many different decay scenarios of the Higgs boson, as summarised in Table 4.3.

Table 4.3: SM predictions of the Higgs boson decay branching ratios for the most relevant decay modes and their relative uncertainties for $m_H = 125$ GeV [26, 100].

Decay channel	Branching ratio	Rel. uncertainty
H $\rightarrow$ b$\bar{\text{b}}$	0.58	$+3.2\%$ -3.3%
H $\rightarrow W^+W^-$	0.21	$+4.3\%$ -4.2%
H $\rightarrow \tau^+\tau^-$	6.3×10^{-2}	$+5.7\%$ -5.7%
H $\rightarrow ZZ$	2.6×10^{-2}	$+4.3\%$ -4.1%
H $\rightarrow c\bar{c}$	2.9×10^{-2}	$+12\%$ -12%
H $\rightarrow \gamma\gamma$	2.3×10^{-3}	$+5.0\%$ -4.9%
H $\rightarrow Z\gamma$	1.5×10^{-3}	$+9.0\%$ -8.9%
H $\rightarrow \mu^+\mu^-$	2.2×10^{-4}	$+6.0\%$ -5.9%

In this phenomenological study, the following decay channels are considered:

- **H $\rightarrow$ b$\bar{\text{b}}$.** The decay channel with the largest branching ratio is the one where the Higgs boson decays into a pair of bottom quarks. Ignoring a fully hadronic final state by requiring the SM top to decay leptonically (see Figure 1.4), the single top component of the signal decays as $tH \rightarrow (b\ell\nu)(b\bar{\text{b}})$ and the $t\bar{t}$ component as $t\bar{t} \rightarrow (bW)(qH) \rightarrow (b\ell\nu)(qb\bar{\text{b}})$. Using a b-tagging algorithm, a final state with exactly one electron or muon and exactly three b-tagged jets can be selected.

- **H $\rightarrow W^+W^-$.** An interesting final state is obtained if the W from the SM top decay as well as the same charged W from the Higgs decay leptonically. In this case,

the decay chain for the single top component goes as $tH \rightarrow (bW^{\pm})(W^{\pm}W^{\mp}) \rightarrow (b\ell^{\pm}\nu)(\ell^{\pm}\nu qq')$, and as $t\bar{t} \rightarrow (bW^{\pm})(qW^{\pm}W^{\mp}) \rightarrow (b\ell^{\pm}\nu)(q\ell^{\pm}\nu q'q'')$ for the $t\bar{t}$ component. This final state contains two leptons of the same charge, a state that is not achieved by any of the major SM background processes.

- **H $\rightarrow \gamma\gamma$.** The $H \rightarrow \gamma\gamma$ decay channel provides a very clear Higgs signature and is considered as the channel most sensitive for Higgs searches. By considering the leptonic decay of the SM top quark, the decay chain for the single top component goes as $tH \rightarrow (b\ell\nu)(\gamma\gamma)$, and as $t\bar{t} \rightarrow (b\ell\nu)(q\ell\nu\gamma\gamma)$ for the $t\bar{t}$ component. An event signature containing exactly one electron or muon, two photons and one jet from a bottom quark, is nearly non-existent in the SM.

4.2   Detector simulation for analysis at the LHC

The event generation of all physics processes is performed by `MADGRAPH5_aMC@NLO` at LO. The events are reweighted such that the normalisation of the different samples matches the value of the NLO cross sections as computed by `MADGRAPH5_aMC@NLO`. The matching to parton showers is done with the help of the `PYTHIA` program. Background event generation includes single and double electroweak boson production, $t\bar{t}$ production, possibly in association with either one or two electroweak gauge bosons or a Higgs boson, and single top production, possibly in association with a Higgs or an electroweak gauge boson. The multijet background is ignored, since its correct treatment requires data-driven methods. Instead, a selection strategy of the different analyses is enforced such that the corresponding contribution is negligible.

A simulated LHC detector, based on the CMS geometry, is used to emulate the electronic signals. The response of the detector is simulated by means of the DELPHES 3 program [99] that includes a reconstruction of all physics objects. As a detector parameterization, the so-called MA5TUNE version [101] of the default CMS detector is used. Instead of using the b-tagging machinery of DELPHES, an analysis-level b-tagging emulation is implemented. Three *b*-tagging working points related to the loose, medium and tight selection criteria of the CSVv2 method (see Section 3.3.5) are considered. The corresponding efficiencies from CMS data of correctly identifying b-jets are fitted by polynomial functions of the jet transverse momentum p_T, based on public results from [102] as

$$\epsilon_b(p_T) = a_b + b_b\ p_T + c_b\ p_T^2 + d_b\ p_T^3\ , \tag{4.2}$$

and the mistagging rates of a jet originating from the fragmentation of a *c*-quark (ϵ_c) or a lighter parton (ϵ_ℓ) with a constant and a linear function of the jet p_T respectively,

$$\epsilon_c(p_T) = a_c \qquad \text{and} \qquad \epsilon_\ell(p_T) = a_\ell + b_\ell\ p_T\ . \tag{4.3}$$

All the coefficients of the above polynomials are given in Tables 4.4 and 4.5 for the different working points.

Table 4.4:   Coefficients related to the parameterization of the b-tagging efficiencies associated with Equation (4.2) for the different working points, based on [102].

Scenario	Jet p_T	a_b	b_b [GeV^{-1}]	c_b [GeV^{-2}]	d_b [GeV^{-3}]
Loose	$p_T \leq 180$ GeV	0.62998	0.00422927	$-2.28273 \cdot 10^{-5}$	$3.67139 \cdot 10^{-8}$
	$p_T > 180$ GeV	1.02904	-0.00135858	$2.96676 \cdot 10^{-6}$	$-2.28016 \cdot 10^{-9}$
Medium	$p_T \leq 180$ GeV	0.41136	0.00754662	$-5.25669 \cdot 10^{-5}$	$1.14005 \cdot 10^{-7}$
	$p_T > 180$ GeV	0.78317	-0.00026665	$-5.61403 \cdot 10^{-8}$	0
Tight	$p_T \leq 180$ GeV	0.28795	0.00528263	$-3.67968 \cdot 10^{-5}$	$7.98035 \cdot 10^{-8}$
	$p_T > 180$ GeV	0.54822	-0.00018666	$-3.92982 \cdot 10^{-8}$	0

Table 4.5:   Coefficients related to the parameterization of the mistagging rates of a jet originating from the fragmentation of a c-quark and of a lighter parton (u, d, s or g) a b-jet, for the different working points (see Equation (4.3)), based on [102].

Scenario	Jet p_T	a_c	a_ℓ	b_ℓ [GeV^{-1}]
Loose	$p_T \leq 50$ GeV	0.350	0.230	0
	50 GeV$< p_T \leq 100$ GeV	0.350	0.110	0
	$p_T > 100$ GeV	0.350	0.093	0.000071
Medium	$p_T \leq 50$ GeV	0.200	0.0160	0
	$p_T > 50$ GeV	0.200	0.0104	$3.2 \cdot 10^{-5}$
Tight	$p_T \leq 50$ GeV	0.055	0.001800	0
	$p_T > 50$ GeV	0.055	0.001144	$3.52 \cdot 10^{-6}$

An analysis method dedicated to the modeling of the fake tagging of a jet as an electron or photon has been implemented. A jet for which the ratio of the hadronic to electromagnetic calorimetric energy is smaller than 0.15 is considered to have a non-vanishing probability to satisfy electron reconstruction and isolation[3] criteria, depending on its tranverse momentum and pseudorapidity.

4.3   Sensitivity to the top-FCNC processes

The strategy of each analysis consists in optimizing simple event selection criteria in such a way that the signal significance is maximal. The signal significance is defined as $S/\sqrt{S+B}$, where S is defined as the number of signal events and B the number of background events surviving the selection requirements. The selection criteria are optimized based on kinematic properties of selected objects as well as reconstructed event properties. After this series of baseline selection criteria, a second optimisation of the signal significance is performed by means of a multivariate analysis (MVA). A set of variables that carry some discriminating power between signal and background are combined via an MVA technique called *Boosted Decision Trees* (BDTs) [103] (see Section 5.3.2) to construct a single event discriminant. Background-like events will have low BDT discriminant values, whereas signal-like events end up at higher values. The signal significance is then optimized by requiring each event's BDT discriminant value to surpass a certain threshold. A short description of the baseline event selection criteria and MVA is given below for each considered top-FCNC signal process.

gqt-FCNC

Exactly one isolated lepton (electron or muon) is requested with $p_T > 30$ GeV and $|\eta| <$ 2.5. The criteria on relative isolation ($iso < 0.2$) is defined for a cone of 0.4 around the lepton direction. A veto is applied to reject the event if there is another isolated lepton with $p_T > 20$ GeV or an isolated photon with $p_T > 10$ GeV in the η acceptance ($|\eta| <$ 2.5). Only events with one or two jets are considered in this analysis, these jets should have $p_T > 20$ GeV, $|\eta| <$2.4, a fraction of hadronic/electromagnetic energy larger than 0.15 and an angular separation with the lepton $\Delta R(jet, lepton) > 0.4$. In particular the leading jet is requested to have $p_T > 30$ GeV and pass the tight criterium of the b-tagging algorithm in order to reduce the $W + jets$ contribution. Two additional cuts are applied in order to suppress the non-considered multijet background, the first one requests the missing transverse energy to be $\not{E}_T >$30 GeV, the second one is on the transverse mass of the system lepton+$\vec{\not{E}}_T$ aiming to reconstruct the mass of the W:

---

[3]In the context of this phenomenological study, the lepton isolation is a measure for the ratio of the sum of non-electromagnetic energy deposits and hadrons' transverse momenta in a cone around the lepton direction, to the transverse momentum of the lepton. For a lepton to be isolated, this ratio is required to be below a certain threshold.

$M_T(lepton, \vec{\rlap{/}E}_T) > 50$ GeV. The SM background composition for this baseline selection is dominated by 44% of W + jets, 31% of $t\bar{t}$ and 22% of single top production.

It was found that a set of nine variables gave the best signal-to-background discrimination power in a BDT. The most discriminating power is driven by the mass of the top system. The other eight variables are the transverse mass $M_T(lepton, \vec{\rlap{/}E}_T)$, the transverse momenta of the W and of the b-tagged jet, the η and charge of the lepton, and three angular distances between different particles: $\Delta\phi(b, W)$, $\Delta\phi(l, MET)$, $\Delta\phi(l, b)$. The *gct* and *gut* signals have similar behavior for all those variables, except for the charge and the η of the lepton as the *gut* scenario provides a charge asymmetry but not *gct*.

Zqt-FCNC

Exactly three isolated leptons (electron or muon) are required with $p_T > 20$ GeV and $|\eta| < 2.4$ (2.5). The selection criteria aim to optimally select both the $t\bar{t}$ and single top component of the signal. The criteria on relative isolation, $iso < 0.2$, is defined for a cone of 0.4 around the lepton direction. There should be at least one jet with: $p_T > 40$ GeV, $|\eta| < 2.4$, a fraction of hadronic/electromagnetic energy larger than 0.15 and an angular separation with the lepton $\Delta R(jet, lepton) \geq 0.4$. At least one of the jets is required to pass the medium criterium of the b-tagging algorithm. In order to reconstruct a Z boson, there should be at least one opposite sign same flavour lepton pair with an invariant mass inside a Z mass window of 15 GeV. Further, the reconstructed FCNC top should have an invariant mass inside the SM top mass window of 35 GeV. One additional cut is applied in order to suppress the non-considered multijet background, the transverse mass of the system lepton+$\vec{\rlap{/}E}_T$ aiming to reconstruct the mass of the W: $M_T(lepton, \vec{\rlap{/}E}_T) > 50$ GeV. The SM background composition for this baseline selection is 43% of diboson, 30% of tZq, 17% of $t\bar{t}$ and 10% of $t\bar{t}$ in association with a boson.

Nine variables are used in the BDT for the Zut-FCNC signal and eight for the Zct-FCNC signal. For both BDT cases, the transverse momentum of the reconstructed Z boson is the most discriminating one, followed by the invariant mass of the lepton + leading b jet system. Defining j_1 as the leading jet, b as the leading b-jet, Z as the reconstructed Z boson, t as the SM top and ℓ_{top} the lepton from the SM top decay, the remaining BDT variables are: $p_T(j_1)$, transverse mass of Z, $\Delta R(\ell_{top}, Z)$, $\Delta R(\ell_{top}, b)$, $\Delta R(t, Z)$ and $\Delta R(\vec{\rlap{/}E}_T, Z)$. The ninth variable used only for the Zut case is the relative lepton charge, defined as $Q(\ell_{top}) \equiv Q(\ell_{top}) \times |\eta(\ell_{top})|$.

Hqt-FCNC

- **H $\rightarrow$ b$\bar{\text{b}}$.** Exactly one isolated electron (or muon) is requested with $p_T > 30$ GeV and $|\eta| < 2.4$ (2.5). The criteria on relative isolation, $iso < 0.1$ (0.12), is defined for a cone of 0.4 around the lepton direction. A veto is applied to reject the event

if there is another isolated lepton with $p_T > 10$ GeV in the η acceptance ($|\eta| < 2.5$), with an isolation $iso < 0.2$ for a cone of 0.4 around the lepton direction. In order to study the sensitivity to the $t\bar{t}$ signal-component on the one hand and the single top component on the other hand, two mutually exclusive event categories are defined. The $t\bar{t}$ oriented analysis selects events with at least four jets, of which at least three pass the medium criterium of the b-tagging algorithm. The single top oriented analysis selects events with exactly three jets that are b-tagged according to the medium criterium. In both categories, selected jets should have $p_T > 40$ GeV, $|\eta| < 2.4$, a fraction of hadronic/electromagnetic energy larger than 0.15 and an angular separation with the lepton $\Delta R(jet, lepton) \geq 0.4$. It was found that selection criteria based on characteristics of the reconstructed Higgs boson, hadronically and leptonically decaying top quark did not improve the signal significance. Top quark pairs decaying semi-leptonically are the dominating background in both cases, representing 97% and 95% of the total amount of selected background events for the $t\bar{t}$ and single top oriented categories respectively.

An event reconstruction was considered in order to correctly identify the two b jets (b_H^1 and b_H^2) from the Higgs decay, the b-jet (b_t^{lep}) from the leptonic top quark decay and the light jet ($j_{u/c}$) from the anomalous top quark decay. The jet assignment was performed doing a modified χ^2-minimisation on the masses of the reconstructed objects, using as figure of merit

$$\chi^2_{\text{mod.}} = \frac{[m(b_H^1, b_H^2) - m_H]^2}{(16 \text{ GeV})^2} + \frac{[m(b_H^1, b_H^2, j_{u/c}) - m_{top}]^2}{(30 \text{ GeV})^2} + \frac{[m_T(b_t^{lep}, \vec{\not{E}}_T, lep) - m_{top}]^2}{(30 \text{ GeV})^2},$$

where m_H and m_{top} are respectively the mass of the Higgs boson and top quark, and m_T represents the reconstructed mass in the transverse plane. The denominators are a tuned[4] estimation of the mass resolution[5], in GeV, of the object represented in the numerator. Note that in the single top oriented analysis, the second term gets dropped. This modified χ^2 minimisation proved to be about 65% efficient at correctly associating b jets to the Higgs decay, but only about 22% efficient in picking up the correct $j_{u/c}$ and b_t^{lep}. The efficiency of correctly reconstructing the full process is less than 15%.

The selected objects from the event reconstruction are used to reconstruct kinematic properties of the event. Several of these reconstructed variables have been considered for an MVA, such as: $m(b_H^1, b_H^2)$, $m(b_H^1, b_H^2, j_{u/c})$, $p_T(b_t^{lep}, \vec{\not{E}}_T, \text{lep})$, the sum of the transverse momenta of all jets. No significant additional discrimination

---

[4]The numbers are tuned such that the importance of correctly reconstructing the Higgs boson is increased with respect to reconstructing the top quarks in the event. This choice is motivated by the fact that the presence of a Higgs boson in the signal process is the main difference with respect to the $t\bar{t}$ background process.

[5]The mass resolution is estimated from the spread with respect to the central value in dijet and trijet mass distributions in simulated events, such as in Figure 3.3.

power could be gained from an MVA, as demonstrated in Figure 4.2. Figure 4.3 illustrates the signal to background discrimination power for $m(b_H^1, b_H^2)$. However, no heavy flavour tagging discriminators were used as MVA inputs in this phenomenological study, as the framework does not provide such variables. As previous Hqt-FCNC searches in this final state topology from the ATLAS [10] and CMS [104] collaborations make use of b-tagging discriminators in an MVA, they are expected to benefit ATLAS and CMS analyses during the LHC Run 2 as well.
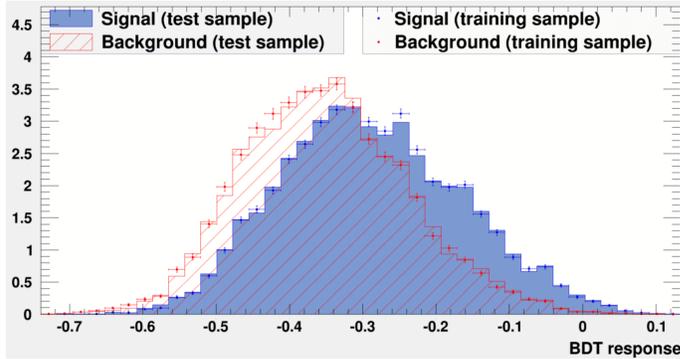
Figure 4.2: The normalized distributions of a BDT event discriminant for the $t\bar{t}$ oriented selection, demonstrating almost no discriminating power is gained from applying a BDT.
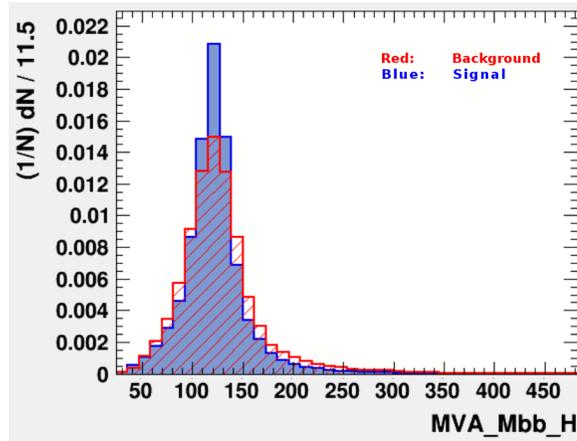
Figure 4.3: The normalized distribution of $m(b_H^1, b_H^2)$ for the $t\bar{t}$ oriented selection.

- **H $\rightarrow W^+W^-$.** Exactly two same sign isolated leptons are required, with $p_T >= 26(20)/15(11)$ GeV for leading and 2nd leading electron (muon) respectively and $|\eta| < 2.5$ (2.4). A veto is applied to reject the event if invariant mass m_{ll} of the 2 leptons is ≤ 12 GeV, and also if $|Z_{mass} - m_{ll}| \leq 15$ GeV to reject Z-boson

events. The criteria on relative isolation ($iso < 0.2$) is defined for a cone of 0.4 around the lepton direction. Only events with at least 4 jets are considered in this analysis, these jets should have $p_T > 50/30/20/20$ GeV for $1^{st}, 2^{nd}, 3^{rd}$ and 4th jet respectively and $|\eta| < 2.4$ and a fraction of hadronic/electromagnetic energy larger than 0.15 and an angular separation with the lepton $\Delta R(jet, lepton) > 0.4$. In particular at least one jet is requested to pass the loose criterium of the b-tagging algorithm. This event selection is optimized for the $t\bar{t}$ component of the signal. The SM background composition for this baseline selection is 74% of $t\bar{t}$, 14% of W + jets, 5% of diboson, 4% of Z + jets and 3% of $t\bar{t}H$.

The set of observables giving the best discriminating BDT output is formed by the following 7 observables: the mass of leading lepton + leading b-tagged jet, the total sum of transverse momenta of all jets, the transverse momenta p_T of the two leptons, spatial separation ΔR between the two leptons, ΔR between leading lepton and leading b-tagged jet and the number of jets.

- **H $\rightarrow \gamma\gamma$.** At least two photons with $p_T(\gamma_1) > 30$ GeV and $|\eta| < 2.5$ need to be present. The criteria on relative combined isolation of calorimeter energy and track momentum is defined for a cone of $\Delta R = 0.4$ around the photon momentum direction and is required to be less than 0.1. One additional lepton with $p_T > 20$ GeV and $|\eta| < 2.5$ is required, with an isolation $iso < 0.1$ for a cone of 0.4 around the lepton. Jets that overlap with the selected leptons or photons are removed for further consideration. At least 2 jets with $p_T > 20$ GeV and $|\eta| < 2.5$ and a fraction of hadronic/electromagnetic energy larger than > 0.15 are required of which one is b-tagged with the tight working point. Finally, the top mass is reconstructed with two photons and one jet taking into account all combinations of jets. The jet which gives the closest mass $M_{j\gamma\gamma}$ to the top quark mass is taken, accepting only events for which $163 < M_{j\gamma\gamma} < 183$ GeV. This event selection is optimized for the $t\bar{t}$ component of the signal. The SM background is dominated by $t\bar{t}$+jets events.

  No MVA is performed, as the invariant mass of the di-photon system shows great discriminatory power between signal and background events, as demonstrated in Figure 4.4 for the Hct-FCNC signal case.

The sensitivity to top-FCNC signals in pp-collisions at a center-of-mass energy of 13 TeV is projected towards a collected 100 fb^{-1} of collision data, a benchmark scenario that will be achieved during Run 2 of the LHC. The branching ratio of the considered anomalous top decay for which a signal significance of 2 is achieved, is used as a figure of merit for the sensitivity[6]. An overview of the sensitivity projections for each of the above discussed cases is given in Table 4.6. In this table branching ratios are shown

---

[6]A significance of 2 corresponds to a 2 σ deviation from the predicted background, where σ represents the analysis uncertainty. Such a deviation of 2 σ means that, under the considered background hypothesis, the signal prediction does not fall in the 95% probability central interval.
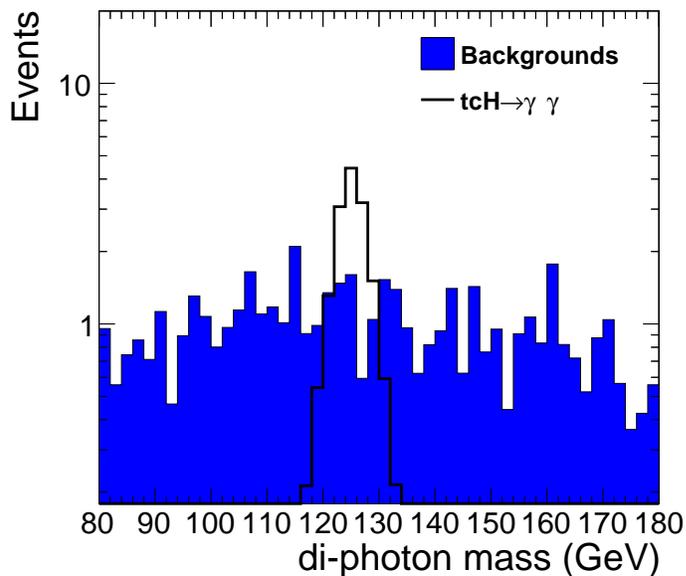
Figure 4.4: Invariant mass ditribution of the di-photon system for the Hct-FCNC signal case with H → γγ, for a projected data set of pp-collisions at 13 TeV with an integrated luminosity of 100 fb^{-1}.

for which a signal significance of 2 is achieved after the baseline selection in the second column and after a cut on the BDT discriminant. These sensitivity projections are shown neglecting any form of systematic uncertainty, since systematic experimental uncertainties can not be treated properly in a phenomenological study.

Though these projections need to be interpreted with a grain of salt due to the lack of systematic uncertainty treatments, they motivate the pursuance of equivalent analyses at CMS and ATLAS during LHC Run 2. In this thesis the Hqt-FCNC signal case is chosen to be sought for in actual pp-collision data recorded by the CMS detector. Some new-physics models, such as two-Higgs doublet models, predict t → qH branching ratios matching the projected sensitivity in this study. Of the considered Hqt-FCNC final state topologies, the H → b$\bar{\text{b}}$ decay channel is chosen. The two considered baseline event selections for this channel separately already provide a sensitivity that is competitive to the H → γγ and H → W^+W^- topologies. A combination of the two event categories increases the sensitivity even further, reaching $\mathcal{B}(t \to uH) = 1.2 \times 10^{-3}$ and $\mathcal{B}(t \to cH) = 1.4 \times 10^{-3}$ at a significance of 2. Assuming b-tagging discriminants developed by the CMS collaboration add considerable discrimination power in an MVA, the sensitivity to the H → b$\bar{\text{b}}$ decay channel is expected to increase in the actual analysis at CMS.

Table 4.6: Overview of the branching ratios $\mathcal{B}$ for the considered top-FCNC signal cases, separating the tqH-FCNC case with H $\to$ b$\bar{\text{b}}$ into two exclusive event categories. The branching ratios where a signal significance of 2 is achieved are shown here for a benchmark scenario for pp-collisions at a center-of-mass energy of 13 TeV and a projected 100 fb^{-1} of collision data. The second column shows the projection for a cut-and-count approach after the baseline event selection, whereas the third column quotes the projection for events surviving a cut on the final BDT discriminant.

Decay channel	$\mathcal{B}^{sig.=2}_{baseline}$	$\mathcal{B}^{sig.=2}_{BDT}$
gqt-FCNC	$\mathcal{B}(t \to ug) = 4.3 \times 10^{-6}$	$\mathcal{B}(t \to ug) = 2.9 \times 10^{-6}$
	$\mathcal{B}(t \to cg) = 1.2 \times 10^{-5}$	$\mathcal{B}(t \to cg) = 9.3 \times 10^{-6}$
Zqt-FCNC	$\mathcal{B}(t \to uZ) = 2.0 \times 10^{-4}$	$\mathcal{B}(t \to uZ) = 1.5 \times 10^{-5}$
	$\mathcal{B}(t \to cZ) = 4.9 \times 10^{-4}$	$\mathcal{B}(t \to cZ) = 4.9 \times 10^{-4}$
Hqt-FCNC$_{\geq 4j, \geq 3b}$: H $\to$ b$\bar{\text{b}}$	$\mathcal{B}(t \to uH) = 2.1 \times 10^{-3}$	N.A.
	$\mathcal{B}(t \to cH) = 1.9 \times 10^{-3}$	N.A.
Hqt-FCNC$_{=3j, =3b}$: H $\to$ b$\bar{\text{b}}$	$\mathcal{B}(t \to uH) = 1.6 \times 10^{-3}$	N.A.
	$\mathcal{B}(t \to cH) = 1.7 \times 10^{-3}$	N.A.
Hqt-FCNC: H $\to W^+W^-$	$\mathcal{B}(t \to uH) = 5.6 \times 10^{-3}$	$\mathcal{B}(t \to uH) = 1.4 \times 10^{-3}$
	$\mathcal{B}(t \to cH) = 5.6 \times 10^{-3}$	$\mathcal{B}(t \to cH) = 1.3 \times 10^{-3}$
Hqt-FCNC: H $\to \gamma\gamma$	$\mathcal{B}(t \to uH) = 4.6 \times 10^{-3}$	N.A.
	$\mathcal{B}(t \to cH) = 4.4 \times 10^{-3}$	N.A.

Chapter 5

Event selection criteria and kinematic reconstruction

The collision data recorded by the CMS collaboration is formed by a large variety of different physics processes. An account of the recorded data and simulated processes considered in this thesis is given in Section 5.1. The main bulk of those processes end up in event signatures that are hardly relevant in the search for the signal process, namely top-FCNC involving a Higgs boson decaying into a pair of b quarks. In order to substantially discard such background signatures, a series of selection steps are introduced. The baseline event selection applied in this top-FCNC search is described in Section 5.2.

The selected final state topology is experimentally represented by a collection of well defined physics objects, being charged leptons, jets and missing transverse energy. The origin of each of these objects however, is not always clear. One hard process particle may culminate into several objects in the final state, provoking ambiguities when reconstructing the hard process from final state objects. In order to resolve the combinatorics arising from these ambiguities, an event reconstruction algorithm is developed as discussed in Section 5.3. This event reconstruction aids in understanding the underlying physics process of a collision event and thereby in discriminating signal from background events.

5.1 Recorded data and simulated processes

In the course of 2016, the CMS collaboration collected collision data of protons colliding at 13 TeV with a total recorded integrated luminosity of 35.9 fb^{-1} $\pm$ 2.5% [105]. It is in this sample of collected collision data that the search for top-FCNC involving a Higgs boson decaying into a pair of b quarks is performed. Throughout the whole year of data taking, CMS has maintained a high performance, collecting about 92% of the luminosity delivered by the LHC, as demonstrated in Figure 5.1.
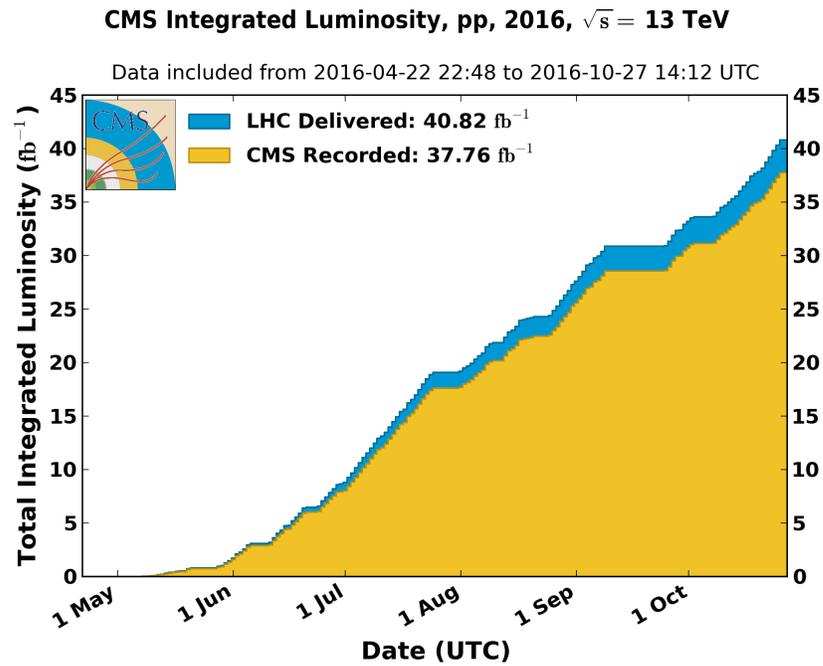
Figure 5.1: Recorded integrated luminosity evolving throughout 2016 as delivered by LHC (blue), and recorded by CMS (orange) for pp-collisions at 13 TeV center-of-mass energy [39]. Of the 37.6 fb^{-1} recorded integrated luminosity by CMS, only 35.9 fb^{-1} got certified for analysis.

5.1.1   Simulated Standard Model processes

A large list of SM processes may mimic the signal event signature, either due to actual topological similarities or flaws in object reconstructions. A short description of the simulation of collision processes has been provided in Chapter 3, but a specific overview of the simulated SM processes used in the analysis is given in Table 5.1. The dominant background contribution comes from the $t\bar{t}$+jets process. In this thesis $t\bar{t}$+jets events are split into three exclusive components according to the flavour of the additional jets. Two components are categorized as $t\bar{t}$ in association with a b quark pair ($t\bar{t} + b\bar{b}$) and $t\bar{t}$ in association with a pair of charm quarks ($t\bar{t} + c\bar{c}$). The remainder of the $t\bar{t}$+jets process is categorized as $t\bar{t}$ + light flavoured jets ($t\bar{t} + lf$). For convenience, all non-$t\bar{t}$ processes are merged and considered as one background labelled *other*.

   The production of QCD multijets is not considered in this analysis. It has been estimated from a data-driven estimate that the contribution of QCD multijet events is negligible for the considered baseline event selection. This estimate showed that the QCD multijet contribution was less than 1% in single muon events, and less than 3% in single electron events. Those estimates are in accordance with their uncertainty and are therefore ignored in this analysis.

5.1.2   Simulation of top-FCNC processes involving a Higgs decay to b quarks

The simulation of the signal processes at generator level is performed as summarised in Section 4.1. The cross sections for the single top and $t\bar{t}$ components of the signal, respectively referred to as *ST Hqt* and *TT Hqt*, are summarised in Table 5.2 for FCNC coupling strengths set to unity ($\kappa_{Hqt} = 1$). These cross sections take into account the Higgs $\rightarrow$ b$\bar{b}$ decay branching ratio $\mathcal{B}_{H\rightarrow b\bar{b}} = 0.577$ and SM leptonic decay (electron, muon or τ) of (one of) the top quark(s) as $\mathcal{B}_{t\rightarrow bW\rightarrow b\ell\nu} = 1/3$. The Feynman diagrams of the respective components have been illustrated in Figure 1.4.

   As more up quarks than charm quarks are present inside a proton, the anomalous single top production cross section via the Hut vertex is significantly higher than via the Hct vertex. A similar mechanism is responsible for the difference between the tH production versus $\bar{t}H$, as more particles than anti-particles compose a proton. The anomalous decay of top quarks in the $t\bar{t}$ component of the signal however, is indifferent to the Hut or Hct vertex and, as both the top and anti-top decay anomalously with an equal rate, no difference in cross section exists between the up (charm) or anti-up (anti-charm) production in the decay. These facts are illustrated in Figure 5.2, where the quadratic dependence of the signal cross section on the coupling strength is demonstrated.

Table 5.1: Summary of simulated SM processes considered in this thesis. The respective generators of each process are given and the corresponding theoretical cross section at a center-of-mass of 13 TeV, indicating to what order they are calculated. Cross sections for electroweak processes have been calculated by programs such as `MCFM 6.6` [106] and `FEWZ` [107]. The $t\bar{t}$+jets process cross section has been calculated by `Top++v2.0` [108]. Single top cross sections have been calculated by `Hathor v2.1` [109, 110]. The quoted uncertainties on $t\bar{t}$+jets and single top cross sections include uncertainties from: renormalisation and factorisation scale, pdf + α_s and top mass.

process	Generator	cross-section [pb]
$t\bar{t}$+jets	POWHEG	$832^{+5.6\%}_{-6.1\%}$ (NNLO)
Single top		
tW-channel (t)	POWHEG	$36 \pm 2.5\%$ (NNLO)
tW-channel ($\bar{t}$)	POWHEG	$36 \pm 2.5\%$ (NNLO)
t-channel (t)	POWHEG	$136^{+4.0\%}_{-3.4\%}$ (NLO)
t-channel ($\bar{t}$)	POWHEG	$81^{+5.0\%}_{-4.5\%}$ (NLO)
s-channel (leptonic W-decay)	MadGraph/MadEvent	$3.4 \pm 3.6\%$ (NLO)
W$\to \ell\nu$		
W + 1 jet	MadGraph/MadEvent	9.82×10^3 (NLO)
W + 2 jets	MadGraph/MadEvent	3.20×10^3 (NLO)
W + 3 jets	MadGraph/MadEvent	944 (NLO)
W + 4 jets	MadGraph/MadEvent	494 (NLO)
Drell-Yan $\to \ell\ell$		
$m_{\ell\ell} > 50$ GeV	MADGRAPH5_aMC@NLO	$5.77 \times 10^3 \pm 1.7\%$ (NNLO)
10 GeV $< m_{\ell\ell} < 50$ GeV	MadGraph/MadEvent	18.6×10^3 (NLO)
$t\bar{t}$+g	MADGRAPH5_aMC@NLO	3.7 (NLO)
$t\bar{t}$+W	MadGraph/MadEvent	0.7 (NLO)
$t\bar{t}$+Z	MadGraph/MadEvent	0.8 (NLO)
$t\bar{t}$+H($\to$ b$\bar{b}$)	POWHEG	0.3 (NLO)
WW	POWHEG	50 (NNLO)
ZZ	MADGRAPH5_aMC@NLO	3.2 (NLO)
WH($\to$ b$\bar{b}$)	MADGRAPH5_aMC@NLO	0.3 (NLO)
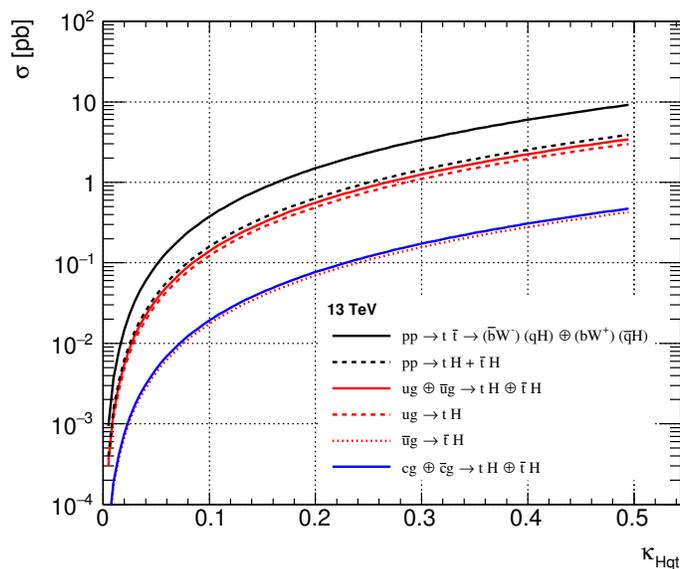
Figure 5.2: Cross section as a function of FCNC coupling strength for the relevant components of the top-FCNC signal processes. The cross section of the $t\bar{t}$ component matches for the Hut and Hct vertex and is represented by the solid black line. The dashed black line represents the summed cross sections of the single top components of the Hut and Hct vertices. The single top component of the Hut vertex is depicted by the solid red line, broken down in its two components represented by the dashed and dotted red lines. The single top component of the Hct vertex is represented by the solid blue line.

Table 5.2: Cross sections for the single top and $t\bar{t}$ components of top-FCNC processes for coupling strength equal to unity ($\kappa_{Hqt} = 1$, $q = u, c$). Branching ratios for $W \to \ell\nu$ and $H \to b\bar{b}$ are included. In case of the $t\bar{t}$ component, only one of the top-quarks is forced to decay via FCNC vertex.

Process	Coupling	Cross section [pb]
tH $\oplus$ $\bar{t}$H	κ_{Hut}	14
tH $\oplus$ $\bar{t}$H	κ_{Hct}	1.9
t$\bar{t}$	κ_{Hqt}	37

5.2   Baseline event selection

The baseline event selection aims to substantially reject SM background events, whilst maintaining a high signal selection efficiency. An initial optimisation of event selection criteria has been performed in the phenomenological study (see Chapter 4), from which it was concluded a final state with exactly one lepton (electron or muon) and at least three b-tagged jets provides a solid selection baseline. The kinematic constraints on the selected physics objects were set rather loosely, as the topological requirements proved to be sufficient to efficiently reject most background processes except for the semi-leptonic decay of top quark pairs. For analysing the collision data recorded by the CMS experiment in 2016, a similar baseline is defined by requiring each event to:

1. Survive event cleaning filters;

2. Have at least one well reconstructed primary vertex;

3. Pass single lepton high-level triggers;

4. Contain exactly one good lepton (muon or electron);

5. Veto on extra loose leptons (veto-muon and veto-electron);

6. Have at least three jets;

7. Have at least two jets passing the CSVv2 medium b-tag requirement.

Steps 1 to 3 are briefly summarised in Section 5.2.1. The lepton selection criteria (steps 4 and 5) and jet requirements (steps 6 and 7) are detailed in respectively Sections 5.2.2 and 5.2.3.

Even though simulated processes provide a good description of the actual collision data, factors that depend on specific running conditions of the LHC or detector calibrations are not (and can not be) included in the simulation process. Because these factors depend on the recorded data, they need to be corrected for in simulations at the analysis-level. The corrections are implemented in the form of event scale factors, which reweight each simulated event in order to match the actual recorded conditions.

The event scale factors considered in this search for top-FCNC are discussed in Section 5.2.4.

5.2.1 Trigger and event cleaning

The main bulk of collision events from uninteresting processes are mostly filtered out by the CMS trigger system. Still, a large variety of detector signatures are triggered, of which most do not overlap with the final state of interest. Therefore, dedicated trigger paths are defined in the trigger system, each one designed to single out events with specific detector signatures. In this thesis three single-lepton trigger paths have been used, as given in Table 5.3. Two muon trigger paths are used to select events with at least one global or tracker muon with $p_T > 24$ GeV. The electron trigger path selects events with at least one electron with a $p_T > 32$ GeV and $|\eta| < 2.1$.

Table 5.3: Trigger paths for the single lepton triggers used in data and simulation at the HLT.

Lepton selection	Trigger path
Single muon	HLT_IsoMu24 or HLT_IsoTkMu24
Single electron	HLT_Ele32_eta2p1_WPTight_Gsf

Some events may arise from instrumental noise and beam backgrounds. Such non-collision events are omitted in the selection by applying filters [111]:

- **Beam halo filter.** Machine induced interactions may produce particles flying along with the beam. Muons produced in these interactions compose a halo around the beam, which have a non negligible probability to interact in the calorimeters.

- **HBHE noise filters.** The hadronic calorimeter subsystems HB and HE (see Section 2.2.3) are known to sporadically record noise at a fixed rate, independent of beam conditions. The geometrical patterns and pulse shape information of the known noise are used in various algorithms to eliminate events with such detector signals.

- **Bad muon filter.** Some events contain muons that are not reconstructed as PF muons. These muons can be misinterpreted as charged hadrons, which in turn skews the missing transverse energy calculation. By looking for a compatibility between non-PF muon candidates and PF charged hadrons, these events can be filtered out.

On top of the event filters, the presence of at least one well reconstructed primary vertex is required. Its longitudinal distance to the beam spot d_z is required to be maximally 24 cm and its transverse distance to the beam spot d_{xy} to be maximally 2 cm.

5.2.2 Lepton selection criteria

Events with exactly one good electron or muon and a veto on additional loose electrons or muons get selected. One of the main concerns in defining the good lepton selection criteria is its ability to only select a so-called *prompt* lepton, i.e. a lepton that is produced in the hard process such as in the leptonic decay of a W boson. Leptons coming from fragmentation processes, such as electrons or muons from B meson decays, can mimic prompt leptons. By defining tight lepton selection criteria however, background processes such as multijet production where non-prompt leptons are produced in abundance, are drastically reduced. A further reduction is achieved by vetoing on additional loose leptons. Below, an account of the selection and veto criteria is given for the muon and electron objects, based on recommendations from [112] and [113] respectively.

Muon criteria

The selected muon has to be reconstructed as a global PF muon, with a normalized[1] χ^2 of its track fit smaller than 10. The tracker trajectory of the muon should have a transverse impact parameter $d_{xy} < 2$ mm and a longitudinal distance $d_z < 5$ mm with respect to the primary vertex. At least one muon-chamber hit should be included in the global-muon track fit and muon segments should be present in at least two muon stations. Inside the tracker system, at least one pixel hit and a minimum of five tracker layers with hits should be recorded. The muon is required to be isolated from energy contributions of photons, pileup, charged and neutral hadrons. The combined relative isolation, also known as the $\Delta\beta$-corrected relative isolation, is defined as

$$I_{rel}^{\Delta\beta} = \frac{\sum p_T^{\text{charged hadron}} + \max\left(0, \sum p_T^{\text{neutral hadron}} + \sum p_T^{\text{photon}} - 0.5 \sum p_T^{\text{pileup}}\right)}{p_T(\mu)}, \quad (5.1)$$

where the respective sums run over all corresponding PF particles (i.e. charged hadrons, neutral hadrons and photons) in a cone of $\Delta R = 0.4$ around the muon direction. Muons need only be isolated from particles of the primary collision event, hence pileup effects in the cone of $\Delta R = 0.4$ around the muon direction are subtracted. Since charged hadron subtraction (see Section 3.3.7) takes care of the charged pileup contributions, the neutral pileup contributions need to be subtracted explicitly. The neutral pileup contribution is estimated as half the charged pileup contribution [84], hence the factor of 0.5 in front of $\sum p_T^{\text{pileup}}$. The $I_{rel}^{\Delta\beta}$ is required to be smaller than 0.15 for the selected muon. Furthermore, the muon should be restricted to $|\eta| < 2.1$ and have a $p_T > 30$ GeV. The p_T theshold is set above the threshold defined by the HLT to cover for the fact that a muon at the HLT is defined slightly different.

---

[1]The normalized χ^2 is the χ^2 divided by the number of degrees of freedom, or $\chi^2/ndof$.

Additional muons are vetoed if they are PF muons that are reconstructed as a global or tracker muon, with $I_{rel}^{\Delta\beta} < 0.25$. They are restricted to $|\eta| < 2.4$ and have a $p_T > 10$ GeV.

Electron criteria

The CMS collaboration defines three working points for the identification of electrons, conveniently called the Tight, Medium and Loose working point. For each of the working points a cut-based approach is defined on a set of variables, which are found to be discrminatory between real electrons and hadrons mimicking electrons. In this thesis, the Tight working point is used for the electron selection, which has a 70% efficiency of selecting real electrons, whilst rejecting over 99% of non-prompt electrons. The cuts applied for the Tight working point are summarised in Table 5.4. The values differ for electrons of which the super-cluster (see Section 3.3.3) lies in the barrel and endcap. The $\sigma_{i\eta i\eta}$ is a representation of the width of the electron shower in the η direction, calculated from an array of 5×5 ECAL crystals around the energy deposit in the ECAL. The azimuthal, $\Delta\phi$, and pseudorapidity, $\Delta\eta$, separations between the super-cluster and GSf track are considered as well. One of the most discriminating variables is the ratio between the energy deposits in the HCAL to the ECAL. Finally, an isolation requirement is put as well on the electron. The isolation for electrons is defined similar to Equation (5.1) as

$$I_{rel}^{EA} = \frac{\sum p_T^{\text{charged hadron}} + \max\left(0, \sum p_T^{\text{neutral hadron}} + \sum p_T^{\text{photon}} - \rho A_{eff}\right)}{p_T(e)}, \qquad (5.2)$$

for PF particles in a cone of $\Delta R = 0.3$ around the electron. In this equation the neutral pileup contribution has been subtracted using effective areas A_{eff} and the transverse-momentum density ρ of the event. The effective area [114] is an estimate of the area in the isolation cone that does not originate from the electron footprint. This type of pileup subtraction is similar to the one applied in the L1 part of jet energy corrections (see Section 3.3.4).

Table 5.4: Requirements for the selected electron according to the Tight working point of the cut-based electron identification.

Variable	Barrel	Endcap
$\sigma_{i\eta i\eta} <$	0.00998	0.0292
$|\Delta\eta(\text{super-cluster,GSf-track})| <$	0.00308	0.00605
$|\Delta\phi(\text{super-cluster,GSf-track})| <$	0.0816	0.0394
$\frac{Hcal}{Ecal} <$	0.0414	0.0641
$I_{rel}^{EA} <$	0.0588	0.0571

On top of the Tight electron working point cuts, the following criteria are applied as well to the selected electron:

- The most inner expected hit of the reconstructed GSf track is not missing;

- The electron does not fall in the EB-EE gap ($1.4442 < |\eta| < 1.5660$);

- $d_z < 0.10$ (0.20) cm in the barrel (endcaps);

- $d_{xy} < 0.05$ (0.10) cm in the barrel (endcaps);

- $|\eta| < 2.1$;

- $p_T > 35$ GeV.

The p_T threshold is raised above the HLT threshold to avoid possible bias in the analysis selection, as the electron definition at the HLT slightly differs from the analysis level definition.

The vetoed additional electrons are defined according to the Loose working point for cut-based electron identification, as summarised in Table 5.5. Furthermore, their transverse momenta are required to be larger than 10 GeV and $|\eta| < 2.5$.

Table 5.5: Requirements for the vetoed electrons according to the Loose working point of the cut-based electron identification.

Variable	Barrel	Endcap
$\sigma_{i\eta i\eta} <$	0.011	0.0314
$\|\Delta\eta(\text{super-cluster,GSf-track})\| <$	0.00477	0.00868
$\|\Delta\phi(\text{super-cluster,GSf-track})\| <$	0.222	0.213
$\frac{Hcal}{Ecal} <$	0.298	0.101
$I_{rel}^{EA} <$	0.0994	0.107

5.2.3   Jet selection criteria

From the collection of PF jets in each event, at least three jets are selected [115]. All selected jets should be composed of at least two PF particles, of which at least one should be a charged particle. Neither neutral hadrons, nor photons or electrons, can represent more than 99% of the total jet energy, whereas a minimal fraction of the jet energy should be represented by charged hadrons. Each jet should be spatially separated from the selected lepton by requiring $\Delta R(\text{jet,lepton}) > 0.4$. Finally, the jets are required to have $p_T > 30$ GeV and $|\eta| < 2.4$.

The identification of b flavour jets is performed using the CSVv2 algorithm. At least two jets are required to pass the medium working point, defined as 0.8484, of the CSVv2 discriminator.

5.2.4 Corrections to simulations

In the analysis, simulated samples are weighted such that their normalisation corresponds to the total recorded integrated luminosity of 35.9 fb^{-1}. This normalisation is ensured by reweighting each event as

$$SF_{norm} = \frac{\mathcal{L}_{int} \times \sigma_{process} \times \mathcal{A}}{N}, \tag{5.3}$$

where N represents the number of events in the simulated sample, $\sigma_{process}$ the cross section of the process and $\mathcal{A}$ the acceptance, which represents the allowed phase space for the process by considering the geometrical and kinematic restrictions. Samples that are generated at NLO with `MADGRAPH5_aMC@NLO` are corrected for negative event weights (see Section 3.2) by applying an additional normalisation factor that accounts for the difference in number of events with positive event weights N_+ versus negative event weights N_- as

$$SF_{\texttt{MADGRAPH5\_aMC@NLO}} = \pm \frac{N_+ + N_-}{N_+ - N_-}, \tag{5.4}$$

where the sign of this factor corresponds to the event's positive or negative weight. Besides the normalisation event scale factor, the following correction factors are considered as well:

- Pileup reweighting;

- Lepton scale factors;

- CSVv2 shape correction;

- Jet energy corrections.

The last item in this list does not correspond to an event scale factor, but to an alteration of the jet energy to match the calibrated jet energy scale and resolution as discussed in Section 3.3.4.

Pileup reweighting

The number of simulated pileup interactions differs from the actual number of pileup interactions, as a simple distribution of number of pileup interactions is assumed. Such a simple distribution has to be assumed in simulations due to the fact that the number of actual pileup interactions depends on the instantaneous luminosity for each bunch crossing, which vary slightly whenever a new round of pp-collisions is started up at the LHC. The instantaneous luminosity for each so-called *lumi-section* of the recorded data is measured and by multiplying it with the total inelastic cross section of pp-collisions at 13 TeV, the distribution of number of pileup interactions is determined. The preliminary measured total inelastic cross section is 69.2 mb $\pm$ 4.6% [116]. A

normalized distribution comparison for the number of pileup interactions between simulation and data is given in Figure 5.3. The simulated profile for number of pileup interactions overestimates the high multiplicity of pileup interactions. Consequently simulation events with more than 35 pileup interactions are consistently downweighted up to the point that simulated events with over 45 pileup interactions get a negligible contribution.
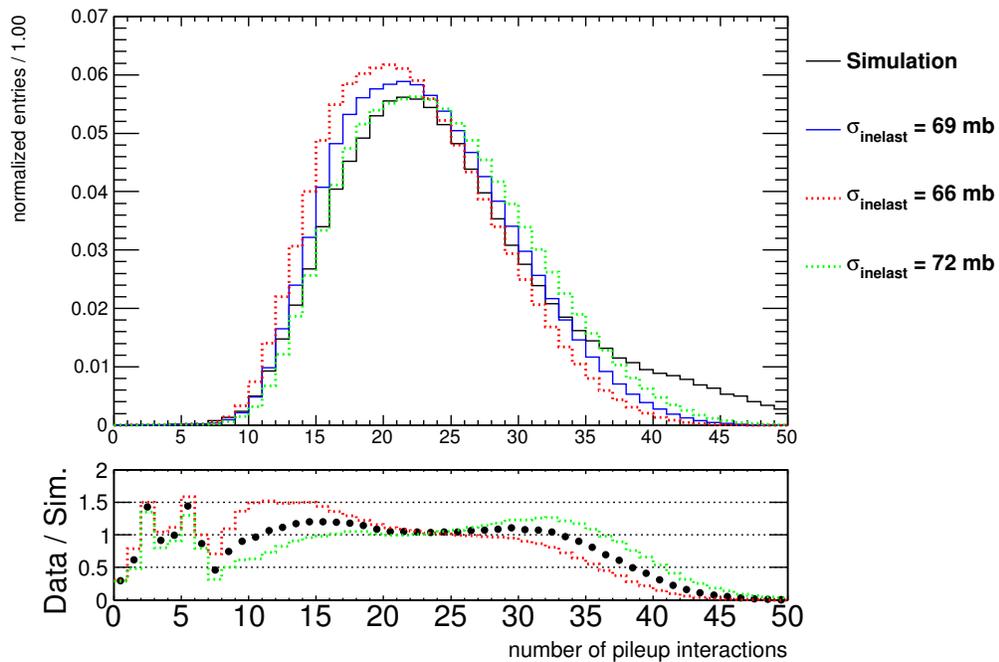
Figure 5.3: Normalized distributions for the number of pileup interactions in simulated samples and as measured in data for the nominal inelastic cross section. The dashed lines represent systematic up- and down variations according to the inelastic cross section uncertainty of 4.6%. The ratio between the normalized data and simulation corresponds to the event scale factors.

The effects of reweighting simulated events according to their number of pileup interactions manifest themselves mainly in the distribution of number of reconstructed primary vertices, as shown in Figure 5.4. Even though pileup reweighting improves the agreement between data and simulations, an obvious mismatch remains. The mismatch is due to the fact that during certain periods of data taking, the CMS tracker was less efficient at recording hits. Such tracking inefficiencies that only affect a subset of the recorded data can not be properly corrected for in simulation. It has been investigated that this mismatch does not affect the analysis in any way.
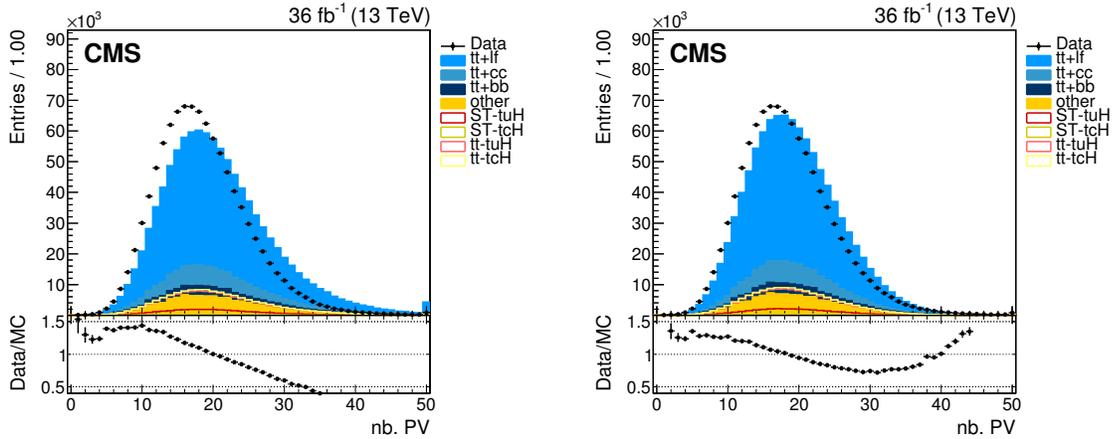
Figure 5.4: The distribution of number of reconstructed primary vertices for events in the baseline selection before (left) and after (right) pileup reweighting.

Lepton scale factors

Several factors determine the selection of electrons and muons in the analysis and each one of those factors have a potentially different effect on actually detected leptons and simulated leptons. Four such factors are found to behave differently under simulated and real detection conditions: tracking-, object reconstruction-, isolation- and trigger efficiencies. The tracking and object reconstruction efficiencies are hard to disentangle and are therefore combined into one factor. This defines a total event scale factor due to the lepton selection criteria as

$$SF_\ell = SF(\text{track \& reco}) \times SF(\text{iso}) \times SF(\text{trig}), \tag{5.5}$$

where each component is determined as

$$SF = \frac{\epsilon(\text{data})}{\epsilon(\text{simulation})}, \tag{5.6}$$

with ϵ representing the efficiency in data and simulated events. The efficiencies in data are measured based on a tag-and-probe method with leptonically decaying Z bosons. The scale factors are determined for the electron and muon objects separately and are dependent on the p_T and η of the object, as respectively provided in [117, 118]. Scale factors for muons in this analysis vary between 0.96 and 0.99, whereas scale factors for electrons lie in the range of [0.85;1], with an average around 0.95.

The application of lepton scale factors has a small effect on the normalisation of the simulated processes up to the percent level, whereas it hardly affects the shapes of p_T and η distributions of muons and electrons as demonstrated in Figures 5.5 and 5.6.
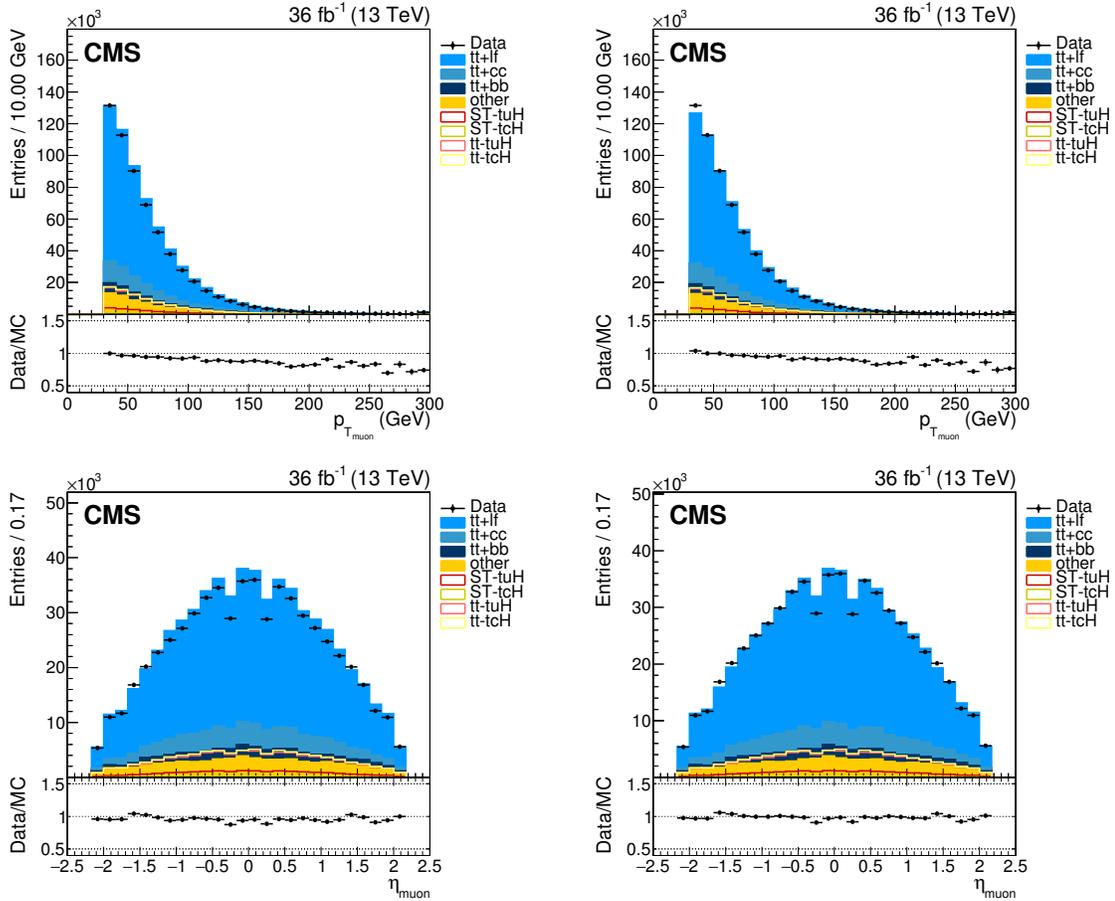
Figure 5.5: Muon p_T (top plots) and η (bottom plots) distributions for single muon events in the baseline event selection before (left) and after (right) applying muon scale factors.
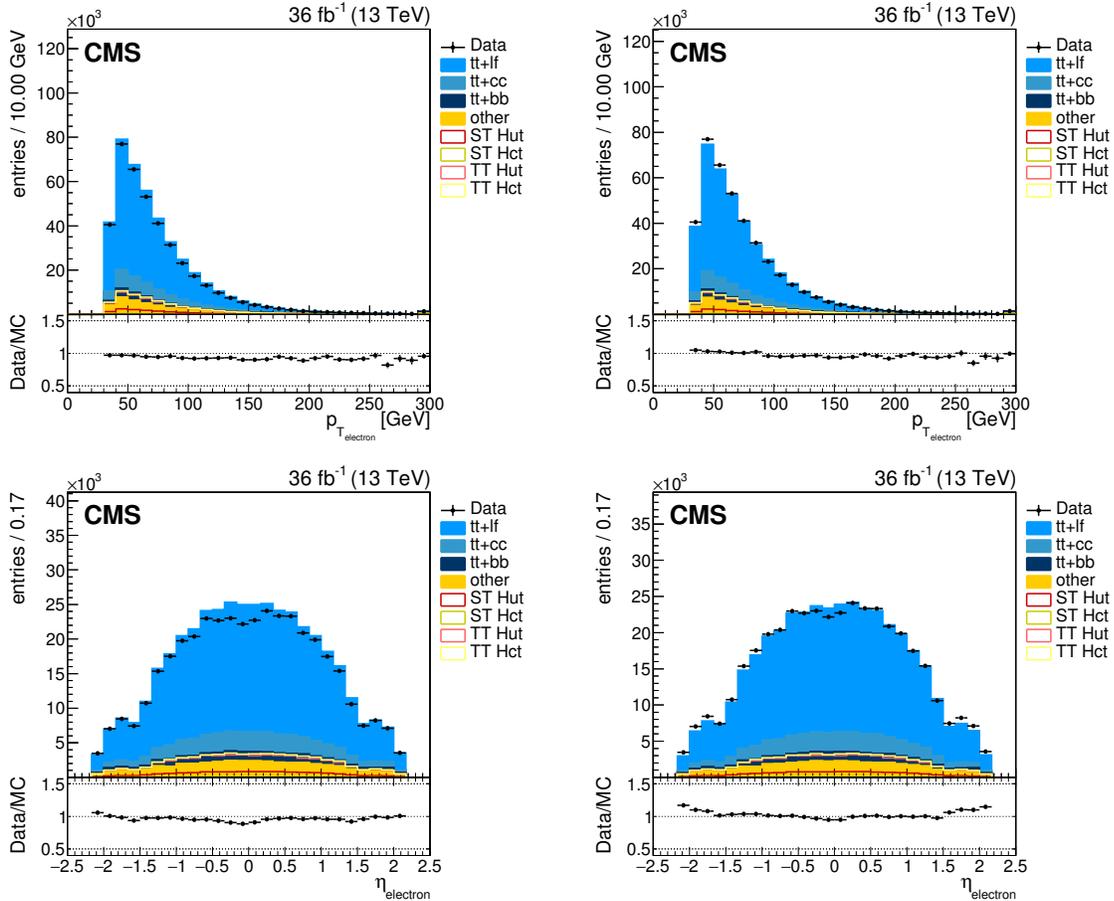
Figure 5.6: Electron p_T (top plots) and η (bottom plots) distributions for single electron events in the baseline event selection before (left) and after (right) applying electron scale factors.

CSVv2 shape correction

The identification of jets originating from bottom quarks is a crucial factor in this analysis, as the final state topology of the signal contains three b quarks. In order to correct for known discrepancies between data and simulation in the shape of the CSVv2 discriminant, event scale factors as introduced in Section 3.3.5 need to be applied. As it was concluded in the phenomenological study that the use of the b-tagging discriminant of jets might provide additional separation power in an MVA, it is necessary to correct the CSVv2 discriminant shape on a jet-by-jet basis. The individual scale factors per jet are calculated as given in Equation (3.10) for the three different jet-flavour conditions. The total event scale factor is obtained by multiplying the individual jet scale factors as

$$SF_{CSVv2} = \prod_{i}^{N_{\text{b flavour}}} SF_i^{\text{b flavour}} \prod_{j}^{N_{\text{light flavour}}} SF_j^{\text{light flavour}} \prod_{k}^{N_{\text{c flavour}}} SF_k^{\text{c flavour}}. \qquad (5.7)$$

Besides the shape correction of the CSVv2 discriminant of the jets, the CSVv2 shape correction has a large impact on the number of selected events. The effect of the CSVv2 shape correction on the three leading jets is demonstrated in Figure 5.7.

Control plots

It is vital to the analysis to ensure simulations describe the data well after the baseline event selection, especially for event properties that play an important role in the analysis. From the phenomenological study (see Chapter 4) the notion of classifying the event according to number of jets and b-tagged jets came forward. A proper agreement between data and simulations for these variables is achieved, as well as for the missing transverse energy, as shown in Figure 5.8. The slight disagreement at high jet multiplicity arises from the simulation of $t\bar{t}$ events with POWHEG. This is a known feature of the POWHEG simulation of $t\bar{t}$ events[2] [77], caused by a slightly mistuned value for α_s. Since the disagreement is very small and only affects the high jet multiplicity, no corrections are implemented. An excellent agreement is achieved for the kinematic properties of the three leading jets, as demonstrated in Figure 5.9 for their p_T and η distributions. Only small mismatches arise towards the high-p_T regimes, a trend which is also present for high missing transverse energies and lepton-p_T.

---

[2]Other event generators for the $t\bar{t}$ process have been considered, such as MADGRAPH5_aMC@NLO, but show worse agreement in overall normalisation and number of b-tagged jets.
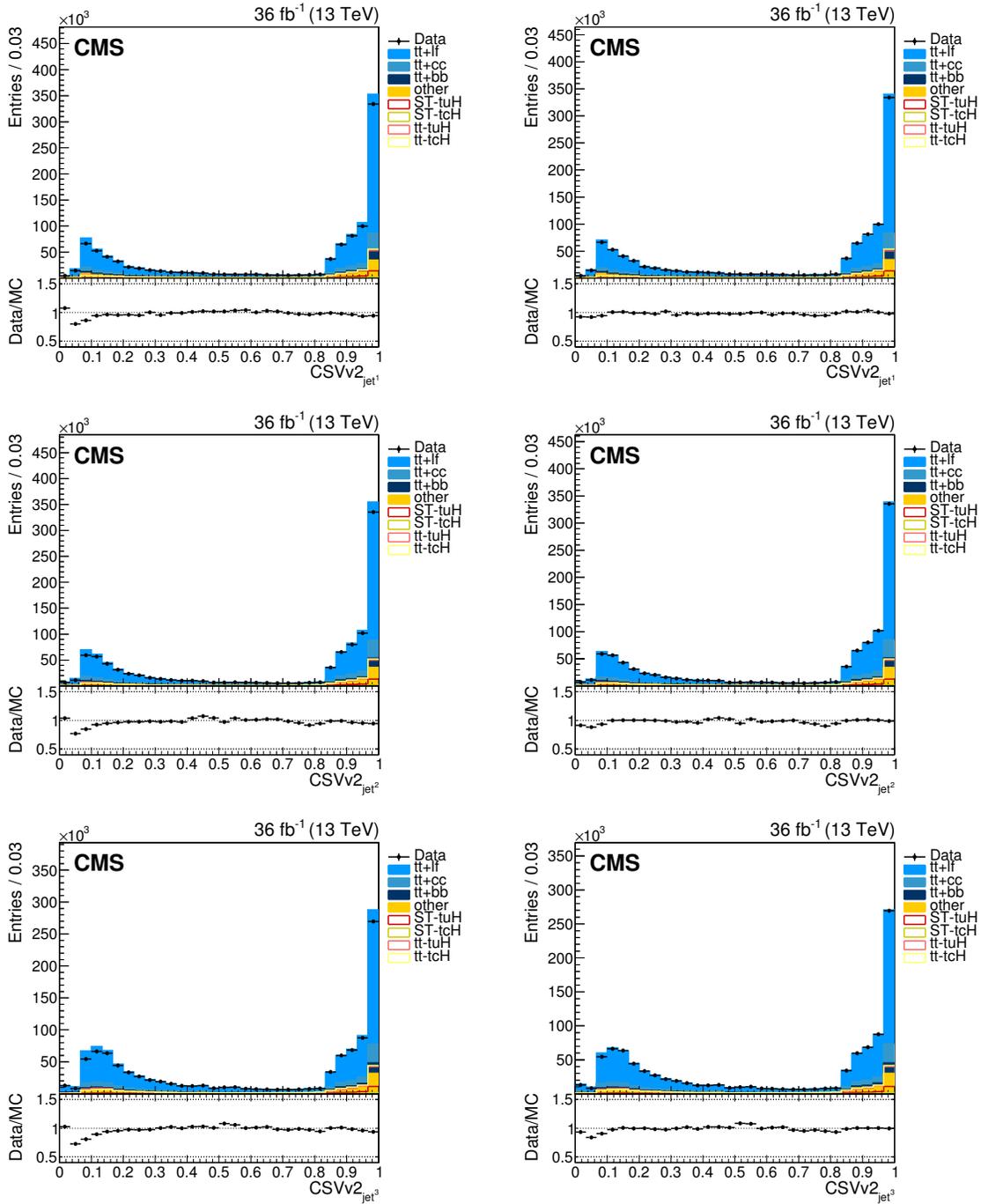
Figure 5.7: The CSVv2 discriminant of the leading (top), sub-leading (middle) and third leading (bottom) jet before (left) and after (right) applying the CSVv2 shape correction to the baseline event selection.
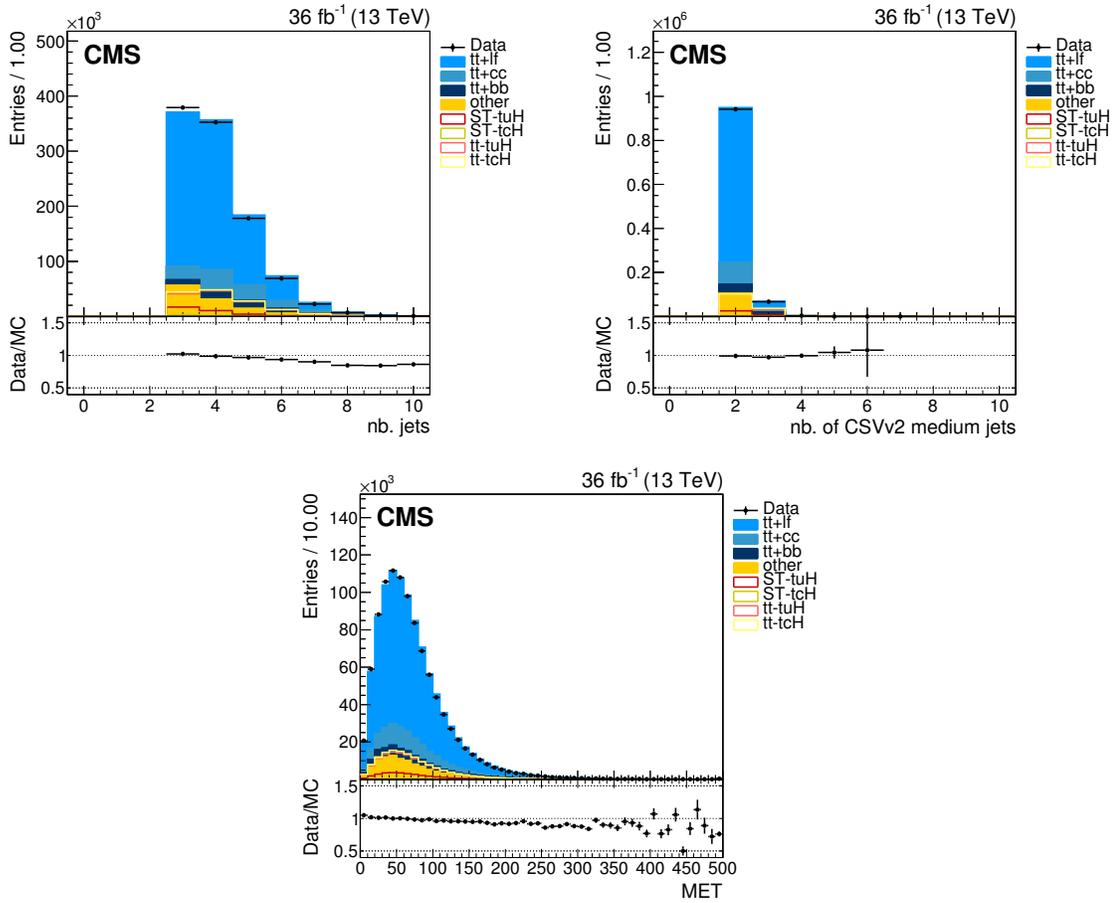
Figure 5.8: Control plots for the number of jets (upper left), number of b-tagged jets according to the medium working point of CSVv2 (upper right) and the missing transverse energy (bottom) for events in the baseline selection.
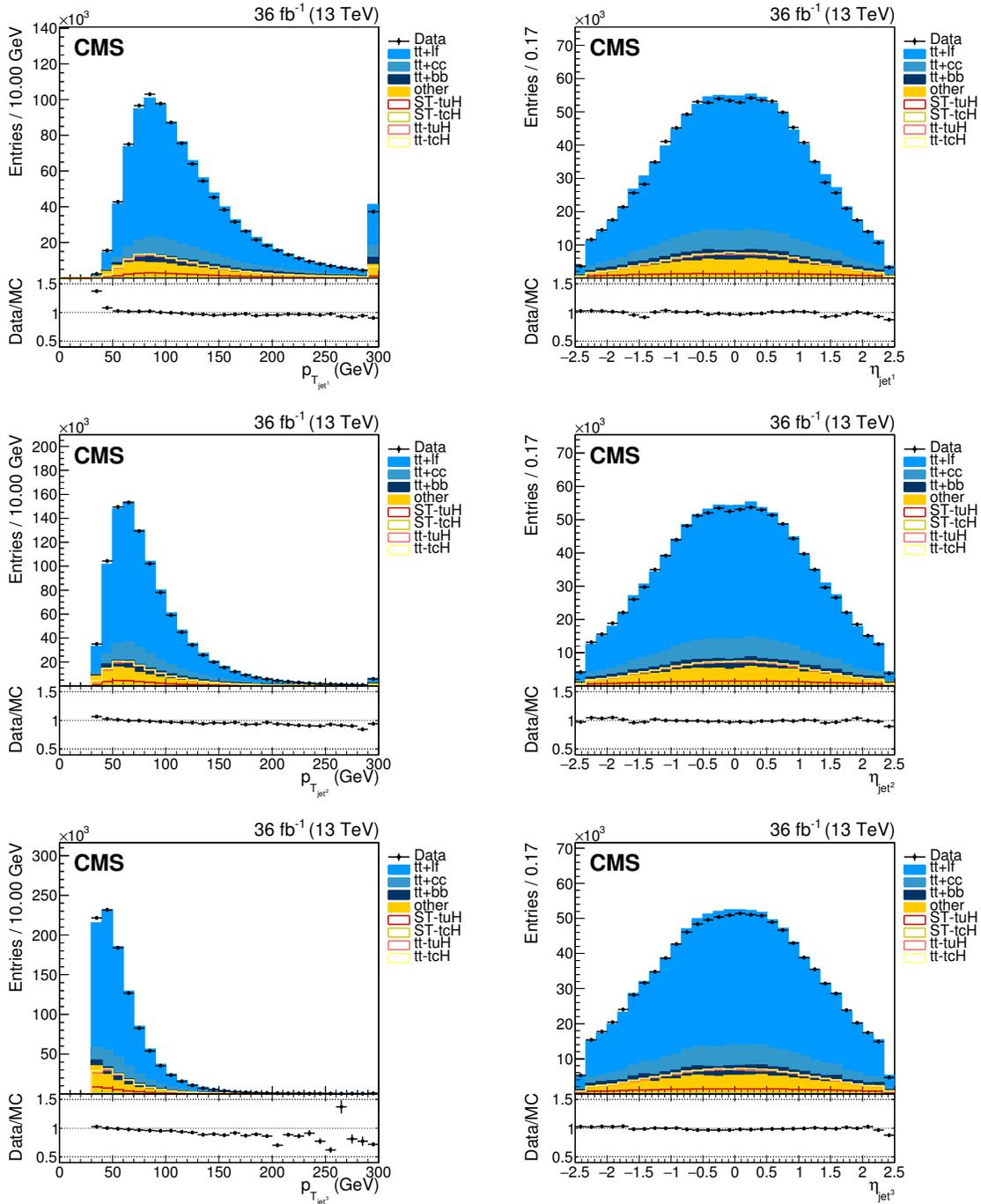
Figure 5.9: The distributions of transverse momentum (left) and pseudorapidity (right) for the leading (top), sub-leading (middle) and third leading (bottom) jet in the baseline event selection.

5.3   Event recontruction

The SM background is composed for about 90% of semi-leptonic $t\bar{t}$ events, which proves to be an irreducible background for the ST-tqH and TT-tqH signal components. The only difference between the signal and semi-leptonic $t\bar{t}$ event topology is the presence of two jets that originate from a Higgs boson decay in signal events compared to a W boson decay in $t\bar{t}$ events. Correctly ascertaining those two jets in the jet collection is of utter importance in the prospect of discriminating signal from background events. Rudimentary approaches at identifying those jets, such as a χ^2-minimization of the di-jet invariant mass with respect to the Higgs boson mass, have been attempted in the phenomenological study (Chapter 4). However, these proved to be less than 20% effective of correctly reconstructing the full hard process (with a 65% efficiency of picking up the correct b jets from the Higgs decay). A more convoluted method is used in this analysis, by means of kinematically fitting the reconstructed objects to the hard process objects as explained in Section 5.3.1. The kinematically reconstructed objects are thereupon used in an MVA to solve the jet combinatorics as detailed in Section 5.3.2.

5.3.1   Reconstruction of event kinematics

Three event signatures are of significant importance to the analysis and are considered as separate hypotheses for a full kinematic event reconstruction, labelled as:

1. TOPTOPLEPHBB: $t\bar{t}$ component of the signal;

2. TOPHLEPBB: single top component of the signal;

3. TOPTOPLEPHAD: $t\bar{t}$ background.

The TOPTOPLEPHBB hypothesis consists of one leptonically decaying W boson coming from a SM top quark decay, respectively labelled as W^{lep} and t^{lep}, and a Higgs boson coming from an anomalous top decay, respectively labelled as H and t^{had}. Defining the b quark from the SM top decay as b_1, the b quarks from the H boson decay as b_2 and b_3, the up or charm quark from the anomalous top decay as q, the lepton and neutrino from W^{lep} as respectively ℓ and ν, the kinematic constraints[3] for the hard process are constructed from the process $t^{lep}t^{had} \rightarrow (W^{lep}b_1)(Hq) \rightarrow ((\ell\nu)b_1)((b_2b_3)q)$ as

---

[3]The kinematic constraints are constructed by requiring the final state objects to satisfy the invariant mass constraints imposed by their mother particles.

$$\text{TOPTOPLEPHBB} : \begin{cases} m_{W^{lep}}^2 = (E(\ell) + E(\nu))^2 - (p_x(\ell) + p_x(\nu))^2 - \\ \qquad (p_y(\ell) + p_y(\nu))^2 - (p_z(\ell) + p_z(\nu))^2, \\ m_H^2 = (E(b_2) + E(b_3))^2 - (p_x(b_2) + p_x(b_3))^2 - \\ \qquad (p_y(b_2) + p_y(b_3))^2 - (p_z(b_2) + p_z(b_3))^2, \\ m_{t^{lep}}^2 = (E(\ell) + E(\nu) + E(b_1))^2 - (p_x(\ell) + p_x(\nu) + p_x(b_1))^2 - \\ \qquad (p_y(\ell) + p_y(\nu) + p_y(b_1))^2 - (p_z(\ell) + p_z(\nu) + p_z(b_1))^2, \\ m_{t^{had}}^2 = (E(q) + E(b_2) + E(b_3))^2 - (p_x(q) + p_x(b_2) + p_x(b_3))^2 - \\ \qquad (p_y(q) + p_y(b_2) + p_y(b_3))^2 - (p_z(q) + p_z(b_2) + p_z(b_3))^2, \end{cases} \quad (5.8)$$

where m and E represent the rest mass and energy of the particle and (p_x, p_y, p_z) its momentum.

Using an analogous notation as for the TOPTOPLEPHBB hypothesis, taking into account the H boson doesn't originate from an anomalous top decay, the kinematic constraints for the TOPHLEPBB-hypothesis, $q \to t^{lep}H \to (W^{lep}b_1)(b_2b_3) \to ((\ell\nu)b_1)(b_2b_3)$, become

$$\text{TOPHLEPBB} : \begin{cases} m_{W^{lep}}^2 = (E(\ell) + E(\nu))^2 - (p_x(\ell) + p_x(\nu))^2 - \\ \qquad (p_y(\ell) + p_y(\nu))^2 - (p_z(\ell) + p_z(\nu))^2, \\ m_H^2 = (E(b_2) + E(b_3))^2 - (p_x(b_2) + p_x(b_3))^2 - \\ \qquad (p_y(b_2) + p_y(b_3))^2 - (p_z(b_2) + p_z(b_3))^2, \\ m_{t^{lep}}^2 = (E(\ell) + E(\nu) + E(b_1))^2 - (p_x(\ell) + p_x(\nu) + p_x(b_1))^2 - \\ \qquad (p_y(\ell) + p_y(\nu) + p_y(b_1))^2 - (p_z(\ell) + p_z(\nu) + p_z(b_1))^2. \end{cases} \quad (5.9)$$

The TOPTOPLEPHAD hypothesis is similar to the TOPTOPLEPHBB hypothesis when the anomalous top decay is replaced by the SM hadronic decay of a top quark. Two b quarks come from top quark decays and are labelled as b_1 and b_2 and the hadronic decay of a W boson results in two light quarks q_1 and q_2. Finally, the set of kinematic constraints for the TOPTOPLEPHAD hypothesis, $t^{lep}t^{had} \to (W^{lep}b_1)(W^{had}b_2) \to ((\ell\nu)b_1)((q_1q_2)b_2)$, becomes

$$\text{TOPTOPLEPHAD} : \begin{cases} m_{W^{lep}}^2 = (E(\ell) + E(\nu))^2 - (p_x(\ell) + p_x(\nu))^2 - \\ \qquad (p_y(\ell) + p_y(\nu))^2 - (p_z(\ell) + p_z(\nu))^2, \\ m_{W^{had}}^2 = (E(q_1) + E(q_2))^2 - (p_x(q_1) + p_x(q_2))^2 - \\ \qquad (p_y(q_1) + p_y(q_2))^2 - (p_z(q_1) + p_z(q_2))^2, \\ m_{t^{lep}}^2 = (E(\ell) + E(\nu) + E(b_1))^2 - (p_x(\ell) + p_x(\nu) + p_x(b_1))^2 - \\ \qquad (p_y(\ell) + p_y(\nu) + p_y(b_1))^2 - (p_z(\ell) + p_z(\nu) + p_z(b_1))^2, \\ m_{t^{had}}^2 = (E(q_1) + E(q_2) + E(b_2))^2 - (p_x(q_1) + p_x(q_2) + p_x(b_2))^2 - \\ \qquad (p_y(q_1) + p_y(q_2) + p_y(b_2))^2 - (p_z(q_1) + p_z(q_2) + p_z(b_2))^2. \end{cases} \quad (5.10)$$

At the reconstruction level, where quarks are interpreted as jets and $(p_x(\nu), p_y(\nu))$ as the missing transverse energy components (E_x^{miss}, E_y^{miss}), the longitudinal momentum of the neutrino is expressed as

$$p_z^\nu = \frac{cb \pm \sqrt{c^2 b^2 - a^2(d^2 e^2 - c^2)}}{a^2}, \tag{5.11}$$

where $a^2 = p_x(\ell)^2 + p_y(\ell)^2$, $b = p_z(\ell)$, $c = m_W^2/2 + p_x(\ell)E_x^{miss} + p_y(\ell)E_y^{miss}$, $d^2 = (E_x^{miss})^2 + (E_y^{miss})^2$ and $e = E(\ell)$. Events are labelled as *fully reconstructed* when a solution for the longitudinal momentum of the neutrino is found. Events where this is not the case are labelled *partially reconstructed*, which represent about 6% of the all events. The longitudinal component is calculated for each possible jet combination.

Kinematic discriminant

Due to misreconstructions, particles escaping detection and resolution effects, the four-momenta of reconstructed objects don't match the generated objects. In order to improve the procedure to find a valid solution for p_z^ν, the four-momenta of reconstructed leptons and jets, as well as E_T^{miss}, are independently recalculated several times according to *toy events*. These toy events are generated from the original event, where the four-momenta of reconstructed objects are altered according to probability density functions (PDFs) that represent the differences between generated and reconstructed objects. The PDFs are estimated from simulated events for the three considered hypotheses separately by fitting summed gaussian distributions to the measured differences. For the following kinematic variables PDFs were calculated: (E_x^{miss}, E_y^{miss}), $(p_x^b, p_y^b, p_z^b, E^b)$, $(p_x^{non-b}, p_y^{non-b}, p_z^{non-b}, E^{non-b})$, $(p_x^e, p_y^e, p_z^e, E^e)$, $(p_x^\mu, p_y^\mu, p_z^\mu, E^\mu)$. Additionally, this list is extended to also include PDFs with the reconstructed distributions of top quark and Higgs boson masses, as well as the generated mass of the W boson. A subset of the measured PDFs for the TOPTOPLEPHBB-hypothesis are presented in Fig 5.10.

For each jet permutation in an event, the kinematic properties of reconstructed objects are varied as explained above and the following *TopKinFit discriminant* is computed:

$$D = -2 \cdot ln(\prod_i P_i) = \sum_i L(P_i), \tag{5.12}$$

where P_i denotes a probability extracted from each PDF corresponding to the reconstructed mass spectrum, while $L(P_i)$ is the result likelihood. The discriminants for the three hypotheses are defined as:

$$D(\text{TOPTOPLEPHBB}) = L(m_W^{lep}) + L(m_t^{lep}) + L(m_H) + L(m_t^{had}), \tag{5.13}$$

$$D(\text{TOPHLEPBB}) = L(m_W^{lep}) + L(m_t^{lep}) + L(m_H), \tag{5.14}$$

$$D(\text{TOPTOPLEPHAD}) = L(m_W^{lep}) + L(m_t^{lep}) + L(m_W^{had}) + L(m_t^{had}). \tag{5.15}$$
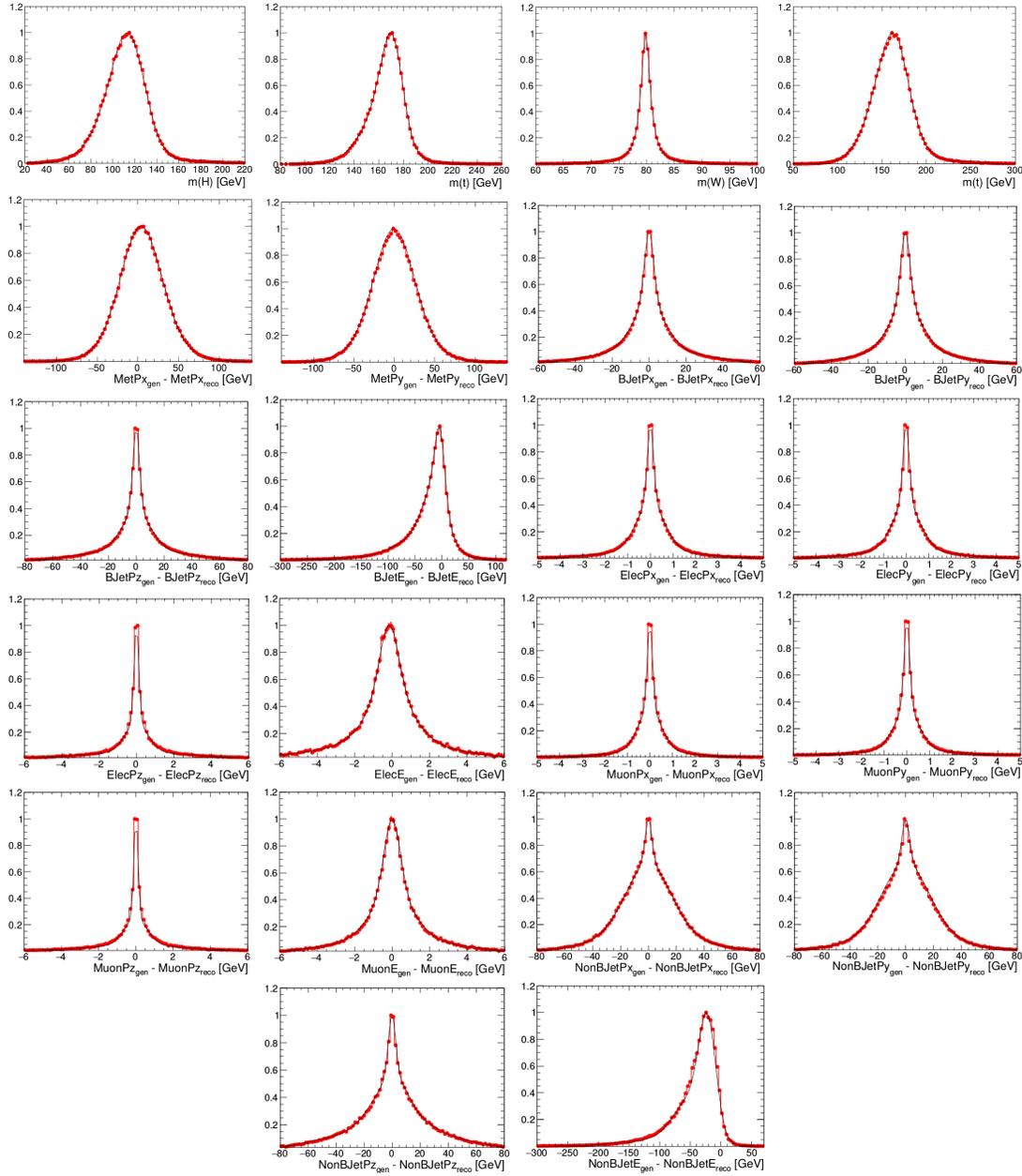
Figure 5.10: Probability distribution functions for reconstructed objects used in the kinematic event reconstruction in the TOPTOPLEPHBB hypothesis. From left to right, top to bottom: m_H, m_t^{lep}, m_W^{lep}, m_t^{had}, (E_x^{miss}, E_y^{miss}) $(p_x^b, p_y^b, p_z^b, E^b)$, $(p_x^e, p_y^e, p_z^e, E^e)$, $(p_x^\mu, p_y^\mu, p_z^\mu, E^\mu)$, $(p_x^{non-b}, p_y^{non-b}, p_z^{non-b}, E^{non-b})$.

Among all toy events the variation with the lowest value of the TopKinFit discriminant is chosen to correspond to the resolved p_z^ν for a given jet permutation. In this thesis, 100 toy events are considered per jet permutation in each event.

5.3.2   Jet assignement strategy with Boosted Decision Trees

The collection of selected jets in an event supports several combinations that can be matched to the three hard process hypotheses. The decision for selecting the correct jet combination in each event can be based on the TopKinFit discriminant, in which case the jet permutation with the lowest discriminant value is chosen as correct combination. A multivariate analysis however, can enhance the efficiency of picking out the correct jet combination. Boosted Decision Trees (BDT), which are briefly introduced below, are used as an MVA technique in this thesis in order to resolve the jet combinatorics. More specifically, BDTs are implemented from the `TMVA Toolkit` [119] in this analysis. The use of BDTs in order to correctly assign jets to the underlying hard process quarks is explained further down.

Boosted Decision Trees

Recognizing patterns in data via artificial intelligence is indispensable for a wide range of purposes, whether it is to classify pp-collision events as being originated from an FCNC signal or a SM background event, or to determine if an e-mail is spam or not. This field of study is developed under the common denominator of *machine learning*. Machine learning techniques mostly adopt a multivariate approach that learns patterns from an input sample, also referred to as *training sample*, in order to classify an observed phenomenon. The technique of BDTs is an example of such a multivariate approach, aiming to classify an event as signal (S) or background (B) based on a BDT event discriminant that maximizes the signal-to-background separation power.

The concept of BDTs is rooted in single Decision Trees. A decision tree takes as input a number of variables that provide some discrimination power between S and B. The learning procedure constructs *nodes* at which the training sample is split by placing a prerequisite on the most discriminating variable in order to separate as many background from signal events. The split decision in this analysis is based on the *Gini Index*, defined as

$$G_{index} = P \cdot (1 - P), \tag{5.16}$$

where $P = S/(S + B)$ represents the purity at each node. The Gini index is maximal if there is an equal amount of signal and background in the sample[4], and becomes zero if the sample is pure in either signal or background. The most discriminating variable and its prerequisite at a node are chosen such that the difference in Gini index between the parent node and the sum of Gini indices of the two daughter nodes achieves a

---

[4]Note that the training sample is reweighted such that the decision tree starts with a maximal Gini index, i.e. signal or background events are reweighted such that $S = B$ (or equivalently $P = 0.5$).

maximum. When the decision tree reaches a predefined depth, which corresponds to the number of consecutive nodes, the learning stops. At that point, the *leaves* of the tree are classified as S or B based on the purity of the leaf. The decision tree process is visualized in Figure 5.11.
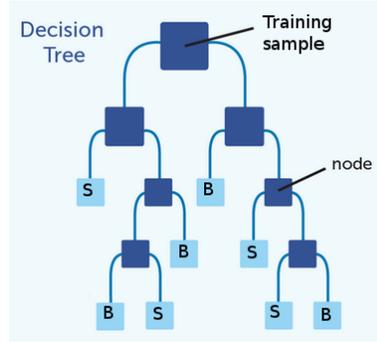
Figure 5.11: A visualisation of a Decision Tree.

Usually a small depth is defined in combination with multiple decision trees. The exact choice is use-case dependent and is based on a combination of reasons: limited CPU time and power, number of input variables and size of the training sample. One has to choose wisely in order to not *overtrain* the decision trees, which means the procedure learns too much details that are specific to the training sample, but not necessarily representative of the full sample.

As impure leaves of a decision tree are classified purely as S or B, boosting is invoked in order to reduce influences of these impurities. The boosting process starts from a simple decision tree and sequentially builds more decision trees, such that the training sample for the next tree gets boosted by only considering incorrectly classified events as training sample. In this analysis, the adaptive boosting or *AdaBoost* is the adopted boosting method. Adaptive boosting takes the rate error at each tree, defined as the number of wrongly classified training events divided by the total number of training events, and defines the degree of boosting as

$$\alpha = \frac{1 - \text{err}}{\text{err}}. \tag{5.17}$$

The degree of boosting for each tree is taken logarithmically into account in the final BDT discriminant as

$$\text{BDT}_{\text{disc}}(\vec{x}) = \frac{1}{N_{\text{trees}}} \sum_i^{N_{\text{trees}}} ln(\alpha_i) h_i(\vec{x}), \tag{5.18}$$

with $\vec{x}$ the set of input variables and $h(\vec{x})$ the result of a decision tree, encoded as $h(\vec{x}) = +1$ and $h(\vec{x}) = -1$ for events categorized in respectively signal and background

leaves. Events with a BDT discriminant value close to $+1$ are considered signal-like, whereas values close to -1 are more background-like.

In the BDT learning procedure, variables are ranked according to the number of times they are used for splitting the training sample. Variables that are ranked high contribute the most to the discrimination power of the final BDT. On the other hand, variables that are ranked low could have a negligible contribution to the discrimination power of the final BDT. Such variables could be considered to be discarded from the training, as the number of variables complicates the training procedure.

Jet assignment strategy

Additionally to the kinematic event reconstruction, a BDT is used to define the most probable jet combination corresponding to the correct assignment of reconstructed jets to generated quarks. The training sample for the BDT is composed of all possible jet combinations, where signal is defined as the correct and background as the wrong jet combinations. Four variables from the kinematically reconstructed event are used as input for the BDT for the three different hypotheses. Jet combinations from fully reconstructed and partially reconstructed are trained separately. An overview of the input variables for the different hypotheses is given in Table 5.6. For partially reconstructed events, variables based on the reconstructed leptonic top are projected to the transverse plane, due to the fact that the longitudinal momentum of the neutrino did not get reconstructed. In this partial reconstruction, masses of objects are replaced by transverse masses m_T and spatial separations ΔR by $\Delta\phi$. The jet combination with the highest BDT discriminant value is chosen as the correct jet assignment.

Table 5.6: A summary of the sets of four variables used as input for the BDT training for each of the three event hypotheses, distinguishing between fully and partially reconstructed events.

Hypothesis	Reconstruction	Variables for training
TOPTOPLEPHBB	Full	$m(H)$, $m(t^{lep})$, $\Delta R(H,t^{lep})$, $p_T(t^{lep})$
	Partial	$m(H)$, $m_T(t^{lep})$, $\Delta\phi(H,t^{lep})$, $p_T(t^{lep})$
TOPHLEPBB	Full	$m(H)$, $m(t^{lep})$, $\Delta R(H,t^{lep})$, $p_T(t^{lep})$
	Partial	$m(H)$, $m_T(t^{lep})$, $\Delta\phi(H,t^{lep})$, $p_T(t^{lep})$
TOPTOPLEPHAD	Full	$m(t^{had})$, $m(t^{lep})$, $\Delta R(t^{had},t^{lep})$, $p_T(t^{lep})$
	Partial	$m(t^{had})$, $m_T(t^{lep})$, $\Delta\phi(t^{had},t^{lep})$, $p_T(t^{lep})$

The three hypotheses represent topologies with a different number of jets and b jets. The efficiency of proper assignment of initial quarks to reconstructed jets depends on the algorithm used to identify b jets. In order to study this dependence, several options are considered for composing the jet collection from which the correct jet combination

is picked[5]:

- Truth: only reconstructed jets matched to initial quarks are used. This jet collection can only be considered for simulated events and serves as a reference of the performance of the method, meaning the performance of the method on this jet collection represents the maximal performance of the method.

- HighestCSVv2: sort jets according to decreasing CSVv2 discriminant value and consider the first two (three) highest CSVv2 value jets as b jets for the TOPTOPLEPHAD (TOPTOPLEPHBB and TOPHLEPBB) hypothesis. The remaining jets are considered as light jets.

- CSVv2L: select b jets as those passing the Loose working point of the CSVv2 algorithm.

- CSVv2M: select b jets as those passing the Medium working point of the CSVv2 algorithm.

- CSVv2T: select b jets as those passing the Tight working point of the CSVv2 algorithm.

- All: consider all jets for permutations without imposing b-tagging requirement.

Note that the TOPTOPLEPHBB and TOPTOPLEPHAD hypothesis are only considered for events with at least 4 jets. In case all jets in an event satisfy the CSVv2L, CSVv2M or CSVv2T options, the last (two) jet(s) in the collection, ordered according to decreasing CSVv2 value, are considered as light jets for the TOPTOPLEPHBB (TOPTOPLEPHAD) hypothesis.

Figures 5.12 and 5.13 illustrate the discrimination power between correct and wrong jet combinations of the BDT input variables for the TOPTOPLEPHBB hypothesis in the different jet collections, respectively for fully reconstructed and partially reconstructed events. The corresponding outcomes of the BDT are illustrated in Figures 5.14 and 5.15. Figures 5.14 and 5.15 also serve as a visual check for overtraining, as both the BDT outcome for the training sample as a validation sample are shown. If an obvious difference exists between the BDT distributions of the training and validation sample, this would indicate overtraining. No indication of overtraining is found for these trainings. A good discrimination power between correct and wrong jet combinations is found in both the input variables as the BDT discriminants. The performance of the method on the different jet collections is compared in Figure 5.16, which illustrates the background rejection efficiency as a function of signal efficiency for a scan over the BDT discriminant. This clearly indicates that the method performs much worse for the All jet collection with respect to jet collections in which b-tagging information is used.

---

[5]For the jet assignment study, the requirement on the number of b-tagged jets in the baseline selection is dropped.
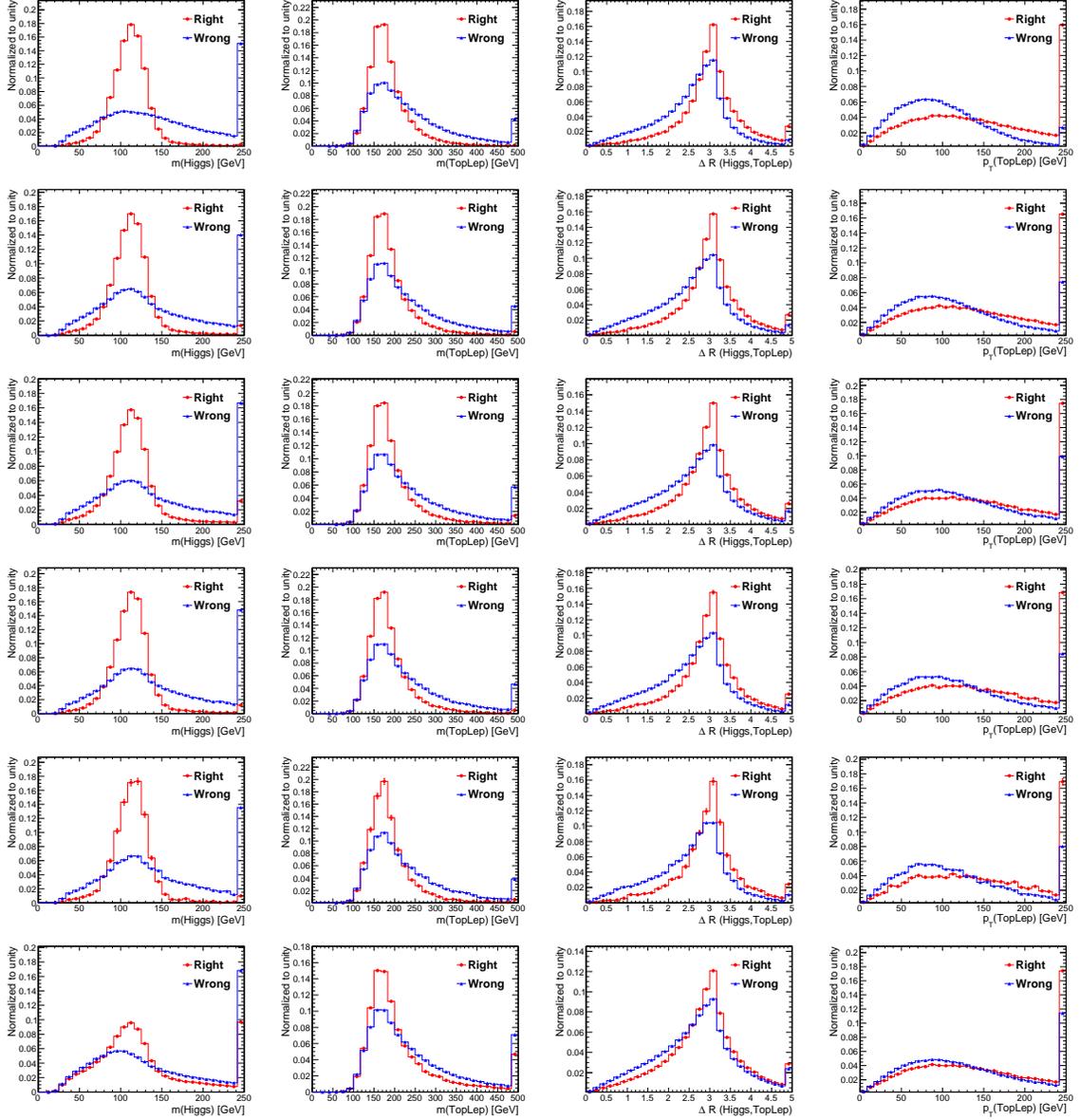
Figure 5.12: Normalized BDT input variables for fully reconstructed events in the TOPTOPLEPHBB hypothesis for correct (red) and wrong (blue) jet combinations. Jet collection used from top to bottom row: Truth, HighestCSVv2, CSVv2L, CSVv2M, CSVv2T, All.

For each of the choices of the b tagging selection to be considered for the analysis, the corresponding efficiencies to select the correct jet assignment are computed. These efficiencies are presented in Figures 5.17—5.19 for all considered hypotheses and whether one is interested in correctly matching jets with all hard process quarks in the left plots or just with the hard process b quarks in the right plots. To indicate the gain
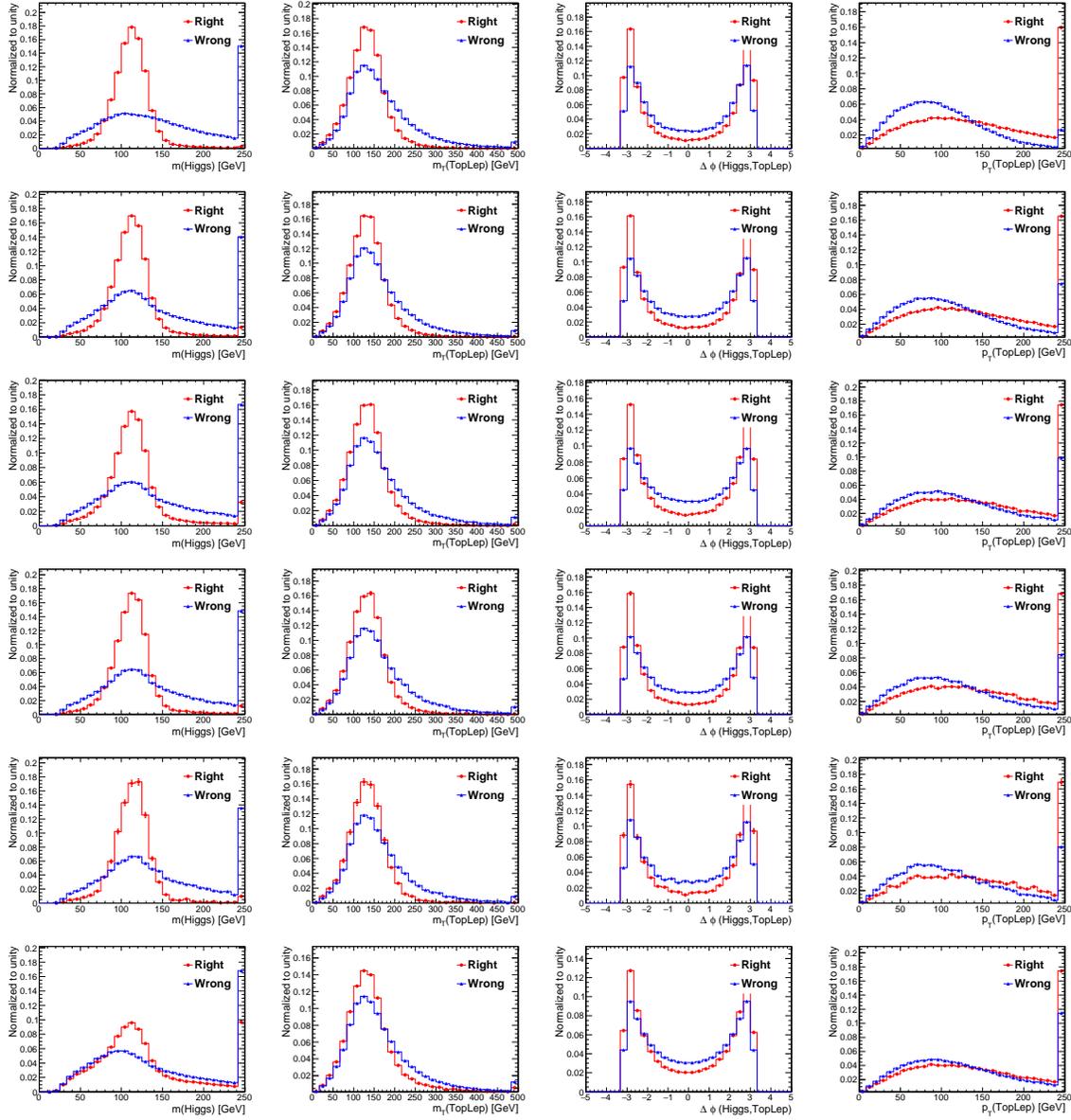
Figure 5.13: Normalized BDT input variables for partially reconstructed events in the TOPTOPLEPHBB hypothesis for correct (red) and wrong (blue) jet combinations. Jet collection used from top to bottom row: Truth, HighestCSVv2, CSVv2L, CSVv2M, CSVv2T, All.
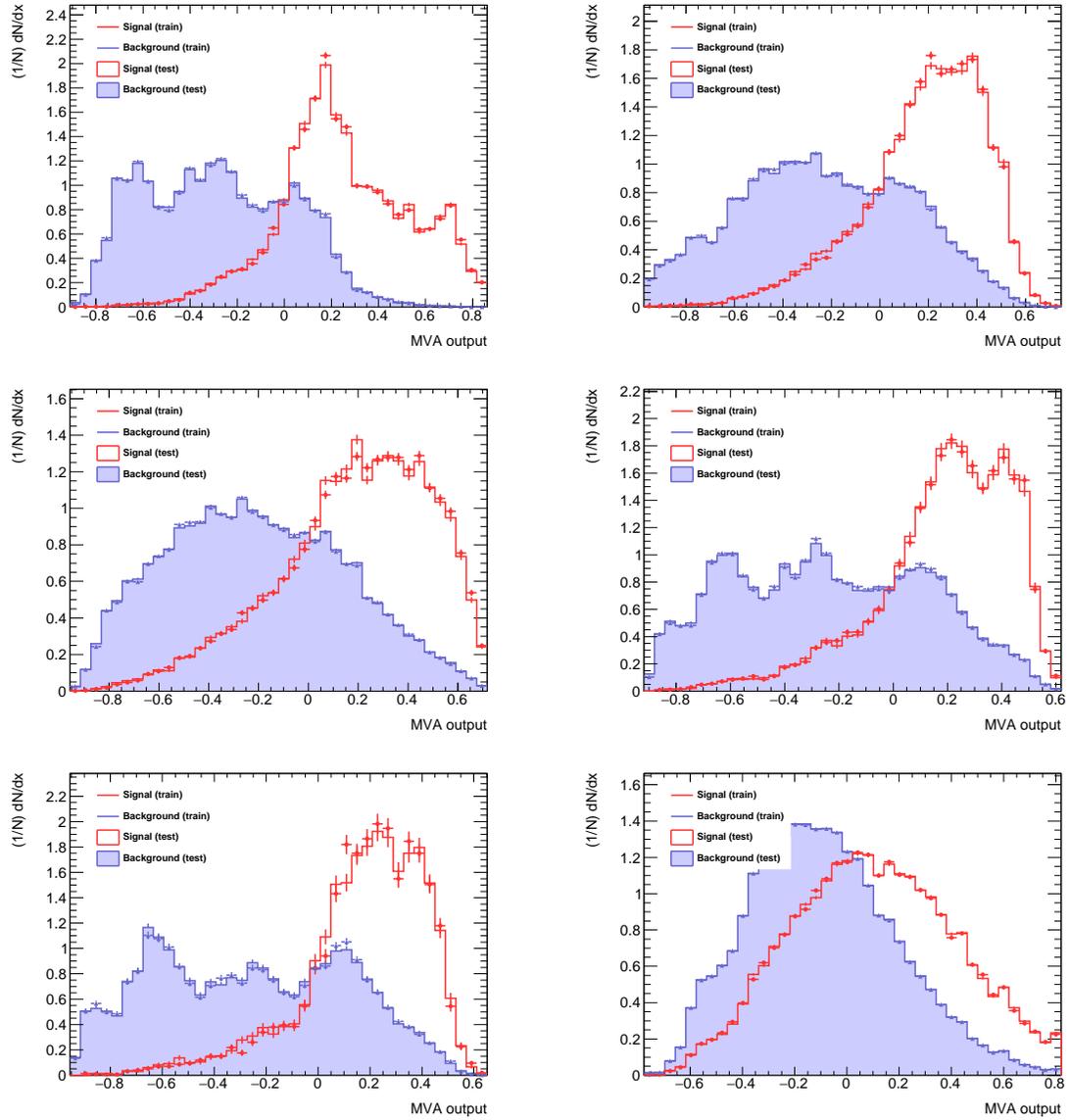
Figure 5.14: Resulting discriminants of BDT training for fully reconstructed events in the TOPTOPLEPHBB hypothesis. The distribution for the trained sample is indicated by individual points, whereas the validation sample is represented by the unbroken line. Jet collection used from top to bottom, left to right: Truth, HighestCSVv2, CSVv2L, CSVv2M, CSVv2T, All.
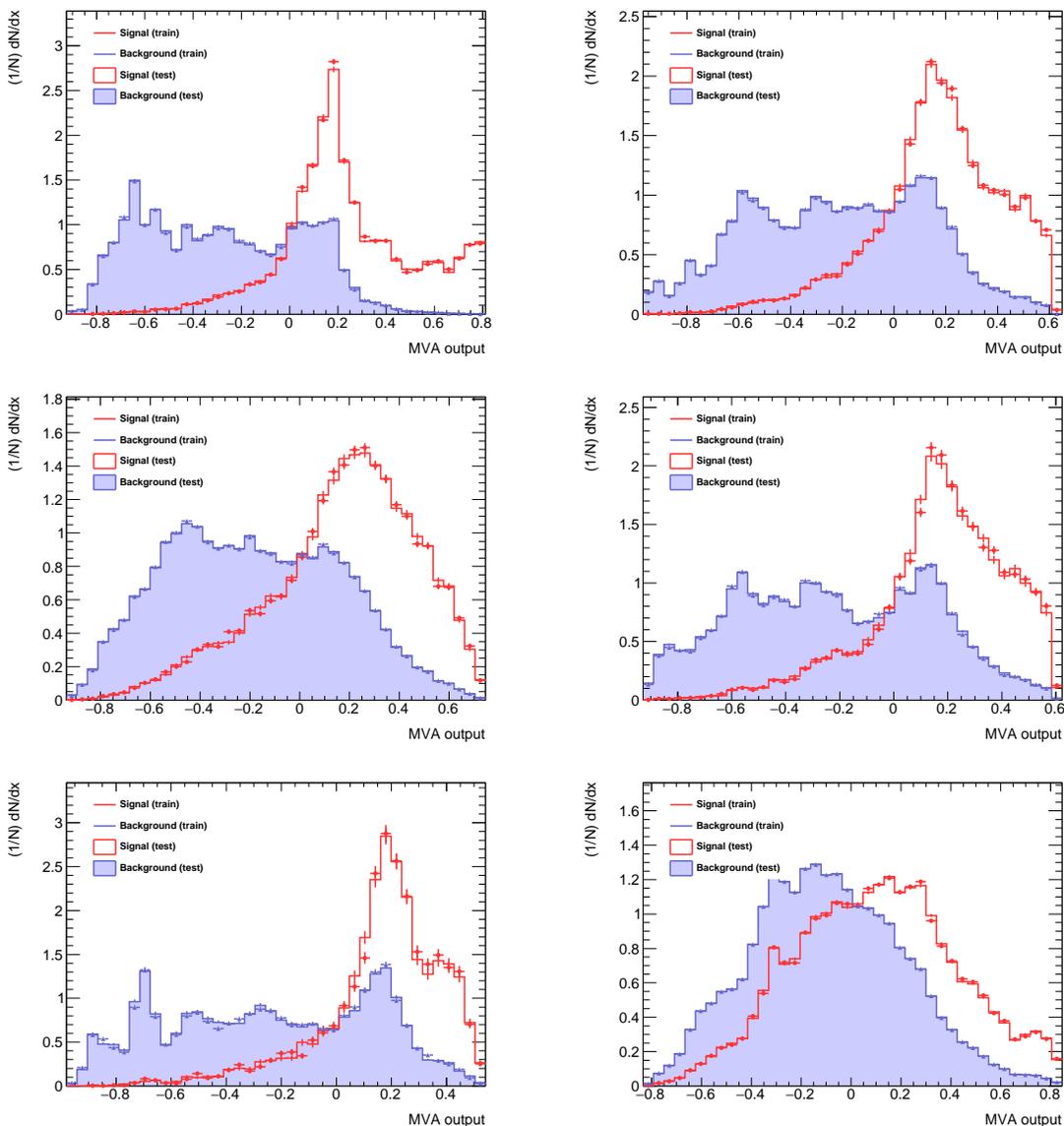
Figure 5.15: Resulting discriminants of BDT training for partially reconstructed events in the TOPTOPLEPHBB hypothesis. The distribution for the trained sample is indicated by individual points, whereas the validation sample is represented by the unbroken line. Jet collection used from top to bottom, left to right: Truth, HighestCSVv2, CSVv2L, CSVv2M, CSVv2T, All.
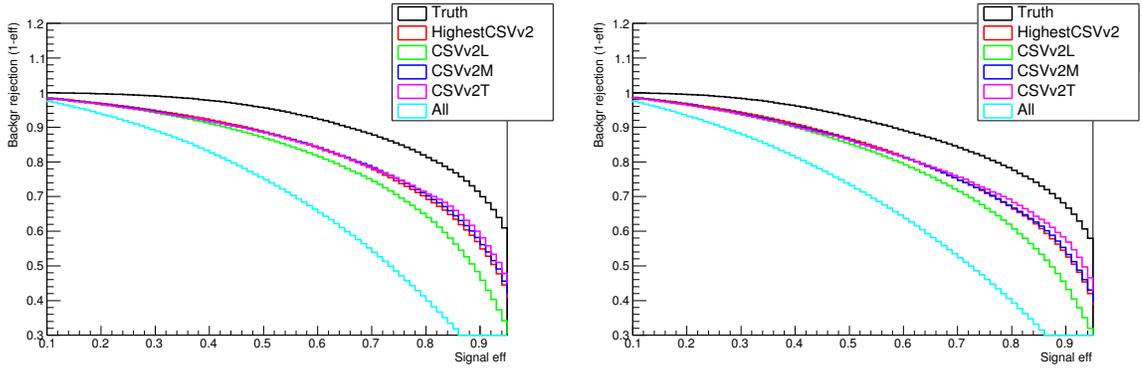
Figure 5.16: Performance comparison of the trained BDTs for fully reconstructed events (left) and partially reconstructed events (right), showing the background rejection efficiency as a function of signal efficiency for the considered jet collections. The efficiencies are obtained by performing a cut-scan over the BDT discriminants and calculating the number of signal (background) events surviving the cut versus the initial number of signal (background) events.

of using a BDT, the efficiency is also shown for if the jet combination would be chosen as the one giving the lowest TopKinFit discriminant (Equations (5.13)—(5.15)). For all hypotheses and all considered jet collections, the gain of using a BDT on top of the kinematic reconstruction is apparent.

The Truth jet collection reconstruction efficiency can be interpreted as the maximally possible reconstruction efficiency of the method. If one is interested in only correctly assigning jets to b quarks from the hard processes, all jet collection compositions render reconstruction efficiencies close to the Truth collection, except for the All collection. However, the full hard process reconstruction indicates that only the CSVv2M and CSVv2T jet collections come close to the Truth jet collection reconstruction efficiency.

As additional information, the event selection efficiencies for the three CSVv2 working points are shown in Fig 5.20 for the three different hypotheses. Although the CSVv2L working point is the most efficient way of maximizing the number of selected events, and the algorithm associated to the CSVv2L collection is almost as efficient as for CSVv2M and CSVv2T, the CSVv2L is not initially considered for tagging b jets in this analysis. This is due to the fact that this b-tagging working point is much more subject to misidentifying light jets as b jets, which would make the analysis sensitive to SM processes, such as multijet events, that could otherwise be neglected (see Chapter 6). On the other hand, the CSVv2T working point rejects more than twice the number of events that get selected via the CSVv2M working point, while the algorithm efficiency is similar for the two working points. Therefore, as a compromise between the amount of selected statistics and the efficiency of the event reconstruction, the

b-tagging selection that corresponds to CSVv2M requirement is initially used in the main analysis.
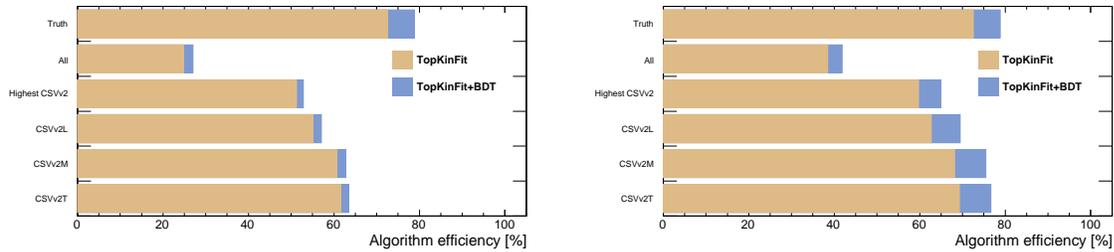
Figure 5.17: Efficiency of the jet assignment algorithm for all hard process quarks (left) and b quarks only (right) in TOPTOPLEPHBB hypothesis. The efficiency based on the TopKinFit discriminant only is indicated in brown, with the gain of using the BDT indicated in blue.
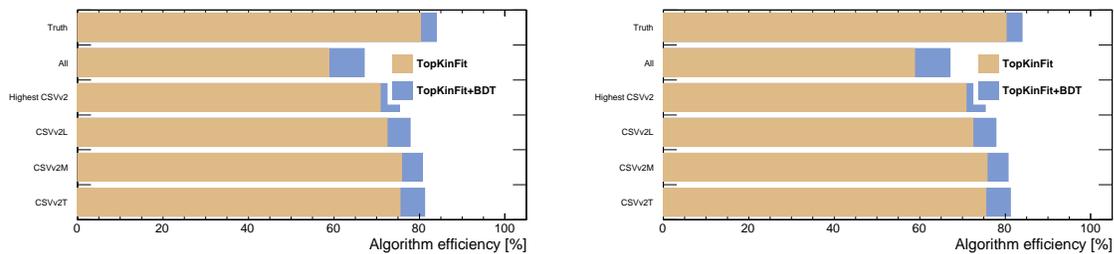
Figure 5.18: Efficiency of the jet assignment algorithm for all hard process quarks (left) and b quarks only (right) in TOPHLEPBB hypothesis. The efficiency based on the TopKinFit discriminant only is indicated in brown, with the gain of using the BDT indicated in blue.
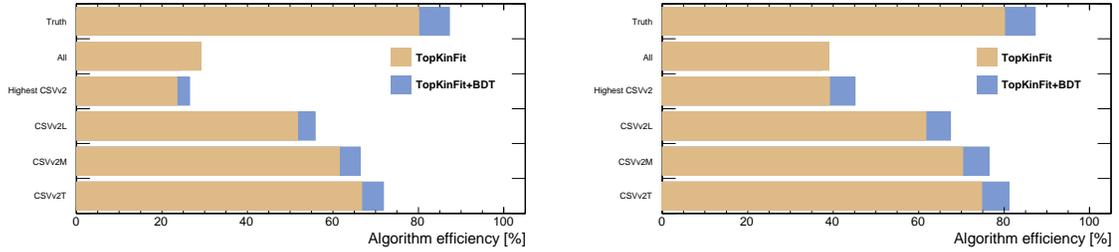
Figure 5.19: Efficiency of the jet assignment algorithm for all hard process quarks (left) and b quarks only (right) in TOPTOPLEPHAD hypothesis. The efficiency based on the TopKinFit discriminant only is indicated in brown, with the gain of using the BDT indicated in blue.
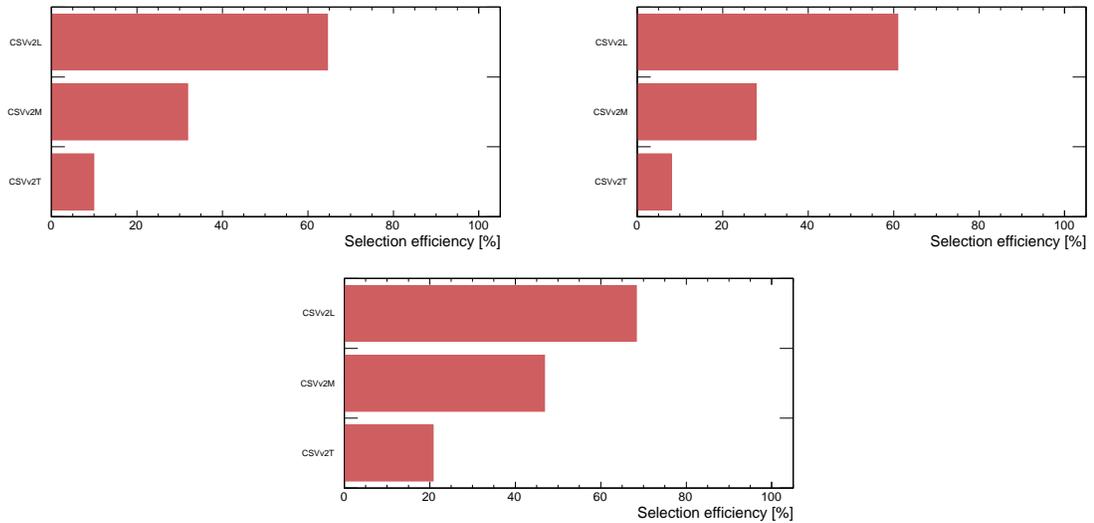
Figure 5.20: Event selection efficiency for TOPTOPLEPHBB (top left), TOPHLEPBB (top right) and TOPTOPLEPHAD (bottom) hypotheses.

Chapter 6

Search for top-FCNC with H → b$\bar{\text{b}}$

A dedicated search for rare phenomena depends on a correct description of the bulk of the collision data with theory, simulation and calibrations. The baseline event selection and applied scale factors applied to simulation prove this to be the case, as shown in Section 5.2. On top of that, the main event signatures can be reconstructed with high efficiencies. With these basic ingredients, the search for respectively the *Hut* and *Hct* vertex in top-FCNC involving a Higgs boson decay into a pair of b quarks, can be pursued.

The particular challenge in this search is the fact that the predicted number of t$\bar{\text{t}}$ background events is immensely large and very much alike in kinematical and topological properties to the FCNC signal processes. In order to capitalize on existing, yet subtle, differences between signal and background, the analysis is factorized into exclusive event categories as detailed in Section 6.1. Each event category is defined such that the contribution of one of the following three event signatures is maximized: *ST FCNC Hqt*, *TT FCNC Hqt* or SM t$\bar{\text{t}}$. A dedicated signal-to-background discrimination is performed in each category by means of BDTs, as explained in Section 6.2.

The processes and reconstructed objects that are used in the construction of the BDTs are known with a particular degree of certainty. The known sources of systematic uncertainty that affect this analysis are summarised in Section 6.3. Finally, by means of a combined template fit, where the constructed BDT discriminants of all event categories are simultaneously fitted and systematic uncertainties are treated as nuisance parameters, limits are extracted on the branching ratios of both t → uH and t → cH. The limit extraction and final results are discussed in Section 6.4.

Throughout this chapter, unless explicitly mentioned otherwise, the nominal FCNC signal processes assume a coupling strength $\kappa_{Hqt} = 1$, which corresponds to a branching ratio $\mathcal{B}(\text{t} \rightarrow \text{qH}) = 0.14$ (see Equation 1.25 in combination with Table 4.1).

6.1   Event categorisation

The number of reconstructed jets and b-tagged jets has been cited as a source of discrimination power between SM tt̄ background events and the *ST Hqt* and *TT Hqt* signal events on numerous occasions throughout Chapters 4 and 5. Categorizing events according to the number of jets and b-tagged jets allows to factorize the contributions of those three event signatures. Categories with exactly three jets are created to single out *ST Hqt* signal events as much as possible, whereas categories with at least four jets increase the relative contribution of *TT Hqt* signal events. A further subcategorisation increases the signal-to-background ratio by requiring exactly three b-tagged jets, as SM tt̄ events expect only two b-tagged jets from the leading-order process. Categories with exactly two b-tagged jets on the other hand, even though they are expected to decrease the signal-to-background ratio, can serve as *control categories* for the dominating tt̄ background events. The remaining events of the baseline event selection are grouped in a category with at least four b-tagged jets. Even though none of the three main event signatures support this choice for a fifth category on the basis of their leading-order topology, the charm jet in the *TT Hct* process may be misidentified as a b jet with a probability of about 12%. The five mutually exclusive event categories, pictorially represented in Figure 6.1, are labelled as:

- *b2j3*: Exactly two b-tagged jets and exactly three jets. This is a control category for SM tt̄.

- *b2j4*: Exactly two b-tagged jets and at least four jets. This is a control category for SM tt̄.

- *b3j3*: Exactly three b-tagged jets and exactly three jets. This is a signal category, mainly for *ST Hut* and *ST Hct*.

- *b3j4*: Exactly three b-tagged jets and at least four jets. This is a signal category, mainly for *TT Hut* and *TT Hct*.

- *b4j4*: At least four b-tagged jets and four jets. This is a signal category, mainly for *TT Hct*.

It is expected that 3-jet categories are more sensitive to the *ST Hut* than the *ST Hct* signal, due to its higher cross section. The tt̄ components of the signal will have non-negligible contributions in those 3-jet categories as well. As a general trend, one can expect to select more *TT Hct* events than *TT Hut* events due to the higher misidentification probability of the charm jet as a b jet. These trends are confirmed in Tables 6.1 & 6.2, which summarise the event yields for every process in each of the five categories. The indicated uncertainties represent the quadratic sum of all relative systematic uncertainties in the analysis, as will be introduced in Section 6.3. An illustration of the excellent agreement between the simulated processes and actual data in all five categories is shown in Figure 6.2.
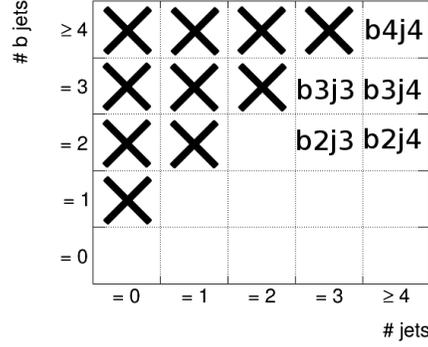
Figure 6.1: Representation of the phase-space coverage of the event categories in the analysis in the number of jets and b jets.

Table 6.1: Event yields in event categories with exactly two b-tagged jets. The change in event yield due to the total up ($+1\sigma$) and down (-1σ) systematic uncertainty, calculated as the quadratic sum of all systematic uncertainties as summarised in Section 6.3, is indicated as well.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$
Data	365 890	575 500
ST Hct	2 271$^{+5.1\%}_{-4.3\%}$	1 843$^{+5.4\%}_{-5.8\%}$
TT Hct	36 138$^{+5.7\%}_{-4.8\%}$	63 393$^{+6.2\%}_{-5.3\%}$
ST Hut	13 273$^{+5.2\%}_{-4.8\%}$	10 863$^{+6.5\%}_{-5.7\%}$
TT Hut	34 601$^{+6.3\%}_{-5.1\%}$	65 226$^{+5.4\%}_{-4.9\%}$
ST ⊕ TT (Hct)	38 409$^{+5.6\%}_{-4.7\%}$	65 235$^{+6.2\%}_{-5.3\%}$
ST ⊕ TT (Hut)	47 874$^{+6\%}_{-4.9\%}$	76 089$^{+5.6\%}_{-5\%}$
t$\bar{\text{t}}$+b$\bar{\text{b}}$	8 649$^{+15\%}_{-14\%}$	31 262$^{+18\%}_{-15\%}$
t$\bar{\text{t}}$+c$\bar{\text{c}}$	23 130$^{+15\%}_{-14\%}$	75 352$^{+18\%}_{-15\%}$
t$\bar{\text{t}}$+lf	269 322$^{+14\%}_{-13\%}$	424 367$^{+17\%}_{-15\%}$
other	53 212$^{+11\%}_{-9.2\%}$	49 794$^{+10\%}_{-8.5\%}$
Total background	354 313$^{+13\%}_{-12\%}$	580 775$^{+16\%}_{-14\%}$

Table 6.2: Event yields in event categories with exactly three and at least four b-tagged jets. The change in event yield due to the total up $(+1\sigma)$ and down (-1σ) systematic uncertainty, calculated as the quadratic sum of all systematic uncertainties as summarised in Section 6.3, is indicated as well.

Process	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
Data	13 481	53 352	2 764
ST Hct	$663^{+14\%}_{-10\%}$	$798^{+14\%}_{-11\%}$	$50^{+24\%}_{-19\%}$
TT Hct	$8\ 146^{+14\%}_{-10\%}$	$27\ 507^{+13\%}_{-10\%}$	$2\ 805^{+26\%}_{-18\%}$
ST Hut	$3\ 453^{+14\%}_{-10\%}$	$4\ 134^{+13\%}_{-10\%}$	$197^{+19\%}_{-21\%}$
TT Hut	$5\ 971^{+14\%}_{-10\%}$	$24\ 095^{+14\%}_{-11\%}$	$1\ 128^{+20\%}_{-18\%}$
ST $\oplus$ TT (Hct)	$8\ 810^{+14\%}_{-10\%}$	$28\ 305^{+13\%}_{-10\%}$	$2\ 855^{+26\%}_{-18\%}$
ST $\oplus$ TT (Hut)	$9\ 424^{+14\%}_{-10\%}$	$28\ 230^{+14\%}_{-11\%}$	$1\ 325^{+20\%}_{-18\%}$
t$\bar{\text{t}}$+b$\bar{\text{b}}$	$1\ 131^{+20\%}_{-19\%}$	$11\ 414^{+21\%}_{-17\%}$	$1\ 371^{+32\%}_{-21\%}$
t$\bar{\text{t}}$+c$\bar{\text{c}}$	$1\ 251^{+31\%}_{-25\%}$	$9\ 765^{+29\%}_{-23\%}$	$468^{+53\%}_{-40\%}$
t$\bar{\text{t}}$+lf	$9\ 291^{+26\%}_{-22\%}$	$28\ 874^{+27\%}_{-23\%}$	$524^{+58\%}_{-53\%}$
other	$1\ 503^{+25\%}_{-19\%}$	$4\ 323^{+19\%}_{-15\%}$	$273^{+25\%}_{-20\%}$
Total background	$13\ 177^{+25\%}_{-20\%}$	$54\ 376^{+24\%}_{-20\%}$	$2\ 636^{+34\%}_{-27\%}$
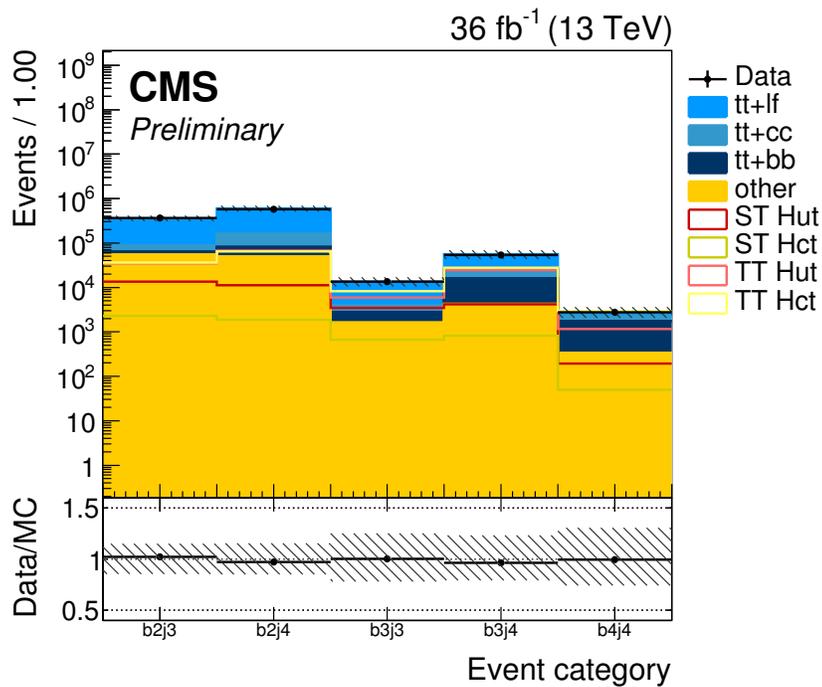
Figure 6.2: Comparison between the number of events in data and simulations after the baseline event selection. The signal processes are drawn individually, overlaying the SM processes. The total uncertainty on the SM processes, representing the quadratic sum of all statistical and systematic uncertainties as summarised in Section 6.3, is indicated by the shaded area.

6.2   Signal to background discrimination with Boosted Decision Trees

The phenomenological study indicated that the differentiation between signal and background based on kinematic properties of the event is a nigh-impossible feat, even if BDTs are invoked. That conclusion however, only holds for the methods applied in that study, which does not include the use of b-tagging discriminants as input variables. Besides the fact that b-tagging discriminants are available in this study, the chosen jet assignment method outperforms those considered in the phenomenological study. Therefore, the use of BDTs in the analysis as a way to discriminate signal events from background events is reconsidered.

BDT training

A BDT training is performed with similar settings as discussed in Section 5.3.2. The signal in the training sample is composed of the sum of the single top and t$\bar{\text{t}}$ component, weighted according to their respective normalisation scale factor (Equation (5.3)). The background in the training sample is defined as the sum of all SM background contributions, again weighted according to their respective normalisation scale factor. Training of the BDT for the *Hut* and *Hct* vertex in top-FCNC are performed individually. As the five event categories represent different signal and background compositions, the BDT training is executed in each category separately. However, due to insufficient signal statistics in the training sample of the *b4j4* category for the *Hut* FCNC vertex, that category is not considered for BDT training as it would induce overtraining and is excluded from the analysis. This totals to nine distinctive BDT trainings, as sketched in Figure 6.3.
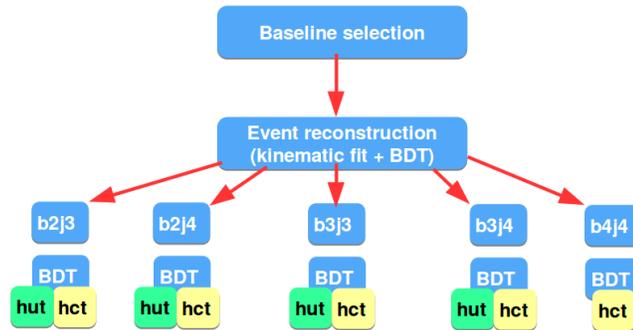
Figure 6.3: Schematic representation of the analysis strategy with nine BDT trainings: four for the *Hut* vertex and five for the *Hct* vertex in top-FCNC.

Input variables

A large collection of variables has been considered for the training procedure, based on the three kinematic event reconstruction hypotheses from Section 5.3.1. This collection has been reduced and two sets have been constructed for training purposes: a single top oriented set and a t$\bar{\text{t}}$ oriented set. The single top oriented set of input variables is used in the 3-jet categories, whereas the t$\bar{\text{t}}$ oriented set is used in the 4-jet categories' trainings. The reduction of the number of input variables is based on the need to minimize the tendency of overtraining the BDT, dropping variables that score low in the BDT variable ranking. The final sets of variables for the 3-jet and 4-jet categories are listed in Table 6.3, respectively ranked according to their BDT ranking in the *b3j3* and *b3j4* training categories of the *Hut* FCNC signal. The same notations as in Equations (5.13)—(5.15) are adopted and the BDT discriminant from the jet assignment procedure is labelled as *bMVA*. It must be noted that one more variable is used in the *Hut* FCNC signal with respect to the *Hct* signal, namely the charge of the lepton. This variable is extremely powerful for the *ST hut* signal, due to the large difference in top versus anti-top anomalous production cross section, as shown in Figure 5.2, resulting in a higher production rate of positively versus negatively charged leptons.

Table 6.3: List of input variables for training the event BDT in all five categories. Two separate sets are defined for 3-jet and 4-jet categories. For each variable, the signature hypothesis that is used for the variable reconstruction is indicated.

b2j3, b3j3		b2j4, b3j4, b4j4	
b_1^H CSVv2	(TOPHLEPBB)	b_1^H CSVv2	(TOPHLEPBB)
$b_{t^{lep}}$ CSVv2	(TOPHLEPBB)	m_H	(TOPHLEPBB)
m_H	(TOPHLEPBB)	$\Delta R(b_1^H, b_2^H)$	(TOPHLEPBB)
$\Delta R(b_1^H, b_2^H)$	(TOPHLEPBB)	bMVA	(TOPHLEPBB)
bMVA	(TOPHLEPBB)	$q_1^{W_{had}}$ CSVv2	(TOPTOPLEPHAD)
$m_{t^{lep}}$	(TOPHLEPBB)	bMVA	(TOPTOPLEPHAD)
η_H	(TOPHLEPBB)	$b_{t^{lep}}$ CSVv2	(TOPHLEPBB)
$p_T(t^{lep})$	(TOPHLEPBB)	$m_{t^{had}}$	(TOPTOPLEPHAD)
b_2^H CSVv2	(TOPHLEPBB)	η_H	(TOPHLEPBB)
Lepton charge	(Hut only)	$b_{t^{had}}$ CSVv2	(TOPTOPLEPHAD)
$\eta_{t^{lep}}$	(TOPHLEPBB)	Lepton charge	(Hut only)
$\Delta R(t^{lep}, H)$	(TOPHLEPBB)	b_2^H CSVv2	(TOPHLEPBB)
		$q_2^{W_{had}}$ CSVv2	(TOPTOPLEPHAD)

The normalized distributions of the variables listed in Table 6.3 are shown in Fig-

ures 6.4 & 6.5. Most of the variables show little discrimination power between signal and background, except for: b_1^H CSVv2, m_H, lepton charge and bMVA in the TOPHLEPBB hypothesis. As an illustration the excellent data to simulations agreement for those four variables is shown in Figure 6.6 for the *b3j3* category, where the individual signal contributions are scaled such that their normalisation matches the total number of simulated background events. The same level of agreement between simulations and data is achieved for all input variables across all nine BDT event trainings. The input variables show little correlation, as illustrated in Appendix A.
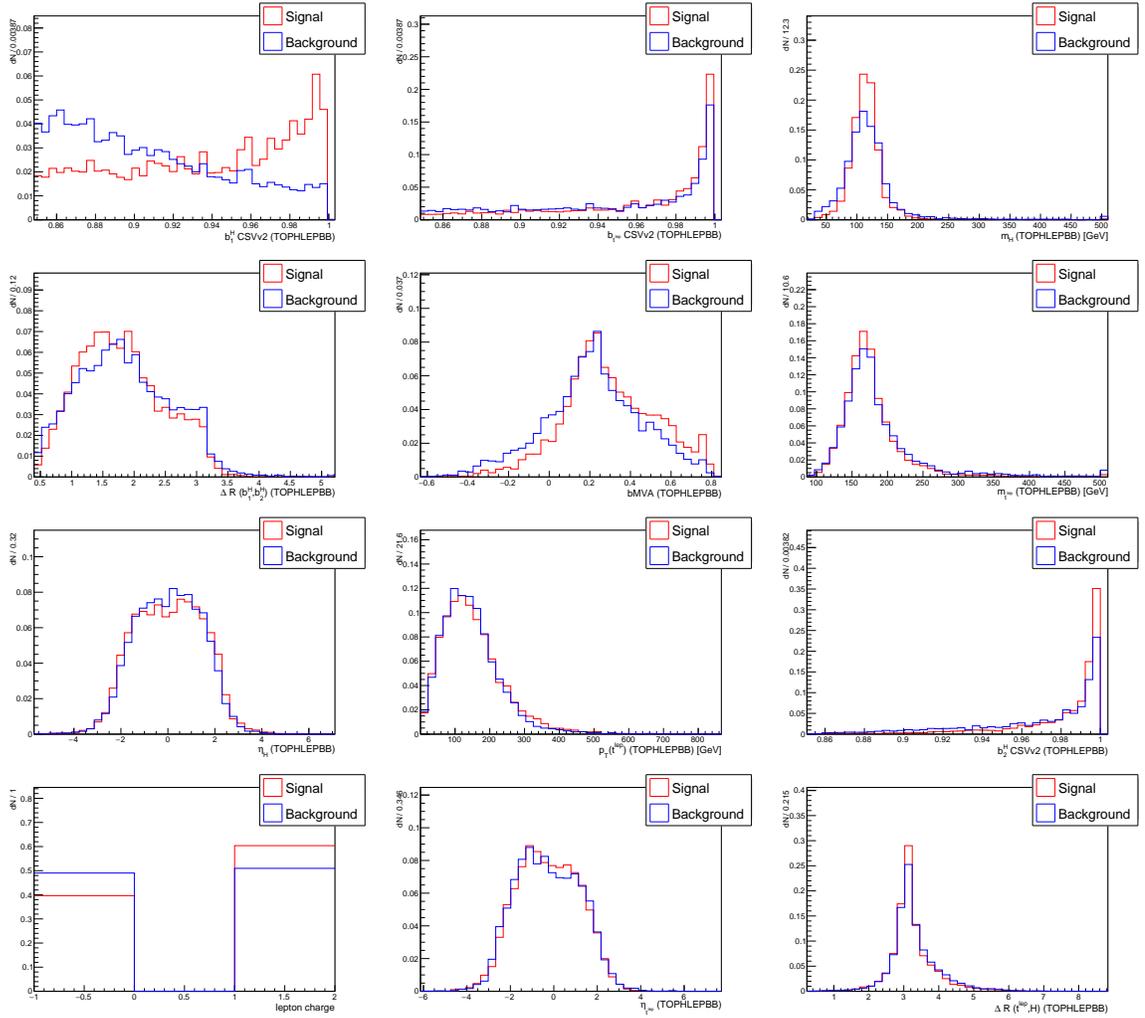
Figure 6.4: Normalized distributions of the input variables for the event BDT training in the *b3j3* category. The signal (red) consists of the top-FCNC *Hut* processes and the background (blue) of SM processes.
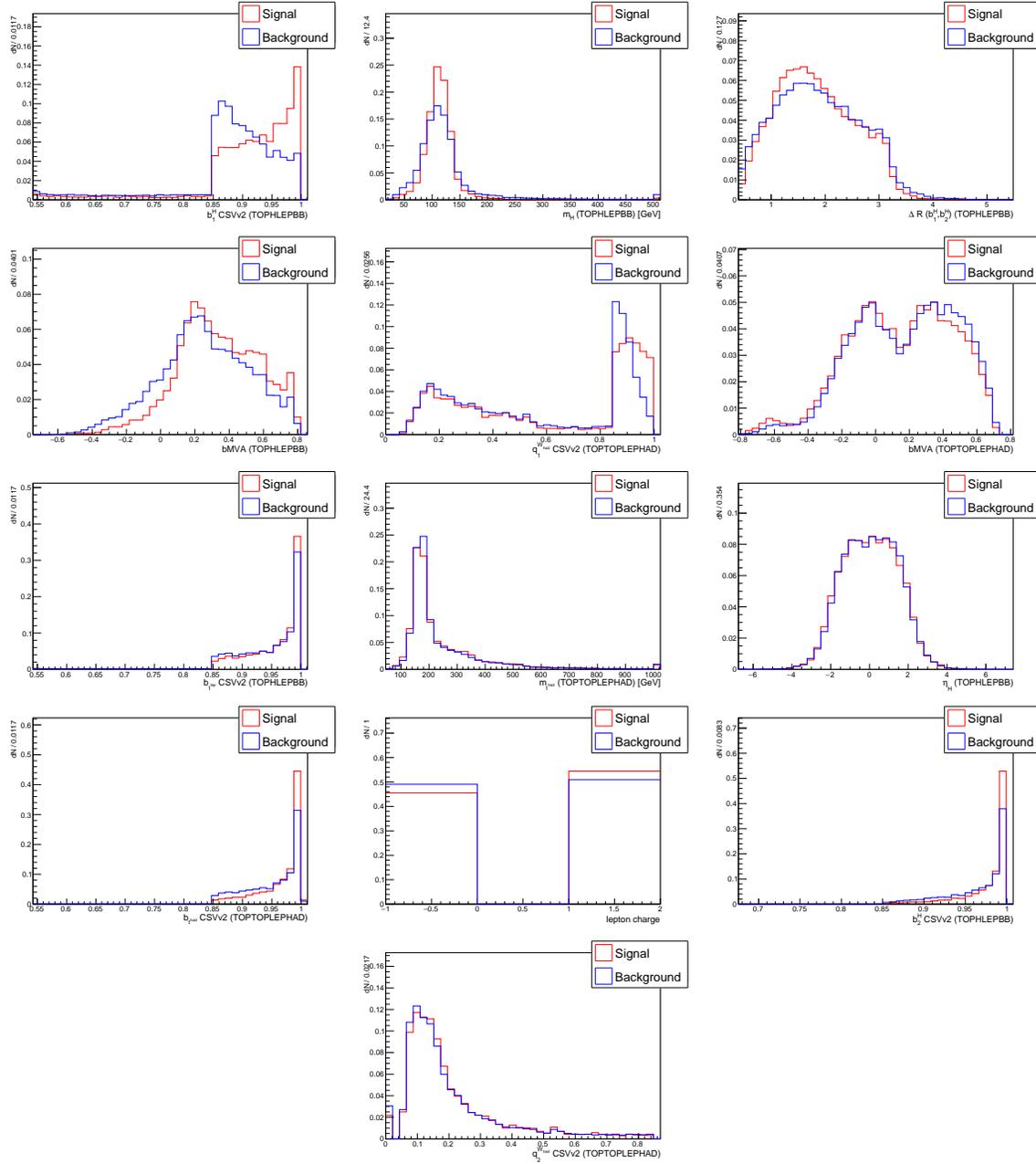
Figure 6.5: Normalized distributions of the input variables for the event BDT training in the *b3j4* category. The signal (red) consists of the top-FCNC *Hut* processes and the background (blue) of SM processes.
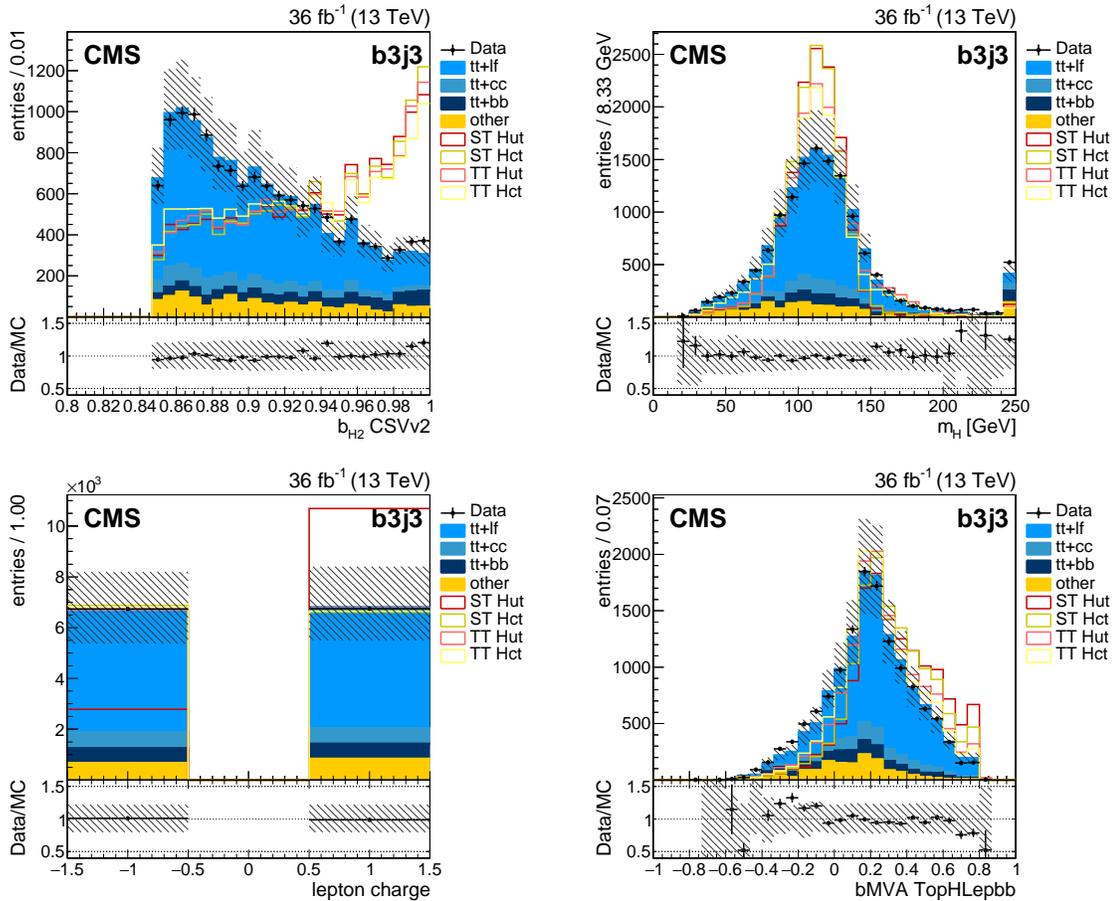
Figure 6.6: Illustration of the simulation to data agreement for four of the BDT input variables in the *b3j3* category. The total uncertainty, representing the quadratic sum of all statistical and systematic uncertainties as summarised in Section 6.3, is indicated by the shaded area. The individual signal contributions are scaled such that their normalisation matches the total number of simulated background events.

Event discriminant

The final BDT event discriminants for the four *Hut* and five *Hct* categories are shown in Figures 6.7 & 6.8 respectively. The tendency for overtraining of the event BDT discriminants is negligible, as illustrated in Appendix A. No significant deviations in data from simulations are visible in the event discriminants.
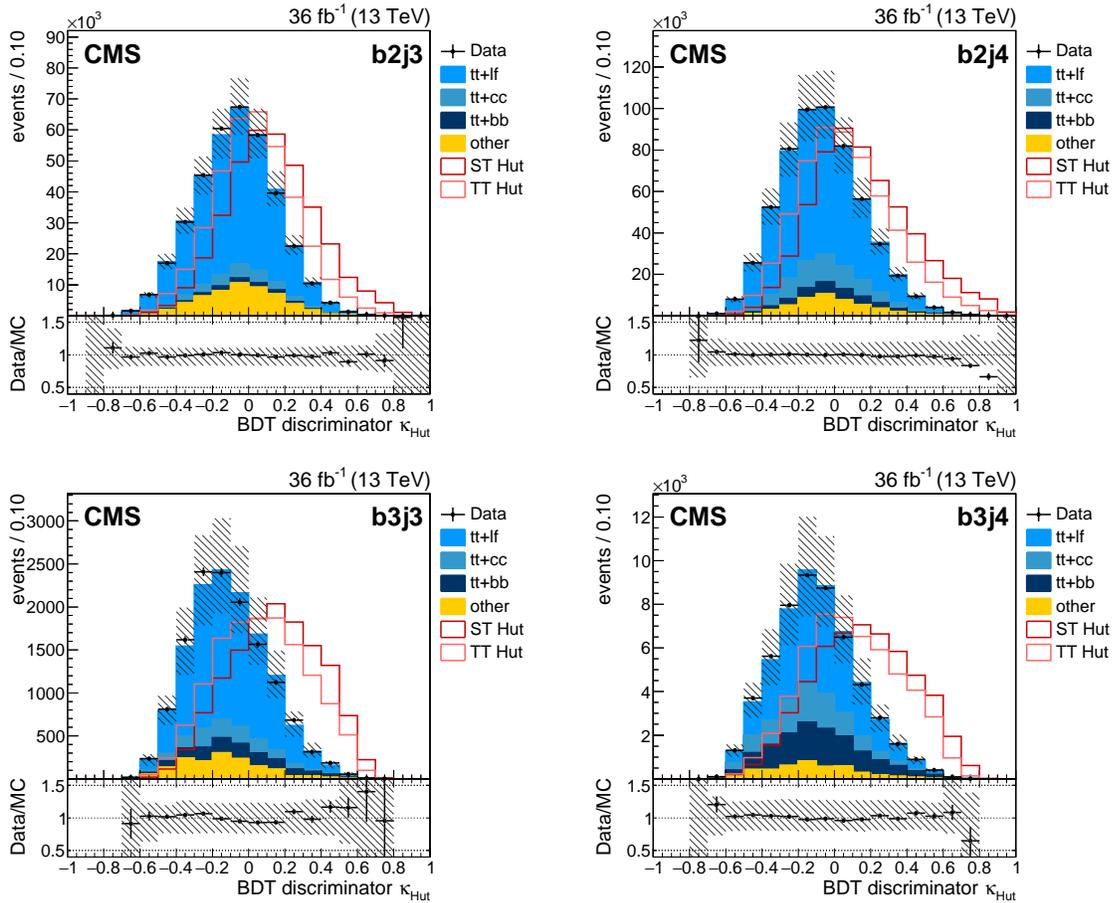
Figure 6.7: BDT output distributions for *Hut* signal in the four considered event categories. The total uncertainty, representing the quadratic sum of all statistical and systematic uncertainties as summarised in Section 6.3, is indicated by the shaded area. The individual signal contributions are scaled such that their normalisation matches the total number of simulated background events.
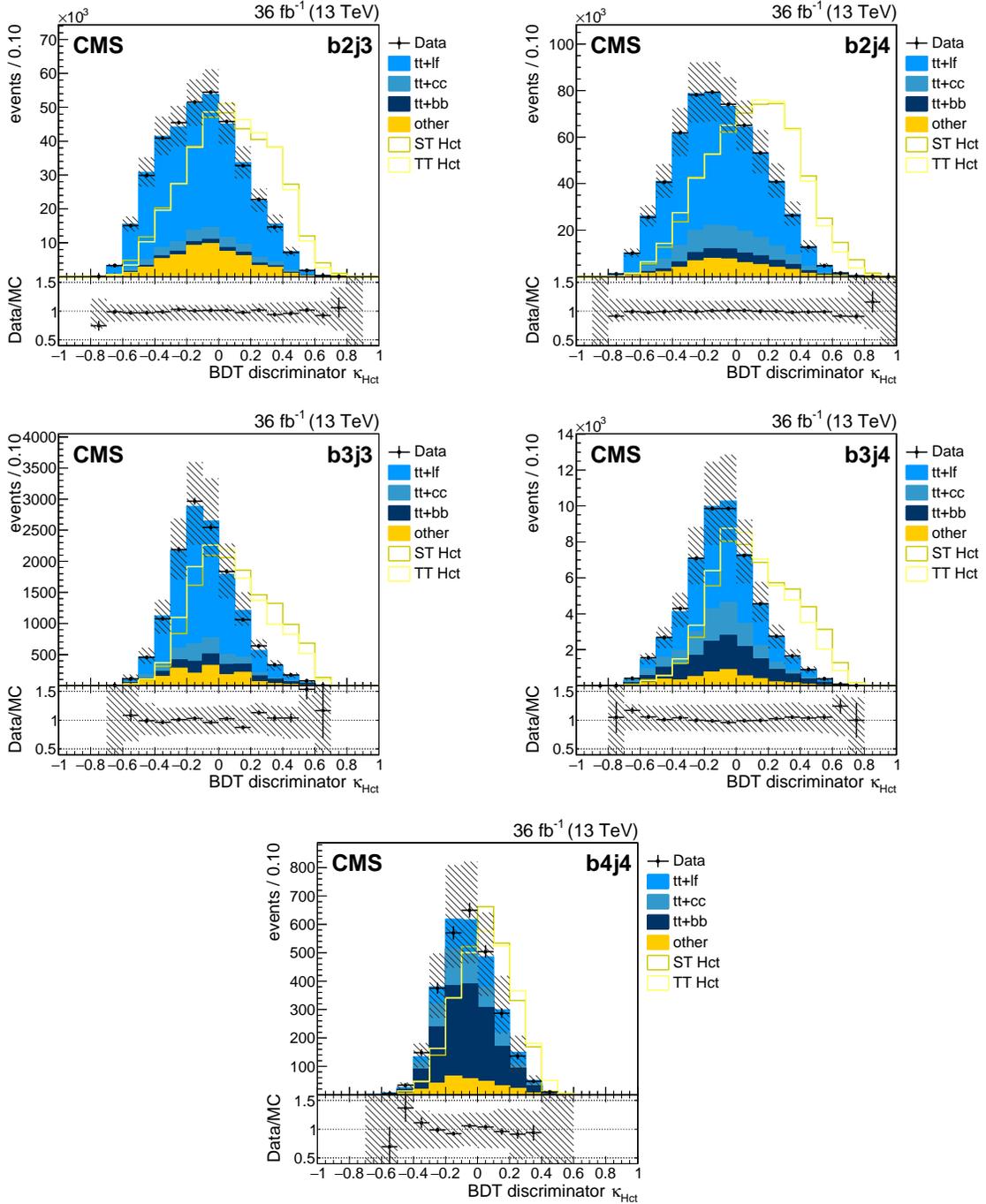
Figure 6.8: BDT output distributions for *Hct* signal in the five considered event categories. The total uncertainty, representing the quadratic sum of all statistical and systematic uncertainties as summarised in Section 6.3, is indicated by the shaded area. The individual signal contributions are scaled such that their normalisation matches the total number of simulated background events.

6.3   Systematic uncertainties

Various effects, either theoretical or experimental, can affect the shape and/or normalisation of the distributions of observables. The effect of systematic uncertainties is studied by varying their sources by one standard deviation ($\pm\sigma$). All sources of uncertainties listed below are treated as uncorrelated, unless mentioned otherwise. Table 6.4 summarises the relative changes in event yields in each category for every systematic source listed below for the dominating background, namely semi-leptonic t$\bar{\text{t}}$+jets events. In event categories with exactly two b-tagged jets, the ME QCD scale & PS scale uncertainty are dominant, whereas categories with high b-tagged jet multiplicity are dominated by the CSVv2 shape reweighting uncertainty.

Table 6.4: Overview of the relative changes in event yields for every source of systematic uncertainty in each category for the total t$\bar{\text{t}}$+jets background, as well as its statistical uncertainty. The total systematic uncertainty is given as the quadratic sum of all individual uncertainties.

Source of unc.	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
Int. luminosity	2.5%	2.5%	2.5%	2.5%	2.5%
Cross section	5.6%	5.6%	5.6%	5.6%	5.6%
t$\bar{\text{t}}$+c$\bar{\text{c}}$ & t$\bar{\text{t}}$+b$\bar{\text{b}}$ rates	50%	50%	50%	50%	50%
Pileup	−0.36% +0.35%	−0.13% +0.13%	+0.28% −0.28%	+0.66% −0.61%	+1.7% −1.5%
Lepton SF	+0.33% −0.33%	+0.36% −0.36%	+0.33% −0.33%	+0.37% −0.37%	+0.39% −0.39%
JES	−0.46% +0.53%	+4.9% −4%	+0.69% +0.96%	+5.2% −4.4%	+7.9% −7.3%
JER	−0.21% +0.094%	+0.3% +0.27%	+0.41% +0.73%	+0.26% +0.52%	−1.1% −1.5%
ME QCD scale & PS scale	+12% −12%	+14% −13%	+13% −12%	+15% −13%	+20% −16%
ME-PS matching scale	−0.9% +0.91%	+2.7% −1.7%	+2.6% +3.5%	+3.4% −2.1%	+4.4% −4%
PDF	+0.092% −0.046%	+0.097% −0.086%	+0.28% −0.5%	+0.073% −0.068%	+0.092% −0.11%
Underlying event	+0.074% −0.62%	+1.6% +0.83%	+0.24% +0.0038%	+1.8% +0.85%	+0.71% −0.38%
CSVv2 shape	+7.9% −6.1%	+7.2% −5.6%	+21% −17%	+18% −15%	+27% −22%
Total systematic uncertainty	+13% −13%	+14% −13%	+21% −18%	+17% −16%	+25% −20%
Statistical uncertainty	± 0.55%	± 0.43%	± 2.5%	± 1.3%	± 5.2%

- **Integrated luminosity.** The integrated luminosity of the recorded data is measured with an uncertainty of 2.5% [105]. This systematic uncertainty is correlated across all samples.

- **Cross sections of background processes.** In this analysis the dominating background is the production of top-quark pairs. The systematic uncertainty on its cross section is calculated as $^{+5.6\%}_{-6.1\%}$ to NNLO accuracy [108] (see Table 5.1). Two additional rate uncertainties[1] of 50% are assigned to the heavy flavoured t$\bar{\text{t}}$+c$\bar{c}$ & t$\bar{\text{t}}$+b$\bar{\text{b}}$ samples respectively. This number is motivated by the typical uncertainty of the NLO QCD corrections and available measurement results. CMS experimental results have measured the t$\bar{\text{t}}$+b$\bar{\text{b}}$ cross section, which was 50% higher than the MADGRAPH prediction [120]. Theoretical cross section calculations for the t$\bar{\text{t}}$+b$\bar{\text{b}}$ cross section range uncertainties from 20% to 100% [121, 122].

  All other SM processes are considered sub-dominant and their cross sections, calculated to at least NLO accuracy, will be treated as 100% correlated with the same conservative systematic uncertainty, being 10%.

- **Pileup reweighting.** Varying the minimum bias cross section, used to calculate the pileup distribution in data, by ± 4.6% [116] gives a systematic shift in the number of pileup events distribution as demonstrated in Section 5.2.4. The relative change in event yields is given in Table 6.5.

Table 6.5: Relative event yield changes due to pileup uncertainty.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
ST Hct	−0.63% +0.61%	+0.018% −0.04%	−0.75% +0.7%	−0.4% +0.39%	−0.71% +0.78%
TT Hct	−0.4% +0.4%	−0.2% +0.17%	−0.28% +0.22%	−0.4% +0.42%	+0.43% −0.31%
ST Hut	−0.58% +0.57%	−0.07% +0.04%	−0.63% +0.56%	−0.28% +0.29%	+1.2% −1.2%
TT Hut	−0.46% +0.47%	−0.25% +0.24%	−0.52% +0.47%	−0.35% +0.36%	+0.84% −0.74%
t$\bar{\text{t}}$+b$\bar{\text{b}}$	−0.27% +0.27%	−0.2% +0.21%	−0.17% +0.13%	−0.33% +0.38%	+0.33% −0.26%
t$\bar{\text{t}}$+c$\bar{c}$	−0.28% +0.28%	−0.12% +0.11%	−0.18% +0.19%	+0.5% −0.43%	+1.6% −1.4%
t$\bar{\text{t}}$+lf	−0.37% +0.35%	−0.12% +0.12%	+0.4% −0.4%	+1.1% −1.1%	+5.4% −5%
other	−0.18% +0.38%	+0.27% −0.25%	+1.8% −1.7%	+0.32% −0.46%	+2.5% −2.5%

- **Lepton scale factors.** The systematic uncertainties are related to the measured uncertainties on the three components of the lepton scale factors (Equation (5.5)): tracking & object reconstruction, isolation and trigger scale factors. The relative change in event yields is given in Table 6.6.

- **Jet Energy Scale (JES).** The uncertainty on the jet energy scale is estimated according to [89, 90]. The JES uncertainty requires a re-calculation of all jet-

---

[1]These uncertainties are not shown in the uncertainty bands on all plots, neither in Tables 6.1, 6.2 and 6.4.

Table 6.6: Relative event yield changes due to lepton scale factor uncertainties.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
ST Hct	+0.34% −0.34%	+0.35% −0.35%	+0.33% −0.33%	+0.35% −0.35%	+0.38% −0.38%
TT Hct	+0.36% −0.36%	+0.38% −0.38%	+0.34% −0.34%	+0.37% −0.37%	+0.38% −0.38%
ST Hut	+0.34% −0.34%	+0.36% −0.36%	+0.33% −0.33%	+0.36% −0.35%	+0.36% −0.36%
TT Hut	+0.35% −0.35%	+0.38% −0.38%	+0.34% −0.34%	+0.37% −0.36%	+0.38% −0.38%
t̄t+bb̄	+0.35% −0.35%	+0.38% −0.38%	+0.34% −0.34%	+0.38% −0.38%	+0.4% −0.4%
t̄t+cc̄	+0.34% −0.34%	+0.37% −0.37%	+0.34% −0.34%	+0.37% −0.36%	+0.39% −0.39%
t̄t+lf	+0.33% −0.33%	+0.36% −0.36%	+0.33% −0.33%	+0.36% −0.36%	+0.38% −0.38%
other	+0.98% −0.97%	+1.2% −1.1%	+0.92% −0.91%	+0.97% −0.95%	+0.94% −0.92%

and $\not{E}_T$-related kinematical observables for the up- and down variation. This has an effect on the shape of the BDT and the event acceptance, as the number of selected jets may be altered. The relative change in event acceptance is given in Table 6.7. The 4-jet categories are affected more by this uncertainty. This is understood by considering that varying the JES upwards, the p_T of all jets varies upwards, making it possible for additional softer jets to survive the minimal requirement of $p_T > 30$ GeV. In this way, events in 3-jet categories transfer to 4-jet categories. As there are less events with exactly two jets than with three jets, the transfer of events into the 3-jet category is less pronounced, which could even result in a decrease of number of events in 3-jet categories. A similar argumentation is valid for the downwards variation of JES.

- **Jet Energy Resolution (JER).** The transverse momenta of jets are altered according to the JER uncertainty as mentioned in Table 3.1. The relative change in event yields is given in Table 6.8.

- **ME QCD scale & PS scale.** The value of the factorisation scale μ_R and renormalisation scale μ_F (see Section 3.1.1) used in the simulations may impact both the normalisation as the shape of the final BDT discriminant. By varying μ_R twice (by a factor 2 and 0.5) at fixed μ_F and varying μ_F twice (by a factor 2 and 0.5) at fixed μ_R, four uncertainty variations at matrix element-level are obtained. Two more variations at matrix element-level are retrieved by varying μ_F and μ_R simultaneously by a factor 2 and 0.5. At parton shower level, the scale uncertainties manifest themselves as an uncertainty [77] on α_s in the initial and final state radiation that are controlled by PYTHIA. To estimate the variations at parton shower level, four dedicated samples are used where α_s has been varied up and down at ISR and FSR from the nominal value of $\alpha_s = 0.115^{+0.0145}_{-0.0142}$. The

Table 6.7: Relative event yield changes due to JES uncertainties.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
ST Hct	+1.3% −1.3%	+4.4% −4.8%	+4.2% −2.3%	+6.5% −5.7%	+8.6% −7.1%
TT Hct	−1.1% +0.68%	+4.4% −3.8%	+0.75% −0.19%	+5.2% −5.2%	+7.3% −3.5%
ST Hut	+0.91% −1.8%	+5.4% −4.7%	+3.4% −1.7%	+5.5% −4.7%	+3.3% −11%
TT Hut	−0.81% +0.46%	+4.1% −3.6%	+0.14% −0.29%	+5.6% −5.1%	+7.1% −4.6%
$t\bar{t}$+b$\bar{b}$	−2.5% +2.6%	+3.3% −2.9%	+1.6% +2.2%	+5.6% −3.3%	+9.1% −5.8%
$t\bar{t}$+$c\bar{c}$	−2% +1.8%	+4.1% −3.2%	−0.92% +1.2%	+4.5% −3.3%	+3.7% −8.2%
$t\bar{t}$+lf	−0.26% +0.35%	+5.2% −4.2%	+0.8% +0.77%	+5.3% −5.2%	+8.2% −10%
other	+3.8% −4.1%	+5.7% −5.3%	+8.3% −2.2%	+6.3% −4.7%	+0.35% −4.7%

Table 6.8: Relative event yield changes due to JER uncertainties.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
ST Hct	−0.16% +0.17%	+0.12% −0.88%	−0.14% +1%	−0.14% +0.21%	+0.003% +2.2%
TT Hct	−0.3% −0.28%	+0.14% +2.4%	+0.49% +2.6%	−0.52% −0.39%	+1.2% +3.8%
ST Hut	−0.58% −0.26%	+0.043% +0.25%	+0.44% +1%	+0.23% −0.28%	−5.1% −3.3%
TT Hut	−0.44% −0.12%	+0.19% +1.3%	+0.029% +1.3%	−0.22% −0.2%	+0.25% +2.8%
$t\bar{t}$+b$\bar{b}$	+0.072% +0.2%	−0.17% +0.12%	−0.11% +0.55%	+1.5% +1.4%	+0.61% +0.37%
$t\bar{t}$+$c\bar{c}$	−0.11% −0.28%	+0.33% +0.27%	−1.2% −0.82%	+0.82% +0.76%	−3.7% −4.9%
$t\bar{t}$+lf	−0.23% +0.12%	+0.33% +0.28%	+0.69% +0.95%	−0.41% +0.11%	−3.2% −3.5%
other	+0.023% +0.034%	+0.13% −0.35%	−2% −0.31%	+1% +0.17%	+0.79% −4.6%

total uncertainty is then estimated by taking the envelope[2] on the final BDT discriminant of these 10 uncertainty sources. This uncertainty is only considered for the t$\bar{\text{t}}$ background. The relative change in event yields is given in Table 6.9. An increase (decrease) in the energy scale results in a decreased (increased) α_s, which in turn leads to less (more) additional parton radiations and eventually less (more) jets. This tendency is reflected in the table.

Table 6.9: Relative event yield changes due to ME QCD scale & PS scale.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
t$\bar{\text{t}}$+b$\bar{\text{b}}$	+14% −13%	+16% −14%	+14% −14%	+17% −14%	+20% −15%
t$\bar{\text{t}}$+c$\bar{\text{c}}$	+13% −12%	+15% −14%	+17% −15%	+16% −14%	+21% −17%
t$\bar{\text{t}}$+lf	+12% −12%	+14% −13%	+12% −12%	+14% −13%	+19% −20%

- **ME-PS matching scale.** High-p$_T$ radiations in the matching of matrix-element partons and parton showers are regulated in `POWHEG` by damping real emissions with a factor $h^2_{damp}/(p^2_T + h^2_{damp})$ [77]. The nominal value for the h_{damp} parameter is set to $1.581^{+0.66}_{-0.59}m_t$, with the top quark mass $m_t = 172.5$ GeV. Dedicated samples have been made with the up- and down-variation of the h_{damp} parameter. The relative change in event yields is given in Table 6.10. The tendencies in the change of event yields can be understood as follows. Increasing the h_{damp} parameter within its uncertainty correspond to a higher p$_T$ of the real emissions. Hard gluons from such emissions are more likely to split into a pair of b quarks, increasing the relative production of t$\bar{\text{t}}$+b$\bar{\text{b}}$ with respect to t$\bar{\text{t}}$+lf and t$\bar{\text{t}}$+c$\bar{\text{c}}$. Moreover, as the performance of b-tagging increases for jets approaching a p$_T$ of 100 GeV, categories which are sensitive to high b-tag multiplicities are more sensitive to the variation of the ME-PS matching scale. A similar argumentation is valid for the downwards variation of the h_{damp} parameter.

Table 6.10: Relative event yield changes due to ME-PS matching scale uncertainties.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
t$\bar{\text{t}}$+b$\bar{\text{b}}$	+0.53% −0.66%	+4.6% −4.5%	+2.3% −4.2%	+2.8% −5.2%	+12% −3.1%
t$\bar{\text{t}}$+c$\bar{\text{c}}$	−1.3% −1.6%	+4.5% −3.1%	+4.1% +2.3%	+6.5% −2.6%	−8.8% −4.9%
t$\bar{\text{t}}$+lf	−0.91% +1.2%	+2.3% −1.2%	+2.5% +4.6%	+2.5% −0.69%	−3.8% −5.6%

---

[2]The envelope considers each bin of the distribution individually and assumes as value the systematic variation for which the deviation from the nominal distribution is maximal.

- **Parton density functions (PDFs).** The uncertainties in PDFs are propagated to the final BDT discriminants as event weights, which are calculated by `LHAPDF` [123] as

$$w^j = \frac{PDF^j(x_1, f_1, Q) \cdot PDF^j(x_2, f_2, Q)}{PDF^0(x_1, f_1, Q) \cdot PDF^0(x_2, f_2, Q)} \text{ for } 0 \leq j \leq 2n, \qquad (6.1)$$

where x_1 and x_2 represent the momentum fractions of the two partons resolving from the protons, f_1 and f_2 their flavours, Q the interaction scale and n the number of pdf uncertainty parameters. Ending up with $2n + 1$ variations of the final BDT discriminant, the envelope of these variations is taken as the pdf uncertainty. For this analysis, the pdf-uncertainty is only considered for the t$\bar{\text{t}}$ background. The nominal pdf-set for the used t$\bar{\text{t}}$ sample is `NNPDF30_nlo_as_0118` and the set of error PDFs is taken from `PDF4LHC15_nlo_100` [124]. The relative change in event yields is given in Table 6.11.

Table 6.11: Relative event yield changes due to PDF uncertainties.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
t$\bar{\text{t}}$+b$\bar{\text{b}}$	+0.11% / −0.16%	+0.099% / −0.19%	+0.2% / −0.31%	+0.11% / −0.16%	+0.1% / −0.15%
t$\bar{\text{t}}$+$c\bar{c}$	+0.065% / −0.047%	+0.17% / −0.28%	+0.071% / −0.086%	+0.041% / −0.043%	+0.082% / −0.067%
t$\bar{\text{t}}$+lf	+0.094% / −0.042%	+0.083% / −0.043%	+0.32% / −0.58%	+0.071% / −0.041%	+0.073% / −0.059%

- **Underlying event.** Two dedicated samples, representing variations of the underlying event tune CUETP8M2T4 of the t$\bar{\text{t}}$ background sample, are used to estimate the underlying event uncertainty. The relative change in event yields is given in Table 6.12. The propagation of changes in the UE tune to the final event reconstruction and the interpretation of their effects on reconstructed objects is not straightforward. The changes in event yields in Table 6.12 are, however, consistent with the statistical uncertainties of the available samples.

Table 6.12: Relative event yield changes due to underlying event uncertainties.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
t$\bar{\text{t}}$+b$\bar{\text{b}}$	+0.79% / −0.56%	+2.1% / +2.3%	−3.7% / −3.6%	+3.3% / +0.98%	+4.3% / +7.4%
t$\bar{\text{t}}$+$c\bar{c}$	+1.1% / −0.79%	+2.2% / +0.32%	−1.7% / −2%	+2.7% / −0.5%	−6.4% / −6.5%
t$\bar{\text{t}}$+lf	−0.041% / −0.6%	+1.4% / +0.81%	+0.97% / +0.72%	+0.9% / +1.3%	−2.3% / −15%

- **CSVv2 discriminant shape reweighting.** The reweighting of the CSVv2 discriminant shape is based on the method as described in Section 3.3.5 and [98]. Summarized, there are three sources of uncertainty contributing to the measurement of the scale factors: JES, purity of the sample and statistics. The uncertainty associated to the JES is evaluated simultaneously with the JES uncertainty and is 100% correlated to the JES uncertainties. The purity uncertainty is subdivided into two uncorrelated uncertainties, pertaining respectively to the purity of light flavoured and heavy flavoured jet contributions in the scale factor measurements. A $\pm 1\sigma$ shift in each of the two purity uncertainties corresponds to a higher or lower contamination of these flavours in the scale IterativeFit procedure. The statistical uncertainties consist of four uncorrelated sources of uncertainties, two for heavy flavour and two for light flavour. One of these two uncertainties in each flavour category corresponds to a shift consistent with statistical uncertainties on the scale factors. The second uncertainty is propagated in a non-trivial way, such that only the upper and lower ends of the CSVv2 distribution are affected with respect to the center of the distribution. The uncertainty on charm jet scale factors is obtained from the uncertainty on the heavy flavour scale factors, doubling it in size and constructing two nuisance parameters to control the charm flavour scale factors, treating them as independent uncertainties. In total, there are eight uncertainty sources affecting the shape of the CSVv2 distribution, which are all considered as uncorrelated nuisance parameters in this analysis. Table 6.13 sums up (quadratically) the relative changes in event yields caused by the eight different sources (not including the JES effect) of CSVv2 shape reweighting uncertainties. There exists an asymmetry in the upward and downward fluctuations of this uncertainty. This is due to the nature of the determination of the systematic uncertainties, especially the light-flavour contamination uncertainty. As the distribution of the CSVv2 discriminant for light-flavoured jets is asymmetric with respect to the Medium working point, i.e. more light-flavoured jets fail that working point cut than pass it, the variation of the contamination of this flavour is expected to have an asymmetric effect.

- **Top quark p_T reweighting.** The shape of the p_T distribution of individual top quarks in data is softer than predicted by simulations. Dedicated scale factors to correct this behaviour have been measured in a differential $t\bar{t}$ cross section measurement [125]. This top quark p_T shape correction is considered only as a systematic uncertainty for $t\bar{t}$ events, where the $+1\sigma$ variation is obtained by applying an event scale factor as $\sqrt{SF(top) \times SF(anti-top)}$. The scale factors as function of the generated (anti-) top-p_T are calculated as $SF(p_T) = \exp\left(0.0615 - 0.0005 \times p_T\right)$. The -1$\sigma$ variation is set as the nominal value. As the reweighting of the top quark p_T should only affect the top-p_T shape, the applied $+1\sigma$ weights are averaged over all selected events.

Table 6.13: Relative event yield changes due to CSVv2 shape uncertainty as a quadratic sum of the eight uncorrelated sources.

Process	b2j3-$^{+1\sigma}_{-1\sigma}$	b2j4-$^{+1\sigma}_{-1\sigma}$	b3j3-$^{+1\sigma}_{-1\sigma}$	b3j4-$^{+1\sigma}_{-1\sigma}$	b4j4-$^{+1\sigma}_{-1\sigma}$
ST Hct	+4.9% −4.1%	+3.1% −3.1%	+14% −10%	+12% −9%	+23% −17%
TT Hct	+5.6% −4.6%	+3.7% −3.8%	+14% −10%	+11% −8.7%	+24% −18%
ST Hut	+5.1% −4.2%	+3.4% −3.2%	+14% −9.9%	+12% −9%	+19% −15%
TT Hut	+5.9% −4.8%	+3.6% −3.5%	+14% −10%	+12% −8.9%	+22% −17%
t$\bar{\text{t}}$+b$\bar{\text{b}}$	+5.9% −4.8%	+2.6% −2.7%	+14% −10%	+11% −8.1%	+18% −14%
t$\bar{\text{t}}$+c$\bar{c}$	+7.6% −6%	+6.8% −5.6%	+26% −19%	+23% −18%	+48% −32%
t$\bar{\text{t}}$+lf	+8% −6.1%	+7.7% −5.9%	+22% −18%	+22% −18%	+53% −45%
other	+10% −8.1%	+8.1% −6.5%	+24% −18%	+17% −14%	+25% −19%

6.4 Limit extraction results and discussion

The claim for the presence of a new physics signal (or lack thereof) needs to be supported by a well founded statistical measure. A commonly used formalism at LHC collaborations that defines such a statistical test is known as the CLs procedure [126]. A more refined version of the CLs procedure, as described in [127], is used in this analysis and is shortly introduced in Section 6.4.1. This limit setting formalism has been elaborated by the Higgs Combine Group based on the `RooStats` framework [128] and is used in this analysis to extract limits on the t → uH and t → cH branching ratios, based on a combined fit of the BDT discriminants in each event category. The results of the combined fit are discussed in Section 6.4.2.

6.4.1 Limit setting formalism

The assessment of putting limits on a new physics process, which in terms of the statistical testing procedure is referred to as the signal plus background hypothesis, is done with respect to the background-only hypothesis. Limits are extracted from an observable x for which the hypotheses predict different outcomes in an experiment. From the hypotheses′ predictions of x, a probability model is derived. That probability model serves as the basis on which a test statistic is constructed. By means of pseudo-experiments, the distributions of the test statistic for the two hypotheses are generated. The value of the test statistic in the actual data is used to determine the level of confidence in the hypotheses, which is combined in the CLs procedure to determine exclusion limits at a certain confidence level for the signal plus background hypothesis.

Constructing a test statistic

When constructing a histogram with N bins[3] of the measured observable x, the expectation value for the number of events n in the i-th bin can be written as

$$E[n_i] = \alpha S_i + B_i, \tag{6.2}$$

where S_i and B_i represent the number of predicted events in respectively the signal and background model for the i-th bin. The value α represents the overall signal strength, such that the expectation value for $\alpha = 0$ corresponds to the background-only hypothesis and $\alpha = 1$ the nominal signal plus background hypothesis. The predicted events S_i and B_i depend on the probability density functions f_S and f_B of the variable x as

$$S_i = S_{tot} \int_{x_{\text{lower},i}}^{x_{\text{upper},i}} f_S(x, \boldsymbol{\theta}_S) dx, \tag{6.3}$$

$$B_i = B_{tot} \int_{x_{\text{lower},i}}^{x_{\text{upper},i}} f_B(x, \boldsymbol{\theta}_B) dx, \tag{6.4}$$

where S_{tot} and B_{tot} are the total number of predicted events for their respective models and θ_S and θ_B are parameters that define the shape of the probability density functions. The parameters $\boldsymbol{\theta} = (B_{tot}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_S)$ are considered as *nuisance parameters*. Nuisance parameters are caused by the uncertainties of the analysis and are defined such that normalisation uncertainties only affect B_{tot} and shape-only uncertainties are represented by $\boldsymbol{\theta}_B$ and/or $\boldsymbol{\theta}_S$. When an uncertainty affects both rate and shape, the uncertainty is first normalised to the nominal B_{tot} such that the shapes-only uncertainties are treated as $\boldsymbol{\theta}_B$ and/or $\boldsymbol{\theta}_S$, and the normalisation factor is absorbed in B_{tot}.

Given the distribution of observable x according to a measured data set $\mathcal{D}$ in different channels, the likelihood of the hypothesis to describe this distribution is given as a product of Poisson probabilities

$$L(\mathcal{D}|\alpha, \boldsymbol{\theta}) = \prod_c^{\text{channels}} \left[\prod_{j=1}^{N} \frac{(\alpha S_j + B_j)^{n_j}}{n_j!} e^{-(\alpha S_j + B_j)} \right] \cdot \prod_a^{\text{Aux}} f_a(\theta_{a,m}|\theta_a), \tag{6.5}$$

where $f_a(\theta_{a,m}|\theta_a)$ represents the probability density function corresponding to a constrained value $\theta_{a,m}$ of the nuisance parameter θ_a. This constrained value $\theta_{a,m}$ is an estimate of the true value of θ_a from an auxiliary measurement, which is assessed as a log-normal probability distribution for normalisation uncertainties. Shape uncertainties are auxiliary measurements modelled by template morphing techniques [129], which quadratically interpolate between the up- and down systematic templates. Linear interpolation is used beyond the up- and down templates. The profile likelihood

---

[3]The case where $N = 1$ is defined as an unbinned analysis. Limits extracted from such an analysis are referred to as *cut-and-count* limits.

ratio is accordingly defined as

$$\lambda(\mathcal{D}|\alpha) = \frac{L(\mathcal{D}|\alpha, \hat{\hat{\boldsymbol{\theta}}})}{L(\mathcal{D}|\hat{\alpha}, \hat{\boldsymbol{\theta}})}, \tag{6.6}$$

where $\hat{\alpha}$ and $\hat{\boldsymbol{\theta}}$ are maximum likelihood estimators of the true values of α and $\boldsymbol{\theta}$. The numerator represents the maximum likelihood value for a specified signal strength α, such that $\hat{\hat{\boldsymbol{\theta}}}$ is the conditional maximum likelihood estimator of $\boldsymbol{\theta}$, given α. The test statistic, based upon which upper limits are derived, is introduced as

$$q_\alpha = -2\ln\lambda(\mathcal{D}|\alpha), \tag{6.7}$$

for $\alpha \geq \hat{\alpha}$. Since $\alpha < \hat{\alpha}$ represent signal strengths that are weaker than its maximum likelihood value, those need not to be considered for extracting upper limits on α. High values of q_α correspond to a higher incompatibility between $\mathcal{D}$ and signal strength α.

The distribution of q_α is sampled for the signal plus background and background-only hypotheses by means of simulated pseudo-experiments, which can be CPU intensive. When the expected number of events is large enough however, an asymptotic approximation of the pseudo-experiments can be used as demonstrated in [127]. In this approximation, the set of pseudo-experiments is replaced by a single representative data set, called the *Asimov data set*.

CLs procedure

Knowing the value $q_{\alpha,\mathrm{obs}}$ of the actual measured data allows to define the probability of finding a data set with equal or greater incompatibility to the signal plus background hypothesis as

$$p_{S+B} = \int_{q_{\alpha,\mathrm{obs}}}^\infty f_{S+B}(q_\alpha|\alpha)dq_\alpha, \tag{6.8}$$

where $f_{S+B}(q_\alpha|\alpha)$ is the probability density function describing the q_α distribution of signal plus background hypothesis. This value is also referred to as the p-value of a hypothesis. Similarly, the p-value for the background-only hypothesis becomes

$$1 - p_B = \int_{-\infty}^{q_{\alpha,\mathrm{obs}}} f_B(q_\alpha|0)dq_\alpha. \tag{6.9}$$

The CLs procedure defines a confidence level (CL) of 95% of excluding the signal plus background model if the CLs value

$$CLs = \frac{p_{S+B}}{1 - p_B}, \tag{6.10}$$

is below 0.05. By defining the confidence level in the exclusion limits as such, one prevents downward fluctuations in the background model that would exclude models

for which the observable has little sensitivity. Such downward fluctuations are especially dangerous if the background-only and signal plus background hypotheses have similar test statistic distributions. In this case one should not expect to be able to exclude the signal plus background hypothesis, which is prevented in the CLs definition as then the CLs value would be close to unity.

Expected exclusion limits are obtained by replacing the $q_{\alpha,\text{obs}}$ in Equations (6.8) and (6.9) with the value of the test statistic that represents the 50% quantile of $f_B(q_\alpha|0)$. Similarly the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands for the expected limits are determined from the test statistic values that give the 68% and 95% quantiles of $f_B(q_\alpha|0)$.

6.4.2   Results and discussion

The limit setting procedure as introduced above is applied to this analysis. Results are obtained in the event categories separately, to indicate the sensitivity gained from each category, as well as for the combination of all categories.

In order for the maximum likelihood fit to converge, a couple of simplifications in the treatment of nuisance parameters is necessary. It was therefore decided to consider systematic uncertainties that have a negligible contribution as a flat normalisation uncertainty instead of a shape uncertainty. The uncertainties that are treated like this are: pileup reweighting, lepton scale factors, Top-p_T reweighting, PDFs and UE. The rate uncertainty on the *other* process is decomposed into five uncorrelated uncertainties, each for one of the five event categories. This choice is based on the fact that the composition of the *other* process is different in each of the five categories. Finally, the 50% additional rate uncertainties for t$\bar{\text{t}}$+b$\bar{\text{b}}$ and t$\bar{\text{t}}$+c$\bar{\text{c}}$ are decorrelated for 3- and 4-jet categories. These rate uncertainties only have vague estimations from both theoretical and experimental sides and mostly correspond to 4-jet selections and are quite conservative. Taking this into account together with the difference in kinematics between 3- and 4-jet categories for t$\bar{\text{t}}$ events, this decorrelation choice is well motivated.

Signal strength

At first it can be noted that no signal is observed, neither for the *Hut* and *Hct* vertex. The maximum likelihood estimator of their signal strengths[4] μ ($= \hat{\alpha}$) is compatible with zero, as summarised in Figure 6.9. The *b2j4*, *b3j3*, *b3j4* and *b4j4* categories indicate equal sensitivity to the *Hct* signal, greatly surpassing the sensitivity in the *b2j3* category. The main sensitivity to the *Hut* signal comes from the *b2j3*, *b3j3* and *b3j4* categories, whereas the *b2j4* category have larger uncertainties on the fitted signal strength. In the control categories, it is clear that the *b2j3* category is more sensitive to the *Hut* than *Hct* vertex, which can be attributed to the single top component of the respecitve signals. In the *b2j4* category the situation is reversed, due to the fact

---

[4]The signal strengths are defined such that a signal strength of unity corresponds to the FCNC process occurring with a branching ratio of 0.01.

that the t$\bar{\text{t}}$ component of the *Hct* signal is dominant over the t$\bar{\text{t}}$ component of the *Hut* signal.
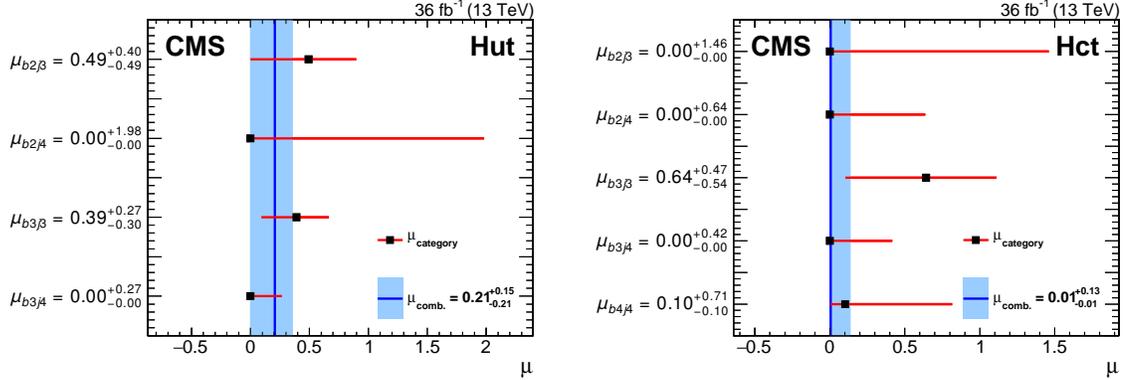
Figure 6.9: The maximum likelihood estimators for the signal strength per event category, as well as the combination, for κ_{Hut} (left) and κ_{Hct} (right).

Post-fit systematic uncertainties

The values of the maximum likelihood estimators for the nuisance parameters, representing the systematic uncertainties, are shown in Figure 6.10. The values obtained from the background-only and signal plus background fits are in good agreement, as expected since there is no signal observed in the data. Several of the nuisance parameters' uncertainties get constrained. These constraints are determined mostly by the *b2j4* and *b2j3* categories, which proves the use of these categories as control categories. Some estimations of the nuisance parameters' true values deviate from the expectation of the auxiliary measurement (Equation (6.6)), like the JER uncertainty for example.

The maximum likelihood estimations of the nuisance parameters are used to recreate the BDT event discriminant distributions in each category, reshaping and normalising the templates of the background processes accordingly. Figures 6.11 & 6.12 show, respectively for the *Hut* and *Hct* vertex, how well the maximum likelihood fitted parameters describe the data. Tables 6.14 and 6.15 summarises the event yields of the background processes, with their total uncertainty, obtained from the combined post-fit results in each category.
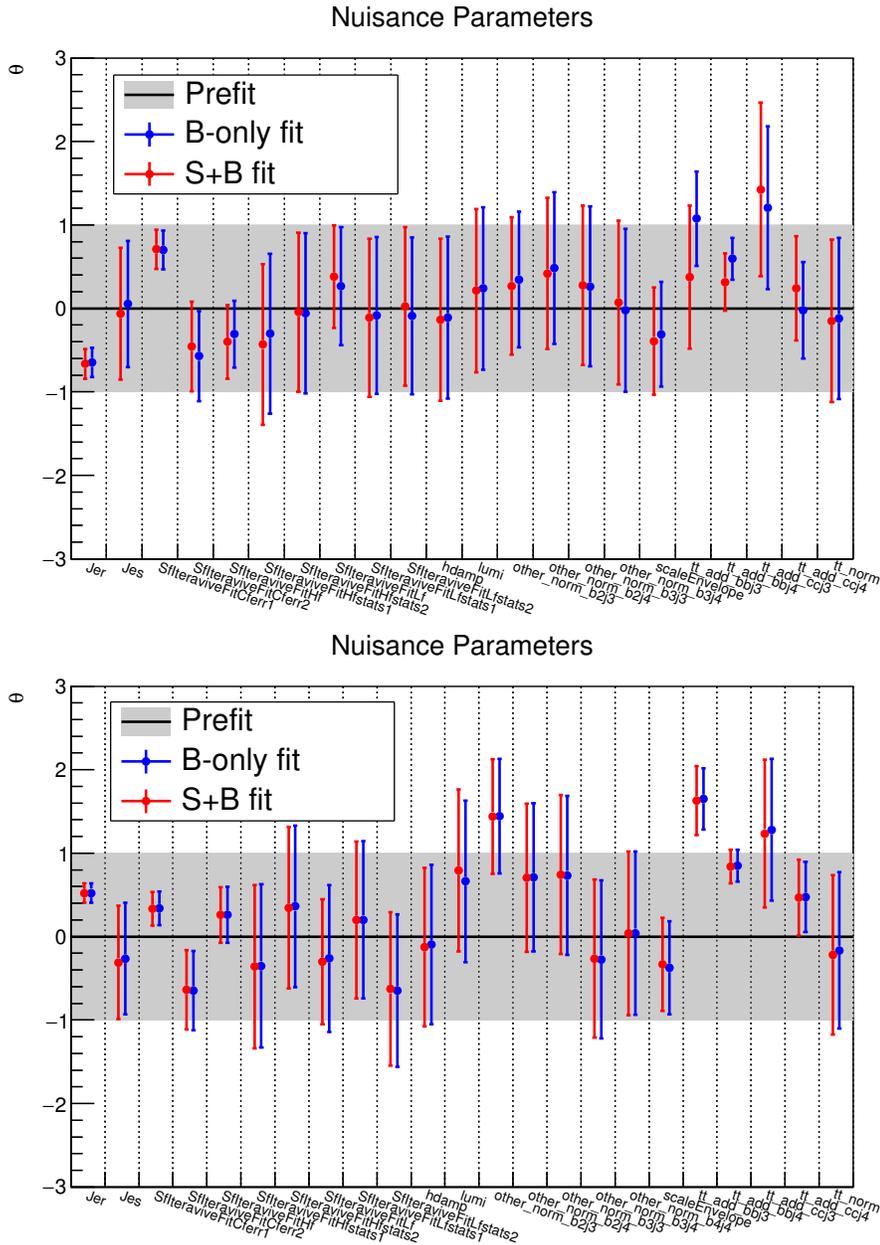
Figure 6.10: The maximum likelihood estimators of the nuisance parameters for κ_{Hut} (top) and κ_{Hct} (bottom) in the signal plus background (S+B) and background-only (B-only) hypotheses. From left to right, the nuisance parameters are: JER, JES, CSVv2 discriminant shape reweighting (8 sources), ME-PS matching scale, Luminosity, cross section of *other* processes (5 sources), ME QCD scale & PS scale, 50% additional rate for t̄t+b̄b and t̄t+c̄c (decorrelated in 3- and 4-jet categories), cross section of t̄t + *jets* processes.
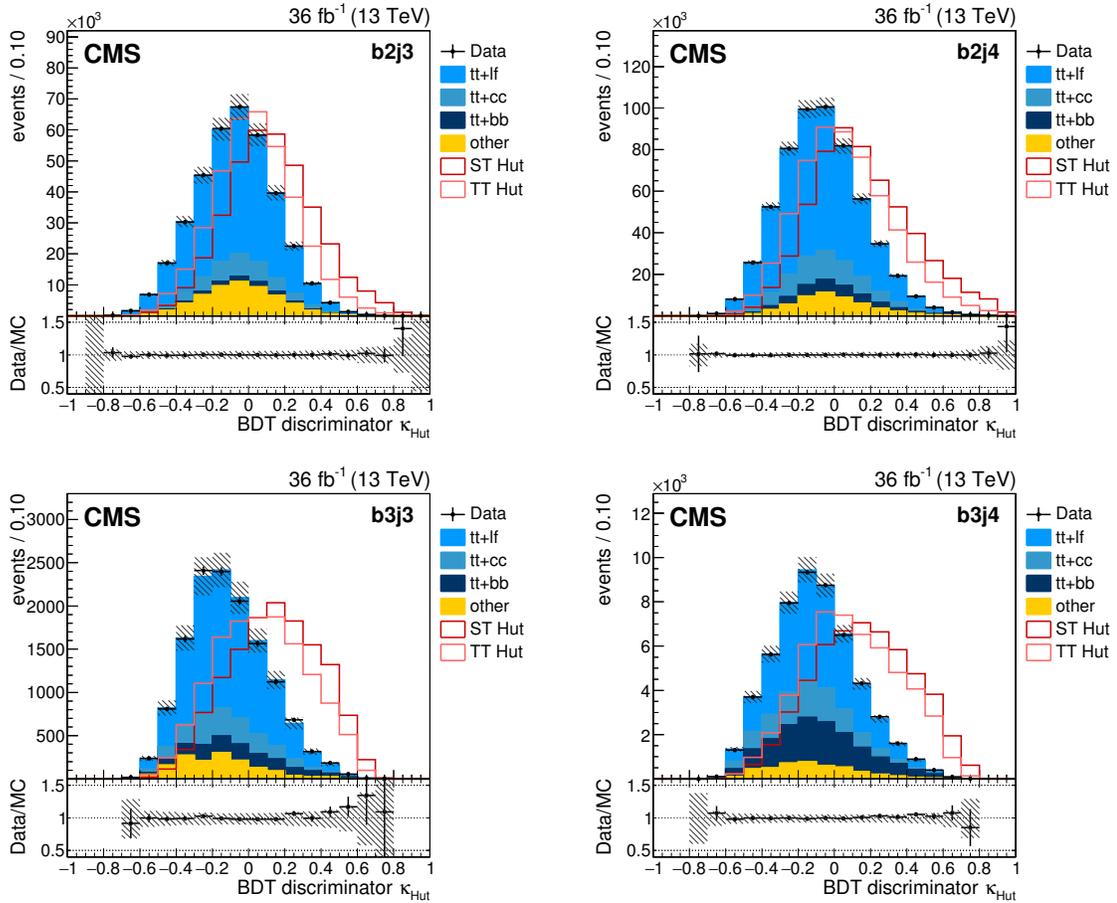
Figure 6.11: BDT event discriminants for the *Hut* vertex in the different event categories, where the background processes are reshaped according to the fitted distributions in the limit setting procedure. The signal processes are rescaled to match the background normalisation.
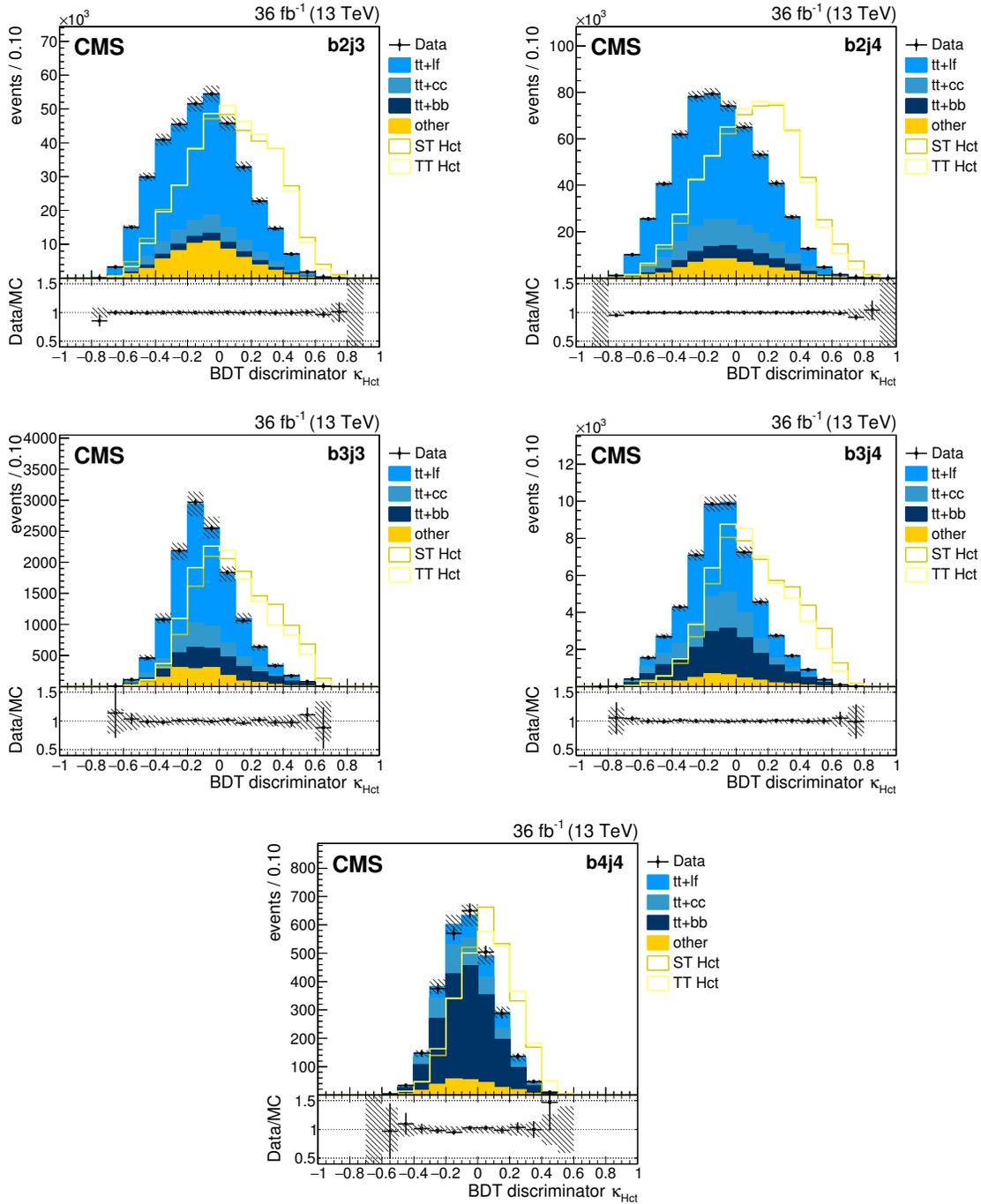
Figure 6.12: BDT event discriminants for the *Hut* vertex in the different event categories, where the background processes are reshaped according to the fitted distributions in the limit setting procedure. The signal processes are rescaled to match the background normalisation.

Table 6.14: Event yields in each category together with its total relative uncertainty as obtained from the maximum likelihood fit to data for the Hct vertex.

Process	b2j3 ± 1σ	b2j4 ± 1σ	b3j3 ± 1σ	b3j4 ± 1σ	b4j4 ± 1σ
Data	365 890	575 500	13 481	53 352	2 764
tt̄+bb̄	1 0176 ± 19%	34 174 ± 11%	1 367 ± 20%	12 897 ± 8.2%	1 517 ± 8.5%
tt̄+cc̄	33 210 ± 36%	102 186 ± 15%	1 674 ± 37%	12 280 ± 15%	521 ± 20%
tt̄+lf	258 679 ± 3.4%	385 395 ± 2.8%	8 349 ± 5.4%	24 083 ± 4.7%	383 ± 18%
other	62 887 ± 9.1%	52 134 ± 12%	1 742 ± 23%	3 513 ± 16%	262 ± 19%
Total	364 952 ± 4.6%	573 889 ± 3.2%	13 132 ± 7.3%	52 773 ± 4.4%	2 682 ± 6.9%

Table 6.15: Event yields in each category together with its total relative uncertainty as obtained from the maximum likelihood fit to data for the Hut vertex.

Process	b2j3 ± 1σ	b2j4 ± 1σ	b3j3 ± 1σ	b3j4 ± 1σ
Data	365 890	575 500	13 481	53 352
tt̄+bb̄	8 880 ± 41%	30 157 ± 17%	1 214 ± 42%	11 668 ± 15%
tt̄+cc̄	26 035 ± 43%	81 959 ± 22%	1 281 ± 45%	9 753 ± 23%
tt̄+lf	270 989 ± 5.1%	410 028 ± 4%	9 104 ± 7.4%	27 079 ± 6.4%
other	58 991 ± 11%	51 845 ± 12%	1 616 ± 22%	4 269 ± 18%
Total	364 895 ± 6.2%	573 989 ± 4.4%	13 215 ± 9.5%	52 769 ± 6.5%

One-dimensional limits

The limit setting procedure returns the limits on the signal cross sections. As the signal cross sections depend quadratically on the coupling strength, the interpretation of these limits in terms of coupling strength is straightforward. Similarly, using Equation (1.25), the limits can be translated to limits on the branching ratios of t → uH and t → cH.

The expected, and observed, exclusion limits on the *Hut* and the *Hct* coupling strengths, shown in Fig 6.13, are $\kappa_{Hut} \lesssim$ **0.16** and $\kappa_{Hct} \lesssim$ **0.18** at 95% CL, and $\kappa_{Hut} \lesssim$ **0.19** and $\kappa_{Hct} \lesssim$ **0.19** at 95% CL, respectively. The corresponding expected, and observed, branching ratio upper limits are $\mathcal{B}(\text{t} \rightarrow \text{uH}) \lesssim$ **3.4** $\times$ **10^{-3}** and $\mathcal{B}(\text{t} \rightarrow \text{cH}) \lesssim$ **4.4** $\times$ **10^{-3}**, and $\mathcal{B}(\text{t} \rightarrow \text{uH}) \lesssim$ **4.7** $\times$ **10^{-3}** and $\mathcal{B}(\text{t} \rightarrow \text{cH}) \lesssim$ **4.7** $\times$ **10^{-3}**, respectively. A summary of the expected and observed upper limits on the cross sections for the *Hut* and *Hct* vertex is presented in respectively Tables 6.16 & 6.17.

The upper limits on the *Hut* vertex are tighter than on the *Hct* vertex. This is mainly due to the fact that the single top component has a much higher cross section in the *Hut* vertex compared to the *Hct* vertex, where the contribution from the single top component is almost negligible. Consideration of the single top component for signal events provides a 15% relative improvement in the final sensitivity for κ_{Hut} in this search, but no improvement to the κ_{Hct} significance. However, due to the presence of a charm quark in its anomalous decay, the TT Hct signal is much more sensitive to the adopted approach in this analysis than the TT Hut signal. These conclusions are demonstrated in Table 6.18, which shows a comparison of the expected limits on κ_{Hqt} when considering three signal scenarios: single top component, t$\bar{\text{t}}$ component and the sum of the two components. Overall, it can be stated that the t$\bar{\text{t}}$ component is still the most dominant contribution when extracting limits, but the single top component becomes relevant to the limits on κ_{Hut} due to its non-negligible cross section and the lepton charge as a variable discriminating against t$\bar{\text{t}}$ backgrounds.

Table 6.16: Excluded cross sections at 95% CL in pb of t → uH (ST ⊕ TT) for the *b2j3*, *b2j4*, *b3j3* and *b3j4* category and their combination. For the expected upper limit, the limit plus and minus one standard deviation and two standard deviations are also shown. The best fit values for the signal strength are shown in the last column.

Category	$\sigma_{\text{exp}} - 2\sigma$	$\sigma_{\text{exp}} - 1\sigma$	σ_{exp}	σ_{obs}	$\sigma_{\text{exp}} + 1\sigma$	$\sigma_{\text{exp}} + 2\sigma$	best fit $\pm 1\sigma$
b2j3	2.3	3.1	4.4	5.1	6.2	8.3	$0.49^{+0.40}_{-0.49}$
b2j4	2.3	3.0	4.2	3.9	5.9	7.8	$0.00^{+1.98}_{-0.00}$
b3j3	1.2	1.6	2.2	3.2	3.0	4.1	$0.39^{+0.27}_{-0.30}$
b3j4	1.2	1.6	2.3	2.2	3.2	4.4	$0.00^{+0.27}_{-0.00}$
comb	0.7	0.9	1.3	1.7	1.7	2.3	$0.21^{+0.15}_{-0.21}$
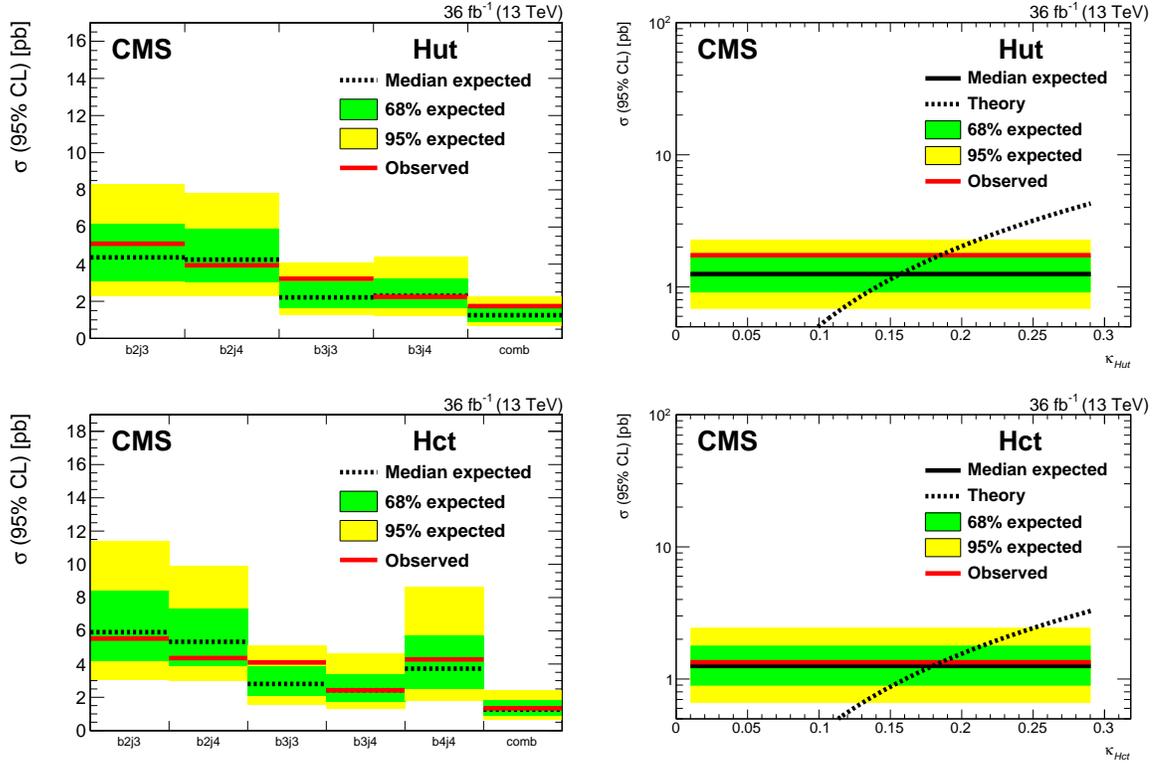
Figure 6.13: Excluded signal (ST ⊕ TT) cross section at 95% CL per event category for κ_{Hut} (top left) and κ_{Hct} (bottom left). The combined excluded signal cross section for κ_{Hut} (top right) and κ_{Hct} (bottom right) is shown with respect to the predicted dependence of the signal cross section on the coupling strength (as illustrated in Figure 5.2). The value of the coupling strength for which the theoretical prediction overlaps with the cross section limit is taken as the 95% CL upper limit on κ_{Hqt}.

Table 6.17: Excluded cross sections at 95% CL in pb of t → cH (ST ⊕ TT) for the *b2j3, b2j4, b3j3, b3j4* and *b4j4* category and their combination. For the expected upper limit, the limit plus and minus one standard deviation and two standard deviations are also shown. The best fit values for the signal strength are shown in the last column.

Category	$\sigma_{\text{exp}} - 2\sigma$	$\sigma_{\text{exp}} - 1\sigma$	σ_{exp}	σ_{obs}	$\sigma_{\text{exp}} + 1\sigma$	$\sigma_{\text{exp}} + 2\sigma$	best fit $\pm 1\sigma$
b2j3	3.1	4.2	5.9	5.5	8.4	11.4	$0.00^{+1.46}_{-0.00}$
b2j4	3.0	3.9	5.3	4.4	7.4	10.0	$0.00^{+0.64}_{-0.00}$
b3j3	1.6	2.1	2.8	4.1	3.8	5.1	$0.64^{+0.47}_{-0.54}$
b3j4	1.3	1.7	2.4	2.4	3.4	4.6	$0.00^{+0.42}_{-0.00}$
b4j4	1.8	2.5	3.7	4.3	5.7	8.7	$0.10^{+0.71}_{-0.10}$
comb	0.7	0.9	1.3	1.3	1.8	2.5	$0.01^{+0.13}_{-0.01}$

Table 6.18: Expected upper limits at 95% on the coupling strengths when considering the sum of the two signal components (combined), the t$\bar{\text{t}}$ component only (TT Hqt) and single top component only (ST Hqt).

Coupling	Combined	TT Hqt	ST Hqt
κ_{Hut}	0.16	0.19	0.21
κ_{Hct}	0.19	0.17	0.85

Two-dimensional limits

Up until now the *Hut* and *Hct* vertex have been considered as separate signal cases. However, these two signal processes may occur simultaneously, meaning a combined search for Hut ⊕ Hct needs to be considered as well. A naive approach at combining the two top-FCNC vertices into one signal has been considered. The two signal cases are merged for 100 benchmark signal scenarios. The 100 benchmark signal samples are constructed from the existing signal samples (ST Hut, ST Hct, TT Hut and TT Hct) as

$$\text{Signal} = \kappa_{Hut}^2 \cdot (\text{ST Hut} + \text{TT Hut}) + \kappa_{Hct}^2 \cdot (\text{ST Hct} + \text{TT Hct}), \qquad (6.11)$$

where κ_{Hqt} may adopt any value from 0 to 0.27 in steps of 0.03. For each of the 100 benchmark scenarios, the whole analysis chain is recycled, meaning that the same event selection, jet assignment technique, five event categories and BDT event discriminant training variables have been used. A dedicated BDT training is performed for each of the benchmarks and combined limits on their signal strength have been calculated. The expected limits on the signal strength at 95% CL of each benchmark scenario are shown in Figure 6.14. Benchmark scenarios where the expected limit on the signal strength is below 1 are expected to be excluded with 95% CL.
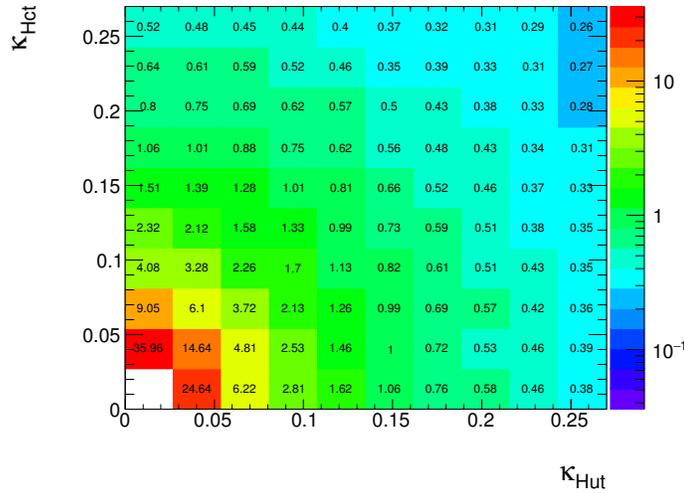
Figure 6.14: Expected limits on the signal strength at 95% CL for each of the 100 benchmark signal scenarios. Benchmark scenarios where the expected limit on the signal strength is below 1 are considered excluded with 95% CL.

From these 100 benchmark scenarios an interpolation of the expected limits from the one-dimensional cases has been extracted. This is done in order to interpret Figure 6.14 for a more fine-grained representation of the 2D limits. This interpolates from the one-

dimensional limits $\kappa_{Hut,1D}$ and $\kappa_{Hct,1D}$ as

$$\text{Limit}_{2D} = \kappa_{Hct,1D}\sqrt{1 - \left(\frac{\kappa_{Hut}}{\kappa_{Hut,1D}}\right)^2}. \tag{6.12}$$

Applying this interpolation to the expected (median, $\pm 1\sigma$ and $\pm 2\sigma$) and observed 1D-limits, one obtains the 2D-results for the coupling strengths and branching ratios as represented in Figure 6.15.
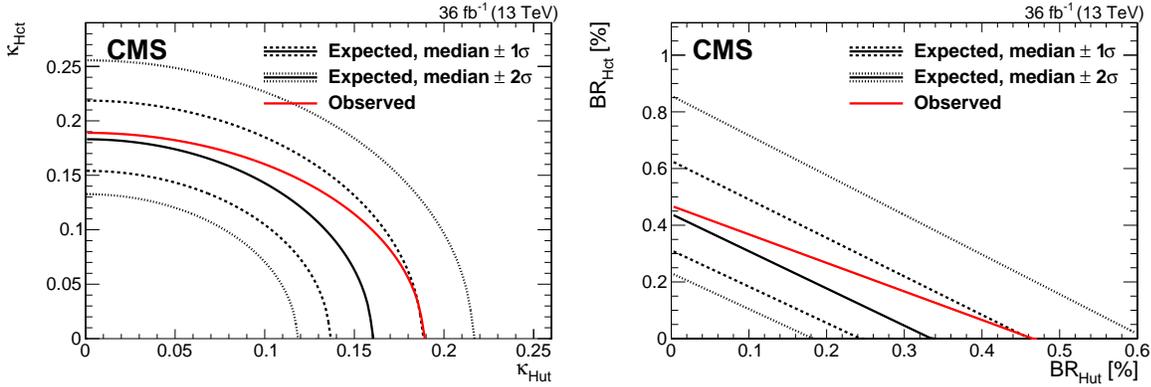
Figure 6.15: Excluded top quark FCNC coupling strengths (left) and branching ratios (right) at 95% CL.

Cut-and-count limits

A so-called cut-and-count approach has been considered as well. In this approach, a value of the BDT event discriminant is chosen such that, if each event is required to surpass this value, the background is rejected maximally whilst selecting as much signal events as possible. The figure of merit used for chosing the optimal BDT cut value is $\sqrt{(1 - \epsilon_s)^2 + \epsilon_b^2}$, where ϵ_s (ϵ_b) is the efficiency of signal (background) events surviving the BDT cut. This figure of merit becomes 1 if no (or all) events survive this cut, and achieves a minimum if background is minimally and signal maximally selected. The upper limits on the coupling strengths determined from this cut-and-count approach are compared to the upper limits from the template fit in Table 6.19. The upper limits from the template fit are far more stringent than those in a cut-and-count approach.

Table 6.19: Overview of the expected and observed upper limits at 95% CL on the coupling strengths κ_{Hut} and κ_{Hct} from the template-fit and the cut-and-count approach.

Method	Expected limit		Observed limit	
	κ_{Hut}	κ_{Hct}	κ_{Hut}	κ_{Hct}
Template fit	0.16	0.18	0.19	0.19
Cut-and-count	0.37	0.40	0.37	0.40

Chapter 7

Conclusions and prospects

7.1 Conclusions

The high suppression of flavour changing neutral currents in the Standard Model is a gold mine for new-physics searches, as much higher occurrence rates of FCNC are predicted by many new-physics models. Though currently many of the new-physics gems remain hidden and no top-FCNC have been observed, the search performed in this thesis has been able to set stringent upper limits on FCNC involving a top quark and Higgs boson, where the Higgs boson decays into a pair of bottom quarks. The data set used in this thesis is a collection of high-energy proton-proton collisions at a center-of-mass energy of 13 TeV, produced by the Large Hadron Collider and recorded by the CMS experiment at CERN in 2016 with an integrated luminosity of 36 fb^{-1}. It is the first time ever that the top-Higgs FCNC couplings are probed in the single top associated production with the Higgs boson, resulting in improved overall sensitivity to the *Hut* vertex.

The observed (expected) upper limits at 95% CL on the branching ratios $\mathcal{B}(\mathbf{t} \rightarrow \mathbf{uH}) \lesssim \mathbf{4.4 \times 10^{-3}} \ (\mathbf{3.4 \times 10^{-3}})$ and $\mathcal{B}(\mathbf{t} \rightarrow \mathbf{cH}) \lesssim \mathbf{4.7 \times 10^{-3}} \ (\mathbf{4.7 \times 10^{-3}})$, are significantly better than the upper limits from the search with the CMS experiment using 8 TeV pp-collisions in the same channel (H $\rightarrow$ b$\bar{\text{b}}$), being $\mathcal{B}(\text{t} \rightarrow \text{uH}) \lesssim 1.92 \times 10^{-2} \ (8.5 \times 10^{-3})$ and $\mathcal{B}(\text{t} \rightarrow \text{cH}) \lesssim 1.16 \times 10^{-2} \ (8.6 \times 10^{-3})$. This large improvement shows that the methods developed in this thesis are much more sensitive to top-FCNC searches than those developed in the 8 TeV search with the CMS experiment. The comparison of the (expected and observed) limits at 95% CL obtained in this thesis with the limits from searches by both the CMS and ATLAS collaboration in the same channel (H $\rightarrow$ b$\bar{\text{b}}$) is shown in Figure 7.1. The limits obtained in this thesis are the most stringent limits to date on FCNC involving a top quark and Higgs boson, where the Higgs boson decays into a pair of b quarks. Figure 7.1 also indicates the branching ratios as predicted in the SM and some new-physics models, which are still beyond the limits presented in this thesis. The t $\rightarrow$ cH upper limits are on the verge of unveiling (or excluding) predictions from flavour-violating two-Higgs doublet models.
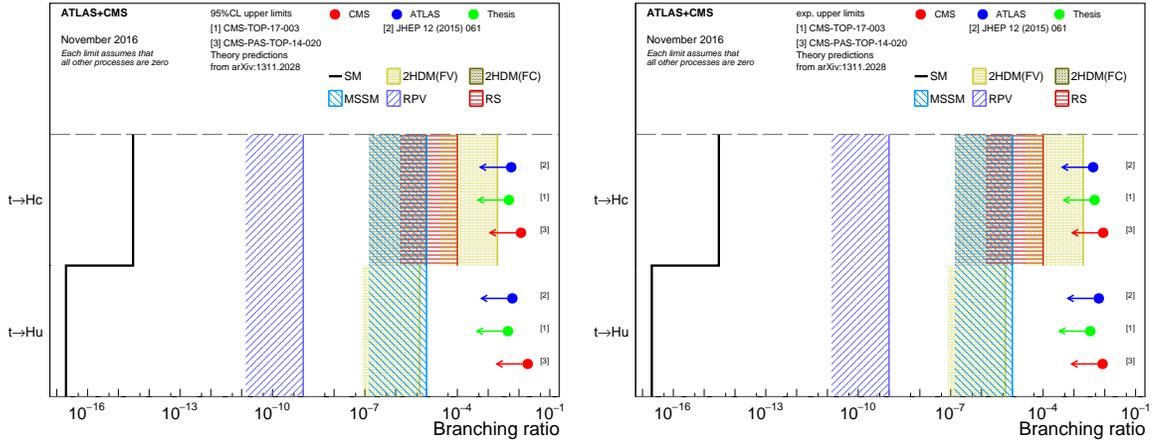
Figure 7.1: Summary of top-FCNC observed (left) and expected (right) upper limits at 95% CL from CMS and ATLAS searches (red and blue arrows) at 8 TeV in the H → b$\bar{\text{b}}$ channel and the results from this thesis (green arrows). A comparison is shown with Standard Model and new-physics predictions. This figure is adapted from [130].

An overview of the current best observed upper limits at 95% CL, as obtained from a combination of search channels, on all different top-FCNC from searches by the ATLAS and CMS collaborations using 8 TeV pp-collision data is shown in Figure 7.2, together with the limits as obtained in this thesis in the H → b$\bar{\text{b}}$ channel. This overview shows that the limits from this thesis are comparable with the best limits as obtained from a combination of limits from several search channels.

A very recent result [131] was submitted for publication by the ATLAS experiment using 13 TeV proton collision data put the most stringent limits to date on top-Higgs FCNC couplings. The observed limits at 95% CL from this search are $\mathcal{B}(\text{t} \rightarrow \text{uH}) \lesssim 2.4 \times 10^{-3}$ and $\mathcal{B}(\text{t} \rightarrow \text{cH}) \lesssim 2.2 \times 10^{-3}$. These limits were obtained in the H → $\gamma\gamma$ channel, considering only the top quark pair production mode for FCNC where the SM decay mode of the top quark was not limited to a leptonic decay. This illustrates its complementarity to the results obtained in this thesis.

7.2   Prospects

The future of the search for FCNC involving a top quark and Higgs boson (H → b$\bar{\text{b}}$) looks bright. Even though the systematic uncertainties are quite large with up to 34% uncertainty in the *b4j4* category, they get significantly constrained in the fitting procedure to less than 10% in that same category. This constraint can be largely attributed to the use of the control categories *b2j3* and *b2j4*. Naively, one would expect an even tighter constraint of the nuisance parameters by expanding the number of control categories by including categories with 1 and 0 b-tagged jets. However, SM

Figure 7.2: Summary of FCNC upper limits at the 95% CL from CMS and ATLAS searches (red and blue arrows), combining results from different channels at 8 TeV, and the results from this thesis (green arrows) in the H → b̄b channel. A comparison is shown with Standard Model and new-physics predictions. This figure is adapted from [130].

processes that have a negligible contribution in the current approach, such as W+jets and multijet events, would be enriched in such categories. This would open up the analysis to a range of systematic uncertainties that could be neglected so far, e.g. uncertainties coming from the estimation of multijet events, making such a strategy undesirable.

The biggest source of systematic uncertainty comes from the reweighting of the CSVv2 discriminant. Of the eight sources contributing to that uncertainty, two are dominant: the contamination of light-flavoured jets and one of the charm flavour nuisance parameters. A more careful estimation of the charm flavour nuisance parameters, which is currently done in a conservative way, could reduce this uncertainty. For example, one could try to actually measure the scale factors for charm-flavoured jets, instead of putting them to unity. Although measuring scale factors for charm-flavoured jets is difficult, methods to do so are currently being developed by the CMS collaboration. The contamination of light-flavoured jets is correlated to the subtraction of light-flavour contributions in the determination of scale factors for b-flavoured jets. Improving the understanding of this uncertainty is not straightforward, but could be achieved by a better simulated description of actual data at high CSVv2 discriminant values (see Figure 3.7).

Regardless of reducing systematic uncertainties, this search is not limited by them. An improvement of the upper limits can be expected if the measurement is performed on a larger to be collected data set. With a simple extrapolation of the current analysis strategy to a data set with an integrated luminosity of 100 fb^{-1}, the upper limits at 95% CL can improve by roughly 15%, as summarised in Table 7.1. An integrated luminosity of 100 fb^{-1} is expected to be achieved by the end of LHC Run II and it is therefore worthwhile to repeat this analysis once Run II is finished.

Table 7.1: Comparison of the expected upper limits at 95% CL on a data set with an integrated luminosity of 36 fb^{-1}, to an extrapolation to a data set with an integrated luminosity of 100 fb^{-1}.

Luminosity [fb^{-1}]	Expected limit at 95% CL	
	$\mathcal{B}(\text{t} \rightarrow \text{uH})$	$\mathcal{B}(\text{t} \rightarrow \text{cH})$
36	3.4×10^{-3}	4.7×10^{-3}
100	2.3×10^{-3}	2.9×10^{-3}

On the part of the analysis strategy itself, some improvements could be introduced as well. For example, instead of using the CSVv2 b-tagger, one could switch to the recently developed DeepCSV b-tagger, which, at the considered Medium working point, has an absolute 6% higher efficiency of correctly identifying b jets. Moreover, a better description of actual data by simulations seems to be achieved in the DeepCSV discriminant with respect to the CSVv2 discriminant. This could result in a reduction of systematic uncertainties for DeepCSV related scale factors. Besides using a different b-tagger, c-tagging discriminants are interesting to consider as well in the future.

As it is clear that b-tagging information has the biggest discrimination power in the event BDT trainings, c-tagging information might have a similar effect. Furthermore, the performance of heavy-flavour tagging is expected to improve even further with the installation of the new pixel detector in the CMS detector in March 2017.

Bibliography

[1] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (Oct, 1970) 1285–1292.

[2] J. J. Aubert *et al.*, *Experimental Observation of a Heavy Particle J*, Phys. Rev. Lett. **33** (Dec, 1974) 1404–1406.

[3] J. E. Augustin *et al.*, *Discovery of a Narrow Resonance in e^+e^- Annihilation*, Phys. Rev. Lett. **33** (Dec, 1974) 1406–1408.

[4] CDF Collaboration, *Evidence for $B_s^0 \to \phi\phi$ Decay and Measurements of Branching Ratio and A_{CP} for $B^+ \to \phi K^+$*, Phys. Rev. Lett. **95** (2005) 031801.

[5] LHCb, CMS Collaboration, V. Khachatryan *et al.*, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68–72.

[6] W. Altmannshofer and D. M. Straub, *New Physics in $B \to K^*\mu\mu$?*, Eur. Phys. J. **C73** (2013) 2646.

[7] CMS Collaboration, *Search for associated production of a Z boson with a single top quark and for tZ flavour-changing interactions in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **07** (2017) 003.

[8] CMS Collaboration, *Search for anomalous single top quark production in association with a photon in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **04** (2016) 035.

[9] ATLAS Collaboration, *Search for single top-quark production via flavour-changing neutral currents at 8 TeV with the ATLAS detector*, Eur. Phys. J. **C76** (2016), no. 2, 55.

[10] ATLAS Collaboration, *Search for flavour-changing neutral current top quark decays $t \to Hq$ in pp collisions at $\sqrt{s}$=8 TeV with the ATLAS detector*, Journal of High Energy Physics **2015** (2015), no. 12, 61.

[11] CMS Collaboration, *Search for top quark decays via Higgs-boson-mediated flavor-changing neutral currents in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **02** (2017) 079.

[12] CMS Collaboration, *Search for the Flavor-Changing Interactions of a Top Quark with the Higgs boson in H → bb̄-Channel at $\sqrt{s}$ = 13 TeV*, TOP-17-003: in progress.

[13] CMS Collaboration, *Identification of b and c jets in the CMS experiment at 13 TeV*, BTV-16-002: in progress.

[14] M. Peskin and D. Schroeder, *An introduction to quantum field theory*. The Advanced Book Program. Addison-Wesley Publishing Company, 1995.

[15] F. Mandl and G. Shaw, *Quantum Field Theory*. A Wiley-Interscience publication. John Wiley & Sons, 2010.

[16] CMS Collaboration, *Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV*, JHEP **06** (2013) 081.

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

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

[19] *The Dutch Paradigm*, `http://thedutchparadigm.org/the-paradigm/electron/electron-point-particle/`, Accessed: 2017-04-06.

[20] S. D. Drell and T.-M. Yan, *Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies*, Phys. Rev. Lett. **25** (Aug, 1970) 316–320.

[21] A. Djouadi, *The Anatomy of electro-weak symmetry breaking (II): The Higgs boson in the minimal supersymmetric model*, Phys. Rept. **459** (2008).

[22] S. P. Martin, *A Supersymmetry primer*, `hep-ph/9709356`. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[23] P. Langacker, *Grand Unified Theories and Proton Decay*, Phys. Rept. **72** (1981) 185.

[24] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (Jun, 1963) 531–533.

[25] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973), no. 2, 652.

[26] Particle Data Group Collaboration, *Review of Particle Physics*, Chin. Phys. **C40** (2016), no. 10, 100001.

[27] J. A. Aguilar-Saavedra, *Top flavor-changing neutral interactions: Theoretical expectations and experimental detection*, Acta Phys. Polon. **B35** (2004) 2695–2710.

[28] J. A. Aguilar-Saavedra, *A Minimal set of top anomalous couplings*, Nucl. Phys. **B812** (2009) 181–204.

[29] J. Gao, C. S. Li, and H. X. Zhu, *Top Quark Decay at Next-to-Next-to Leading Order in QCD*, Phys. Rev. Lett. **110** (2013), no. 4, 042001.

[30] O. S. Brning *et al.*, *LHC Design Report: The LHC main ring.* No. v. 1 in CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[31] The ALICE Collaboration, *The ALICE Experiment at the CERN LHC*, JINST **3 S08002** (2008).

[32] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3 S08003** (2008).

[33] The CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3 S08004** (2008).

[34] The LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3 S08005** (2008).

[35] The TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, JINST **3 S08007** (2008).

[36] The LHCf Collaboration, *The LHCf Detector at the CERN Large Hadron Collider*, JINST **3 S08006** (2008).

[37] S. Russenschuck, *Magnets for Accelerators*, pp. 1–48. 2010.

[38] D. Leroy *et al.*, *The superconducting magnets for LHC*, in *proceedings of the 15th International Conference on Charged Particle Accelerators, Protvino, Russian Federation.* 1996.

[39] *LHC Report: end of 2016 proton-proton operation*, `https://home.cern/cern-people/updates/2016/10/lhc-report-end-2016-proton-proton-operation`, Accessed: 2017-06-02.

[40] CMS Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* Technical Design Report CMS. CERN, Geneva, 2006.

[41] C. Lippmann, *Particle identification*, Nucl. Instrum. Meth. **A666** (2012) 148–172.

[42] D. Sprenger *et al.*, *Validation of Kalman Filter alignment algorithm with cosmic-ray data using a CMS silicon strip tracker endcap*, JINST **5** (2010) P06007.

[43] W. Adam, B. Mangano, T. Speer, and T. Todorov, *Track reconstruction in the CMS tracker*, CERN-CMS-NOTE-2006-041.

[44] S. Cucciarelli, M. Konecki, D. Kotlinski, and T. Todorov, *Track reconstruction, primary vertex finding and seed generation with the pixel detector*, CERN-CMS-NOTE-2006-026.

[45] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering **82** (1960), no. Series D, 35–45.

[46] CMS Collaboration, S. Chatrchyan *et al.*, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, JINST **9** (2014), no. 10, P10009.

[47] CMS Collaboration, *Performance of CMS hadron calorimeter timing and synchronization using test beam, cosmic ray, and LHC beam data*, Journal of Instrumentation **5** (Mar., 2010) T03013.

[48] CMS Collaboration, *CMS reconstruction improvement for the muon tracking by the RPC chambers*, PoS **RPC2012** (2012) 045. [JINST8,T03001(2013)].

[49] V. Halyo *et al.*, *GPU Enhancement of the Trigger to Extend Physics Reach at the LHC*, JINST **8** (2013) P10005.

[50] I. Bird, P. Buncic, *et al.*, *Update of the Computing Models of the WLCG and the LHC Experiments*, Tech. Rep. CERN-LHCC-2014-014. LCG-TDR-002, Apr, 2014.

[51] CMS Collaboration, *CMS physics: Technical design report*, CERN-LHCC-2006-001.

[52] *WLCG: Worldwide LHC Computing Grid*, `http://wlcg.web.cern.ch/documents-reference`, Accessed: 2017-05-02.

[53] M. Dobbs, S. Frixione, E. Laenen, and K. Tollefson, *Les Houches Guidebook to Monte Carlo Generators for Hadron Collider Physics*, `arXiv:hep-ph/0403045`.

[54] J. D. Bjorken, *Asymptotic Sum Rules at Infinite Momentum*, Phys. Rev. **179** (Mar, 1969) 1547–1553.

[55] R. P. Feynman, *Photon-hadron interactions*, vol. 14. WA Benjamin New York, 1972.

[56] *The Coordinated Theoretical-Experimental Project on QCD*,
`http://www.physics.smu.edu/scalise/cteq/`, Accessed: 2017-03-17.

[57] *Martin-Stirling-Thorne-Watt Parton Distribution Functions*,
`https://mstwpdf.hepforge.org/`, Accessed: 2017-03-17.

[58] *Neural Network Parton Distribution Functions*,
`https://nnpdf.hepforge.org/`, Accessed: 2017-03-17.

[59] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, The European Physical Journal C **63** (2009), no. 2, 189–285.

[60] G. Altarelli and G. Parisi Nucl.Phys. **298** (1977).

[61] Y. L. Dokshitzer Sov.Phys. **641** (1977).

[62] V. N. Gribov and L. N. Lipatov Sov.J.Nucl.Phys. **438** (1972).

[63] J. C. Collins, D. E. Soper, and G. Sterman, *Factorization of Hard Processes in QCD*, `arXiv:hep-ph/0409313`.

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

[65] T. Sjostrand, *Jet Fragmentation of Nearby Partons*, Nucl. Phys. **B248** (1984) 469–502.

[66] CDF Collaboration, *The Underlying event in hard scattering processes*, eConf **C010630** (2001) P501.

[67] CMS Collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements*, The European Physical Journal C **76** (2016), no. 3, 155.

[68] J. Alwall *et al.*, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, Journal of High Energy Physics **2014** (2014), no. 7, 79.

[69] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *Matching matrix elements and shower evolution for top-pair production in hadronic collisions*, Journal of High Energy Physics **2007** (2007), no. 01, 013.

[70] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007), no. 11, 070.

[71] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, Journal of High Energy Physics **2004** (2004), no. 11, 040.

[72] S. Alioli, S.-O. Moch, and P. Uwer, *Hadronic top-quark pair-production with one jet and parton showering*, Journal of High Energy Physics **2012** (2012), no. 1, 137.

[73] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics **2010** (2010), no. 6, 43.

[74] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026.

[75] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867.

[76] J. Allison *et al.*, *Geant4 developments and applications*, IEEE Transactions on Nuclear Science **53** (Feb, 2006) 270–278.

[77] CMS Collaboration Collaboration, *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of* $t\bar{t}$ *at* $\sqrt{s} = 8$ *and 13 TeV*, Tech. Rep. CMS-PAS-TOP-16-021, CERN, Geneva, 2016.

[78] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, Comput. Phys. Commun. **185** (2014) 2250–2300.

[79] C. Degrande *et al.*, *UFO - The Universal FeynRules Output*, Comput. Phys. Commun. **183** (2012) 1201–1214.

[80] J. Pumplin *et al.*, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012.

[81] M. L. Mangano *et al.*, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, JHEP **01** (2007) 013.

[82] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, Journal of High Energy Physics **2002** (2002), no. 06, 029.

[83] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, CMS-PAS-PFT-09-001.

[84] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, `1706.04965`.

[85] CMS Collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, CMS-PAS-PFT-10-001.

[86] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering **82** (1960), no. Series D, 35–45.

[87] G. P. Salam, *Towards Jetography*, Eur. Phys. J. **C67** (2010) 637–686.

[88] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k t jet clustering algorithm*, Journal of High Energy Physics **2008** (2008), no. 04, 063.

[89] CMS Collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, JINST **12** (2017), no. 02, P02014.

[90] CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, Journal of Instrumentation **6** (2011), no. 11, P11002.

[91] *JER Scaling factors and Uncertainty for 13 TeV (2015 and 2016)*, `https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#JER_Scaling_factors_and_Uncertai`, Accessed: 2017-05-12.

[92] CMS Collaboration, *Identification of b quark jets at the CMS Experiment in the LHC Run 2*, CMS-PAS-BTV-15-001.

[93] CMS Collaboration, *Identification of c-quark jets at the CMS experiment*, Tech. Rep. CMS-PAS-BTV-16-001, CERN, Geneva, 2016.

[94] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, Journal of Instrumentation **8** (2013), no. 04, P04013.

[95] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, 2003.

[96] S. Haykin, *Neural Networks: A Comprehensive Foundation*. International edition. Prentice Hall, 1999.

[97] R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Adaptive Vertex Fitting*, Tech. Rep. CMS-NOTE-2007-008, CERN, Geneva, Mar, 2007.

[98] CMS Collaboration, *Search for the associated production of the Higgs boson with a top-quark pair*, J. High Energy Phys. **09** (Aug, 2014) 087. 61.

[99] DELPHES 3 Collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, JHEP **1402** (2014) 057.

[100] A. Denner *et al.*, *Standard Model Higgs-Boson Branching Ratios with Uncertainties*, Eur. Phys. J. **C71** (2011) 1753.

[101] B. Dumont *et al.*, *Toward a public analysis database for LHC new physics searches using MADANALYSIS 5*, Eur. Phys. J. **C75** (2015), no. 2, 56.

[102] CMS Collaboration, *Performance of b tagging at $\sqrt{s}$=8 TeV in multijet, ttbar and boosted topology events*, CMS-PAS-BTV-13-001.

[103] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning.* Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[104] CMS Collaboration, *Search for the Flavor-Changing Neutral Current Decay* $t \to qH$ *Where the Higgs Decays to* $b\bar{b}$ *Pairs at* $\sqrt{s} = 8$ *TeV*, Tech. Rep. CMS-PAS-TOP-14-020, CERN, Geneva, 2015.

[105] CMS Collaboration Collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, Tech. Rep. CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[106] J. M. Campbell, R. K. Ellis, and W. T. Giele, *A Multi-Threaded Version of MCFM*, Eur. Phys. J. **C75** (2015), no. 6, 246.

[107] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, *FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order*, Comput. Phys. Commun. **182** (2011) 2388–2403.

[108] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930.

[109] P. Kant *et al.*, *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, Comput. Phys. Commun. **191** (2015) 74–89.

[110] M. Aliev *et al.*, *HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR*, Comput. Phys. Commun. **182** (2011) 1034–1046.

[111] *Utilities for Accessing Pileup Information for Data*, `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2#Moriond_2017`, Accessed: 2017-05-12.

[112] *Baseline muon selections for Run-II*, `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2#Baseline_muon_selections_for_Run`, Accessed: 2017-05-12.

[113] *Cut Based Electron ID for Run 2*, `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2`, Accessed: 2017-05-12.

[114] *Effective areas for electron isolation*,
      `https://indico.cern.ch/event/369239/contributions/874575/`
      `attachments/1134761/1623262/talk_effective_areas_25ns.pdf`, Accessed:
      2017-06-06.

[115] *Jet Identification for the 13 TeV data Run2016*,
      `https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016`,
      Accessed: 2017-05-12.

[116] *Utilities for Accessing Pileup Information for Data*,
      `https://twiki.cern.ch/twiki/bin/viewauth/CMS/`
      `PileupJSONFileforData#Pileup_JSON_Files_For_Run_II`, Accessed:
      2017-05-12.

[117] *Electron efficiencies and scale factors*, `https://twiki.cern.ch/twiki/bin/`
      `view/CMS/EgammaIDRecipesRun2#Electron_efficiencies_and_scale`,
      Accessed: 2017-05-12.

[118] *Reference muon id, isolation and trigger efficiencies for Run-II*,
      `https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffsRun2#`
      `Introduction`, Accessed: 2017-05-12.

[119] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT**
      (2007) 040.

[120] V. Khachatryan and etal The European Physical Journal C **76** (2016), no. 7,
      379.

[121] G. Bevilacqua *et al.*, *Assault on the NLO Wishlist: pp → t anti-t b anti-b*,
      JHEP **09** (2009) 109.

[122] A. Bredenstein, A. Denner, S. Dittmaier, and S. Pozzorini, *NLO QCD
      Corrections to Top Anti-Top Bottom Anti-Bottom Production at the LHC: 2.
      full hadronic results*, JHEP **03** (2010) 021.

[123] D. Bourilkov, R. C. Group, and M. R. Whalley, *LHAPDF: PDF use from the
      Tevatron to the LHC*, in *TeV4LHC Workshop - 4th meeting Batavia, Illinois,
      October 20-22, 2005*. 2006. `hep-ph/0605240`.

[124] J. Butterworth *et al.*, *PDF4LHC recommendations for LHC Run II*, J. Phys.
      G43 (2016) 023001.

[125] CMS Collaboration, V. Khachatryan *et al.*, *Measurement of differential cross
      sections for top quark pair production using the lepton+jets final state in
      proton-proton collisions at 13 TeV*, Phys. Rev. **D95** (2017), no. 9, 092001.

[126] A. L. Read, *Presentation of search results: the CL s technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002), no. 10, 2693.

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

[128] L. Moneta *et al.*, *The RooStats Project*, PoS **ACAT2010** (2010) 057.

[129] H. B. Prosper and L. Lyons, eds., *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, CERN. CERN, Geneva, 2011.

[130] *Top Quark Physics Summary Figures*, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryFigures`, Accessed: 2017-06-30.

[131] ATLAS Collaboration, M. Aaboud *et al.*, *Search for top quark decays $t \to qH$, with $H \to \gamma\gamma$, in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector*, `1707.01404`.

Appendices

A   Correlations between input variables of event BDT training

Most of the input variables used in the training of BDTs to separate FCNC signal events from SM background events show no correlation. In this appendix the correlations between the input variables for the four categories of the *Hut* signal are shown in Figures A.1—A.4 and for the five categories of the *Hct* signal in Figures A.5—A.9. The correlations are shown separately for the FCNC signal, SM background and recorded data sample. These figures also illustrate the tendency for overtraining by showing the comparison of the BDT output shapes between the training and validation sample.
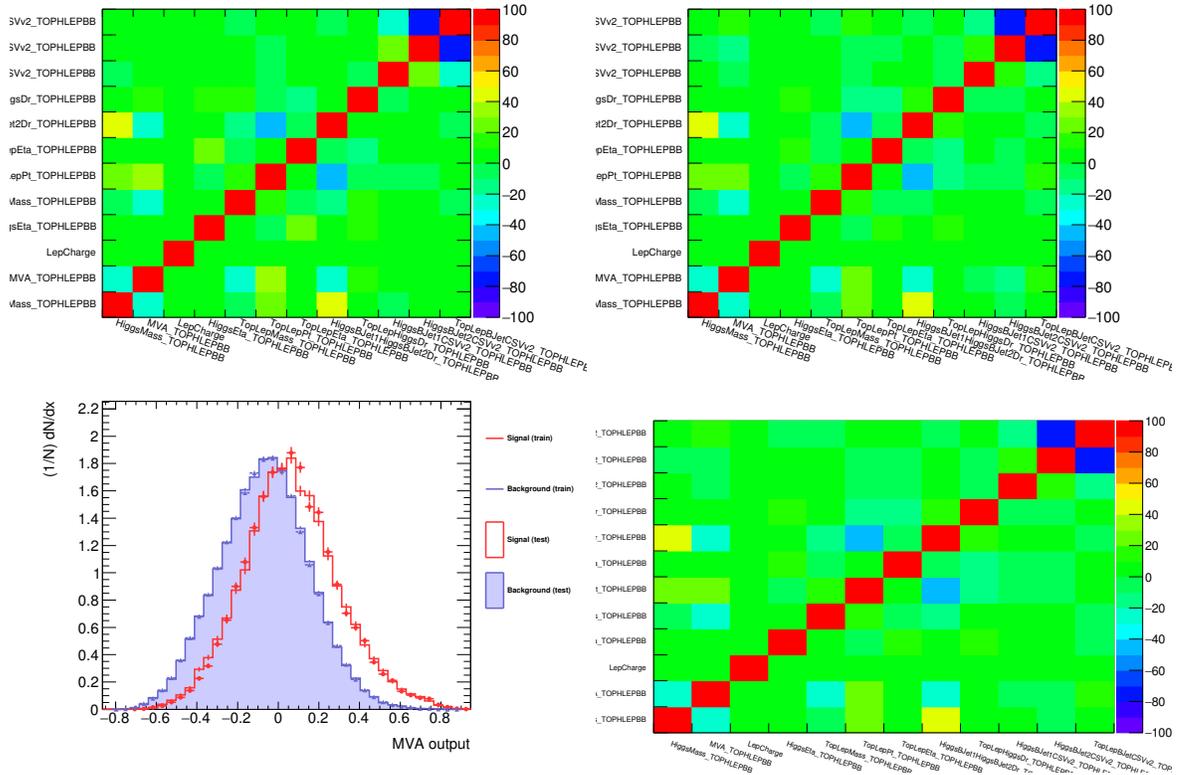
Figure A.1: Correlation plots for signal (top left), background (top right) and actual data (bottom right) events for the input variables of the BDT event training for the *Hut* coupling in the $b2j3$ category. The distribution of the BDT output (bottom left) for the trained sample is indicated by individual points, whereas the validation sample is represented by the unbroken line
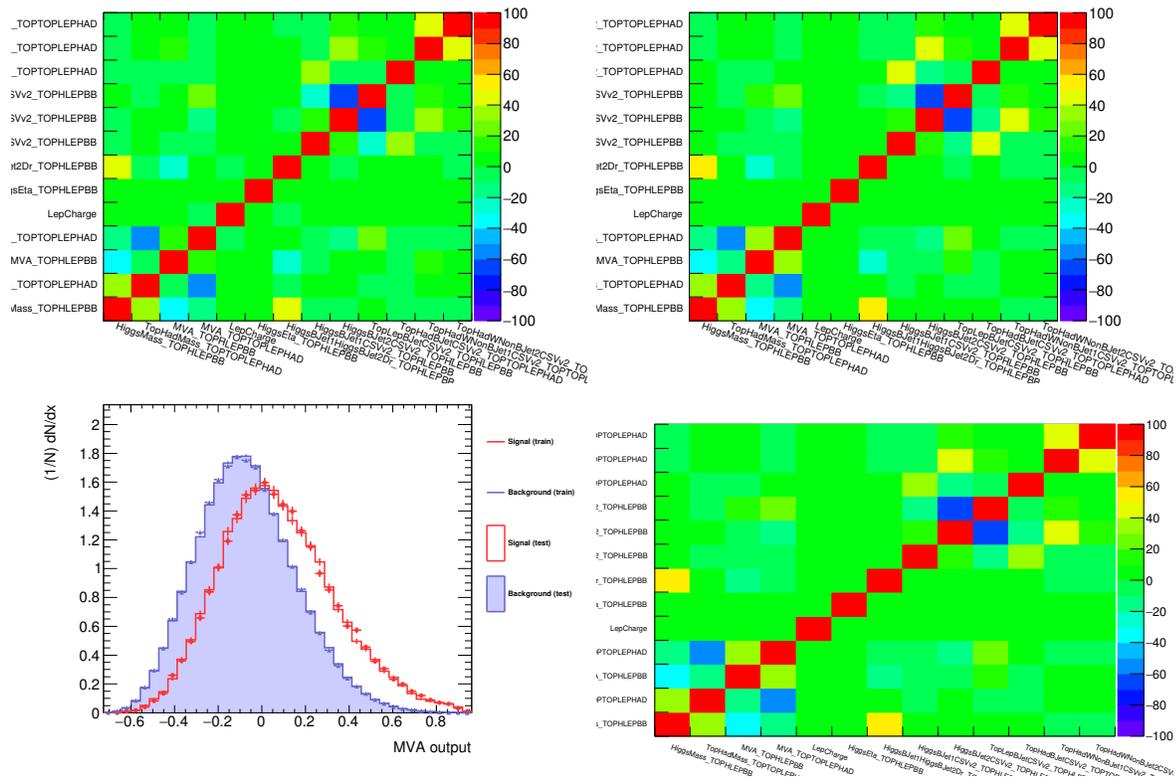
Figure A.2: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the *Hut* coupling in the *b2j4* category.

Figure A.3: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the *Hut* coupling in the *b3j3* category.
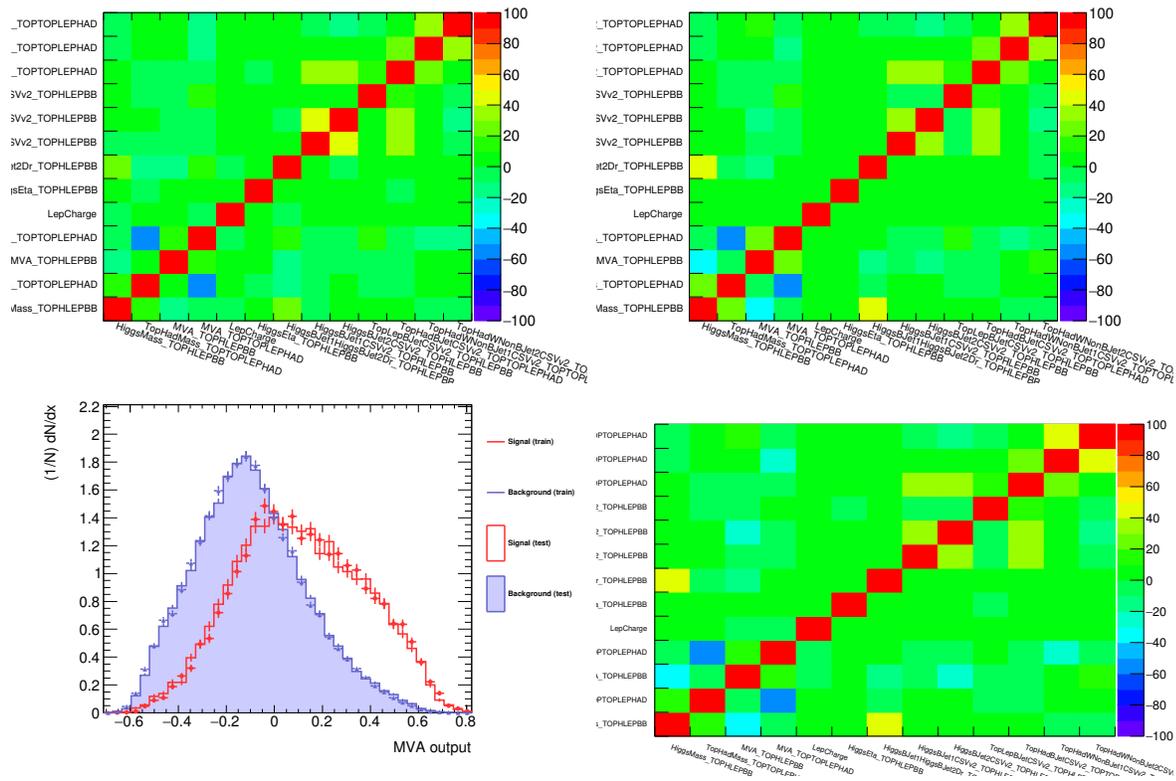
Figure A.4: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the *Hut* coupling in the *b3j4* category.
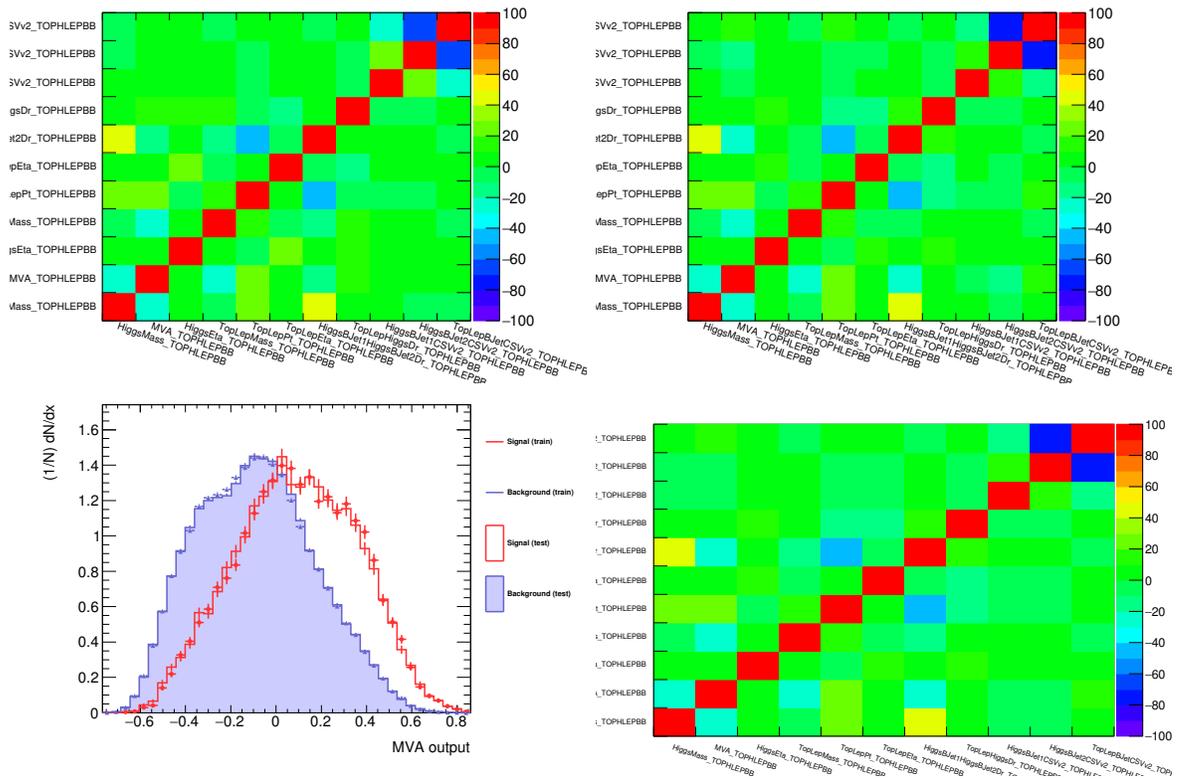
Figure A.5: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the Hct coupling in the $b2j3$ category.
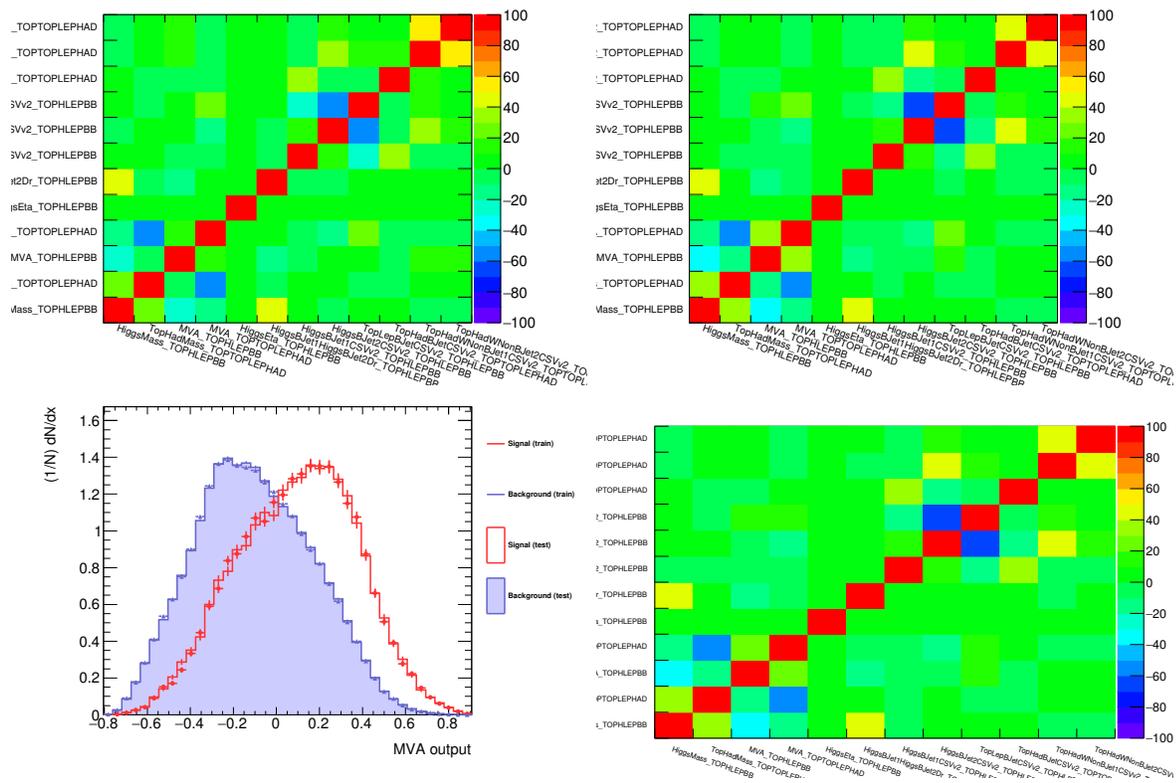
Figure A.6: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the Hct coupling in the $b2j4$ category.
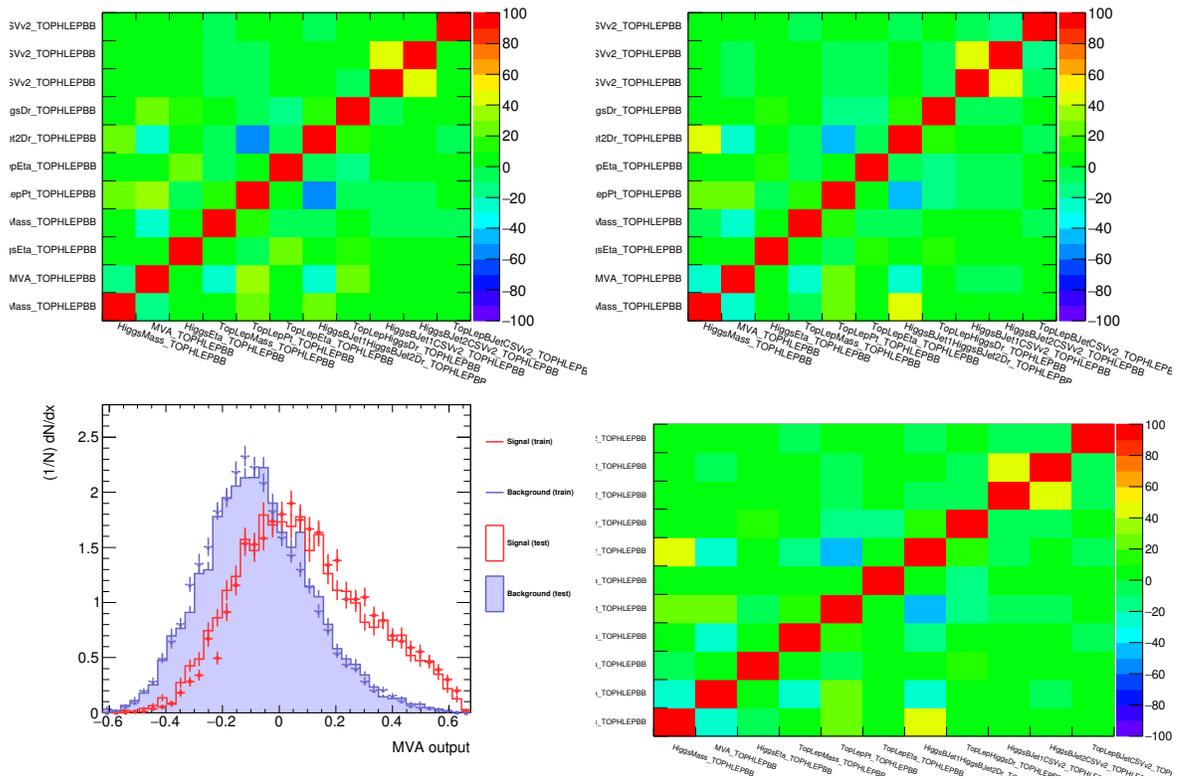
Figure A.7: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the Hct coupling in the $b3j3$ category.
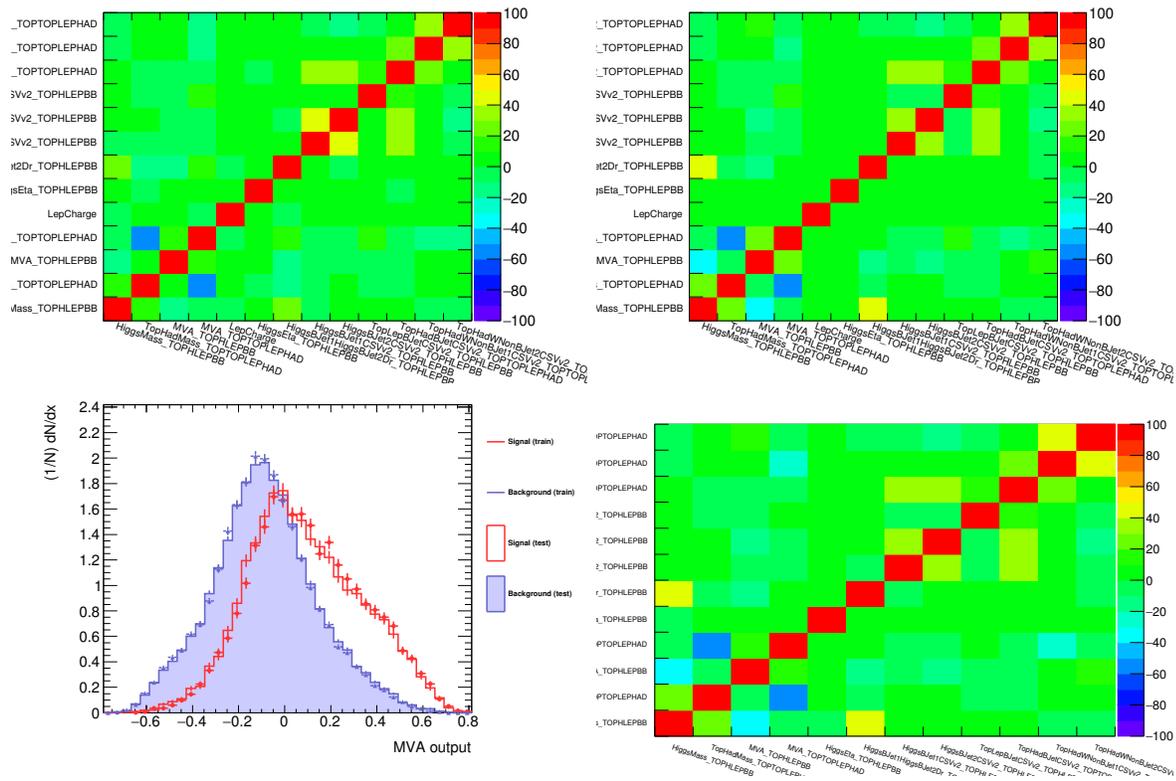
Figure A.8: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the Hct coupling in the $b3j4$ category.
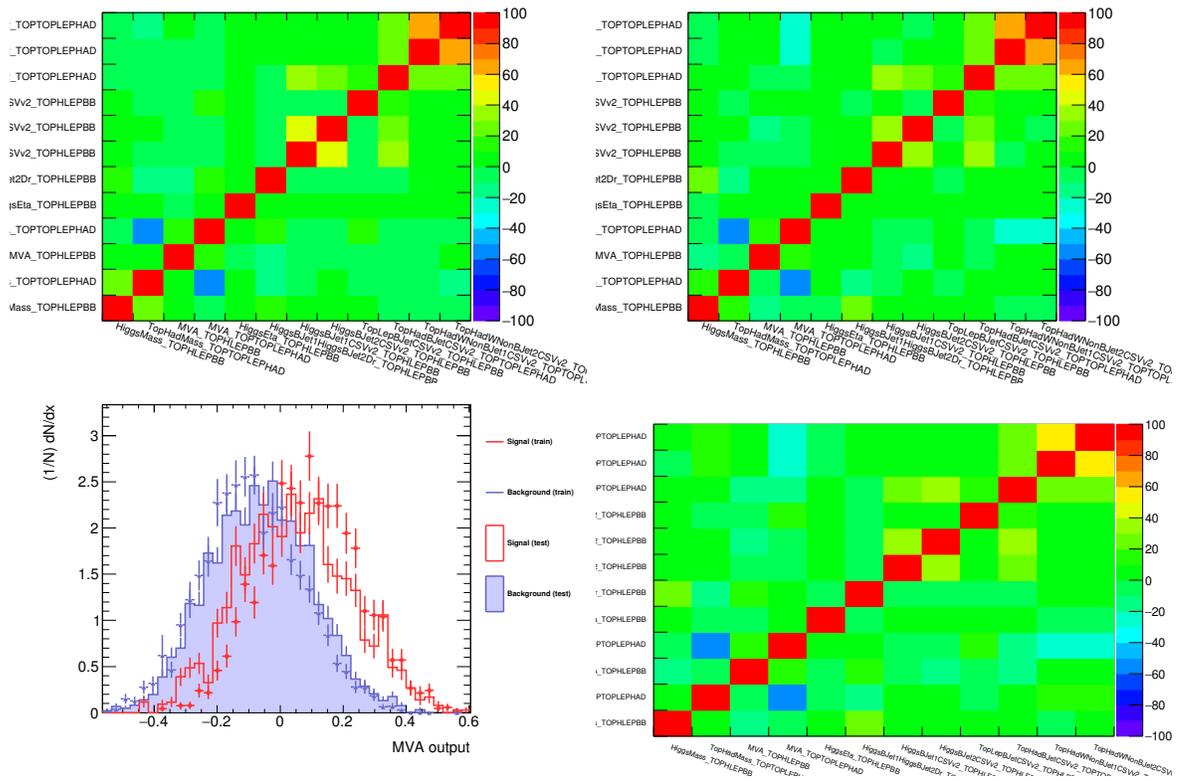
Figure A.9: Correlation plots for signal (left) and background (right) events for the input variables of the BDT event training for the Hct coupling in the $b4j4$ category.

Summary

The Standard Model of elementary particle physics provides a theoretical framework in which the current understanding of the fundemantal building blocks of nature, and their interactions, is encapsulated. It provides an extremely accurate representation of the workings of nature at the smallest length scales, as reachable by current experiments, and has been experimentally validated with exceptional precision at electroweak energy scales. Despite all this, the Standard Model is only able to account for about 4% of the universe's energy content, having no explanation for the nature of Dark Matter and Dark Energy. Furthermore, the Standard Model is only valid up to a certain energy scale, as it does not incorporate the unification of the four fundamental forces of nature at higher energy scales. The challenge for elementary particle physicists exists in expanding the framework of the Standard Model such that it can explain phenomena like Dark Matter and the unification of the four fundamental forces. Many theories stand ready to fill in those gaps where the Standard Model lacks the appropriate description, but have no experimental support so far. One of the most powerful ways to experimentally probe theorized extensions to the Standard Model, is to search for flavour-changing neutral currents (FCNC) in elementary particle interactions. Such interactions change the flavour of quarks, by means of a neutral boson, without changing their electric charge. FCNC are highly suppressed in the Standard Model, but could become strongly enhanced by new physics. A search for FCNC is therefore a strong probe for these models.

From a phenomenolgical study on FCNC involving top quarks, as performed in this thesis and as a result from a collaboration between research groups from the *Vrije Universiteit Brussel*, the *Institut Pluridisciplinaire Hubert CURIEN* at Strasbourg and *Chonbuk National University* in South-Korea, it was found that a search for the FCNC coupling between a top quark and a Higgs boson in proton collisions at a center-of-mass energy of 13 TeV, could become sensitive to new-physics predictions. Among the possible search channels for that specific FCNC coupling, the case where the Higgs boson decays into a pair of bottom quarks proved to be a very powerful channels. In this thesis, FCNC are explored in the highest-energy proton-proton collisions ever recorded by mankind, as produced by the Large Hadron Collider and recorded by the CMS experiment at CERN in 2016. The protons are collided at a center-of-mass energy of 13 TeV and the investigated data set, as recorded by the CMS experiment, represents an integrated luminosity of 36 fb^{-1}. The main bulk of the recorded data set is irrelevant in the search for FCNC and in order to substantially discard such collision events, a series of event selection criteria are introduced. On top of that, events are categorised

in control and signal categories on the basis of their topology. The event categories are defined such that the contribution of one of the following three event signatures is maximized: the anomalous FCNC production of a single top quark, production of a top quark pair followed by the FCNC decay of one of the top quarks, or the production of top quark pair where the decay goes according to known Standard Model processes. This is the first time that the anomalous FCNC production of a single top quark is investigated, made possible due to dedicated event categories and its predicted cross section.

A dedicated study has been performed in order to construct a highly performant algorithm, from which the investigated FCNC event signatures can be reconstructed. This algorithm is used to reconstruct FCNC event characteristics that can be used as a discriminating variable between FCNC events and background processes. A set of such variables is defined for each category and is used as a basis for a machine learning algorithm to construct a *master* variable, where the discriminatory power of each individual variable is combined. From simulations, the distribution of this variable can be predicted for the background-only and signal plus background hypotheses, and compared to the actual distribution as obtained from data.

No deviations from the background-only hypothesis have been observed in data, meaning FCNC involving a top quark and Higgs boson continue to elude experimental observation. However, stringent upper limits on the occurrence rate of FCNC have been extracted in this search, on the basis of a template fit. The observed (expected) upper limit at 95% confidence level on the branching ratios for a top quark t to an up quark u and Higgs boson H is $\mathcal{B}(\mathbf{t \rightarrow uH}) \lesssim \mathbf{4.4 \times 10^{-3}}$ $(\mathbf{3.4 \times 10^{-3}})$. The anomalous FCNC production of a single top quark improves the limits on $\mathcal{B}(t \rightarrow uH)$ with about 21% with respect to not considering it, which illustrates the added value of including this signal process. The observed (expected) upper limit at 95% confidence level on the branching ratios for a top quark t to an up quark c and Higgs boson H is $\mathcal{B}(\mathbf{t \rightarrow cH}) \lesssim \mathbf{4.7 \times 10^{-3}}$ $(\mathbf{4.7 \times 10^{-3}})$. Compared to a more simplified cut-and-count approach, the template fit is able to obtain limits that are more than two times more constrained. The limits, derived with the methods developed in this thesis, are able to significantly improve the upper limits obtained from the search in the same channel at the CMS experiment using 8 TeV collision data. Furthermore, the t $\rightarrow$ cH upper limits are on the verge of unveiling (or excluding) specific predictions from existing new-physics models.

Samenvatting

Een zoektocht naar smaak-veranderende neutrale stromen waarin een top quark en Higgs boson betrokken zijn met het CMS experiment

Het Standaard Model van de elementaire deeltjes fysica voorziet een theoretische omkadering die de huidige kennis van de fundamentele bouwstenen in de natuur, en hun interacties, omvat. Dit model verstrekt een extreem precieze voorstelling van de fundamentele werking van de natuur op de kleinste schaal, zoals momenteel experimenteel toereikbaar is, en is experimenteel gevalideerd met een exceptionele precisie bij de Elektro-Zwakke energie schaal. Desondanks slaagt het Standaard Model er enkel in om ongeveer 4% van de energie-inhoud in het universum te verklaren, waarbij de aard van Donkere Materie en Donkere Energie onverklaard blijven. Verder is het ook duidelijk dat het Standaard Model enkel geldig is tot een bepaalde energie schaal, aangezien het, onder andere, de eenmaking van de vier fundamentele krachten niet voorspelt. De uitdaging voor elementaire-deeltjes fysici bestaat erin om een uitbreiding op het Standaard Model te vinden zodat het fenomenen, zoals Donkere Materie en de eenmaking van de vier fundamentele krachten, kan beschrijven. Vele theorieën zijn reeds bedacht die deze hiaten in het Standaard Model verantwoorden, maar ontbreken dusver een experimentele onderbouwing. Een zeer krachtige manier om experimenteel zulke nieuwe theorieën te testen, is de zoektocht naar smaak-veranderende neutrale stromen, oftewel *flavour-changing neutral currents* (FCNC) in het Engels, in interacties tussen elementaire deeltjes. Zulke stromen komen voor wanneer quarks, door middel van een neutraal boson, van smaak veranderen zonder hun elektrische lading te wijzigen. Zulk proces komt bijna niet voor in het Standaard Model, maar wel versterkt wordt in vele van de nieuwe-fysica modellen.

Uit een fenomenologische studie naar FCNC die betrekking hebben op top quarks, zoals uitgevoerd in deze thesis en die het resultaat is van een collaboratie tussen onderzoeksgroepen aan de *Vrije Universiteit Brussel*, het *Institut Pluridisciplinaire Hubert CURIEN* te Strasbourg en *Chonbuk National University* in Zuid-Korea, bleek dat een zoektocht naar de FCNC koppeling tussen een top quark en Higgs boson in proton botsingen bij een energie van 13 TeV gevoelig kan zijn aan voorspellingen van bepaalde nieuwe theorieën. Verder bleek dat van alle zoek-kanalen naar deze specifieke

169

koppeling, het geval waar het Higgs boson vervalt naar twee bottom quarks een zeer krachtig zoek-kanaal is. In deze thesis zijn FCNC onderzocht in de hoogst-energetische proton botsingen ooit vastgelegd door mensen, geproduceerd door de Large Hadron Collider en geregistreerd door het CMS experiment te CERN in 2016. Deze protonen worden gebotst aan een energie van 13 TeV en de onderzochte data set, zoals geregistreerd door het CMS experiment, vertegenwoordigt een geïntegreerde luminositeit van 36 fb^{-1}. Het grootste gedeelte van deze data set is irrelevant voor de zoektocht naar FCNC en om zulke data substantieel uit te sluiten, wordt er een reeks selectie criteria opgelegd aan elke botsing. Daarbovenop worden botsing gebeurtenissen gecategoriseerd in controle en signaal categorieën op basis van hun topologie. De categorieën zijn zodanig gedefinieerd dat de bijdrage van een van de volgende processen gemaximaliseerd wordt: de anomale FCNC productie van een enkel top quark, de productie van een koppel top quarks waarbij er een vervalt via een FCNC decay, of de productie van een top quark koppel waarbij beide quarks vervallen via gekende Standaard Model processen. Dit is de eerste keer dat de anomale FCNC productie van een enkel top quark wordt onderzocht, mogelijk gemaakt door het gebruik van toegewijde categorieën en diens voorspelde werkzame doorsnede.

Een toegewijde studie is uitgevoerd om een hoog-performant algoritme op te stellen, van waar het onderzochte FCNC proces kan gereconstrueerd worden. Dit algoritme wordt gebruikt om bepaalde karakteristieken van het FCNC proces te reconstrueren. Op basis van deze karakteristieken worden variabelen opgesteld die een discriminerende kracht bieden tussen FCNC en achtergrond processen. Een reeks van zulke variabelen is gedefinieerd voor elke categorie en wordt gebruikt als basis om, via machinaal lerentechnieken, een meester variabele op te stellen. Deze meester variabele combineert en optimaliseert de discriminerende kracht van elke variabele. Vanuit simulaties kan een voorspelling gemaakt worden van hoe de verdeling voor die meester variabele eruit ziet, in zowel de achtergrond als achtergrond plus signaal hypothesen. Deze voorspelde verdeling wordt vergeleken met de verdeling zoals verkregen uit de echte data.

Er zijn geen afwijkingen teruggevonden in data van de verdelingen van de meester variabelen ten opzichte van de achtergrond hypothese. Dit betekent dat FCNC processen waarbij een top quark en een Higgs boson betrokken zijn, niet waargenomen zijn. Er zijn echter wel strikte limieten gesteld op het voorkomen van zulke FCNC processen, op basis van een template fit. De geobserveerde (verwachte) bovenlimiet bij een 95% vertrouwensniveau op de vertakkingsvoorkoming van een top quark t naar een up quark u en een Higgs boson H is $\mathcal{B}(\mathbf{t} \rightarrow \mathbf{uH}) \lesssim \mathbf{4.4} \times \mathbf{10^{-3}}$ ($\mathbf{3.4} \times \mathbf{10^{-3}}$). De anomale FCNC productie van een enkel top quark verbetert de limieten op $\mathcal{B}(t \rightarrow uH)$ met ongeveer 21% ten opzichte van het scenario waar dit signaal niet beschouwd wordt. Dit illustreert de toegevoegde waarde van het inclusief maken van deze analyse aan de anomale FCNC productie van een enkel top quark. De geobserveerde (verwachte) bovenlimiet bij een 95% vertrouwensniveau op de vertakkingsvoorkoming van een top quark t naar een charm quark c en een Higgs boson H is $\mathcal{B}(\mathbf{t} \rightarrow \mathbf{cH}) \lesssim \mathbf{4.7} \times \mathbf{10^{-3}}$ ($\mathbf{4.7} \times \mathbf{10^{-3}}$). Vergeleken met een simpelere snij-en-tel methode, zijn de limieten van

de template fit meer dan twee keer zo strikt. De limieten die afgeleid zijn met de ontwikkelde methoden in deze thesis, verbeteren significant de limieten van de zoektocht naar exact hetzelfde signaal aan het CMS experiment waarbij botsingen aan 8 TeV werden gebruikt. Daarbovenop staan de limieten op t $\rightarrow$ cH aan de rand van het ontdekken (of uitsluiten) van voorspellingen van bestaande nieuwe-fysica modellen.

Acknowledgements

Wow, I did it! I cannot express in any kind of phrasing how relieved and happy I am by having written this thesis, so I won't even try. Although, there is a BIG nuance to be made in the "I" did it, because this thesis is not just the result of my individual efforts. Many people, in many different ways, have contributed to the quality of this thesis and supported me throughout my whole career as a PhD student. For that support I am extremely thankful and I will try to convey my gratitude to all those people here.

Let me start by thanking my promotor Jorgen D'Hondt, whose influence in my career as a particle physicist is unmistakable. There is of course the obvious fact that you were so kind to take me in as a PhD student, but that deed of kindness pales in comparison to how much your enthousiasm drove me to explore the various aspects of particle physics, not just during my PhD career but already in my first years as a Bachelor student. The brainstorm sessions we had were always fruitful, enjoyable and a reminder of how lucky I was to learn from your expertise and refreshing ideas. In the same spirit, I owe a big thank you to my co-promotor Kirill Skovpen as well. Although the start of our collaboration on the CMS analysis only started in June 2015, your influence on the work presented in this thesis is undeniable. We have worked closely ever since on the analysis, a collaboration that only intensified when you joined the office. Your humerous spirit lightened the heavy analysis load (along Abbey Road). A thank you goes out as well to all other members of our analysis group: Predrag, Milos, Gerrit, Junghwan and Taejeong. It was a really pleasant experience to work with the likes of you, making our analysis very dynamic. Furthermore, the feedback from the TOP PAG conveners, Rebeca and Pedro, and members of the analysis review committee, Luca, Javier, Benedikt and Benjamin, was highly appreciated. I would also like to thank the members of the jury for reading the thesis and the valuable discussion during the private defence. Your input and comments undoubtedly were of great benefit to the thesis.

It is often said of a PhD research that it is lonely work. I am happy to state that this is as far from the truth as it can be in my case. Due to the extensiveness of the CMS collaboration and the pheno-collaboration, I was lucky to meet many new people with different cultural backgrounds. I can honestly say that I feel enriched on a cultural and personal level by these international contacts and am happy of the great opportunities these brought forward, like attending a workshop in South-Korea. If there is one thing I will never forget about my time as a PhD student, it is the

amazing atmosphere in the office. Many faces have come and gone in those four years I occupied space in that humongous office, but each and everyone of you really left their mark. I would like to thank all of you for making the office such a great place to work, discuss physics, share laughs, eat cake... Thank you to Gerrit, James, Quentin, Taejong, Gregory, Nadir, Natalie, Annik, Petra, Shimaa, Isabelle, Giannis, Seth, Jarne, Dom, Doug, Simon, Kirill and Denys. You are all wonderful people and I wish you the best of luck in all your future endeavours. From among all of you with whom I've shared the office, there are two ladies who deserve a special acknowledgement, namely Isis and Lieselotte. Our shared experience goes back to the beginning of our physics studies, in 2008. Nine years we have walked the same path: we have worked together, studied together, taught together and even lived together for two months at CERN. I am really glad we have shared those moments and look forward to share some more.

I would be lying if I said these were the only ones who made life fun at the VUB. There are so many more amazing colleagues at the Interuniversity Institute for High Energies. There's of course Marleen, whose cheerful personality is always enjoyable and who eased all the administrative worries. The computing group has done an amazing job in maintaining our computing infrastructure, but also in providing personal technical support. Thank you Olivier, Romain, Samir, Shkelzen and Stephane.

Luckily, my years as a PhD student weren't all PhD research, or I would have gone mad. I was also happy to be able to set my worries aside through all my hobbies. Badminton has been a big aspect of my life since quite some years now and I'm very happy with all the matches I played with my friends at the BAMADI team. Also my music band, Eraser, contributed to necessary distractions. Although the number of rehearsals were very limited the past years, I always felt great relief when they happened. Not to forget, playing football with colleagues/friends at APNELEM United was very amuzing as well. We could have won every match if it weren't for the opponent.

Last, but not least, I would like to express my extreme gratitude to my family. My parents and brother have been my greatest supporters, but especially in encouraging me to pursue my academic goals. I can't thank you enough for everything you've done for me, but know you planted the seeds for this success. My wife, who has beared with me during all my grievances, I laurel with the greatest praises. Thank you for understanding that writing a PhD comes at the cost of postponing some chores, but most of all thank you for your support. There is one boy, whose cheerfulness and great smile have a constant positive influence on my mood. I am proud to be the godfather of that boy, Thorfinn. Above all, though, there is one girl whose size is inversly proportional to the influence she has had on my life, my beautiful, lovely and fun daughter Juune. You are truly amazing and I dare to say that from the day you entered this world, my world got so much bigger and so much more enriched. Your coming helped me get better organized and your smile after a hard day's work made me oblivious to all my worries. Maybe there will be a day when you will understand the contents of this thesis, but I don't really care about that. I feel extremely fortunate to have you as my daughter.