VRIJE UNIVERSITEIT BRUSSEL

Faculty of Science and
Bio-engineering sciences
Physics Department

AIN SHAMS UNIVERSITY

Faculty of Women for Arts,
Science and Education
Physics Department

Search for Top Quark Flavour Changing Neutral Couplings with the CMS Experiment at the LHC

Shimaa AbuZeid

Promotor: Prof. Dr. Jorgen D'Hondt

Promotor: Prof. Dr. Magda Abdelwahab

A dissertation submitted to Ain Shams University and Vrije Universiteit Brussel in accordance with the requirements of the double-degree of PhD
PhD in Physics, Ain Shams University,
PhD in Sciences, Vrije Universiteit Brussel

June 2018

Doctoral examination commission:

Chair: Prof. Dr. Freya Blekman (Vrije Universiteit Brussel)
secretary: Prof. Dr. Alberto Mariotti (Vrije Universiteit Brussel)
Supervisors:
Prof. Dr. Jorgen D'Hondt (Vrije Universiteit Brussel)
Prof. Dr. Magda Abdelwahab (Ain Shams University)
Others:
Prof. Dr. Manal Mahmoud Sirag (Ain Shams University, Egypt)
Prof. Dr. Pascal Vanlaer (Université Libre de Bruxelles, Belgium)
Prof. Dr. Benjamin Fuks (Sorbonne University, France)

Date of Private Defence: 25 May 2018.
Date of Public Defence: 27 June 2018.

Cover information: The photo in the front cover is for the first half of the inner tracker barrel of CMS experiment, it is taken from [1]. The symbols in the back cover means "top-FCNC" in old Ancient Egyptians language.
The cover is kindly designed by my friend Dr. Haiam Adel.

Contents

Contents **iii**
Introduction v

1 Theoretical Basis and Concepts **3**
1.1 The fundamental particles and forces 3
1.2 The Standard Model (SM) 5
1.2.1 Electroweak (EW) interactions 7
1.2.2 Flavour changing interactions in the SM 8
1.2.3 Shortcomings of the SM 9
1.3 The top quark 11
1.3.1 FCNC in the SM 12
1.3.2 FCNC beyond the SM 14
1.4 Overview of experimental searches on top-FCNC 17

2 The CMS Experiment at the LHC **25**
2.1 The CERN accelerator complex 25
2.1.1 The LHC parameters 27
2.2 The CMS Experiment 28
2.2.1 Overview of the CMS detector 29
2.2.2 The Silicon Trackers 29
2.2.3 The Electromagnetic Calorimeter (ECAL) 34
2.2.4 Hadronic Calorimeter (HCAL) 37
2.2.5 The muon detection System 38
2.2.6 Data acquisition 39
2.2.7 Computing at CMS 40

3 Event Generation **45**
3.1 The physics of proton-proton collision 45
3.1.1 Parton Distribution Functions (PDF) 47
3.1.2 Hard scattering 49
3.1.3 Parton Shower 50
3.1.4 Hadronization 51
3.1.5 Underlying event 51
3.1.6 Pile-up 52
3.2 Event generation and detector simulations 52

3.2.1 MC Generators . . . 52
3.2.2 Detector simulators . . . 54
3.3 Signal and background generation . . . 55
3.3.1 FCNC signal generation . . . 57

4 Reconstruction and Identification of Physics Objects 61
4.1 Primary Vertex and track reconstruction . . . 61
4.2 Particle-Flow reconstruction . . . 63
4.3 Muons reconstruction . . . 64
4.3.1 Muon identification and isolation . . . 65
4.4 Electrons reconstruction . . . 66
4.4.1 Charge estimation . . . 69
4.4.2 Electron identification . . . 69
4.5 Jet reconstruction . . . 70
4.5.1 Jet identification and jet-energy corrections . . . 72
4.5.2 Jet energy resolution correction (JER) . . . 74
4.6 Identification of heavy-flavour jets (b & c jets) . . . 75
4.7 Missing energy reconstruction . . . 83

5 The FCNC analysis strategy and techniques 87
5.1 Event pre-selections and cleaning . . . 88
5.1.1 Triggers . . . 90
5.1.2 Filters . . . 90
5.1.3 Jet and lepton cleaning . . . 92
5.2 Baseline selections . . . 93
5.2.1 Lepton selection . . . 93
5.2.2 Jet selection criteria . . . 96
5.2.3 The signal and control regions . . . 96
5.2.4 Corrections and scaling factors . . . 97
5.2.5 Event yields . . . 100
5.3 Discrimination and optimization . . . 101
5.3.1 The Boosted Decision Trees (BDT) . . . 103
5.3.2 The limit setting tool . . . 105

6 The search for FCNC interactions involving a top quark and a H boson 111
6.1 Signal to background discrimination with a BDT method . . . 111
6.2 Systematic uncertainties . . . 126
6.2.1 The effect of systematic uncertainties on backgrounds . . . 128
6.2.2 The effect of systematic uncertainties on the signal . . . 136
6.3 Validation of the limit setting procedure . . . 137
6.4 Results and discussion . . . 138
6.4.1 Postfit distributions and yields . . . 142
6.4.2 Obtained limits at 95% CL . . . 146

7 Conclusion and Prospects **153**
7.1 Conclusion 153
7.2 Prospects 155

A Apppendix A: Variables used in b-tagging algorithms **159**

B Apppendix B: Distributions related to the BDT Training **161**

Achievements and contributions **169**

Bibliography **173**

List of abbreviations **187**

Summary **191**

Samenvatting **193**

Arabic Summary **195**

Acknowledgements **199**

"Knowledge is the conformity of the object and the intellect"

Ibn Rushd (Averroes)

Introduction

Since the beginning of life, curiosity guides humans to discover the universe and develop knowledge regarding it's existence. Historically the ancient Egyptians and the ancient Greeks are the first people who wrote and documented their beliefs about the creation of the universe. Both gave religious or mythological explanations for natural phenomena, proclaimed that every event had a natural cause.

The universe according to ancient Egyptians beliefs is created by a self-created deity god called "Atum". Atum represents the god of preexistence and post-existence where the energy and matter are contained. Atum created himself from the dark primordial waters "Nu". In this abyss everything was formless and inert so he creates the first piece of land "ben-ben" to stand on it, then it creates two other gods "Shu" and "Tefnut" that represent the gods of air and moisture respectively. Then by marriage of these two gods the family increased to be nine gods. Atum then arises to the sky in order to light his creation on the land. On the other hand the ancient Greeks believed that everything is made from four basic elements. These elements are "air, water, earth and fire". The basic idea of both Egyptian and Greek beliefs indicates that both agree that there are one or more basic building blocks of the universe. Now this idea evolved via thousands of scientists from different countries.

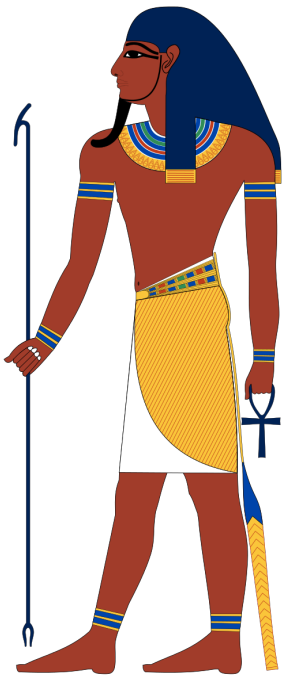


Figure 1: Atum: The god of creation in the ancient Egyptians beliefs.

During the twentieth century, several experimentally confirmed theories gave a description of matter and its composition. These theories are combined into the most successful physics theory called the Standard Model (SM) of particle physics, which explains almost all experimental results and precisely predicted many phenomena. Unfortunately the SM is not the theory of everything, some phenomena are not embraced in the SM like gravity, the composition of dark matter and the neutrino masses. These shortcomings of the SM motivate scientists to search for new physics Beyond the Standard Model (BSM). The main purpose of these BSM theories is to establish a more general understanding of particle physics especially in the

conditions where the SM fails, such as at very high energies.

The top quark is an excellent candidate for searching for such new physics phenomena. Since it is the heaviest particle in SM, physicists believe that it has an enhanced sensitivity to various new particles and interactions suggested by BSM theories. One of interesting phenomenon is the presence of flavour-changing neutral current (FCNC) interactions involving the top- quark in which the flavour of the top quark is changed by means of neutral bosons. The FCNC interactions are highly suppressed in the SM by the so-called Glashow–Iliopoulos–Maiani (GIM) mechanism [2].

FCNC interactions were confirmed for the first time in 2005 by the Collider Detector at Fermilab (CDF) experiment in the $B_s^0 \to \phi\phi$ decay process[3]. The results of this search matched the SM predictions. After the construction of the Large Hadron Collider (LHC) the FCNC interactions involving b-quarks are observed in CMS and LHCb experiments in $B_s^0 \to \mu^+\mu^-$ [4]. This observation also agrees with the SM prediction and has put stringent limits on BSM theories.

Although the FCNC interactions of the top quark are suppressed in the SM by the GIM mechanism, many BSM theories predict the existence of these processes with higher branching ratios, where some are within the reach of the current experiments. The observation of a SM-like Higgs boson by the ATLAS and CMS experiments in 2012 [5] initiated the study of FCNC interactions involving a top quark and a Higgs bosons. Up to date the results of these studies are still far from being sensitive to Standard Model predictions. This leaves a large phase space to confirm or exclude the presence of such new physics phenomena.

In this thesis a research for FCNC interactions involving a top quark and a Higgs boson is presented with a signature of two leptons in the final state with equal electric charges. The analysis is done using data of pp-collisions collected in 2016 by the CMS experiment at a center-of-mass energy of 13 TeV and an integrated luminosity of $36 fb^{-1}$. We probe the FCNC interactions in top-quark pair decays as well as in the single top quark production through Hqt-couplings, where the quark q is either an up-quark or a charm-quark. The thesis has seven building blocks called chapters which explain the process of the doctoral study. This process needs some theoretical bases which will be introduced in the first chapters. In order to perform this study an experimental setup has to work with certain conditions that will be described in the second chapter. The skeleton of the proton-proton collisions will be demonstrated in chapter three. The reconstruction of the final state particles which are emerged from these collisions will be explained in chapter four. The analysis of the collected information will be illustrated in chapters five and six. Finally the author presents conclusions and prospects in chapter seven.

"It is inconceivable that inanimate brute matter should, without the mediation of something else which is not material, operate upon and affect other matter without mutual contact"

Isaac Newton

Chapter 1

Theoretical Basis and Concepts

One of the recurrent dreams in elementary particles physics is that of a possible fundamental synthesis between electromagnetism and weak interactions.
This qoute had been said by A. Salam in Weak and Electromagnetic Interactions Elementary Particle Theory (1968) [6].

1.1 The fundamental particles and forces

Almost all commonly encountered physical phenomena can be described in terms of the electron, electron neutrino, proton and neutron, interacting by the electromagnetic, strong and weak forces.

The negative electrons are bound to the positive nucleus via the electrostatic attraction, which is the low-energy manifestation of the fundamental theory of electromagnetism, namely Quantum Electrodynamics (QED). Inside the atomic nucleus, the protons and neutrons are bound together by the strong nuclear force, that is a manifestation of the fundamental theory of strong interactions, called Quantum Chromodynamics (QCD). The fundamental interactions of particle physics are completed by the weak force, which is responsible for the nuclear β-decays of radioactive isotopes in which the electron and another particle called the electron neutrino ν_e are produced. This picture is completed by gravity, which although extremely weak, is always attractive and is therefore responsible for large-scale structures in the Universe. This represents a simple physical model with just four "fundamental" particles and four fundamental forces. However, at higher energy scales, further structures are observed. For example, the protons and neutrons are found to be bound states of smaller fundamental particles called quarks [7].

The Standard Model (SM) of particle physics describes all the known elementary particles and their interactions. It provides a successful description of all current experimental data and represents one of the triumphs of modern physics.

Within this model all particles are built from a number of fundamental, spin $\frac{1}{2}$, particles called fermions (quarks and leptons) and the interactions among those particles take place through bosons, which are described in terms of particles of spin 1 and work as mediators of forces between the fermion constituents. For each fermion there

is an anti-fermion, which has the same mass but it is oppositely charged. Furthermore, quarks carry colour charge while leptons do not. From the upper limits on the possible neutrino masses, neutrinos ν have been recently observed to have non-zero mass unlike previously assumed. They have masses at least nine orders of magnitude lighter than the other fermions. Furthermore, because they are neutral, the question whether each neutrino is its own antiparticle (Majorana particles) or not (Dirac particles) remains unanswered. In general the different types of quarks and leptons are also characterized as flavours.

According to the SM fermions are categorized into three generations, each identical to the other except for their mass and flavour quantum numbers. The first generation combines the up and down quarks, together with the electron and the electron-neutrino, while the second generation consists of the charm and strange quarks, muon and muon-neutrino and finally, the top and bottom "or beauty" quarks together with the tau lepton and tau-neutrino make up the third generation as can be seen in Figure 1.1. Ordinary matter on earth is essentially composed of particles from the first generation: up and down quarks in the nucleus, and electrons in the electron shells.

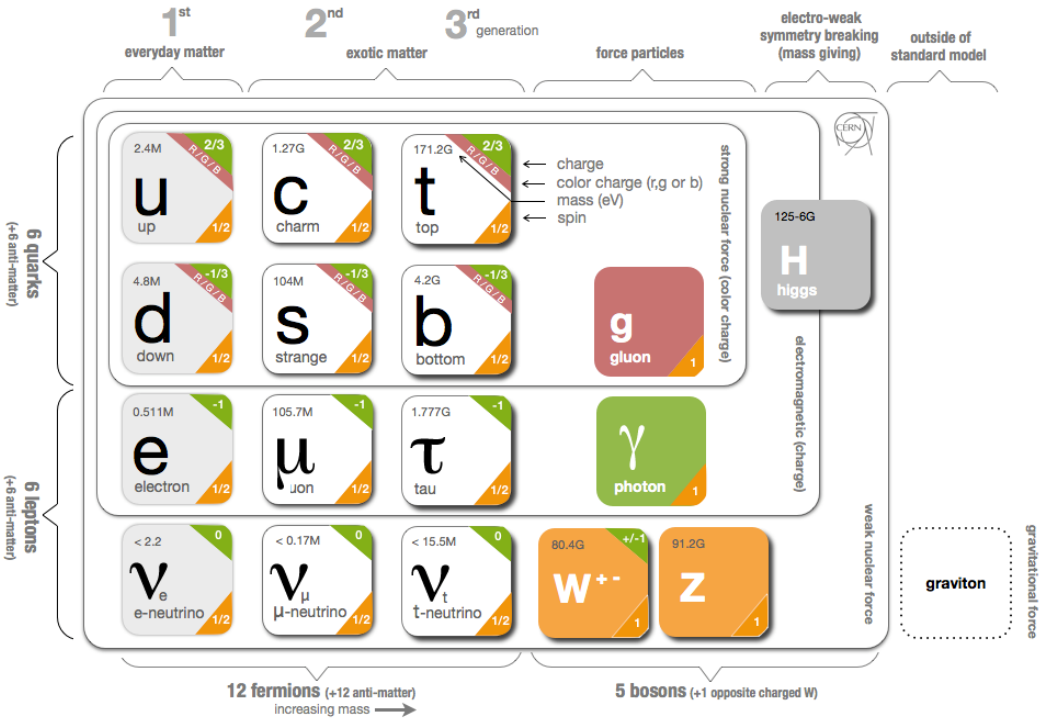


Figure 1.1: The particle content of the SM. All twelve fundamental particles carry the charge of the weak interaction, known as weak isospin, so they undergo weak interactions. With the exception of the neutrinos, the other nine particles are electrically charged and participate in the electromagnetic interaction of QED. Only the quarks undergo the strong force interactions because they carry colour charge (the QCD equivalent of electric charge). The figure taken from CERN website [8]

Each of the three forces of relevance to particle physics is described by a Quantum Field Theory (QFT) corresponding to the exchange of a spin-1 force-carrying particle,

known as a gauge boson. The strong interaction is responsible for binding the quarks. This interquark force is mediated by gauge bosons called "gluons" which are massless, electrically neutral but carry colour charge. The massless photons γ, mediate the electromagnetic interaction between charged particles. The quanta of the weak interaction fields are the charged $W^{\pm}$-boson and the neutral Z^0-boson. These last two carry mass and make the weak interaction short-ranged. On the contrary, the electromagnetic interaction has an infinite range because of the massless mediator. Gluon fields despite being massless are confined, so the strong force is not observed as a long range force. The SM is completed by the discovery of the Brout-Englert-Higgs boson (commonly called Higgs boson) which has been confirmed by the CMS and the ATLAS experiments in July 2012 [5]. The Higgs particle is the only known spin-0 boson, also called a scalar particle, and is responsible for giving the mass of all aforementioned particles.

Due to the QCD confinement of quarks, quarks can not exist as free particles so they are bound through gluons into *hadrons* which are divided into two categories. *Baryons*, such as protons and neutrons, are composed of three quarks while *mesons*, such as π^0 and K^0 are composed of a quark and an anti-quark. Almost all baryons (except the proton) and mesons are unstable (short-lived) particles that decay via $W^{\pm}$ or Z^0 bosons, the latter particles usually decay into a charged lepton and neutrino or into quarks. Z^0 can also decay into 2 charged leptons.

1.2 The Standard Model (SM)

The SM is a gauge field theory where the particles are represented by fields. Their interactions and kinematics are described by mathematical constructions called Lagrange densities or shortly Lagrangians. The SM Lagrangian is invariant under Lorentz transformation that implies the conservation of energy and momentum.

The SM theory is based on the $SU(3)_C \times SU(2)_L \times U(1)_Y$ group. $SU(3)_c$ describes the strong interactions and the index C refers to the color charge which is conserved under $SU(3)$ symmetry. The conservation of color charge implies the existence of eight gauge bosons that are gluons. Gluons interact with each other in a similar way as quarks do. Similarly to quarks they can not be found freely. Under $SU(3)_c$ quarks are color triplets while leptons are color singlets; therefore quarks carry a color index ranging between one and three, whereas leptons do not take part in strong interactions.

$SU(2)_L \times U(1)_Y$ describes the non-abelian gauge group of the electroweak interaction. L refers to the left chiral nature of the $SU(2)$ coupling and Y refers to the weak hypercharge which is carried by the gauge field B_μ and is related to the electric charge Q and the weak isospin T_3 by

$$Y = 2(Q - T_3). \tag{1.1}$$

The three gauge fields W^1_μ, W^2_μ and W^3_μ are associated to $SU(2)_L$ with three generators that can be written as half the Pauli matrices:

$$T_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, T_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, T_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.2}$$

The generators T^a satisfy the Lie algebra:

$$[T_a, T_b] = i\epsilon_{abc}T^c \ , \ [T_a, Y] = 0 \tag{1.3}$$

where ϵ_{abc} is a fully anti-symmetric tensor.

In the SM, fermions (quarks and leptons) are described by matter fields or Dirac (spinor) fields, denoted by ψ, that are split in left and right handed parts (ψ^L & ψ^R). Left handed charged leptons and the corresponding neutrino spinors are grouped in weak isospin doublets. Similarly for the quark fields, where left handed up-type quark fields are grouped with the corresponding down-type quark fields.

$$\Psi_q^L = \begin{pmatrix} \psi_{u_q}^L \\ \psi_{d_q}^L \end{pmatrix}, \psi_{u_q}^R, \psi_{d_q}^R. \tag{1.4}$$

Similarly for leptons

$$\Psi_l^L = \begin{pmatrix} \psi_{\nu_l}^L \\ \psi_l^L \end{pmatrix}, \Psi_l^R = l_R \tag{1.5}$$

where u, d represent up-type and down-type quarks, Ψ_l^L is the left-handed lepton field, l is an index over the three lepton families $(e, \nu_e), (\mu, \nu_\mu), (\tau, \nu_\tau)$ and Ψ_l^R is the right-handed field. One should notice that in the SM there are no right-handed neutrinos. The SM Lagrangian can be written in a condensed form as [9]:

$$\begin{aligned} \mathcal{L}_{SM} &= \mathcal{L}_{fermions} + \mathcal{L}_{bosons} + \mathcal{L}_{Yukawa} + \mathcal{L}_{Higgs} \\ &= i\bar{\Psi}^i\gamma^\mu D_\mu\Psi^i - \frac{1}{4}F^a_{\mu\gamma}F^{a\mu\gamma} + i\bar{\Psi}^i Y_{ij}\Psi^j\Phi + (|D_\mu\Phi|^2 - V(\Phi)) \end{aligned} \tag{1.6}$$

where the first term describes the interaction between the fermions and gauge bosons, and Ψ^i is given by Equations 1.4 and 1.5 for quarks and leptons respectively [10]. γ^μ represents the Dirac matrices [1] and D_μ is the covariant derivative is given by:

$$D_\mu = \partial_\mu - \frac{1}{2}ig_1 Y B_\mu - ig_2\mathcal{T}_k W_\mu^k \ - \frac{1}{2}ig_3\lambda_k G_\mu^k, \tag{1.7}$$

where g_1, g_2, g_3 are the coupling strengths and Y, $\mathcal{T}_k$, λ_k are the hypercharge, Pauli matrices and the Gell-Mann matrices representing the generators of $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$ respectively.

The second term of Equation 1.6 describes self-interactions of the gauge bosons $F^a_{\mu\gamma}$ is called the stress-energy tensor. The third term represents the Yukawa sector, which is responsible for giving masses to fermions and to $W^\pm$ and Z^0 bosons. V and Φ are the Higgs potential and field respectively. The fourth term represents the Higgs sector which is responsible for the spontaneous symmetry breaking that will be described in next section.

[1] Dirac matrices are 4×4 matrices given by $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$, $g^{\mu\nu} = diag(+ - - -)$ Minkowski metric.

1.2.1 Electroweak (EW) interactions

The theory of electroweak (EW) interactions arises from the unification of the electromagnetic and weak interactions. It was proposed by Steven Weinberg and Abdus Salam independently in 1967 [6, 11]. It is also called the Glashow-Salam-Weinberg theory of electroweak interactions, since it came after the proposal of Sheldon Lee Glashow for the $SU(2) \times U(1)$ symmetry group in 1961 [12].

In the EW theory the chirality and the isospin T_3 are strongly related. The fermions can be grouped according to their isospin as given by Equations 1.4 and 1.5. The left-handed fermions carry an isospin of $T_3 = \pm 1/2$ while the right-handed fermions have $T_3 = 0$. Therefore left-handed fermions form isospin doublets of $SU(2)_{T_3}$ while right handed fermions form singlets of $SU(2)_{T_3}$. Up-type fermions $(u, c, t, \nu_e, \nu_\mu, \nu_\tau)$ have $T_3 = +1/2$ while the down-type fermions (d, s, b, e, μ, τ) have $T_3 = -1/2$.

According to the $SU(2)_L \times U(1)_Y$ symmetry group of the EW interactions there are four massless gauge bosons denoted as $(W^1_\mu, W^2_\mu, W^3_\mu)$ that arise from the $SU(2)_L$ symmetry and B_μ which arises from the $U(1)_Y$ symmetry. These four gauge fields are represented in the 1^{st} and 2^{nd} terms of Equation 1.6. The linear combination of these massless gauge bosons gives as a result the $W^\pm$ and Z^0 bosons and the photon as described by the following equations:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \tag{1.8}$$

$$W^\pm_\mu = \sqrt{\frac{1}{2}}(W^1_\mu \mp i W^2_\mu),$$
$$\tan\theta_W = \frac{g\prime}{g} \tag{1.9}$$

where A_μ and Z_μ represent the photon and the Z boson respectively. The coupling strengths g, for $SU(2)_L$, and $g\prime$, for $U(1)_Y$, of W^3_μ and B_μ are related to the weak mixing angle, Weinberg angle, θ_W.

The EW symmetry breaking

It was proved and verified experimentally that the $W^\pm$ and Z^0 bosons, quarks, charged leptons and neutrinos have mass. However without the fourth term in SM Lagrangian in Equation 1.6 all these particles are massless. These particles can gain mass only via spontaneous symmetry breaking through what is called the Brout-Englert-Higgs mechanism [13, 14] . The Higgs term of the SM Lagrangian contains a scalar field Φ, which is a weak isospin doublet given by

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.10}$$

The Higgs potential $V(\Phi)$ in Equation 1.6 can be written in the form

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2. \tag{1.11}$$

The self-interaction of the scalar field is described by the term λ which should always be positive. For $\mu^2 < 0$ the scalar fields will gain a non-zero vacuum expectation value (vev) and therefore the electroweak symmetry is broken. The minimum energy states that satisfy $\Phi^\dagger\Phi = vev/2$ where $vev = \sqrt{\mu^2/\lambda}$ are infinitly degenerate. One can choose the scalar field as

$$\Phi = \begin{pmatrix} 0 \\ \frac{vev}{\sqrt{2}} \end{pmatrix} + \phi. \tag{1.12}$$

This choice allows one to expand the Higgs term of the SM Lagrangian in a perturbative series of the scalar field Φ around the (vev) which is measured to be around 246 GeV [15]. Consequently the $W^\pm$ and Z^0 bosons gain masses that can be given as:

$$m_W = \frac{vev}{2}g \qquad , \qquad m_Z = \frac{vev}{2}\sqrt{g^2 + g'^2}. \tag{1.13}$$

1.2.2 Flavour changing interactions in the SM

Since the weak interaction is the only known force for which the incoming and outgoing fermions are different, the weak charged current interaction is particularly important when considering particle decays as it introduces a change of flavour.

Experimentally it was observed that the coupling of quarks through $W^\pm$ bosons changes the flavour of the quark. This coupling can change the quark to another quark in the same or a different generation. However the coupling between flavours of quarks from different generations is suppressed compared to the flavour changing process within the same generation.

In 1963, Nicola Cabibbo proposed the first description of flavour changing. However at that time only few quarks (u, d, s) were known, therefore he described the charged weak current as a coupling between the up (u) quark and (d_{weak}), where d_{weak} is a linear combination of the down (d)-quark and the strange (s) quark ($d = \cos\theta_c d + \sin\theta_c s$). The linear combination is a direct consequence of the chosen rotation:

$$\begin{pmatrix} d_{weak} \\ s_{sweak} \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} = \mathcal{R} \begin{pmatrix} d \\ s \end{pmatrix} \tag{1.14}$$

where θ_c is the rotation angle, also known as the Cabibbo angle. $\mathcal{R}$ represents the rotational matrix. This defines the charged weak current between u and d quarks as:

$$J_\mu = \bar{u}\gamma_\mu(1 + \gamma_5)d_{weak}, \tag{1.15}$$

where γ_μ is the Dirac matrix and γ_5 matrix is defined as $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. However in this approach s_{weak} is left uncoupled. This leads three scientists [2] to postulate the existence of a fourth quark which can couple to the s_{weak} quark [2, 16]. This quark is known as the charm (c) quark which was then discovered in 1974 in the J/Ψ-meson decay [17, 18]. Therefore the new definition, actually it is a redefinition of Equation 1.15, for the charged current was introduced to be:

$$J_\mu = \begin{pmatrix} \bar{u} & \bar{c} \end{pmatrix} \gamma_\mu(1 + \gamma_5)\mathcal{R} \begin{pmatrix} d \\ s \end{pmatrix} \equiv \bar{U}\gamma_\mu(1 + \gamma_5)\mathcal{R}D. \tag{1.16}$$

[2] The three scientists are Glashow, Iliopoulos and Maiani, hence the name of GIM mechanism.

where the matrices $\bar{U}$ and D represent the column matrices of the up-type quarks (u and c) and the down-type quarks (d and s) respectively. The neutral (weak) current is defined as:

$$J_3 = \bar{D}\gamma_\mu(1+\gamma_5)(\mathcal{R}^\dagger\mathcal{R})D \tag{1.17}$$

This is diagonal in flavour space which means that no flavour changing neutral current (FCNC) at tree-level exist in the SM.

The introduction of a third generation of quarks to the SM leads Kobayashi and Maskawa to generalize the Cabibbo rotation matrix to be

$$\begin{pmatrix} d_{weak} \\ s_{weak} \\ b_{weak} \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \mathcal{V}_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.18}$$

where $\mathcal{V}_{CKM}$ is a 3×3 matrix, called the CKM matrix. It controls the mixing of the weak states of down-type quarks. $\mathcal{V}_{CKM}$ is a unitary matrix, $\left(\mathcal{V}^\dagger_{CKM}\mathcal{V}_{CKM} = \mathcal{I}\right)$, which implies that the sum of the squared elements in one row (column) equals unity. The squared values of the matrix elements V_{ij} represent the transition probability of a quark i going to a quark j, and it is experimentally measured to be [15]:

$$\mathcal{V}_{CKM} = \begin{pmatrix} 0.97425 \pm 0.00022 & 0.2253 \pm 0.0008 & (4.13 \pm 0.49) \times 10^{-3} \\ 0.225 \pm 0.008 & 0.986 \pm 0.016 & (41.1 \pm 1.3) \times 10^{-3} \\ (8.4 \pm 0.6) \times 10^{-3} & (40.0 \pm 2.7) \times 10^{-3} & 1.021 \pm 0.032 \end{pmatrix} \tag{1.19}$$

The numbers given in the matrix are obtained through a global fit from the combination of many experiments. One can directly observe that the diagonal elements of the $\mathcal{V}_{CKM}$ are close to one, whereas the other elements are small. This means that mixing between quarks of different generations is suppressed in the SM. Also the mixing of the first two generations is significantly larger than the mixing of the second and third generations. The mixing of the first and the third generations is the smallest.

1.2.3 Shortcomings of the SM

The SM is considered a very successful theory capable of describing many physics phenomena. Over many years several predictions made by the SM have been experimentally confirmed such as the top quark and the Higgs boson. However there are a number of unexplained theoretical and experimental phenomena which are not covered by the SM which can be summarized below.

- **Dark Matter problem**:
  According the SM of cosmology, the total mass-energy of the known universe contains 4.9% ordinary (baryonic) matter, 26.8% dark matter and 68.3% dark energy. However the density of cold non-baryonic matter is determined by the Plank experiment [19]. Cold dark matter is assumed only sensitive to the weak and gravitational force, which makes neutrinos to be the only possible candidates from the SM. However neutrinos are too light to account for the huge measured fraction of dark matter. Therefore other models are needed for that.

- **Neutrino oscillations**:
  According to the SM, neutrinos are massless. However experiments for solar, accelrators and reactors neutrinos have established that neutrinos can oscillate and can change flavour during flight. These oscillations are only possible if neutrinos have masses. Therefore the flavour neutrinos (ν_e, ν_μ, ν_τ) are linear expressions of the fields of at least three mass eigenstates (ν_1, ν_2, ν_3) [20, 21].

- **Matter-antimatter asymmetry**:
  According to the Big Bang theory an equal amount of matter and antimatter should have been created in the early universe. However, there is not much antimatter to be found. The amount of antimatter is extremely small compared to the amount of matter in the universe. This is known as the matter-antimatter asymmetry in the universe [22]. The SM does not predict this observation, neither provides a mechanism that could generate it.

- **Unification of forces**:
  The idea of the grand unification theory (GUT) where the three gauge interactions of the SM which define the electromagnetic, weak and strong interactions (or forces) are merged into one single force at some high unification or GUT scale, above which the three interactions unify [23]. This is not possible in the SM. Instead the SM prediction is that the three forces do not converge at the Planck scale.

- **Naturalness problem**:
  One of objections to the SM theory is the high number of arbitrary parameters. The SM has 19 free parameters that are not predicted by the theory. The arbitrarity of parameters in the SM introduces the problem known as the naturalness problem. A "natural" theory is characterized by free parameters with values almost of the same order of magnitude which is not the case of SM. In the SM the top quark mass is 10^5 times larger than the mass of up quark. Such huge difference in arbitrary parameters indicates unknown principles underlying a more complete theory that covers the SM.

- **Hierarchy problem**:
  The hierarchy problem can be described as the instability of the value of the Higgs boson mass (m_H) when radiative corrections are included in the presence of a physical energy scale cutoff, Λ_{UV}, taken at energies far above the electroweak scale. This is related to the large difference between the electroweak symmetry breaking scale ($\approx 10^2\,\text{GeV}$) and the Plank scale ($\approx 10^{19}\,\text{GeV}$) [24]. Beyond this Plank scale quantum and gravitational effects are not reconcilable. The hierarchy problem appears when calculating m_H to all orders of perturbation. This can be written as

$$m_H^2 = m_{H_0}^2 + \Delta m_H^2 \tag{1.20}$$

  where m_H is the physical mass of the Higgs boson, m_{H_0} its bare mass and Δm_H is quantum loop correction. The quantum loop correction from fermions coupling

to the Higgs boson is given by

$$(\Delta m_H^2)_f = \frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2, \tag{1.21}$$

where λ_f is the Yukawa coupling of the fermion. The dominant contribution is coming from the top quark because it has the largest Yukawa coupling due to its large mass. The the correction from scalar particle s of mass m_s that couples to the scalar field with a Lagrangian term $-\lambda_s|\Phi|^2|s|^2$ is given by

$$(\Delta m_H^2)_s = \frac{|\lambda_s|^2}{16\pi^2}\left(\Lambda_{UV}^2 - 2m_s^2 ln\left(\frac{\Lambda_{UV}}{m_s}\right) + ...\right) \tag{1.22}$$

As can be seen from these equations the correction term to m_H^2 is larger than the mass of the physical mass of the scalar boson. In order to cancel this divergence in Λ_{UV}^2 there must be new scalar particles at the TeV scale which couple to the scalar boson to avoid this fine-tuning.

1.3 The top quark

The top (t) quark is the heaviest known fundamental particle with a mass of around 172.5 GeV. Its existence was predicted in 1973 by Makoto Kobayashi and Toshihide Maskawa in order to explain the charge parity (CP) violations observed in kaon decay [25]. The top quark was discovered at Fermilab in March 1995 by the $D\emptyset$ and the CDF collaborations [26, 27]. Due to its huge mass compared to other quarks, the top quark has a very short lifetime ($\approx 5 \times 10^{-25} s$) which much shorter than the typical hadronisation process time. Therefore it is impossible for the top quark to form hadrons. The top quark is believed to be a sensitive probe to physics beyond the Standard Model given the closeness of its mass to the electroweak scale. The study of its properties is therefore an important topic in the experimental program of the LHC to test the SM.

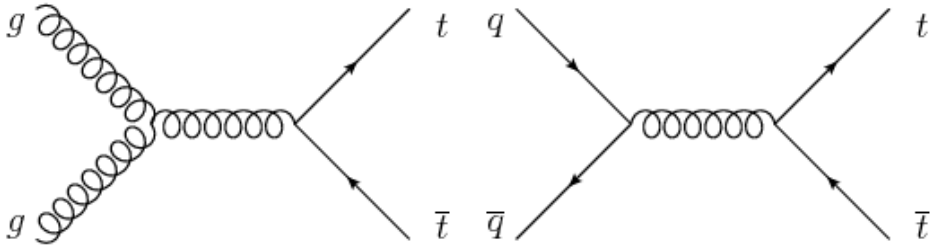
There are two main mechanisms to produce top quarks at the LHC:

1. **Via the strong interaction**: In this case a top and anti-top quark pair, $t\bar{t}$, is produced. At hadron colliders, the dominating ways to produce $t\bar{t}$ are via gluon fusion ($gg \rightarrow t\bar{t}$) or quark fusion ($q\bar{q} \rightarrow t\bar{t}$) as can seen in Figure 1.2. At the LHC, the gluon fusion is the main contributor to the $t\bar{t}$ production cross section compared to quark fusion at the Tevatron. The process $gg \rightarrow t\bar{t}$ contributes around 85% to the total $t\bar{t}$ production cross-section at the LHC at centre-of-mass energies of $7 - 13$ TeV.

2. **Via the electroweak interaction**: In this case a single top quark is produced. There are three processes, called channels, that can produce single top quarks, namely the t-channel production, the s-channel production and the Wt-channel production as shown in Figure 1.3 [3].

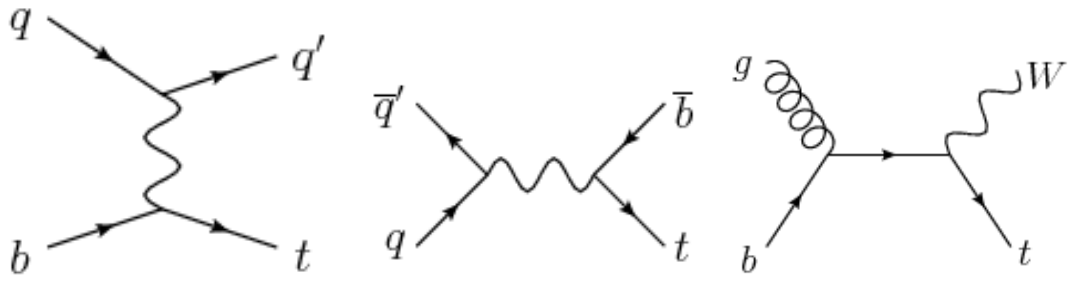
[3]The naming of s- and t-channels comes from the Mandelstam variables.

The single top-quark production cross-section is much smaller than the $t\bar{t}$ production cross-section and therefore it is more difficult to measure. An overview plot of the LHC and Tevatron top quark pair production cross section measurements as a function of the centre-of-mass energy is shown in Figure 1.4. These measurements are in agreement with their SM predictions.



(a) (b)

Figure 1.2: The figure illustrates the leading order Feynman diagrams of the $t\bar{t}$ production via gloun fusion at (a) and via quark and anti-quark fusion at (b).



(a) (b) (c)

Figure 1.3: The figure illustrates the leading order Feynman diagrams of the electroweak production of single top quarks in t-channel at (a), s-channel at (b) and Wt-channel at (c).

From the CKM matrix , Equation 1.19, we learn that top quarks predominantly decay via charged weak currents, and due to the value of $V_{tb} \approx 1$ the top quarks decay almost exclusively to a W boson and a b quark. Also due to its huge mass the top quark is the only quark that can produce an on-shell W boson. The resulting W boson itself decays further into either a quark–antiquark pair or it decays into a lepton and a neutrino. Hence depending on the type of decay products of the W boson, the final states of the top quark are usually called the hadronic channel or leptonic channel. In Figure 1.5 the top quark decay according to the SM is illustrated.

1.3.1 FCNC in the SM

Physics beyond the Standard Model (BSM) refers to the theoretical developments needed to address one or several of the shortcomings of the SM. Over the years, many BSM extensions have been proposed. The main purpose of these BSM theories is to

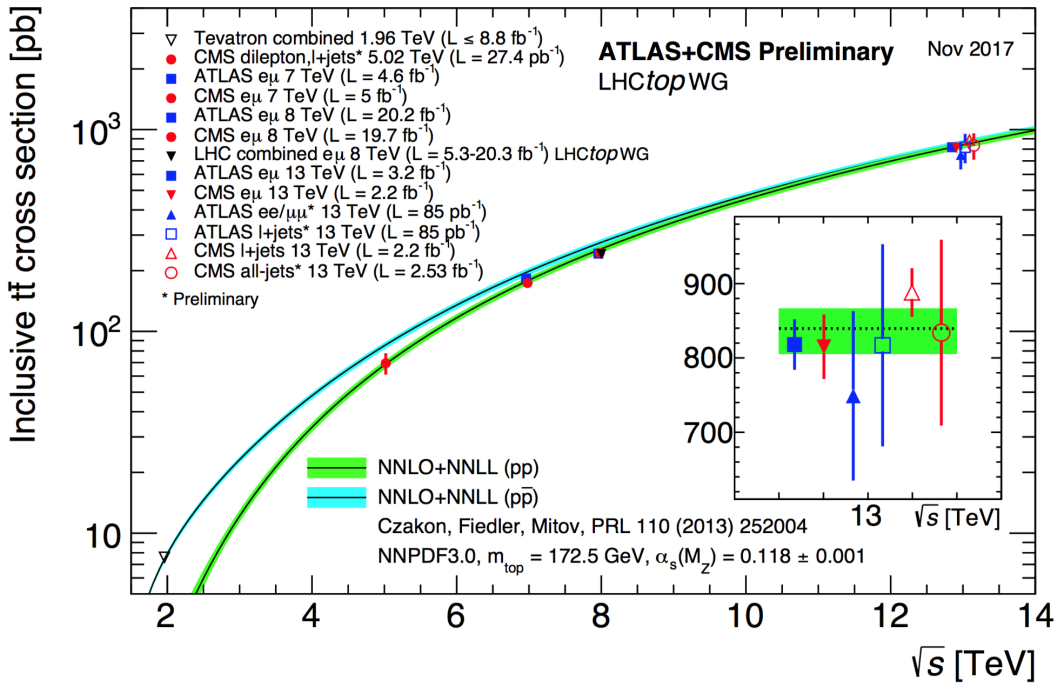
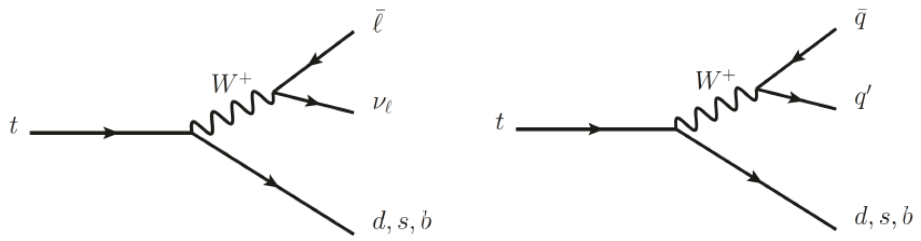


Figure 1.4: An overview of LHC and Tevatron measurements of the top-quark pair production cross-section as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NNLL resummation at $m_{top} = 172.5\,\text{GeV}$. The figure is taken from [28].



(a) (b)

Figure 1.5: The figure illustrates the leading order Feynman diagrams of the top quark decaying (according to the SM) leptonically at (a) and hadronically at (b).

establish a more general understanding of particle physics especially in the conditions where the SM fails, such as at very high energies. BSM theories can be categorised in different ways depending on the drawbacks of the SM that they try to address. As mentioned before, Flavour Changing Neutral Current (FCNC) processes can not occur at tree level in the SM, but they can occur in higher order loop diagrams within the SM framework. Several BSM scenarios predict the presence of FCNC contributions already at tree level.

In this section an overview is presented of the theoretical concepts behind the FCNC interactions in the SM. The absence of FCNC interactions at tree level in the SM is

demonstrated. Also the theoretical predictions of the branching ratios of BSM models will be described.

In the SM, only the charged $W^{\pm}$ bosons can change the flavour of the quarks through (FCCC) at tree level. The neutral ($Z^0, \gamma,\ g$ & H) bosons, called neutral currents, cannot change the flavour of fermions at tree level. This because there is no allowed vertex in the SM that couples a neutral current with two fermions of different flavours.

An example of an FCNC process allowed in the SM is shown in Figure 1.6 where the top quark changes its flavour into a u or c quark while emitting a Z boson is through a loop diagram where a W boson is exchanged. In order to calculate the branching ratio of the top quark through FCNC one has to take into account all other possible loop level diagrams involving d and s quarks.

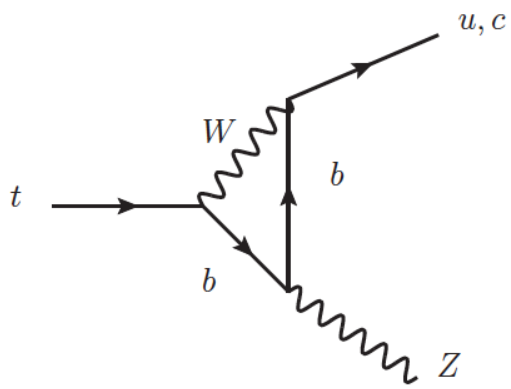


Figure 1.6: The diagram illustrates the one-loop FCNC decay of the top quark.

Sheldon Lee Glashow, John Iliopoulos and Luciano Maiani proved that when adding the diagrams with all internal quarks (b, s, d), the total amplitude is significantly reduced due to destructive interference. This is due to the fact that the internal quarks have much smaller masses compared to the top quark mass. This suppression mechanism of the FCNC processes in the SM is called the Glashow–Iliopoulos–Maiani (GIM) mechanism [2]. The predicted branching ratio of the FCNC processes involving top quarks in the SM is of the order of 10^{-13}. The expected Branching Ratios (BR) for a top quark decaying via a FCNC interaction in the SM are given in Table 1.1. This is beyond the reach of today's particle accelerators.

1.3.2 FCNC beyond the SM

FCNC in top quark interactions is highly suppressed in the SM. However many extensions of the SM predict significantly higher cross-sections and branching ratios for FCNC processes involving top quarks. The observation of such processes would be clear evidence of new physics. Hence, searches for FCNC processes at high-energy colliders are one of the best strategies to test BSM phenomena. Several BSM models predict strong enhancements of top-FCNC branching ratios, of which some are summarized in Table 1.1.

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

Table 1.1: The predicted branching ratios for FCNC decays involving the top quark in the SM and in some BSM theories, namely Minimal Supersymmetric SM (**MSSM**), Two-Higgs Doublet Model (**2HDM**), and Quark Singlet models (**QS**) [29].

The impact of such new physics can be written in a model independent way by means of an effective field theory (EFT) that is valid up to an energy scale Λ. Therefore the general SM Lagrangian becomes

$$\mathcal{L}_{SM+EFT} = \mathcal{L}^4_{\text{SM}} + \sum_i \frac{\bar{c}_i^{(5)}}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i \frac{\bar{c}_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \mathcal{O}\Big(\frac{1}{\Lambda^3}\Big) \ , \tag{1.23}$$

where $\mathcal{O}_i^{(n)}$ are operators of dimension n. The Wilson coefficients $\bar{c}_i^{(n)}$ are dimensionless coupling constants corresponding to each operators. Their values depend on the considered theory and on the way that new physics couples to the SM particles.

In the Warsaw basis (described in details in [30]), a set of independent operators of dimension 5 and 6 are built out of the SM fields and are consistent with the SM gauge symmetries. For simplicity, we assume that new physics effects are described by dimension-6 operators only, and neglecting the effects described by dimension-5 operators. In the fully gauge symmetric case, the EFT Lagrangian is then given by

$$\mathcal{L}_{SM+EFT} = \mathcal{L}^4_{\text{SM}} + \sum_i \frac{\bar{c}_i}{\Lambda^2}\mathcal{O}_i + \mathcal{O}\Big(\frac{1}{\Lambda^3}\Big) \ , \tag{1.24}$$

Since Λ is considered very large, the contribution by powers of Λ greater than two are suppressed. Moreover all four-fermion operators are omitted for the research that is presented in this thesis. Under these assumptions, the Warsaw basis is adopted for the independent effective operators. Then all the effects of new physics relevant for the flavor-changing neutral interactions of the top quark can be parameterized by the effective Lagrangian:

$$\begin{aligned}
\mathcal{L}^t_{\text{eff}} = {} & \frac{\bar{c}_{uG}}{\Lambda^2}\Phi^\dagger\cdot\big[\bar{Q}_L\sigma^{\mu\nu}T_a u_R\big]G^a_{\mu\nu} + \frac{\bar{c}_{uB}}{\Lambda^2}\Phi^\dagger\cdot\big[\bar{Q}_L\sigma^{\mu\nu}u_R\big]B_{\mu\nu} + \frac{2\bar{c}_{uW}}{\Lambda^2}\Phi^\dagger\mathcal{T}_i\cdot\big[\bar{Q}_L\sigma^{\mu\nu}u_R\big]W^i_{\mu\nu} \\
& + i\frac{\bar{c}_{hu}}{\Lambda^2}\big[\Phi^\dagger\overleftrightarrow{D}_\mu\Phi\big]\big[\bar{u}_R\gamma^\mu u_R\big] + i\frac{\bar{c}^{(1)}_{hq}}{\Lambda^2}\big[\Phi^\dagger\overleftrightarrow{D}_\mu\Phi\big]\big[\bar{Q}_L\gamma^\mu Q_L\big] \\
& + i\frac{4\bar{c}^{(3)}_{hq}}{\Lambda^2}\big[\Phi^\dagger\mathcal{T}_i\overleftrightarrow{D}_\mu\Phi\big]\big[\bar{Q}_L\gamma^\mu\mathcal{T}^i Q_L\big] + \frac{\bar{c}_{uh}}{\Lambda^2}\Phi^\dagger\Phi\ \Phi^\dagger\cdot\big[\bar{Q}_L u_R\big] + \text{h.c.} \ ,
\end{aligned} \tag{1.25}$$

where all flavour indices are implied, $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu,\gamma^\nu]$, Q_L are the left-handed $SU(2)_L$ doublets of the quark fields, u_R are the right-handed quark fields and Φ is the $SU(2)_L$-doublet of Higgs fields, and

$$\begin{aligned}
B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \ , \\
W^k_{\mu\nu} &= \partial_\mu W^k_\nu - \partial_\nu W^k_\mu + g\epsilon_{ij}{}^k\ W^i_\mu W^j_\nu \ , \\
G^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f_{bc}{}^a\ G^b_\mu G^c_\nu \ ,
\end{aligned} \tag{1.26}$$

where $f_{bc}{}^a$ is the structure constant of the $SU(3)_C$ group and $\epsilon_{ij}{}^k$ is the structure constant of $SU(2)_L$. The gauge covariant derivatives are defined as

$$D_\mu\Phi = \partial_\mu\Phi - \frac{1}{2}ig'B_\mu\Phi - ig\mathcal{T}_k W^k_\mu\Phi \ , \tag{1.27}$$

with the conventions of D and T given in Section 1.2. 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 in Equation 1.25 can be written as

$$\begin{aligned}\mathcal{L}^{t}_{\text{eff}} = \frac{\sqrt{2}}{2}\sum_{q=u,c}\Bigg[&\frac{e\kappa_{\gamma qt}}{\Lambda}A_{\mu\nu}\bar{t}\sigma^{\mu\nu}(f^{L}_{\gamma q}P_L + f^{R}_{\gamma q}P_R)q + \frac{g\kappa_{Zqt}}{2\cos\theta_W\Lambda}Z_{\mu\nu}\bar{t}\sigma^{\mu\nu}(f^{L}_{Zq}P_L + f^{R}_{Zq}P_R)q \\ &+ \frac{\sqrt{2}g\zeta_{Zqt}}{4\cos\theta_W}\,\bar{t}\gamma^{\mu}(\tilde{f}^{L}_{q}P_L + \tilde{f}^{R}_{q}P_R)qZ_{\mu} + \frac{g_s\kappa_{gqt}}{\Lambda}\,\bar{t}\sigma^{\mu\nu}T_a(f^{L}_{gq}P_L + f^{R}_{gq}P_R)qG^{a}_{\mu\nu} \\ &+ \kappa_{Hqt}\,\bar{t}(\hat{f}^{L}_{Hq}P_L + \hat{f}^{R}_{Hq}P_R)qH + \text{h.c.}\Bigg]\end{aligned} \tag{1.28}$$

This equation gives the FCNC interactions of the top quark that do not exist in the SM. Where P_L and P_R are the left-handed and right-handed chirality projectors, e stands for the electromagnetic coupling constant. The value of the FCNC couplings at scale Λ are represented by κ_{Xqt} and ζ_{Zqt}. These are assumed to be real, positive and dimensionless. The complex chiral parameters are normalized according to $|f^L|^2 + |f^R|^2 = 1$. The unitary gauge is adopted and the scalar field is expanded around its vacuum expectation value vev with H being the SM Higgs boson.

The EFT model described in this section is one will be used for the search for FCNC interactions in top quark presented in this thesis.

1.4 Overview of experimental searches on top-FCNC

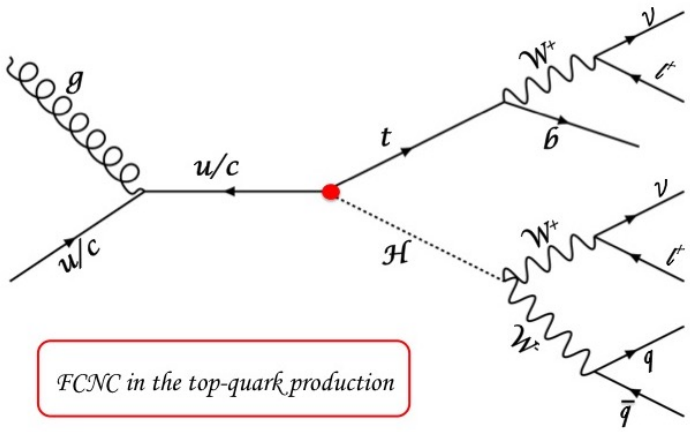
Equation 1.28 gives the description for the presence of top-FCNC in terms of coupling strengths κ_{Xqt} for different neutral currents. However, experimental physicists put limits on branching ratios which allows an easier interpretation across different EFT models. The branching ratio is related to the coupling strength as follows

$$\mathcal{BR}(t \to\ qX) = \frac{\delta^2_{Xqt}\quad \Gamma_{t\to\ qX}}{\Gamma_t} \tag{1.29}$$

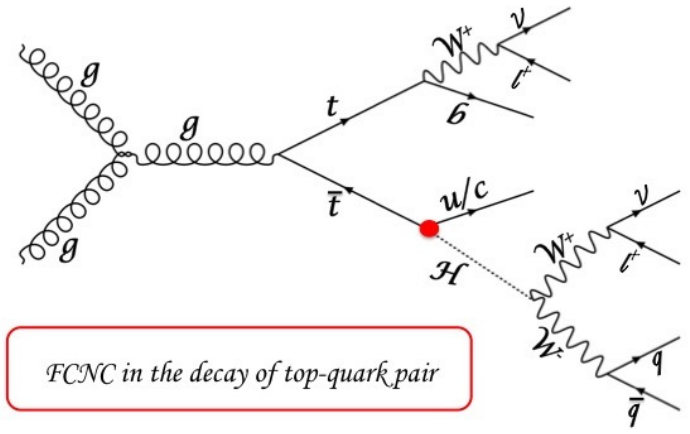
where Γ_t is the full SM decay width of the top quark. For coupling strength $\delta^2_{Xqt} = 1$ and top quark mass $= 172.5\,\text{GeV}$, $\Gamma^{SM}_t = 1.32158\,\text{GeV}$ as calculated in [31].

In general searches focus on one of the FCNC vertices described in EFT model by putting all other FCNC coupling strengths to zero. Many experimental searches for top-FCNC have been performed by the ATLAS and CMS experiments. Searches usually adopt a search strategy depending on the experimental set-up and the FCNC interaction of interest. A search is usually sensitive to one of the following two search modes: The search for FCNC interactions in the production of a single top quark and the search for FCNC interactions in $t\bar{t}$ decay. Figure 1.7 illustrates the two modes for the top-FCNC (Hqt)-coupling, where the Higgs boson decays into two $W^{\pm}$ bosons.

The presence of top-FCNC interactions has not yet been observed. Then experiments have so far only been able to put upper bounds on the branching ratios. An



(a)



(b)

Figure 1.7: The figure illustrates the leading order Feynman diagrams for the processes with a FCNC Hqt interaction, where the FCNC interaction vertex is indicated with a red point. The FCNC interactions in the production of a single top quark (a), and (b) FCNC interactions in $t\bar{t}$ decay.

overview of the current best limits for different couplings is provided in Table 1.2. A comparison between the current best limits set by ATLAS and CMS with respect to several BSM model benchmark predictions is shown in Figure 1.8. One can see that FCNC searches involving a H or Z boson are close to excluding or confirming several BSM theories. The searches for top-FCNC performed by CMS at 8 TeV are summarised in Figure 1.9.

The physics analysis of our search for top-FCNC presented in this thesis considers only the Hqt coupling where the Higgs boson decays into two $W^{\pm}$ bosons which result into two same-sign leptons in the final state.

Process	**Search mode**	**Observed** $\mathcal{BR}$ limit (95%) CL	**Expected** $\mathcal{BR}$ limit (95%) CL	**Experiment**
$t \to uH$	$t\bar{t}$ decay	2.4×10^{-3}	1.7×10^{-3}	**ATLAS** [32]
$t \to uZ$	$t\bar{t}$ decay	1.7×10^{-4}	2.4×10^{-4}	**ATLAS** [33]
$t \to u\gamma$	Single top production	1.3×10^{-4}	1.9×10^{-4}	**CMS** [34]
$t \to ug$	Single top production	2.0×10^{-5}	2.8×10^{-5}	**CMS** [35]
$t \to cH$	$t\bar{t}$ decay	2.2×10^{-3}	1.6×10^{-3}	**ATLAS** [32]
$t \to cZ$	$t\bar{t}$ decay	2.3×10^{-4}	3.2×10^{-4}	**ATLAS** [33]
$t \to c\gamma$	Single top production	2.0×10^{-3}	1.7×10^{-3}	**CMS** [34]
$t \to cg$	Single top production	2.4×10^{-4}	1.8×10^{-4}	**CMS** [36]

Table 1.2: Overview of the most stringent observed and expected upper limits on top-FCNC branching ratios at 95% confidence level.

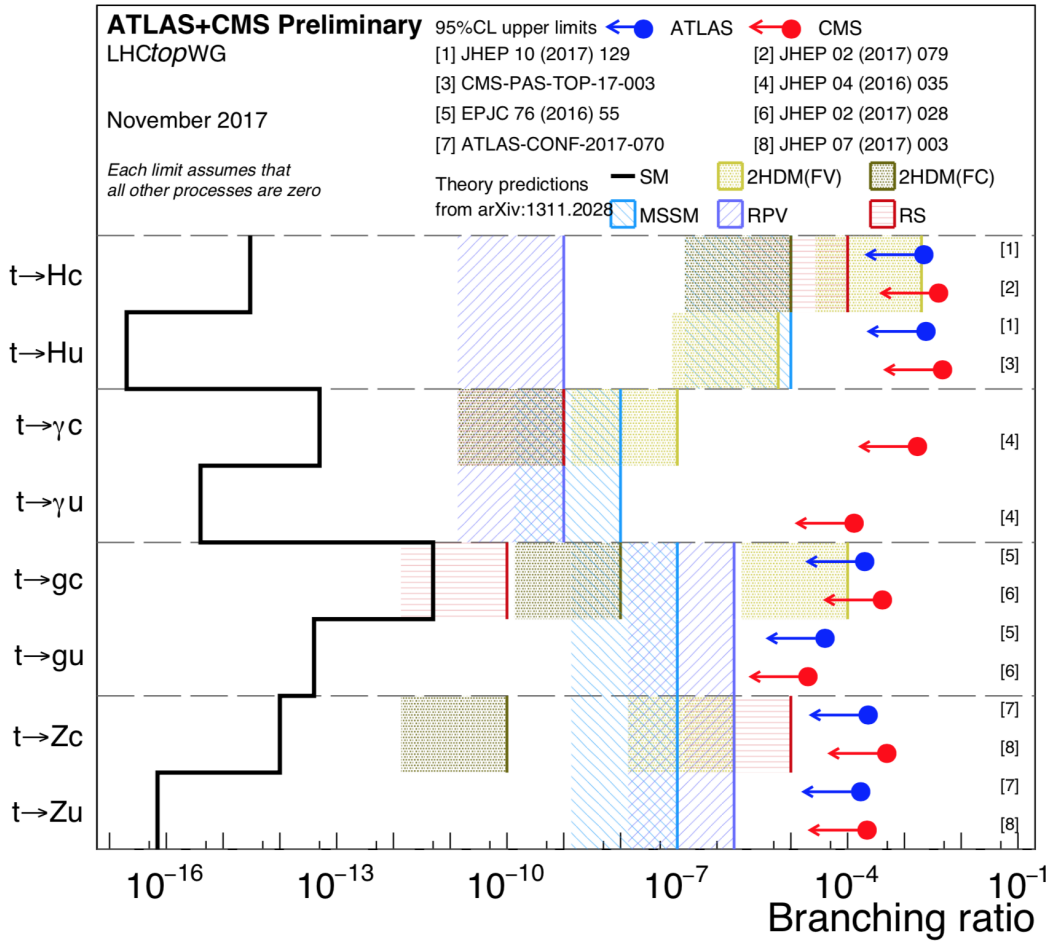


Figure 1.8: Current best limits at 95% confidence level on the branching ratios set by CMS and ATLAS for top-FCNC interactions. Figure is taken from [28].

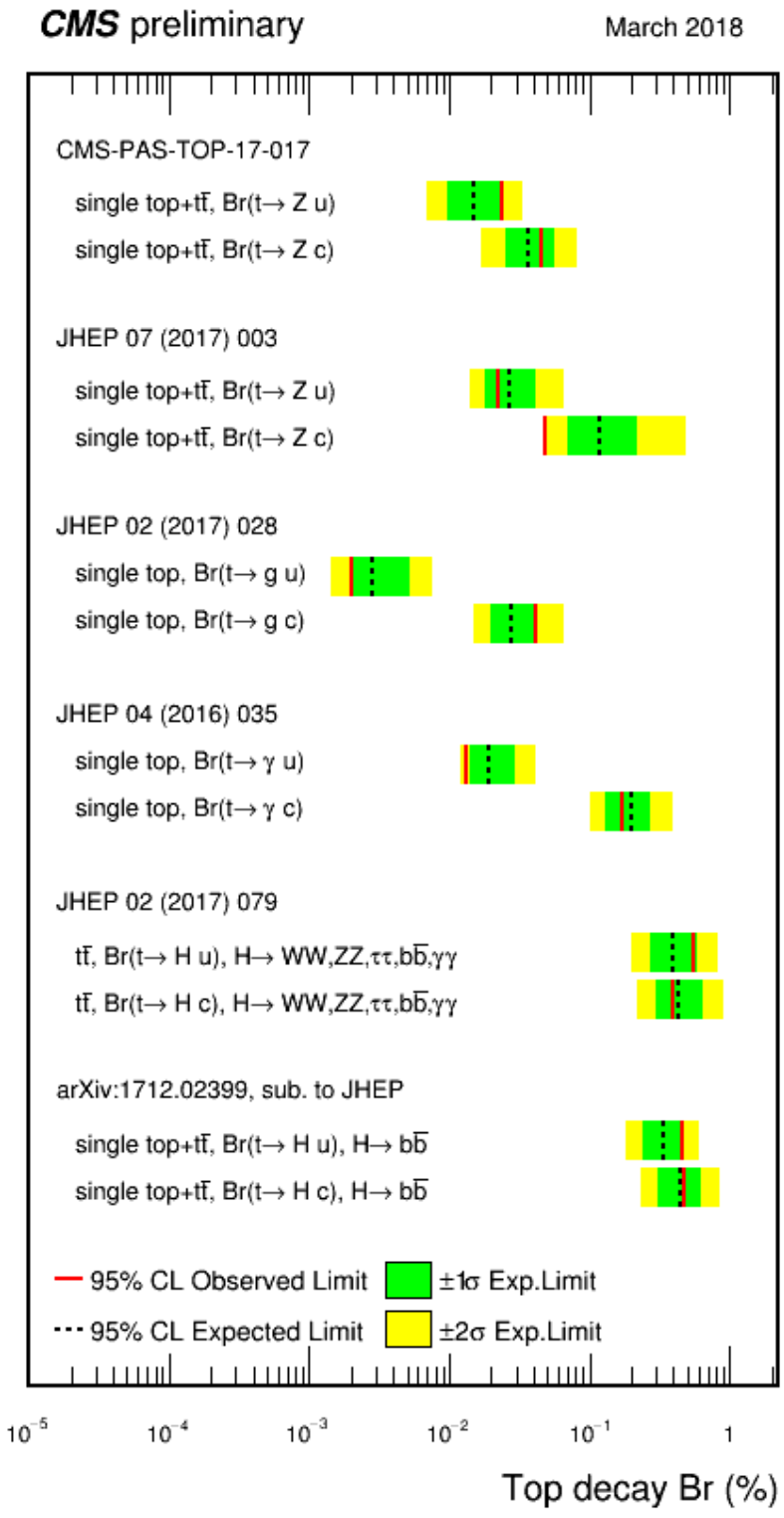


Figure 1.9: Summary of the FCNC branching ratios from CMS searches at a centre-of-mass energy of 8 TeV. Figure is taken from [37].

"As scientists, we step on the shoulders of science, building on the work that has come before us - aiming to inspire a new generation of young scientists to continue once we are gone."

Stephen Hawking

Chapter 2

The CMS Experiment at the LHC

The Large Hadron Collider (LHC) at CERN (the European Organization for Nuclear Research), in Geneva, Switzerland, is the world's largest and most powerful tool for particle physics research. It was installed in the existing 26.7 km tunnel, 100 meters below the French-Swiss border constructed between 1984 and 1989 for the CERN LEP (the Large Electron Positron) collider [38]. The main goal of the LHC was the search for the Brout-Englert-Higgs boson. The LHC is a two-ring-superconducting-hadron accelerator and collider which is designed to collide proton beams with a centre-of-mass energy of up to 14 TeV with a luminosity of $10^{34}\,\mathrm{cm}^{-2}s^{-1}$, and to collide heavy (Pb) ions with an energy of 2.8 TeV per nucleon and a peak luminosity of $10^{27}\,\mathrm{cm}^{-2}s^{-1}$ [39].

2.1 The CERN accelerator complex

Acceleration of the protons is not done inside the LHC directly; but protons follow an acceleration journey through a complex chain of accelerators as shown in Figure 2.1. During this journey each accelerator injects the particle beam into the next accelerator in the chain, which brings the beam to an even higher energy. In this way, the particles are increasingly accelerated at each stage of the accelerator complex before they reach the main accelerator ring, the LHC.

First protons are obtained via the ionization process of hydrogen atoms using a pulsed ion source called "Duoplasmatron". The protons are produced inside it and accelerated to 1.4% of speed of light. Then they start their acceleration journey as follows:

- Protons are initially accelerated up to 50 MeV in the first accelerator of the complex, the linear accelerator LINAC 2, with this energy protons are injected to the second accelerator, the Proton Synchrotron Booster (PSB).

- At the PSB (or Booster) proton packs are accelerated up to 1.4 GeV and each pack is split up in twelve bunches with 25 or 50 ns spacing. Then the proton bunches are injected into the Proton Syncrotron (PS) where they are accelerated up to 25 GeV. At this energy they are injected into the Super Proton Synchrotron (SPS) where they gain more energy up to 450 GeV. Proton bunches are then injected

into the two transfer lines to the LHC, resulting in two proton beams traveling in opposite directions inside the LHC beam pipes.

- The proton beams at the LHC travel in separate vacuum beam pipes and are maintained in a fixed orbit using 1232 superconducting dipole magnets. Eight superconducting RF cavities operating at 400 MHz generate the electric fields used for the particle acceleration. The superconducting dipole magnets are cooled to 1.9 K using liquid helium. They provide magnetic field strengths of up to 8.33 T. The beams are also squeezed (focused) using 392 main superconducting quadrupole magnets.

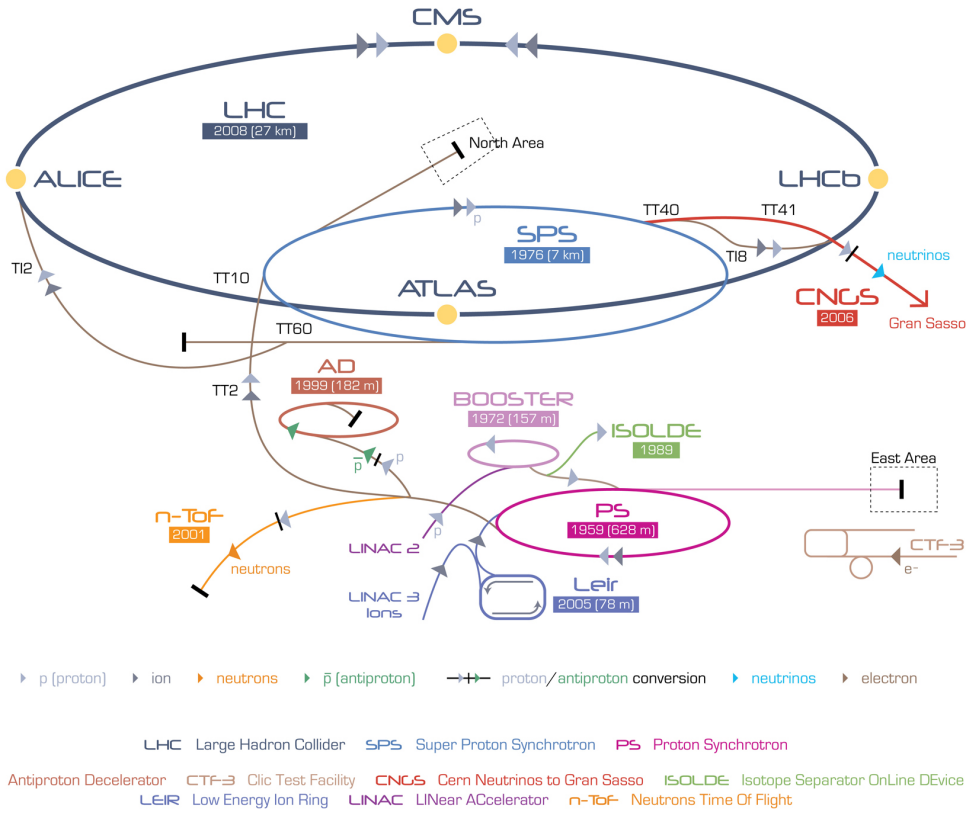


Figure 2.1: Schematic overview of the accelerators complex of CERN. The figure is taken from [40]

Inside the LHC the protons are accelerated in bunches. After accelerating up to the final energy (6.5 TeV in Run 2), proton beams are brought to collision at four interaction points (IPs) along the LHC ring. Each of them is surrounded by an experiment (detector):

1. CMS (Compact Muon solenoid): It is a multi-purpose detector with wide physics objectives. In addition to its capability to find the Higgs boson with masses up to 1 TeV other objectives are defined for example, the search of new physics, dark matter candidates or supersymmetric particles, more details about this detector will be described in Section 2.2.

2. ATLAS (A Toroidal LHC ApparatuS): It is also a multi-purpose detector and has the same objectives as CMS but exploits different technical solutions, including a large toroidal magnet [41].

3. LHCb (Large Hadron Collider beauty): It aims to study matter-antimatter asymmetry via CP violation, through measurements involving b quarks [42].

4. ALICE (A Large Ion Collider Experiment): It is a general-purpose, heavy-ion detector at the LHC which focuses on QCD studies, the strong-interaction sector of the Standard Model. It is designed to study the physics of strongly interacting matter and the quark-gluon plasma at extreme values of energy density and temperature in nucleus-nucleus collisions [43].

Although ATLAS and CMS analyze independent datasets; their respective results can cross-check each other, and can also be combined to increase the precision of the measurements.

In addition to these four main experiments, several smaller ones are located close to the interaction points. These experiments are constructed with the purpose of studying particles emitted during proton collisions in the very forward region. For example MoEDAL [44] searching for magnetic monopoles, LHCf [45] that studies hadron interaction models used in cosmic ray analyses and TOTEM for the total cross-section, elastic scattering and diffraction dissociation at the LHC [46].

2.1.1 The LHC parameters

The ***luminosity*** is an important parameter to characterize the performance of the LHC, it is a measure of the collision rate, i.e. the number of collisions that can be produced in a detector per cm^2 and per second. ***Instantaneous luminosity*** ($\mathcal{L}$) can be calculated from the beam parameters using

$$\mathcal{L} = \frac{N^2}{t\ S_{eff}} \tag{2.1}$$

where N is number of protons in one bunch, the squared is due to the fact that each particle in a bunch might collide with another particle from the bunch approaching head on, t is the time between bunches and S_{eff} is the effective section of the collision that depends on the head-on cross section of the bunch σ ("effective" because the beam profile doesn't have a sharp edge). The formula for S_{eff} is given as $S_{eff} = 4\ \pi\ \sigma^2$, If we use the bunch crossing frequency f and considering different number of protons per crossing bunches, and the x and y components for σ separately, this becomes

$$\mathcal{L} = \frac{f\ N_1\ N_2}{4\ \pi\sigma_x\sigma_y}. \tag{2.2}$$

One can determine the integrated luminosity ($\mathcal{L}_{int}$) by the integration of the instantaneous luminosity over the time, and then for the particular process the produced

number of events, N_{events} can be calculated by knowing the cross-section σ of this process

$$N_{events} = \mathcal{L}_{int} \times \sigma. \tag{2.3}$$

The total integrated luminosity ($\mathcal{L}_{int}$) for CMS Run 2 during 2016 is shown in Figure (2.2). The delivered luminosity accounts for the luminosity delivered from the start of stable beams [1] until the LHC requests CMS to turn off the sensitive detectors to allow a beam dump or beam studies.

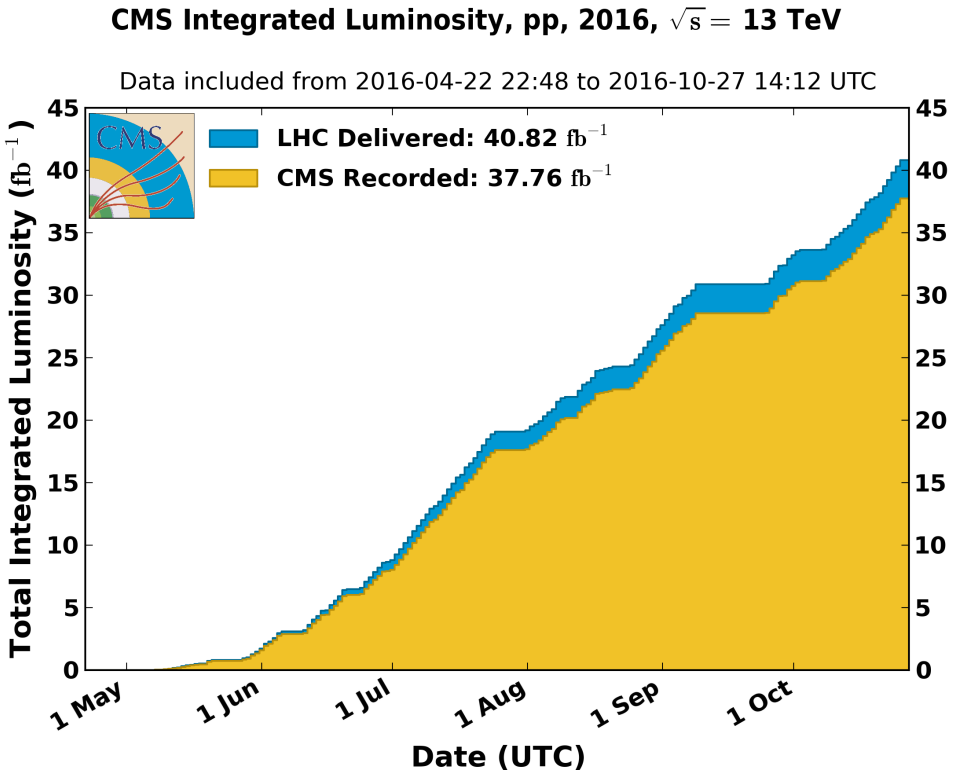


Figure 2.2: Cumulative integrted luminosity measured online versus day delivered to CMS (blue), and recorded by CMS (orange) during stable beams and for pp collisions at 13 TeV centre-of-mass energy in 2016 [47].

2.2 The CMS Experiment

The CMS detector is a multi-purpose detecting system, designed to explore the physics at the TeV-scale. This is the energy region where physicists believe they will find answers to the central questions in particle physics for which solutions are predicted by theories Beyond Standard Model.

CMS gets its name from its specifications: limited size (compact) considering its complexity, its sophisticated muon system, and its solenoidal superconducting magnet.

[1] Stable beam status is given by experts which means the proton beams are stable to start collecting the data from pp-collisions

2.2.1 Overview of the CMS detector

The design of CMS provides a good muon identification and momentum resolution, a good charged-particle reconstruction efficiency and momentum resolution, a good energy-resolution of electromagnetic calorimeter , a good missing transverse energy and di-jet mass resolution that allow to reach the objectives of the research programme.

The CMS detector is composed of different subdetector layers, arranged in a central cylinder, the *barrel*, and closed by two *endcaps*. It is 28.7 m long, has a diameter of 15.0 m and weighs 14000 t.

The origin of the right-handed coordinate system of CMS is at the nominal collision point, while the x-axis points radially towards the center of the LHC, the y-axis points vertically and the z-axis is directed along the beam direction (towards the Jura mountains from LHC Point 5) which is orthogonal to the other axes.

The azimuthal angle ϕ is the angle measured from the x-axis in the (x, y)-plane, and the radial coordinate in this plane is denoted r, the polar angle θ measured from the z-axis, in a (r, z)-plane orthogonal to the (x, y)-plane,The momentum component transverse to the beam axis, denoted p_{T}, is computed from the x and y components. The transverse energy is defined as $E_{\mathrm{T}} = E \sin\theta$. The polar angle θ is used to define another variable which is called pseudorapidity η and defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.4}$$

Pseudorapidity η is widely-used in accelerator physics because it has the advantage of being invariant under Lorentz boosts along the beam axis (z axis) and is used as a measure of the angle of a particle with respect to the z axis. For particles with $E >> m$, η is a good approximation for the rapidity y which is defined as

$$y = \frac{1}{2}\ln\left(\frac{E + P_z}{E - P_z}\right). \tag{2.5}$$

The difference of the rapidities of two particles is invariant under a Lorentz boost in the z-direction.

The key elements of the CMS detector, described in the next sections and illustrated in Figure 2.3 are, from the innermost to the outermost part: an inner tracking system, an electromagnetic calorimeter (ECAL), a hadronic calorimeter and a muon system. The tracker and the largest part of the calorimeters are enclosed within a superconducting solenoid magnet, which provides a 3.8 T magnetic field parallel to the beam axis to bend the tracks of charged particles. This strong magnetic field is contained by using an iron return yoke in which the muon chambers are embedded.

2.2.2 The Silicon Trackers

The tracker is the innermost subdetector layer close to the interaction point at CMS. Its name comes from the main function that is to reconstruct tracks from charged particles produced from pp-collisions, i.e. recording the curved trajectories of charged particles. The trajectories can be used to estimate the momentum of the charged particles with

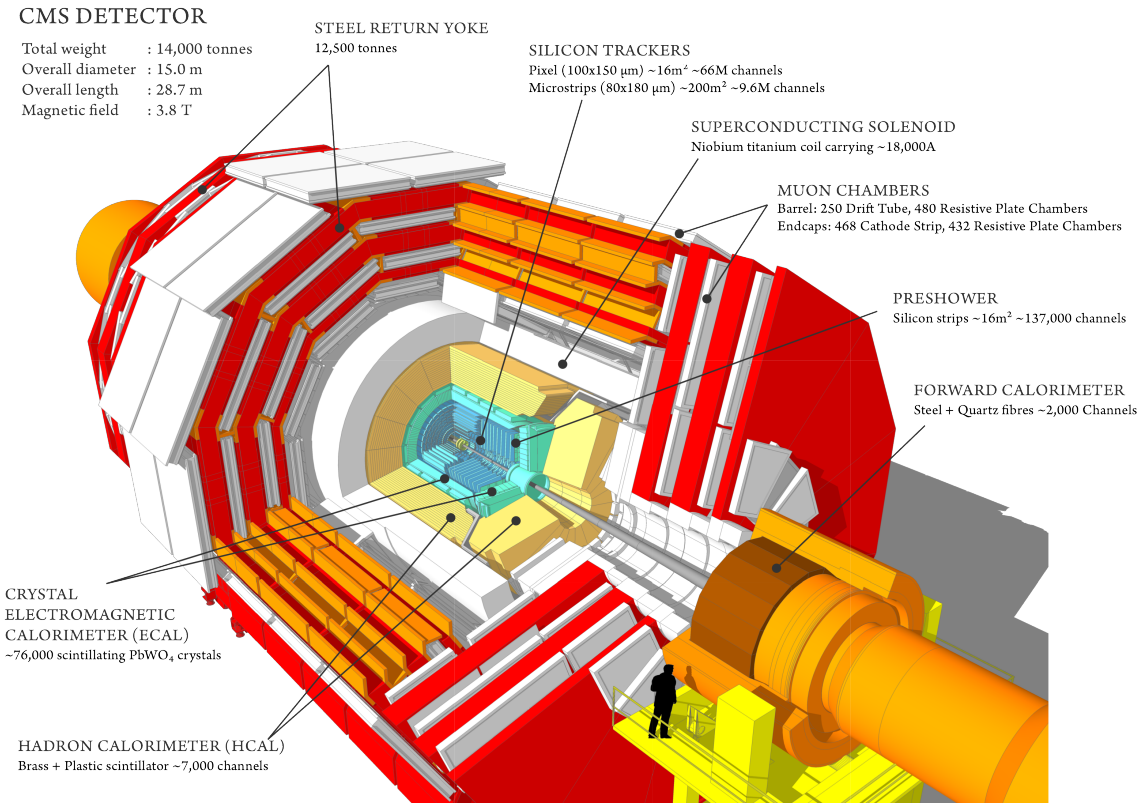


Figure 2.3: Schematic of CMS subdetector layers. The figure is taken from [48]

great precision and to reconstruct secondary vertices from long-lived particle decays [49].

The CMS tracker occupies a cylindrical volume 5.8 m in length and 2.5 m in diameter, with its axis closely aligned to the LHC beam line. Because of the high particle density environment from the proton-proton collision, it is of primary importance that the tracker has a high granularity. While the quantity of material has to be kept low to limit photon conversion, multiple scattering, bremsstrahlung and nuclear interactions, the tracker had to have a high power density of electronics and to be radiation resistant to achieve this. The tracker comprises a large silicon strip detectors with a small high-granularity silicon pixel detector inside it close to the proton-proton collision point, as schematically shown in Figure 2.4. In the central pseudorapidity region, the pixel tracker consists of three co-axial barrel layers and the strip tracker consists of ten co-axial barrel layers. Both subdetectors are completed by endcaps on each side of the barrel, each consisting of two disks in the pixel tracker, and twelve disks in the strip tracker (three small plus nine large disks). The endcaps extend the acceptance of the tracker up to a pseudorapidity of $|\eta| < 2.5$.

The pixel detector consists of cylindrical barrel layers at radii of 4.4, 7.3 and 10.2 cm, and two pairs of endcap disks at $z = \pm 34.5$ and ± 46.5 cm. It provides three-dimensional position measurements of the hits arising from the interaction of charged particles with its sensors. The hit position resolution is approximately 10 μm in the transverse coordinate, and $20 - 40$ μm in the longitudinal coordinate, while the third

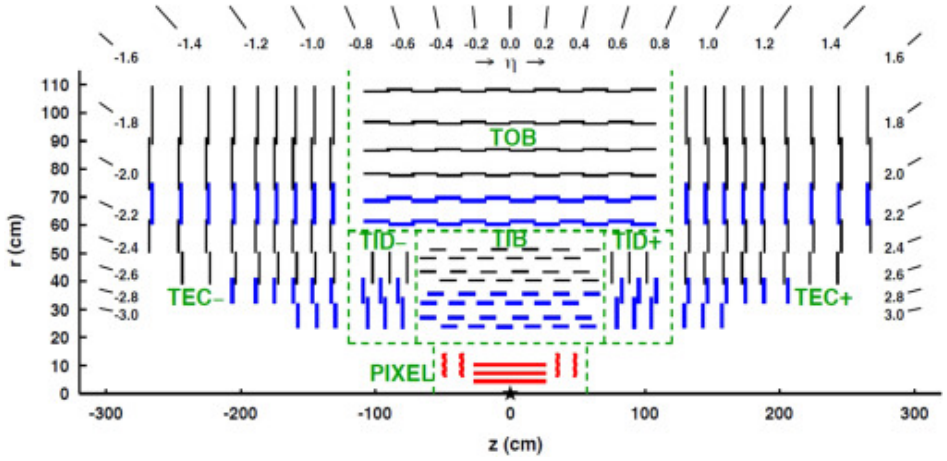


Figure 2.4: Schematic cross section through the CMS tracker in the $r - z$ plane. In this view, only the top half is shown, the tracker is symmetric around the horizontal line $r = 0$. The "star" symbole represents the center of the tracker where is the pp-interaction point. Pixel modules, shown by the red lines, provide 3-D hits. Within a given layer, each module is shifted slightly in r or z with respect to its neighbouring modules, which allows them to overlap, thereby avoiding gaps in the acceptance. Strip tracker modules that provide 2-D hits are shown by thin, black lines, while those provid in 3-D hits are shown by thick, blue lines. The latter actually each consist of two back-to-back strip modules, in which one module is rotated through a 'stereo' angle [50].

coordinate is given by the sensor plane position. In total, its 1440 modules cover an area of about $1\,\mathrm{m}^2$ and have 66 million pixels [49]. The pixel detector not only forms high quality seeds for the offline track reconstruction algorithm, but is also used to do fast tracking online in the high-level trigger (HLT) for primary vertex reconstruction, electron/photon identification, muon reconstruction, tau identification and b-tagging [51, 52].

The strip tracker has 15148 silicon modules, which in total cover an active area of about $198\,\mathrm{m}^2$ and they have 9.3 million strips. It is composed of four subsystems. On each side of the **Tracker Inner Barrel (TIB)**, **Tracker Inner Disks (TID)** are placed. The TIB is composed of four barrel layers that covers up to $r < 55\,\mathrm{cm}$ and $|z| < 65\,\mathrm{cm}$, while the TID is composed of three disks of strip modules oriented perpendicular to the beam line and covering up to $|z| < 124\,\mathrm{cm}$. These provide position measurements in $r\phi$ with a resolution of approximately $13 - 38\mu\,\mathrm{m}$. **The Tracker Outer Barrel (TOB)** consists of six barrel layers covering $r > 55\,\mathrm{cm}$ and $|z| < 118\,\mathrm{cm}$ providing position measurements in $r\phi$ with a resolution of approximately $18 - 47\mu\,\mathrm{m}$. **The Tracker EndCaps (TEC)** cover the region $124 < |z| < 282\,\mathrm{cm}$. Each TEC is composed of nine disks, each containing up to seven concentric rings of silicon strip modules, yielding a range of resolutions similar to that of the TOB. The principal characteristics of the tracker are summarized in Table 2.1

Tracker subsystem	**Layers**	**Pitch**	**location**
Pixel tracker barrel	3 cylindrical	$100 \times 150 \mu\,\mathrm{m}^2$	$4.4 < r < 10.2\,\mathrm{cm}$
Pixel tracker endcap (TEC)	2 disks	$100 \times 150 \mu\,\mathrm{m}^2$	$34.5 < \lvert z\rvert < 46.5\,\mathrm{cm}$
Strip Tracker Inner Barrel (TIB)	4 cylindrical	$80 - 120 \mu\,\mathrm{m}$	$20 < r < 55\,\mathrm{cm}$
Strip Tracker Outer Barrel (TOB)	6 cylindrical	$122 - 183 \mu\,\mathrm{m}$	$55 < r < 116\,\mathrm{cm}$
Strip Tracker Inner Disks (TID)	3 disks	$140 - 141 \mu\,\mathrm{m}$	$58 < \lvert z\rvert < 124\,\mathrm{cm}$
Strip Tracker Outer Disks (TID)	9 disks	$97 - 184 \mu\,\mathrm{m}$	$124 < \lvert z\rvert < 282\,\mathrm{cm}$

Table 2.1: Characteristics of the Silicon Tracker geometry. Pitch for Pixels means the size of each cell while for strips means the distance between neighbouring strips. The table taken from [49].

The first data collisions, so-called Run 1 period, started on 2011 until February 2013 after which the CERN accelerator complex shut down for two years for maintenance and consolidation during so-called Long Shutdown-1 (LS1). The new data taking period known as Run 2 started in March 2015. According to the LHC programme, Run 2 will ended on when the Long Shutdown-2 (LS2) will begin by the end of 2018 and continue for 2 years.

During (LS1) damaged silicon pixel and strip modules in the tracker were repaired or replaced thereafter the barrel pixel detector reached about 99% operational efficiency. Also during LS1 the tracker cooling system was improved to reach temperatures down to -15 and -10 ^oC for strips and pixels respectively. This increases the damage mitigation against radiation [53, 54].

Track Reconstruction

Charged particles produce **"hits"** in the silicon sensors arranged in the cylindrical layers of the trackers. The trajectory of the charged particle or its "**track**" can be reconstructed from these hits.

The tracking software of CMS is commonly referred to as the Combinatorial Track Finder (CTF), which is an adaptation of the combinatorial Kalman filter which in turn is an extension of the Kalman filter to allow pattern recognition and track fitting to occur in the same framework [55–57]. The collection of reconstructed tracks is produced by multiple passes (iterations) of the CTF track reconstruction sequence, in a process called **"iterative tracking"**. Each iteration proceeds in four steps:

1. **Seed Generation**: The seed defines the initial estimate of the trajectory parameters and their uncertainties. It provides initial track candidates found using only a few (2 or 3) 3-D hits, where a **"3-D hit"** is defined to be any hit that provides a 3-D position measurement. To limit the number of hit combinations, seeds are required to satisfy certain weak restrictions, for example, on their minimum p_T and their consistency to originate from the pp-interaction region. Seeds are constructed in the inner part of the tracker and the track candidates are built outwards. This approach is followed because of three reasons. First, although the track density is much higher in the inner region of the tracker, the high granularity of the pixel detector ensures that the channel occupancy (fraction of channels that are hit) of the inner pixel layer is much lower than that of the outer strip layer. Second, the pixel layers produce 3-D spatial measurements, which provide more constraints and better estimates of trajectory parameters. Finally, generating seeds in the inner tracker leads to a higher efficiency for reconstructing tracks [58].

2. **Track Finding Module**: The track-finding module of the CTF algorithm is based on the Kalman filter method. The filter begins with a coarse estimate of the track parameters provided by the trajectory seed, and then builds track candidates by adding hits from successive detector layers, updating the parameters at each layer. The information needed at each layer includes the location and uncertainty of the detected hits, as well as the amount of material crossed, which

is used to estimate the effects of multiple Coulomb scattering and energy loss. All resulting track candidates found at each layer are then propagated to the next compatible layers, and the procedure is repeated until a termination condition is satisfied. However, to avoid a rapid increase in the number of candidates, only a limited number (default is 5) of the candidates with the most hits are kept.

3. **Track-Fitting Module**: For each trajectory, the track-finding stage yields a collection of hits and an estimate of the track parameters. The estimate can be biased by constraints, such as a beam spot constraint applied to the trajectory during the seeding stage. The trajectory is therefore refitted using a Kalman filter and smoother. The Kalman filter is initialized at the location of the innermost hit, with the trajectory estimate obtained by performing a Kalman filter fit to the innermost hits (typically four) on the track. The fit then proceeds in an iterative way through the full list of hits, from the inside outwards, updating the track trajectory estimate sequentially with each hit. To obtain the best precision, this filtering and smoothing procedure uses a Runge-Kutta propagator [2] to extrapolate the trajectory from one hit to the next. This not only takes into account the effect of material, but it also accommodates an inhomogeneous magnetic field and this is most important in the region $|\eta| > 1$, where the magnetic field inhomogeneity is greatest.

4. **Track Selection**: This module sets quality flags, and discards tracks that fail certain specified criteria. For events containing jets, the track-finding procedure yields a significant fraction of fake tracks. A fake track is defined as a reconstructed track not associated with a charged particle. This fake rate can be reduced substantially through quality requirements. Tracks are selected on the basis of the number of layers that have hits, whether their fit yielded a good χ^2/dof, and how compatible they are with originating from a primary interaction vertex. For performance optimization several requirements are determined as a function of track p_{T} and η.

2.2.3 The Electromagnetic Calorimeter (ECAL)

The electromagnetic calorimeter (ECAL) is installed inside the $3.8T$ CMS superconducting solenoid. It has been designed to ensure that its excellent performance extends over a very wide range of energies up to electron and photon energies of 1 TeV and beyond. It is a hermetic, homogeneous calorimeter that is composed of fine-grained lead tungstate ($PbWO_4$) scintillation crystals shown in Figure 2.5. The choice of an homogeneous medium was made to obtain a better energy resolution by minimizing sampling fluctuations and the choice of $PbWO_4$ is due to its characteristics. An extremely short radiation length of $X_0 = 0.85\,\mathrm{cm}$ and a small Molière radius $R_M = 2.19$ cm ensure the compactness of the detector and the radiation hardness necessary to cope with the harsh environment of the LHC collisions [60, 61].

[2] The Runge-Kutta propagator divides the distance to be extrapolated into many small steps. It extrapolates the track trajectory over each of these steps, using a mathematical technique for solving first-order differential equations, called the fourth-order Runge-Kutta method [59]

The crystals are arranged into three sections the *ECAL Barrel (EB)* closed by two *ECAL Endcaps (EE)* and *ECAL Preshowers (ES)* as schematically shown in Figure 2.6:

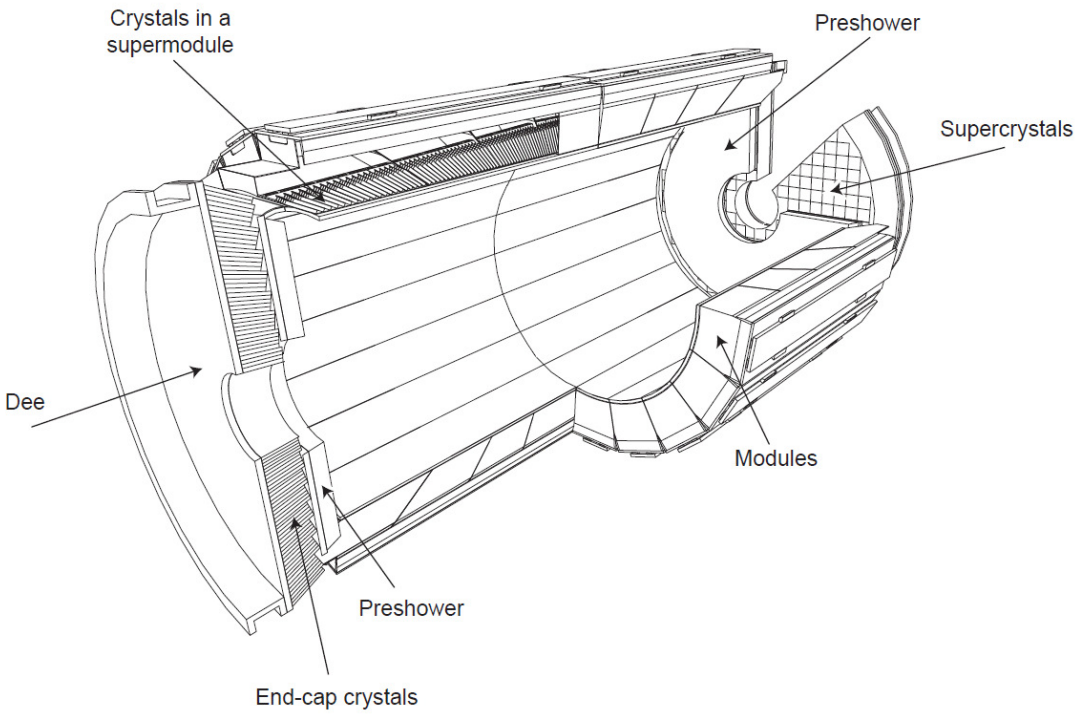


Figure 2.5: The ECAL system contains 75848 $PbWO_4$ crystals. Crystals are contained in a thin-walled alveolar structure (submodule) which are assembled into modules. Each module contains 400 or 500 crystals and every four modules are assembled in a supermodule which contains 1700 crystals. The ECAL is composed of 36 supermodules. Supermodules at EndCaps are composed of (5×5) crystals with 7324 crystals in total. Preshowers are installed in front of the EndCaps [62].

- **ECAL Barrel section (EB)**: 61200 crystals are included in this section each with a truncated pyramidal shape ($2.2 \times 2.2\,\mathrm{cm}^2$ on the frontal face, with a length of 23 cm, corresponding to approximately $26X_0$) and are organized in 36 supermodules (each containing 1700 crystals) covering the pseudorapidity range $|\eta| < 1.48$. Crystals are positioned slightly off-pointing relative to the interaction point (IP) to avoid cracks aligned with particle trajectories.
- **ECAL Endcaps section (EE)**: Each Endcap consists of two dees, with 3662 crystals each. Crystals also have a truncated pyramidal shape ($2.86 \times 2.86\,\mathrm{cm}^2$ on the frontal face, with a length of 22 cm, corresponding to approximately $25X_0$) and are arranged in a (x-y)-grid which extends the coverage of the ECAL up to $|\eta| < 3$.
- **ECAL Preshower (ES)**: This is a sampling calorimeter composed of lead absorbers equipped with silicon strip sensors (4288 sensors, 137216 strips, $1.90 \times$

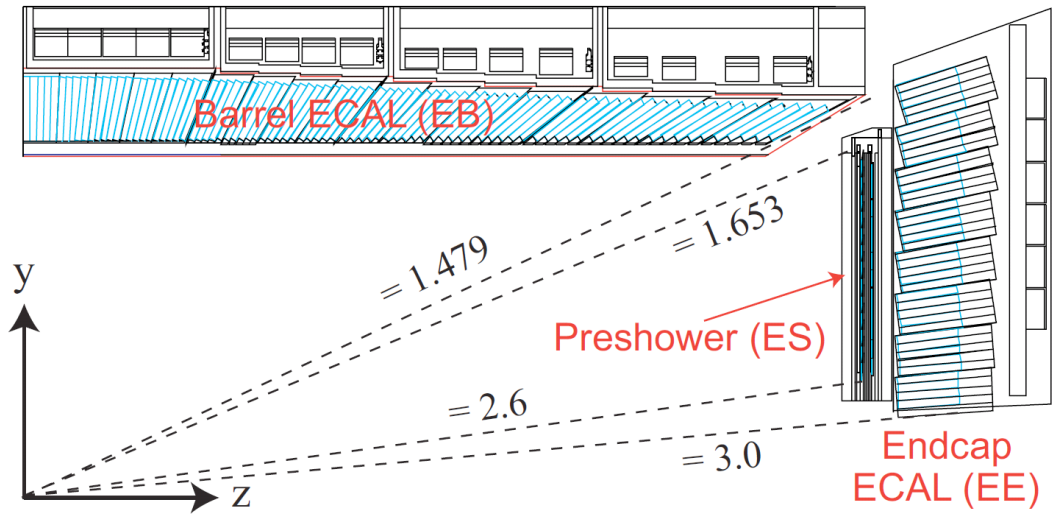


Figure 2.6: Sectional view of a quarter of the ECAL. The figure is taken from [63]

61 mm^2 with x-y view). It is placed in front of the EE, covering the region $1.65 < |\eta| < 2.6$. The total thickness of the ES is about $3X_0$. The ES helps in improving the $photon - \pi^0$ discrimination.

The scintillation light is detected by avalanche photodiodes (APDs) in the EB, and by vacuum phototriodes (VPTs) in EE, yielding an average of 4.5 photoelectrons per MeV deposited in the crystals.

The most important challenges of the ECAL operation at LHC Run 2 are the increased PU collision rate and the reduced LHC bunch spacing of 25 ns. They increase the probability of single calorimeter cells to be hit by a particle in successive bunch crossings and makes it more difficult to differentiate contributions from preceding and trailing bunches [64].
The performance of the ECAL has been measured with test beams [**?**], with no magnetic field, nor material in front of the calorimeter. The ECAL barrel energy resolution measured at the test beam is given by:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E(\text{GeV})}} \oplus \frac{b}{E(\text{GeV})} \oplus c \tag{2.6}$$

where $a = 2.8\%$ is the stochastic term for the statistical fluctuations on the number of secondary particles produced, $b = 12\%$ is noise term from the electronics and digitisation and $c = 0.3\%$ is the irreducible constant that includes contributions from the non-uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter and single-channel response uniformity [3]. In order to keep the excellent intrinsic energy resolution of the ECAL, the constant term c should be kept within 0.5%.

[3]Channel response changes with time due to the radiation damage of the crystals and detector instabilities.

2.2.4 Hadronic Calorimeter (HCAL)

The hadronic calorimeter (HCAL) has been designed to measure the energy of hadrons emerging from the proton-proton collisions. It is particularly important for the reconstruction of jets and neutrinos or exotic particles resulting in apparent missing transverse energy $\not{E}_T$.

The CMS HCAL is a hermetic sampling calorimeter that is composed of scintillators interleaved with absorbers. It consists of four major sections as shown in Figure 2.7: the HCAL Barrel (HB), the HCAL Endcap (HE), the HCAL Outer (HO), and the HCAL Forward (HF).

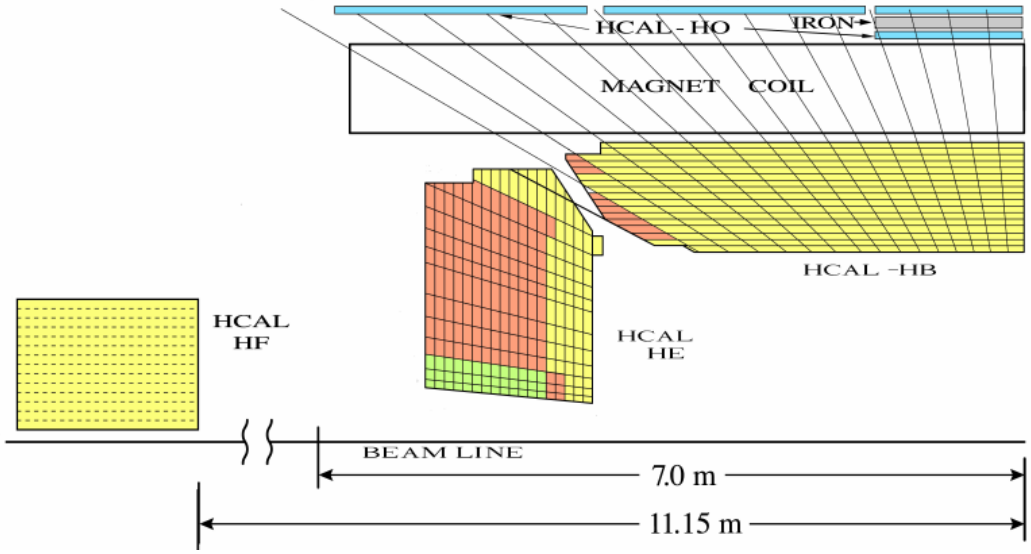


Figure 2.7: Schematic view of one quarter of the HCAL. The figure is taken from [65]

The HCAL central barrel (HB) and endcap (HE) subdetectors are joined hermetically covering $|\eta| < 3.0$. They completely surround the ECAL and are fully immersed within the high magnetic field of the solenoid. Therefore they are restricted to the radial dimensions of $1.77 < r < 2.95\,\text{m}$. Since the outer extent of the electromagnetic calorimeter (r = 1.77 m) and the inner extent of the magnet coil (r = 2.95 m) [62].

The HB consists of 36 identical azimuthal wedges which form the two half-barrels (HB+ and HB–). The wedge is composed of flat brass absorber plates parallel to the beam axis. The innermost and outermost absorber layers are made of stainless steel for structural strength. In total 17 active plastic scintillator tiles are interspersed between the stainless steel and brass absorber plates. The HE has a similar system of alternating absorber and plastic scintillator. The scintillators in both the HB and HE have a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and 0.17×0.17 for $|\eta| > 1.6$ [63].

Due to the restricted dimensions the HCAL and ECAL do not always contain all of the energy from the particle showers. An outer hadron calorimeter (HO) functions as a tail-catcher for hadronic showers, and is placed outside the solenoid effectively increasing the thickness of the barrel calorimeter. The two forward Cherenkov light

detectors (HF) made of quartz fibers, which run longitudinally through the absorber and collect Cherenkov light, are embedded within a 165-cm-long steel absorber. Each of the HF is placed at 11.2 m from the interaction point extending the pseudorapidity from $|\eta| < 2.3$ to $|\eta| < 5.2$. Since the forward calorimeter is covering the region $2.9 < |\eta| < 5.2$, it plays an important role in identifying forward jets, enables better measurements of missing transverse energy, and helps measuring the luminosity [66].

Neutral hadronic interactions are the most important consideration for the energy resolution of the HCAL as, unlike charged particles, no additional information can be obtained from the tracker to combine with the calorimetry measurement to improve the resolution. The HCAL energy resolution was measured in a pion test beam to be:

$$\frac{\sigma}{E} = \left(\frac{110\%}{\sqrt{E}}\right) \oplus 9\%, \tag{2.7}$$

where E is expressed in GeV.

2.2.5 The muon detection System

The CMS muon system is composed of four stations of gas-ionization muon detectors installed outside the solenoid and sandwiched between the layers of the steel return yoke. It was designed for muon reconstruction with the aid of information collected by the Silicon Trackers. The iron yoke is instrumented with a muon spectrometer for muon identification, and for momentum measurement and for better muon triggering [67, 68]. The muon system consists of detectors with three different technologies. As seen in Figure (2.8): **Drift Tube (DT)** chambers in the barrel region $|\eta| < 1.2$, **Cathode Strip Chambers (CSC)** to handle the higher rates and non-uniform magnetic field in the endcap region $0.9 < |\eta| < 2.4$ and **Resistive Plate Chambers (RPC)** are located in both barrel and endcap regions covering the range of $|\eta| < 1.2$. There are 4 stations in the barrel and in each endcap labeled $MB1 - MB4$ and $ME1 - ME4$, respectively, where the "station" is an assembly of chambers around a fixed value of r (in the barrel) or z (in the endcap). Chambers in the barrel are divided into five "wheels", with Wheel-0 centered at $z = 0$, two wheels in the $+z$ direction and two wheels in the $-z$ direction. Similarly in the r-direction in the endcaps, there are "rings" labeled $ME1/n - ME4/n$, where the integer n increases with the radial distance from the beam line [70].

During the LS1 of the LHC, three types of interventions are done [71]:

- All three DT, CSC and RPC systems engaged in an extensive campaign of reparations, in order to recover dead channels due to the electronics or the high voltage system.
- Relocation of the Sector Collector (SC) electronics [4] of the DT system, and a simpler copper-to-optical (CuOF) converter has been installed within the CMS detector towers and 3500 optical links were routed to the counting room where optical-to-copper converters (OFCu) feed the relocated SC. This was done to

[4] A system of VME crates for data concentration and optical transmission installed inside the CMS cavern

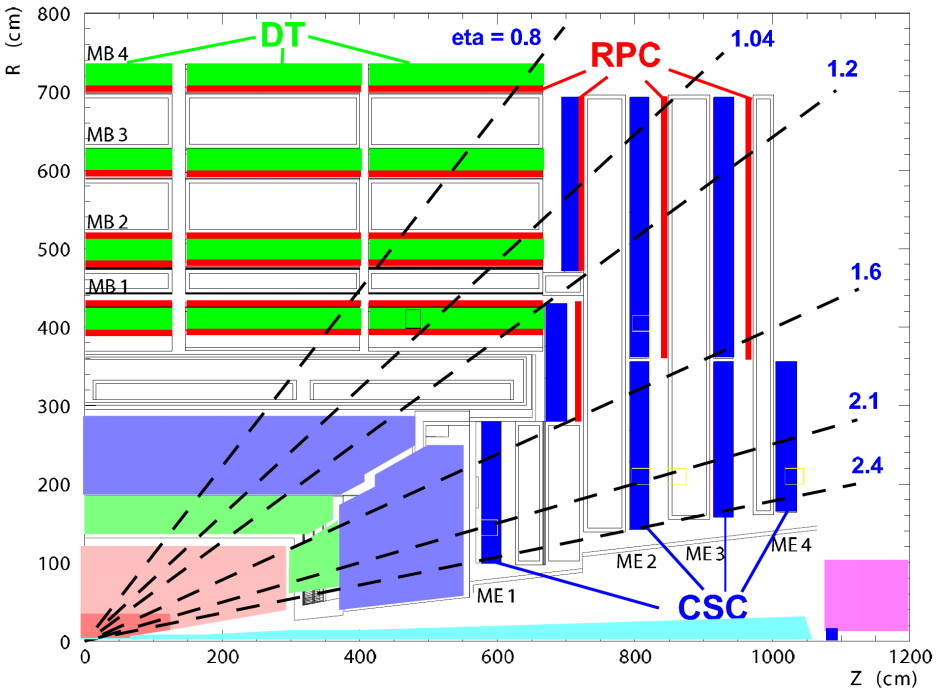


Figure 2.8: Schematic view of one quarter of the Muon System showing different parts with DT in the Barrel, CSC in the EndCap and RPC in both. The figure is taken from [69]

overcome access limitations to the experimental cavern, hence the full DT data is available on optical links in the counting room, paving the way for further Level-1 trigger upgrades.

- CSC and RPC chambers had been installed in the fourth endcap disks, this represents the completion of the original CMS muon system design. The increased redundancy will allow the trigger performance to be improved at high luminosity, in terms of a better p_{T} resolution and a better rejection of fake muons and other backgrounds.

2.2.6 Data acquisition

The LHC's bunch crossing frequency for proton-proton interactions is 40 MHz, and for Run 2 approximately 25 inelastic collisions occur every 25 ns corresponding to an interaction rate of the order of 1 GHz. For such huge datasets there is neither the readout bandwidth nor the storage capability to store all of the data produced. Therefore these amounts of data must be reduced down to about 100 Hz. This reduction is achieved by trigger systems. The purpose of the trigger system is to identify interesting event candidates fulfilling a predefined set of criteria.

The trigger system of CMS works in two sequential stages: the Level-1 (L1) Trigger and the High-Level Trigger (HLT).

- **The Level-1 (L1) Trigger**: A high bandwidth, fixed latency (3.2 μs) system based on FPGAs (Field Programmable Gate Arrays) and ASICs (Application

Specific Integrated Circuits) that has to take the decision to accept or reject an event within a few microseconds after a collision. The L1 Trigger uses only calorimetric and muon data, since the bandwidth requirement for the tracker data is too large to allow read out of every bunch-crossing. It is read out only once the L1 Trigger is passed. For safe operation the Global Trigger electronics are located away (on average $\approx$ 90 m) from the detector, in the counting room adjacent to the experimental underground cavern.

- **High-Level Trigger (HLT)**: A multi-stage iterative algorithm that uses full detector information to reproduce the L1 Trigger decision and then to iteratively improve on this decision by the staged introduction of fine-grained calorimetry information and tracking information. The aim of the HLT is to reduce the data rate so that it can be recorded to disk and tape storage at approximately 100 Hz for further physics analyses.

2.2.7 Computing at CMS

The selected data by the above mentioned triggers need to be stored and transferred to research institutes for further processing. This is achieved via the Worldwide Large Hadron Collider Computing Grid (WLCG) [72]. It combines the computing resources of many of the institutes participating to LHC experiments around the world. WLCG consists of three tier-levels that are distributed among worldwide centers as shown in Figure 2.9.

The data collected by the experiments are processed as well as stored at Tier-0. Then the data is distributed to 14 Tier-1 and around 150 Tier-2 computer centres. These Tier-1 and Tier-2 centres take care of the data reprocessing and storage. They are also used for the production of simulated data. Physicists around the world can access and analyse the data from Tier-2 centres.

For CMS experiment the Tier-0 located at CERN. The data is then distributed from the Tier-0 centre to 7 Tier-1 centres located in France, Germany, Italy, Spain, Taiwan, UK, and the US. Thereafter the data is further distributed from Tier-1 centres to over 40 Tier-2 centres.

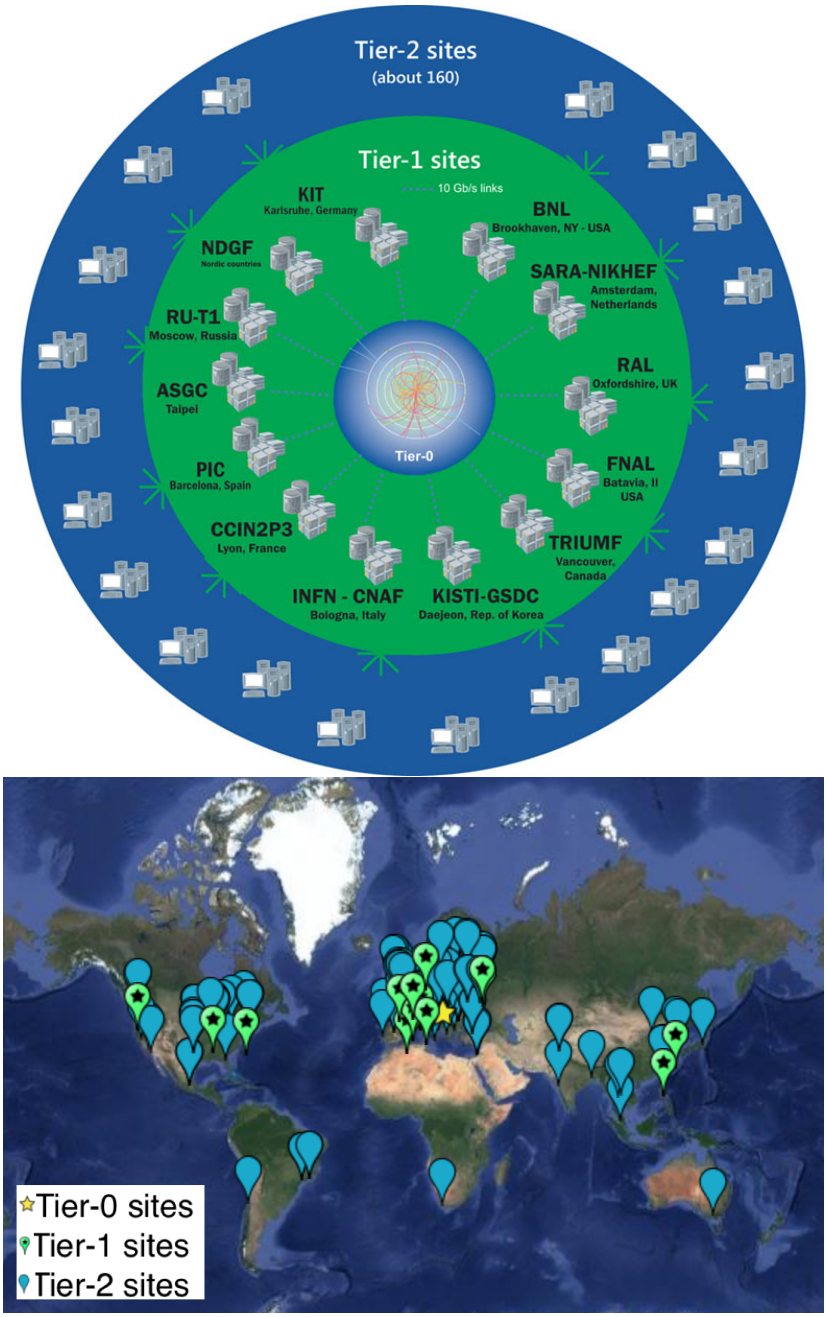


Figure 2.9: The structure of the WLCG tier-levels (top) and the mapping of them among worldwide (bottom). Figures are taken and adapted from [73].

"Look deep into nature, and then you will understand everything better"

Albert Einstein

Chapter 3

Event Generation

Many models in particle physics are developed based on fundamental hypotheses. In order to confront any theoretical model with empirical observations its induced phenomena should be predicted in a form that can be compared with experimentally observed features. This is the function of simulation programs. The Monte-Carlo (MC) event simulation at colliders is considered as a transformer of the theoretical model into a form that can be analyzed like the real data collisions.

Through different software (generators and simulators) we can get simulated samples which are called MC-samples. The MC generators are used to simulate Standard Model (SM) background as well as our required signals of new physics phenomena, while simulators are used to simulate the detector response.

In this chapter we will explain how the collision event looks like at the LHC and how it proceeds until final state particles that will interact with materials of different types of detectors in the CMS experiment as described in Chapter 2. Secondly we will overview the MC programs used in this research.

3.1 The physics of proton-proton collision

The generation of a pp $\rightarrow$ X collision event means generating the whole process starting from the incoming protons at the LHC to the experimentally observable final states X. These final states are detected with different kinds of detectors depending on the type of the final state. The process of generating the event takes place through sequential steps [74] (subprocesses) as illustrated in Figure 3.1 and as summarized below.

- **Parton Distribution Functions**
  Each of the incoming protons consists of 3 valence quarks as well as sea quarks and glouns. Physicists call each of these constituents a "parton". The longitudinal momentum of the proton is distributed among its constituents (partons) according to what is called the Parton Distribution Functions (PDF) that will be described in more details in Section 3.1.1

- **Hard interaction**
  A parton from one direction collides with another parton coming from the other

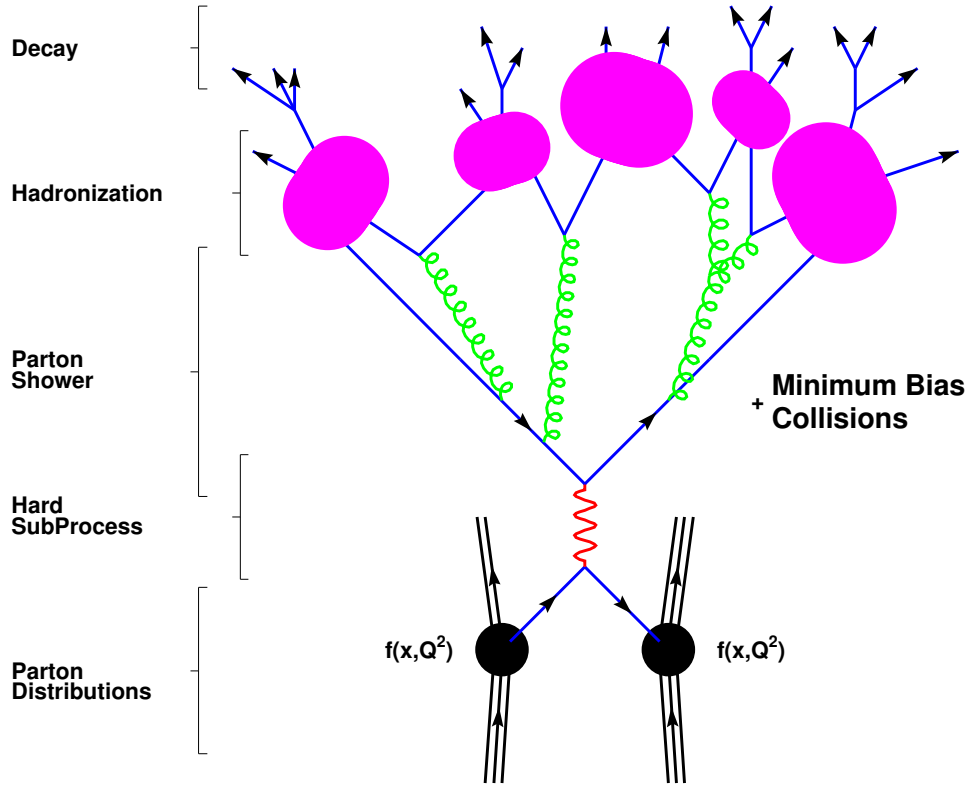


Figure 3.1: Schematic representation of the structure of the event generation chain for proton collisions - the figure is taken from [75].

direction producing the interaction of interest which is called the hard interaction or scattering. As a result new "colour-charged" particles can be produced. Among these particles some might have a very short lifetime such as top quarks, H, $W^{\pm}$ and Z^0 -bosons , their decay will be considered as a part of this subprocess. Their intrinsic properties will be propagated from the production down to their daughter particles. This will be described in more details in Section 3.1.2.

- **Parton Showers**
  The produced partons from the hard scattering can branch into other partons, that can branch themselves also, resulting in a shower of secondary partons. The same holds for the incoming partons. The radiation governed by the QCD process, that comes out of the hard process is categorized into two types according to its origin. The first category is called the initial state radiation (ISR) which originates from the incoming partons while the radiation of a second category from the outgoing particles is called final state radiation (FSR). The perturbative method to include the parton shower (PS) radiation in the event generation is controlled by the coupling constant α_S (with $\alpha_S < 1$) and is discussed in Section 3.1.3.

- **Hadronization**
  Because the phenomenon of colour confinement, the generated partons are pre-

vented to stay free. It forces partons to fragment and group into hadrons. The phase which results in colour neutral hadrons is known as hadronization. These newly-created hadrons are usually unstable. They will further decay at various timescales until stable or sufficiently long-lived particles are created which can be observed directly in the detectors. This hadronisation is dictated by the non-perturbative method using some phenomenological models as will be described in Section 3.1.4.

- **Underlying event:**
  The colliding partons take in general only a small fraction of the energy of the incoming protons. Much of the energy remains in the beam remnant, which continues to travel essentially in the original direction. These remnants also carry colour charges that compensate the colour taken away by the colliding partons. Also in the previous steps we have considered the interaction from only one parton from each of the incoming protons that participates into the hard collision while the proton is made up of more partons. Hence more partons can collide with other partons from the other proton, resulting in multi-parton interactions (MPI). The hadronization of beam-remnant and the multi-parton interaction (MPI) phenomena are referred to as the *underlying event* (UE) which will be described in more details in Section 3.1.5.

At this point, we have to match and combine the event-generator description to a detector simulation framework. It is only after this point that experimental information can be retrieved and used to reconstruct the initially states and interactions.

For the simulation of the different steps in the event generation chain described above many software programs have been developed. Some MC techniques are used to reproduce the stochastic nature of each step in the event generation chain and others are used for simulating the paths of the outgoing particles through the detector after which the response of the detector is as well simulated. More details about these MC techniques will be introduced in Section 3.2.

3.1.1 Parton Distribution Functions (PDF)

At high-energy hadron colliders, such as the LHC, the collisions involve composite particles (protons) with internal structure (quarks and gluons). The proton momentum is split among its , sea and valence; partons.

The hard interaction (process of interest) comes out of a collision between two partons i and j, one from each direction. These partons carry momentum fraction x_i, x_j of the incoming protons i, j. At this partonic level the center-of-mass energy $\sqrt{s}$ of the interaction gets reduced to $\sqrt{Q^2} \equiv \sqrt{s} = \sqrt{2x_i x_j E_{proton}}$, where Q^2 is the factorisation scale more commonly denoted as μ_F. This factorisation scale represents the energy at which the hadronic interaction can be expressed as a product of the partonic cross section and the process independent PDF. From the QCD factorization theorem the hadronic cross section of a particular process $\sigma_{pp\to X}$ can be determined as a function of the differential partonic cross-section $d\hat{\sigma}_{ij\to X}$ weighted with the parton distribution functions $f(x, Q^2)$ [76] by the following equation:

$$\sigma_{pp\to X} = \sum_{i,j=q,\bar{q},g} \iint dx_i dx_j f_i(x_i, Q^2) f_j(x_j, Q^2) \hat{\sigma}_{ij\to X(x_i,x_j,\alpha_S(Q^2))} \tag{3.1}$$

where the sum is over the possible initial parton flavors. The parton distribution functions $f(x, Q^2)$ determine the probability to find a parton with fraction x of the proton momentum when it is probed at an energy scale Q^2, as can be represented in Figure 3.2.

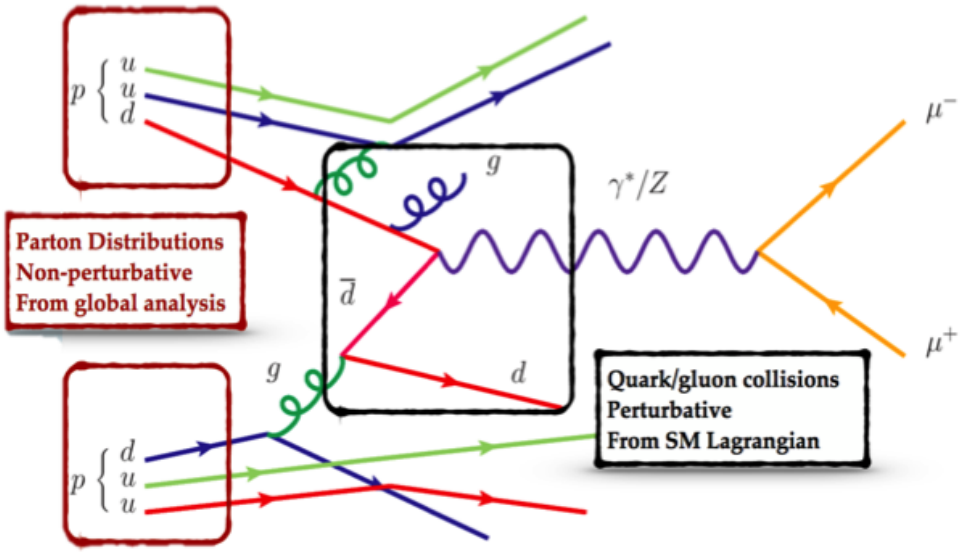


Figure 3.2: Schematic representation of the pp hard interaction at the LHC. The calculations of the hadronic cross-sections are a combination of perturbative quark/gluon-initiated processes, and non-perturbative parton distributions information. The figure is taken from [77].

Perturbative QCD does not work below energy scales of about 1 GeV, because at low energy the strong coupling constant α_S becomes larger than one. Since the PDFs contain a non-perturbative part that can not be determined from first principles, they has to be estimated from measurements. The PDFs are determined and included in global PDF sets that combine the experimental data from Deep Inelastic Scattering (DIS) experiments from the HERA collider where the structure of the proton is probed with electrons [78, 79] together with proton-antiproton collisions from the Tevatron [80], and proton-proton collisions data at the LHC (Run 1) [81]. These combined PDF sets known as the PDF4LHC combined sets that are recommended for new physics searches. Their measurement at scale Q^2 is extrapolated to higher energies by use of the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) equations [82]. Also a new generation of global PDF sets has been developed for use at the LHC Run 2 [83].

A statistical combination of PDFs from NNPDF [84], CT14 [85] and MMHT2014 [86] are used for the construction of the PDF4LHC15 combined sets that is used in the signal production used for the physics analysis in thesis.

An example of PDFs from the NNPDF collaboration is shown in Figure 3.3 for two interaction energy scales $Q^2 = 10\,\text{GeV}^2$ and $Q^2 = 10^4\,\text{GeV}^2$. It can be seen that for

small values of the momentum fraction x, the gluon density dominates over the quark densities. Hence, it is easier to probe the gluons than the quarks. For x values close to 1, the parton densities of the up and down quarks (the valence quarks of the proton) dominate over the gluon density. The PDFs for anti-up and anti-down quarks are in general lower, since these sea quarks originate in the proton only from gluon splitting. The uncertainty on the parton distributions is evaluated using the Hessian technique [87].

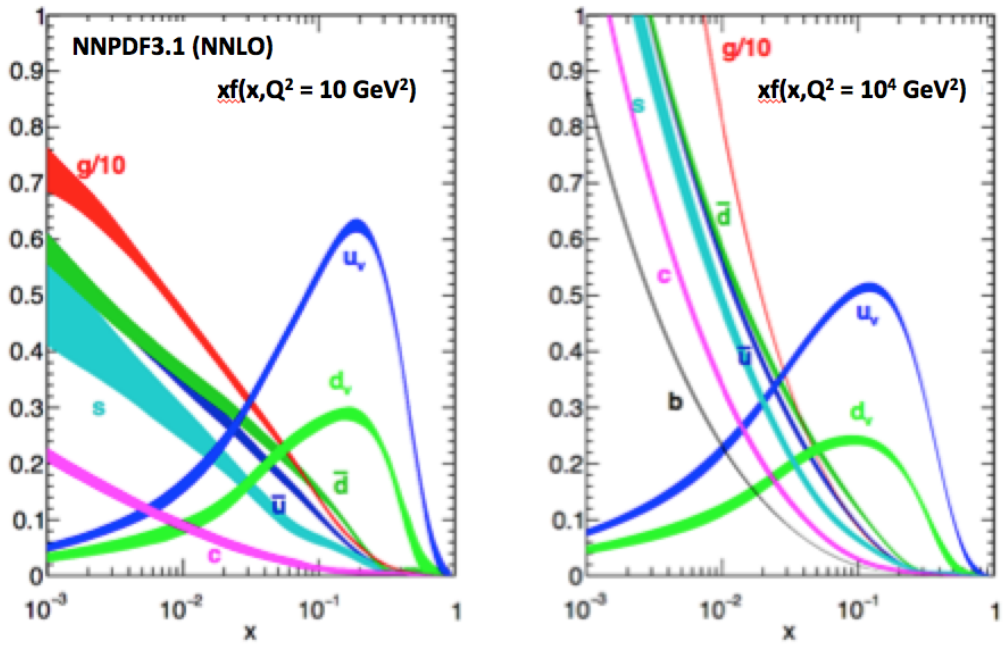


Figure 3.3: The NNPDF3.1 NNLO PDFs evaluated for parton flavours at $Q^2 = 10\,\mathrm{GeV}^2$ (left) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (right). The plots are taken from [88].

3.1.2 Hard scattering

The hadronic cross-section as given by Eq. 3.1 is a function of the differential partonic cross section or the cross section of the hard interaction $\hat{\sigma}_{ij\to X}$ between two partons. The differential cross section of this process for specific phase space can then be written as:

$$d\hat{\sigma}_{ij\to X} = \frac{1}{2s_{real}} \frac{|\mathcal{M}_{ij\to X}|^2}{8(2\pi)^2}\, dps \tag{3.2}$$

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

Also the hadronic cross section $d\hat{\sigma}_{pp\to X}$ can be expanded in the strong coupling constant α_S. If the energy scale of the interaction is high enough (equivalently, if the strong coupling constant α_S is small enough) a perturbative calculation of the total

partonic cross section is possible. Hence the matrix element can be calculated up to a certain perturbative order in α_S.

The quantum fluctuations can cause ultraviolet divergences at high energies. Therefore a certain renormalization scale μ_R needs to be chosen to deal with these divergences. Typically μ_R is used to redefine physical quantities in the way which makes the theory able to describe the experimental regime. Consequently the coupling constants will run as a function of μ_R and beyond μ_R the divergences due to the high energy corrections, like loop corrections to the propagators are absorbed in the physical quantities through a renormalization of the field. Particularly for the strong coupling constant α_S it was found to be:

$$\alpha_S = \frac{\alpha_S(\mu_0^2)}{1 + \alpha_S(\mu_0^2)\frac{33-2n_f}{12\pi}\left(\frac{|\mu_R|^2}{\mu_0^2}\right)} \tag{3.3}$$

where n_f the number of quarks and μ_0 the refrence scale at which the coupling is known. From equation 3.3, it can be noticed that the coupling strength inversely proportional to the renormalization scale, this is known as asymptotic freedom. Moreover by following the behavior of $\alpha_S(\mu_R^2)$ it was found that a limit $\Lambda_{QCD} \approx 200\,\text{MeV}$ for which α_S becomes larger than one. Additionally Infrared and collinear divergences (caused by the emission of respectively soft (low-p_T) gluons and gluons collinear to the partons) are also encountered in the calculation of the matrix element.

Some event generators like `MadGraph5_aMC@NLO` [89, 90] and `POWHEG` [91] and `MC@NLO` [90] can generate events up to next-to-leading order (NLO) that will be summarized in Section 3.2.1.

3.1.3 Parton Shower

The physics of additional radiated partons emitted by the incoming and outgoing partons of the hard interaction is described by the parton showering formalism. Successive splitting processes occur before and after the hard scattering and result in showers of partons. The showering continues until the energy of the partons reaches values below Λ_{QCD} for which the perturbative showering approach is not valid. Once this happens hadronization models are used to further describe the non-perturbative regime.

Three different types of splittings exist. The splitting of a gluon into a quark and an anti-quark ($g \to q\bar{q}$) , the splitting of a quark into a quark and gluon ($q \to qg$) or the splitting of a gluon into two gluons ($g \to gg$). At each splitting, the momentum of the mother parton is divided among the daughter partons. Consider mother parton "a" splits into two daughters partons "b" and "c". If b takes fraction z of the mother "a" energy then c will take fraction (1-z) of the mother energy. These daughters can in turn branch to other partons at a lower Q^2 scale. An accurate description of parton showering at leading order can be provided by the matrix element event generators.

A perturbative evolution of the parton fragmentation functions is possible for the energy scales of the hard interaction using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) formalism

3.1.4 Hadronization

As mentioned above the parton-shower cascade continues down to the splitting energy Λ_{QCD}. At this point the hadronization of partons starts. Because the hadronization process is theoretically not yet fully understood from first principles, different phenomenological models have been developed to simulate the process. Currently there are mainly two phenomenological approaches. One based on string fragmentation while the other is based on the cluster model. In PYTHIA the hadronization is simulated using the Lund string model [92, 93] based on the idea of a physical picture of a colour flux tube "string" connecting a quark q and anti-quark $\bar{q}$. As the two quarks move apart, the string gets stretched and potential energy $V(r) = kr$ builds up, where r is the distance between q and $\bar{q}$. The string constant k corresponds to the amount of energy per unit length and is estimated to be $\approx 1\,\mathrm{GeV}/fm$. When $V(r)$ becomes high enough a new pair of quarks $q'\bar{q'}$ is produced with a mass m. In this way two new colour singlets $q\bar{q'}$ and $q'\bar{q}$ are formed. Further breakings might occur provided that the invariant mass of the new colour singlets is high enough [94]. The splitting process continues until only colour-neutral hadrons with an on-shell mass remain and at each splitting the probability of string-breaking is proportional to

$$exp\left(\frac{-\pi(m^2 + p_T^2)}{k}\right), \tag{3.4}$$

where p_T is the transverse momentum of the quarks in the pair. Light quarks (up, down and strange) are included in this model while heavy quarks (charm, bottom) are ignored as the probability to create a heavy quark pair is heavily suppressed. Gluons in the string model are set as small snags on the string between two quarks. The Lund string model is infrared and collinear safe [93].

3.1.5 Underlying event

The hadronization of the beam-remnant and the multi-parton interaction (MPI) phenomena are referred to as the *underlying event* (UE) [95]. These processes are not independent from that in the hard-scattering event, because color connections between these processes exist.

The activity coming from MPI is softer in transverse momentum ($p_T \leq 3\,\mathrm{GeV}$) than that of the hard interaction, hence producing mainly low energetic hadrons. The perturbative MPI differential cross-section diverges as p_T^{-4} of the outgoing partons at center-of-mass frame. This divergence is regulated in QCD MC models by including a smooth phenomenological cutoff $p_{T,0}$ as follows:

$$p_T^{-4} \rightarrow (p_T^{-2} + p_{T,0}^{-2})^{-2}. \tag{3.5}$$

This formula approaches the perturbative result for large p_{T}, and a finite result as $p_{\mathrm{T}} \rightarrow 0$. It also regulates the divergence of the strong coupling α_S at low p_{T}. The cutoff $p_{T,0}$ depends on the center-of-mass energy $\sqrt{s}$. In PYTHIA this energy dependence is regulated as

$$p_{T,0}(\sqrt{s}) = p_{T,0}^{ref}\frac{\sqrt{s}}{\sqrt{s_0}}. \tag{3.6}$$

where $p_{T,0}^{ref}$ is the reference cutoff at reference energy $\sqrt{s_0}$. The amount of MPI depends on the PDF and the value of $p_{T,0}$. As a result the simulation of the UE is tuned by $\sqrt{s_0}$, PDF, p_{T0}^{ref} and α_S [96].

We must distinct the MPI from the Pile-Up events where several proton pairs collide during the bunch-crossing. Furthermore, it is obvious that each of these collisions can be coupled with ISR and FSR.

3.1.6 Pile-up

In-time pile-up (PU) collisions originate from the scattering of protons in the same bunch crossing of the hard process of interest. They mainly consist of soft QCD interactions. *Out-of-time pile-up* collisions originate from the other (previous/next) proton bunch crossings. Both are modeled in a similar way as the UE.

3.2 Event generation and detector simulations

There are many MC event generators used to generate full collision events as described in previous sections. The MC method allows these steps to be considered sequentially. Within each step a set of rules is defined and used iteratively to construct a more and more complex state. This ends with hundreds of particles moving out from the interaction point in different directions.

In real life, machines produce events that are stored by the data acquisition system of a detector. In the virtual reality, event generators like `MadGraph` and `POWHEG` [97] in combination with `PYTHIA` play the role of machines like the LHC. Detector simulation programs like `GEANT4` [98] play the role of detectors like CMS [75].

3.2.1 MC Generators

The main event generators used for the production of simulated samples used in the physics analysis in this thesis are the following:

1. `MadGraph`
   `MadGraph` is one of the leading-order (LO) event generators that calculates the ME from tree-level Feynman diagrams. Additional partons are included (as extra legs in the Feynman diagram) to get an approximate description of ISR and FSR. It takes PDF sets as input which describe the kinematics of the incoming partons. Next it generates all Feynman diagrams for a particular process and evaluates each ME for a given phase space point. The number and type of partons and the kinematics of the event are generated. Based on these generated ME, the hard process events are generated by `MadEvent` and LO cross sections can be calculated. In CMS the `MadGraph/MadEvent` LO matrix element generator is very widely used for the simulation of both the known SM processes and BSM physics processes.

   The `MadGraph5_aMC@NLO` package [99] uses both tree-level and one-loop Feynman diagrams. It can simulate events at next-to-leading order (NLO) in perturbative

QCD. These additional corrections from higher-order Feynman diagrams make the simulation more accurate compared to its LO counterparts. This package also includes ISR and FSR.

Since higher order perturbations are included in the cross-section calculations, it is necessary to consider terms which interfere destructively. This is achieved for such samples by assigning negative weights to some events within the generator.

The matching of MC@NLO with parton shower programs is more ambiguous due to the fact that the matrix element includes up to N+1 partons. The final N+1 parton state can be produced either from the N parton matrix element or from the N+1 parton matrix element. Analytically The MC@NLO program calculates how the showering of N parton matrix element populates the N+1 parton phase space [99]. This calculation is subtracted from the N+1 parton matrix element showering to obtain the "true" N+1 events. The difference is considered as being part of the N-parton final state. Now the N-parton and (N+1)-parton event classes are defined. Hence double counting can be avoided by applying full showering to both the N-parton and (N+1)-parton final states separately.

Since virtual loops are not fully included in LO generators, the quantity and properties of additional partons in the final state might not be described accurately. A LO cross section can be scaled to the NLO level using a so-called k-factor. It is defined as the ratio of theoretical NLO and LO cross section calculations. Caution is needed, however, as such a k-factor may depend on the kinematic phase space and the probed energy scale.

2. POWHEG
   POWHEG is the abbreviation of Positive Weight Hardest Emission Generator. It generates the hardest process in the event from QCD computations for a selected set of physics processes [100]. The matching between POWHEG and hadronization programs such as PYTHIA is based on the p_{T} ordering of the parton emissions. The highest p_{T} showers are matched to the hard interaction first. This is used to avoid the double counting of the low-p_{T} radiation by applying p_{T}-veto.

3. PYTHIA
   The PYTHIA [101, 102] program is one of the most widely used generators amongst LHC experimentalists. It produces particularly good agreement with data compared to other PS generators. It takes the parton-level event generated by ME-generators to produce the parton shower (PS) after adding soft emissions from ISR and FSR state particles. PYTHIA performs the fragmentation and hadronisation of quarks and gluons as well. It can also simulate everything standalone including the initial protons fragmentation, MPI, hadronisation and underlying event.

3.2.1.1 Matching matrix elements with parton showers

"Matching" or "merging" refers to the method of combining the well separated particles of the hard scattering output with parton showers. The p_{T} threshold at which partons

from the ME are matched to the PS is known as the ME-PS threshold which can be summarized as follows [103]:

- Consider N partons in the hard processes. Another $n \leq N$ matrix-element partons are generated in a separate sample with certain requirements on the kinematics of the partons in the final state. The requirements are: $p_{\mathrm{T}}^{parton} > p_{\mathrm{T}}^{min}$, $|\eta| < |\eta^{max}|$ and an angular separation $\Delta R_{i,j} > \Delta R_{min}$ between partons i, j in (η, ϕ) plane where $\Delta R_{i,j} = \sqrt{(\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2}$.

- The events with n partons are passed to `PYTHIA` for showering. The shower evolution is impelled down to lower Q^2 scales and subsequently clustered. The clustering algorithm, described in Section 4.5, is applied with a jet radius parameter ΔR_{jet}^{min}. Additionally jets are required to have a minimum transverse energy E_{T}^{min}. The parameters ΔR_{jet}^{min} and E_{T}^{min} are related to the parton level parameters ΔR_{min} and p_{T}^{min} but not necessarily identical.

- The procedure continues by associating every generated parton to one and only one of the jets, by taking the hardest parton (maximum p_{T}^{parton}) and by selecting the closest jet. If the distance between the jet axis and the parton is smaller than ΔR_{jet}^{min} the parton and jet are matched. The matched jet and parton are then removed from the list and the next parton is matched.

- For $n < N$ parton configurations, events are rejected when partons are not matched to jets or when there are more jets than partons. For the $n = N$ parton configurations, events are only rejected if there are unmatched partons and/or there are unmatched jets which are harder than the softest parton.

Several schemes exist for matching samples with a different final state multiplicities. In `MadGraph` two types of matching are used. For LO matrix elements the MLM-merging [104] scheme is used, while FxFx-merging is used for NLO matrix elements. For `MadGraph/MadEvent` generated processes in the physics analysis of this thesis.

3.2.2 Detector simulators

After the parton shower and hadronization, the particles pass through the detector. Therefore their interactions with the detector material needs to be simulated. For a detailed detector simulation, `GEANT4` is used. This toolkit includes the active detector regions (i.e. the sensitive layers of the detector that are designed to detect traversing particles) and the dead material regions (e.g. gaps between detector components, support structures and cables). `GEANT4` provides a mapping of the magnetic field as well, which is crucial to simulate the curvatures of charged particles. The interactions of the traversing particles with the detector material, the detector response and its conversion into electric signals are also simulated. Based on well understood underlying interaction mechanisms, the energy losses of the traversing particles are simulated as well as their trajectories in the detector. The simulated effects include for instance Bremsstrahlung of charged particles and photon conversions to an electron-positron

pair. The showering of electrons, photons and hadrons in the calorimeters are simulated as well.

In GEANT4 pileup interactions are simulated from extra (soft) pp-interactions and by processing their corresponding detector hits. The number of simulated pileup interactions is chosen such that it approximates the observed distribution of the number of expected additional soft pp-collisions. This number will be reweighted once the true distribution of number of pileup interactions per bunch crossing is measured in actual data. The number of simulated pileup interactions usually follows an expected distribution. The detector response simulation is very CPU intensive and can take (depending on the process) several minutes for one event.

3.3 Signal and background generation

The MC sample for the FCNC process of interest is called *Signal*, while all other MC samples are considered as *Background*. In the physics analysis of this thesis the parton distribution functions used for the production of the Monte-Carlo samples are taken from the CTEQ collaboration [105].

The generation of SM background events follows in steps illustrated in Figure 3.4.

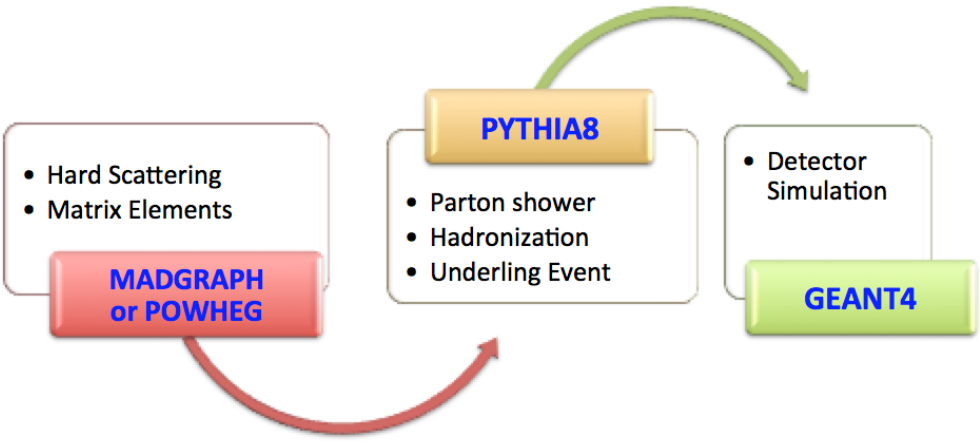


Figure 3.4: The successive processes of the MC event generation.

MadGraph/MadEvent is used for the simulation of some background processes (e.g. $t\bar{t}+V$, DY+Jets, W+jets). POWHEG is used to generate $t\bar{t}+jets$, $t\bar{t}+H$, and single top quark production in both the t-channel and the tW-channel. PS including ISR, FSR and MPI, hadronization and UE is simulated with PYTHIA8. Finally GEANT4 is used for the CMS detector simulation. All SM background events in the physics analysis presented in this thesis are simulated using this full simulation of the CMS detector.

An overview of these samples is given in Table 3.1. The generators of each process are listed as well as the corresponding theoretical cross-sections at 13 TeV. The $t\bar{t}+jets$ cross section has been calculated with Top++v2.0 program [106]. Single top quark cross-sections have been calculated by HATHOR v2.1 [107–109]. Cross sections for electroweak processes have been calculated by programs like MCFM 6.6 [110].

Process	**Generator**	$\sigma(pb)$	**Order**
$t\bar{t} + jets$ (inclusive)	POWHEG	$831.8^{+5.6}_{-6.1}$	$NNLO$
Single top (SM-Stop):			
t-channel (t)	POWHEG	$136.0^{+5.4}_{-4.6}$	NLO
t-channel ($\bar{t}$)	POWHEG	$80.9^{+4.1}_{-3.6}$	NLO
tW-channel (t)	POWHEG	35.9 ± 2.6	NLO
tW-channel ($\bar{t}$)	POWHEG	35.9 ± 2.6	NLO
s-channel	MC@NLO	$10.1^{+0.4}_{-0.4}$	NLO
DY+jets$\rightarrow ll$			
$10\,\text{GeV} < m_{ll} < 50\,\text{GeV}$	MadGraph/MadEvent	18610	NLO
$m_{ll} > 50\,\text{GeV}$	MadGraph/MadEvent	5765.4	NLO
Di-bosons			
$WZ \rightarrow 3l\nu$	POWHEG	4.43	NLO
$ZZ \rightarrow 4l$	POWHEG	1.3	NLO
WW	POWHEG	12.2	NLO
$t\bar{t}$**V**			
$t\bar{t} + W$	MadGraph/MadEvent	0.20	NLO
$t\bar{t} + Z$	MadGraph/MadEvent	0.25	NLO
$t\bar{t}H$, $(H \rightarrow no\ b\bar{b})$	POWHEG	0.2586	NLO

Table 3.1: Summary of simulated SM processes considered for the physics analysis in this thesis. V in $t\bar{t} + V$ means vector boson and inclusive means all decay channels are considered. The order indicates to what order the cross-section is calculated. The quoted uncertainties on $t\bar{t}$+$jets$ and single top quark cross sections include uncertainties from renormalisation and factorisation scale, top mass and PDF+ α_S.

3.3.1 FCNC signal generation

The FCNC signal is composed of two components. The first component describes events where the FCNC interactions is considered in single top quark production whereafter the top quark decays according to the Standard Model. The second component describes events where the SM production of $t\bar{t}$ is followed by an FCNC decay for one of the top quarks, while the other decays according to the Standard Model.

The FeynRules package [111] and its Universal FeynRules Output (UFO) format [112] are used to implement the FCNC Lagrangian (Equation 1.28) in `MadGraph5_aMC@NLO` (version 2.2.2). The chiral parameters are chosen and fixed to be $f_{Xq}^{R} = 1$ and $f_{Xq}^{L} = 0$. The single top quark and $t\bar{t}$ signals are generated independently using `MadGraph5_aMC@NLO` with the following options:

- For the FCNC in the single top quark production process, no additional partons are included in the generation to avoid contributions from FCNC top quark pairs into the single top quark signal.
- For the FCNC interactions in the $t\bar{t}$ decay process, up to two additional partons are added in the initial hard process at LO.

The cross sections are calculated by convolution of the hard scattering matrix elements with the LO set of NN2.3LO parton density functions. Then the generated hard scattering events are matched to parton showers via `PYTHIA8`, where the simulation of the QCD environment relevant to hadronic collisions is done as well.

The branching ratio for the anomalous coupling is obtained from the partial width. The latter is estimated via `MadGraph5_aMC@NLO` and consequently the branching ratio $BR(t \to\ qX)$ is calculated using the following equation:

$$BR(t \to\ qX) = \frac{\Gamma_{t\to\ qX}}{\Gamma_t^{SM} + \Gamma_t^{FCNC}} \approx \frac{\Gamma_{t\to\ qX}}{\Gamma_t^{SM}} \tag{3.7}$$

where ($\Gamma_t^{SM} = 1.32158\,\text{GeV}$) is the full SM decay width of the top quark as calculated in [31] for a top quark mass of 172.5 GeV. As the top quark width corresponding to FCNC, Γ_t^{FCNC}, is very small compared to that of SM, it can be ignored in the denominator. The partial width for the Hqt coupling is calculated for κ_{Hut} and κ_{Hct} and it is the same for u and c quarks. The production of anomalous single top quark signals is calculated by the event generator at LO. The NLO effect is obtained by multiplying the LO cross-section with a global k-factor. The k-factor for each coupling and cross-sections at 13 TeV is given in Table 3.2.

As can be noticed from Table 3.2, the cross-section for the anomalous Hut coupling is significantly higher than that of the Hct coupling. This is related to the proton composition that contains more up-quarks than charm-quarks.

For FCNC $t\bar{t}$ signal, the SM-$t\bar{t}$ production cross section is calculated with `MadGraph5_aMC@NLO` at NLO at a center-of-mass energy of 13 TeV which is $\sigma_{t\bar{t}} = 674.1\ pb$. The total cross-section of FCNC $t\bar{t}$ signal is given by considering $t\bar{t} \to bWqH$ with the branching ratio $BR(t \to Hq)$ equal to its current 95% CL. upper limit. The total cross-section is the same of both κ_{Hut} and κ_{Hct} and equal to 13.53 pb. For the SM top quark decay only the leptonic decay is considered i.e. $t_{SM} \to Wb \to l\nu b$ where

Process	**Anomalous coupling**	$\Gamma_{t\to\ qX}(\text{ GeV})$	$\sigma(pb)$	k-factor	$\sigma_t/\sigma_{\bar{t}}$
$pp \to tH \oplus \bar{t}H$	κ_{Hut}	0.19 κ^2_{Hut}	5.06	1.5	0.1287
$pp \to tH \oplus \bar{t}H$	κ_{Hct}	0.19 κ^2_{Hct}	0.7	1.8	0.4983

Table 3.2: LO partial width for the anomalous H(u/c)t couplings. the $\kappa_{Hqt} = 1$ is chosen for the signal production.

$l = e, \mu$ and τ, knowing that $BR(t_{SM} \to Wb) \simeq 1$ and $BR(W \to l\nu) = 0.324$. For the Higgs boson decay only $H \to W^+W^-$ is considered since $BR(H \to ZZ) = 0.027$ and $BR(H \to \tau^+\tau^-) = 0.063$ that are very small relative to $BR(H \to WW) = 0.215$.

"scientific truth is not a frozen dogma waiting to be unveiled , it is a patient construction of the human mind based on the observation of an objective reality interpreted according to a rational analysis .."

Serge Haroche

Nobel Laureate 2017

Chapter 4

Reconstruction and Identification of Physics Objects

As discussed in Chapter 2 the purpose of the CMS detector is to detect, distinguish and identify the vast amount of particles created in the proton-proton collisions provided by the LHC. The interaction point of these collisions are located in the center of the detector. The first step in order to be able to identify the particles relies on the information gathered from the raw data provided by the different components of the detector. This information is then processed in order to form the physics objects relevant to the analysis presented in this dissertation. In CMS, the Particle-Flow (PF) algorithm is used for reconstructing a complete description of all stable particles traversing the detector in a given collision event.

In this Chapter, the main physics objects considered in the analysis will be described. The physics analysis uses the standard object selection criteria recommended by the CMS physics object groups (POG). The PF reconstruction technique will be explained in Section 4.2 [113]. More details about the reconstruction and identification of muons and electrons will be discussed in Section 4.3 and 4.4 respectively. The charged and neutral hadrons are clustered into jets as explained in Section 4.5. Finally the missing energy reconstruction will be summarized in Section 4.7.

4.1 Primary Vertex and track reconstruction

As discussed in Chapter 3 many interactions happen at each bunch crossing. As a result a large number of vertices is to be reconstructed along the z-direction as illustrated in Figure 4.1. The goal of the primary-vertex reconstruction is to measure the location of all proton-proton interaction vertices in each event. This is done by using the available reconstructed tracks. It is important to mention that these reconstructed vertices include the 'signal' vertex and all vertices from typically pileup collisions [**?**]. The reconstruction consists of three steps:

1. Selecting tracks consistent with being produced promptly in the primary interaction region.

2. Clustering tracks that appear to originate from the same interaction vertex based on their z-coordinates at their point of closest approach to the center of the

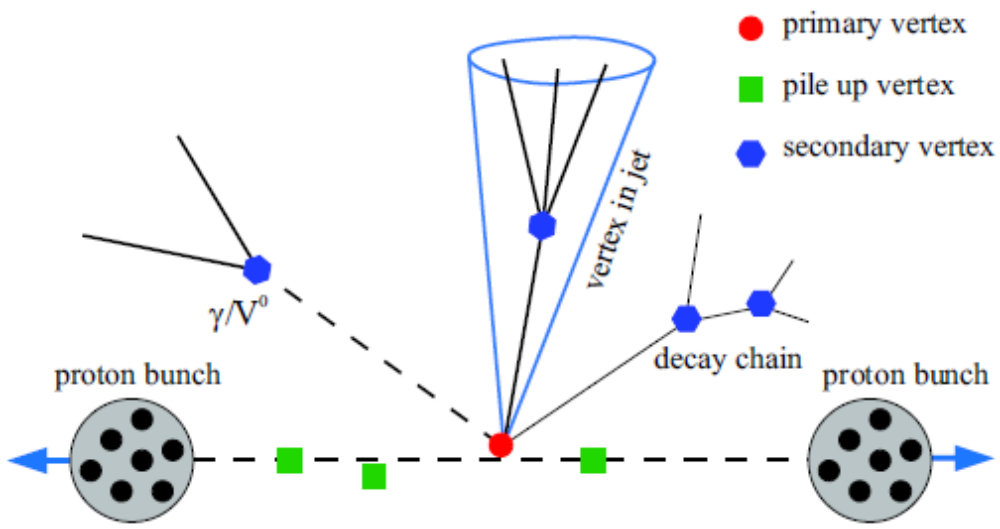


Figure 4.1: Schematic diagram showing the different types of reconstructed vertices. This figure is taken from the CERN website [114].

beam spot. This clustering approach allows reconstruction of any number of pp interactions in the same LHC bunch crossing.

3. Fitting the position of each vertex using its associated tracks. The resolution of the reconstructed primary-vertex position depends strongly on the number of tracks used to fit the vertex and the transverse momenta of those tracks.

Primary vertices are then ranked according to the sum of the momentum squared of all the tracks considered to have originated from that vertex. The vertex with the largest sum is regarded as the signal vertex, i.e. the most interesting proton collision.

The aim of tracking is to determine the charged particle properties: origin, transverse momentum, and direction. This is done using a combinatorial track finder (CTF) algorithm (described in detail at Section 2.2.2). Track reconstruction described in the following sections is essential in the reconstruction of isolated muons, identifying energetic and isolated hadronic τ decays and for tagging b-quark jets.

The reconstructed trajectories from random particle hits in the tracker are considered to be *fake tracks*. Any charged hadron missed by the tracking algorithm would be merely detected by the calorimeters as a neutral hadron. However it comes with reduced efficiency [1], largely degraded energy resolution and biased direction due to the bending of its trajectory in the magnetic field. Increasing the track reconstruction efficiency while keeping the mis-reconstructed rate unchanged is therefore critical for PF event reconstruction. On the other hand the nuclear interactions in the tracker material may lead to either a kink in the original charged particle trajectory, or to the production of a number of secondary particles.

[1]The reconstruction efficiency is defined as the fraction of simulated tracks reconstructed with at least 50% of the associated simulated hits, and with less than 50% of unassociated simulated hits. The mis-reconstruction rate is the fraction of reconstructed tracks that cannot be associated with a simulated track.

4.2 Particle-Flow reconstruction

The function of the particle-flow algorithm is to collect and combine all information from the CMS subsystems. The algorithm uses these data to reconstruct the whole event by identifying all final state particles traversing the detector as shown in Figure 4.2. The combined information gives more precise measurements of particle momenta compared to using information from a single sub-detector. The PF reconstruction algorithm starts by considering fundamental elements namely charged-particle tracks and calorimeter clusters. The tracks are obtained from the iterative-tracking procedure described in Section 2.2.2. The muon information from the muon chambers 2.2.5. The calorimeter clusters from the ECAL 2.2.3 and HCAL calorimeters 2.2.4 are formed, in steps, for each calorimeter separately as following. First, a calorimeter-cell seed is defined by energy deposits above a given energy threshold. Next, topological clusters are formed from the seeds by joining adjacent cells if their corresponding energy is large enough. The energy thresholds are chosen to suppress contributions from electronic noise.

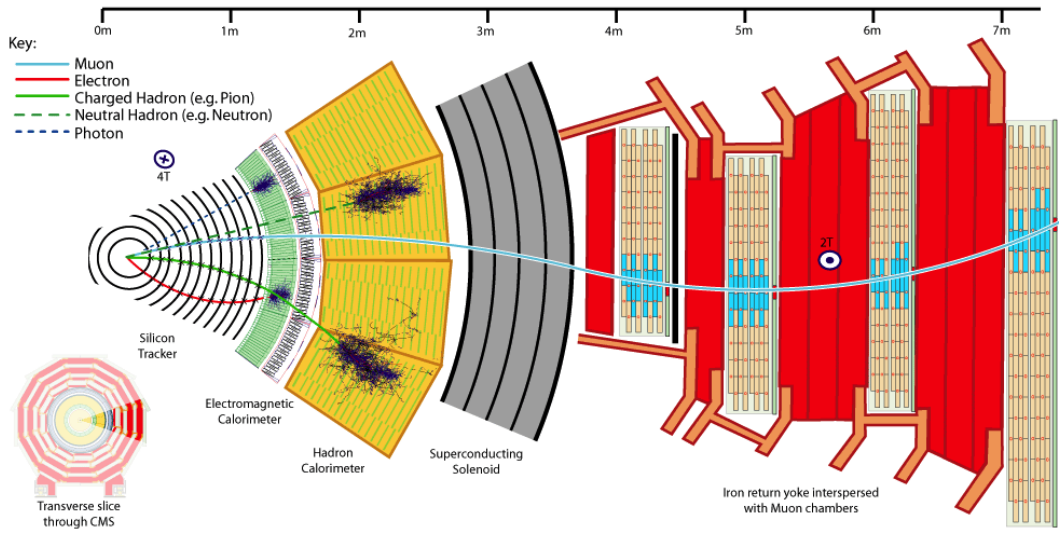


Figure 4.2: A sketch presents the behavior of different particle interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector. The muon and the charged pion are positively charged, and the electron is negatively charged. Neutral hadrons leave no tracks in the tracker but have deposits in the ECAL and HCAL. Photons leave deposits in the ECAL but not the HCAL. The figure is taken from [115].

The PF algorithm links the elements, i.e. the charged particles tracks and the clusters from calorimeters, into "blocks". Then the algorithm interprets them in terms of muons, electrons, photons and charged and neutral hadrons. The particle-flow algorithm reconstructs objects in a well defined order starting from the easiest as sketched in Figure 4.3. Once a block is found, the corresponding elements are removed from the event. Subsequently the next link is sought for, and so on until no more elements are left. The first objects to be reconstructed are PF muons. The most difficult objects

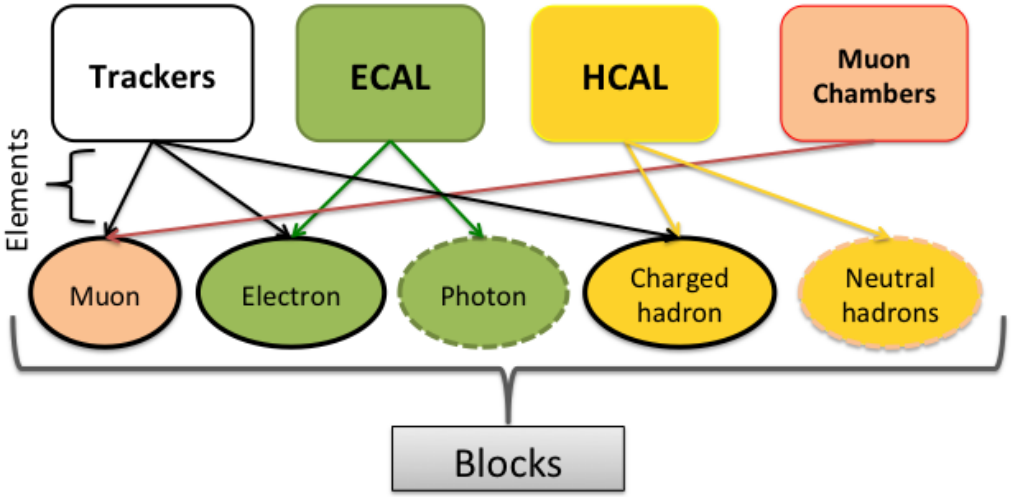


Figure 4.3: The first objects to be reconstructed are PF muons then electrons. After all muons and electrons are reconstructed the isolated photons in the event are identified and removed from the PF blocks. Then identification and reconstruction of charged hadrons starts, and finally neutral hadron.

such as neutral hadrons are one of the last objects to be reconstructed because their properties can be constrained from the previously reconstructed objects. Each muon identified from the muon chambers is associated to compatible hits in the tracker. This associated track and muon chamber hits are then removed from their respective collection. Tracker and ECAL elements are combined for the final identification of a PF electron. Again the track and ECAL clusters are removed from their respective collections. From the remaining charged-particle tracks and PF clusters photons, charged and neutral hadrons are reconstructed as follows. For tracks linked to PF clusters, the track momentum is compared to the PF-cluster momentum. If both momenta are compatible within uncertainties, a charged PF hadron is identified. If the 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 photon or hadron energy deposit. The decision depends on whether the calorimeter excess is measured in the ECAL or HCAL respectively [2]. After all charged hadrons in the event are found, the remaining calorimeter clusters in the HCAL and ECAL are reconstructed as neutral hadrons and photons respectively.

4.3 Muons reconstruction

Muon tracks are reconstructed independently in the inner CMS tracker producing what is a so-called *tracker track* and from (hits and segments) in the muon chambers what

[2] A charged hadron cannot 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.

is known as a *standalone-muon.*Track segments are formed from combinations of hits in the drift-tubes (DT), the cathode strips chambers (CSC) and the resistive plate chambers (RPC). The track segments in the innermost muon chambers are considered as seeds. Then a Kalman filter technique is used to propagate the track inside-out towards the outermost chamber. At each layer, the track is updated, and the final fitted track is refitted outside-in using the available track segments. Based on the tracker track and standalone-muon, two reconstruction approaches are used [116]:

- ***Global muon reconstruction (outside-in):*** A global-muon track is fitted combining hits from the tracker track and the standalone-muon track, using the Kalman-filter technique [55]. For each standalone-muon track, a matching tracker track is found by comparing parameters of the two tracks propagated onto a common surface.

- ***Tracker muon reconstruction (inside-out):*** In this approach all tracker tracks with $p_T > 0.5\,\text{GeV}$ and total momentum $p > 2.5\,\text{GeV}$ are considered as possible muon candidates. They are extrapolated to the muon system taking into account the magnetic field, the average expected energy losses, and multiple Coulomb scattering in the detector material. If at least one muon segment matches the extrapolated track, the corresponding tracker track qualifies as a Tracker muon. Track-to-segment matching is performed in a local (chamber) coordinate system, where local x is the best-measured coordinate (in the $r-\Phi$ plane) and local y is the coordinate orthogonal to it. The extrapolated track and the segment are considered to be matched if the distance between them in local x is less than 3 cm or if the value of the pull for local x is less than 4, where the pull is defined as the difference between the position of the matched segment and the position of the extrapolated track, divided by their combined uncertainties.

At large transverse momenta, $p_T > 200\,\text{GeV}$, the global-muon reconstruction can improve the momentum resolution compared to the tracker-only reconstruction. Tracker muon reconstruction is more efficient at low momenta, $p_T \leqslant 5\,\text{GeV}$, because it requires only a single muon segment in the muon system.

4.3.1 Muon identification and isolation

One important step in the muon reconstruction is to ensure that the reconstructed muon is well isolated from the surrounding hits. The latter comes from charged hadrons, neutral hadrons, photons and pileup. The measure of this isolation is denoted as the $\Delta\beta$-corrected Relative Isolation ***(RelIso*** $I_{rel}^{\Delta\beta}$***)*** and is given by:

$$I_{rel}^{\Delta\beta} = \frac{\Sigma p_T^{chargedhadron} + max\left(0, \Sigma p_T^{neutralhadron} + \Sigma p_T^{photon} - 0.5\Sigma p_T^{pileup}\right)}{p_T^{\mu}} \tag{4.1}$$

where the sum run over all PF particles in a cone of $\Delta R = 0.4$ around muon direction. Muons need only to be isolated from particles of the primary collision event. Therefore pileup effects in the same cone size around the muon direction is subtracted.

The max() function ensures that the corrected neutral hadronic energy is never defined as negative. The factor of 0.5 in front of the Σp_T^{pileup} term motivated by from observations that the neutral pileup contribution is estimated as half of the charged pileup contribution [113].

The muon identification proceeds by a set of selections based on the global and tracker muon properties. In Run 2 *tight*, *medium* and *Loose* ID working points (WP) are used to define three different identification criteria for the isolated muons reconstruction. One chooses a working point depending on the physics analysis envisaged. For each working point different requirements (i.e. cuts) are applied on variables such as those summarized in Table 4.1. The muon candidates are required to be global muons, with a small normalized χ^2 (i.e. $\chi^2/ndof$, which is the χ^2 divided by the number of degrees of freedom of the fit). The transverse impact parameter $d0$ in the $\phi-$plane and the longitudinal impact parameter dz of the muon with respect to the leading primary vertex should be small to ensure a good rejection of cosmic muons and muons originating from additional proton collisions (pileup) in the event. To ensure that the muon candidate is well separated from a requirement $\Delta R_{\mu,jet} > 0.4$[3] is added.

Compared to medium or loose muons tight muons have tighter requirements on various quantities including their transverse momentum because as the muon momentum decreases as is harder to distinguish it from other particles. They are in general used when defining selection requirements on how many muons should be in the event from the signal process. Loose muons are in general used to veto additional objects which are still likely to be muons but could be a misidentified object such as a pion. Also the loose WP is used for selecting muons from heavy and light quark decays. The loose criteria will capture more real muons in its selection but with a lower purity. Medium muons have the same criteria as loose muons with additional track-quality and muon-quality requirements. In this physics analysis the tight working point is used for the muon selection.

4.4 Electrons reconstruction

Electrons[4] are reconstructed by associating a track reconstructed in the silicon detector with a cluster of energy in the ECAL. Due to the lower mass compared to muons by two orders of magnitude, the reconstruction of well-isolated electrons is more complicated. Electrons suffer from large energy loss due to Bremsstrahlung in the tracker material. This is one of the biggest challenges in reconstructing electrons. These Bremsstrahlung photons can convert into e^-e^+ pairs in the tracker material creating secondary electrons which must be distinguished from electrons coming from the hard process. Therefore the Bremsstrahlung radiation needs to be taken into account to accurately measure the electrons momentum.

The traditional electron seeding strategy is called the ECAL-driven approach. The starting point of this approach is a seed crystal. The seed defined as the one containing most of the energy deposited in any considered region, that has a minimum E_T of

[3] $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$

[4] In this thesis "electron" includes both charges, i.e. electron and positron.

Requirements	**Tight WP**	**Medium WP**	**Loose WP**
Particle-Flow muon id	Yes	Yes	Yes
Which Muon	Global	Global or tracker	Global or tracker
$\chi^2/ndof$ of track	<10	<3	-
RelIso	< 0.15	< 0.25	< 0.25
d0 (cm)	<0.2	-	-
dz (cm)	<0.5	-	-
#pixel-hits	>0	-	-
#trackers with muon hits	> 5	-	-
#hits in muon-chamber	>0	-	-
#muon-stations with Muon segments	>1	-	-
$\Delta R_{\mu,jet}$	>0.4	>0.4	>0.4

Table 4.1: The table shows the muon criteria for each working point [117].

4 GeV ($E^{min}_{T,seed} = 4\,\text{GeV}$). Subsequently the energy of the radiated photons is collected that mainly spread along the $\phi-$direction. The spread in the $\eta-$direction is usually negligible because of the bending of the electron trajectory in the magnetic field. Two clustering algorithms exist. The "hybrid" algorithm in the barrel and the "$multi\ 5\times5$" in the endcaps, are used for this purpose, and are described in [118]. Arrays of crystals are added around the seed crystal, in a range of N_{steps} crystals in both directions of ϕ if their energies exceed a minimum threshold of E^{min}_{array}. The adjoining arrays are grouped into clusters. Each distinct cluster is required to have a seed array with energy greater than a threshold of $E^{min}_{seed-array}$ in order to be collected in the final global cluster, called the supercluster (SC). The energy of the supercluster is simply taken as the sum of the energy of all its constituent clusters, whereas its position is computed as the energy-weighted mean of the cluster positions.

The energy and position of the associated supercluster of electrons in jets are often biased because of the overlapping contributions from other particle deposits. This leads to large inefficiencies. Similarly for electrons with small p_T the radiated energy is spread over such an extended region that the supercluster cannot include all deposits. The missed deposits bias the position of the supercluster and prevent it from being matched with the proper hits in the innermost tracker layers.

Another issue is the propagation from the supercluster backward to the interaction region is likely to be compatible with many hits from other charged particles in the innermost tracker layers causing a substantial mis-reconstruction rate. In order to integrate the electrons missed by the ECAL-driven approach in PF reconstruction a tracker-driven electron seeding method is used.

The tracker-driven approach starts with all tracks reconstructed by iterative tracking (Section 2.2.2) and the ones with $p_T > 2\,\text{GeV}$ are considered as electron track seeds. In order to reduce fake identification of charged hadrons as electrons a pre-identification is applied considering two separate cases:

- *Case 1*: If the energy radiated by the electron is small, Bremsstrahlung is negligible. The KF algorithm collects hits up to the ECAL inner surface, where the KF track is well matched to the closest ECAL cluster. Its momentum is measured with good precision taking into account that to form an electron, the ratio of the cluster energy to the track momentum is required to be compatible with unity.

- *Case 2*: If significant Bremsstrahlung is present, the KF algorithm cannot follow the change of curvature of the electron trajectory because of the Bremsstrahlung. It either stops collecting hits, or keeps collecting them, but with a bad quality identified through a large value of χ^2_{KF}. Another alternative track filter is applied to such tracks using a Gaussian-sum filter (GSF) [119]. In the GSF approach all tracks with a small number of hits or a large χ^2_{KF} are therefore refitted using a Gaussian-sum filter (GSF). The number of hits and the quality of the KF track χ^2_{KF}, the quality of the GSF track χ^2_{GSF}, and the geometrical and energy matching of the ECAL and tracker information are used in a multivariate analysis (MVA) [5] to select the tracker seed as an electron seed.

[5]The multivariant analysis technique will be explained latter in Section 5.3.

The seeds from the ECAL-driven and tracker-driven approaches are collected into a list. Then all are 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 using a multivariate analysis.

4.4.1 Charge estimation

Another important challenge is the measurement of the electric charge of a reconstructed electron. Normally the electric charge identification for electrons, just like for muons, is derived from the direction of the curvature of the reconstructed track. However, the measurement of the electron charge is affected by Bremsstrahlung followed by photon conversions. Generally the electric charge is estimated from the sign of the GSF track curvature. Unfortunately this can be altered by the mis-identification probability in the presence of conversions, especially for $|\eta| > 2$, that can reach about 10% for reconstructed electrons from Z boson decays.

In order to improve the electron-charge measurement two other charge estimates are combined. One is based on the associated KF track matched to a GSF track when at least one hit is shared in the innermost region. The second is evaluated using the SC position where the charge-estimate defined as the sign of the difference in ϕ between the vector joining the beam spot to the SC position and the vector joining the beam spot and the first hit of the electron GSF track. Depending on these charge estimates two methods can be used to estimate the charge of the electron [118]:

- ***Majority method***: In this method the electron charge is defined by the sign shared by at least two of the three estimates.
- ***Selective method***: This is a higher purity method that is obtained by requiring all three estimates to agree.

The selective method is used mainly in analyses where the charge estimate is crucial for example to study of charge asymmetry in supersymmetry searches or other new physics searches that are using same-charge di-leptons as the case of our search for FCNC. The charge misidentification probability decreases strongly when the identification selections become more restrictive mainly because of the suppression of photon conversions. This is done in our study and will be presented in details in Section 5.2.1.1.

4.4.2 Electron identification

The electron candidates are required not to originate from a photon conversion to an electron-positron pair. The isolation for electrons is similar to muons with a small change in the pileup correction. In the case of electrons called an effective-area corrected relative isolation I_{rel}^{EA} is defined as:

$$I_{rel}^{EA} = \frac{\Sigma p_T^{chargedhadron} + max\left(0, \Sigma p_T^{neutralhadron} + \Sigma p_T^{photon} - \rho A_{eff}\right)}{p_T^e} \tag{4.2}$$

Here the term ρA_{eff} represents the neutral pileup contribution where ρ is defined as the transverse-momentum density of the event, and A_{eff} is the effective area. These

two variables are used in jet energy corrections which will be explained later in this chapter.

As for the muon case, in CMS three ID working points (WP) are defined for electron candidates named *Tight*, *Medium* and *Loose*. In this thesis the cut-based approach working points are used for selecting electrons. The variable $\sigma_{i\eta i\eta}$ represents the width of the electron shower in the η direction, calculated from an array of $5X5$ crystals around the energy deposit in the ECAL. The pseudorapidity differece $\Delta\eta(SC, GSFtrack)$ and the azimuthal difference $\Delta\phi_{in}$ are separations in angle between the SC and the GSF track. Electron candidates are rejected if the most inner expected hit of the reconstructed track is missing. The cuts differ for electrons associated to superclusters in the barrel and the endcap of the ECAL subdetector as summarized in Table 4.2 and 4.3 respectively.

In the cut-based approach d0 and dz cuts are not explicitly part of the tuned ID. This is because it is found that the efficiency strongly depends on the physics of the event. They are applied in addition, see Table 4.4.

Variable	**Tight WP**	**Medium WP**	**Loose WP**
$\sigma_{i\eta i\eta} <$	0.00998	0.00998	0.011
$\Delta\eta(SC, GSFtrack) <$	0.00308	0.00311	0.00477
$\Delta\phi(SC, GSFtrack) <$	0.0816	0.103	0.222
$Hcal/Ecal <$	0.0414	0.253	0.298
$I_{rel}^{EA} <$	0.0588	0.0695	0.0994
$\lvert 1/\mathrm{E} - 1/\mathrm{p}\rvert <$	0.241	0.134	0.0129
Conv. veto	yes	yes	yes
expected missing inner hits $\leq$	1	1	1

Table 4.2: The table shows the electron criteria for each working point in the barrel region ($|\eta_{SC}| \leq 1.479$) [120].

4.5 Jet reconstruction

Jets are the experimental signatures of energetic partons (quarks and gluons) produced in high-energy processes. When partons hadronise as discussed in Sections 3.1.4, they form a number of charged and neutral hadrons travelling in approximately the same direction as the original parton. These final state particles can be clustered into what is known as a jet. This is achieved by using various reconstruction algorithms. All jet-clustering algorithms start from a set of particles or calorimeter deposits and cluster

Variable	**Tight WP**	**Medium WP**	**Loose WP**
$\sigma_{i\eta i\eta} <$	0.0292	0.0298	0.0314
$\Delta\eta(SC, GSFtrack) <$	0.00605	0.00609	0.00868
$\Delta\phi(SC, GSFtrack) <$	0.0394	0.045	0.213
$Hcal/Ecal <$	0.0641	0.0878	0.101
$I_{rel}^{EA} <$	0.0571	0.0821	0.107
$\lvert 1/\mathrm{E} - 1/\mathrm{p}\rvert <$	0.0129	0.13	0.14
Conv. veto	yes	yes	yes
expected missing inner hits	1	1	1

Table 4.3: The table shows the electron criteria for each working point in the EndCap region ($|\eta_{SC}| > 1.479$) [120].

variable	**Barrel**	**EndCap**
$d0$ (cm)	0.05	0.10
dz (cm)	0.10	0.20

Table 4.4: These cuts are recommended by the cut-based approach for Run 2 2016 data.

them together in jets. A jet serves then as the high-level physics object representation of the parton in the final event analysis.

The two important properties which are required for any theoretically jet-clustering algorithm are the insensitivity to additional soft radiation (infrared safety) and the insensitivity to the collinear splitting of a hard parton (collinear safety). Therefore the most common algorithm used within the CMS collaboration is the so-called anti-k_T algorithm which is infrared and collinear safe.

The anti-k_T algorithm starts with a high p_T deposit in the calorimeter and considers it as a seed. Then it uses the distance definition of Equation 4.3 to find the nearest deposit to merge with.

$$d_{ij} = min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{4.3a}$$

$$d_{iB} = k_{T,i}^{2p}, \tag{4.3b}$$

where d_{ij} is the distance between particles i and j , d_{iB} is the distance between particle i and the beam. The values $k_{T,i}$ and $k_{T,j}$ represent the transverse momenta of particles i, j respectively, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, R is the radius parameter and p is a parameter to govern the relative power of the transverse momenta with respect to the geometrical Δ_{ij}^2 scale.

If the distance between the seed particle and the beam, d_{iB}, is smaller than the distance to another hit, d_{ij}, the particle is merged with the beam, otherwise the particle is merged with the nearest hit, according to the value of d_{ij}. The algorithm terminates when all particles i have $d_{iB} < d_{ij}$ and the merged particles are considered to be a jet [121].

4.5.1 Jet identification and jet-energy corrections

In the PF event reconstruction for CMS Run 2, a distance (cone) parameter of R = 0.4 is used to reconstruct jets. For the anti-k_T algorithm, the parameter p assumes a value of -1. This choice for p = -1 ensures soft particles are more likely to be associated to a closeby hard particle. Particle-flow jets (PF jets) are reconstructed by clustering the four-momentum vectors of PF candidates. Using tracker information additional to the calorimeter information improves the jet momentum and spatial resolution. The high granularity of the ECAL improves the energy resolution through the independent measurements of charged hadrons and photons inside a jet. The Particle Flow (PF) JetID criteria is a set of cuts which was constructed in order to reject fake, badly reconstructed and noise jets while retaining $98 - 99\%$ of the real jets. PFJetID is used in all CMS analyses which make use of PF jets. Two sets of JetID criteria are defined with certain cuts as described in Table 4.5. These cuts are recommended by the JetMET Physics Object Group (JME POG).

In clustering hadrons into jets different sources can affect the jet energy scale so some corrections have to be applied. Furthermore the jet energy resolution in simulated samples is different from that in real data, so an additional calibration has to be done for the simulated jet energy resolution.

Variable	**Tight Id**	**Loose Id**
Neutral Hadron Fraction	< 0.9	< 0.99
Neutral EM Fraction	< 0.9	< 0.99
Number of Constituents	> 1	> 1
Charged Hadron Fraction	> 0	> 0
Charged Multiplicity	> 0	> 0
Charged EM Fraction	< 0.99	< 0.99

Table 4.5: The table shows the two Jet-Id criteria. The cuts are applied for jets at $|\eta| < 2.4$ while for jets within $2.4 < |\eta| < 2.7$ only the first 3 cuts are applied. The table is taken from [122]

When comparing energies of the generated jet to the reconstructed jet it is observed that they are not identical. A fraction of the difference between them is due to the non-linearity of the detector response and the presence of pileup particles. Consequently reconstructed jets need to be calibrated in order to have the correct energy scale (i.e. the reconstructed jet energy matches the energy of the generated jet). This is the aim of the jet energy corrections (JEC).

A jet energy correction procedure is used to bring the jet energy response to unity and to remove any dependence on p_{T} and η. The jet energy response is defined as the mean ratio of the reconstructed jet energy to the reference (generated) jet energy. In CMS a factorized approach is used to apply a set of corrections sequentially and this with a fixed order. As can be seen in Figure 4.4, at each correction level the jets four-momentum vector is scaled with a scale factor. Each scale factor (SF) depends on various jet related quantities (p_T, η, flavour, etc.). The corrections can be summarized as follows [123]:

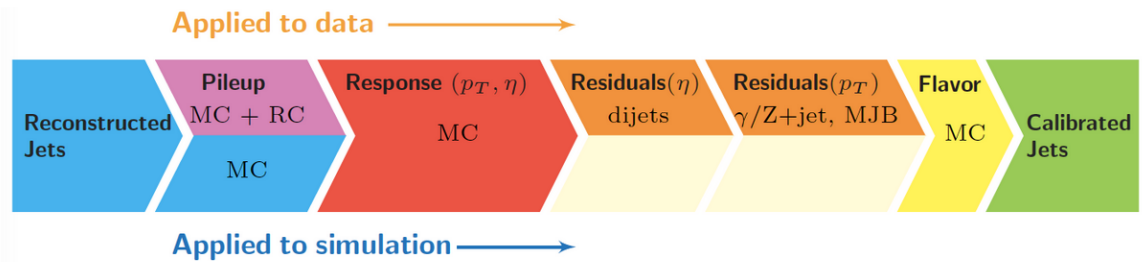


Figure 4.4: Jet energy correction levels

1. **L1 Pile Up**: At this level the energy coming from pile-up events is removed from the jet in both data and simulated events. It removes any dataset depen-

dence on luminosity. So the following corrections are applied upon a luminosity independent sample. Here the correction is estimated from simulated samples. Since the PU effects are introduced in detector simulation, these effects can be estimated by comparing the fully reconstructed jets to the generator jets. The corrected transverse momentum $p_{T,j}^{L1corr}$ of a PF jet j is calculated by subtracting the PU offset from the uncorrected transverse momentum as follows

$$p_{T,j}^{L1corr} = p_{T,j}^{uncorr} - PU\text{Offset}(p_{T,j}^{uncorr}, \eta_j, \rho, A_j) \tag{4.4}$$

where ρ is the offset energy density and A_j is the jet area.

2. **L2L3 MC-truth corrections**: Despite the L1 correction, some differences remain between reconstructed jet and generated jet energies. MC calibration factors are added to $p_{T,j}^{L1corr}$ in order to obtain a uniform (flat) response in both p_T and η. These calibration factors are determined from QCD multijet simulated events by measuring the jet energy response $\mathcal{R}$, $\mathcal{R} = p_{T,j}^{L1corr}/p_{T,GenJet}^{L1corr}$, as a function of $p_{T,j}^{L1corr}$ and η. The result will be $p_{T,j}^{L2L3}$ and it is calculated using Equation 4.5:

$$p_{T,j}^{L2L3} = \frac{p_{T,j}^{L1}}{\langle \mathcal{R}(p_{T,j}^{L1}, \eta_j) \rangle} \tag{4.5}$$

3. **L2L3 Residuals**: At this level corrections are added to correct the remaining small differences in jet energy response for data only. L2 residual correction (η-dependent) are determined from dijet events, where one of the jets is required to lie in the barrel region of the detector. 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 correction (p_{T}-dependent) is determined from data by studying Drell-Yan $Z/\gamma^* + jets$ events where jets are produced back-to-back. The leptonically decaying Z or γ can be used for jet calibration. Since the transverse momentum of leptons and photons is measured more precisely relative to that of jets it can be used to estimate the transverse momentum residual dependence of the jet corrections.

Additional jet energy (optional) correction steps have been developed in the CMS collaboration. For example to correct for jet-flavour dependencies of the energy response and the underlying event activity. These optional corrections are not used in this thesis, as they generally make small corrections compared to the recommended L1, L2 and L3 corrections.

The jet energy scale (JES) corrections explained above are associated to uncertainties that can be propagated as systematic uncertainties in physics analyses.

4.5.2 Jet energy resolution correction (JER)

Measurements find that the jet energy resolution (JER) in data is worse than in the simulation. Hence the simulated jets need to be smeared to describe the data.

In this analysis the scaling method is utilized to smear reconstructed simulated jets. As a result their p_T resolution matches that of reconstructed jets observed in data [124]. With this method the corrected four-momentum of a reconstructed jet in simulation is re-scaled with a factor c_{JER} which is defined by the following equation

$$c_{JER} = 1 + (s_{JER} - 1)\frac{p_T - p_T^{ptcl}}{p_T} \tag{4.6}$$

where p_T is transverse momentum, p_T^{ptcl} is the transverse momentum of the corresponding jet clustered from generator-level particles, and s_{JER} is the data-to-simulation core resolution scale factor which is measured in bins of η as given in Table 4.6.

$\|\eta\|$	s_{JER}	**uncertainty**
$0 - 0.5$	1.109	0.008
$0.5 - 0.8$	1.138	0.013
$0.8 - 0.1.1$	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

Table 4.6: Jet energy resolution scaling factors as a function of $|\eta|$ and their uncertainties for 13 TeV (2015 and 2016) data [125].

4.6 Identification of heavy-flavour jets (b & c jets)

In general it is not possible to know which was the flavour of the quark that produced the jet. In particle physics, the heavy-flavour hadrons are defined as those contain a "heavy" quark. These kind of hadrons have relatively long lifetimes so they travel on average up to a few millimeters before decaying. Jets of b-quark and c-quark hadrons are good representation of heavy-flavour jets. In the CMS software jets containing at least one b hadron are defined as b jets; the ones containing at least one c hadron and no b hadron are defined as c jets. The other jets (labeled “udsg”) are considered to be light-flavour jets. The identification of b-quark jets is a crucial part of this analysis and for many other searches for new physics beyond the Standard Model. Hence it will be discussed in more details.

The b-quark hadron is relatively long lived with a lifetime of order $1.5 \times 10^{-12} s$. This means that the B hadron travels on average a few millimeters before decaying. The decays of B hadrons often produce more than one charged particle and because of the relatively large mass of the b-quark, the decay products can be produced at a relatively large angle with respect to the original b-quark direction. Therefore tracks coming from B meson decays have larger Impact Parameter (IP) values than reconstructed tracks in light-flavour jets, where the impact parameter of a track is defined as the transverse distance of closest approach (DCA) of the track to the primary vertex. So the experimental signature for a b-quark is a jet of particles emerging from the primary vertex (PV) and a secondary vertex (SV) from the b-quark hadron decay which is displaced from PV by up to several millimeters, as indicated in Figure 4.5.

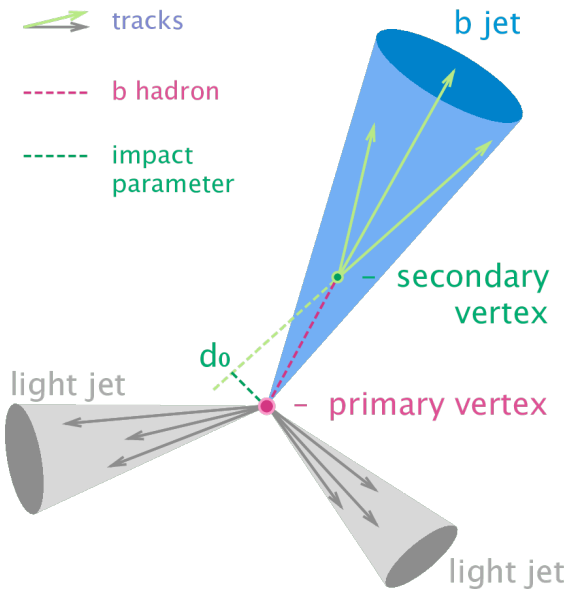


Figure 4.5: Illustration of a b-quark jet with a secondary vertex from the decay of the B meson resulting in charged particle displaced tracks. The distance between primary vertex and the secondary vertex is the flight distance of B meson while d0 is impact parameter. The figure is taken from [126].

The impact parameter of a track can be used to distinguish the decay products of a b hadron from prompt tracks. A more practical observable is the so-called *track impact parameter significance* IP/σ_{IP} where σ_{IP} is the measured uncertainty of the IP. The impact parameter has a positive or a negative sign. A positive sign is given for tracks that originate along the jet direction and a negative sign for the other tracks. Also the presence of a secondary vertex and the kinematic variables associated with this vertex can be used to discriminate between b and non-b jets.

In the CMS collaboration several algorithms have been developed for b-quark jet identification. They are called "b-tagging" algorithms or "b-taggers". The main inputs for these algorithms are the properties of the tracks clustered within the jet after applying appropriate selection criteria. For Run 2 the following algorithms have been developed [127]:

- **Jet probability algorithms:**
  The Jet Probability (JP) algorithm uses the signed impact parameter significance of the tracks associated with the jet to obtain a probability for the jet to originate from the primary vertex. Particularly it uses an estimate of the likelihood that all tracks associated to the jet come from the primary vertex. This probability is based on the probability density functions for the IP significance of prompt tracks that are extracted from data for different track quality classes. This is done using the shape of the negative part of the impact parameter significance distribution [128]. The variant to the JP algorithm, is the Jet B Probability (JBP) algorithm. The latter gives more weight to the tracks with the highest IP significance, up to a maximum of four of such tracks, which matches the average number of reconstructed charged particles from b-hadron decays.

- **Combined Secondary Vertex algorithms:**
  The CSVv2 b-tagging algorithms use a multivariate analysis techniques to combine both secondary vertex and track impact parameter information together in a likelihood discriminant. The discriminant output value ranges from zero to one. It distinguishes between b-jets and jets originating from "dusg" and "c" quarks.

  Two variants of the CSVv2 algorithm exist according to whether *Adaptive Vertex Reconstruction* (AVR) or *Inclusive Vertex Finding* (IVF) vertices are used. AVR fits tracks to a vertex, using only good-quality tracks from the jet in its procedure. When a vertex is found in the jet the associated tracks are removed and a next vertex is sought for with the remaining tracks [129]. In the IVF technique the track-clustering step is done before passing tracks through adaptive vertex fitting as follows. First all tracks with $p_T \geq 0.8\,\text{GeV}$ and longitudinal IP < 0.3 cm are collected and track-seeds are selected depending on their IP and IP significance. These tracks are then clustered based on the compatibility with the track-seed. If at least 70% of the tracks are shared between two IVF secondary vertices and the distance significance between the two vertices is less than 2, one of the vertices is removed. The efficiency of SV reconstruction for b-quarks using the IVF algorithm is 10% higher compared to reconstructing a secondary vertex with the AVR algorithm. For Run 2 IVF vertices are used in the CSVv2 algorithm.

  The CSVv2 algorithm is trained on inclusive multijet events in jet p_T and η bins. A set of powerful discriminating variables are combined into a neural network in the following vertex categories:

  - *RecoVertex*: The jet contains one or more secondary vertices.
  - *PseudoVertex*: No secondary vertex is found in the jet but a set of at least two tracks with a 2D IP significance and a combined invariant mass of at least 50 MeV away from the K_S^0 mass are found.
  - *NoVertex*: Containing jets not assigned to one of the previous two categories.

  Some examples of track-related discriminating variables can be seen at Figure 4.6, while discriminating vertex-related variables can be seen at Figure. 4.7. The full set of these variables and their definitions can be found in Appendix (A).

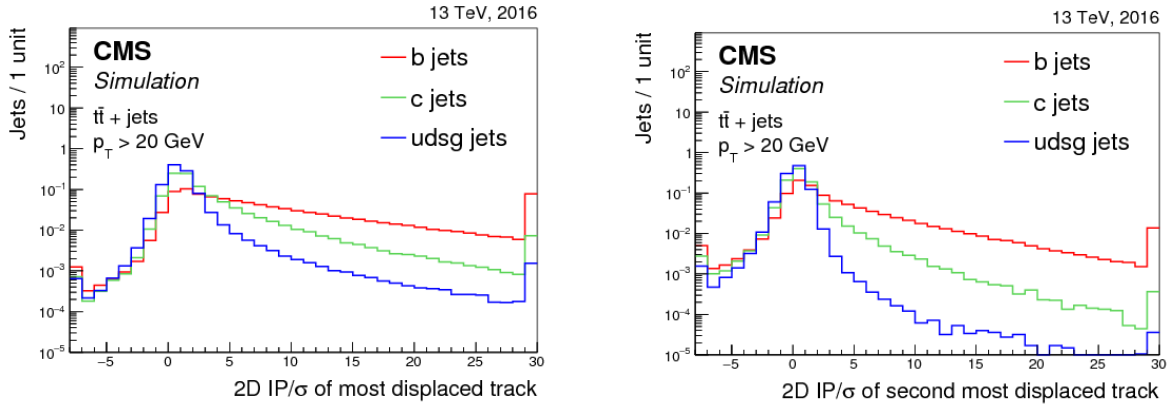


Figure 4.6: The distributions of the 2D impact parameter significance for the track with the highest (left) and second-highest (right) 2D impact parameter significance for jets of different flavours. The distributions are normalized to unit area. The first and last bin include the underflow and overflow entries, respectively. The figures are taken from [127].

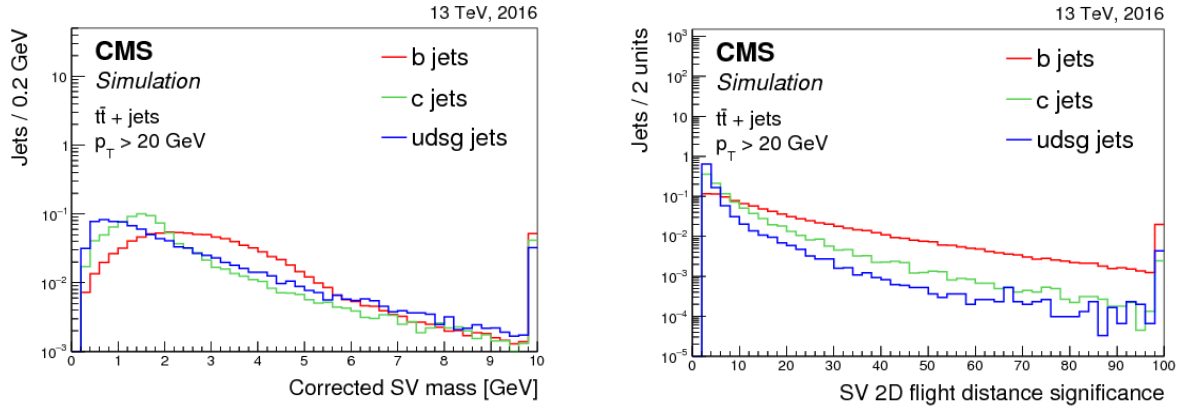


Figure 4.7: The distribution of the corrected secondary vertex mass (left) and of the secondary vertex 2D flight distance significance (right) for jets containing an IVF secondary vertex. The distributions are normalized to unit area. The figures are taken from [127].

Two dedicated neural network trainings are performed, one with c jets, and one with light-flavour jets as background. The final value of the discriminator is a linear combination of the output of these two trainings with relative weights of 1 : 3 for the output of the network trained against c and light-flavour jets, respectively. The values of the discriminator for the three vertex categories are combined with a likelihood ratio taking into account the fraction of jets of each flavour expected in $t\bar{t}$ events.

- **Soft-Lepton algorithms:**
  In about 20% of B hadron decays, a soft lepton is produced. For such B hadron decays, the information related to the charged lepton is used to construct a soft-

electron (SE) and soft-muon (SM) tagger. In this approach a Boosted Decision Trees (BDT) is used to construct the SE and SM discriminator. The input variables for the BDT containing lepton track(s) and their relation to the jet-axis.

- **Combined MVA (cMVAv2) algorithm:**
  The combined tagger, cMVAv2, uses the previous b-quark jet discriminators namely JP, JBP, CSVv2(IVF) ,CSVv2(AVR), SE and SM as input variables for an MVA to get a single but more performant discriminator. The training is performed using the open source SCIKIT-LEARN package [130].

- **The DeepCSV tagger:**
  The DeepCSV tagger is a recently developed version of the CSVv2 which uses a deep neural network for training with more hidden layers. The tracks, IVF secondary vertices and input variables that are used in this approach are the same as for the CSVv2 tagger. The only difference is that for the track-based variables up to six tracks are used in the training of the DeepCSV. In this approach the relative ratio of the number of jets of each flavour is set to 2 : 1 : 4 for b : c : udsg jets. Also a mixture of multijet and $t\bar{t}$ events is used in the training step to reduce the possible dependency of the training on the heavy-flavour quark production process. This discriminator was not included in cMVAv2 algorithm because it was developed after the cMVAv2 tagger.

The performance of a b-tagger, (usually illustrated with the ROC curve), is define as the b jet identification efficiency versus the mis-identification probability for either c or light-flavour jets. The efficiency of the b-tagger to correctly tag a jet with flavour f is defined as the number of jets of flavour f passing the tagging requirement divided by the total number of jets of flavour f . While the mis-identification probability is the probability to wrongly tag a jet as a b jet. The performance of different b-taggers is estimated in simulated $t\bar{t}$ events as shown in Figure 4.8 for jets with $p_T > 20\,\text{GeV}$ and $\eta < 2.4$.

The c-quark identification algorithm, is being studied and developed within the CMS collaboration only since Run 2. It is considered very promising to explore new-physics searches. The distributions of the tagging variables for c jets lie in between the distributions for b and light-flavour jets as can be seen in Figures 4.6 and 4.7. Therefore the techniques for identifying c-quark jets are very similar to those used for b-quark jet identification. The c-jet identification algorithm uses properties related to displaced tracks, secondary vertices, and soft leptons inside the jets [131].

The efficiency to tag b-jets and the rate of mis-identification of non-b jets depend on the working point chosen. Here again there are three working points (WP) for each tagger named *Tight* (T), *Medium* (M) and *Loose* (L) and are set corresponding to thresholds on the discriminator after which the mis-identification probability is around $10\%, 1\%$ and 0.1% respectively, for light-flavour jets. The working points and corresponding efficiency for b jets are summarized in Table 4.7. For JP and JBP taggers there are no specified working points to be used in the analysis directly. This is because for Run 2 these taggers are used only as input for cMVAv2 and in the calculations of the scaling factors for other taggers.

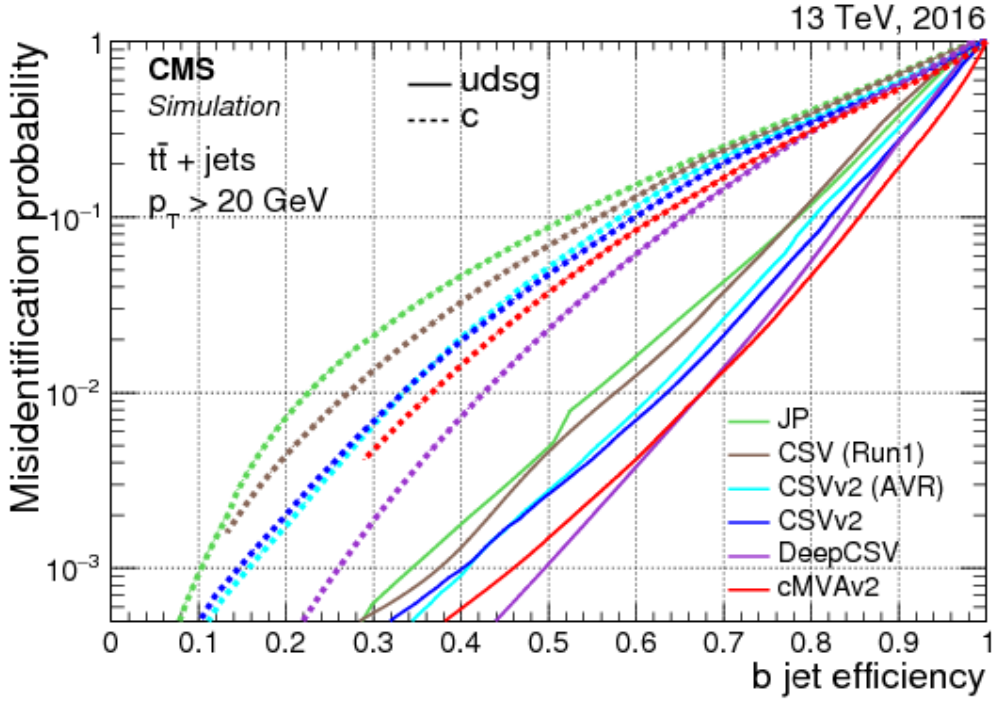


Figure 4.8: The b-jet identification efficiency and misidentification probability for c and light-flavour (dusg) jets for various b-tagging algorithms.

Tagger	Working points **label**	**Discr. cut**	ϵ_b%	ϵ_c%	ϵ_{dusg}%
Jet probability (JP)	JP L		78	37	9.6
	JP M		56	12	1.1
	JP T		36	3.3	0.1
Combined secondary vertex (CSVv2)	CSVv2 L	0.5426	81	37	8.9
	CSVv2 M	0.8484	63	12	0.9
	CSVv2 T	0.9535	41	2.2	0.1
Combined MVA (cMVAv2)	cMVAv2 L	-0.5884	84	39	8.3
	cMVAv2 M	0.4432	66	13	0.8
	cMVAv2 T	0.9432	46	2.6	0.1
Deep combined secondary vertex (DeepCSV)	DeepCSV L	0.2219	84	41	11
	DeepCSV M	0.6324	68	12	1.1
	DeepCSV T	0.8958	50	2.4	0.1

Table 4.7: The table shows the discriminator cut and efficiency corresponding to each working operating point provided by the BTV group for each b-taggers for the 2016 data [132].

The performance of the taggers was studied on simulated samples. It provides an accurate description and prediction of the measured data for variables used in b-quark taggers. However, small deviations are found in the final discrimination variables between data and simulation as can be seen in Figure 4.9.

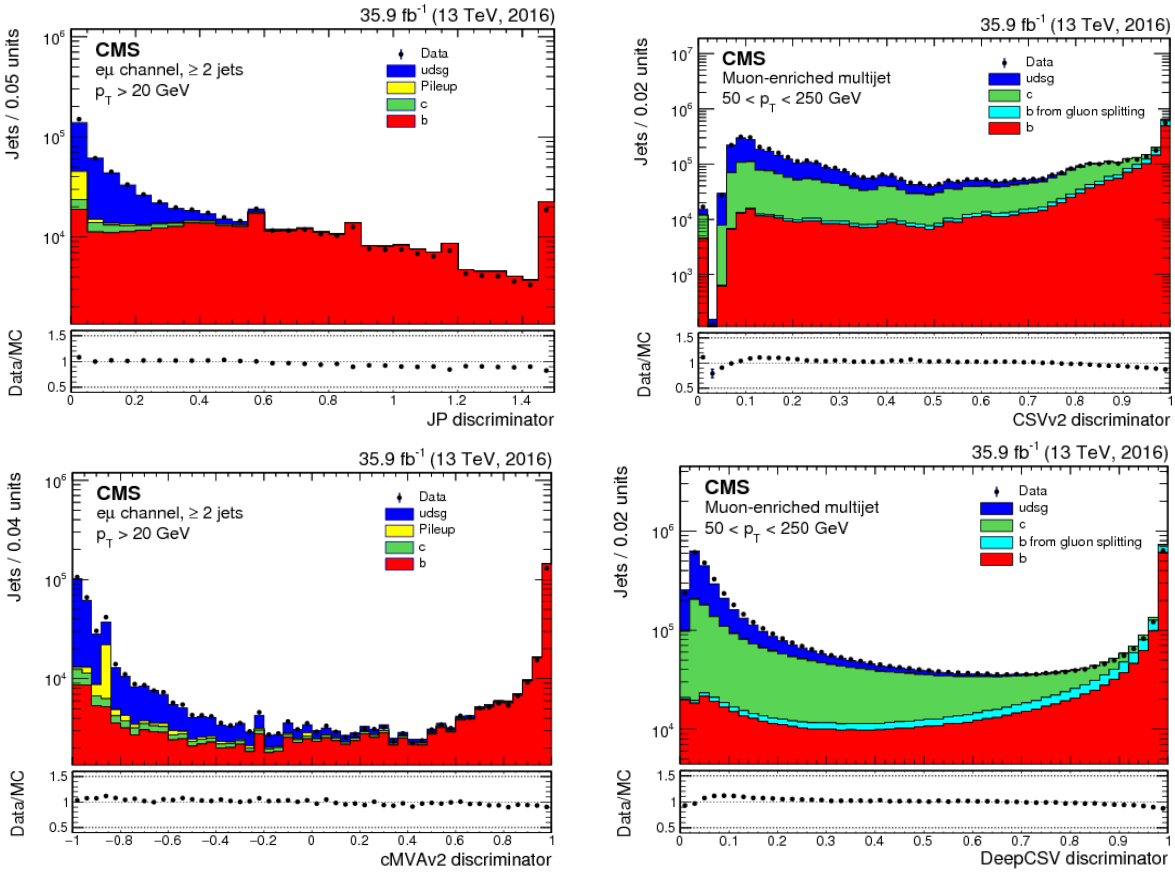


Figure 4.9: Data to simulation agreement for the JP (upper left), the CSVv2 (upper right), cMVAv2 (bottom left) and DeepCSV (bottom right) discriminators in a multijet sample, enriched in soft muons. The first and last bin of each histogram contain the underflow and overflow entries, respectively.

These deviations can cause a small discrepancy in tagging performances between data and simulated samples. In order to overcome these deviations scale factors need to be applied to simulated events. The scale factors are measured as a function of p_T and η, and the discriminator value of a jet of flavour f according to the following equation:

$$SF_f(\eta, p_T, disc.) = \frac{\epsilon_f^{data}(\eta, p_T, disc.)}{\epsilon_f^{sim}(\eta, p_T, disc.)} \tag{4.7}$$

The flavour of jets in simulations is determined from the matched generated hadrons. In data they 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 measure and applied in different ways. For analyses in which the full distribution of the b-tagging

discriminator values is used, a technique of iterative fitting (IterativeFit) is used to measure data-to-simulation scale factors [133]. If the full discriminator distribution is used, the distribution using jets in simulated events has to be corrected to match the one observed in data. scale factors are applied to calibrate the simulation to data.
The scale factors are derived from events with two oppositely charged leptons within the tracker acceptance. They have to satisfy the tight identification and isolation requirements. For the events with same-flavour leptons, the di-lepton invariant mass is required to be away from the Z-boson mass, i.e. $|M_{ll} - M_Z| > 10\,\text{GeV}$. These requirements are applied to reduce the contribution from Z + jets events and to enhance the dominating di-lepton $t\bar{t}$ 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-flavour scale factor measurements. The contamination of non-b quark jets is estimated from simulation and is subtracted from the measurement. Alternatively to enrich the sample in light-flavour jets, a control region is defined to select Z+jets events. Selecting Z+jets events is achieved by inverting the $t\bar{t}$ selection requirements. Requiring one of the two jets not to 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. The scale factors are determined here as well iteratively (hence IterativeFit). The iterative procedure measures the scale factors when no scale factors are applied to the simulations used for the contamination estimation first. Then 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. By the end, the scale factors are obtained from following equations:

$$SF^{bflavour}(\eta, p_T, disc.) = \frac{N_{data}^{bflavour} - N_{sim}^{bflavour}}{N_{sim}^{bflavour}}, \tag{4.8a}$$

$$SF^{dusgflavour}(\eta, p_T, disc.) = \frac{N_{data}^{dusgflavour} - N_{sim}^{dusgflavour}}{N_{sim}^{bflavour}} \tag{4.8b}$$

4.7 Missing energy reconstruction

Missing transverse momentum (MET) is the imbalance in the transverse momentum of all visible particles, particles which interact via the electromagnetic or strong forces, in the final state of the collisions. Because momentum is conserved in each direction, the MET is the transverse momentum that must have been carried out by something invisible, for example neutrinos. Therefore MET is an estimate of the transverse momentum of neutrinos. Since a momentum balance along the beam axis cannot be used, because the longitudinal momentum of the centre-of-mass system of the colliding partons is not known on an event-by-event basis. The missing transverse energy vector

in the PF event is calculated from the vector sum of the transverse momenta $\vec{p_T}$ of all reconstructed particles [134].

$$\vec{\not{E}}_{T,raw} = -\sum_{i=1}^{n} \vec{p}_{T,i} \tag{4.9}$$

where n is number of reconstructed particles. Since the energy of jets are corrected, as explained in Section 4.5, these corrections have to be taken in account so that the uncorrected transverse momentum of each PF jet with $p_T > 10$ GeV is replaced by the corrected transverse momentum $\vec{p}_{T,j}^{corr}$, so the corrected MET is given by:

$$\vec{\not{E}}_T = -\sum_{i=1}^{n} \vec{p_{T,i}} - \sum_{j=1}^{n_{jets}} (\vec{p}_{T,j}^{corr} - \vec{p}_{T,j}) \tag{4.10}$$

"Do not go where the path may lead.
Go instead where there is no path and leave a trial"

Ralph Waldo Emerson

Chapter 5

The FCNC analysis strategy and techniques

The typical physics objects, which are the outcome of the FCNC interactions described in Section 3.3.1 are electrons, muons, missing tranverse momentum due to a neutrino from the leptonic decay of the W bosons and jets of which some of them originate from b-quarks. Some of the SM processes are expected to have a similar final-state topology as the signal samples and are considered as background in the FCNC analysis. The aim of a typical analysis in particle physics is to reject these background processes as much as achievable in order to increase the signal significance. This can be achieved by a simple *cut and count* approach in which some specific cuts on the kinematics of physics objects are applied on an event-by-event basis. More discrimination between signal and background can be obtained with the help of Machine Learning (ML) techniques.

The Analysis Strategy

As described in Section 3.3 the CMS Monte Carlo Generator group produces a huge number of simulated samples. During 2016 the CMS collaboration collected proton collisions data at 13 TeV with a total recorded integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ as shown in Figure 2.2. This represents $\sim 92\%$ of the luminosity delivered by the LHC machine during the same period [135]. The datasets used in our search FCNC interactions of a top quark with two same-sign leptons in the final state, are DoubleMuon, DoubleEG and MuonEG (in 03Feb2017 MINIAOD format).

The strategy of this analysis is divided into four main sequential steps which are outlined below.

1. *Event Pre-selection and Cleaning*: This first level of the event selection criteria is applied on physics objects and some filters are applied to remove instrumental noise and beam backgrounds. This will be explained in Section 5.1.

2. *Baseline selection*: At this stage the basic cut and count approach is applied in order to remove backgrounds and to reconstruct some kinematics to be used for the next step. This will be discussed with more details in Section 5.2.

3. *Discrimination and optimization*: At this level a MVA technique is used in order to have more discrimination between signal and background processes. The MVA tool will be discussed in Section 5.3 and outputs will be illustrated in Section 6.1.

4. *Exclusion Limits*: The so-called Higgs combined tool is used to estimate the signal strength or to set limits on the branching ratio of FCNC couplings. The tool used to calculate exclusion limits will be described in Section 5.3.2, and the results will be presented in section 6.3.

The main analysis anatomy (from step 2 to 4) after pre-selection and cleaning, is illustrated in the flowchart of Figure 5.1. Remember that in this analysis we search for FCNC interactions of a top quarks with the Higgs boson and a u or c quark, hence denoted Hut and Hct respectively. The same strategy is applied for both and our selection is $t\bar{t}$-oriented. In this chapter the output from step one and two will be presented. Only the basis and the techniques that are used for steps three and four will be explained in this chapter while the output and results will be overviewed in next chapter.

5.1 Event pre-selections and cleaning

Before applying the baseline event selection, collections of physics objects have to be identified and subjected to some filters to be ready for the main selection cuts. The main objects are given in table 5.1.

Collection	p_{T} (GeV)	$\lvert\eta\rvert$	**Notes**
Muons	> 12	< 2.1	The same cuts are applied for loose and tight muons as explained in Section 4.3.1
Electrons	> 15	< 2.5	The same cuts are applied for loose and tight electrons as explained at Section 4.4.2
Jets	> 25	< 2.4	All jets have to pass the tight-ID as explained Section 4.5.1

Table 5.1: The table shows the initial criteria for physics objects that will be used in the physics analysis. Loose and tight muons (electrons) are the collections that pass the criteria for loose and tight WP for muon (electrons) respectively.

The first requirement to accept the event is the presence of at least one well reconstructed primary vertex described in Section 4.1. The reconstructed vertex with the

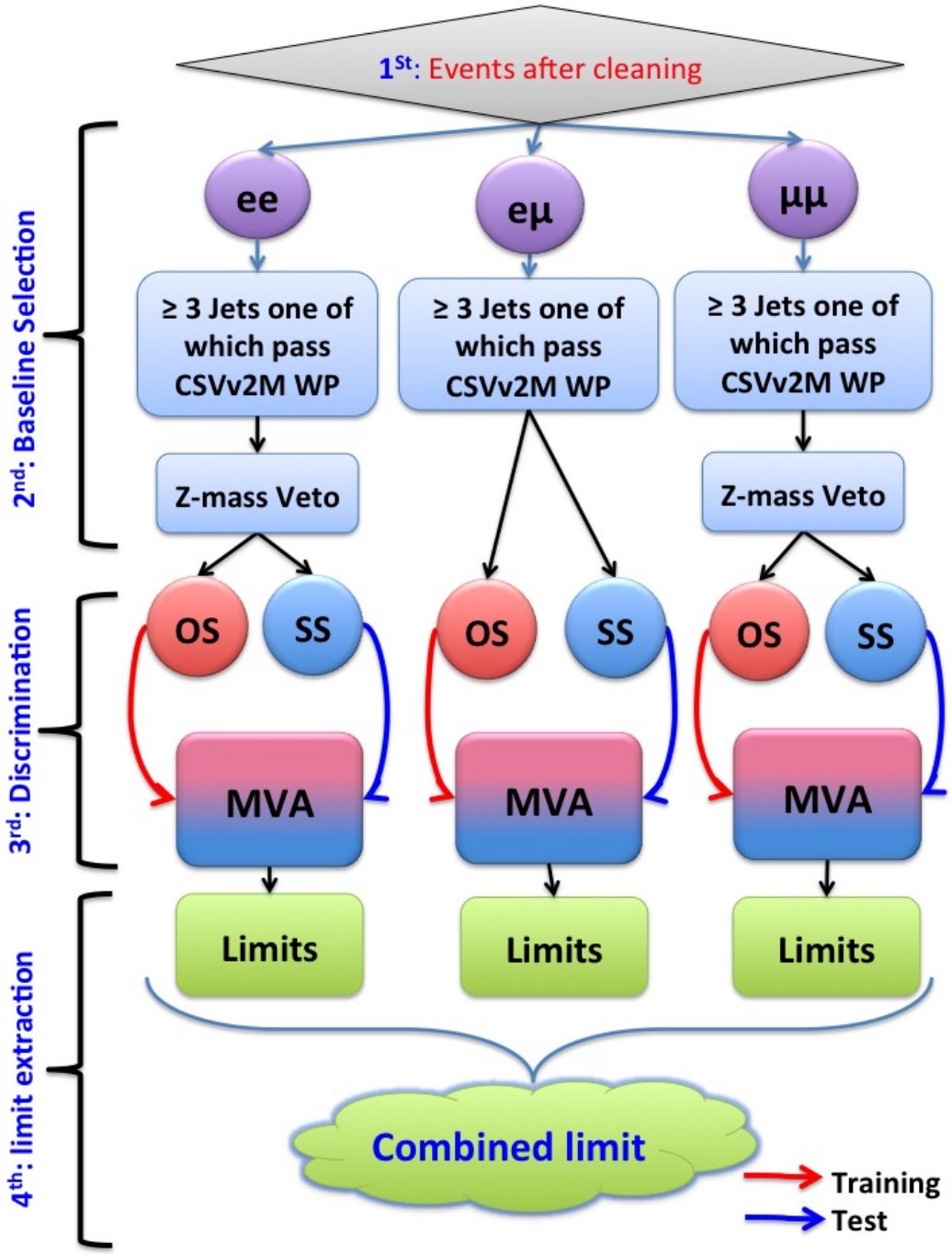


Figure 5.1: The cleaned events are divided into three di-lepton channels (ee, eμ and $\mu\mu$), then the other selection cuts are applied. An MVA technique is used for each channel separately where for the background samples the events with two opposite-sign (OS) leptons in the final state are used for training and those with two same-sign (SS) leptons are used for testing, while for the signal sample events with SS leptons are used for both training and testing. Finally the BDT discriminator distributions are passed to the combined limit setting tool to obtain exclusion limits.

largest value of the summed physics-object p_{T}^2 is taken to be the primary pp interaction vertex in the event. The physics objects chosen are those that have been defined using information from the tracking detector, including jets, the associated missing transverse momentum, which was taken as the negative vector sum of the p_{T} of those jets, and charged leptons [136].

The reconstructed vertex is required to be within $dz \leq 24\,\mathrm{cm}$, the longitudinal distance from the beam spot, and $d_{xy} \leq 2\,\mathrm{cm}$, transverse distance to the beam spot. Another important step is to ensure the consistency of the electron-charge measurement. So for our analysis electrons have to pass the selective-method requirement for the charge measurement which has been described in Section 4.4.1.

5.1.1 Triggers

Most of the collision events from uninteresting processes are removed by the CMS HLT-trigger system as discussed in Section 2.2.6, and the dedicated trigger paths are defined in the trigger system to single out the events with the required detector signature for the search presented in this thesis. The trigger paths that are used in this thesis are based on the online trigger objects with at least two leptons and are summarized in Table 5.2. Furthermore to ensure the full trigger efficiency the offline p_{T} threshold for electrons and muons are chosen to be higher than that used in online triggers. For simulated samples the event is considered when it passes one of the trigger paths. However for data double counting of the same event has to be taken into account and a procedure to avoid double counting is applied in the way of vetoing in a given dataset the events that are already selected in another.

5.1.2 Filters

Although the high-precision settings in the operation of the LHC and CMS during the data taken, some of the instrumental noises and beam backgrounds still exist and cause unwanted events to infiltrate in the analysis. In order to remove such events so-called filters are used in the offline measurements.

- **Beam halo filter:** As the proton bunches circle the LHC, protons can induce reactions with the residual gas particles or the beam collimators and as a result many secondary particles are produced and form a halo around the beam. Among these secondary particles muons are considered to be the most threatening as they have long lifetimes and are the most penetrating, meaning they are capable of traversing the entire longitudinal dimension (parallel-to-beam) of the CMS detector as can be seen in Figure 5.2 [138].

- **Bad muon filter:** In some events there are muons that are not reconstructed as PF muons. These muons are labeled as bad muons that can be misinterpreted as charged hadrons, which in turn distort the missing transverse energy calculation. Such events can be filtered by removing these bad muons which is achieved by so-called bad muon filters.

leptons channel	**Trigger path**
Di-electrons (ee)	$HLT\_Ele23\_Ele12\_CaloIdL\_TrackIdL\_IsoVL\_DZ$
Di-muons ($\mu\mu$)	$HLT\_Mu17\_TrkIsoVVL\_Mu8\_TrkIsoVVL\_DZ\|\|$ $HLT\_Mu17\_TrkIsoVVL\_TkMu8\_TrkIsoVVL\_DZ$
Electron-muon (eμ)	$HLT\_Mu23\_TrkIsoVVL\_Ele12\_CaloIdL\_TrackIdL\_IsoVL\|\|$ $HLT\_Mu23\_TrkIsoVVL\_Ele12\_CaloIdL\_TrackIdL\_IsoVL\_DZ\|\|$ $HLT\_Mu8\_TrkIsoVVL\_Ele23\_CaloIdL\_TrackIdL\_IsoVL\|\|$ $HLT\_Mu8\_TrkIsoVVL\_Ele23\_CaloIdL\_TrackIdL\_IsoVL\_DZ\|\|$ $HLT\_Mu12\_TrkIsoVVL\_Ele23\_CaloIdL\_TrackIdL\_IsoVL\|\|$ $HLT\_Mu12\_TrkIsoVVL\_Ele23\_CaloIdL\_TrackIdL\_IsoVL\_DZ$

Table 5.2: The table shows the trigger paths recommended by the top quark group [137] for the di-leptons HLT triggers used for data and simulation.

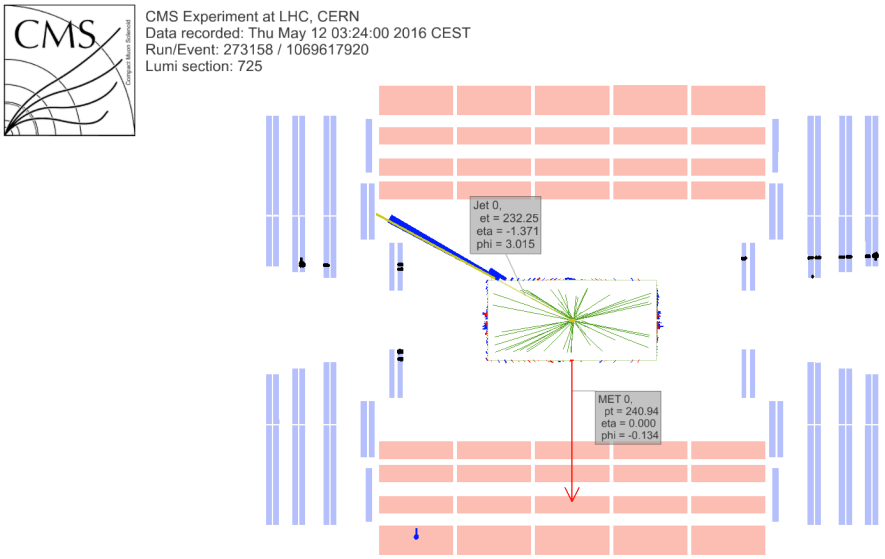


Figure 5.2: The figure illustrates an event display for a beam halo event with collinear hits in the CSC (black), $\not{E}_T = 240\,\text{GeV}$ and a jet with transverse energy $= 232\,\text{GeV}$. The hadronic deposit is spread in η but narrow in Φ. The figure is taken from [139].

- **HBHE noise filter:** HBHE noise filters are algorithms based on the pulse shape, the timing and the topology of HCAL digis/rechits in HB and HE channels described in Section 2.2.4. A HCAL Digi represents the signal in one readout channel, and consists of ten coded integers, each representing the charge deposited

in a 25 ns time bin. The timing is tuned to maximize the amount of charge in the fifth and sixth time bins. The aim to eliminate anomalous signals due to rare electronics noise activity.

In this physics analysis all these filters are applied on an event-by-event basis.

5.1.3 Jet and lepton cleaning

The charged leptons (electrons and muons) produced in decays of heavy particles, such as W and Z bosons are called *"prompt leptons"*. They are spatially isolated from the hadronic activity in the event, while leptons produced in hadron decays or from photon conversions, as well as hadrons misidentified as leptons are known as *"nonprompt leptons"* and usually embedded in jets. The main isolation is done by the PF algorithm variable I_{rel} as given by Equations 4.1 and 4.2 for muons and electrons respectively. Extra isolation is required between leptons and jets to reduce the fake rate. Technically we call this extra isolation "cleaning". In our analysis we do cleaning for both leptons and jets with the recipe taken from a SUSY analysis [140] which summarized below. In this thesis we define Lepton cleaning process as removing jets that are misidentified as leptons "fake leptons". Jet cleaning is removing leptons within a specific cone size around the jet.

- Leptons cleaning: For each lepton (electron/muon),

  1. Define a variable p_{T}^{ratio} as the ratio of the lepton p_{T} to that of closest jet within a distance of $\Delta R = 0.4$.

  $$p_{\mathrm{T}}^{ratio} = \frac{p_{\mathrm{T}}(l)}{p_{\mathrm{T}}(jet)}. \tag{5.1}$$

  This variable is set to one if there is no jet within this distance.

  2. Define p_{T}^{rel} as the transverse momentum of the lepton relative to the residual momentum of the closest jet after lepton momentum subtraction. It is calculated by the following equation:

  $$p_{\mathrm{T}}^{rel} = \frac{|(\vec{p}(jet) - \vec{p}(lep)) \times \vec{p}(lep)|}{|\vec{p}(jet) - \vec{p}(lep)|}. \tag{5.2}$$

  The p_{T}^{rel} variable allows the identification of leptons that accidentally overlap with jets.

  3. A lepton is considered to be isolated if it satisfies the following condition:

  $$I_{rel} < I_1 \quad and \quad (p_T^{ratio} > I_2 \quad or \quad p_{\mathrm{T}}^{rel} > I_3). \tag{5.3}$$

  The values of I_i, with $i = 1, 2, 3$ depend on the lepton flavour. Since the probability to misidentify a lepton is higher for electrons, tighter isolation values are used in this case. The values of isolation for electrons and muons are given in Table 5.3 .

- Jet-cleaning: This step is done as an extra cleaning by removing any leptons existing in a cone with radius $R = 0.4$ around the jet axis.

Isolation	Electrons	Muons
I_1	0.12	0.16
I_2	0.80	0.76
I_3	0.72	0.72

Table 5.3: The table shows the values of the isolation parameters used for electrons and muons.

5.2 Baseline selections

As mentioned before the analysis discussed in this dissertation is focused on FCNC involving a top quark and a Higgs boson. The signal is composed of two sectors, as described in Section 3.3.1, where $t\bar{t} \rightarrow Wb + Hq(q = u, c) \rightarrow l\nu b + q\ l\nu QQ$ and $tH \rightarrow Wb + H \rightarrow l\nu b + l\nu\ QQ$, where $H \rightarrow W^+W^-$. For both the top quark pair and single top quark signals the SM top-quark decays leptonically. Since for the single top signal the (u/c) quark is in the initial state while for $t\bar{t}$ signal appears in final state, the final state for the signals has the following objects: two same-sign leptons with their associated neutrinos + one b-quark jet + one (u- or c-) quark jet (for $t\bar{t}$ signal) + two light-quark jets. In searching for same-sign di-leptons in the final state of proton collisions we have many challenges which will be discussed in the following sections.

5.2.1 Lepton selection

In our analysis the events are divided into three channels depending on the flavour of leptons in the final state. The three channels are: di-electron (ee), di-muon ($\mu\mu$) and electron-muon pair ($e\mu$). The two leptons must satisfy the tight WP criteria. Therefore veto is applied for extra leptons that pass the loose WP [1]. The event is rejected if it contains more than two leptons in the final state. In case of same flavour leptons additional requirements on the invariant mass are applied.

- **Di-electrons** (ee)

  Events with two electrons are selected if the electrons have $p_{\mathrm{T}} \geq 25, 15\,\mathrm{GeV}$ for the leading and the second leading electron respectively. The invariant mass of the two electrons should be $m_{ee} \geq 12\,\mathrm{GeV}$ in order to reject events coming from low resonance masses.

- **Di-muons** ($\mu\mu$)

  Events with two muons are selected if the muons have $p_{\mathrm{T}} \geq 20, 15\,\mathrm{GeV}$ for the leading and the second leading muon respectively. Again here the invariant mass

[1] The leptons that pass the loose WP criteria not necessary pass the tight WP criteria.

of the two muons is required to be $m_{\mu\mu} \geq 12\,\text{GeV}$ in order to reject events coming from low resonance masses.

- **Electron-muon ($e\mu$)**

  In this channel there are two cases for the leading and the second leading lepton. If $p_{\text{T}}(electron) > p_{\text{T}}(muon)$, then $p_{\text{T}} \geq 25, 15\,\text{GeV}$ is chosen for the leading and the second leading lepton. If $p_{\text{T}}(muon) > p_{\text{T}}(electron)$, then $p_{\text{T}} \geq 25, 20\,\text{GeV}$ is chosen for the leading and the second leading lepton.

5.2.1.1 The electron charge mis-identification background

One source of backgrounds in the ee and $e\mu$ channels comes from the mis-identification of the electron charge. This background arises from processes with prompt opposite-sign electrons like DY+jets where the charge of one electron is wrongly measured. In this physics analysis a data driven method is used for the estimation of the probability of mis-identifying electron charge (i.e. the probability is extracted from data). Studies like [141] show that the probability of charge mis-idetification for muons is so small that it can be ignored.

The probability of charge mis-identification for electrons is measured from the control region around the mass of the Z-boson ($m_Z = 91.19\,\text{GeV}$) with the recipe below.

- Select events with two electrons that pass the $|m_Z - m_{ee}| \leq 10\,\text{GeV}$ requirement. As illustrated in Figure 5.3 the region between the blue dashed lines represent the peak in the mass distribution of the Z boson. Electrons in this region come from real opposite-sign charge electron-pairs coming from DY+jets background. Consequently if the electron-pair has the same electric charge it means that the charge of one of these electrons is mis-identified.

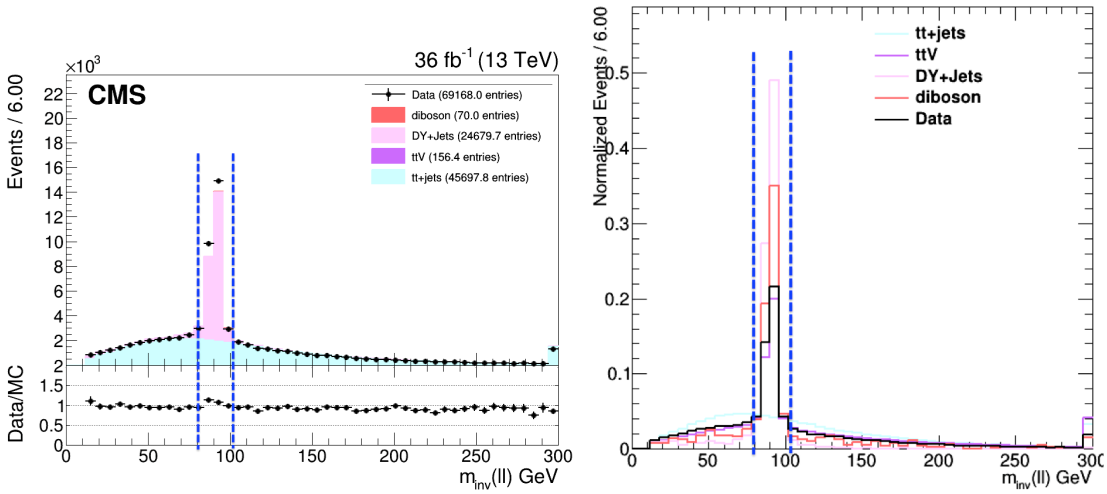


Figure 5.3: The events in the area between the blue dashed lines are considered in the Z boson mass window and are used for the measurement of the charge mis-identification. The left figure shows the data/MC agreement while the right figure illustrates the difference between the normalized distribution of different SM backgrounds.

- If we assume that each electron has a probability $p_{misId}(q_e)$ to have its charge mis-identified then the number of events N_{SS} with a same-sign electron pair from the Z boson decays is given by

$$N_{SS} = 2 \times p_{misId}(q_e)(1 - p_{misId}(q_e)) \times N_{tot} \tag{5.4}$$

  where N_{tot} is the total number of events from Z boson decays. In this case the number of events N_{OS} with opposite-sign electron pairs from Z boson decays is given by

$$N_{OS} = (p^2_{misId}(q_e) + (1 - p_{misId}(q_e))^2) \times N_{tot} \tag{5.5}$$

  and in case $p_{misId}(q_e) << 1$ this can be reduced to

$$p_{misId}(q_e) = \frac{N_{SS}}{2N_{OS}} \tag{5.6}$$

- The events are divided into six (p_{T}, η) regions. In both barrel $(0 \leq |\eta| < 1.479)$ and the endcap $(1.479 \leq |\eta| < 2.5)$ regions the events are divided according to electron's p_{T} like $p_{\mathrm{T}} \leq 25, 25 < p_{\mathrm{T}} \leq 50$ and $p_{\mathrm{T}} \geq 50\,\mathrm{GeV}$ ending with 6 different regions for the electron pairs. In each region $p_{misId}(q_e)$ is measured using Equation 5.6 . The contribution from the WZ simulated sample (which contains real electrons) in the OS and SS regions is subtracted.

The measured mis-identification probabilities in data (which are dominated by DY+jets events) and in the simulated DY+jets sample are given in Table 5.4. A good agreement between data and simulation is obtained [2].

		Mis-identification probabilities %		
	η	$p_{\mathrm{T}} \leq 25\,\mathrm{GeV}$	$25 < p_{\mathrm{T}} \leq 50\,\mathrm{GeV}$	$p_{\mathrm{T}} \geq 50\,\mathrm{GeV}$
Data	**Barrel**	0.029	0.021	0.029
	Endcap	0.057	0.1	0.23
DY+jets	**Barrel**	0.03	0.025	0.037
	Endcap	0.059	0.1	0.25

Table 5.4: The table shows the values of the electron charge mis-identification probabilities for data and simulated DY+jets.

In order to estimate the background from the mis-identification of the electron charge, for the ee and the $e\mu$ channels we apply the calculated probabilities to the events

[2]These results have been presented and discussed in the EGamma POG of the CMS collaboration meeting and they were approved by experts.

after the same selection cuts [3] as for the signal region but with opposite-sign electrons. Since we apply a veto on the mass of the Z-boson and because the probabilities in different regions are very small, we did not get any event from this background.

As described in Section 4.4.1 there are two, majority and selective, methods that can be used for the estimation of the electron charge. Therefore another check is to determine the mis-identification probabilities from data in case of applying each method. The results are shown in Table 5.5. These measured values clearly confirm that using the selective method for the estimation of the electron charge in this analysis is better for rejection the background coming from electron-charge mis-identification.

		Mis-identification probabilities %		
	η	$p_{\mathrm{T}} \leq 25\,\mathrm{GeV}$	$25 < p_{\mathrm{T}} \leq 50\,\mathrm{GeV}$	$p_{\mathrm{T}} \geq 50\,\mathrm{GeV}$
Selective method	**Barrel**	0.029	0.021	0.029
	Endcap	0.057	0.1	0.23
Majority method	**Barrel**	0.28	0.21	0.36
	Endcap	0.42	0.73	0.97

Table 5.5: The table shows the values of the electron charge mis-identification probabilities in data in case of the selective and the majority method.

5.2.2 Jet selection criteria

The selected events must have at least three jets in the final state with $p_{\mathrm{T}} \geq 30\,\mathrm{GeV}$ for the leading, the 2^{nd} and the 3^{rd} leading jets. These criteria are defined to include the events from both the single-top quark and top quark pair signal process. Since FCNC signals have one b quark in the final state one of these three jets should be coming from the hadronization of a b-quark. In this analysis the identification of b-flavour jets is performed using the CSVv2 algorithm described at Section 4.6. Exactly one (out of the three jets) is required to pass the criteria of the medium WP of the CSVv2 discriminator which is previously illustrated in Table 4.7.

5.2.3 The signal and control regions

One of main backgrounds in the searches in di-lepton final states comes from the DY-process ($pp \rightarrow Z/\gamma \rightarrow ll$). In order to remove this background a veto is applied on the appearance of a lepton pair in the event in a window around the Z boson mass. In

[3] The application step is done after all baseline selection cuts.

this way the event is rejected if it contains two leptons with an invariant mass within $[81.1, 101.1]$ GeV.

As described in Section 3.3.1 in the signal production only the leptonic decay is considered for SM top quark. Because we look for events with two same-sign leptons in the final state. This can be obtained only if the W boson from the H boson decay with the same charge as the W from SM top decays leptonically.

Therefore only 50% of FCNC signals will have two same-sign leptons in the final state while the rest of the FCNC signal events will have two opposite-sign leptons in the final state. In this analysis we do not consider these FCNC opposite-sign lepton events, and focus on the same-sign lepton signature.

In this physics analysis we define two main regions:

- **SSSR**: This represents the signal region (SR) where the events passed all baseline selection cuts and have two leptons with same-sign (SS) electric charge.

- **OSCR**: The control region (CR) where the events passed all baseline selection cuts but have two leptons with opposite-sign (OS) electric charge.

From now on we will use the terms SSSR and OSCR to refer to the signal and control regions respectively.

The reconstruction of different kinematic variables in both SSSR and OSCR shows that the shape of the distributions of these variables are very similar, as can be seen from Figures 5.4. Hence we can benefit from this behavior when we want to further discriminate between signal and backgrounds. This will be done using an MVA technique as will be explained in Section 5.3. Together with the good agreement between data and simulation this motivates us to use OSCR events for the training step in the MVA technique.

5.2.4 Corrections and scaling factors

Some parameters used in the simulation of collision processes are not precisely known. Even though these parameters are tuned to get the best matching between the simulation and the observed data, some residual discrepancies remain when comparing data with simulation. In order to obtain a better agreement, some corrections or *scale factors* (SF) are applied to the simulated samples. Most of these SFs depend on the p_{T} and η values of the reconstructed particles involved and may depend on the jet flavour as well. These SFs are determined using large datasets and large simulated samples, they are applied typically event-by-event on the simulated events. The following corrections or scaling factors are considered in the physics analysis of this thesis:

- **Luminosity reweighing**
  The simulated samples are normalized corresponding to the total recorded integrated luminosity of the data (described in Section 2.1.1). In this analysis the integrated luminosity adds up to $\mathcal{L}_{int} = 35.9\,\mathrm{fb}^{-1}$. For each simulated sample a SF is calculated as:
  $$SF_{lumi-weight} = \frac{\sigma_{process} \times \mathcal{L}_{int}}{N_{events}} \tag{5.7}$$

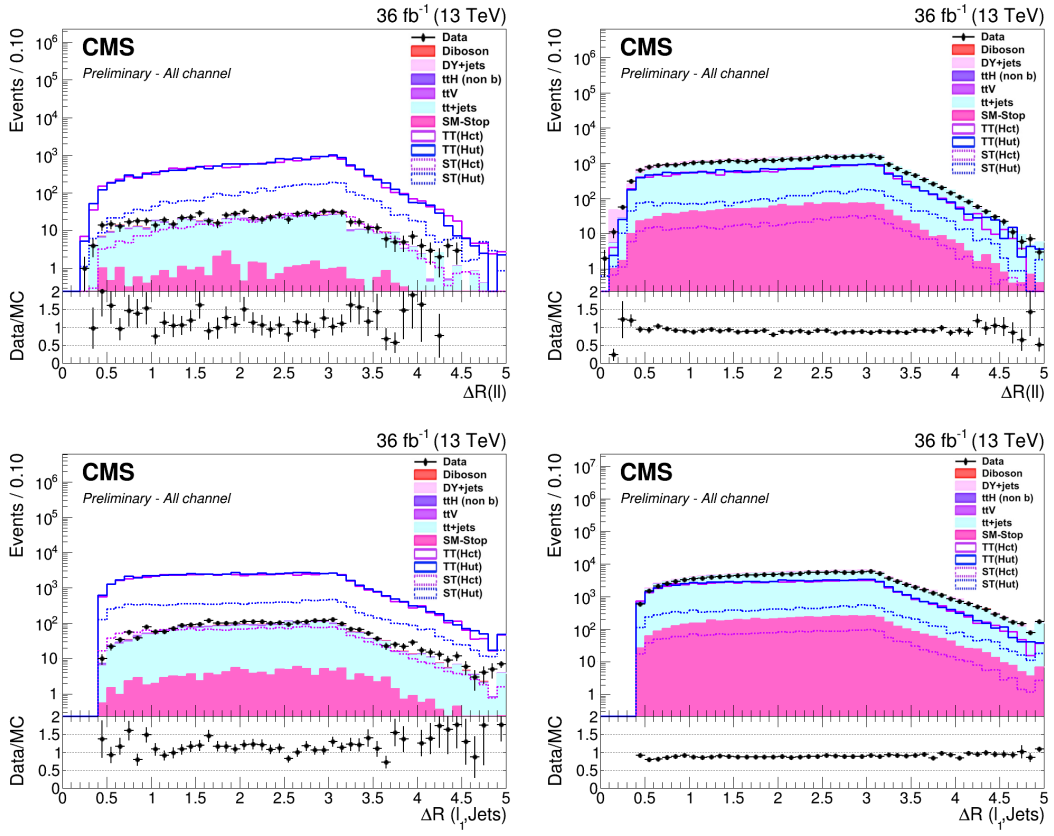


Figure 5.4: The plots at the top represent the distribution of the ΔR variable between two leptons in the SSSR (top left) and in the OSCR (top right) regions. The plots at the bottom represents the distribution of the ΔR between 2^{nd} leading lepton and all the jets in the events in the SSSR (bottom left) and in OSCR (bottom right) regions.

where $\sigma_{process}$ is the cross-section of a process involved and N_{events} is the number of events in the simulated sample corresponding to this process.

In order to simulate the differential cross section correctly, The samples that are generated by `MadGraph5_aMC@NLO` are scaled using the negative event weight, W_{neg}^{event}, defined by Equation:

$$W_{neg}^{event} = \frac{N_{total}}{N_{eff}} = \frac{N_{pos} + N_{neg}}{N_{pos} - N_{neg}} \tag{5.8}$$

where N_{eff} is the effective number of events that are produced by the generator and N_{total} the total number of events produced. N_{pos} is number of events with positive event weights and N_{neg} is number of events with negative event weights.

- **Pileup Reweighing**
  The number of PU interactions in simulated events N_{PU}^{Sim} varies from that in the real pp collisions collected of the CMS experiment. In simulation the number of PU is described by a simple distribution. However in real collisions the number of PU interactions N_{PU}^{Data} depends on the instantaneous luminosity. For Run 2 the

minimum bias inelastic cross-section of 69.2 mb is measured [142, 143]. Therefore the simulated samples have to be corrected by reweighting them such that the distribution of the number of reconstructed primary vertices agrees with that of the data.

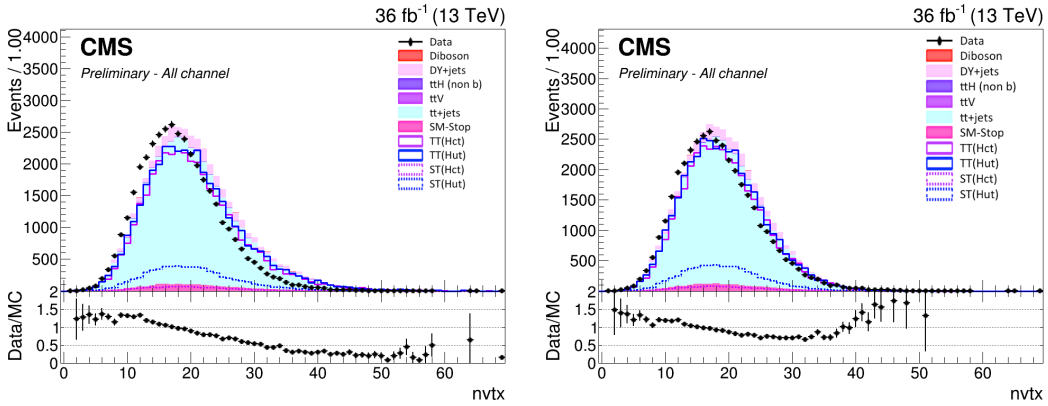


Figure 5.5: The figure shows the number of reconstructed vertices before (left) and after (right) applying the correction factors on the simulation.

The PU reweighing manifests itself in the distribution of number of reconstructed vertices as can be seen in Figure 5.5. As can be seen from the figure although pileup reweighting is applied, an observable mismatching between data and simulation still exist. This is known for POGs and will be considered as a systematic uncertainty.

- **Lepton SF**
  The selection efficiency for leptons in simulation differs from that in data. This difference is related to the fact that different parameters are used for the lepton selection as described in Section 4.4 and 4.3 for electrons and muons respectively. In order to overcome this mismatch the simulated events are weighted by what is called the lepton scaling factors SF_{lep} given by

$$SF_{lep} = \frac{\epsilon_{Data}}{\epsilon_{Sim}} \tag{5.9}$$

  where ϵ_{Data} and ϵ_{Sim} are the selection efficiencies in data and simulation respectively. These SFs are calculated for each electron and muon in the event separately based on their p_{T} and η as recommended by the EGM [144] and Muon [145] POGs. Moreover SFs are obtained for the identification, isolation, tracking and trigger efficiencies and the total SF per event is a multiplication of these four SFs as given by

$$SF_{lep}^{total} = \prod_{i}^{lep} SF_{ID}(p_{Ti}, \eta_i) SF_{Iso}(p_{Ti}, \eta_i) SF_{track}(p_{Ti}, \eta_i) SF_{trig}(p_{Ti}, \eta_i). \tag{5.10}$$

  The effect of the lepton SF on lepton variables (electrons and muons) can be seen in Figure 5.6 .

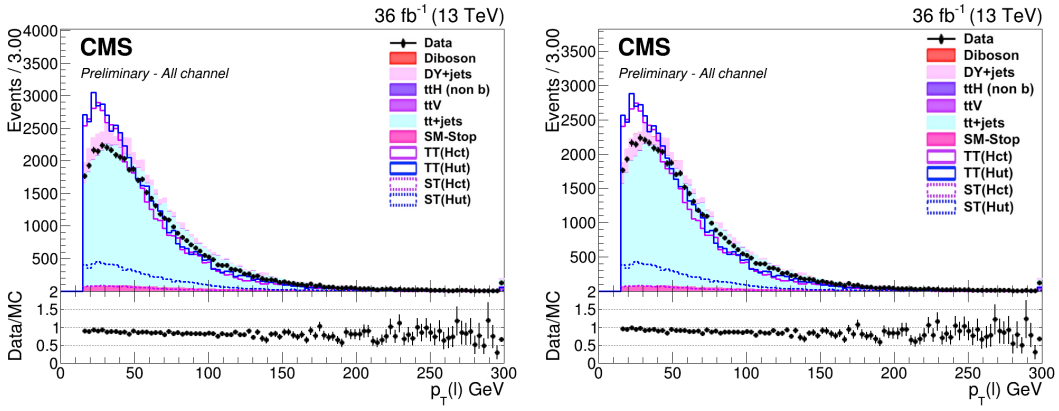


Figure 5.6: The figures show the p_{T} distribution for all leptons (electrons and muons) before (left) and after (right) applying SF_{lep}^{total}.

- **CSVv2 shape SF**
  Even though only one jet is required to be tagged as a b-quark jet, b-tagging plays an important role in our analysis. This is because some information of b-tagged jet will be used as an input variable for the MVA to separate signal from backgrounds. Since the CSVv2 discriminant shown in Figure 5.7 is used for tagging jets as b-quark jets, the shape of this discriminant has to be corrected to remove the discrepancies between data and simulation. This is achieved by applying b-tag SFs discussed in Section 4.6 on a jet by jet basis. Therefore the total b-tag discriminant shape SF is given by Equation 5.11 as the product of individual (b, c and dusg)-quark jet's SFs

$$SF_{CSVv2} = \prod_{i}^{N_b} SF_i^b \quad \prod_{j}^{N_c} SF_j^c \quad \prod_{k}^{N_{dusg}} SF_k^{dusg} \tag{5.11}$$

  where N_b, N_c and N_{dusg} are the number of b-, c- and light-quark jets in the event respectively.

  The effect of the b-tag SF on the discriminator distribution of the highest p_{T} jet used in this physics analysis can be seen in Figure 5.8

- **Jet energy corrections**
  For both data and simulations, the jet energy corrections (JEC) are applied. These JECs are taken from a pre-measured database as described in Section 4.5.2.

5.2.5 Event yields

The event yields in the SSSR region for each di-lepton channel are given in Table 5.6. The main contribution is coming from the SM $t\bar{t}$+jets background. The event yields in the ee and $e\mu$ channels are affected by the tight selection cut on the electron charge. The event yields of FCNC signal samples are corresponding to the branching ratio and cross-sections used for the production described in Section 3.3.1. One can notice that

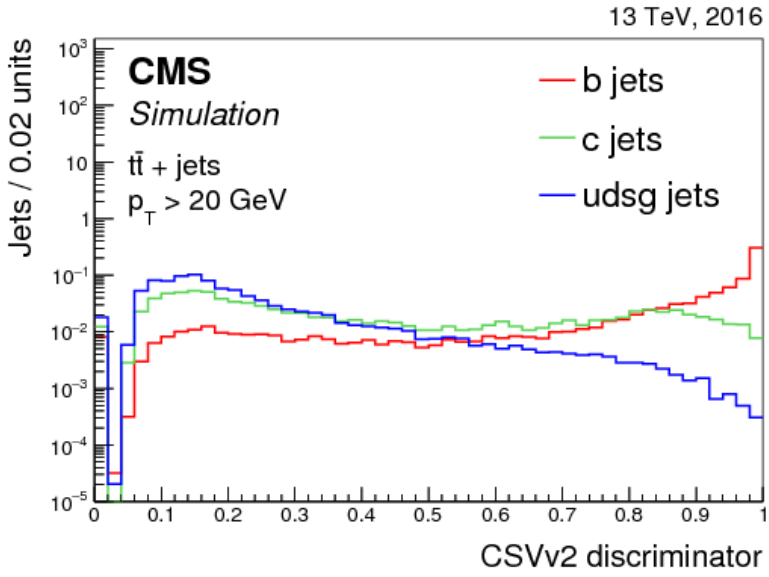


Figure 5.7: The figure shows the distribution of the CSVv2 discriminator for b-, c- and dusg (*light*)-quarks, the plot is taken from [127].

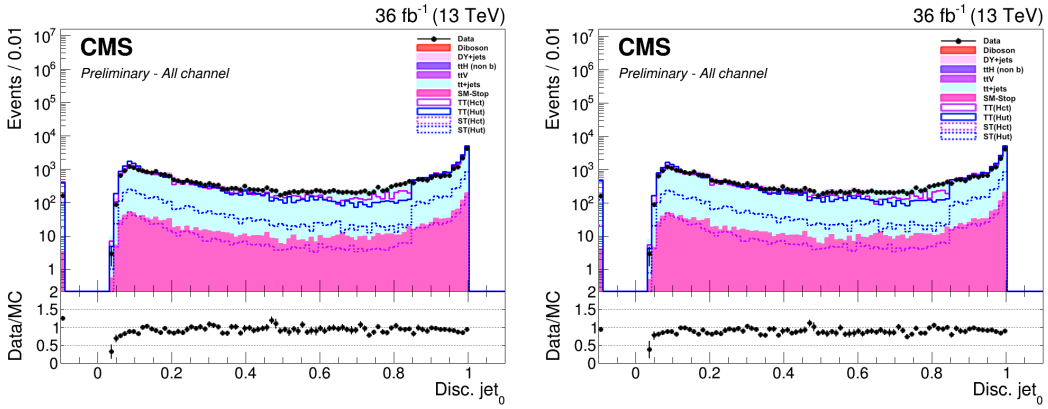


Figure 5.8: The figure shows the CSVv2 b-discriminator for the highest p_{T} jet before (left) and after (right) applying SF_{CSVv2}. The events at zero reflect the underflow from a default value saved as b-tag discriminator.

for the FCNC signals the contribution from single top quark Hut-coupling is higher than that in Hct-coupling. This is due to the proton composition that contain u-quark more than c-quark. Furthermore the FCNC signals for both Hut and Hct couplings are dominated by the top quark pair decay.

5.3 Discrimination and optimization

The multivariate analysis (MVA) method is one of the most popular machine learning techniques that are used in particle physics research. The machine learning is the paradigm for automated learning from data, using computing algorithms. The primary

Process	**ee** channel	$\mu\mu$ channel	$e\mu$ channel
$t\bar{t}$+jets	70 ± 5	357 ± 12	118 ± 7
DY+jets	4.2 ± 4.2	–	–
$t\bar{t}$V	6.0 ± 0.1	10.5 ± 0.1	7.84 ± 0.1
$t\bar{t}$H	4.7 ± 0.1	9.9 ± 0.2	6.5 ± 0.1
Diboson	1.9 ± 0.4	4.5 ± 0.6	2.9 ± 0.5
SM S-top	2.5 ± 1.6	19.3 ± 3.1	8.1 ± 2.9
FCNC-ST (Hut)	68.5 ± 2.5	140 ± 3.6	107 ± 3.2
FCNC-TT (Hut)	354 ± 9	835 ± 14	573 ± 12
FCNC-ST (Hct)	12.2 ± 0.4	25.3 ± 0.6	19 ± 0.5
FCNC-TT (Hct)	360 ± 9	790 ± 14	546 ± 11
total Bkg.	109 ± 7	401 ± 12	143± 7.
Data	102 ± 10	458 ± 21	168 ± 13

Table 5.6: The table shows the expected event yields for the SM background processes and the FCNC signal processes as well as the selected yield in the Run 2 (2016 data) after the baseline selection cuts in the SSSR region.

goal of learning from previous data is to respond correctly to future data. The MVA technique is used to achieve a better discrimination power between the signal and the backgrounds with respect to a simpler analysis based on individual selection criteria (cut and count) or poorly discriminating variables. Many software tools exist to apply an MVA technique in a physics analyses. For the physics analysis in this thesis, the Tool for Multivariate Analysis (TMVA) is used [146]. TMVA is an open source project included into ROOT [147]. The Boosted Decision Trees (BDT) method is employed for the classification of events as implemented in the TMVA framework [148]. In this section we provide an overview of the BDT classifier and how we use it for selection optimization.

5.3.1 The Boosted Decision Trees (BDT)

Decision Trees are an important type of algorithm for predictive modeling machine learning. The output of each decision tree is what is called a classifier or discriminator. The boosting algorithm is a procedure that combines many weaker classifiers and gives a final more powerful classifier or discriminator [149, 150].

First each sample (signal and background) is divided into two parts. The first part, *the training sample*, is used to train the decision trees. The second part, *the test sample*, is used to test the final classifier (discriminator) after the training. For each event in the sample there are some variables $var_1, var_2, var_3, ...var_n$ used as input variables, which are used to distinguish between signal (*Sig.*) and background (*Bkg.*) [151]. The determination of the final discriminator is done by a sequential growing of binary trees as illustrated in Figure 5.9 and explained as follows.

1. Starting from an initial number of sample events at a *node*, for each input variable the events are ordered depending on the value of the variable.

2. Picking one variable say, var_1,for each event, the algorithm finds the value (d) of the variable that gives the best separation into one side having mostly signal and the other side mostly background, each side is called a *branch*.

3. The algorithm continues with repeating step 2 for each input variable in turn, i.e. selecting a variable and finding the splitting value which gives the best separation.

4. Considering each branch as a new node the algorithm repeats steps 2 and 3, and keeps repeating it until a given number of final branches, *leaves*, are obtained, each leaf is pure signal or pure background, or has too few events to continue.

The decision of which branch will define the next node depends on the purity P of the separation. Of course the one which gives the highest purity is chosen to be the next node to split. The purity at each node can be calculated as

$$P = \frac{\#Sig.}{\#Sig. + \#Bkg.} \tag{5.12}$$

where $Sig.$ and $Bkg.$ are the number of the signal and background events respectively after the branching at the node. In the physics analysis of this thesis the criterion used

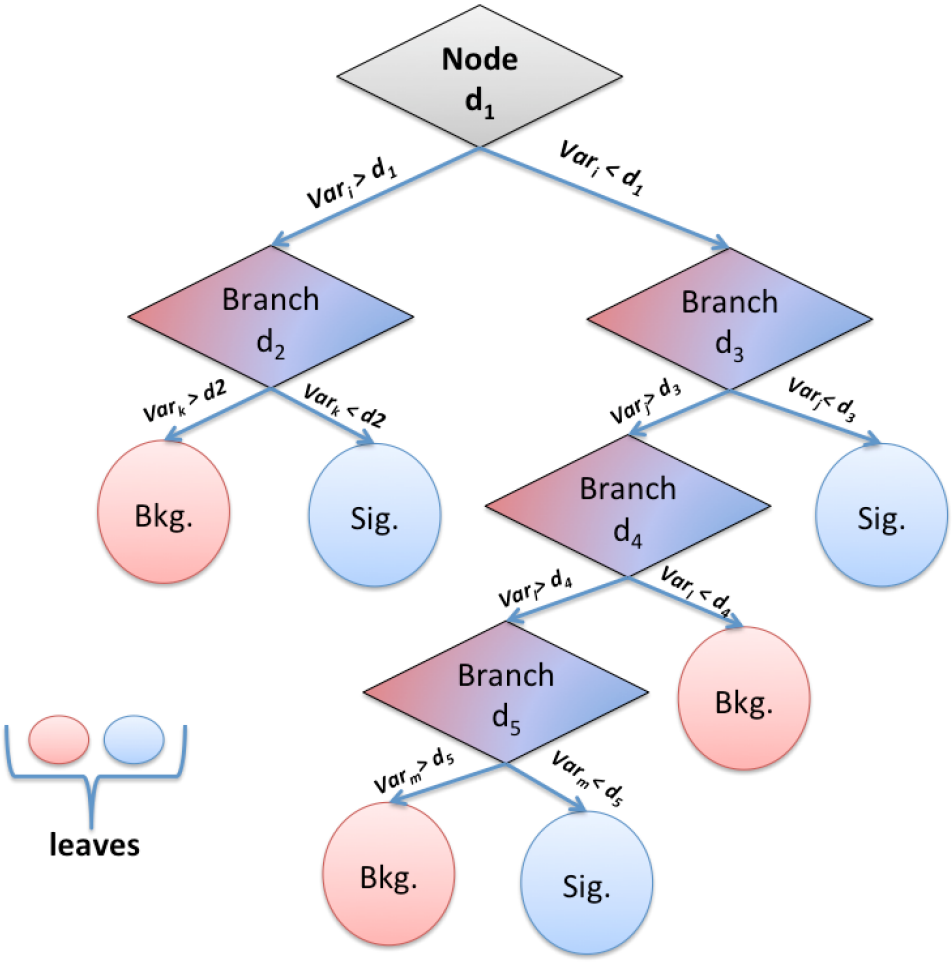


Figure 5.9: Schematic view of the decision tree. Starting from the node a sequence of binary splits based on the value of (d) into secondary nodes (branches) using the discriminating (input) variables $var_{i,j,k,l,m}$. the leaf nodes at the bottom end the decision tree.

for the decision is based on what is called the *Gini*-index which is defined as

$$Gini = \sum_{i}^{n} W_i \, P(1-P) \tag{5.13}$$

where n is the number of events at that branch, and W_i is the weight given to each event. If the sample is pure signal or pure background $P(1-P) = 0$. The highest purity is obtained and accordingly the most discriminant variable is chosen when the difference between the Gini index of the parent node and the sum of Gini indices of the two daughter nodes reach a maximal value. When the decision tree reaches a predefined depth, the learning stops. The training sample is reweighted such that the decision tree starts with a maximal Gini index, and at this case S = B (or equivalently P = 0.5). If a leaf has a purity greater than 0.5 then it is called a signal leaf and if it is less it is considered as a background leaf.

Events are classified as signal or background according to the leave where they landed on. If a training event is mis-classified, i.e. a signal event lands on a background leaf or the reverse then the weight of that event is increased (boosted). Thereafter a second tree is built using the new weights. Hence repeating the procedure considering only mis-classified events with their weights boosted. Several decision trees can be built and then combined into a forest where the final output BDT discriminant is determined by the majority vote of all trees. This stabilizes the decision trees against statistical fluctuations. In this physics analysis, the adaptive boosting (AdaBoost) method [150] is chosen. The final BDT discriminant is ranging between -1 and 1, where the events with discriminant value near 1 are considered as signal-like events and those near to -1 are considered as background-like events. The performance of the BDT training is given in the form of Receiver Operating Characteristic (ROC) curves. The ROC curve illustrates the background rejection versus the signal efficiency [146].

5.3.2 The limit setting tool

In particle physics in order to claim the existence or absence of a new physics phenomenon a *statistical test* is performed involving two hypotheses. Clearly the *null*-hypothesis is tested against the *alternative*-hypothesis. In the physics analysis of thesis the statistical method implemented in the so-called "Higgs combine tool" is used [152]. This tool is developed by the ATLAS and CMS collaborations specifically for the benefit of the LHC Higgs Combination working group (LHC-HCG) [153, 154]. It provides a command line interface for many different statistical techniques available inside RooFit/RooStats that are used widely inside CMS [155].

For purposes of excluding a signal process (i.e. setting limits) the two hypotheses are defined as:

- **Null hypothesis** H_0: The model which describes all processes including both the background and the new physics signal. This is often referred to as the signal-plus-background (S+B) hypothesis.

- **Alternative hypothesis** H_1: The model which describes the background processes only and so is referred to as the background-only (B) hypothesis.

The two hypotheses can be generalized by introducing a signal strength μ which acts as a multiplicative factor to the signal yield.

Consider a histogram of the measured observable x with a number of bins N. The expected number of events in the i^{th} bin can be calculated from:

$$E[n_i] = \mu S_i + B_i, \tag{5.14}$$

where S_i and B_i are the expected number of events of signal and background in the i^{th} bin respectively. Therefore, the B hypothesis corresponds to $\mu = 0$ whereas the (S+B) hypothesis corresponds to $\mu = 1$. S_i and B_i are calculated from:

$$S_i = S_{total} \int f_S(x, \theta_S) dx, \tag{5.15}$$

$$B_i = B_{total} \int f_B(x, \theta_B) dx, \tag{5.16}$$

where S_{total} and B_{total} are the total event yield of signal and background respectively that are predicted by simulation, while $f_S(x, \theta_S)$ and $f_B(x, \theta_B)$ are their probability density functions (pdfs) integrated over the relevant bin i in the variabe x. These pdfs are functions of variable x and a collection of *nuisance parameters* (θ). Nuisance parameters (NP) are related to the systematic uncertainties of the analysis.

In particle physics the Bayesian and/or frequentist statistical methods are widely used to estimate (approximate) confidence levels (CLs) [156].

In the physics analysis of this thesis the modified frequentist approach is used for confidence level approximation [157]. In this approach a binned likelihood is used to quote the exclusion limit. Assuming a Poisson distribution in each bin, the binned likelihood is given by the product of the Poisson likelihoods of the individual bins of the variable x

$$\mathbf{L}(\mathcal{D}|\mu, \theta) = \prod_{i=1}^{N} \frac{(\mu S_i + B_i)^{n_i}}{n_i!} e^{(\mu S_i + B_i)} . \prod_j pdf_j(\tilde{\theta}_j|\theta_j) \tag{5.17}$$

where $pdf_j(\tilde{\theta}_j|\theta_j)$ is the probability density function of nuisance parameter θ_j with nominal value $\tilde{\theta}_j$. The value of $\tilde{\theta}_j$ is an approximation of the true value of θ_j and is estimated from auxiliary measurements. $\mathcal{D}$ can be real observed data or pseudo-data. All systematic uncertainties present in this analysis are incorporated in the likelihood. They are considered as 100% uncorrelated or correlated. In case the uncertainty is partially correlated it is splitted into sub-components to fit the requirements. The morphing technique [158] is used for modeling the shape uncertainties.

In order to test the hypothesized value of μ the profiled likelihood ratio is defined as

$$\lambda(\mathcal{D}|\mu) = \frac{\mathbf{L}(\mathcal{D}|\mu, \hat{\hat{\theta}})}{\mathbf{L}(\mathcal{D}|\hat{\mu}, \hat{\theta})} \tag{5.18}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the maximum-likelihood estimators **MLE** (i.e they maximize the likelihood in the denominator). The value $\hat{\hat{\theta}}$ is the value of θ that (conditionally) maximizes the likelihood for a specified μ for a given (pseudo/real) data $\mathcal{D}$. For many analyses the signal contribution to the mean number of events is assumed to be non negative which means μ must be positive. However, the definition of $\hat{\mu}$ should be the value of μ that maximized the likelihood even if this gives a negative value for $\hat{\mu}$ if it provides positive values for the Poisson mean, $\mu S_i + B_i$. At the LHC the test statistic q_μ is defined as

$$q_\mu = -2 \ln \lambda(\mathcal{D}|\mu). \tag{5.19}$$

The distribution of q_μ is sampled for the (S+B) and B-only hypotheses by means of simulated pseudo-experiments. The calculation of q_μ can be CPU intensive so to reduce computing time the Asymptotic CL method is used. In this method replace Monte Carlo samples are replaced by one representative dataset, called the *Asimov*

dataset which is constructed such that all observed quantities are set equal to their MLE values. More details about this procedure can be found in [159].

Confidence Level (CLs) procedure

In most of the physics analysis the CLs method is used to exclude regions of phase space where CLs $< (1-\alpha)$, where α is the desired confidence level. For the real data, the probability of finding a data set with equal or greater incompatibility to the signal plus background hypothesis is determined by p-value associated to (S+B) hypothesis which is given by P_{S+B}:

$$P_{S+B} = \int_{q_{\mu,obs}}^{\infty} f_{S+B}(q_\mu|\mu, \hat{\theta}_\mu)\ dq_\mu. \tag{5.20}$$

where $q_{\mu,obs}$ is test statistic observed for real measured data, $f_{S+B}(q_\mu|\mu)$ is the probability density function describing the distribution of q_μ for the (S+B) hypothesis which is constructed by pseudo-data with nuisance parameters fixed to $\hat{\theta}_\mu$. Similarly, for the background-only hypothesis:

$$1 - P_B = \int_{q_{\mu,obs}}^{\infty} f_B(q_\mu|\mu = 0, \hat{\theta}_{\mu=0})\ dq_\mu. \tag{5.21}$$

The signal is excluded at 95% confidence level (CLs) of

$$CLs = \frac{P_{S+B}}{1 - P_B} \leq 0.05 \tag{5.22}$$

The observed limit is defined as the upper limit on μ at the 95% CLs which is obtained by adjusting μ until CLs= 0.05. The expected median upper limit and the $\pm 1\sigma$ and $\pm 2\sigma$ band are generated by a set of pseudo-data and the CLs and the value of at 95% CL for each of them. The median expected value is obtained by replacing the $q_{\mu,obs}$ with the value of the test statistic that represents the 50% quantile of the cumulative distribution function. The $\pm 1\sigma$ and $\pm 2\sigma$ bands on the expected limit are obtained when replacing $q_{\mu,obs}$ with the 68% and 95% quentiles of the cumulative distribution function respectively.

"When you become comfortable with uncertainty, infinite possibilities open up in your life."

Eckhart Tolle

Chapter 6

The search for FCNC interactions involving a top quark and a H boson

In the previous chapter the baseline selection cuts are applied in order to discriminate the FCNC signal from the SM background. However more discrimination can be obtained by using the BDT training as described in Section 5.3. In this chapter the results of the BDT training will be presented in Section 6.1. The exclusion limits on the branching ratios of the FCNC signals will be overviewed in Section 6.4.

Additional to the OSCR and SSSR regions defined in the previous chapter we use for simplification as well the ST+TT abbreviation for the combined single-top quark and top quark pair FCNC signal.

6.1 Signal to background discrimination with a BDT method

In the physics analysis of this thesis, a BDT method within the TMVA framework is trained on simulated signal and background samples in the three different channels ($ee, \mu\mu, e\mu$) separately for both couplings Hut and Hct in the way illustrated in Figure 6.1.

- **Signal sample**: This includes both FCNC signals (ST + TT) for the Hut or Hct couplings.
- **Background samples**: This includes all other simulated samples of SM background processes.

As a result we obtain six different BDT discriminants (three di-lepton channels for each coupling).

Different input variables are used for the training of the BDT each coupling. Initially we start the BDT-training process with a large number of input variables. The TMVA framework provides a list of input variables ranked according to their importance in the training. This ranked list is used to minimize the number of input variables to be used for the training. A reduced set of input variables not only saves computational time but also reduces the over-training. For each training with a reduced set of input variables we verify the performance of the resulting BDT-classifier to ensure the stability and

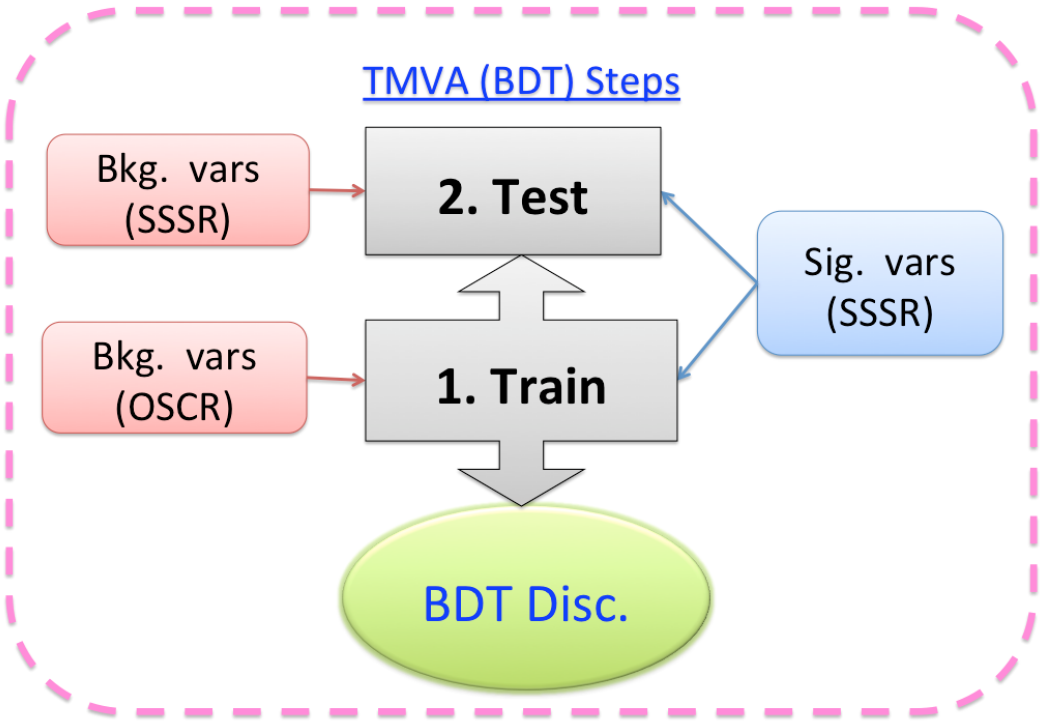


Figure 6.1: Schematic view of the BDT training steps.

the purity of the training. Finally we use only the top ranked variables that result in an acceptable performance. The input variables used for the BDT training for the Hct coupling are listed in Table 6.1. Their normalized distributions and mutual linear correlations are showed in Figures 6.2, 6.3 and 6.4. The variables used for the BDT training for the Hut coupling are listed in Table 6.2 and their distributions are showed in Figures 6.5, 6.6 and B.6. The correlations between the input variables for signal and background are almost the same which is a positive indication of good choice of discriminating variables.

The most discriminating input variable is the ΔR_{ll} between the two leptons. This should be related to the TT-component of the FCNC signal, the dominant component, which is produced back-to-back. In turn the final state leptons will be spatially separated with $\Delta R_{ll} \approx \pi$. This can be seen for the signal as a peak around this value while it is smoother for backgrounds. Smaller variations appear between the shape of signal and background distributions, but most of them are related to the spatially separation between the leptons and jets in the event. Another feature that can be noticed are the p_{T} distributions of leptons that indicate leptons from signal have lower values than those from the backgrounds. This can be understood as one lepton comes from the Higgs boson decay while the other comes directly from top-quark decay.

For this analysis, the variable distributions of the background events in the OSCR region are used for the BDT training. For the testing step background events from the SSSR region are used. For the input variables of FCNC signal events, the simulated events in the SSSR region are used for both the training and testing steps. We choose this approach for the BDT training process because as overviewed in Table 5.6 we have

too few SSSR simulated background events to perform a precise training.

In order to verify if this makes sense we perform the whole BDT process in the OSCR region, i.e. using signal and background simulated events for training and testing steps from the OSCR region. The distribution of the output BDT discriminators resulting from this training are compared with the ones from the training in the SSSR region. The distributions are illustrated in Figures 6.8 and 6.9 for the Hct and Hut coupling respectively. For the ee and $e\mu$ channels, very small shifts can be noticed between the train and testing distributions for background when the training step is done on OSCR events and the testing is done on SSSR events. This small loose in performance to separate signal and background is expected but not dramatic for the analysis.

The performance of each BDT training is given by the ROC curve shown in Figure 6.10 for the Hct coupling and in Figure 6.11 for the Hut coupling. As can be seen for the training on OSSR events and testing on SSSR events in the ee and $e\mu$ channels the performance of testing is better than of training, while for the case where both training and testing using OSCR events the performance is the same. This is due to the better discriminating power on some input variables that is shown for example ΔR_{ll} in Figure 6.12.

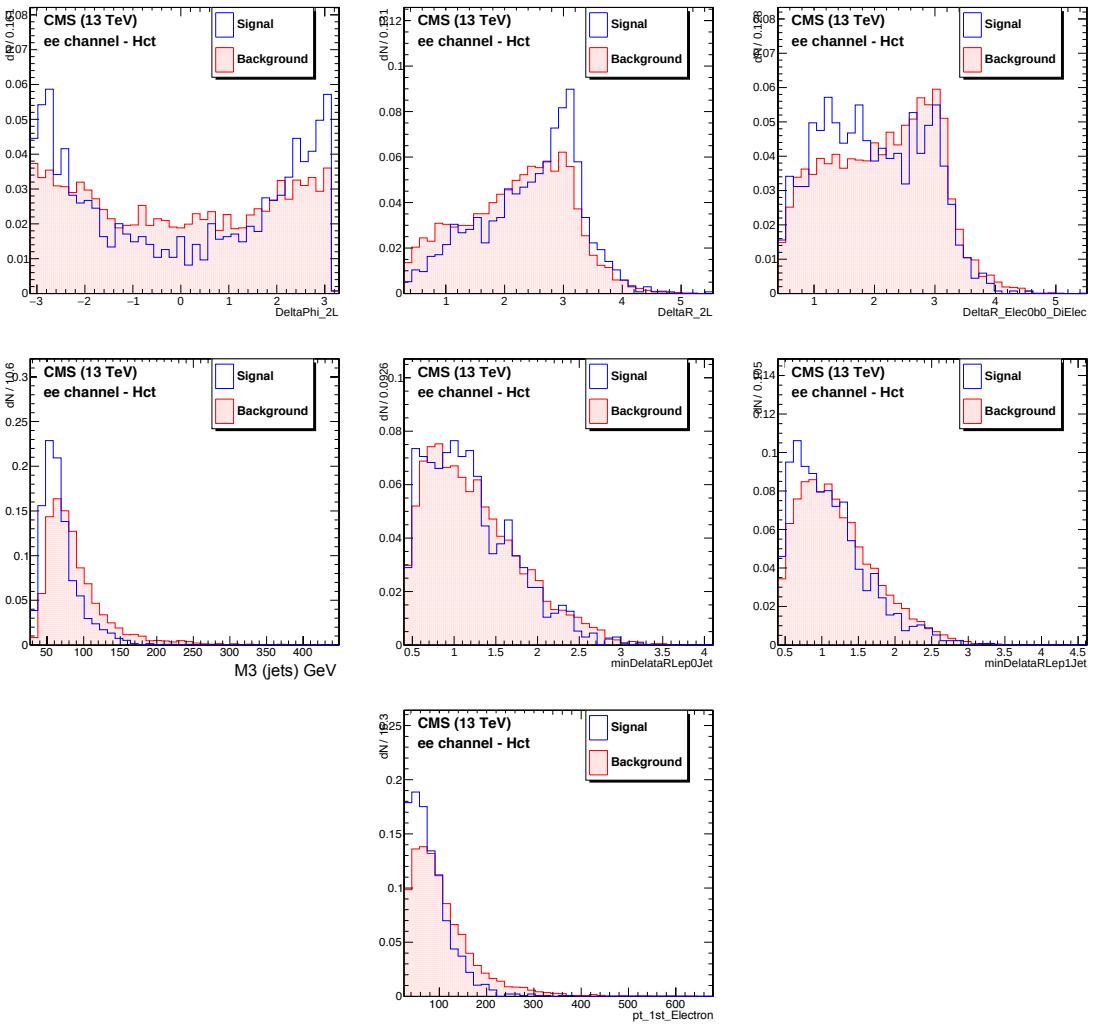
The plots of the input variables and the correlations in case the training is done in the OSCR region can be found in Appendix B.

i	**Variable**	$Definition$
1	$\Delta R(l_0, l_1)$	the $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ between the two leptons
2	$\Delta\phi(l_0, l_1)$	the difference in azimuthal angle (ϕ) between the two leptons
3	$\Delta R_{min}(l_0, jet)$	the ΔR between the leading lepton and the nearest jet
4	$\Delta R_{min}(l_1, jet)$	the ΔR between the 2^{nd} leading lepton and the nearest jet
5	$\Delta R(l_0, b)$	the ΔR between the leading lepton and the b-tagged jet
6	$m3(jets)$	the invariant mass of the three highest-p_{T} jets
7	$p_{\mathrm{T}}(l_0)$	the transverse momentum of the leading lepton

Table 6.1: The table overviews the input variables used in the BDT training for the Hct coupling. The rank of each variable in the BDT training may differ from one di-lepton channel to another.

i	**Variable**	$Definition$
1	$\Delta R(l_0, l_1)$	the ΔR between the two leptons
2	$\Delta\phi(\not{E}_T, l_1)$	the difference in azimuthal angle (ϕ) between the $\not{E}_T$ and the 2^{nd} leading lepton
3	$\Delta R_{min}(l_0, jet)$	the ΔR between the leading lepton and the nearest jet
4	$\Delta R_{min}(l_1, jet)$	the ΔR between the 2^{nd} leading lepton and the nearest jet
5	$\Delta R(l_0, b)$	the ΔR between the leading lepton and the b-tagged jet
6	$H_T(jets)$	the sum of p_{T} of all jets in the event
7	$p_{\mathrm{T}}(l_0)$	the transverse momentum of the leading lepton
8	$p_{\mathrm{T}}(l_1)$	the transverse momentum of the 2^{nd} leading lepton

Table 6.2: The table overviews the input variables used in the BDT training for the Hut coupling. The rank of each variable in the BDT training may differ from one di-lepton channel to another.



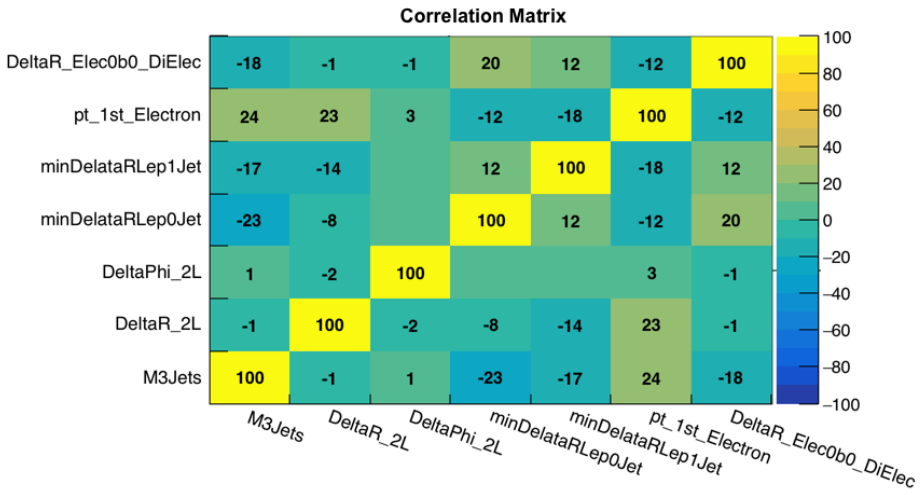
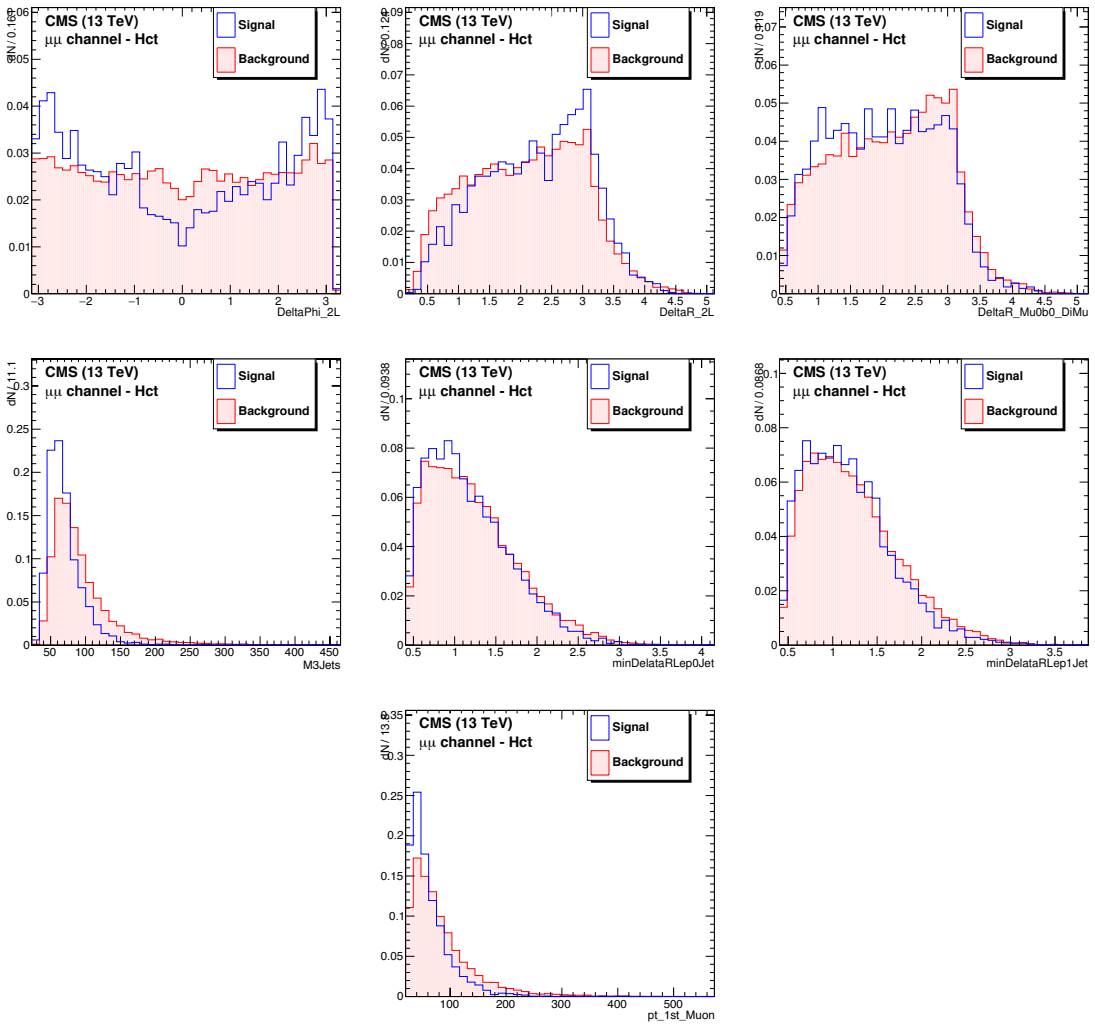


Figure 6.2: The figures show the normalized distributions of the input variables for the *ee* channel (Hct coupling). The correlation between these variables is shown (bottom) for signal events.



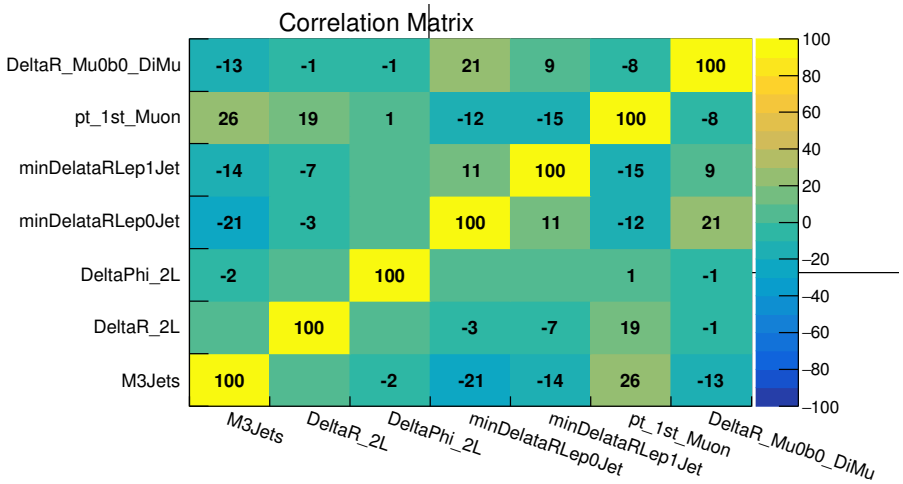
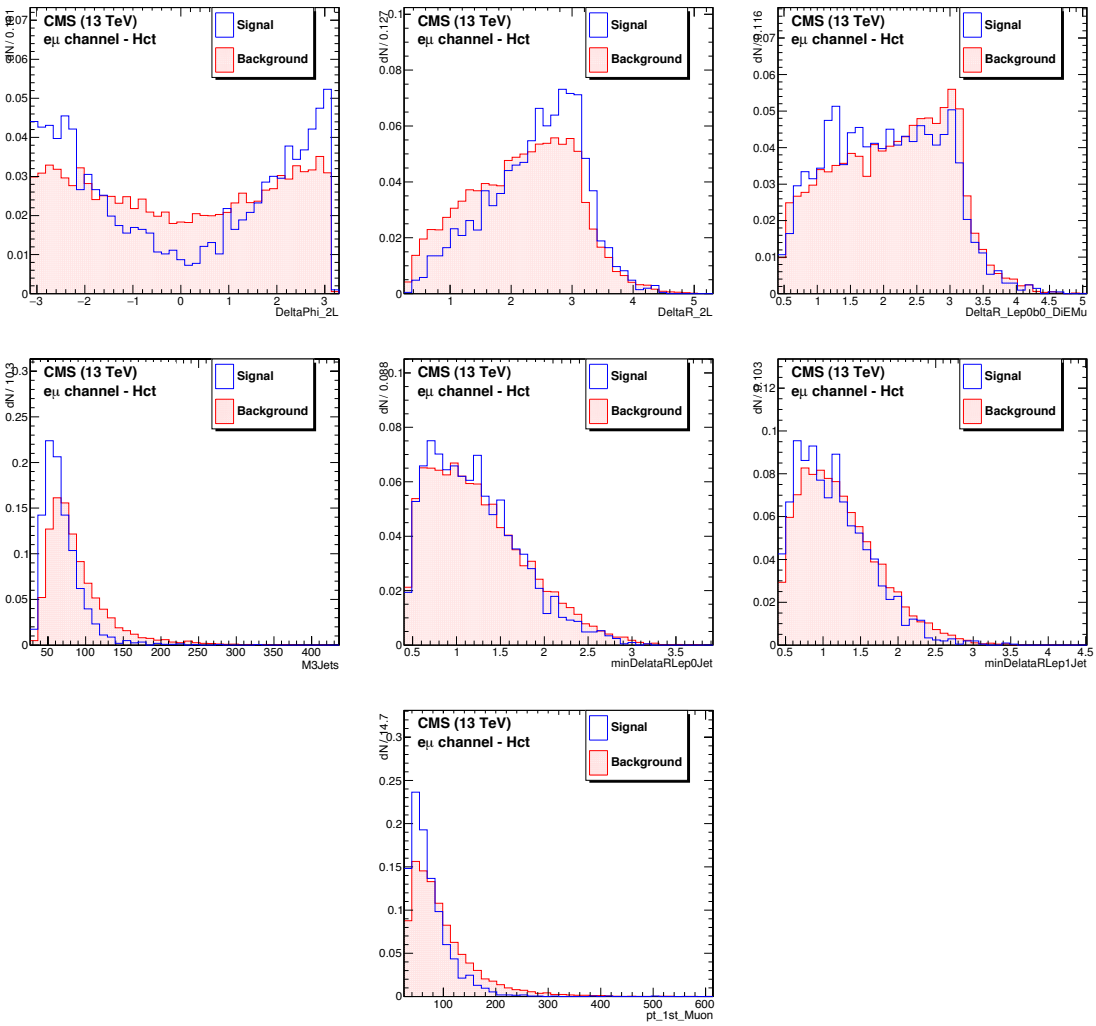


Figure 6.3: The figures show the normalized distributions of the input variables for the $\mu\mu$ channel (Hct coupling). The correlation between these variables is shown (bottom) for signal events.



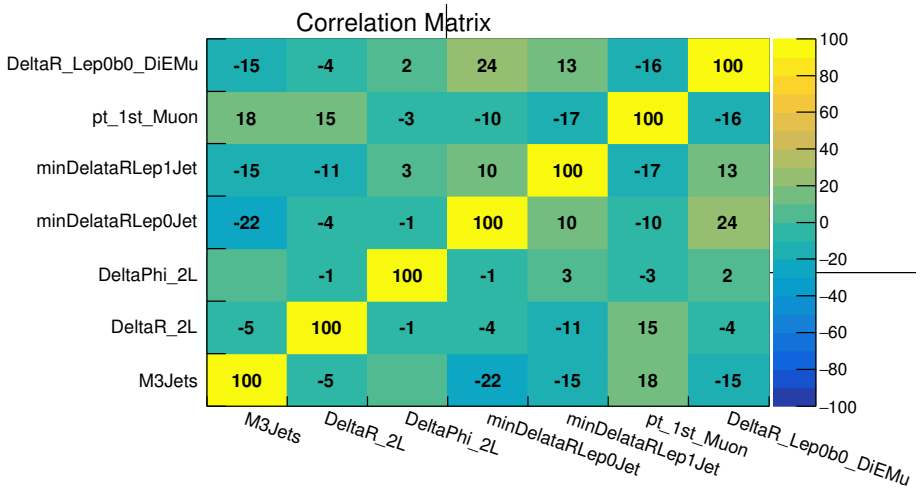
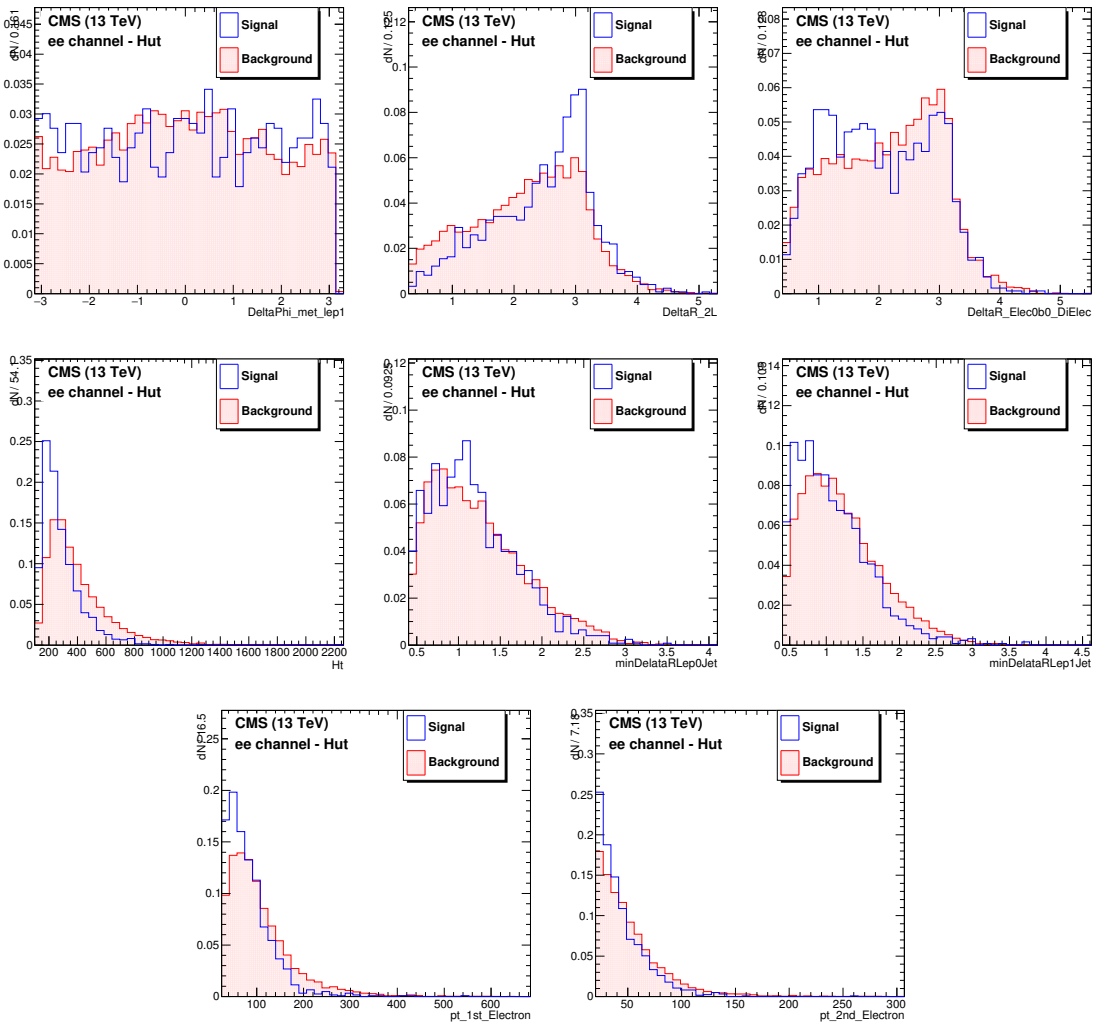


Figure 6.4: The figures show the normalized distributions of the input variables for the $e\mu$ channel (Hct coupling). The correlation between these variables is shown (bottom) for signal events.



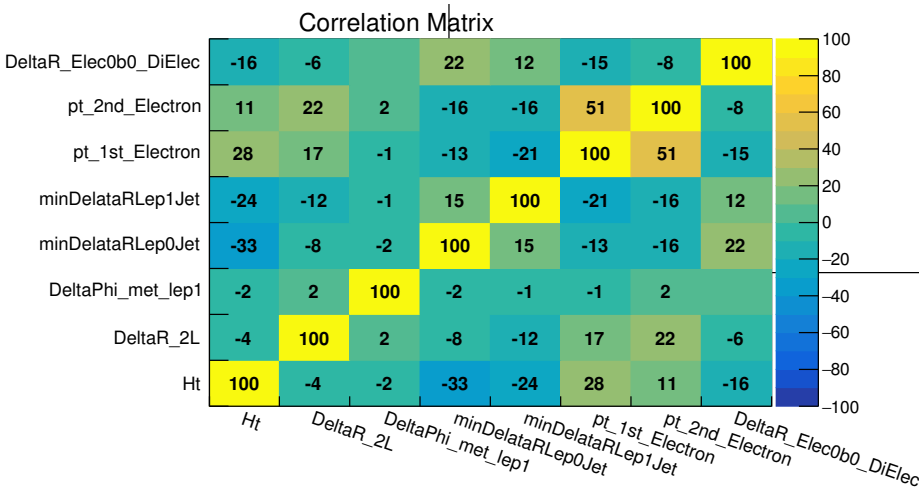


Figure 6.5: The figures show the normalized distributions of the input variables for the *ee* channel (Hut coupling). The correlation between these variables is shown (bottom) for signal events.

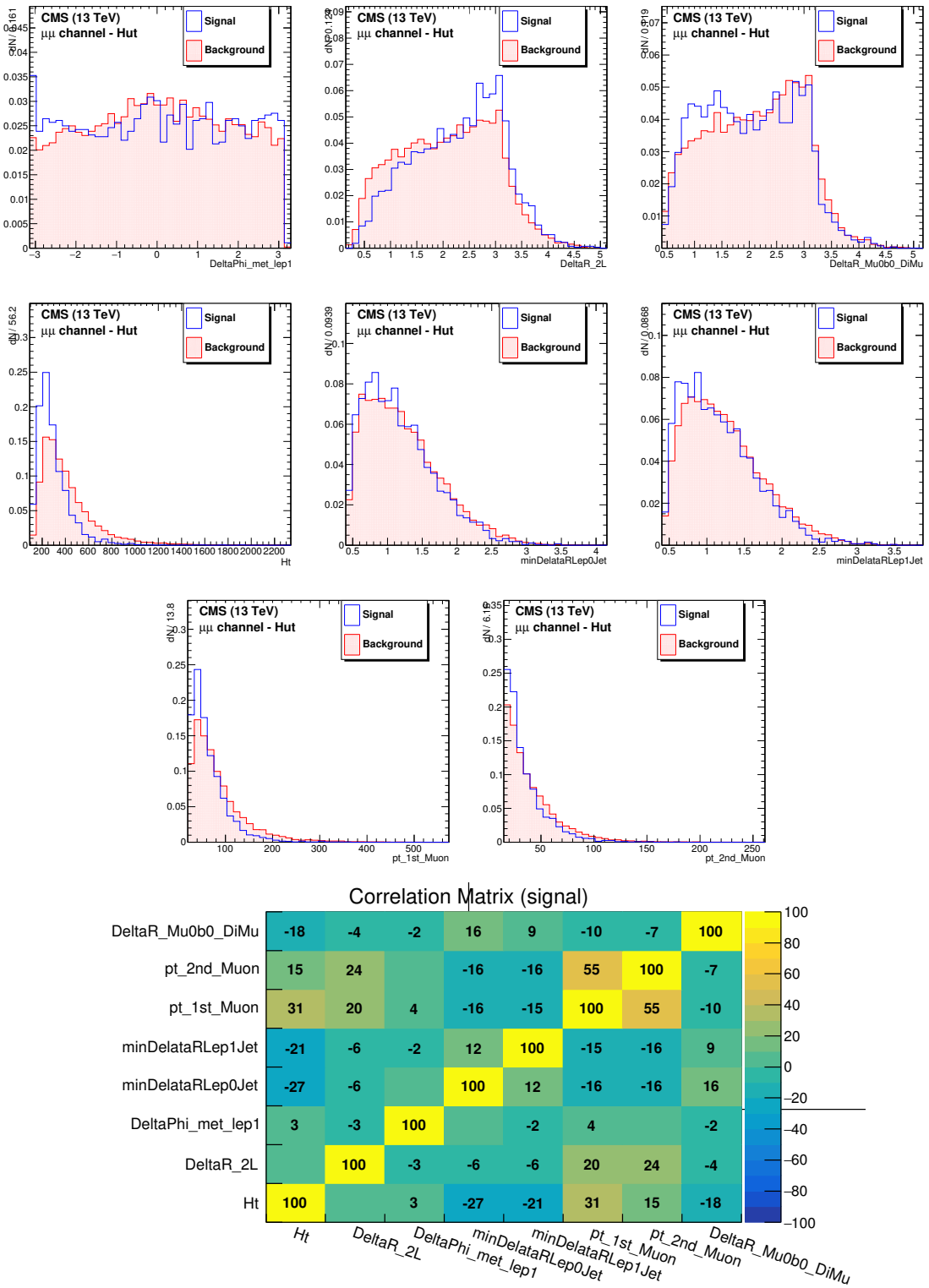
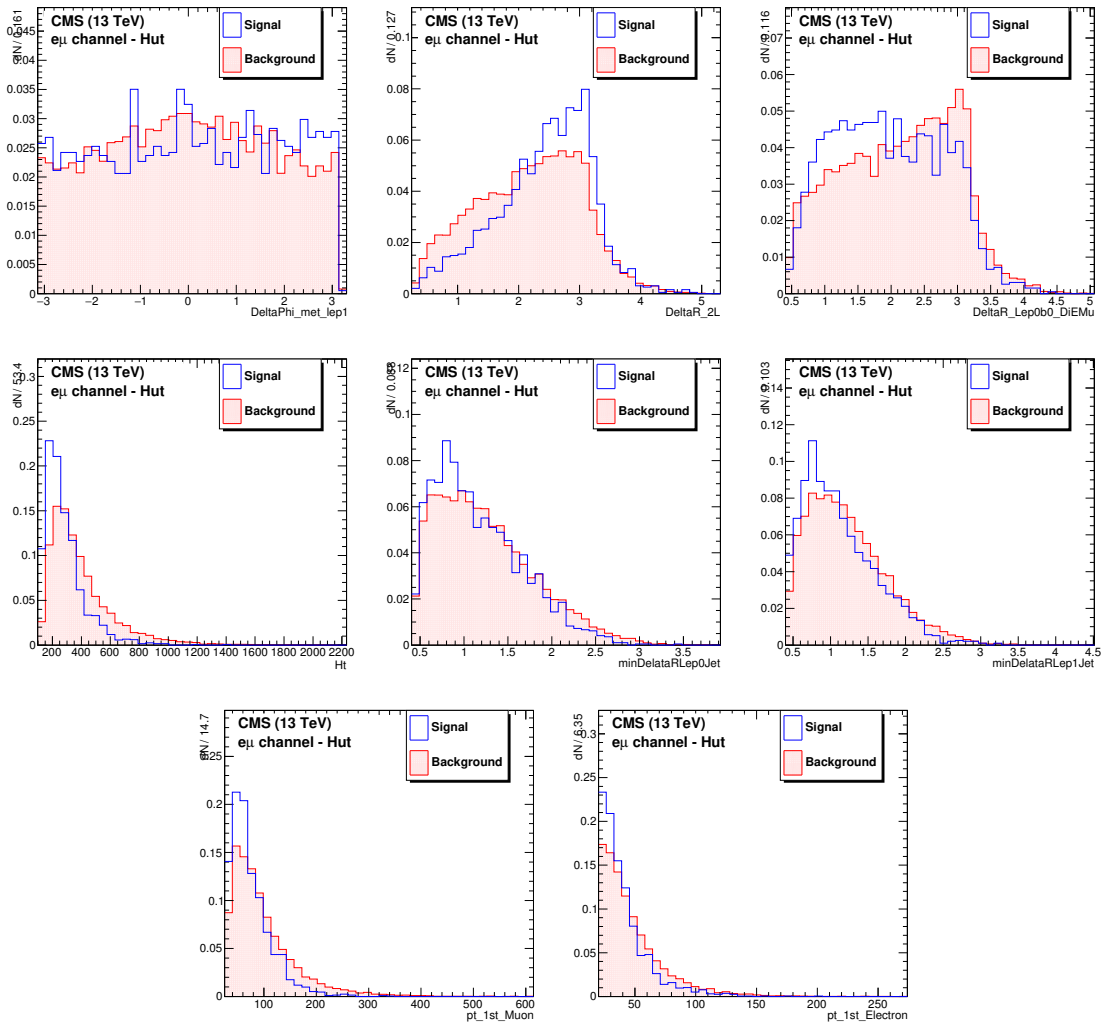


Figure 6.6: The figures show the normalized distributions of the input variables for the $\mu\mu$ channel (Hut coupling). The correlation between these variables is shown (bottom) for signal events.



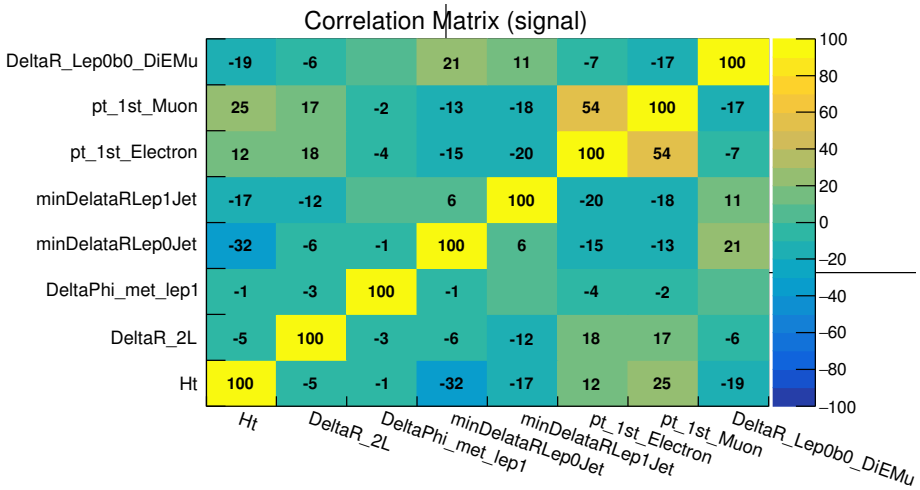


Figure 6.7: The figures show the normalized distributions of the input variables for the $e\mu$ channel (Hut coupling). The correlation between these variables is shown (bottom) for signal events.

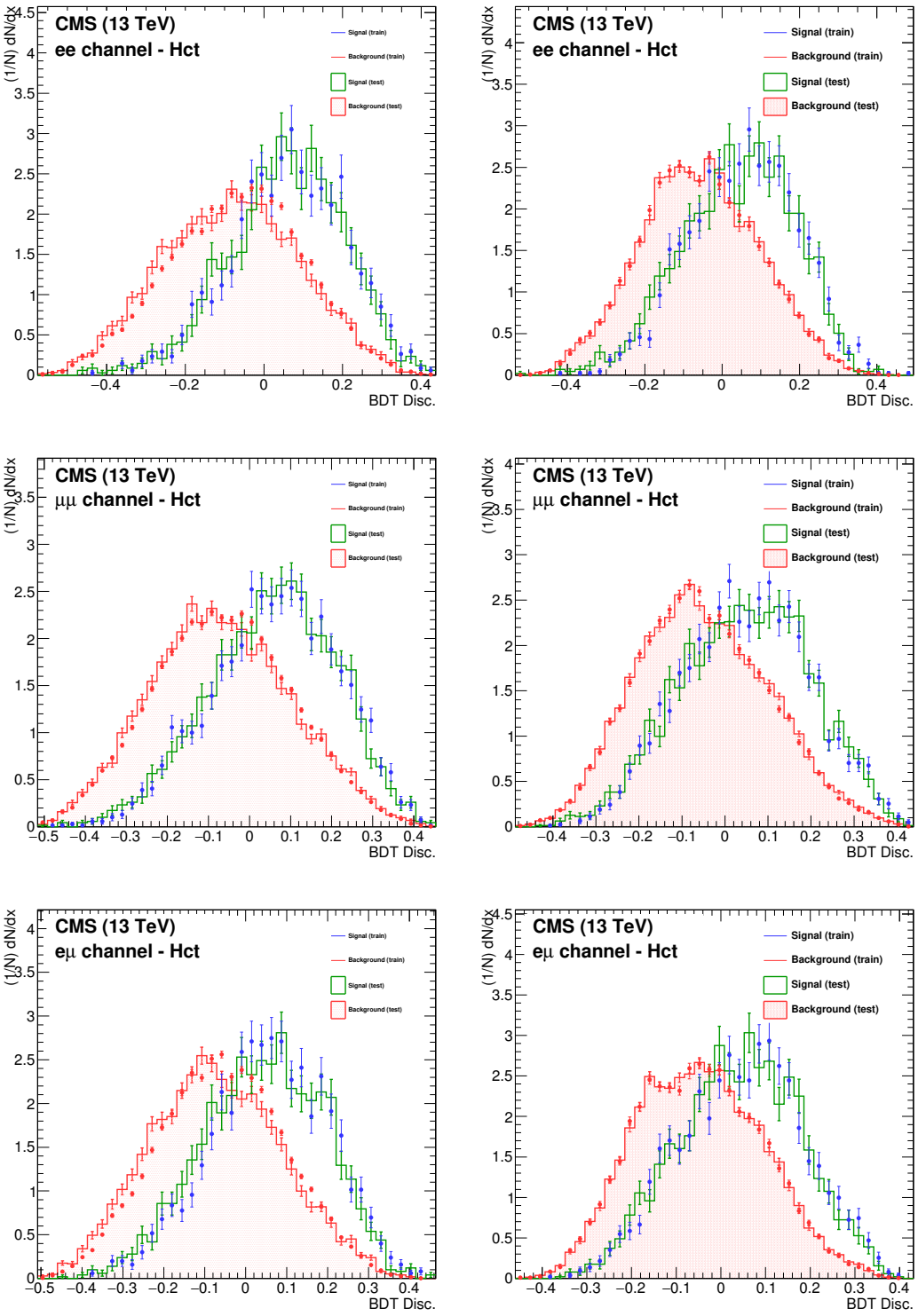


Figure 6.8: The figure shows the normalized distributions of BDT discriminators. For the Hct coupling in the *ee* channel (top), the $\mu\mu$ channel (middle) and the $e\mu$ (bottom). For each channel, the left plot is when for the training background events come from the OSCR region and for testing events are taken from the SSSR region. The right plots represent the case when the whole training is done with events from the OSCR region.

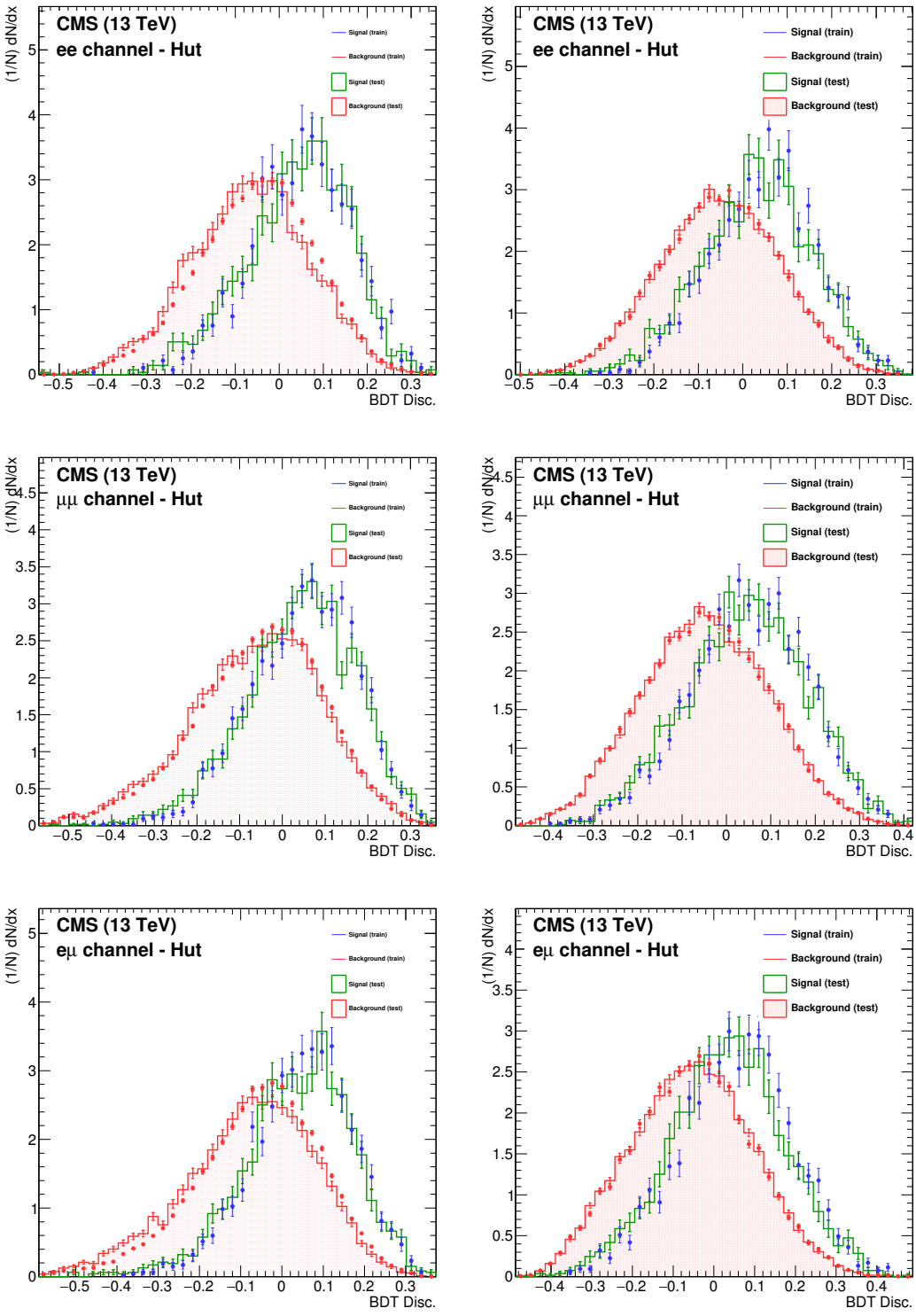


Figure 6.9: The figure shows the normalized distributions of BDT discriminators. For the Hut coupling in the ee channel (top), the $\mu\mu$ channel (middle) and the $e\mu$ (bottom). For each channel, the left plot is when for the training background events come from the OSCR region and for testing events are taken from the SSSR region. The right plots represent the case when the whole training is done with events from the OSCR region.

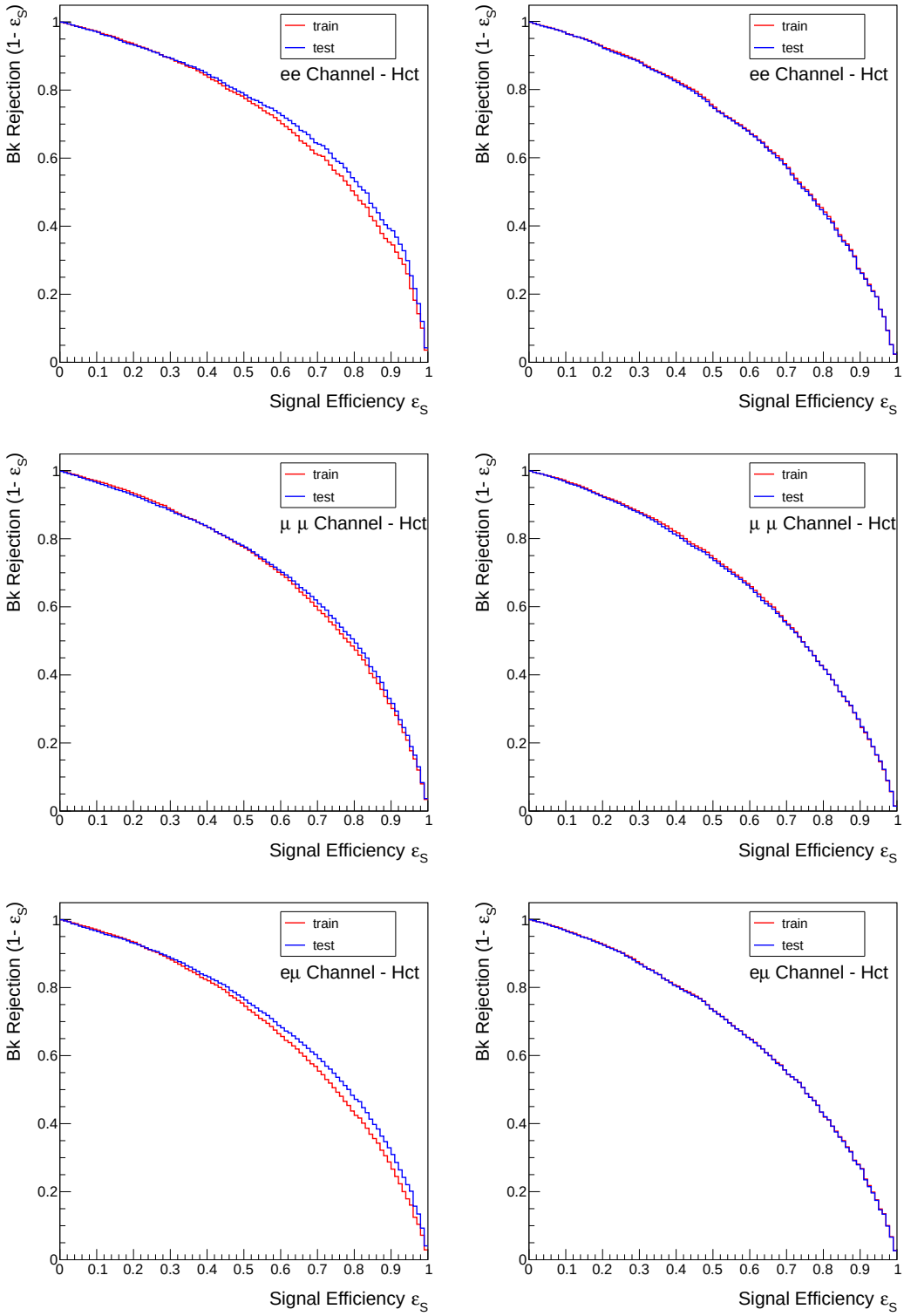


Figure 6.10: The figure shows the performance (ROC curve) of the BDT training for the Hct coupling in the ee channel (top), the $\mu\mu$ channel (middle) and the $e\mu$ channel (bottom). For each channel, the left plot is when for the training background events come from the OSCR region and for testing events are taken from the SSSR region. The right plots represent the case when the whole training is done with events from the OSCR region.

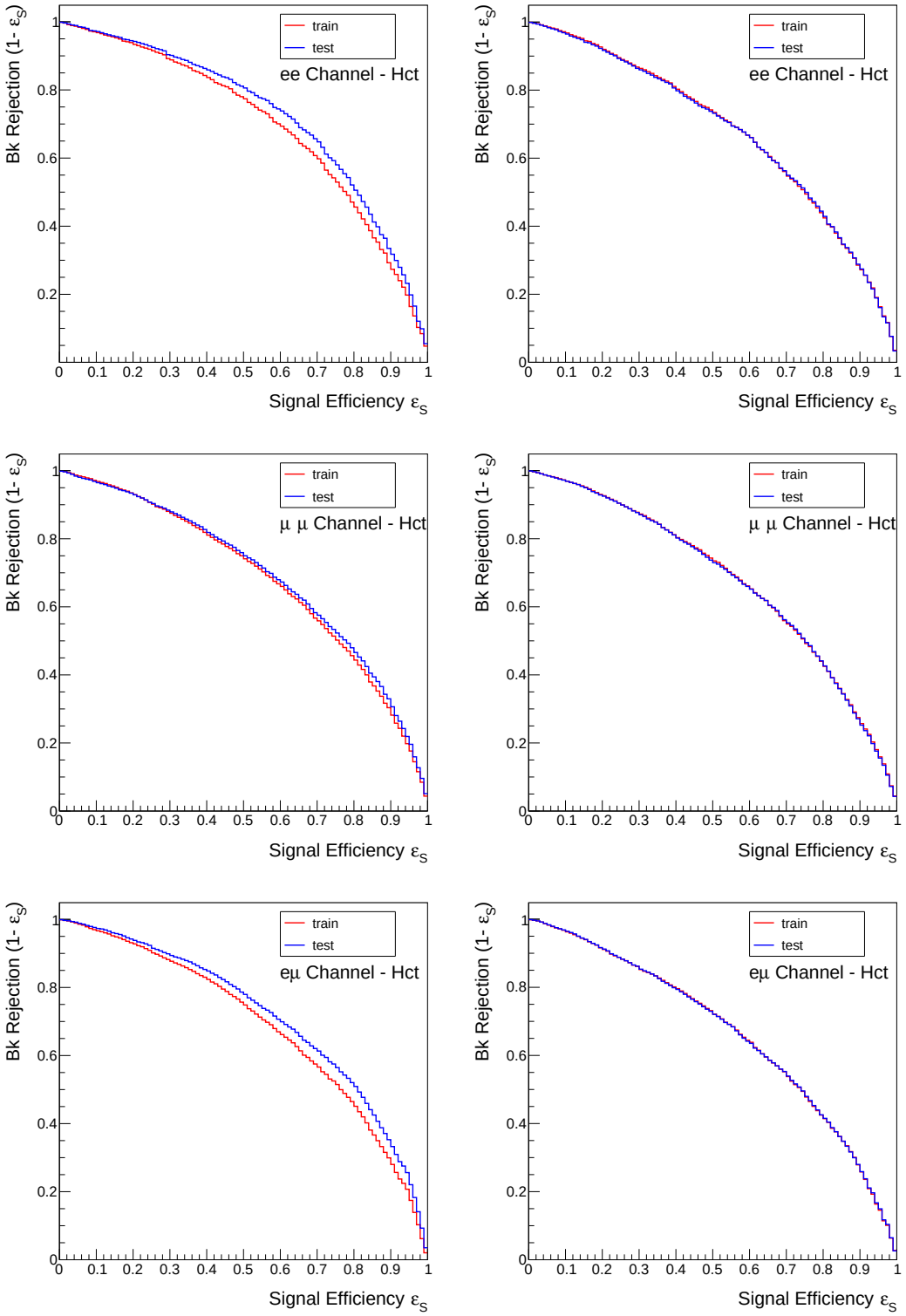


Figure 6.11: The figure shows the performance (ROC curve) of the BDT training for the Hut coupling in the ee channel (top), the $\mu\mu$ channel (middle) and the $e\mu$ channel (bottom). For each channel, the left plot is when for the training background events come from the OSCR region and for testing events are taken from the SSSR region. The right plots represent the case when the whole training is done with events from the OSCR region.

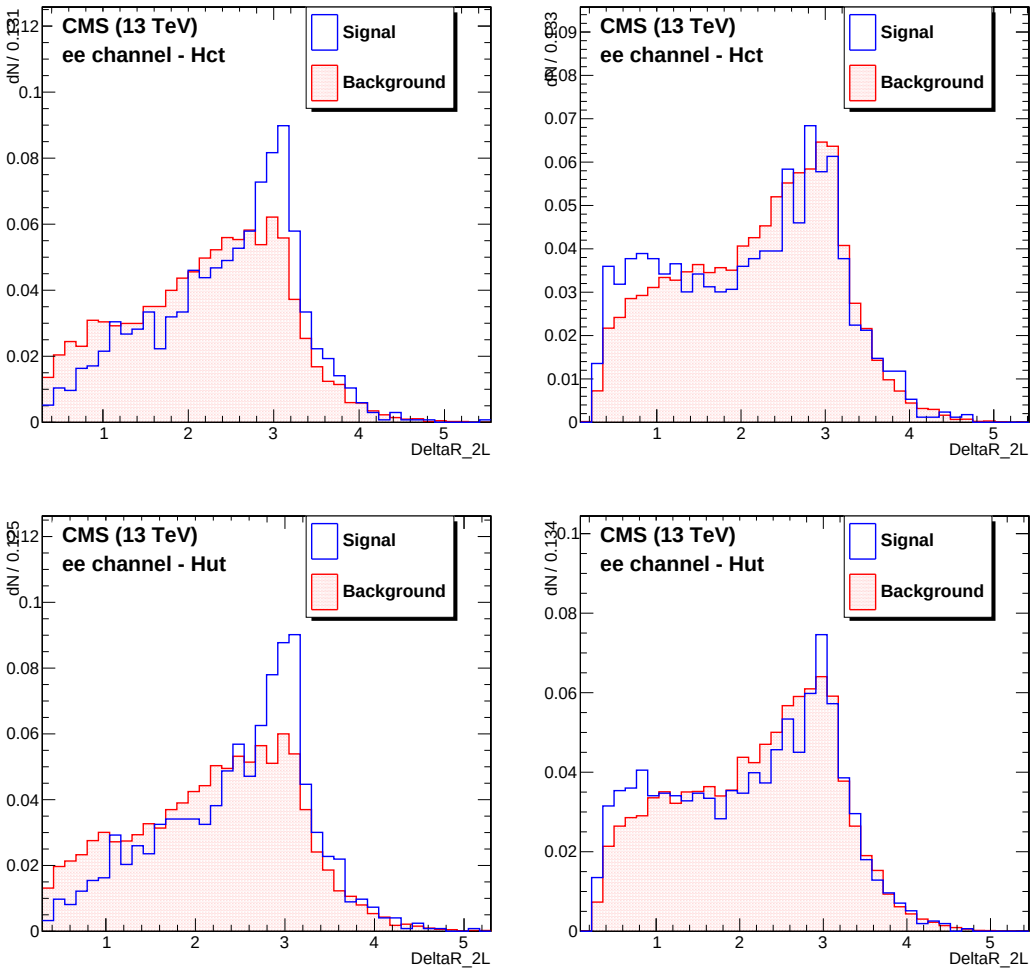


Figure 6.12: The figure shows the distribution of ΔR_{ee} with events from the SSSR region (left) or with events from the OSCR region (right) for the Hct (top) and the Hut (bottom) couplings.

6.2 Systematic uncertainties

The experimental systematic uncertainties in this physics analysis come from potential and observed mismatches between simulated and real observed proton collisions. These systematic uncertainties are divided into two categories. One that affects the number of the events passing the baseline selection and these are called normalization uncertainties. Those that affect as well the shape of the BDT discriminant distributions are called shape uncertainties. The systematic uncertainties that are considered in the physics analysis are listed in Table 6.3 and are described shortly below. A log-normal probability distribution is used as prior for the normalisation uncertainties while the shape uncertainties are auxiliary measurements modelled by template morphing techniques [159]. The effect of different systematic uncertainties on the distributions of the BDT discriminators for background and signal events will be presented in Sections 6.2.1 and 6.2.2 respectively.

Systematic	**Type**	**Systematic variation**
int. luminosity	normalization	2.5%
$t\bar{t}$ cross-section	normalization	5.6%
SM single-top quark cross-section	normalization	2.5%
other background processes cross-sections	normalization	30%
PileUp	shape	$\pm\sigma$ of minimum bias cross section
JER	shape	$\pm\sigma(p_{\mathrm{T}}, \eta)$
JES	shape	$\pm\sigma(p_{\mathrm{T}}, \eta)$
lepton SF	shape	$\pm\sigma SF(p_{\mathrm{T}}, \eta)$
b-tagging SF	shape	$\pm\sigma(p_{\mathrm{T}}, \eta)$
PDF	shape	PDF4LHC recipe
renormalization and factorization scales	shape	variation of μ_F and μ_R
ME-PS scaling	Shape	$\pm\sigma$ of ME-PS matching scale

Table 6.3: The table shows the different sources of systematic uncertainties.

6.2.1 The effect of systematic uncertainties on backgrounds

- **Integrated luminosity**:
  The overall uncertainty of the luminosity measurement for the 2016 data taking period is estimated to be 2.5% [135].

- **Cross-sections of background processes**:
  The dominant background in this physics analysis comes from $t\bar{t} + jets$ process. The systematic uncertainty on its cross section of 5.6% is taken from the NNLO accuracy obtained in [106]. The uncertainty on the cross-section of the other simulated backgrounds is taken to be 30% of their calculated cross-section such that it covers all uncertainties at NLO accuracy.

- **PileUp (PU)**:
  The systematic shift in the distribution of the number of pileup events is obtained by varying the minimum bias cross section [143], that is used to calculate the pileup distribution demonstrated in Section 5.2.4, by ±4.6%. The effect of uncertainties due to pileup on the BDT distribution is shown in Figure 6.13.

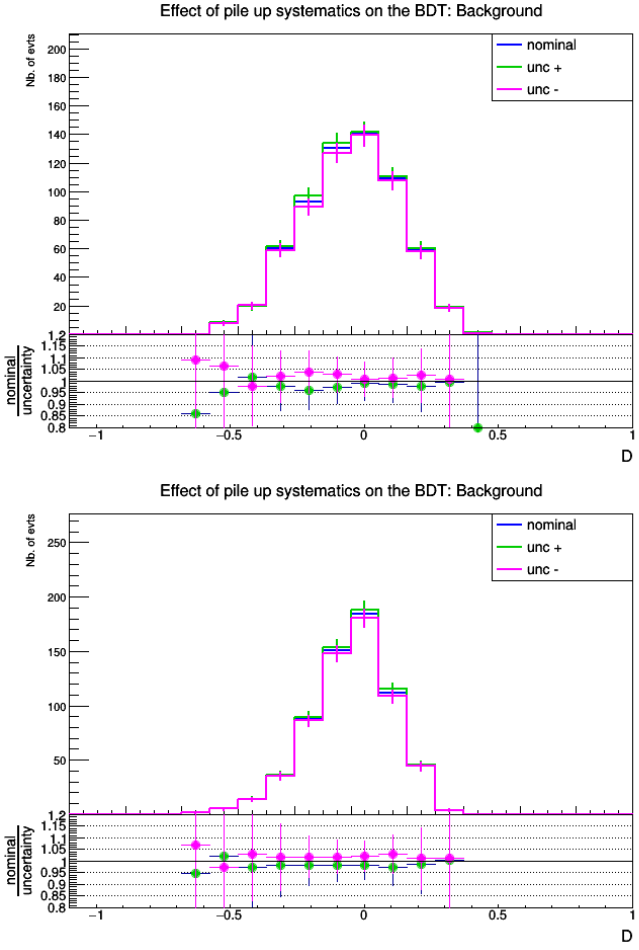


Figure 6.13: The effect of the PU systematic uncertainty on the BDT discriminator for the Hct- coupling (top) and Hct-coupling (bottom) analysis.

- **Lepton-SF systematic uncertainties**:
  The systematic uncertainty due to the lepton (electron and/or muon) scale factors consists of three sources: identification, isolation and tracking. Their systematic impact on the BDT discriminator distribution is obtained by varying the applied scale factors independently within one standard deviation ($\pm\sigma$) of their measurement. The effect of the uncertainties due to the lepton scaling factors on the BDT distribution is shown in Figure 6.14.

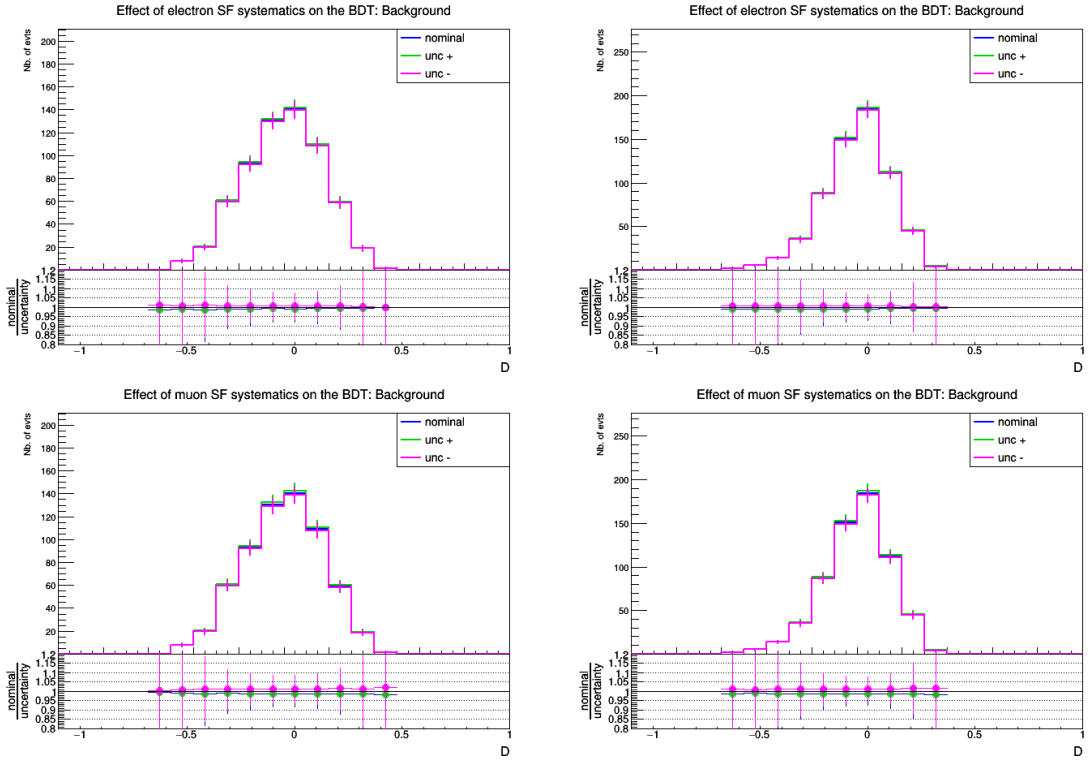


Figure 6.14: The impact of the electron-SF systematic uncertainties on the BDT discriminator for the Hct-coupling (top-left) and Hut-coupling (top-right) and the effect of the uncertainties from the muon-SF for Hct (bottom-left) and Hut (bottom-right) couplings respectively.

- **Jet energy corrections (JER/JES)**:
  Corrections and smearing are applied on jets in order to match the overall energy scale and resolution of simulation and data as described in Section 4.5.2. The systematic uncertainties due to these scale factors are estimated by varying them within their uncertainties.

  The estimation of the systematic effect of JES/JER uncertainties require repeating the measurements of all jet-related and $\not\!\!E_T$-related kinematical observables for up and down variations. For the evaluation of uncertainties due to the JES the recipe documented in [160] is used.

  Systematic differences in the JER are treated by scaling the difference of the reconstructed jet-p_{T} to the matched GenJet-p_{T} by a certain scaling factor as

given in Table 4.6. The effect of the uncertainties arising from the jet energy corrections on the shape of the distribution can be seen in Figure 6.15. The uncertainties from the JES have a large effect on the BDT distribution shape and normalization for both couplings. The effect is smaller than 15% for the Hct-coupling and 10% for the Hut-coupling. This is because when varying the JES upwards, the p_{T} of all jets in the event varies upward so that it is possible for additional jets to pass the minimal requirement of $p_{\mathrm{T}} \geq 30\,\mathrm{GeV}$ in the event selection.

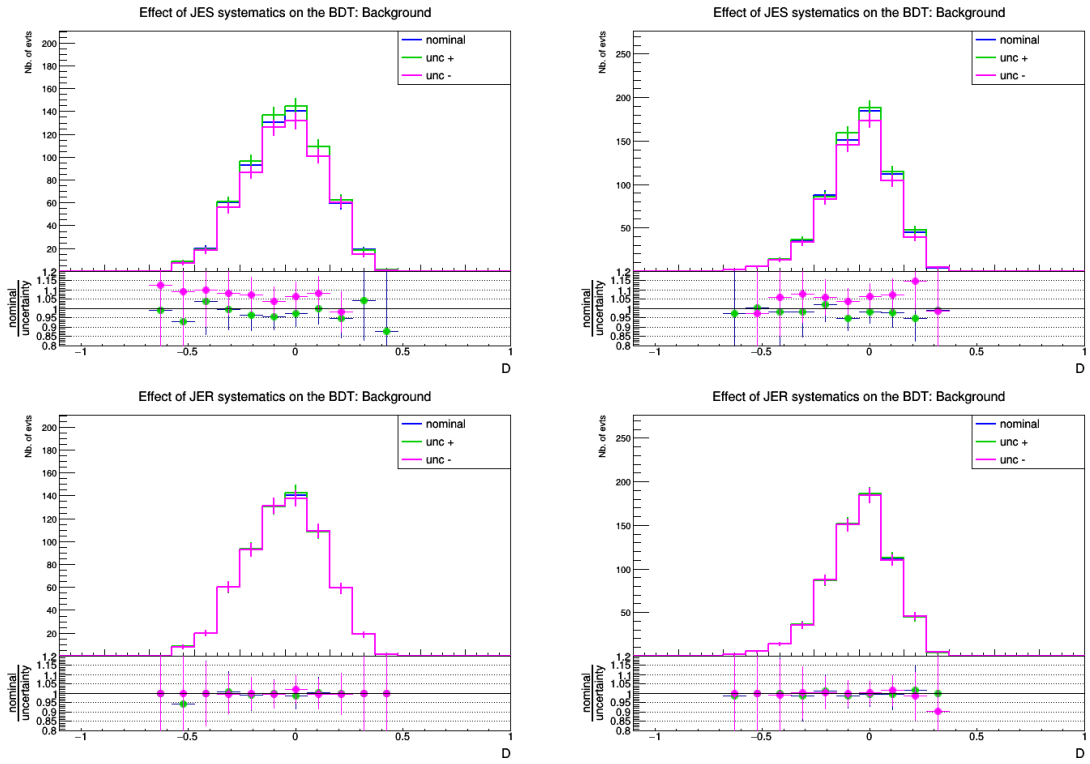


Figure 6.15: The effect of JES systematic uncertainties on BDT discriminator for Hct-coupling (top-left) and Hut-coupling (top-right) and from JER Hct- coupling (bottom-left) and Hut-coupling (bottom-right).

- **B-tagging uncertainties**
  There are eight sources of systematic uncertainties that affect the shape of the CSVv2 distribution, which are all considered as uncorrelated nuisance parameters in this physics analysis [161]. These uncertainties are divided into three sources. First the uncertainties associated with the JES which are evaluated simultaneously with the JES uncertainty since it is 100% correlated to the JES uncertainties. The second, systematic uncertainties due to the purity of the b-tagging performance measurements. It consists of two uncorrelated uncertainties one corresponding to the light flavour (lf) and the other to the heavy flavour (hf) jet contributions in the measurement of b-tagging SF. A $\pm\sigma$ shift in each of the two purity uncertainties corresponds to a higher or lower contamination of these flavours in the scale IterativeFit procedure.

  The third, the statistical uncertainty which consists of four uncorrelated systematic uncertainties, two for heavy flavour (hfstats1 and hfstats2) and two for light flavour (lfstats1 and lfstats2). Where (hfstats1,lfstats1) are uncertainties corresponding to a shift with statistical uncertainties on the scaling factors. The second uncertainties (hfstats2, lfstats2) are 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 the charm jet scale factors are obtained from the uncertainty on the heavy flavour scale factors. This is by doubling it in size and constructing two nuisance parameters (cferr1 and cferr2) corresponding to linear and quadratic uncertainties to control the charm flavour scale factors and treating them as independent uncertainties. The effect of the uncertainties due to b-tagging SF are shown in Figures 6.16 for the Hct-coupling and 6.17 for Hut-coupling. For both couplings the total effect of the uncertainties due to b-tagging is very small.

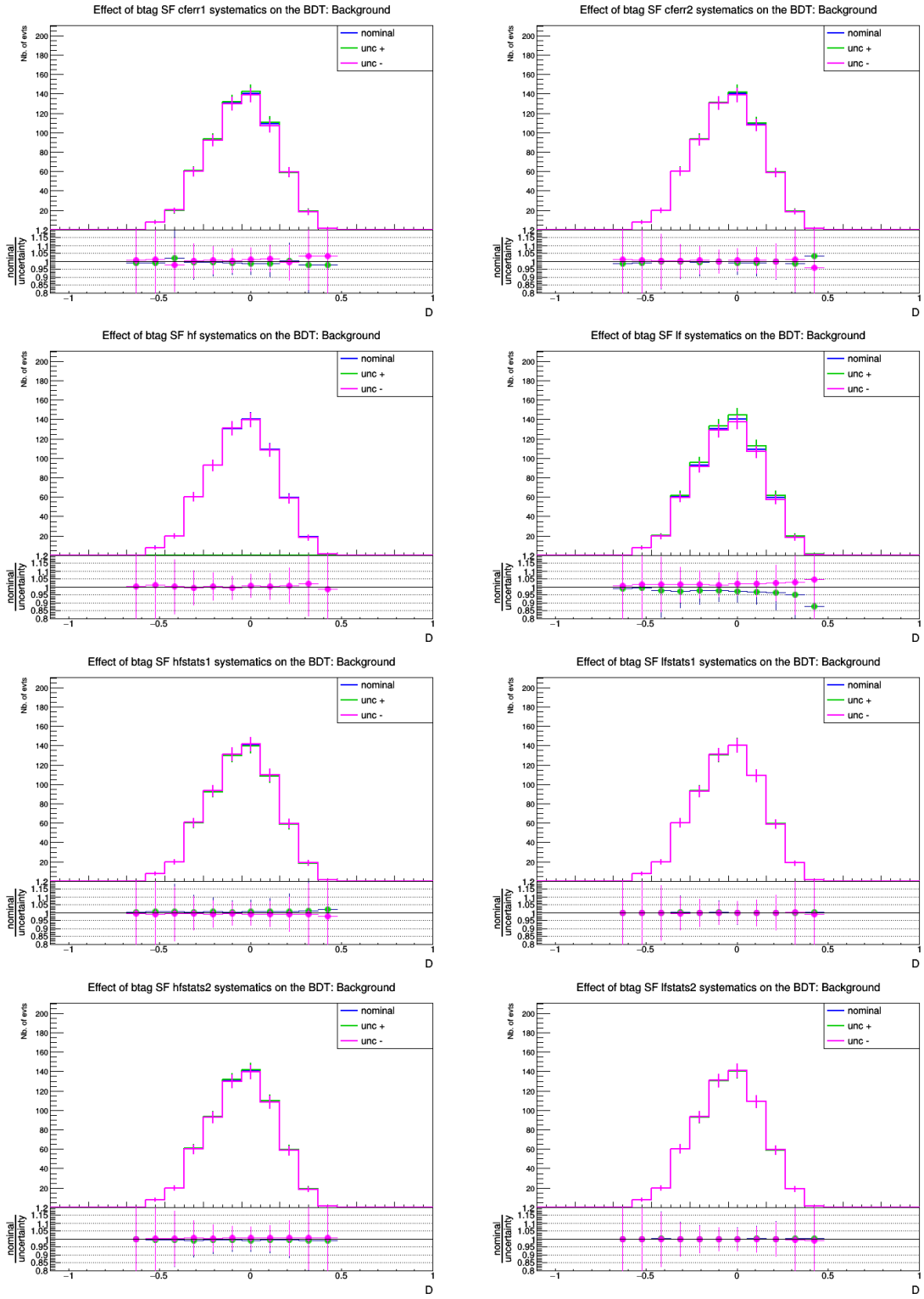


Figure 6.16: The effect of systematic uncertainity on BDT discriminator from different b-tagging systematic uncertainties

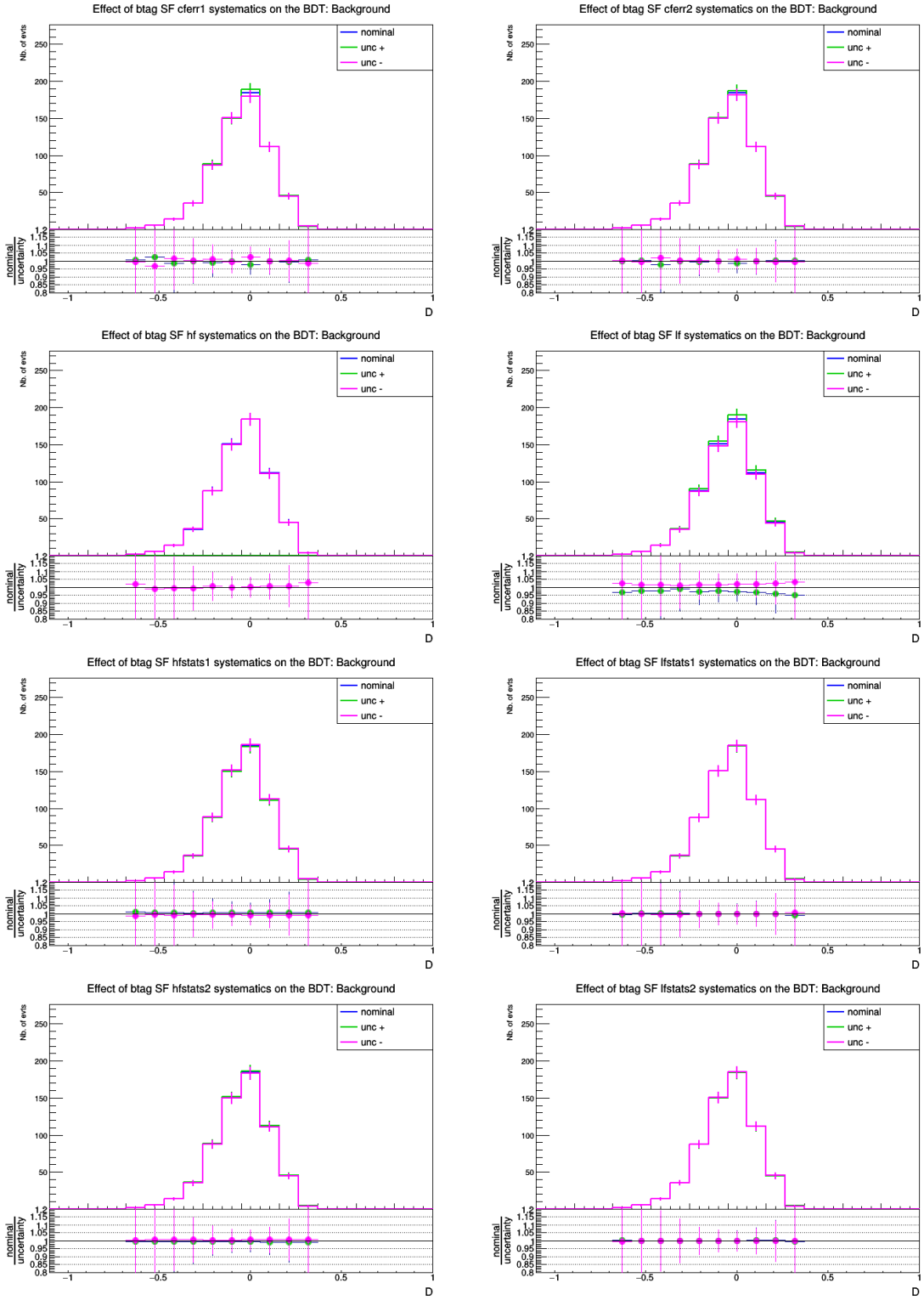


Figure 6.17: The effect of systematic uncertainity on BDT discriminator from different b-tagging systematic uncertainties

- **PDF uncertainties**

  Theory uncertainties due to parton density functions used for the modelling of the background are estimated only for the $t\bar{t}+jets$ process to account for the effect on the shape of the BDT distributions. This is done using the PDF4LHC recipe described in [88], which combines the MMHT14, CT14, and NNPDF3.0 PDF sets. The nominal pdf-set for the used $t\bar{t}$ sample is taken from $NNPDF30\_nlo\_as\_0118$ while the set of error PDFs is taken from $PDF4LHC15\_nlo\_100$. The effect of this on the BDT distribution of $t\bar{t}+jets$ simulated events is shown in Figure 6.18.

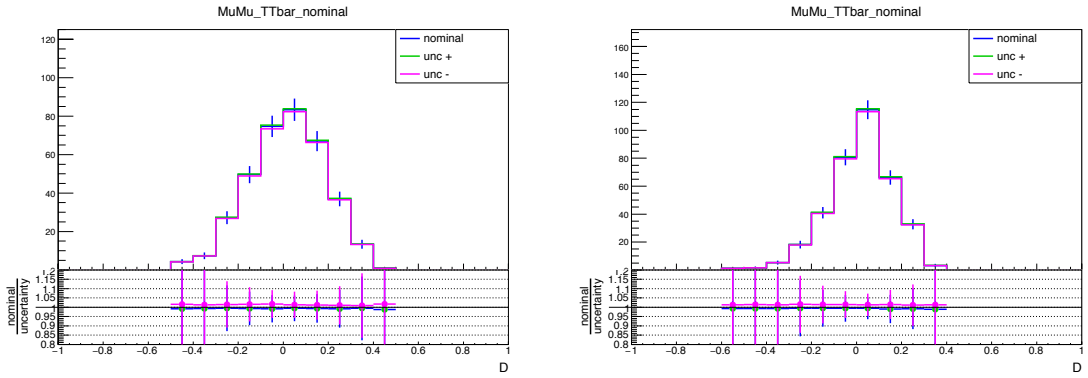


Figure 6.18: The effect of the PDF systematic uncertainties on the BDT discriminator distribution for the Hct-coupling (left) and Hut-coupling (right) analysis.

- **Renormalization and Factorization scales**

  The effect of theory uncertainties due to the factorization scale μ_F and renormalization scale μ_R is estimated by varying the μ_F by a factor 0.5 and 2 at fixed μ_R then varying μ_R by a factor 0.5 and 2 at fixed μ_F. Two more variations at matrix element-level are retrieved by varying μ_F and μ_R simultaneously by a factor 2 and 0.5. The envelope of these variations is used as an uncertainty on the final BDT distribution. This uncertainty is only considered for the $t\bar{t}+jets$ background. The uncertainties due to renormalization and factorization scales have the largest effect on the BDT distribution but still within 20% for both the Hct and Hut coupling as shown in Figure 6.19.

- **ME-PS matching scale**

  Other systematic uncertainties from the modeling of simulated events arise from the matching of matrix-element (ME) partons and parton showers (PS) in `POWHEG`. The high-p_{T} radiations during this matching process are regulated by damping real emissions with a factor of $h^2_{damp}/(p^2_{\mathrm{T}}+h^2_{damp})$. The nominal value for the h_{damp} parameter was calculated to be $1.58^{0.66}_{-0.59}$ m_t for a top of mass $m_t = 172.5\,\mathrm{GeV}$ [162]. The differences between the nominal and the up/down h_{damp} variations are estimated using dedicated weighted samples for which the information stored in the simulated samples produced by `POWHEG`. In this analysis the ME-PS systematic uncertainty is applied for $t\bar{t}$ simulated samples and its effect on the BDT distribution is shown in Figure 6.20.

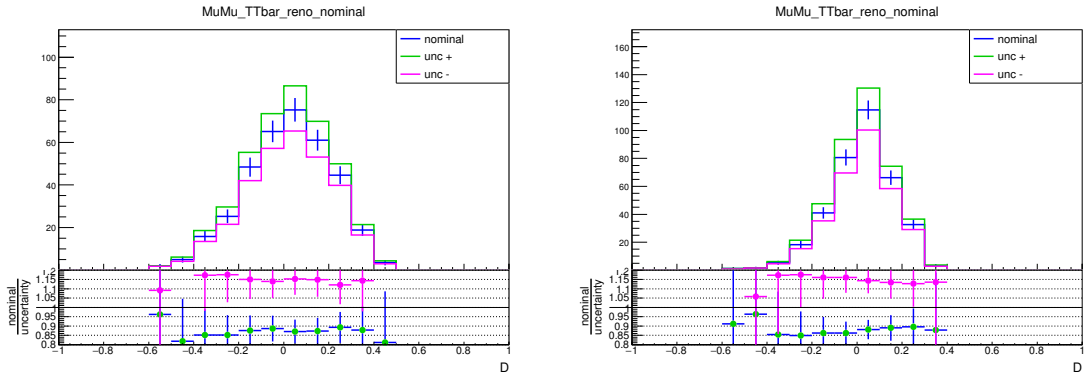


Figure 6.19: The effect of systematic uncertainties due to the renormalization and factorization scales on the BDT discriminator distribution for the Hct-coupling (left) and the Hut-coupling (right) analysis.

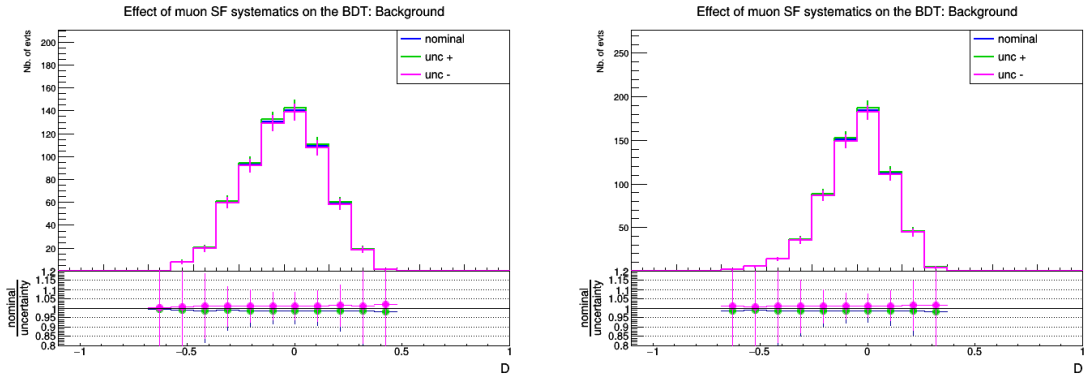


Figure 6.20: The effect of the ME-PS matching scale systematic uncertainties on the BDT discriminator distribution for Hct-coupling (left) and Hut-coupling (right) analysis.

6.2.2 The effect of systematic uncertainties on the signal

The effect of the systematic uncertainties on the BDT discriminator distribution for the signal is also measured. The uncertainties due to pileup, lepton identification, jet energy resolution, the b-tagging, as well as those from theory are within 1% to 3%. The systematic uncertainty due to JES corrections has the largest effect on the BDT discriminator distribution. As shown in Figure 6.21 it is still within 15% for the Hct-coupling and 10% for the Hut-coupling analysis.

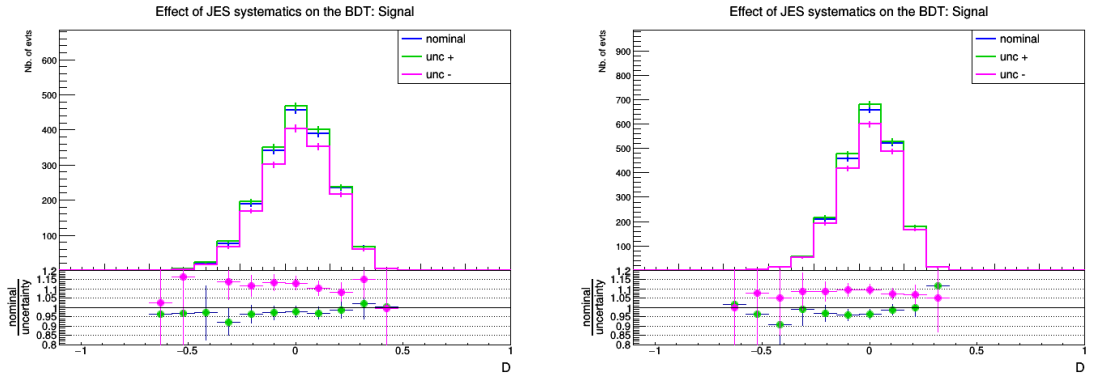


Figure 6.21: The effect of the JES systematic uncertainty on the BDT discriminator distribution of signal events for the Hct- coupling (left) and the Hut-coupling (right) analysis.

6.3 Validation of the limit setting procedure

In this physics analysis we want to assure that no bias is introduced into the measurements. Therefore the analysis strategy has been established using a blinded methodology where the limit setting procedure has been validated by using a pseudo dataset. We perform a signal injection test in which the pre-set signal strength, from a pseudo dataset, is injected. As can be seen from Figure 6.22 for both the Hct and the Hut couplings, for each pre-set signal strength, $\mu = 0.1, 0.2, 0.3, 0.4, 0.5$, the maximum likelihood estimate returns an estimate value $\hat{\mu}$ that agrees within its uncertainty with the expected value.

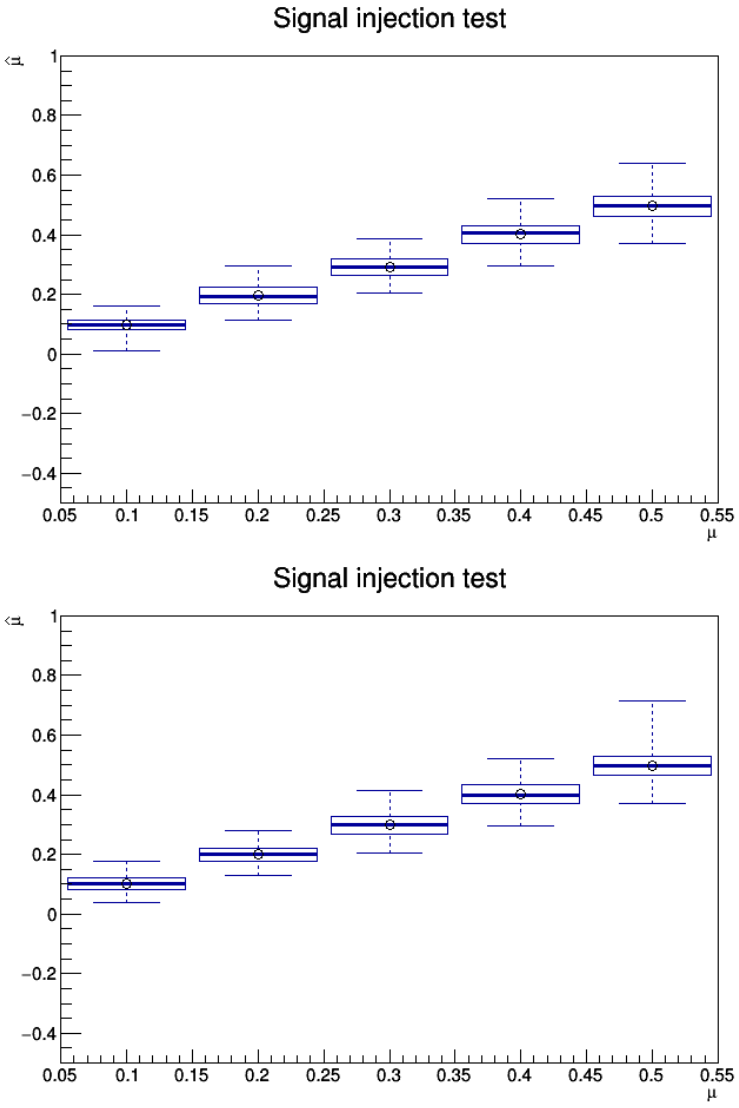


Figure 6.22: The estimated signal strength with the Maximum Likelihood method $\hat{\mu}$ is in agreement with the signal strength μ used to generate the pseudo-data set for the Hct (top) and Hct (bottom) couplings. The solid line represents the input signal strength while the open marker is the "measured" signal strength for pseudo-data.

6.4 Results and discussion

The limit setting procedure explained in Section 5.3.2 is applied for each di-lepton channel separately and for the combination of the three channels as well. The results are obtained in terms of signal strength (μ) as shown in Figure 6.23. For both the Hut and the Hct couplings the maximum likelihood estimator for their signal strengths (μ) is compatible with zero. This means no signal is observed, neither for the Hut nor the Hct couplings in this analysis. As expected the FCNC analysis has more sensitivity to the Hut vertex than to the Hct vertex. This is due to the attribution of the single-top quark component of the FCNC signal where the production cross-section with the Hut vertex is higher than that with the Hct vertex.

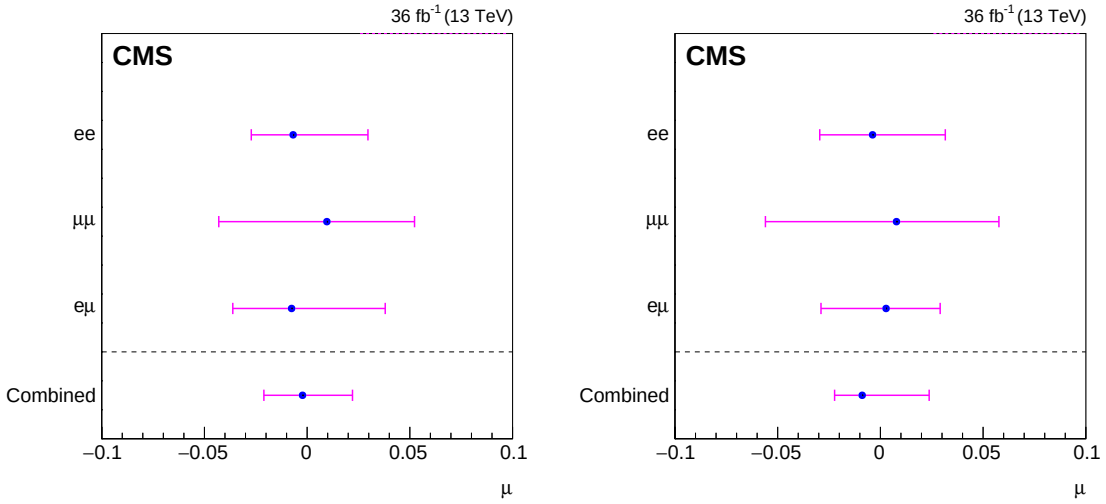


Figure 6.23: The maximum likelihood estimators for the signal strengths for the Hct coupling (left) and Hut coupling (right) for each di-lepton channel as well as for the combination of all channels.

The maximum likelihood estimators for the nuisance parameters $\hat{\theta}$ for the Hct and the Hut coupling analysis are shown in Figure 6.24. The values obtained from the signal plus background (S+B) fit are in agreement with those from the background-only fit. This is expected since there is no signal observed in the data.

The impact of the uncertainties on the maximum likelihood estimate of the signal strength is shown in Figures 6.25 and 6.26 for the Hct and Hut couplings respectively. One can see that the uncertainties due to the re-normalization and factorization scales have the largest effect followed by those from the cross-section uncertainties of the $t\bar{t}+jets$ background. Furthermore, the nuisance parameters related to them and those related to the JES uncertainty are shifted with respect to their initial values.

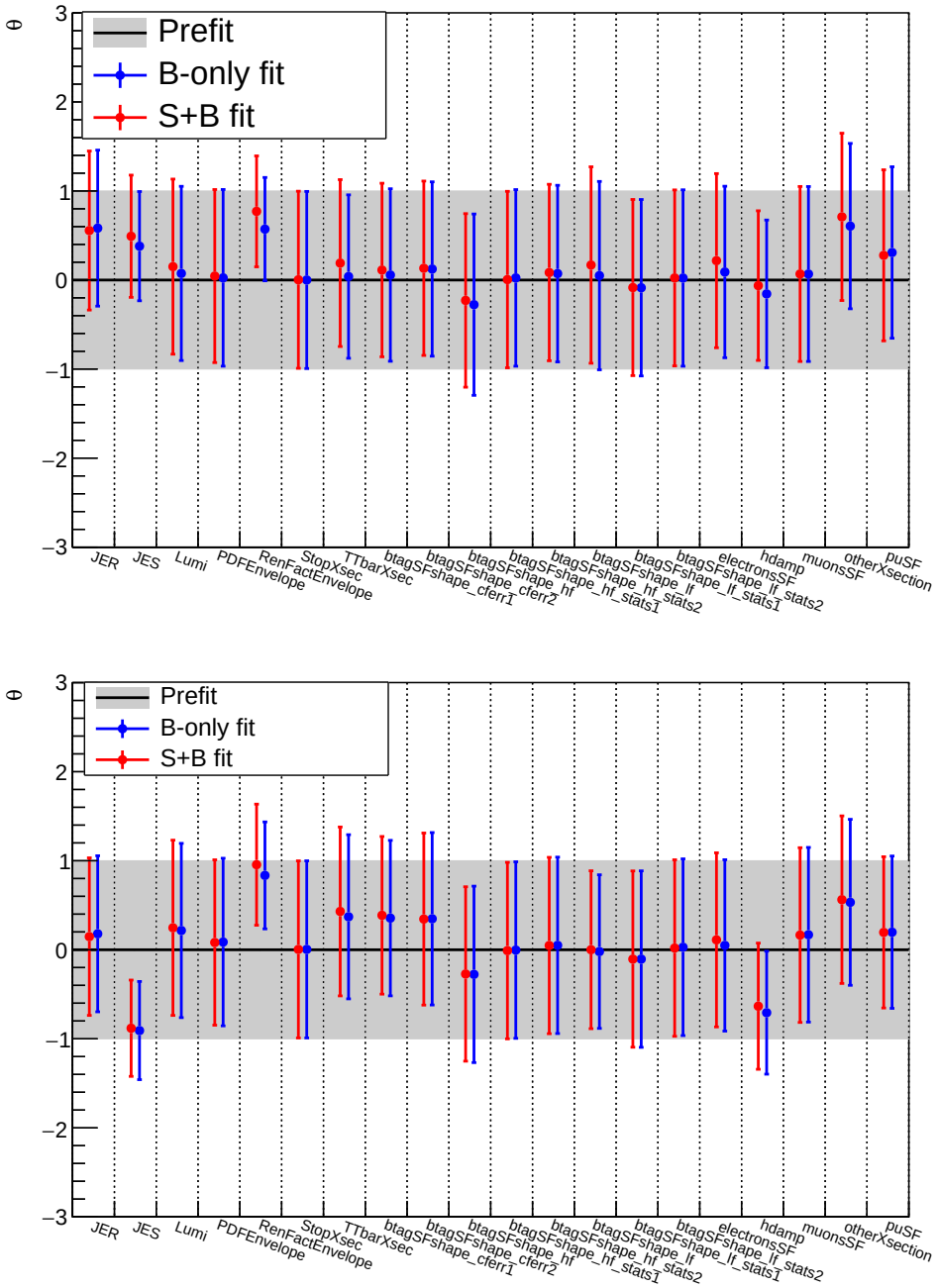


Figure 6.24: The maximum likelihood estimators of the nuisance parameters for the Hct vertex (top) and Hut vertex (bottom) analysis where all di-lepton channels are combined.

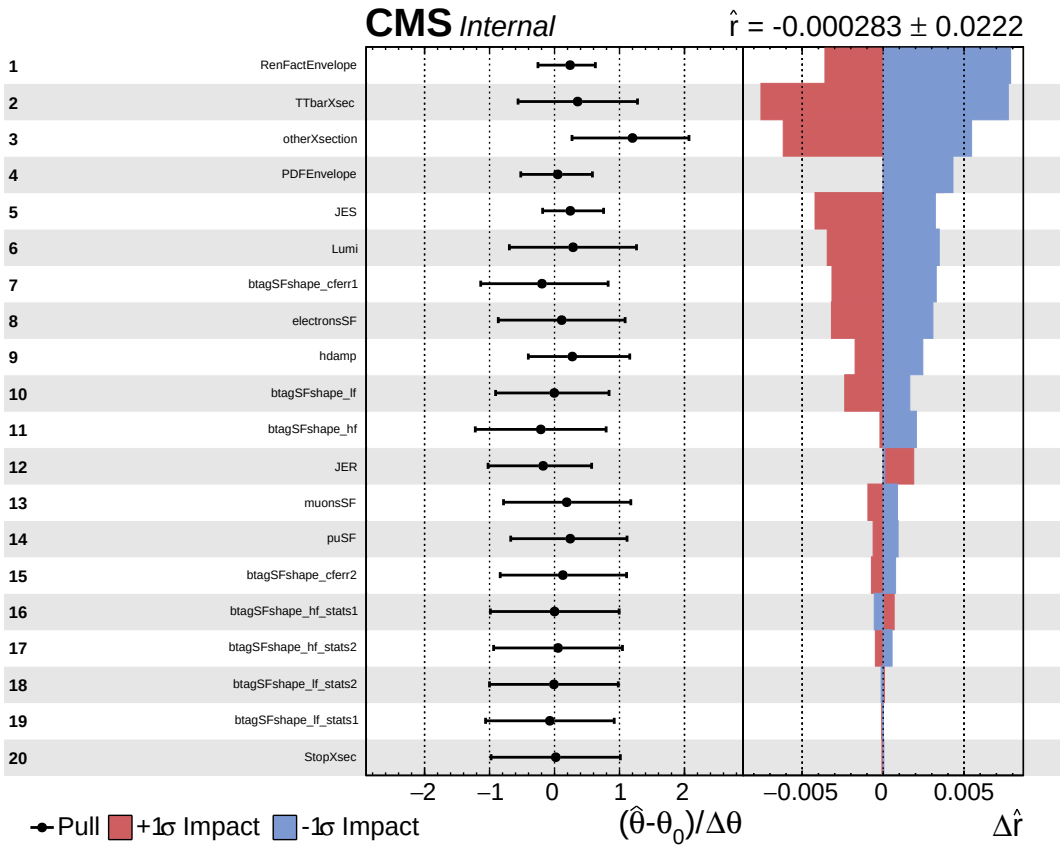


Figure 6.25: The pulls (shifts) of the nuisance parameters and the influence of their uncertainty on the maximum likelihood estimation of the signal strength $\hat{r}$ for the Hct coupling.

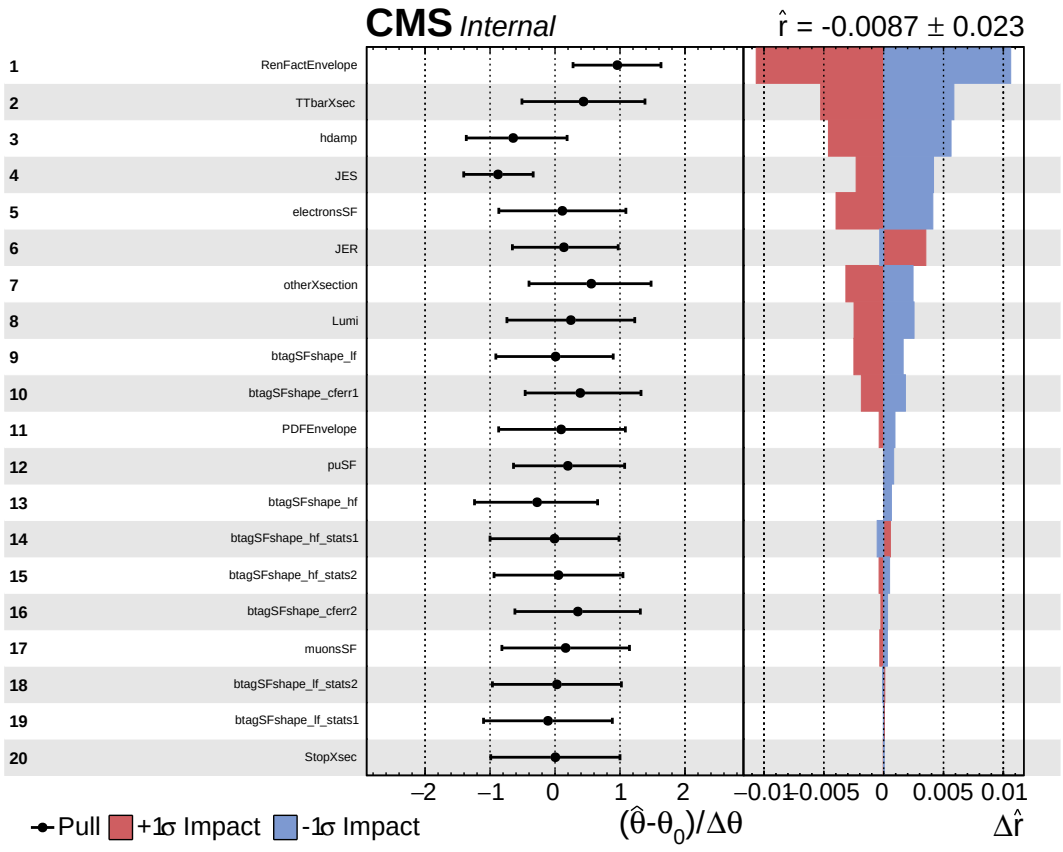


Figure 6.26: The pulls (shifts) of the nuisance parameters and the influence of their uncertainty on the maximum likelihood estimation of the signal strength $\hat{r}$ for the Hut coupling.

6.4.1 Postfit distributions and yields

The distributions of the BDT discriminator in each channel is recreated using the maximum likelihood estimations of the nuisance parameters. This is done in the form of reshaping and normalising the templates of the background processes. The resulting distributions are shown in Figure 6.27 for the Hct coupling and in Figure 6.28 for the Hut coupling. A good agreement between data and simulation is obtained for all channels. The fluctuations in case of ee and $e\mu$ channels are due to low statistics in these channels. In Table 6.4 and Table 6.5, the event yields for each process after the global fit are given for the signal involving the Hct-coupling and Hut-coupling respectively. For both couplings analyses all backgrounds have risen within their uncertainties.

Process	**ee** channel	$\mu\mu$ channel	$e\mu$ channel
$t\bar{t}$+jets	76±9	389 ± 20	127± 11
DY+jets	6.2± 2.5	–	–
$t\bar{t}$V	9.1± 3	15.6 ±4	11.5 ± 3.4
$t\bar{t}$H	7.1± 2.7	14.7±3.8	9.5±3.1
Diboson	3.7±1.9	8.6±2.9	5.5±2.3
SM S-top	3.1± 1.8	23.7±4.9	9.3 ± 3.0
total Bkg.	105 ± 10	452 ± 21	162± 13
Data	102 ± 10	458 ± 21	168 ± 13

Table 6.4: The table shows event yields for SM backgrounds simulated samples as well as Run2 2016 data after the baseline selection cuts in the SSSR region

Process	**ee** channel	$\mu\mu$ channel	$e\mu$ channel
$t\bar{t}$+jets	84±9	411 ±20	140± 12
DY+jets	4.2± 2.1	–	–
$t\bar{t}$V	6.9±2.69	12.2 ±3.5	9.1 ±3.0
$t\bar{t}$H	5.8±2.4	12.0±3.5	7.6±2.8
Diboson	2.5± 1.6	6.8±2.6	4.1± 2.0
SM S-top	2.7±1.7	19.7±4.4	7.7 ± 2.8
total Bkg.	106±10	461±22	169 ± 13
Data	102 ± 10	458 ± 21	168 ± 13

Table 6.5: The table shows event yields for SM backgrounds simulated samples as well as Run2 2016 data after the baseline selection cuts in the SSSR region.

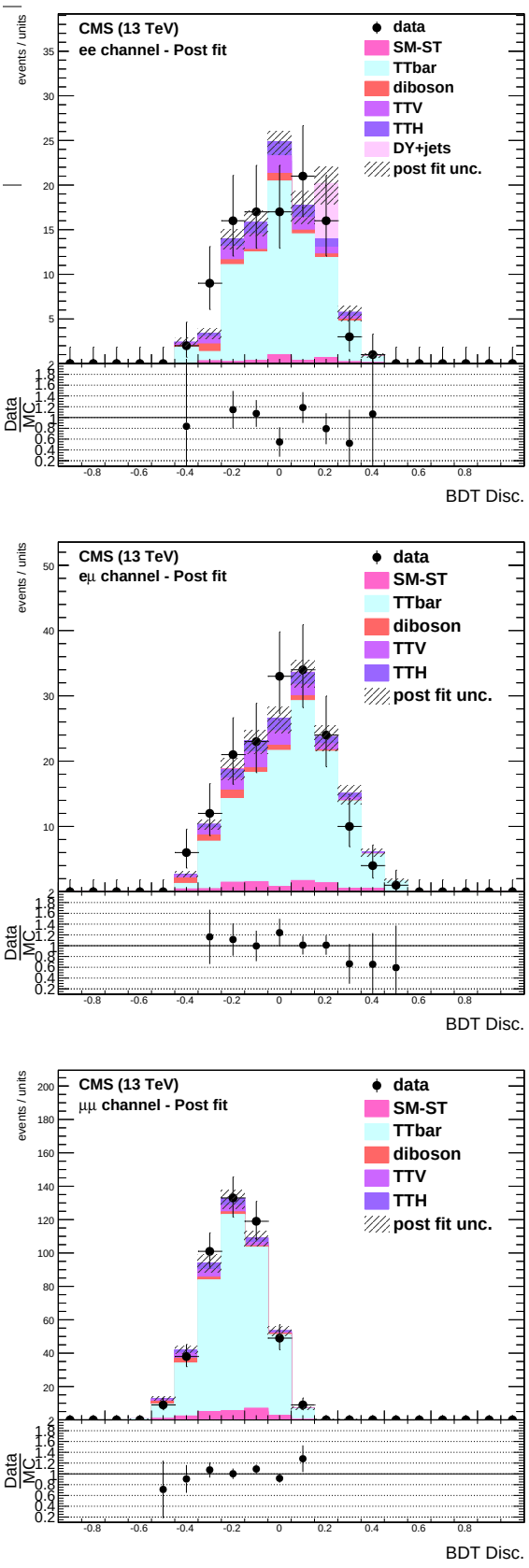


Figure 6.27: The post-fit distribution of BDT discriminator for ee (top), $e\mu$ (middle) and $\mu\mu$ (bottom) channels for Hct coupling.

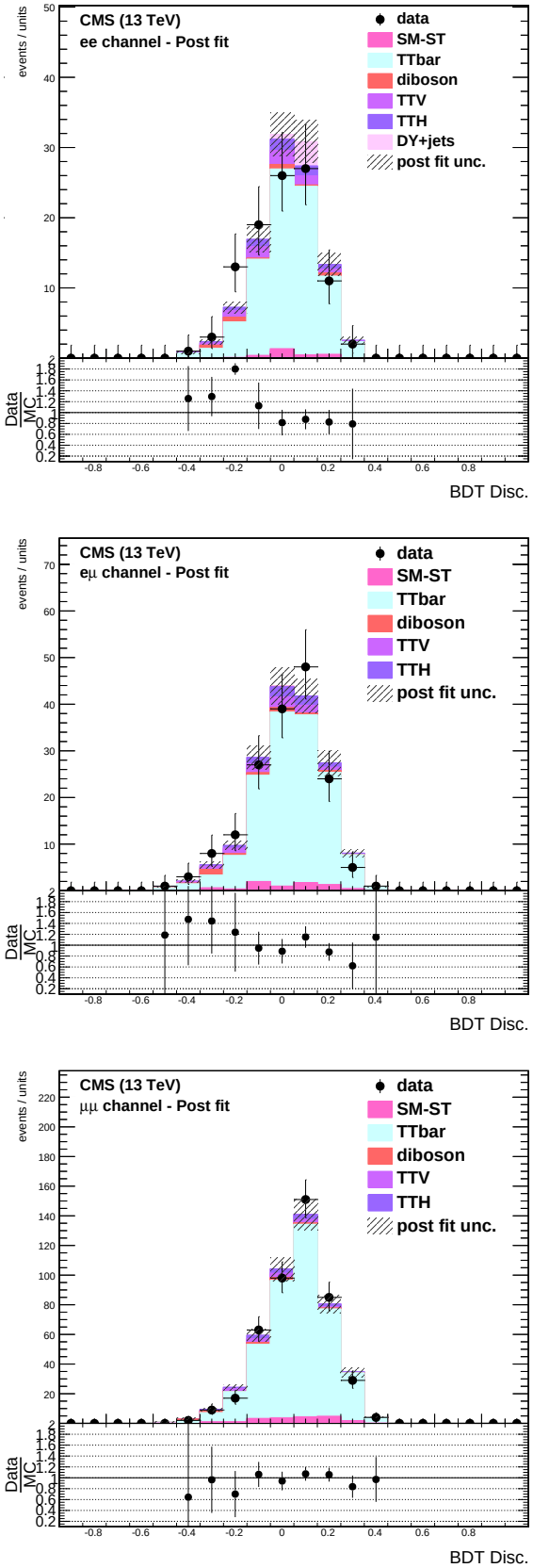


Figure 6.28: The post-fit distribution of BDT discriminator for ee (top), $e\mu$ (middle) and $\mu\mu$ (bottom) channels for Hut coupling.

6.4.2 Obtained limits at 95% CL

The limit setting procedure used in this physics analysis returns limits in terms of the signal strength. As the signal cross sections depend quadratically on the coupling strength, the interpretation of these limits in terms of the coupling strength is straightforward. These limits are also interpreted as a limit on the branching ratio for $(t \to cH)$ and $(t \to uH)$ using Equation 1.29.

6.4.2.1 One-dimensional limits

The expected, and observed, exclusion limits on the Hct and the Hut coupling strengths are shown in Figure 6.29 and Figure 6.30 respectively.

The expected (observed) exclusion limits on the Hct and the Hut coupling strengths at 95% CL are $\kappa_{Hct} \leq 0.22$ (0.218) and $\kappa_{Hut} \leq 0.22$ (0.21) respectively. The corresponding expected (observed) branching ratio upper limits are $\mathcal{B}(t \to cH) \leq 7.0 \times 10^{-3}$ (6.8×10^{-3}) and $\mathcal{B}(t \to uH) \leq 6.70 \times 10^{-3}$ (6.1×10^{-3}). The upper limits on the Hut coupling are slightly tighter than on the Hct coupling. This is due to the fact that the single top quark component has a much higher contribution in the Hut analysis compared to the Hct analysis. For both couplings the contribution from the single top quark component is however small. A summary of the expected and observed upper limits on the cross-sections of the combined (ST+TT) signal for the Hct and the Hut vertex is presented in respectively Tables 6.6 and 6.7.

channel	-2σ	-1σ	**Expected**	**Observed**	$+1\sigma$	$+2\sigma$
ee	0.46	0.64	0.94	0.82	1.42	2.1
$\mu\mu$	0.93	1.26	1.76	2.39	2.48	3.43
$e\mu$	0.55	0.75	1.04	1.16	1.44	1.94
comb	0.36	0.49	0.69	0.68	0.99	1.36

Table 6.6: Excluded cross sections in pb of the combined (ST+TT) FCNC $(t \to cH)$ signal for each of the di-lepton channels and their combination. For the expected upper limit, the limit plus and minus one standard deviation and two standard deviations are shown.

6.4.2.2 Two-dimensional limits

In previous sections we consider the Hut and Hct couplings as separate signals. However in reality these two signal processes may occur simultaneously. One approach is to interpolate the one dimensional limits to a scenario where both couplings are considered. This can be achieved by combining the signal of the two top-FCNC couplings

channel	-2σ	-1σ	**Expected**	**Observed**	$+1\sigma$	$+2\sigma$
ee	0.71	1.0	1.52	1.07	2.32	3.46
$\mu\mu$	1.10	1.47	2.04	3.26	2.87	3.84
$e\mu$	0.56	0.76	1.066	1.12	1.5	2.03
comb	0.45	0.61	0.86	0.78	1.24	1.71

Table 6.7: Excluded cross sections in pb of the combined (ST+TT) FCNC ($t \to uH$) signal for each of the di-lepton channels and their combination. For the expected upper limit, the limit plus and minus one standard deviation and two standard deviations are shown.

into one signal as

$$Signal = \kappa^2_{Hct}(ST + TT)_{Hct} + \kappa^2_{Hut}(ST + TT)_{Hut}. \tag{6.1}$$

For 100 benchmark of signal scenarios an approach explained in [163] has found that the experimental extrapolation formula 6.31

$$limit_{2D} = \kappa_{Hct(1D)}\sqrt{1 - \frac{\kappa_{Hut}}{\kappa_{Hut(1D)}}} \tag{6.2}$$

can be used to interpolate one-dimensional limits $\kappa_{Hct(1D)}$ and $\kappa_{Hut(1D)}$ into the two-dimensional limits on the coupling and branching ratios as shown in Figure 6.31.

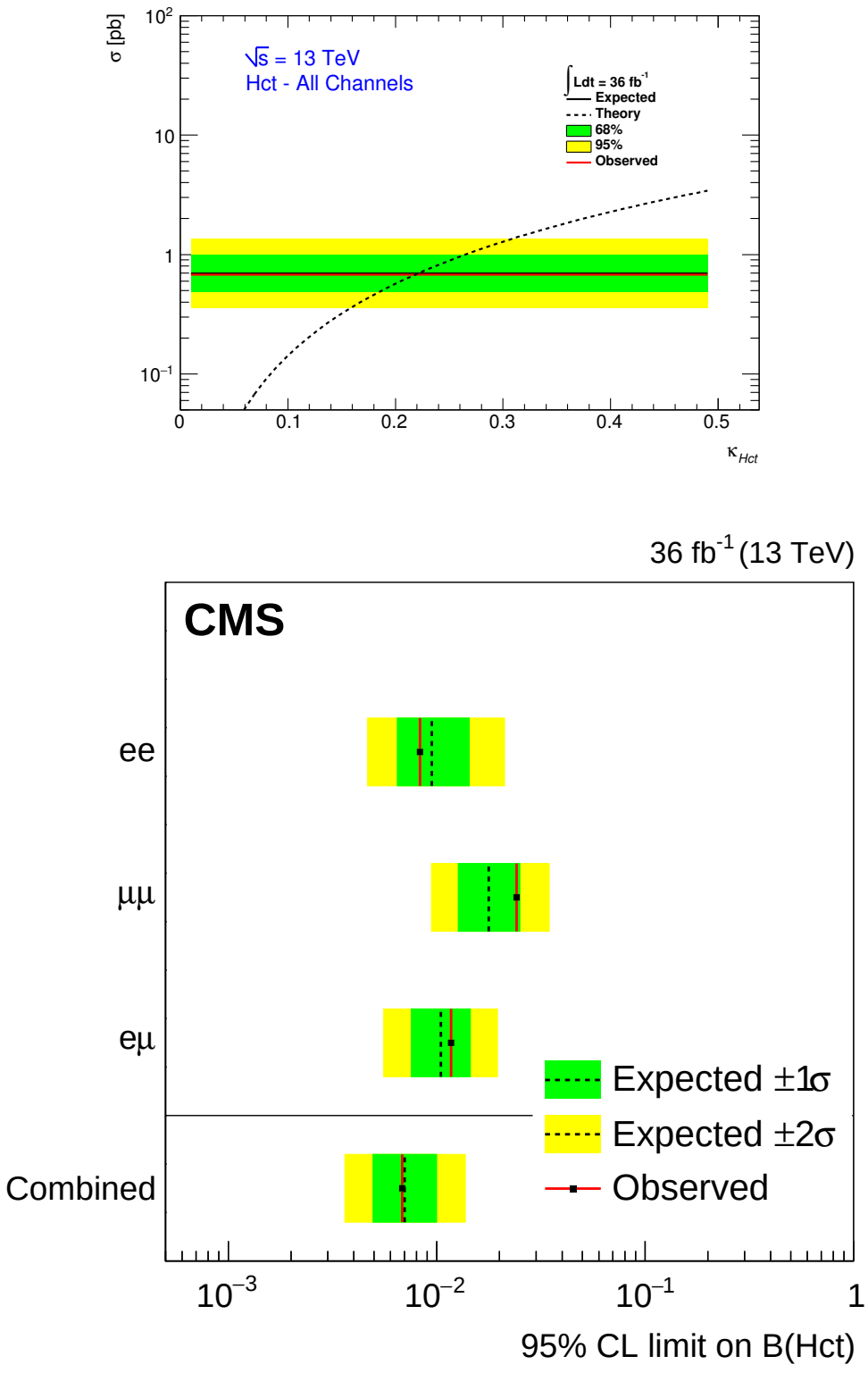


Figure 6.29: Exclusion limits at 95% CL on the coupling strenth (top) as a function of the cross section of the FCNC process. Limits on the FCNC-Hct branching ratios (bottom) for each di-lepton channel and their combination.

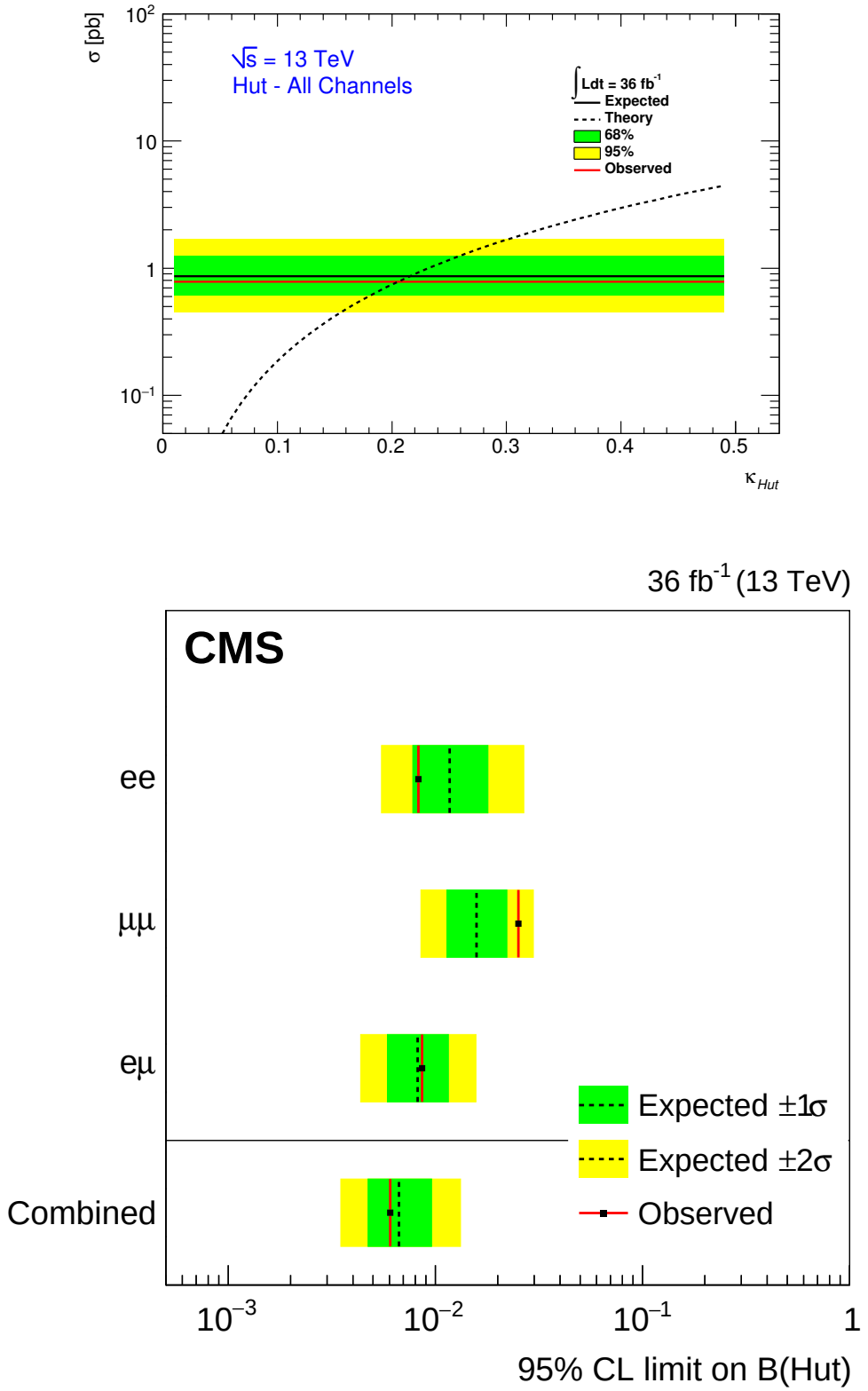


Figure 6.30: Exclusion limits at 95% CL on the coupling strenth (top) as a function of the cross section of the FCNC process. Limits on the FCNC-Hut branching ratios (bottom) for each di-lepton channel and their combination.

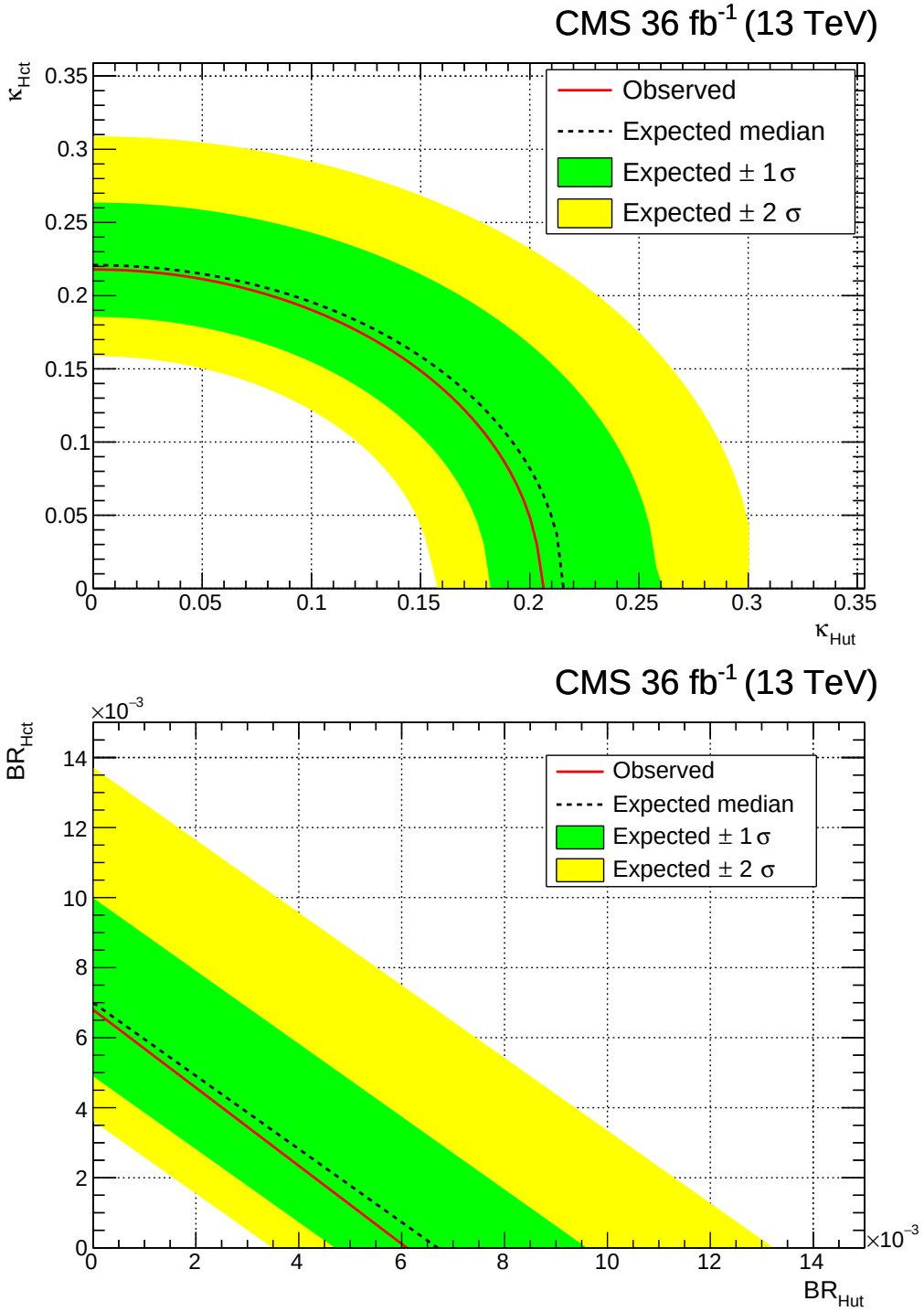


Figure 6.31: Two-dimensional limits on couplings (top) and the branching ratios (bottom) for FCNC interactions involving the Hqt couplings.

"In science we have a referee which is absolute and never makes mistakes .. That is nature."

The author

Chapter 7

Conclusion and Prospects

The physics analysis presented in this thesis is performed using data of pp-collisions collected in 2016 by the CMS experiment at a center-of-mass energy of 13 TeV and an integrated luminosity of $36 fb^{-1}$.

7.1 Conclusion

The physics analysis concerns the search for FCNC interactions involving a top quark and a Higgs boson where the Higgs boson decays into pair of W boson ($H \to WW^*$ with two same-sign leptons in the final state. In this analysis the top-quarck FCNC interactions in $t\bar{t}$-decays and in single-top quark production are considered as a signal. It is the first analysis done in RunII where the top-Higgs FCNC couplings are probed in the single top quark associated production with the Higgs boson, resulting in an improved overall sensitivity to the Hqt coupling.

The expected (observed) upper limits at 95% CL on branching ratio are $\mathcal{B}(t \to cH) \leq 7.0 \times 10^{-3}$ (6.8×10^{-3}) and $\mathcal{B}(t \to uH) \leq 6.7 \times 10^{-3}$ (6.1×10^{-3}). These limits are better than the upper limits obtained from the search with the CMS experiment using 8 TeV pp-collisions in the same-sign dilepton channel that are $\mathcal{B}(t \to cH) \leq 9.3 \times 10^{-3}$ (9.9×10^{-3}) and $\mathcal{B}(t \to uH) \leq 8.5 \times 10^{-3}$ (9.3×10^{-3}).

The upper limits are obtained by ATLAS experiment using 8 TeV were $H \to WW^*$ and $H \to \tau^+\tau^-$ probing multi-lepton at final states associated with the hadronic decay of τ. The The expected (observed) upper limits measured by the ATLAS experiment are $\mathcal{B}(t \to cH) \leq 5.3 \times 10^{-3}$ (7.9×10^{-3}) and $\mathcal{B}(t \to uH) \leq 5.3 \times 10^{-3}$ (7.8×10^{-3}).

The comparison of the expected (observed) upper limits at 95% CL obtained in this thesis with the limits from searches by both the CMS and ATLAS collaboration is shown in Figure 7.1. Our measurements are able to set stringent upper limits on the top-FCNC interactions. The expected values of upper limits are comparable with those obtained by ATLAS and CMS. Moreover the figure shows that the search for the top-FCNC presented in this thesis is expected to have an improved sensitivity if we focus as well as on other leptonic channels considering $H \to \tau^+\tau^-$ and $H \to ZZ^*$ that have multi-leptons in their final states. The figure shows also the branching ratios as predicted in the SM and some new-physics models, that are still beyond the limits presented in this thesis. However upper limits for $t \to cH$ are very close to the edge of the predictions of flavour-violating two-Higgs doublet models.

An overview of the current best observed upper limits at 95% CL, as obtained from a combination of different search channels, for all top-FCNC interactions is shown in Figure 7.2.These results at 8 TeV are compared to the limits obtained in this thesis in the $H \rightarrow WW^*$ channel with the same-sign dilepton final state analysis using 13 TeV pp-collision data. This summary figure shows that the limits from this thesis are comparable with the best limits as obtained from a combination of limits from several search channels.

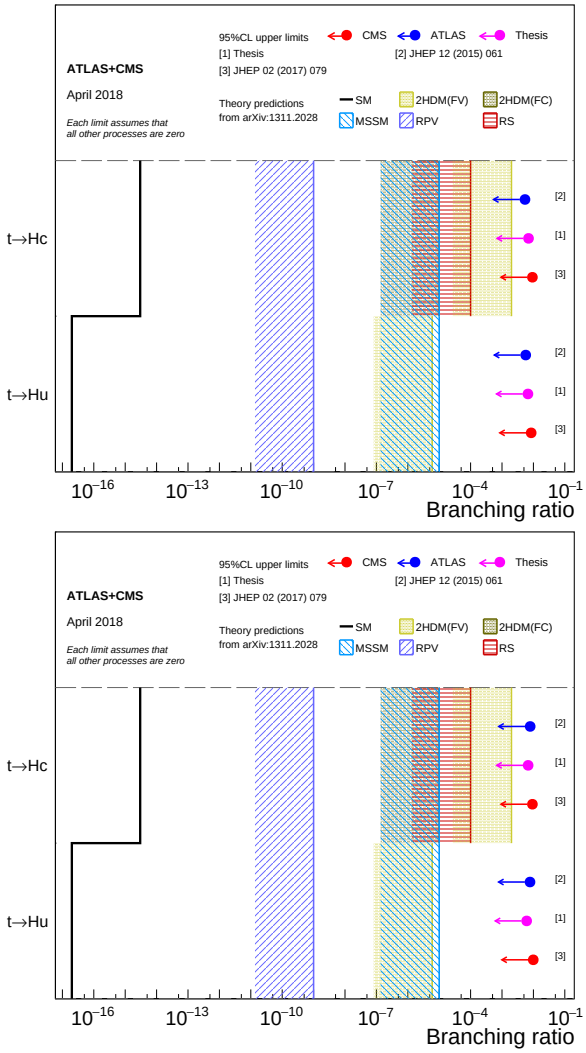


Figure 7.1: Upper expected (top) and observed (bottom) limits on branching ratios at 95% CL for top-FCNC from CMS in the same-sign dilepton channel (red) and ATLAS in multi-lepton channel (blue) at 8 TeV and the results from this thesis (magenta) at 13 TeV. A comparison is shown with SM and new-physics predictions. The figure is adapted from [28, 37]

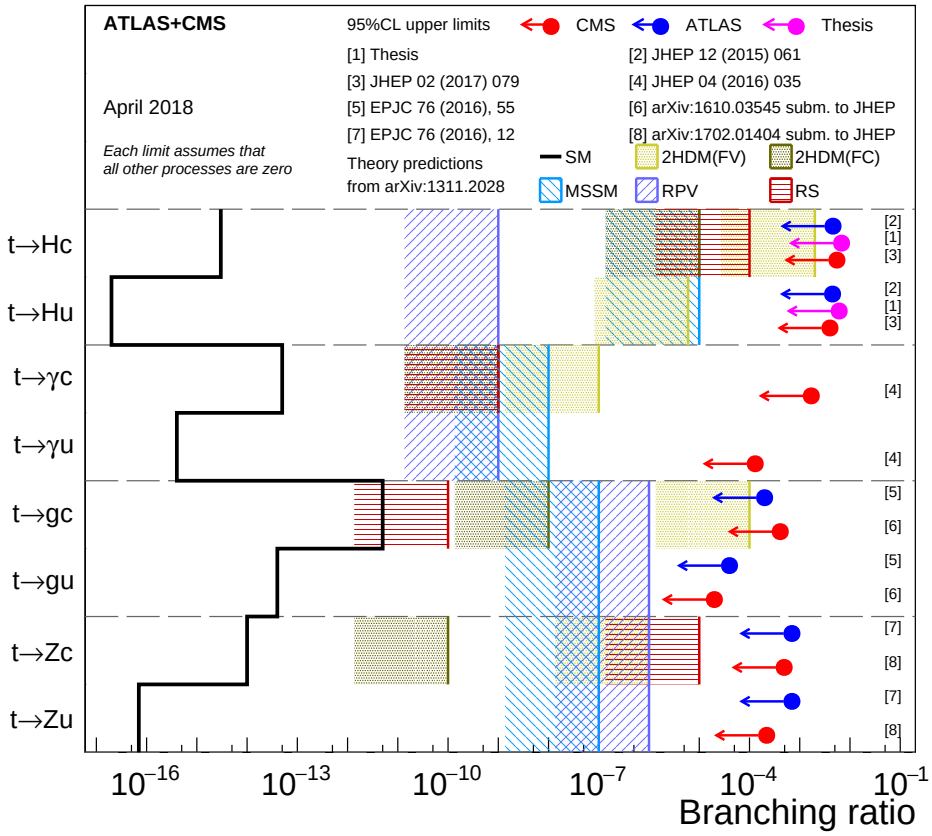


Figure 7.2: Summary of the most stringent upper limits on top-FCNC interactions at 95% CL from CMS (red) and ATLAS (blue) at a centre-of-mass energy of 8 TeV and the results from this thesis (magenta). A comparison is shown with SM and new-physics predictions. The figure is adapted from [28, 37].

7.2 Prospects

The experimental search of top-FCNC interactions involving a top quark and a Higgs boson in both the ATLAS and CMS experiments using RunII pp-collisions data at 13 TeV set stringent limits by use of two other decay-channels of Higgs bosons as illustrated in Table 7.1. In case of the $H \to \gamma\gamma$ analysis top-FCNC interactions are only considered in $t\bar{t}$ decays while for the SM-top quark in the event both leptonic and hadronic decays are considered. For the $H \to b\bar{b}$ and $H \to WW^*$ analyses top-FCNC interactions are considered in both single-top quark production and $t\bar{t}$ decays while only the leptonic decay of the SM-top quark is considered. The results motivate us to combine them in order to obtain more stringent limits on $\mathcal{B}(t \to qH)$.

The search for top-FCNC interactions presented in this thesis is statistically limited. Therefore the search for these FCNC phenomena is expected to have an improved sensitivity when performed on a larger dataset. The expected upper limits on the branching ratios at 95% CL are estimated by extrapolating the current analysis to a dataset related to a higher integrated luminosity of $100 fb^{-1}$ and $300 fb^{-1}$ as expected to be collected for full Run 2 and Run 2+ Run 3 periods respectively. We assume that the

Channel	$\mathcal{B}(t \to uH)$	$\mathcal{B}(t \to cH)$	**Experiment**
$H \to \gamma\gamma$	1.7×10^{-3}	1.6×10^{-3}	ATLAS [32]
$H \to b\bar{b}$	3.4×10^{-3}	4.7×10^{-3}	CMS [164]
$H \to WW^*$	6.1×10^{-3}	6.8×10^{-3}	CMS

Table 7.1: Summary of the most stringent upper limits on top-FCNC interactions at 95% CL from the CMS and ATLAS experiments at a centre-of-mass energy of 13 TeV.

templates for the systematic uncertainties are unchanged. The obtained expected limits together with the result obtained in the presented search for Hut and Hct couplings are shown in Figure 7.3. As can be observed the expected limit on the branching ratios for $\mathcal{L} = 100 fb^{-1}$ is improved with a factor ≈ 2 compared to the upper limits obtained form the analysis presented at this thesis for both the Hut and Hct couplings. The sensitivity is further improved for $\mathcal{L} = 300 fb^{-1}$ and so some of the beyond the Standard Model theories could be confirmed or excluded.

In the physics analysis presented in this thesis the main contribution of FCNC-signal is coming from the decay of the top quark pair. The recent progress in the developments of b and c tagging algorithms are presented during the "CMS Heavy flavour tagging workshop 2018" where the application of deep learning techniques to the identification of heavy flavour identification methods is shown [165]. The performance of c-tagging [166] should help in improving the sensitivity of the analysis to top-FCNC interactions involving the Hct coupling.

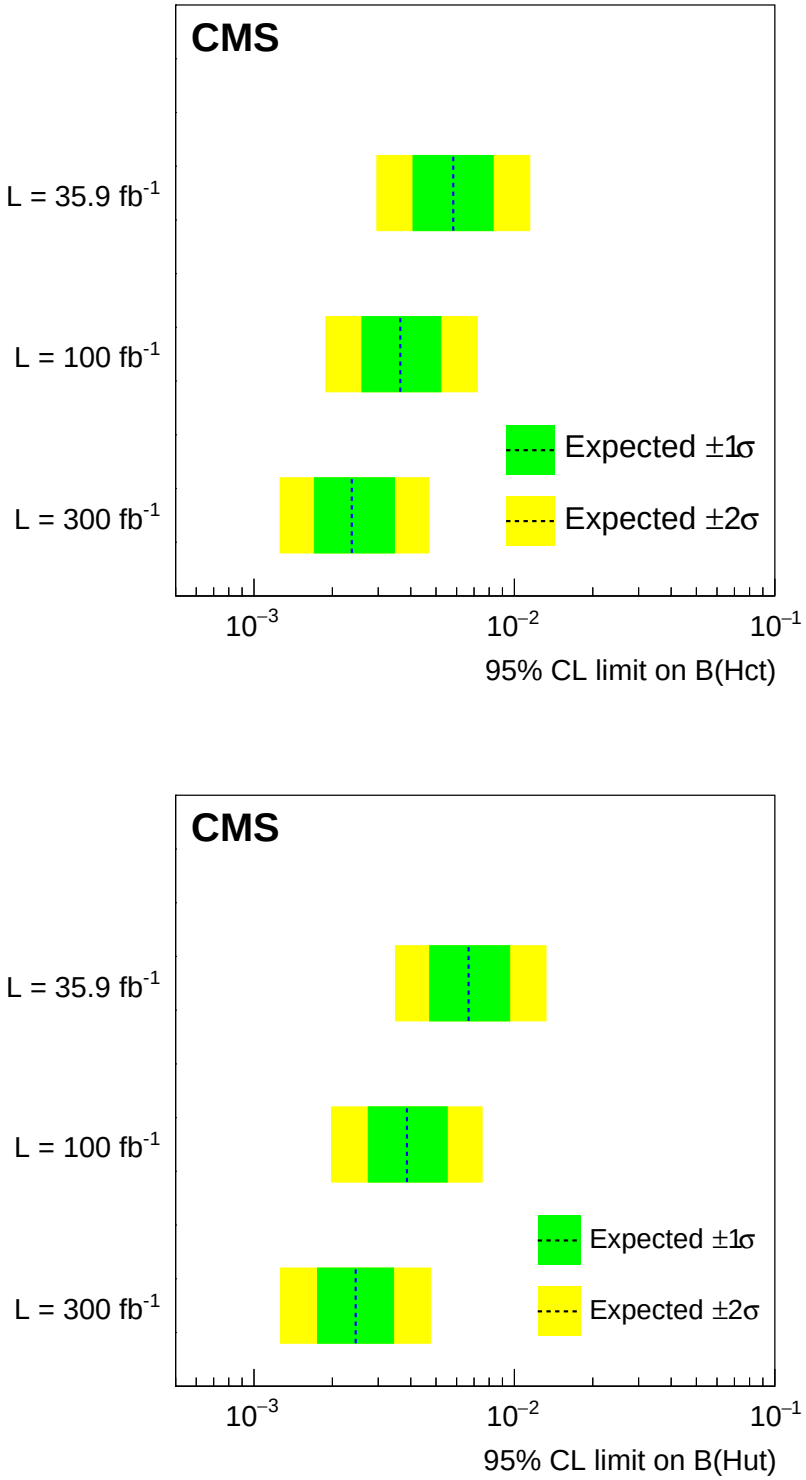


Figure 7.3: The expected upper limits on the branching ratios for top-FCNC interactions at 95% CL for Hct (top) and Hct (bottom) couplings for an integrated luminosity of 100 and 300 fb^{-1} using the same-sign dilepton analysis of this thesis.

Appendix A

Apppendix A: Variables used in b-tagging algorithms

In this appendix an overview is given of the variables used in the multivariate training of different algorithms that are used for the identification of b-quark jets. In addition to the number of selected tracks and the jet p_{T} and η, several kinematic variables related to tracks, the jet and the Secondary Vertex (SV) are defined.

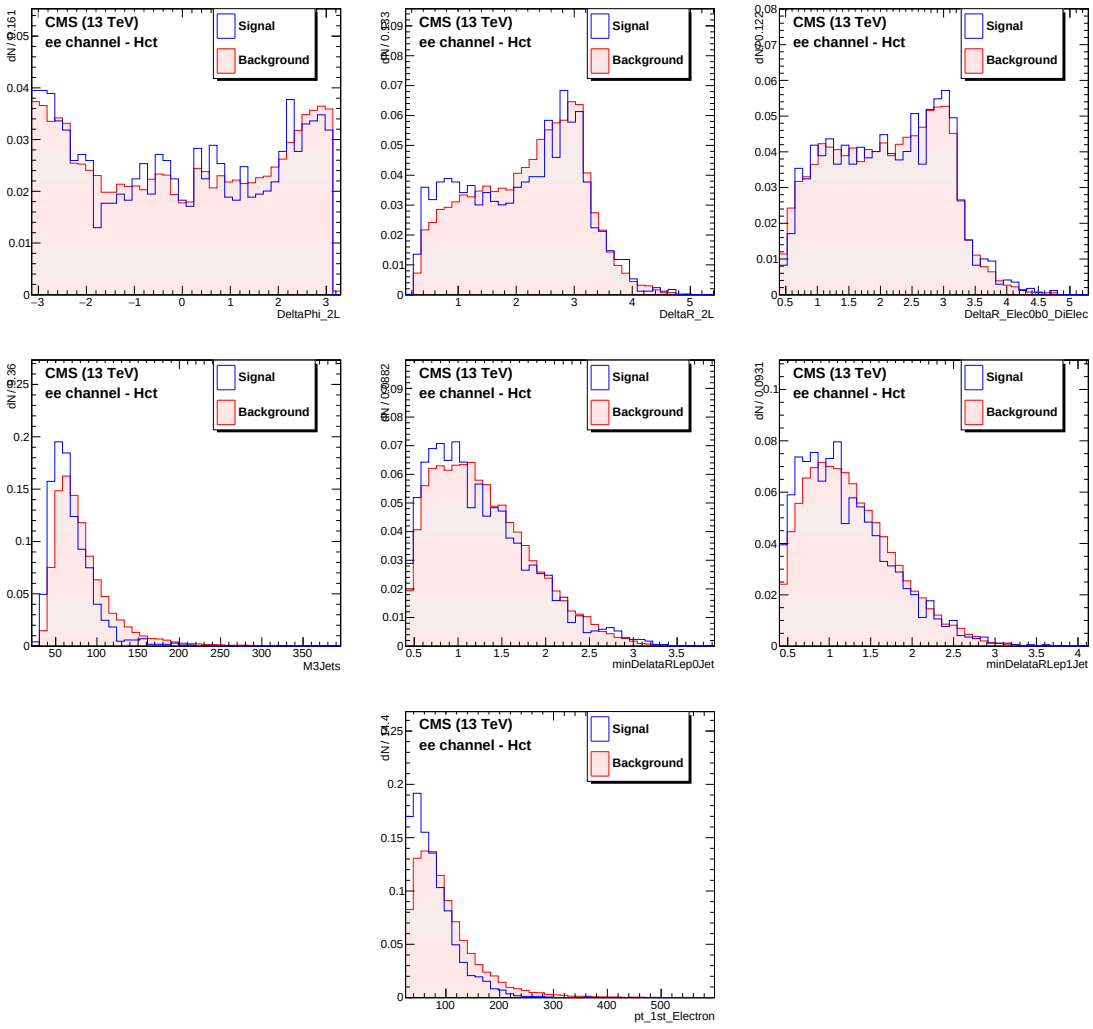
Variable	**Definition**
Number of SV	The number of Secondary Vertices for jets in the RecoVertex category.
Corrected SV mass	The mass of the SV with the smallest uncertainty on its flight distance for jets in the RecoVertex category or the invariant mass obtained from the total summed four-momentum vector of the selected tracks for jets in the PseudoVertex category.
SV 2D flight distance significance	The two-dimensional flight distance significance of the secondary vertex with the smallest uncertainty on its flight distance for jets in the RecoVertex category.
SV energy ratio	The energy of secondary vertex with the smallest uncertainty on its flight distance divided by the total summed four-momentum vector of the selected tracks.
ΔR(SV, jet)	The ΔR between the jet axis and the flight direction of secondary vertex in the RecoVertex category or the total summed four-momentum vector of the selected tracks for in the PseudoVertex category.

Variable	**Definition**
Number of tracks from SV	The number of tracks associated with the secondary vertex for jets in the RecoVertex category or the number of selected tracks for jets in the PseudoVertex category.
Track distance	The distance between the track and the jet axis at their point of closest approach.
Track decay length	The distance between the primary vertex and the track at the point of closest approach between the track and the jet axis.
summed tracks E_T ratio	The transverse energy of the total summed four-momentum vector of the selected tracks divided by the transverse energy of the jet.
ΔR(track, jet)	The ΔR between the track with highest 2D-IP significance and the jet axis.
ΔR(summed track, jet)	The ΔR between the total summed four-momentum vector of the selected tracks and the jet axis.
Track η_{rel}	The pseudorapidity of the track with the highest 2D Impact Parameter (IP) significance relative to the jet axis.
3D IP significance of the first four tracks	The signed 3D-IP significance of the four tracks with the highest 2D-IP significance.
Track $p_{T,rel}$	The track p_{T} relative it the jet axis (i.e. the track-momentum perpendicular to the jet axis).
Track $p_{T,rel}$ ratio	The track $p_{T,rel}$ divided by the track-momentum.
First track 2D-IP significance above c threshold	The 2D-IP significance of the first track that raises the combined invariant mass of the tracks above the c quark mass which is defined as 1.5 GeV.

Appendix B

Apppendix B: Distributions related to the BDT Training

In this appendix the distribution used for BDT training in the OSCR are presented. The distribution of input variables as well as the correlations between these variables are illustrated for both the Hct and Hut couplings in each di-lepton channel.



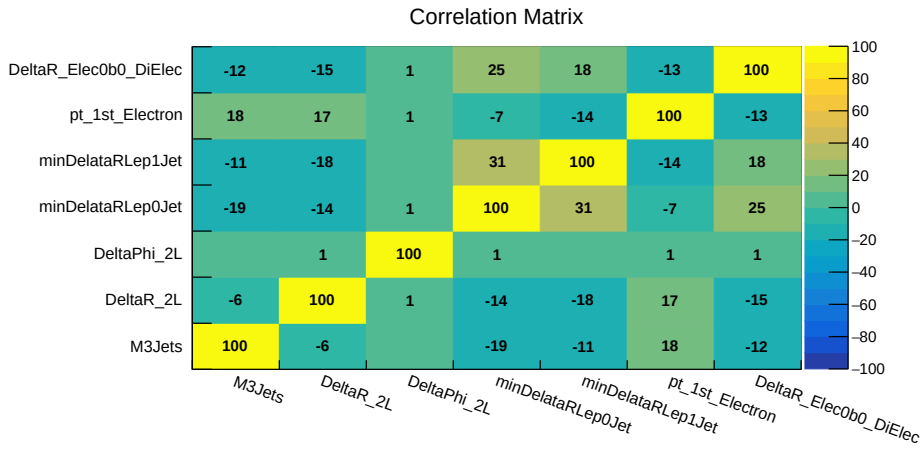
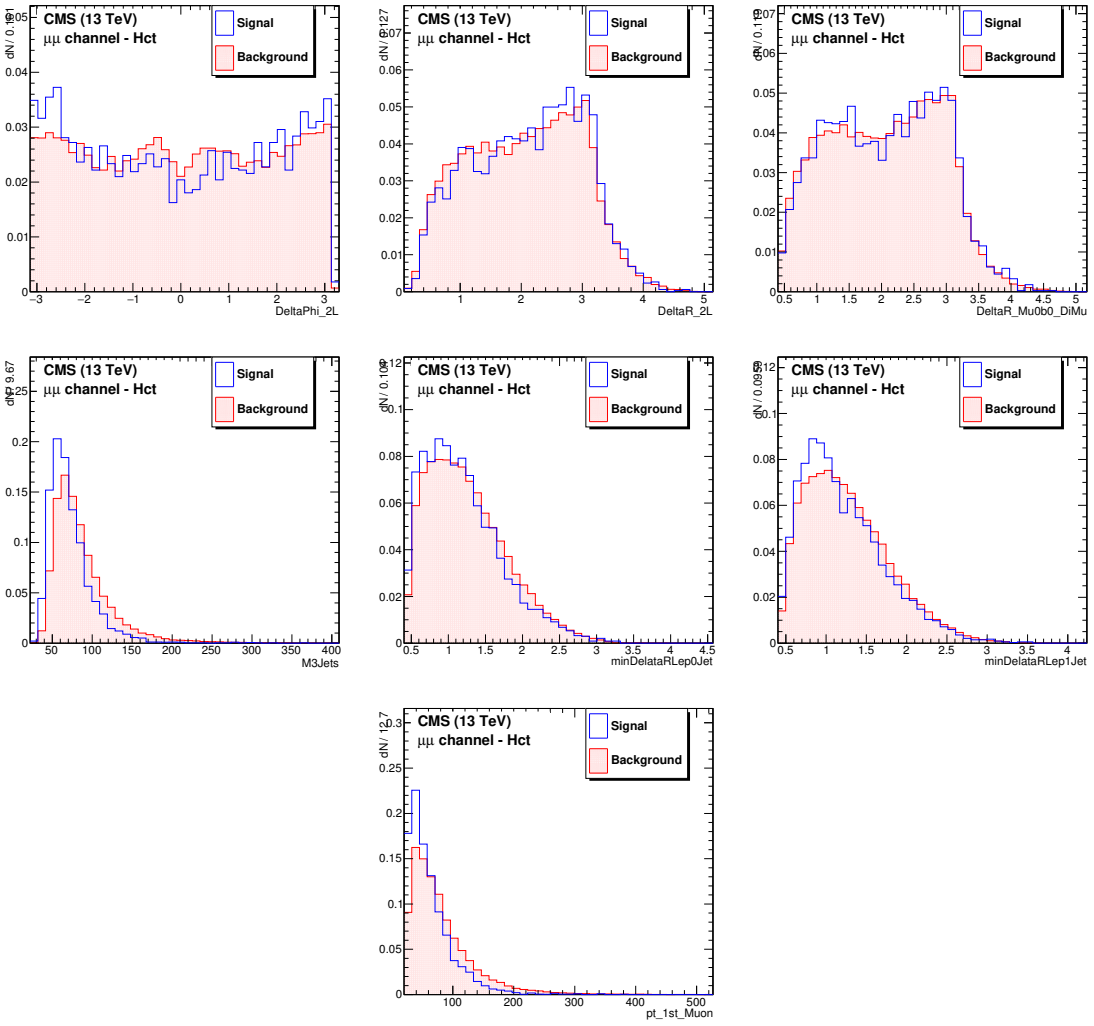


Figure B.1: The figures show the normalized distributions of the input variables for the *ee* channel (Hct coupling). The correlation between these variables is shown (bottom) for signal events.



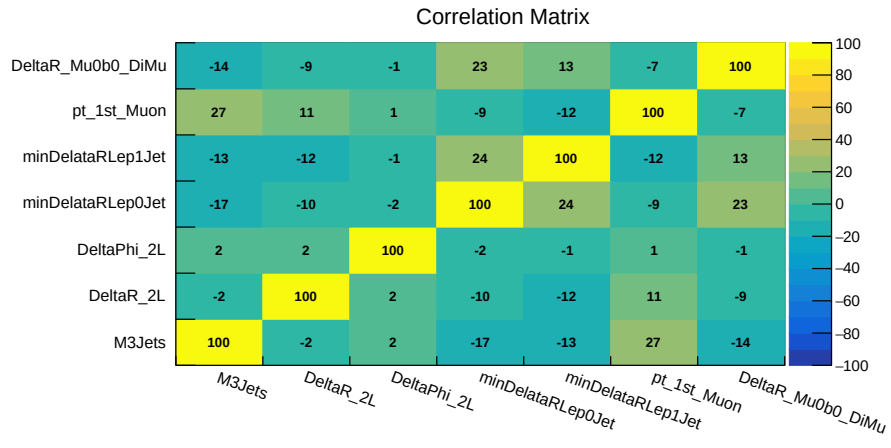
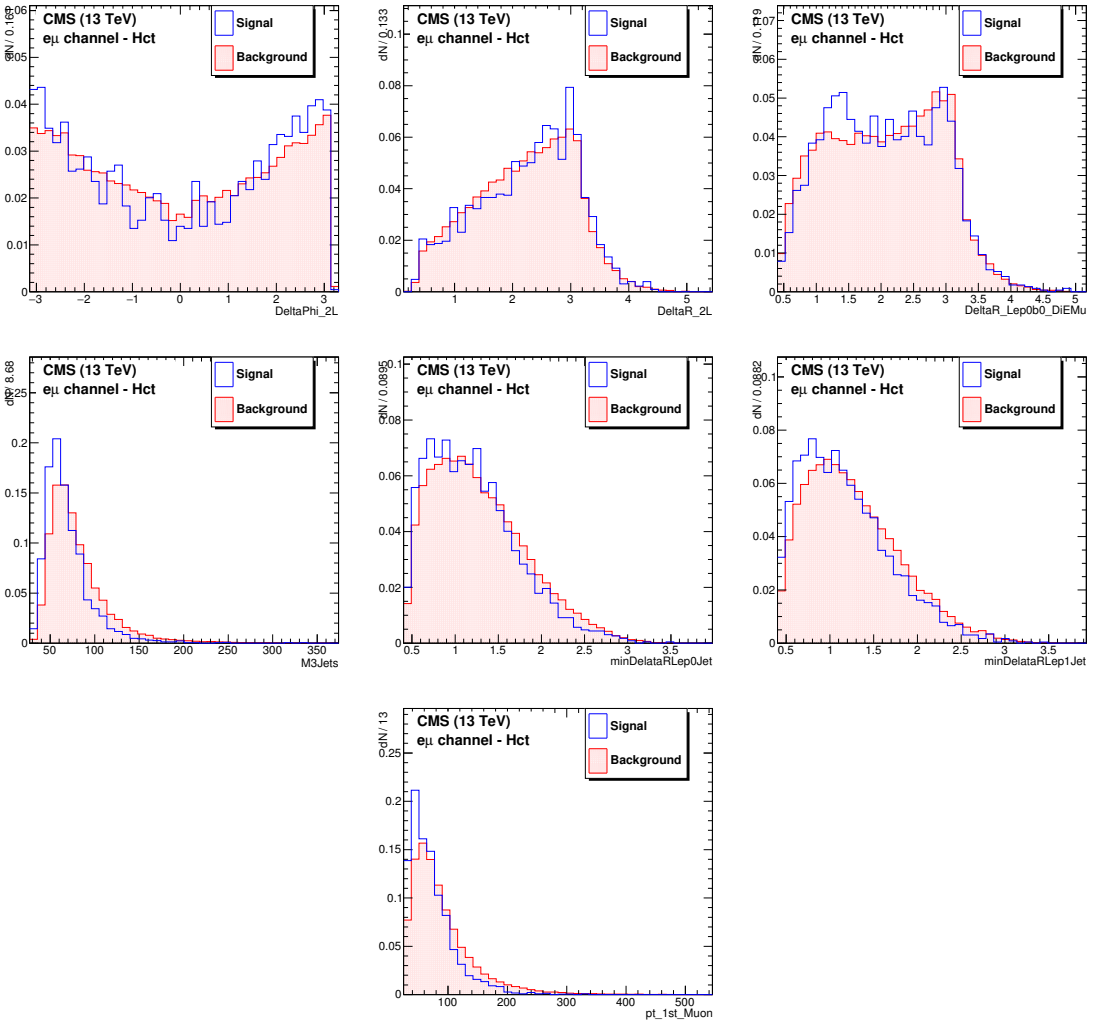


Figure B.2: The figures show the normalized distributions of the input variables for the $\mu\mu$ channel (Hct coupling). The correlation between these variables is shown (bottom) for signal events.



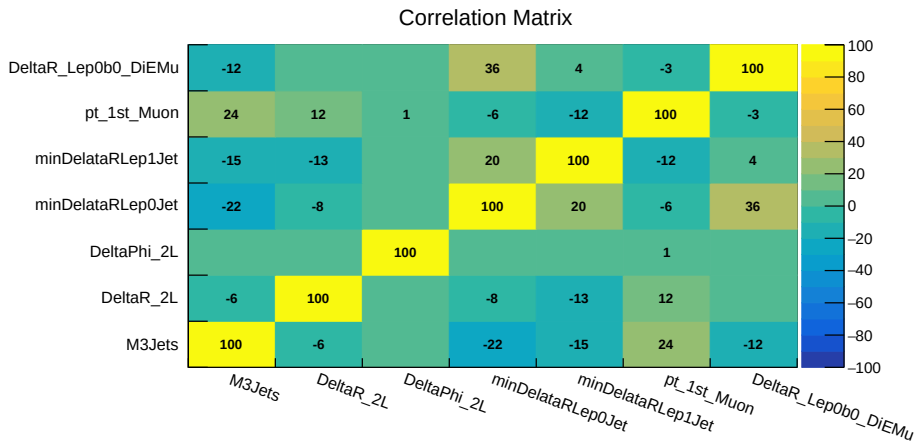
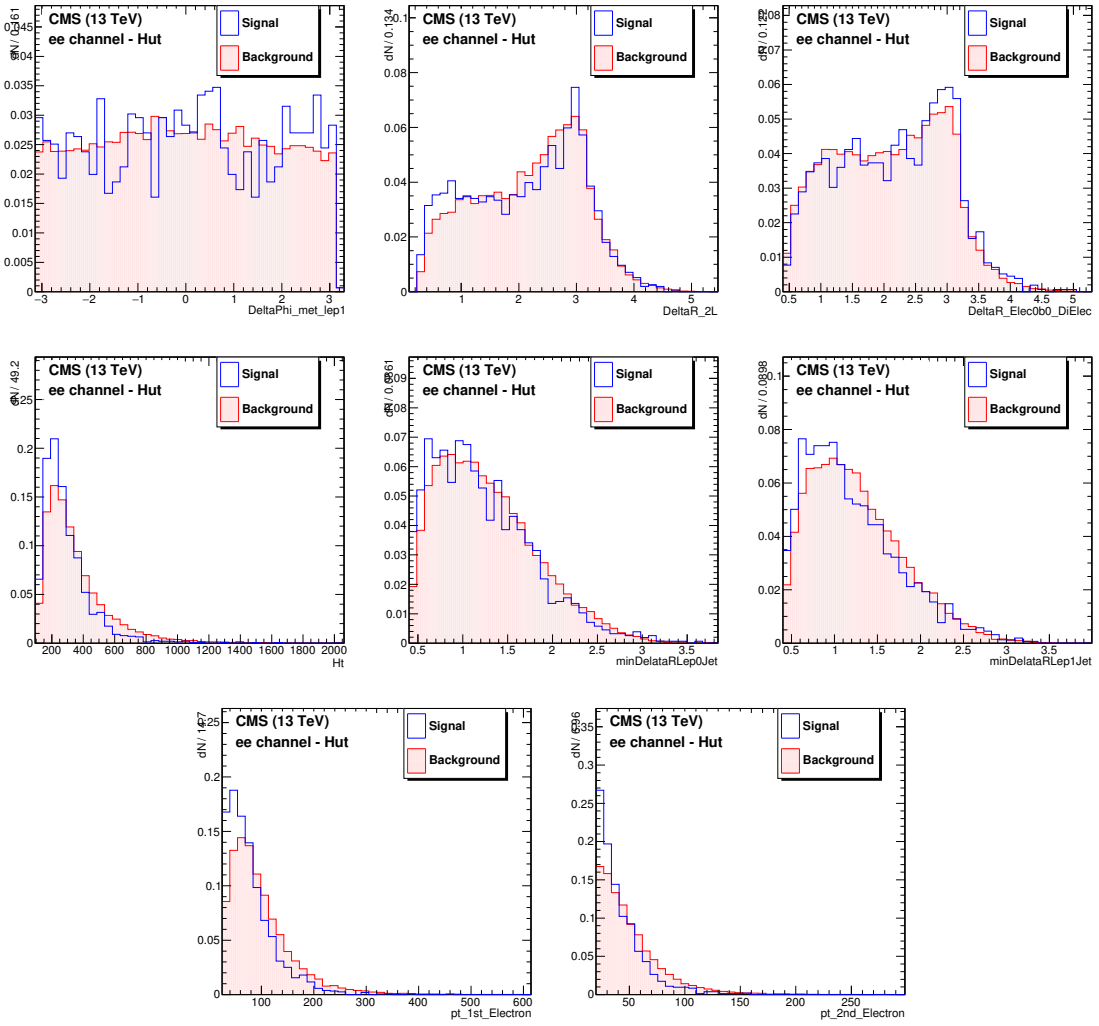


Figure B.3: The figures show the normalized distributions of the input variables for the $e\mu$ channel (Hct coupling). The correlation between these variables is shown (bottom) for signal events.



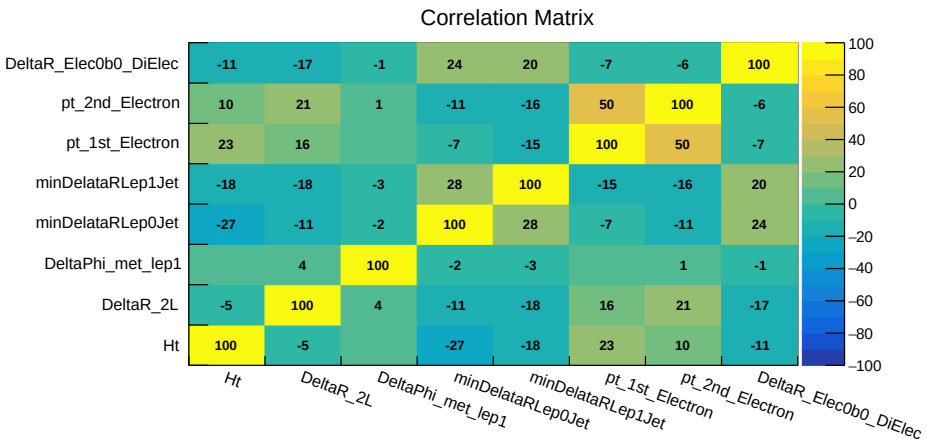
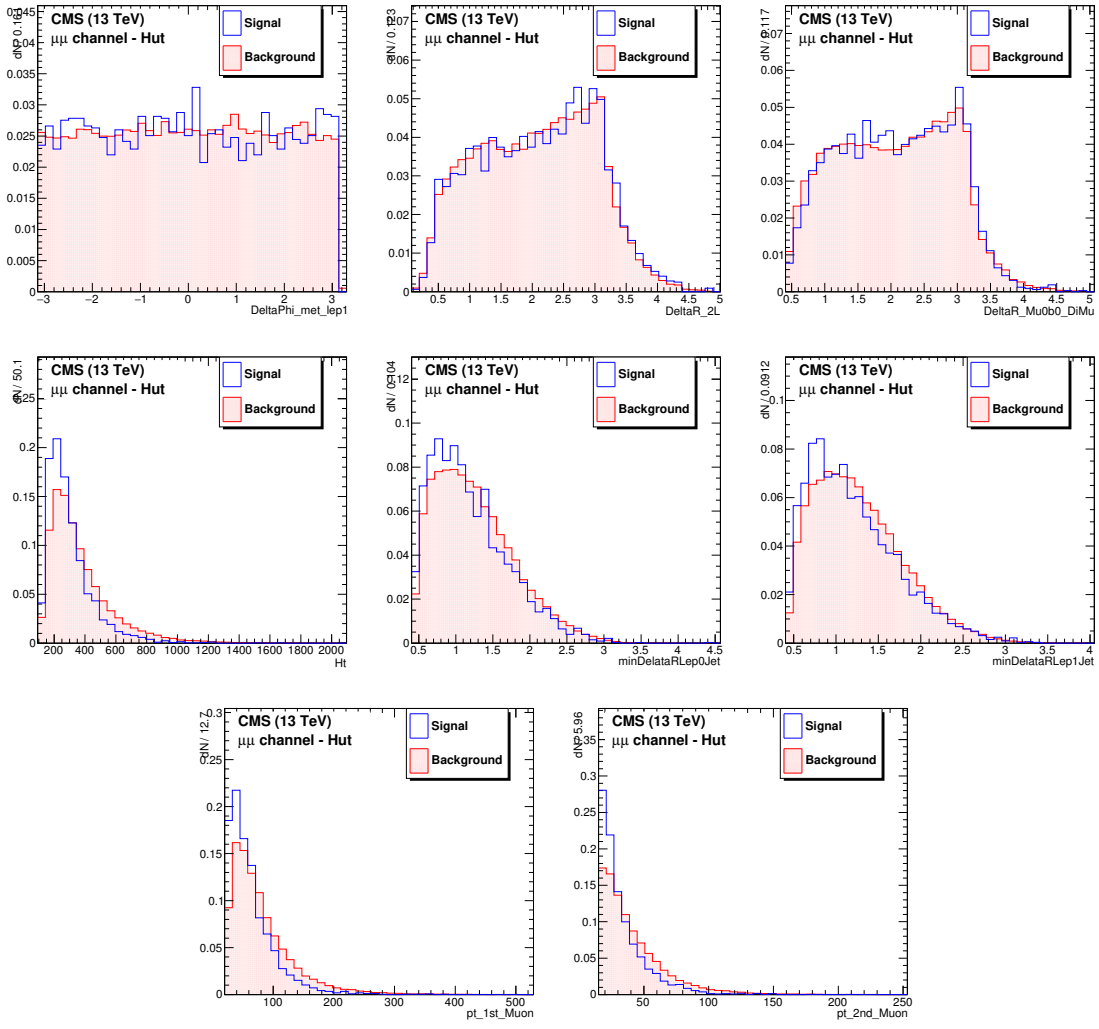


Figure B.4: The figures show the normalized distributions of the input variables for the *ee* channel (Hut coupling). The correlation between these variables is shown (bottom) for signal events.



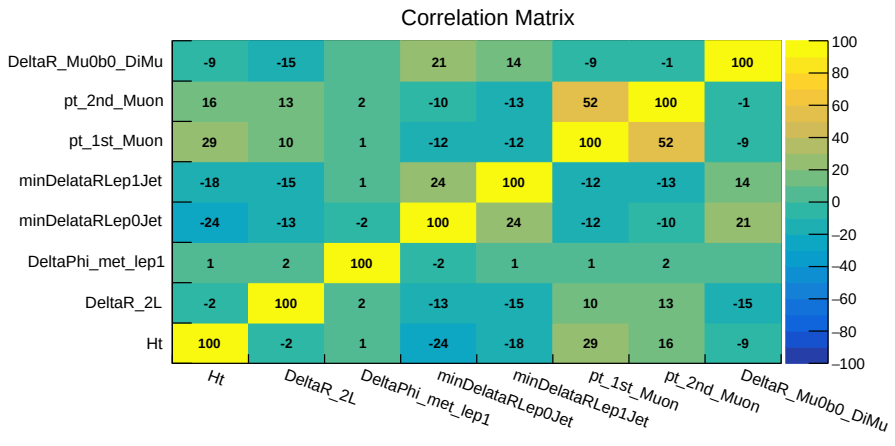
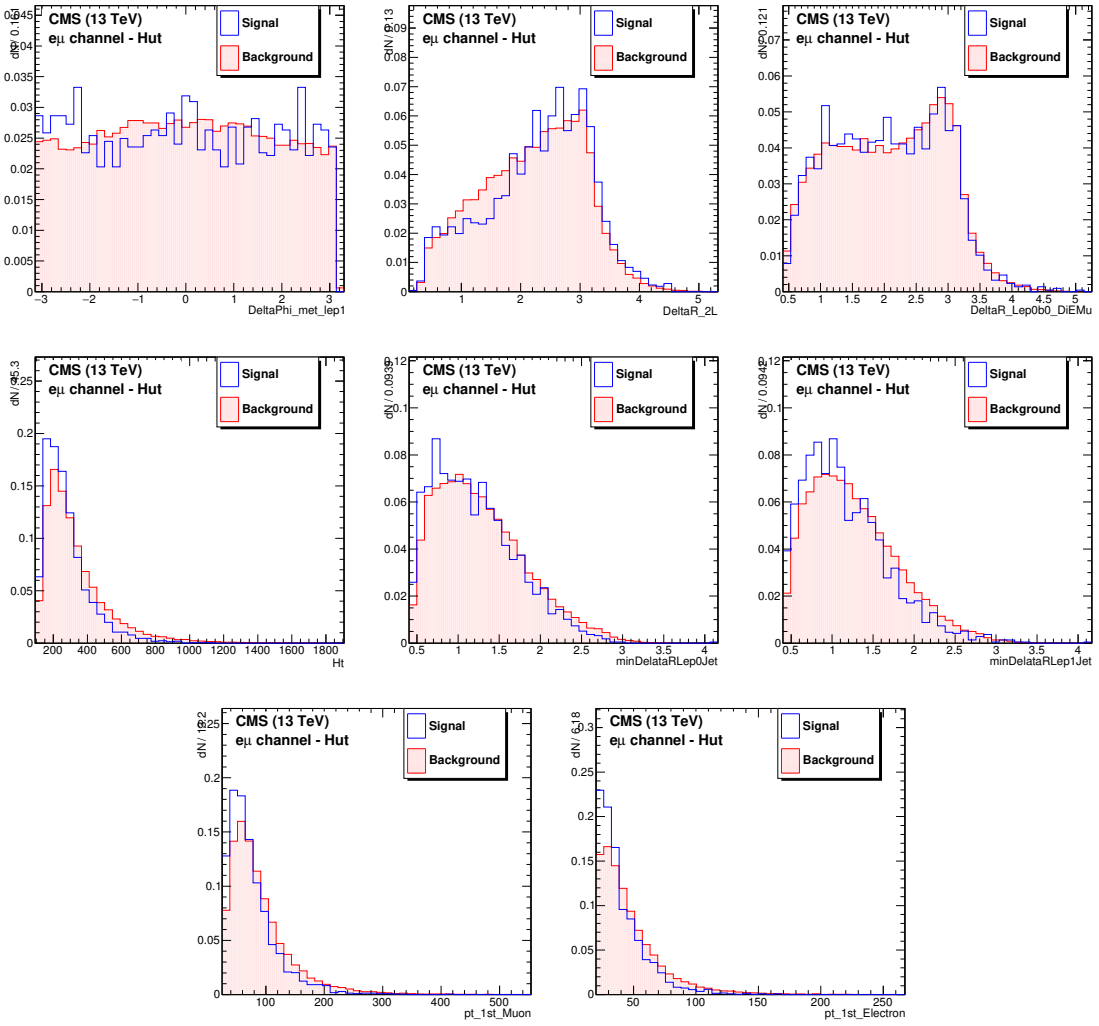


Figure B.5: The figures show the normalized distributions of the input variables for the $\mu\mu$ channel (Hut coupling). The correlation between these variables is shown (bottom) for signal events.



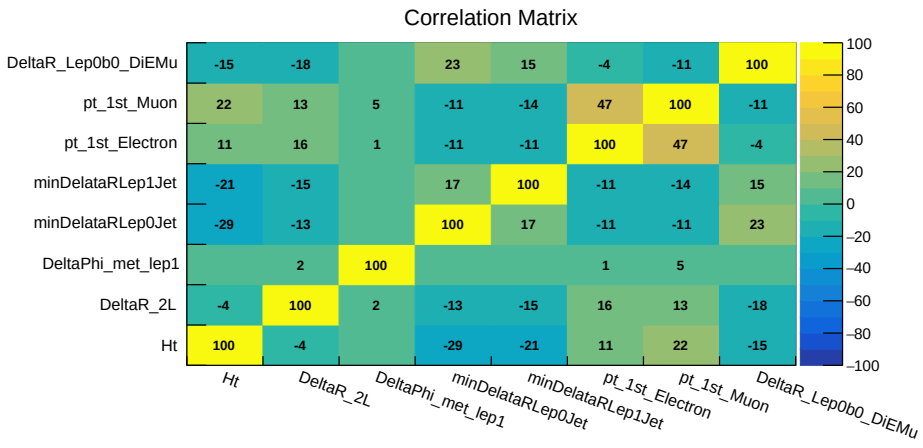


Figure B.6: The figures show the normalized distributions of the input variables for the $e\mu$ channel (Hut coupling). The correlation between these variables is shown (bottom) for signal events.

"Yesterday I was clever, so I wanted to change the world.
Today I am wise, so I am changing myself."

Jalal Ad-Din Rumi

Achievements and contributions

During the period of my doctoral studies, I was engaged in the B-Tagging and Vertexing Physics Object Group (BTV-POG) of the CMS collaboration and have I contributed in the development of the algorithms that are used for the reconstruction of bottom (b) quarks. These techniques help in the identification of the underlying flavour of the quark, which is at the origin of observed hadronic jets. In this part of the project I have contributed in the development and testing of the CSVv2 algorithm that is widely used in different physics analyses in the CMS Collaboration. I have regularly presented progress reports in bi-weekly BTV meetings at CERN. This part has been published as *"Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV"* by the CMS Collaboration [127]. Since the top quark, according to Standard model (SM), decays for almost 100% to a b quark and a W boson, the b-tagging technique is crucial in all top-quark research analyses including the search for Flavour Changing Neutral Current (FCNC) decays of the top quark. A complete list of peer reviewed publications where I am co-author can be found here `https://inspirehep.net/search?ln=en&ln=en&p=shimaa+abu+zeid&of=hb&action_search=Search&sf=earliestdate&so=d&rm=&rg=25&sc=0`.

I have contributed in a phenomenological study of top-FCNC topologies in LHC Run 2 collisions. This study was performed in collaboration between research groups from the Vrije Universiteit Brussel, the Institut Pluridisciplinaire Hubert CURIEN at Strasbourg (France), and the Chonbuk National University in South-Korea. My main contribution was the search for the top-FCNC topology with two same-sign electric charged leptons in the final state. Proceeding on this collaboration I was one of the speakers at the "Brussels-Strasbourg-CBNU top quark physics workshop on FCNC", that was held from 28th to 30th May 2015 at Chonbuk N. University. I presented an overview entitled *"Flavor Changing Neutral Currents in top pair decay with same sign di-lepton final state"*. Also I have participated in the "Top FCNC mini-workshop 2016" that was held on 21-22 January 2016 at Eguisheim (France), where I have presented *"Flavor Changing Neutral Currents in the decay of top-pair to Higgs and (u/c) quark"*.

The research of top-FCNC presented in this thesis is performed using data from pp-collisions at 13 TeV. Also the physics analysis described in thesis has been endorsed by the CMS collaboration.

As part of my research I attended some workshops. I have attended the Belgian Dutch German summer schools (BND) in 2014 and in 2015. I participated in the student sessions where specific projects were studied and I presented the results of the study. Also I have attended the "7th International workshop on Top Quark Physics (Top2014)" in Cannes (France), and the "9th International Workshop on Top Quark

Physics (TOP 2016)" in Olomouc (Czech Republic).

In the CMS collaboration, it is common to take on experimental physics responsibilities (EPR). I had the opportunity to do some shifts in the control room of the CMS experiment. Being a Detector Control and Safety (DCS) technical shifter means I am the responsible person for controlling the CMS sub-systems and the connection between CMS and the LHC machine during the operation period. During night shifts and weekends safety tours are included in/outside the CMS experiment building in order to assure safety conditions during the activities. Additionally, for my work in the CMS Collaboration, I have followed several safety-training courses to provide a safer working environment in CMS. Lately I have followed a training course to be member of the tour guide team at CMS.

Bibliography

[1] CMS Collaboration, CERN, "A beautiful barrel for CMS", Mar, 2014. http://cds.cern.ch/record/1998635.

[2] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry", *Phys. Rev. D* **2** (Oct, 1970) 1285–1292, doi:10.1103/PhysRevD.2.1285.

[3] CDF Collaboration, "First 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, doi:10.1103/PhysRevLett.95.031801, arXiv:hep-ex/0502044.

[4] LHCb, CMS Collaboration, "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, doi:10.1038/nature14474, arXiv:1411.4413.

[5] 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, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.

[6] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* **C680519** (1968) 367–377. "

[7] M.Thomson, "Modern Particle Physics". 2013.

[8] CERN, "The Standard Model and beyond". https://cds.cern.ch/record/1473657/files/SMinfographic_image.png?subformat=, 2017. Accessed: 2018-01-16.

[9] J. Woithe, G. J. Wiener, and F. F. Van der Veken, "Let's have a coffee with the standard model of particle physics!", *Physics Education* **52** (2017) 034001.

[10] F. Halzen, "Quarks and Leptons: An introductory course in Modern Particle Physics". 1984.

[11] S. Weinberg, "A model of leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[12] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[13] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[14] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[15] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C40** (2016) 100001, doi:10.1088/1674-1137/40/10/100001.

[16] L. Maiani, "The GIM Mechanism: origin, predictions and recent uses", in *Proceedings, 48th Rencontres de Moriond on Electroweak Interactions and Unified Theories: La Thuile, Italy, March 2-9, 2013*, pp. 3–16. 2013. arXiv:1303.6154.

[17] J. E. Augustin et al., "Discovery of a narrow resonance in e^+e^- annihilation", *Phys. Rev. Lett.* **33** (Dec, 1974) 1406–1408, doi:10.1103/PhysRevLett.33.1406.

[18] J. J. Aubert et al., "Experimental observation of a heavy particle j", *Phys. Rev. Lett.* **33** (Dec, 1974) 1404–1406, doi:10.1103/PhysRevLett.33.1404.

[19] P. J. E. Peebles and B. Ratra, "The Cosmological constant and dark energy", *Rev. Mod. Phys.* **75** (2003) 559–606, doi:10.1103/RevModPhys.75.559, arXiv:astro-ph/0207347. [,592(2002)].

[20] Super-Kamiokande Collaboration Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562.

[21] Double Chooz Collaboration Collaboration, "Indication of reactor $\overline{\nu}_e$ disappearance in the double chooz experiment", *Phys. Rev. Lett.* **108** (Mar, 2012) 131801, doi:10.1103/PhysRevLett.108.131801.

[22] A. D. Sakharov, "Violation of cp-invariance, c-asymmetry, and baryon asymmetry of the universe", in *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16*, pp. 84–87. World Scientific, 1998.

[23] P. Langacker, "Grand unified theories and proton decay", *Physics Reports* **72** (1981) 185–385.

[24] G. Burdman, "New solutions to the hierarchy problem", *Braz. J. Phys.* **37** (2007) 506–513, doi:10.1590/S0103-97332007000400006, arXiv:hep-ph/0703194.

[25] M. Kobayashi and T. Maskawa, "Cp-violation in the renormalizable theory of weak interaction", *Progress of Theoretical Physics* **49** (1973) 652–657.

[26] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.

[27] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

[28] LHC-Top-Physics-Working-Group, "LHC Top WG Summary Plots". Accessed on 13/04/2018. https://twiki.cern.ch/twiki/pub/LHCPhysics/LHCTopWGSummaryPlots/fcnc_summarybsm_nov17.pdf.

[29] J. A. Aguilar-Saavedra, "Top flavor-changing neutral interactions: Theoretical expectations and experimental detection", *Acta Phys. Polon.* **B35** (2004) 2695–2710, arXiv:hep-ph/0409342.

[30] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-Six Terms in the Standard Model Lagrangian", *JHEP* **10** (2010) 085, doi:10.1007/JHEP10(2010)085, arXiv:1008.4884.

[31] 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) 042001, doi:10.1103/PhysRevLett.110.042001, arXiv:1210.2808.

[32] ATLAS Collaboration, "Search for top quark decays $t \to qH$, with $H \to \gamma\gamma$, in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector", *JHEP* **10** (2017) 129, doi:10.1007/JHEP10(2017)129, arXiv:1707.01404.

[33] ATLAS Collaboration Collaboration, "Search for flavour-changing neutral current top quark decays $t \to qZ$ in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector", Technical Report ATLAS-CONF-2017-070, CERN, Geneva, Sep, 2017.

[34] 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, doi:10.1007/JHEP04(2016)035, arXiv:1511.03951.

[35] CMS Collaboration, "Search for anomalous Wtb couplings and flavour-changing neutral currents in t-channel single top quark production in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **02** (2017) 028, doi:10.1007/JHEP02(2017)028, arXiv:1610.03545.

[36] 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) 55, doi:10.1140/epjc/s10052-016-3876-4, arXiv:1509.00294.

[37] CMS-Top-Quark-Physics-Group, "Top Quark Physics Summary Figures". Accessed on 13/04/2018. https://twiki.cern.ch/twiki/pub/CMSPublic/PhysicsResultsTOPSummaryFigures/fcnc_upperlimits.pdf.

[38] S. Myers, "The LEP Collider, from design to approval and commissioning". John Adams' Lecture. CERN, Geneva, 1991. Delivered at CERN, 26 Nov 1990.

[39] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[40] E. Mobs, "The CERN accelerator complex. Complexe des accélérateurs du CERN",. General Photo.

[41] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[42] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[43] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[44] J. L. Pinfold, "The MoEDAL Experiment at the LHC – a New Light on the Terascale Frontier", *J. Phys. Conf. Ser.* **631** (2015) 012014, doi:10.1088/1742-6596/631/1/012014.

[45] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006, doi:10.1088/1748-0221/3/08/S08006.

[46] TOTEM Collaboration, "The TOTEM experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007, doi:10.1088/1748-0221/3/08/S08007.

[47] J. Salfeld, "CMS Luminosity - Public Results - 2016 Proton-Proton 13 TeV Collisions". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`, 2017. Accessed: 2017-05-16.

[48] C. Collaboration, "Detector Drawings", (Mar, 2012). CMS Collection.

[49] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, `arXiv:1405.6569`.

[50] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[51] CMS Tracker Collaboration, "Plans and Status of the Phase I Upgrade of the CMS Pixel Tracker", *PoS* **Vertex2013** (2013) 014.

[52] CMS Collaboration, "Construction and commissioning of the Phase I upgrade of the CMS pixel detector", *PoS* **EPS-HEP2017** (2017) 479, doi:10.22323/1.314.0479.

[53] CMS Collaboration, "CMS inner detector: the Run 1 to Run 2 transition and first experience of Run 2", *PoS* **VERTEX2015** (2015) 002.

[54] CMS Collaboration, "Alignment of the CMS Tracker: Latest results from LHC Run-II", *J. Phys. Conf. Ser.* **898** (2017), no. 4, 042014, doi:10.1088/1742-6596/898/4/042014.

[55] R. E. Kalman et al., "A new approach to linear filtering and prediction problems", *Journal of basic Engineering* **82** (1960), no. 1, 35–45.

[56] P. Billoir, "Progressive track recognition with a Kalman like fitting procedure", *Comput. Phys. Commun.* **57** (1989) 390–394, doi:10.1016/0010-4655(89)90249-X.

[57] E. J. Wolin and L. L. Ho, "Covariance matrices for track fitting with the Kalman filter", *Nucl. Instrum. Meth.* **A329** (1993) 493–500, doi:10.1016/0168-9002(93)91285-U.

[58] S. Cucciarelli, M. Konecki, D. Kotlinski, and T. Todorov, "Track reconstruction, primary vertex finding and seed generation with the pixel detector",.

[59] J. R. Cash and A. H. Karp, "A variable order runge-kutta method for initial value problems with rapidly varying right-hand sides", *ACM Trans. Math. Softw.* **16** (September, 1990) 201–222, doi:10.1145/79505.79507.

[60] C. Biino, "The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections", *J. Phys. Conf. Ser.* **587** (2015) 012001, doi:10.1088/1742-6596/587/1/012001.

[61] CMS Collaboration, "Precision crystal calorimetry in LHC Run II with the CMS ECAL", *JINST* **12** (2017) C01069, doi:10.1088/1748-0221/12/01/C01069.

[62] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[63] CMS Collaboration, "CMS Physics",.

[64] CMS Collaboration, "Performance and track-based alignment of the upgraded CMS pixel detector", *PoS* **EPS-HEP2017** (2017) 484, doi:10.22323/1.314.0484.

[65] CMS Collaboration, "Identification and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter", *JINST* **5** (2010) T03014, doi:10.1088/1748-0221/5/03/T03014, arXiv:0911.4881.

[66] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", doi:10.2172/1151651.

[67] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC", *JINST* **8** (2013) P11002, doi:10.1088/1748-0221/8/11/P11002, arXiv:1306.6905.

[68] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[69] CMS Collaboration, "CMS reconstruction improvement for the muon tracking by the RPC chambers", *PoS* **RPC2012** (2012) 045, doi:10.1088/1748-0221/8/03/T03001, arXiv:1209.2646. [JINST8,T03001(2013)].

[70] F. Gasparini, "The CMS muon detector: From the first thoughts to the final design", *Int. J. Mod. Phys.* **A25** (2010) 3121–3154, doi:10.1142/S0217751X10050238. [,351(2010)].

[71] CMS Collaboration, "CMS muon system towards LHC Run 2 and beyond", *Nucl. Part. Phys. Proc.* **273-275** (2016) 1014–1022, doi:10.1016/j.nuclphysbps.2015.09.159.

[72] C. Grandi et al., "CMS Computing Model: The "CMS Computing Model RTAG"", technical report, CERN, Geneva, Dec, 2004.

[73] W. L. C. Grid, "WorldWide LHC Computing Grid". http://wlcg.web.cern.ch/, 2018. Accessed: 2018-04-16.

[74] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[75] M. A. Dobbs et al., "Les Houches guidebook to Monte Carlo generators for hadron collider physics", in *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 26-June 3, 2003*, pp. 411–459. 2004. arXiv:hep-ph/0403045.

[76] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, doi:10.1142/9789814503266-0001, arXiv:hep-ph/0409313.

[77] J. Rojo, "Parton distributions from high-precision collider data", 2017. http://pcteserver.mi.infn.it/~nnpdf/talks/rojo-VUB-NNPDF31.pdf.

[78] H. Abramowicz and A. Caldwell, "HERA collider physics", *Rev. Mod. Phys.* **71** (1999) 1275–1410, doi:10.1103/RevModPhys.71.1275, arXiv:hep-ex/9903037.

[79] H. Lai et al., "New parton distributions for collider physics", *Phys. Rev.* **D82** (2010) 074024, doi:10.1103/PhysRevD.82.074024, arXiv:1007.2241.

[80] S. Holmes, R. S. Moore, and V. Shiltsev, "Overview of the Tevatron Collider Complex: Goals, Operations and Performance", *JINST* **6** (2011) T08001, doi:10.1088/1748-0221/6/08/T08001, arXiv:1106.0909.

[81] J. Rojo et al., "The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II", *J. Phys.* **G42** (2015) 103103, doi:10.1088/0954-3899/42/10/103103, arXiv:1507.00556.

[82] A. D. Martin, "Proton structure, Partons, QCD, DGLAP and beyond", *Acta Phys. Polon.* **B39** (2008) 2025–2062, arXiv:0802.0161.

[83] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *J. Phys.* **G43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.

[84] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[85] S. Dulat et al., "New parton distribution functions from a global analysis of quantum chromodynamics", *Phys. Rev.* **D93** (2016) 033006, doi:10.1103/PhysRevD.93.033006, arXiv:1506.07443.

[86] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, "Parton distributions in the LHC era: MMHT 2014 PDFs", *Eur. Phys. J.* **C75** (2015) 204, doi:10.1140/epjc/s10052-015-3397-6, arXiv:1412.3989.

[87] J. Pumplin et al., "Uncertainties of predictions from parton distribution functions. 2. The Hessian method", *Phys. Rev.* **D65** (2001) 014013, doi:10.1103/PhysRevD.65.014013, arXiv:hep-ph/0101032.

[88] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J.* **C77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[89] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[90] 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", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[91] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[92] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.

[93] T. Sjostrand, "Jet Fragmentation of Nearby Partons", *Nucl. Phys.* **B248** (1984) 469–502, doi:10.1016/0550-3213(84)90607-2.

[94] B. Andersson, S. Mohanty, and F. Soderberg, "Recent developments in the Lund model", in *36th Annual Winter School on Nuclear and Particle Physics (PINP 2002) and 8th St. Petersburg School on Theoretical Physics St. Petersburg, Russia, February 25-March 3, 2002*. 2002. arXiv:hep-ph/0212122.

[95] CDF Collaboration, "The Underlying event in hard scattering processes", *eConf* **C010630** (2001) P501, arXiv:hep-ph/0201192.

[96] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[97] T. Sjostrand, "Monte Carlo Tools", in *Proceedings, 65th Scottish Universities Summer School in Physics: LHC Physics (SUSSP65): St. Andrews, UK, August 16-29, 2009*, pp. 309–339. 2009. arXiv:0911.5286. doi:10.1201/b11865-14.

[98] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[99] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.

[100] C. Oleari, "The POWHEG-BOX", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 36–41, doi:10.1016/j.nuclphysbps.2010.08.016, arXiv:1007.3893.

[101] T. Sjostrand et al., "High-energy physics event generation with PYTHIA 6.1", *Comput. Phys. Commun.* **135** (2001) 238–259, doi:10.1016/S0010-4655(00)00236-8, arXiv:hep-ph/0010017.

[102] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[103] 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.

[104] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[105] Acronym-Finder, "The Coordinated Theoretical-Experimental Project on QCD". Accessed: 2017-12-16. https://www.acronymfinder.com/Co_ordinated-Theoretical_Experimental-Project-on-QCD-(CTEQ).html.

[106] 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, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[107] M. Aliev et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034–1046, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.

[108] P. Uwer, "Hadronic single top-quark production: updated predictions and uncertainty estimates", *PoS* **LHCP2016** (2016) 166.

[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, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

[110] J. M. Campbell, R. K. Ellis, and W. T. Giele, "A Multi-Threaded Version of MCFM", *Eur. Phys. J.* **C75** (2015) 246, doi:10.1140/epjc/s10052-015-3461-2, arXiv:1503.06182.

[111] A. Alloul et al., "FeynRules 2.0 - A complete toolbox for tree-level phenomenology", *Comput. Phys. Commun.* **185** (2014) 2250–2300, doi:10.1016/j.cpc.2014.04.012, arXiv:1310.1921.

[112] C. Degrande et al., "UFO - The Universal FeynRules Output", *Comput. Phys. Commun.* **183** (2012) 1201–1214, doi:10.1016/j.cpc.2012.01.022, arXiv:1108.2040.

[113] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[114] CERN, "Lhc p collisions", 2017. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_p_collisions.

[115] CERN, "Detectors - taking a closer look at lhc", 2017. http://www.lhc-closer.es/taking_a_closer_look_at_lhc/1.detectors.

[116] C. collaboration et al., "Performance of cms muon reconstruction in pp collision events at $\sqrt{S} = 7\,\text{TeV}$", *Journal of Instrumentation* **7** (2012) P10002.

[117] Muon-Physics-Object-Group, "Reference muon id, isolation and trigger efficiencies for RunII". Accessed on 13/12/2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2.

[118] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{S} = 8\,\text{TeV}$", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[119] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC",.

[120] E/gamma-Physics-Object-Group, "Cut Based Electron Identification for Run2". Accessed on15/12/2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2.

[121] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", *Journal of High Energy Physics* **2008** (2008) 063.

[122] JetMET-Physics-Object-Group, "Jet Identification". Accessed on 21/01/2018. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID#Recommendations_for_13_TeV_2016.

[123] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[124] J.-E.-R. and Corrections-Subgroup, "Jet Energy Resolution". Accessed on15/12/2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution.

[125] J.-E.-R. and Corrections-Subgroup, "JER Scaling factors and Uncertainty for 13 TeV (2015 and 2016).". Accessed on 15/10/2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution#JER_Scaling_factors_and_Uncertai.

[126] P. Bites, "Diagram showing the common principle of identification of jets initiated by b-hadron decays.", 2017. https://i1.wp.com/www.particlebites.com/wp-content/uploads/2016/08/B-tagging_diagram.png.

[127] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", arXiv:1712.07158.

[128] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[129] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys.* **G34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.

[130] F. Pedregosa et al., "Scikit-learn: Machine learning in python", *CoRR* **abs/1201.0490** (2012) arXiv:1201.0490.

[131] CMS Collaboration, "Identification of c-quark jets at the CMS experiment",.

[132] B. Group, "Usage of b/c tag objects for 13 tev data in 2016 and 80x mc", 2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco.

[133] CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087, 10.1007/JHEP10(2014)106, arXiv:1408.1682.

[134] C. collaboration et al., "Missing transverse energy performance of the cms detector", *Journal of Instrumentation* **6** (2011) P09001.

[135] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period",.

[136] CMS-MET-POG, "Missing Transverse Energy (MET)". Accessed on 15-01-2018. https://twiki.cern.ch/twiki/bin/view/CMS/MissingET.

[137] CMS-MET-POG, "TOP Trigger (Run2)". Accessed on15-11-2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopTrigger.

[138] R. R., "Beam halo identification for met commissioning", 2010. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideEJTermBeamHaloId.

[139] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector",.

[140] CMS Collaboration, "Search for physics beyond the standard model in events with two leptons of same sign, missing transverse momentum, and jets in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J.* **C77** (2017) 578, doi:10.1140/epjc/s10052-017-5079-z, arXiv:1704.07323.

[141] C. collaboration et al., "Performance of electron reconstruction and selection with the cms detector in proton-proton collisions at $\sqrt{S = 8\,\text{TeV}}$", *arXiv preprint arXiv:1502.02701* (2015).

[142] Physics-Validation-Group, "Estimating Systematic Errors Due to Pileup Modeling". Accessed on 15/12/2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors.

[143] Physics-Validation-Group, "Utilities for Accessing Pileup Information for Data". Accessed on 15/12/2017. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData.

[144] Egamma-Physics-Object-Group, "Electron efficiencies and scale factors". Accessed on 21/01/2018. https://twiki.cern.ch/twiki/bin/viewauth/CMS/EgammaIDRecipesRun2#Electron_efficiencies_and_scale.

[145] Muon-Physics-Object-Group, "Reference muon id, isolation and trigger efficiencies for Run-II". Accessed on 21/01/2018. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2#Introduction.

[146] A. Hocker et al., "TMVA - Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[147] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl. Instrum. Meth.* **A389** (1997) 81–86, doi:10.1016/S0168-9002(97)00048-X.

[148] A. Hoecker et al., "TMVA - Toolkit for Multivariate Data Analysis", *ArXiv Physics e-prints* (2007) arXiv:physics/0703039.

[149] B. P. Roe et al., "Boosted decision trees, an alternative to artificial neural networks", *Nucl. Instrum. Meth.* **A543** (2005) 577–584, doi:10.1016/j.nima.2004.12.018, arXiv:physics/0408124.

[150] A. Mayr, H. Binder, O. Gefeller, and M. Schmid, "The evolution of boosting algorithms", *Methods of information in medicine* **53** (2014) 419–427.

[151] P. C. Bhat, "Multivariate Analysis Methods in Particle Physics", *Ann. Rev. Nucl. Part. Sci.* **61** (2011) 281–309, doi:10.1146/annurev.nucl.012809.104427.

[152] H. working group, "Documentation of the RooStats -based statistics tools for Higgs PAG". Accessed on 15/03/2018. https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit.

[153] CMS Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B710** (2012) 26–48, doi:10.1016/j.physletb.2012.02.064, arXiv:1202.1488.

[154] ATLAS, CMS, LHC Higgs Combination Group Collaboration, "Procedure for the LHC Higgs boson search combination in summer 2011",.

[155] L. Moneta et al., "The RooStats Project", *PoS* **ACAT2010** (2010) 057, arXiv:1009.1003.

[156] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys.* **G28** (2002) 2693–2704, doi:10.1088/0954-3899/28/10/313. [,11(2002)].

[157] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth.* **A434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[158] 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). doi:10.5170/CERN-2011-006.

[159] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].

[160] J.-E.-R. and Corrections-Subgroup, "Jet energy scale uncertainty sources". Accessed on 15/09/2017. https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources.

[161] B.-T. and Vertising-Group (BTV), "Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC". Accessed on 15/09/2017. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80X/.

[162] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV",.

[163] K. Deroover, "A search for flavour changing neutral currents involving a top quark and a Higgs boson with the CMS experiment". PhD thesis, VUB, 2017.

[164] CMS Collaboration, "Search for the flavor-changing neutral current interactions of the top quark and the Higgs boson which decays into a pair of b quarks at $\sqrt{s} = 13$ TeV", `arXiv:1712.02399`. "

[165] CMS-BTV-Group, "CMS Heavy flavour tagging workshop 2018". `https://indico.cern.ch/event/695320`.

[166] CMS Collaboration, "Identification of c-quark jets at the CMS experiment",.

List of Abbreviations

2HDM	Two-Higgs-Doublet Model
ALICE	A Large Ion Collider Experiment at the LHC
ATLAS	A Toroidal LHC ApparatuS
BDT	Boosted Decision Trees
BR	Branching Ratio
BSM	Beyond the Standard Model
CDF	Collider Detector Facility at the FNAL Tevatron
CERN	Conseil Européen pour la Recherche Nucléaire, European laboratory for particle physics
CKM	Cabibbo, Kobayashi and Maskawa, used to refer to the quark-mixing matrix
CMS	Compact Muon Solenoid
CP	Charge and Parity transformations
CSC	Cathode Strip Chamber
CSV(AVR)	Combined Secondary Vertex (Adaptive Vertex Reconstruction) algorithm
CSV(IVF)	Combined Secondary Vertex (Inclusive Vertex Finder) algorithm
CSVv2	Combined Secondary Vertex (v2) algorithm, it is another name of CSV(IVF)
CL	Confidence Level
cMVAv2	combined MVA Algorithm
CTEQ	Coordinated Theoretical-Experimental Project on QCD
DØ	Experiment at the FNAL Tevatron
DAQ	Data AcQuisition
DeepCSV	Deep Combined Secondary Vertex algorithm
DESY	Deutsches Elektronen SYnchrotron laboratory, Hamburg
DGLAP	Dokshitzer, Gribov, Lipatov, Altarelli and Parisi, used to refer to the parton-evolution equations
DT	Drift Tube
EB	ECAL Barrel
ECAL	Electromagnetic CALorimeter
EE	ECAL Endcap
EFT	Effective Field Theory

EM	ElectroMagnetic
ES	ECAL preShower
EWSB	ElectroWeak Symmetry Breaking
FCCC	Flavour-Changing Charged Current
FCNC	Flavour-Changing Neutral Current
FNAL	Fermi National Accelerator Laboratory, USA
FSR	Final-State Radiation
GEANT	Detector simulation framework and toolkit
GRID	Infrastructure for distributed computing
GSF	Gaussian Sum Filter
GUT	Grand Unified Theory
HB	HCAL Barrel
HCAL	Hadron CALorimeter
HE	HCAL Endcap
HF	HCAL Forward
HEP	High-Energy Physics
HO	HCAL Outer
HLT	High-Level Trigger
IIHE	Inter-university Institute for High Energies, Brussels
IP	Impact Parameter
ISR	Initial-State Radiation
JEC	Jet Energy Corrections
JER	Jet Energy Resolution
JES	Jet Energy Scale
JP	Jet probability Algorithm
L1	Level-1 Trigger
LEP	Large Electron-Positron collider at CERN
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty experiment
LHCf	Large Hadron Collider forward experiment
LO	Leading Order
LSP	Lightest Supersymmetric Particle
MadGraph	Multi-particle matrix-element generator
MadEvent	Event generator powered by MadGraph
MC	Monte-Carlo simulation program or technique
ME	Matrix Element
MLE	Maximum Likelihood Estimator
MPI	Multi-Parton Interactions
MVA	Multivariat analysis
MSSM	Minimal SuperSymmetric Model
NNLO	Next-to-Next-to-Leading Order
NLO	Next-to-Leading Order

OS	Opposite Sign
PDF	Parton Density Function
pdf	probability density function
PF	Particle-Flow reconstruction algorithm
PS	Parton Shower or Proton Synchrotron
PU	Pileup
PV	Primary Vertex
PYTHIA	Monte-Carlo event generator for high-energy physics collisions
QCD	Quantum ChromoDynamics
QED	Quantum ElectroDynamics
QFT	Quantum Field Theory
QS	Quark Singlet model
RF	Radio Frequency
ROOT	An object-oriented data-analysis framework
RPC	Resistive Plate Chamber
SM	Standard Model
SPS	Super Proton Synchrotron collider at CERN
SS	Same Sign
SU(n)	Special Unitary group contains all matrices of dimension n
SV	Secondary Vertex
TEC	Tracker Endcap disc
TEVATRON	TeV-range proton–anti-proton accelerator at FNAL
TIB	Tracker Inner Barrel
TID	Tracker Inner Disc
TOB	Tracker Outer Barrel
TOTEM	LHC experiment for the measurement of the total cross section, elastic scattering and diffraction dissociation
TPC	Time Projection Chamber
UA1/2/5	Experiments at the CERN SPS collider
UE	Underlying Event
U(n)	Unitary group contains all matrices of dimension n
UED	Universal Extra Dimensions
VBF	Vector-Boson Fusion
VEV	Vacuum Expectation Value
WLHCG	Worldwide Large Hadron Collider Grid
WP	Working Point

Summary

The top quark plays a key role to search for new physics phenomena beyond the standard Model of particle physics . Being the heaviest particle in the Standard Model (SM), physicists believe that it has an enhanced sensitivity to various new particles and interactions suggested by theories Beyond the Standard Model (BSM). One of interesting possibilities is the presence of flavour-changing neutral current (FCNC) interactions between the top quark and the Higgs boson. The observation of a SM-like Higgs boson by the ATLAS and CMS experiments in 2012 initiated the study of these FCNC interactions.

In the SM, FCNC interactions of the top quark are suppressed. However, many BSM theories predict the existence of these processes with higher branching ratios, where some are within the reach of the current experiments. This thesis concerns an experimental research for FCNC interactions involving a top quark and a Higgs boson with a signature of two leptons in the final state with equal electric charge. The analysis is done using data of pp-collisions collected in 2016 by the CMS experiment at a center-of-mass energy of 13 TeV and an integrated luminosity of $36 fb^{-1}$. We probe the FCNC interactions in top-quark pair decays as well as in the single top quark production through Hqt-couplings, where the quark q is either an up-quark or charm-quark. No significant deviation with respect to the predicted SM background is observed. The expected (observed) upper limits at 95% CL on the branching ratio are $\mathcal{B}(t \to cH) \leq 7.0 \times 10^{-3}$ (6.8×10^{-3}) and $\mathcal{B}(t \to uH) \leq 6.7 \times 10^{-3}$ (6.1×10^{-3}).

The thesis is written in seven chapters that are summarized as:

- **Chapter 1**
  In this chapter the theoretical basis and basic concepts are presented, including an introduction about FCNC interactions in the SM and in BSM theories. Also an overview of experimental searches on top-FCNC interactions is provided.

- **Chapter 2**
  The Large Hadron Collider (LHC) is described as well as the CMS experiment, where the pp collision events take place and the collected data is recorded. The different subsystems of the CMS detector are also explained in more details.

- **Chapter 3**
  In this chapter the author explains how a pp-collision event looks like at the LHC and how it proceeds until final-state particles interact with the materials of the different types of detectors in the CMS experiment. Furthermore an overview of the simulation programs used for the production of event samples is added.

- **Chapter 4**
  The reconstruction of the main physics objects considered in the analysis is described. Also the Particle-Flow (PF) algorithm which is used for the identification of leptons and hadrons and the corrections related to them are explained. A brief overview of different b-tagging algorithms used in CMS experiment is given.

- **Chapter 5**
  In this chapter the analysis strategy is illustrated. The baseline selection cuts to disentangle the top-FCNC signal from the backgrounds are explained in details. The event yield is obtained for both signal and background from simulated samples and compared to the observed yield in data. The MVA technique which is used for further discrimination between FCNC-signal and SM backgrounds is explained. Also the statistical methodology used for limit setting is introduced.

- **Chapter 6**
  The results of the MVA training are discussed. The different systematic uncertainties and their effect on the final results are discussed. The expected (observed) upper limits on the branching ratio of top-FCNC are obtained.

- **Chapter 7**
  The conclusion of this experimental search for FCNC interactions is presented. The comparison between the sensitivity of the search presented in this thesis and those obtained from previous searches at the ATLAS and CMS experiments is illustrated. Finally the authors prospects of the top-FCNC research are presented.

Samenvatting

De top quark heeft een cruciale rol in de zoektocht naar nieuwe fysica fenomenen die niet beschreven worden in het standaardmodel van de deeltjesfysica. Omdat de top quark het zwaarste deeltje is in het standaardmodel, verwachten fysici dat het een belangrijke gevoeligheid heeft voor verschillende nieuwe deeltjes en interacties zoals gesuggereerd in modellen die het standaardmodel uitbreiden. Een van de interessante mogelijkheden is het voorkomen van interacties met smaakwisselende neutrale stromen tussen de top quark en het Higgs boson. De ontdekking van het standaardmodel Higgs boson door de ATLAS en CMS experimenten in 2012 opent het venster voor de studie van deze interacties.

In het standaardmodel van de deeltjesfysica zijn deze smaakwisselende neutrale stromen sterk onderdrukt. In diverse modellen die het standaardmodel uitbreiden komen deze interacties voor met een grotere vertakkingsverhouding, en voor sommige is dit binnen het bereik van huidige experimenten. Deze thesis omvat experimenteel onderzoek naar de smaakwisselende neutrale stromen tussen een top quark en het Higgs boson met een signatuur in de geobserveerde eindtoestand van protonbotsingen van twee leptonen met gelijke elektrische lading. De analyse werd ontworpen en uitgevoerd op proton botsingen geobserveerd door het CMS experiment in 2016 en dit bij een botsingsenergie van 13 TeV en met een geïntegreerde luminositeit van $36 fb^{-1}$. We bekijken het voorkomen van deze tHq smaakwisselende neutrale stromen zowel processen met twee top quarks en processen met slechts een top quark, en waar de quark q ofwel een up-quark is of een charm-quark. Geen significante afwijkingen werden waargenomen ten opzichte van de voorspellingen van het standaardmodel. De verwachte (bekomen) bovenlimiet met 95% betrouwbaarheidsniveau op de vertakkingsverhoudingen zijn $\mathcal{B}(t \to cH) \leq 7.0 \times 10^{-3}$ (6.8×10^{-3}) en $\mathcal{B}(t \to uH) \leq 6.7 \times 10^{-3}$ (6.1×10^{-3}).

Arabic Summary

الفصل الثالث:

هذا الفصل يقدم شرحا فيزيائيا للمراحل المختلفة التي يمر بها تفاعل البرتونات بداية من لحظة حدوث التفاعل وانتشار الجسيمات الناتجة من التفاعل حتى يتم قياسها من خلال تفاعلها مع الاجزاء المختلفة لكاشف CMS. كذلك يعرض برامج المحاكاة (المونت كارلو MC) المستخدمة لانتاج العينات اللازمة للدراسة.

الفصل الرابع:

يشرح هذا الفصل كيفية بناء وتشكيل الشارات الفيزياء الاساسية والبرنامج (الخوارزم) المستخدم لتصنيف هذه الشارات وتحديد هويتها كونها ليبتونات او هادرونات. كذلك يشرح التصحيحات الخاصة بها.

الفصل الخامس:

تناول هذا الفصل شرح مفصل لخطة الدراسة وتطبيقها على البيانات الخاصة بكل من العينات المستخدمة المنتجة ببرامج المحاكاة MC والبيانات الحقيقية الناتجة من تفاعل البرتونات Data لفصل الاشارات الناتجة من FCNC عن الناتجة من الخلفية المتوقعة من SM وبالنهاية يتم عرض نتائج تحليل البيانات. كذلك يقدم شرحا تفصيلاً لتقنية تحليل متعدد المتغيرات MVA المستخدمة لتحسين الحسابات بالاضافة للطريقة الاحصائية المستخدمة لحسابات الحدود العليا لنسب التفرع.

الفصل السادس:

عرض هذا الفصل النتائج كما تم شرح وتطبيق الاحصائيات الخاصة بالدراسة. كذلك تم حساب تأثير الاحصائيات ونسب عدم التأكد على نتائج الدراسة.

الفصل السابع :

يحتوي هذا الفصل على النتائج النهائية لهذه الدراسة ومقارنتها مع النتائج المقاسة بواسطة تجربتي ATLAS و CMS عند طاقة 8 ت . أ . ف ثم يتم عرض التوقعات المبنية على هذه الدراسة.

ملخص الرسالة

يلعب كوارك القمة (t) دورهام و اساسي في البحث عن الظواهر الفيزيائية الجديدة. نظرا لكونه اثقل الجسيمات في النموذج المعياري (SM) فإن الفيزيائيين يعتقدوا أن لهذا الكوارك حساسية معززة تجاه الجسيمات والتفاعلات الجسيمية الجديدة التي تقترحها نظريات ما بعد النموذج المعياري (BSM). أحد الاحتمالات المثيرة للانتباه والتي يمكن أن تؤكد افتراض وجود نظريات ما بعد النموذج المعياري هي التفاعلات التي تتضمن خاصية "تغيير النكهة بواسطة التيار متعادل الشحنة الكهربية (FCNC)" بين كوارك القمة وبوزون هيجز (H) والمتضمنة كوارك الاعلى (u) وكوارك الساحر (c).
تبعا للنموذج المعياري فإن هذه الخاصية (FCNC) تكاد تكون معدومة الظهور ولكن العديد من نظريات ما بعد النموذج المعياري تتنبأ بوجود هذه الظاهرة باحتمالات أكبر مما يسمح بدراستها وقياسها بواسطة التجارب العلمية الحالية. كما أن اكتشاف بوزون الهجيز بواسطة تجربتي الملف اللولبي الميوني المدمج (CMS) و أطلس ATLAS فتح مجال اوسع لدراسة هذه التفاعلات بصورة دقيقة.
تختص الدراسة المقدمة بهذه الرسالة بالدراسة العملية لظاهرة FCNC بين كوارك القمة وبوزون الهيجز والتي تحتوي على زوج من الليبتونات متساويي الشحنة الكهربية. وأجريت الدراسة على البيانات المجمعة بواسطة تجربة CMS خلال عام ٢٠١٦ والناتجة عن تفاعل البرتونات عند طاقة ١٣ ت. أ. ف.
علما بأن الدراسة المقدمة خلال هذه الرسالة قد اشتملت على دراسة خاصية FCNC في كلا من انتاج كوارك القمة الاحادي واضمحلال زوج كوارك القمة.
الدراسة لم تلاحظ اي انحراف عن الخلفية المتوقعة بواسطة النموذج المعياري. كما وضعت الحدود العليا المتوقعة والمقاسة لنسبة التفرع عند نسبة مصداقية ٩٥٪.

تشتمل الرسالة على سبعة فصول والتي يمكن تلخيصها كالأتي:

الفصل الاول:

يتضمن هذا الفصل مقدمة عامة عن المفاهيم والنظريات الأساسية الخاصة بالدراسة، وكيف تناولت نظريات SM وBSM ظاهرة FCNC . كذلك استعرض موجز للدراسات العملية السابقة لهذه الظاهرة.

الفصل الثاني:

يتناول هذا الفصل شرح توضيحي للمُصادم الهدروني الكبير LHC وتجربة CMS شاملا شرح الكواشف المختلفة التي تضمها تجربة CMS والتي يتم بداخلها تفاعل البرتونات. كذلك يوضح كيفية تسجيل البيانات الخاصة بناتج التفاعل.

"Gratitude is not only the greatest of virtues,
but the parent of all the others."

Melody Beattie

Acknowledgements

Finally finished.. I did it !!

It was a big challenge to leave my family, friends and my home country but I am lucky to have survived and the proof of this success is between your hands. Living for more than four years alone in a different society with a different language and culture is like taking part in an extensive workshop where I learned many lessons, not only in science but also on life.

First, I would like to thank my Belgian promotor Prof. Jorgen D'Hondt. Words are not enough to express my gratitude to you. I have learned from you not only science but also ethics, leadership and management. Thanks for everything you did to help me to achieve every step in my PhD, starting from accepting me as student in your research team until the day of my PhD defense. You wrote many supporting letters to help me secure funds to complete my studies and you were always there to advise me when problems arose, not only in my research but also with administrative issues. Scientific discussions with you always inspired and motivated me to explore the various aspects of my research in particle physics. Talking with you about anything always gives me positive energy. I am and will always be proud to be one of your students. I would also like to express my thanks and gratitude to my Egyptian promotor Prof. Magda Abdelwahab. Thanks for your kindness and for encouraging me ever since I was an undergraduate student. Thanks for accepting and respecting my choices regarding my research and academic career. Additionally I would like to thank my co-promotor in Egypt, Dr. Ahmed Ali, for his encouragement. I will not forget that you were the only one support me in the decision to change my research to the field of particle physics.

Besides my promotors, I would like to thank the jury members for reading my thesis. Their discussion and insightful comments during the private defense helped me to improve the quality of my work. I would also like to thank TOP PAG conveners for our discussions on physics and technical questions, as well as for their feedback, which incented me to widen my research from various perspectives.

IIHE (Inter-university Institute for High Energies) is the best environment that I have worked in so far, surrounded by gracious people. I have considered you (IIHE members) as my family here and each of you has touched my heart.

To my friends and labmates, I have spent most of my daily life in Belgium with you. During four years and eight months, I worked with many physicists who have all made an impression in my mind and in my heart. Annik, Gerrit, Giannis, Gregory, James, Nadir, Natalie, Taejong: you are all wonderful people. I wish you a happy life and the best of luck in the future. Denys, Dom, Doug, Emil, Isabelle, Isis, Jarne, Kevin, Kirill, Petra, Quentin, Seth and Simon: I have shared with you all kinds of events, both

happy and stressful times. Thanks for being available when I needed your support. I will always remember our discussions and crazy ideas as well as eating delicious home-made cakes and cookies and all the fun we had over the past four years. Special thanks to "Petra": working with you has allowed me to master the basic elements in data analysis. Your friendly hugs in hard times were like a painkiller for me. I believe that your role in my life has been remarkable. Also special thanks to Denys, even though you are the newest post-doc in our office, you are one of best people I have met. Your help, kindness and encouragement, as well as hugs from your lovely son, will not be forgotten. I would like to thank Nadjieh for her support since the first day we met at the Top2014 workshop. Staying with her when I had shifts at the "CMS control room" was amazing and memorable. Also thanks to Reza for helping me with the different techniques I used in my analysis.

I am also grateful to the IT group at IIHE (AbdelHaq, AbdelHakim, Olivier, Romain, Samir, Stéphane, Shkelzen, and Adriano) for always helping out even during weekends and holidays. You are a rescue team for any technical issue related to computing. Marleen, I could not imagine how my stay at IIHE would have been without you. You are the best secretary I have ever seen. Thanks for everything you did for me.

My thanks and regards to the members of the Physics department at Faculty of Women, Ain Shams University. Thanks for supporting and encouraging me. Without your precious support it would not be possible to conduct this work in Brussels.I would like to thank Prof. Amr El-Shalakany, from the Faculty of Medicine at Ain shams University, for all his help and support.

On the Egyptian side, I would like to express my thanks and gratitude to my faithful friends Aya, Eman, Haiam, Hayam, Nada, Noha, Radwa and Reham. Your existence in my life cannot be described in a few words. Thanks for being in my life and asking Allah to preserve our friendship forever.

Thanks to my friend Wahid for calming me down when I needed it and for encourage me when I was depressed, as well as for helping me to choose quotes for the chapters of the thesis. Thanks to Ahmed Hammad, PhD student at Bern University, for explaining some theoretical basis discussions.

Outside work, I would like to thank my friend Marie for her kindness, encouragement and for always praying for me. Talking with you helped me to get rid of any negative energy I might have felt.

During my PhD study I received two different scholarships: one from Erasmus Mundus for the first two years of the PhD, and the other from the Egyptian Mission Department for the last two years. Without this precious support it would not have been possible to conduct my research. I would therefore like to express my gratitude and regards towards both institutions for providing the funding for this work.

Last but not least, unlimited thanks to my family: my parents, my brothers, sister, nephews and nieces for supporting me spiritually throughout my stay abroad and my life in general. Special thanks to my hero - my father - who is the backbone in my life.

And before anything I bow to Allah for giving me all of these gifts in my life and more. Without His mercy I could not achieve anything in my life... So All thanks and praises are to Allah always and forever...

VRIJE UNIVERSITEIT BRUSSEL
كلية العلوم وعلوم الهندسة الحيوية
قسم الفيزياء

AIN SHAMS UNIVERSITY
كلية البنات للآداب والعلوم والتربية
قسم الفيزياء

البحث عن الاقترانات متعادلة الشحنة الكهربية في تغيير نكهة كوارك القمة بواسطة تجربة "الملف اللولبي الميوني المُدمج" المُقام على "المُصادم الهدروني الكبير"

رسالة مقدمة للحصول على درجة الدكتوراة المزدوجة
دكتوراة الفلسفة في العلوم من الجامعة الحرة ببروكسل ـ بلجيكا
دكتوراة الفلسفة في العلوم تخصص فيزياء نووية من جامعة عين شمس ـ مصر

إعداد

شيماء أبوزيد

إشراف

أ.د. ماجدة عبدالوهاب
استاذ الفيزياء النووية ـ
جامعة عين شمس

أ.د. يورهن دونت
استاذ الفيزياء ـ
الجامعة الحرة ببروكسل

يونيو ٢٠١٨