Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Master's Degree in Advanced Physics: Particles, Astrophysics, Nanophysics and Quantum Materials

Master's Thesis

Measurement of the production cross section of top quarks produced with bosons using events with two leptons in the final state

Alejandro Soto Rodríguez

Supervisors:

Jose Enrique Palencia Cortezón

Santiago Folgueras Gómez

Academic Course 2020-2021

ABSTRACT

Measurement of the production cross section of top quarks produced with bosons using events with two leptons in the final state

A measurement of the inclusive and differential cross sections for the production of single top quarks in association with a W boson (tW process) in proton-proton collisions at a centre-of-mass energy of 13 TeV is presented in this Master's Thesis. For this measurement, data recorded with the CMS detector in 2016, 2017 and 2018 corresponding to an integrated luminosity of 137.2 fb^{-1} is analysed. Events containing one electron and one muon with opposite charge are selected. With this requirement, different regions are considered based on the number of jets and b tagged jets. For the inclusive measurement, two Boosted Decision Trees are trained to separate the signal from the dominant $t\bar{t}$ background. For the differential measurement, a fiducial region is defined by the requirement of exactly one b tagged jet with a veto to less energetic ones. The resulting distributions are unfolded to particle-level and compared with predictions at next-to-leading order in perturbative quantum chromodynamics.

Contents

List of Figures

List of Tables

Introduction

This Master's Thesis presents the study of the production of single top quarks in association with a W boson (tW) in proton-proton collisions at 13 TeV of energy in the centre-of-mass. The main objective is to obtain a measurement of the inclusive and differential cross sections of this process using data recorded with the Compact Muon Solenoid (CMS) detector in 2016, 2017 and 2018.

The measurement of the tW process is complex due to the overwhelming top-antitop quark pair production ($t\bar{t}$) background. This process produces very similar final states to those of tW, making it difficult to distinguish between them. For that, two Boosted Decision Trees (BDT) are trained using the kinematic properties of the events. Using the output distributions of the BDTs a likelihood fit is performed to extract the signal contribution and measure the inclusive cross section. For the differential measurement signal events are obtained by substracting the background from data. The measured distributions at particle-level are obtained performing the unfolding in three different variables. These results are compared with SM predictions at next-to-leading order in perturbative quantum chromodynamics (QCD).

This work is a continuation of previous measurements done by the CMS Collaboration ([1, 2]) of the tW process which only used 2016 data. Here, the full Run 2 is analysed which corresponds to about 4 times more data. At the moment, this analysis is in the last stages for its publication and it is being documented in an internal analysis note for the CMS Collaboration. These results are not public yet so the results that are presented in this Master's Thesis for the full Run 2 do not include data. However, it is possible to present the results but only with the 2016 data since it has already been published.

This document is structured as follows. In Chapter 1, the Standard Model (SM), the theoretical framework in which this study is based, is presented. An specific section is dedicated to top quark physics to describe the most important processes of this analysis.

Chapter 2 describes the experimental setup for this measurement. The first section is dedicated to the Large Hadron Collider (LHC). Where the main components and parameters of this accelerator are presented. In the second section, the CMS detector and its trigger system are described.

Chapter 3 is dedicated to describe the event reconstruction and simulation. Event reconstruction allows to extract the physical information from the raw data collected by the detectors. On the other hand, event simulation provides the expected results from the SM to compare with experimental data.

Chapters 4 and 5 are the central part of this work. In Chapter 4 the selection criteria and the strategies for both measurements, the inclusive and differential, are presented. And, in Chapter 5, the experimental results for both measurements are presented.

Lastly, the conclusions of the analysis are presented in Chapter 6.

Chapter 1

The Standard Model of particle physics and beyond

1.1 The Standard Model of particle physics

The SM of particle physics is a theory that describes the elementary particles that compose matter and their interactions. It combines in a single framework the Electroweak Theory (EW) and QCD, and describes the electromagnetic, weak and strong interactions. Formally, the SM is a $U(1) \times SU(2) \times SU(3)$ gauge invariant Quantum Field Theory (QFT). Figure 1.1 shows the particles that compose the SM. There are two main categories of particles in terms of their spin: **fermions** and **bosons**. Fermions have half-integer spin and bosons integer spin. At the same time, fermions are subdivided into **leptons** and **quarks**. There are three generations of leptons and quarks (columns in Fig. 1.1 ordered from less to more massive particles). From these generations, only the first are stable particles and, thus, they are the only ones that conform matter. In the SM, particles are excitations of quantum fields and they interact by the exchange of bosons, which are the mediators of the interactions. The electromagnetic interaction has associated only one charge, the electrical charge which can be positive or negative. On the other hand, there are three charges (red, blue and green) for the strong interaction. These charges are called colour charges. There are also anti-colour charges: anti-red, anti-blue and anti-green.

The quark family consists of the up (**u**) and down (**d**) quarks in the first generation, the charm (**c**) and strange (**s**) in the second generation and the top (**t**) and bottom (**b**) in the third generation. They can interact electromagnetically, weakly and strongly.

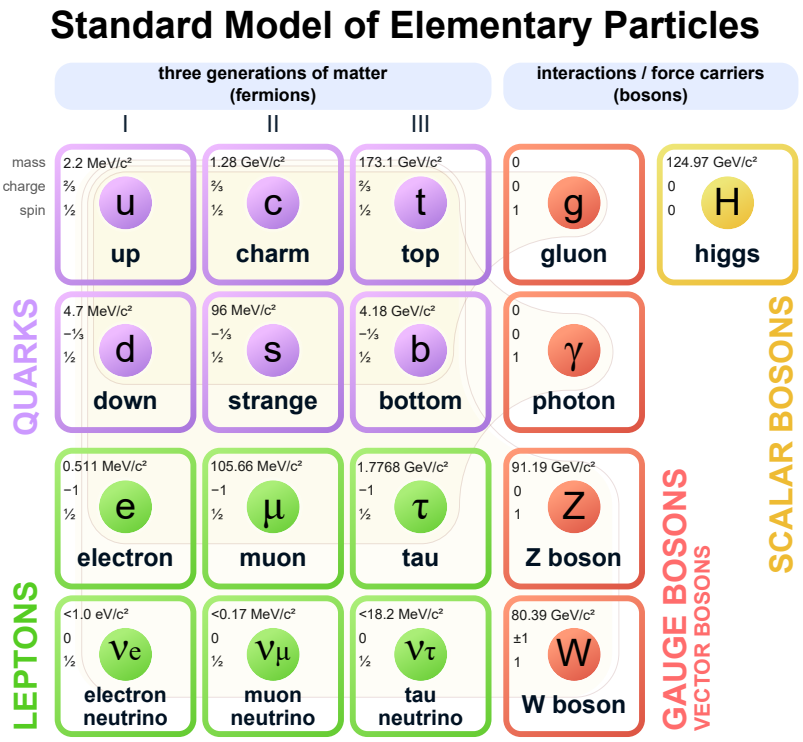

Figure 1.1: Fundamental particles of the Standard Model of Particle Physics and its properties [3].

Regarding the electromagnetic charge, the quarks u, c and t have charge $\frac{2}{3}$ and the quarks d, s and b $\frac{-1}{3}$ in units of the electron charge. From the quark family, only the up and down quarks are stable.

The other group of fermions are the leptons. They are the electron (**e**) and electron neutrino ($\boldsymbol{\nu_e}$) for the first generation, the muon ($\boldsymbol{\mu}$) and muon neutrino ($\boldsymbol{\nu_\mu}$) for the second generation and the tau ($\boldsymbol{\tau}$) and tau neutrino ($\boldsymbol{\nu_\tau}$) for the third generation. All leptons interact via weak interaction and, only the electron, muon and tau via electromagnetic interaction.

Every fermion has an almost identical partner with the same mass and opposite electric charge. They are known as **antiparticles**. Their existence was proposed by Paul Dirac in 1928 [4] when he interpreted the negative energy solutions of its equation (Eq. 1.1) as a new type of particle. The Dirac equation 1.1 was the result of unifying the quantum theory with special relativity.

$$(i\not{\partial} - m)\psi = 0 \qquad \text{in natural units}^1\ (\hbar = c = 1), \tag{1.1}$$

where $\not{\partial} = \gamma^\mu \partial_\mu$, γ^μ the Dirac matrices and ψ a Dirac spinor.

On the other hand, there are five different bosons in the SM: the photon ($\boldsymbol{\gamma}$), mediator of the electromagnetic interaction, the $\boldsymbol{W^\pm}$ and $\boldsymbol{Z^0}$ bosons of the weak interaction, gluons[2] ($\boldsymbol{g}$) of the strong interaction and the Higgs boson ($\boldsymbol{H^0}$) which gives mass to particles coupled to it via the Higgs Mechanism.

The next two subsections are used to explain with more detail the electroweak and strong interactions.

1.1.1 The electroweak interaction and the Higgs mechanism

The electroweak theory is the unified description of electromagnetic and weak interactions. It was proposed by Sheldon Glashow, Abdus Salam and Steven Weinberg in 1979 and it is known as the GWS theory [5, 6]. Mathematically, electromagnetism and weak interactions are unified under the gauge group $SU(2)_T \times U(1)_Y$, where T is the weak isospin and Y the weak hypercharge. This theory contains four massless gauge bosons W_1, W_2, W_3 and B. These bosons are not physical fields until electroweak symmetry breaking effected by the Brout-Englert-Higgs mechanism [7, 8] occurs. This produces the $W^\pm$ and Z^0 bosons and the photon. After electroweak symmetry breaking only the photon remains massless as the theory lost the $SU(2)_T \times U(1)_Y$ symmetry by $U(1)_{\text{EM}}$.

For a fermion field belonging to a general $SU(2)$ representation, with $U(1)$ charge Y, the covariant derivative takes the form:

$$D_\mu = \partial_\mu - igW_\mu^a T^a - ig'YB_\mu. \tag{1.2}$$

T^a are the generators of $SU(2)$ in the representation considered.

To introduce the Higgs mechanism, consider a scalar field ϕ in the spinor representation of $SU(2)$ and with $Y = \frac{1}{2}$. Then, the covariant derivative acting on ϕ:

$$D_\mu \phi = (\partial_\mu - igW_\mu^a \tau^a - i\frac{1}{2}g'B_\mu)\phi. \tag{1.3}$$

Where g and g' are the coupling constants of $SU(2)$ and $U(1)$, $a = (1, 2, 3)$ and $\tau^a =$

[1] Natural units will be used in the rest of the Master's Thesis.

[2] There are eight types of gluons inside the SM. As much as independent combinations of colour-anticolour pairs.

$\sigma^a/2$, being σ^a the Pauli matrices defined as

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.4}$$

Using the definition of the covariant derivative 1.3, the Lagrangian for the scalar field is:

$$\mathcal{L} = |D^\mu \phi|^2 - V(\phi); \quad V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{1.5}$$

Then if the field ϕ acquires a vacuum expectation value (VEV) of the form

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{1.6}$$

the gauge bosons acquire mass. To see this, we introduce the expansion of Eq. 1.6 into the Lagrangian 1.5. Looking only at the quadratic terms of the gauge field (mass terms):

$$\Delta\mathcal{L}_{mass} = \frac{1}{2}\frac{v^2}{4}\left[g^2(W_\mu^1)^2 + g^2(W_\mu^2)^2 + (-gW_\mu^3 + g'B_\mu)^2\right]. \tag{1.7}$$

There are three massive vector bosons:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \quad \text{with mass} \quad m_W = g\frac{v}{2} \tag{1.8}$$

$$Z_\mu^0 = \frac{1}{\sqrt{g^2 + g'^2}}\left(gW_\mu^3 - g'B_\mu\right) \quad \text{with mass} \quad m_Z = \sqrt{g^2 + g'^2}\frac{v}{2}. \tag{1.9}$$

And one massless vector boson orthogonal to Z_μ^0:

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}\left(g'W_\mu^3 + gB_\mu\right) \quad \text{with mass} \quad m_A = 0, \tag{1.10}$$

which is the photon of quantum electrodynamics.

The covariant derivative of Eq. 1.2 in terms of the mass eigenstates:

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}(W_\mu^+ T^+ + W_\mu^- T^-) - i\frac{1}{\sqrt{g^2 + g'^2}}Z_\mu(g^2T^3 - g'^2Y) - i\frac{gg'}{\sqrt{g^2 + g'^2}}A_\mu(T^3 + Y), \tag{1.11}$$

where $T^{\pm} = (T^1 \pm iT^2)$.

We identify the coefficient of the electromagnetic interaction as the electron charge

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}}, \tag{1.12}$$

and the electric charge quantum number

$$Q = T^3 + Y. \tag{1.13}$$

To simplify the expression 1.11 further, we define the weak mixing angle, θ_ω:

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_\omega & -\sin\theta_\omega \\ \sin\theta_\omega & \cos\theta_\omega \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix}, \tag{1.14}$$

that is,

$$\cos\theta_\omega = \frac{g}{\sqrt{g^2 + g'^2}}, \sin\theta_\omega = \frac{g'}{\sqrt{g^2 + g'^2}}. \tag{1.15}$$

Then, we can rewrite the covariant derivative 1.11 in the form

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}(W_\mu^+ T^+ + W_\mu^- T^-) - i\frac{g}{\cos\theta_\omega} Z_\mu (T^3 - \sin^2\theta_\omega Q) - ieA_\mu Q, \tag{1.16}$$

where

$$g = \frac{e}{\sin\theta_\omega}. \tag{1.17}$$

We see that the couplings of all of the weak bosons are described by two parameters: the electron charge and θ_ω. The exchange of W and Z bosons will involve the masses of these particles. However, these masses are not independent since:

$$m_W = m_Z \cos\theta_\omega. \tag{1.18}$$

At the end, all effects of W and Z exchange processes can be written in terms of: e, θ_ω and m_W.

Up to now we have only worked with the kinetic part of 1.5. If we expand the potential

in terms of 1.6, we get:

$$\mathcal{L}_V = -\frac{1}{2}m_h^2 h^2 - \sqrt{\frac{\lambda}{2}}m_h h^3 - \frac{1}{4}\lambda h^4, \tag{1.19}$$

where

$$m_h = \sqrt{2}\mu = \sqrt{2\lambda}v. \tag{1.20}$$

Thus $h(x)$ is a scalar field with mass 1.20. This particle is known as the Higgs boson. It was proposed in 1964 and discovered in 2012 by the CMS [9] and ATLAS [10] collaborations at CERN. The actual measured mass of the Higgs boson is [11]:

$$m_h = 125.10 \pm 0.14 \text{ GeV}. \tag{1.21}$$

Coupling to Fermions

The covariant derivative 1.16 uniquely determines the coupling of the W and Z^0 fields to fermions, once the quantum numbers of the fermions fields are specified. To determine this quantum numbers we need to use the fact that the W boson only couples to left-handed helicity states of quarks and leptons.

When we couple a fermion field ψ to a gauge field, we can assign ψ_L and ψ_R to different representations of the gauge group. In the GWS model we can use this technique to ensure that only the left-handed components of the quark and lepton fields couple to the W bosons. We assign the left-handed fermion fields to doublets of $SU(2)$, while making the right-handed fermion fields singlets under this group. For the right-handed fields, $T^3 = 0$, and so we reproduce the standard electric charges by assigning Y to equal the electric charge. For example, for the right-handed u quark field, $Y = +2/3$; for e_R^-, $Y = -1$. For the left-handed fields,

$$E_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \quad Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \tag{1.22}$$

the assignments $Y = -1/2$ and $Y = +1/6$, respectively, combine with $T^3 = \pm 1/2$ to give the correct charge assignments.

We are going to ignore the fermion masses now to write the Lagrangian for the weak

interaction between leptons and quarks:

$$\mathcal{L} = \bar{E}_L(i\not{D})E_L + \bar{e}_R(i\not{D})e_R + \bar{Q}_L(i\not{D})Q_L + \bar{u}_R(i\not{D})u_R + \bar{d}_R(i\not{D})d_R. \tag{1.23}$$

A right-handed neutrino would have zero coupling both to $SU(2)$ and to $U(1)$, so we have simply omitted it from Eq. 1.23.

Fermion mass terms

The construction of fermion mass terms is not trivial because they combine the left and right-handed components of the fermion. This is forbidden by global gauge invariance. If we want to include their masses, we need to invoke again the mechanism of spontaneous symmetry breaking. This allows us to write the gauge-invariant term for the u and d quarks

$$\Delta\mathcal{L} = -y_d\bar{Q}_L\phi d_R - y_u\epsilon^{ab}\bar{Q}_{La}\phi_b^\dagger u_R + \text{h.c.} \tag{1.24}$$

Where ϵ^{ab} is the Levi-Civita symbol.

Since there are not right-handed neutrinos in the SM, the mass term for the electron would be:

$$\Delta\mathcal{L} = -y_e\bar{E}_L\phi e_R + \text{h.c.} \tag{1.25}$$

In general, the parameter y_f, where f refers to a fermion, is called the Yukawa coupling of the Higgs boson to the fermions. If we replace ϕ by its VEV, we obtain for every fermion (except neutrinos):

$$\Delta\mathcal{L} = -\frac{1}{\sqrt{2}}y_f v\bar{f}_L f_R + \text{h.c.} \tag{1.26}$$

This is a mass term for the fermion. The size of the mass is set by the VEV of ϕ and y_f:

$$m_f = \frac{1}{\sqrt{2}}y_f v. \tag{1.27}$$

The Cabibbo-Kobayashi-Maskawa (CKM) matrix

In the last discussion was only considered the first generation of fermions. As it was mentioned before, the SM contains three generations of fermions. In principle, it is possible to consider a mixing term between generations. This is true for the quarks, in

this case, the Yukawa couplings must be a 3×3 matrix that contain off diagonal terms. We can diagonalise the Higgs couplings by choosing a new basis for the quark fields. This will cause a complication in the gauge couplings since now the mass eigenstates are different from the flavour eigenstates. Denoting with

$$u_L^i = (u_L, c_L, t_L), \quad d_L^i = (d_L, s_L, b_L), \tag{1.28}$$

the flavour eigenstates and, with $u_L^{'i}$ and $d_L^{'i}$ the mass eigenstates. This last basis is the physical one as it diagonalise the mass matrix. The two bases are related by the unitary transformations:

$$u_L^i = U_u^{ij} u_L^{'j}, \quad d_L^i = U_d^{ij} d_L^{'j}. \tag{1.29}$$

This change of basis only affects the interactions of the W boson:

$$J_W^{\mu+} = \frac{1}{\sqrt{2}} \bar{u}_L^i \gamma^\mu d_L^i = \frac{1}{\sqrt{2}} \bar{u}_L^{'i} \gamma^\mu (U_u^\dagger U_d)_{ij} d_L^{'i} = \frac{1}{\sqrt{2}} \bar{u}_L^{'i} \gamma^\mu V_{ij} d_L^{'i}. \tag{1.30}$$

Where V_{ij} is the CKM matrix

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{1.31}$$

and $J_W^{\mu+}$ the W^+ boson current (a similar one exists for $J_W^{\mu-}$).

The off-diagonal terms allow for weak transitions between quarks of different generations, however, the diagonal elements of the CKM matrix are bigger than the off-diagonal. This has the effect that transitions between the same generation are more likely to occur. An especial case, that makes the top quark decay almost always to the bottom quark, is the value of V_{tb}, which is $0.999172 \pm^{0.000024}_{0.000035}$ [11]. The elements of the CKM matrix, although related, they have to be determined experimentally.

1.1.2 The strong interaction

The quantum field theory that describes the strong interaction is QCD. It is a $SU(3)_C$ non-abelian gauge theory with C the conserved colour charge. In the SM quarks are the only fermions that have colour charge, therefore, they have a non-trivial representation

under $SU(3)$. Each quark appears as a triplet of 3 colours, whereas leptons are singlets. This theory has eight gauge bosons, corresponding to gluons named G^a_μ, where $a = 1, \cdots, 8$. In this theory, the covariant derivative takes the form:

$$D_\mu = \partial_\mu - ig_s G^a_\mu t^a. \tag{1.32}$$

Where g_s is the coupling constant and t_a the generators of the group, in this case, the Gell-Mann matrices. In this Master's Thesis, I will refer to the strong coupling constant as $\alpha_s = \frac{g_s^2}{4\pi}$. The generators of $SU(3)$ satisfy the following Lie algebra:

$$[t^a, t^b] = if^{abc}t^c, \tag{1.33}$$

where f^{abc} are a set of numbers called structure constants. The field strength tensor is defined by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu. \tag{1.34}$$

Finally, the QCD Lagrangian is

$$\mathcal{L}_{QCD} = \sum_k^{\text{all quarks}} \bar{\psi}_k \left(i\not{D} - m_k\right) \psi_k - \frac{1}{4}(G^a_{\mu\nu})^2. \tag{1.35}$$

The renormalization group equations predict an scaling of α_s (1-loop beta function) as

$$\frac{d\alpha_s}{dQ^2} = -\frac{\alpha_s^2}{12\pi Q^2}(33 - 2n_f) \tag{1.36}$$

Where n_f is the number of flavours and Q the scale of the interaction. For $n_f = 3$, $\alpha_s \to 0$ at high energies ($Q^2 \to \infty$). The vanishing of the QCD coupling for large values of Q^2 is referred to as asymptotic freedom. One important consequence of asymptotic freedom is that QCD processes at high energies can be reliably computed in a pertubation expansion in α_s. At low energies $\alpha_s \sim 1$ and pertubation expansions will not converge. In the low energy regime quarks will form bounded colourless states named hadrons. This phenomenon is called colour confinement, and implies that quarks and gluons will not be observed as free particles. These states are mainly observed in groups of three quarks (baryons) or as quark-antiquark pairs (mesons). The process by which quarks and gluons form hadrons is called hadronisation.

In high energy collisions the hadrons that are produced are in general unstable. To achieve stability, they decay into more particles producing a shower called jet that is experimentally observable. The exception to this is the top quark, which due to its

high mass, decays before hadronising.

1.1.3 Cross section and Feynman diagrams

In particle physics one of the most important quantities is the cross section, usually represented with the letter σ. It expresses the probability of a certain process to occur. For example, the probability of producing a Higgs boson in a collision of two protons. It has units of area and can be derived from the theory, in our case, the SM. This involves the calculation of the so-called matrix element $\mathcal{M}_{fi}$, which relates to the cross section as $\sigma \propto |\mathcal{M}_{fi}|^2$.

Computing exactly the matrix element is in practice impossible. Instead, what is done is a perturbative expansion in terms of Feynman diagrams. Feynman diagrams are not only a very useful representation of the processes, but also a very powerful tool to make calculations. Regarding their topological structure, they are made of external/internal legs and vertices like the ones in Fig. 1.2. In Fig. 1.3, a Feynman diagram for the Drell-Yan process is represented. For the interactions in the SM, the expansions are made in powers of the electromagnetic, weak and strong couplings: α_{EM}, α_W and α_S respectively. Each vertex of the diagram introduces a $\sqrt{\alpha}$ factor. The coupling constants quantify the strength of the interactions. These expansions are always possible if $\alpha \ll 1$. This is the perturbative regime of the theory. For the strong interaction the low-energy (long distance) regime is non-perturbative and $\alpha_S \sim 1$.

In a collision experiment between fundamental particles, generally, the cross sections decrease with the centre of mass energy ($\sqrt{s}$) as $\sigma \propto 1/s$. This implies that the probability of a collision to occur decreases. In the case of the LHC, collisions are between protons which are non-fundamental particles. For this type of collisions, the

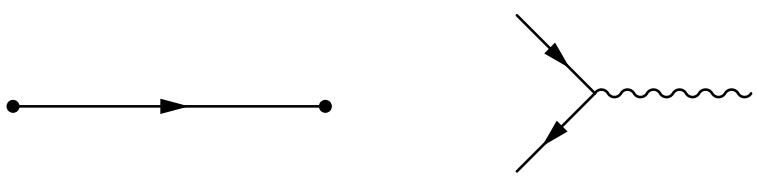

Figure 1.2: External/internal fermion leg (left) and fermion-photon vertex (right) of a Feynman diagram.

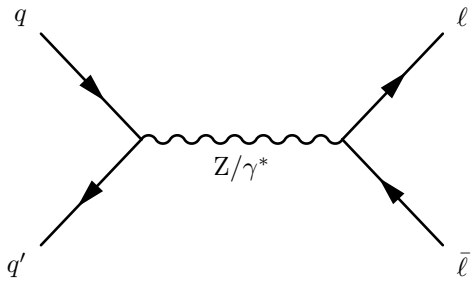

Figure 1.3: Leading order Feynman diagram for Drell-Yan process.

cross section at fixed order takes the form:

$$\sigma(pp \to X) = \sum_{i,j}^{\text{partons}} \int dx_i dx_j f_i(x_i, \mu_F) f_j(x_j, \mu_F) \sigma_{ij}(\hat{s}, \mu_R, \mu_F). \tag{1.37}$$

Where f_i and f_j are the Parton Distribution Functions (PDFs) that express the probability of finding a parton inside the proton with a fraction of momentum $x_i = \frac{|\vec{p}_i|}{|\vec{p}_p|}$, where $|\vec{p}_i|$ and $|\vec{p}_p|$ are the momentum of the parton i and the proton, respectively. $\hat{s} = x_i x_j s$ is the squared partonic centre of mass energy. An important aspect in the calculation of hard scattering processes in hadron colliders is the possibility to separate long distance physics (PDFs) and short distance physics (hard scatter of quarks and gluons). This is parametrised by the factorisation scale μ_F, which can be interpreted as the energy scale at which the soft and hard effects are separated. Finally, μ_R is the renormalization scale that enters in the partonic cross section σ_{ij} to cure divergences of some Feynman diagrams.

Although the cross section $\sigma(\hat{s}, \mu_R, \mu_F)$ decreases with the energy, the probability of occurring the process increase due to the PDFs. This behaviour can be explained with Fig. 1.4. The explanation is that if the centre of mass (CM) energy is increased, less fraction of momentum x is needed to produce the particle of interest. And, if x is smaller, the probability of finding a parton with that energy increases, especially for gluons. Some examples of processes, and the dependence of its cross section with the CM energy, are shown in Fig. 1.5.

The perturbative expansion of the cross section is an infinite series and it has to be truncated somewhere. According to the degree of approximation, we have: Leading Order (LO), Next-to-Leading Order (NLO) and Next-to-Next-to-Leading Order (NNLO) as the first three degrees of approximation. The number of Feynman diagrams increase dramatically with higher order approximations. In particle physics the simulation of the collisions is done using Monte Carlo generators. Besides hard

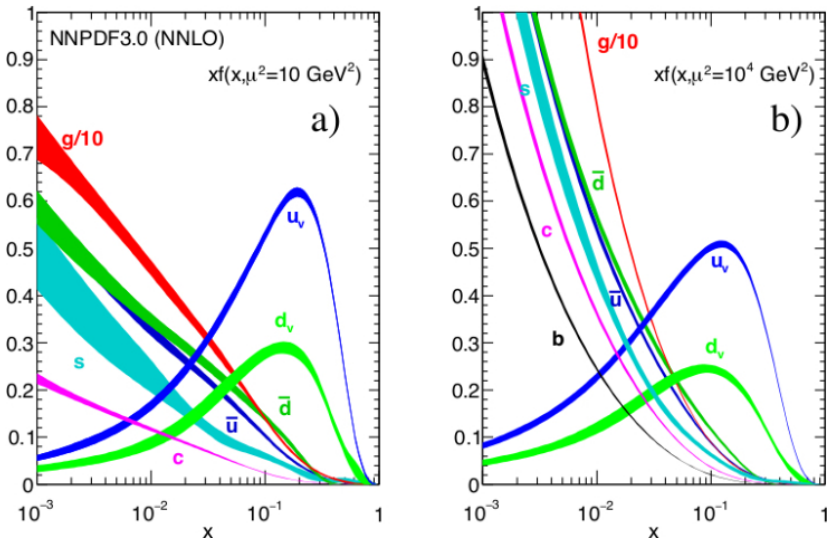

Figure 1.4: Parton Distribution Functions for a proton in terms of $x = \frac{|\vec{p}_i|}{|\vec{p}_p|}$. On the left, with 10 GeV^2 of CM energy squared and, on the right, with 10^4 GeV^2 [11].

scattering, that provides the partonic final state, radiation and decay processes have to be simulated. It is also necessary to simulate the interaction of the final state particles with the detector to compare simulation with real observations. This is a very important aspect in most of the high energy physics experiments and it will be covered later in Chapter 3.

1.2 Top quark physics

The top quark is the most massive elementary particle in the SM. Its phenomenology is driven by its large mass value. It is the only quark that decays before hadronisation, most of the times into a real W boson and a b quark. Its lifetime as predicted by the SM is around 10^{-25} s. Because of this, it is the only quark that can be studied in a free state, where perturbative QCD applies. The large mass of the top quark is due to its strong coupling to the Higgs boson, the biggest of all fermions. For these reasons, the top quark plays an special role in the SM and in many BSM extensions. An accurate knowledge of its properties is important to understand the fundamental interactions at the electroweak symmetry-breaking scale and beyond.

The discovery of the top quark was achieved by the DØ [13] and CDF [14] experiments in 1995. The most precise measurement of its mass up to now is $m_t = 172.8 \pm 0.3$ GeV which is an average of the LHC and Tevatron runs [11].

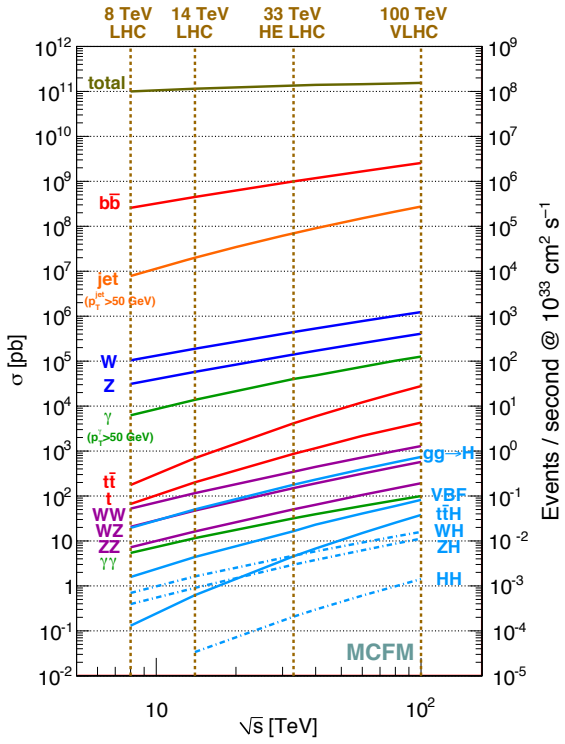

Figure 1.5: Cross section for some SM processes in terms of the CM energy in pp collisions [12].

1.2.1 Top quark production and decay

There are multiple processes that can lead to the production of top quarks in hadron collisions. They can be divided in two categories: top-pair production and single-top production. These are the dominant production modes of top quarks, but there are other less probable processes like the production of top quark pairs in association with a gauge boson ($t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}\gamma$), $t\bar{t}H$ or four top quark production ($t\bar{t}t\bar{t}$).

At the LHC, the single top production is dominated by the t-channel, s-channel and tW-channel. The Feynman diagrams for these processes are shown in Fig. 1.6. In this Master's Thesis I will study the tW process using data recorded with the CMS detector in proton-proton collisions in the LHC. Besides, the other important process due to its similar final states is the top anti-top pair production ($t\bar{t}$ process), which is the main background of the analysis.

Starting with the $t\bar{t}$ production, the leading order Feynman diagrams are shown in Fig. 1.7. From them, at the actual CM energy of 13 TeV in the LHC proton-proton collisions, the gluon fusion mode is the dominant. At this energy, the total inclusive cross section of $t\bar{t}$ at NNLO is $\sigma = 831.76\ ^{+19.77}_{-29.20}\ (\text{scale}) \pm 35.06(\text{PDF})\ ^{+23.18}_{-22.45}\ (m_t)$ pb

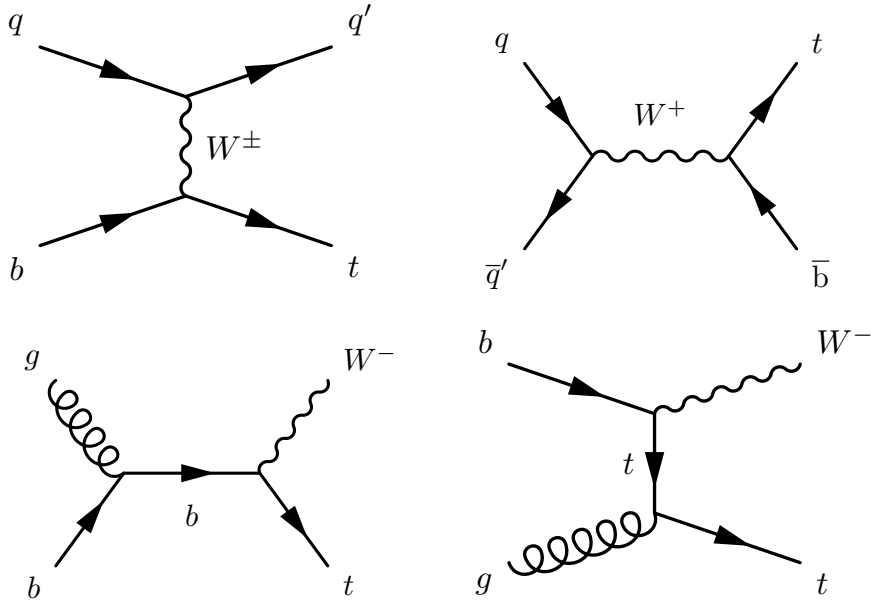

Figure 1.6: Leading order Feynman diagrams for the single top production. The top figures correspond to the t-channel (left) and s-channel (right). On the bottom, two Feynman diagrams for tW are presented.

[15] assuming a top quark mass of 172.5 GeV. The first uncertainty comes from the independent variation of the factorisation and renormalisation scales, the second one is associated to variations in the PDF and α_S and the third one to variations in the top quark mass.

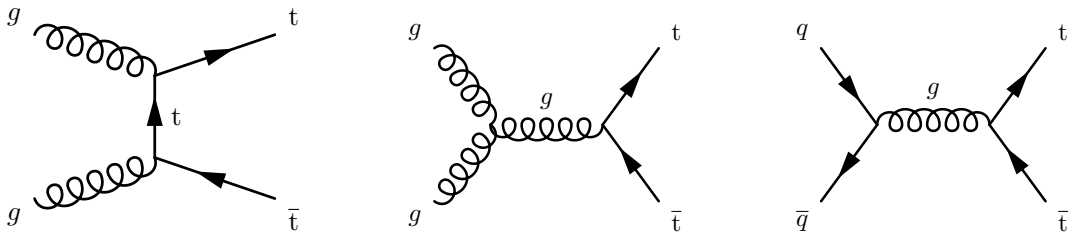

Figure 1.7: Leading order Feynman diagrams for the $t\bar{t}$ production. The two diagrams on the left correspond to the gluon fusion production mode while the one on the right the $q\bar{q}$ mode.

In almost all the events the top quark decays into a W boson and a b quark. Thus, for the $t\bar{t}$ process the first decay is into $bW^+ + \bar{b}W^-$. The possible final states are determined by the W bosons decays, which can decay to a lepton-neutrino pair or to a quark-antiquark pair. Figure 1.8 summarises the possible final states of the $t\bar{t}$ process.

Regarding the tW process, it refers to the production of a top quark together with a W boson. This process is mediated by the electroweak interaction and provides a probe for measuring the V_{tb} coupling. The study of tW production has also interest in

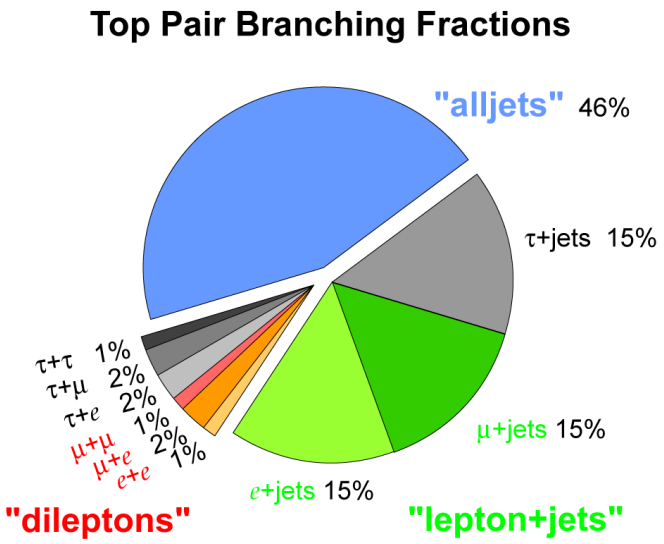

Figure 1.8: Pie chart of the possible final states of the $t\bar{t}$ process.

BSM physics [16]. One remarkable property of this process is the interference with the $t\bar{t}$ production. This interference occurs at NLO in perturbative QCD, where the final states of $t\bar{t}$ and tW are the same. For this reason, the tW process is only well defined at LO where the final states of $t\bar{t}$ and tW are different. To compute higher order corrections to the cross section of tW it is necessary to introduce removal methods that take into account this interference.

In the perturbative computation of tW production, one must consider all partonic processes with final states

$$t + W + \sum_i X_i. \tag{1.38}$$

Where X_i is a set of particles whose multiplicity increases with each order in perturbation theory. At LO this set is empty but at NLO contributions as

$$t + W + \bar{b} \tag{1.39}$$

will appear. With the decay of the $t \to Wb$, this final state is the same as $gg \to t\bar{t} \to WWb\bar{b}$ from the $t\bar{t}$ process. This is shown in Fig. 1.9 where Feynman diagrams for $t\bar{t}$ production and tW with an additional b quark are drawn. The $t\bar{t}$ diagram has two resonant top quarks and it is called a doubly resonant diagram in contrast with the tW diagram on the right which is singly resonant. The squared matrix element for the $WWb\bar{b}$ final state can be written as:

$$|\mathcal{M}_{WWb\bar{b}}|^2 = |\mathcal{M}_{\text{singly}}|^2 + |\mathcal{M}_{\text{doubly}}|^2 + 2Re\left(\mathcal{M}^*_{\text{singly}}\mathcal{M}_{\text{doubly}}\right) \tag{1.40}$$

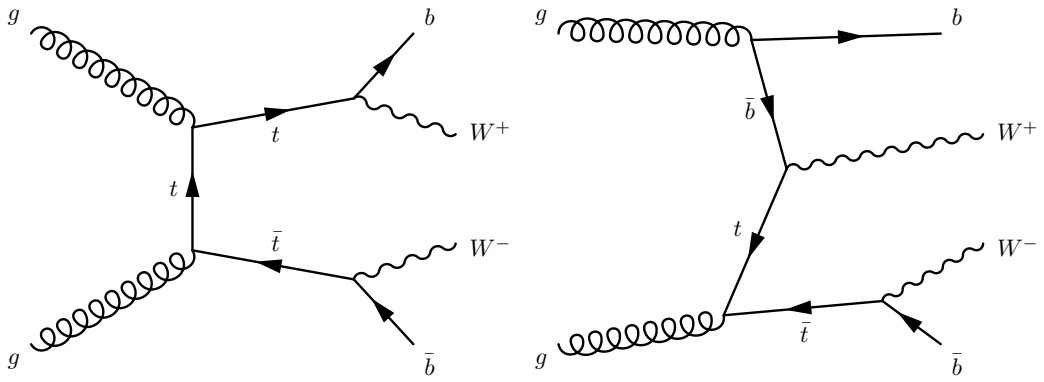

Figure 1.9: Two interfering Feynman diagrams between $t\bar{t}$ (left) and tW production (right). The left diagram shows two resonant top quarks (doubly resonant) and the right diagram only one (single resonant).

Two schemes are proposed to describe the tW signal: diagram removal (DR), where all NLO diagrams which are doubly resonant are excluded from the definition, and diagram subtraction (DS), in which the differential cross section is modified with a gauge-invariant subtraction term that locally cancels the contribution of the $t\bar{t}$ diagrams. The first approach violates gauge invariance, however, in practise, little dependence on the gauge choice is observed. For this analysis, the DR method is used, as it is the convention in the community. The difference between the two methods will be considered as an uncertainty.

The tW production cross section computed at an approximate NNLO is $\sigma = 71.7 \pm 1.8(\text{scale}) \pm 3.4(\text{PDF})$ pb [17] at 13 TeV in proton-proton collisions and assuming a top-quark mass of 172.5 GeV. The possible final states, as in the $t\bar{t}$ case, are determined by the W bosons decay tW $\rightarrow bW^+W^-$. This process was first observed by the CMS [18] and ATLAS [19] collaborations in the dilepton channel. The results showed agreement with the SM predictions.

1.3 Physics beyond the Standard Model

The SM is a renormalizable and mathematically self-consistent quantum field theory. Its precision has been tested in an enormous range of energies without observing any deviations from the predictions. However, it does leave some unexplained phenomena.

Gravity: The SM does not include gravity as QFT. It is expected that the SM would fail at energies or distances where the quantum gravity effects are expected to dominate, this is known as the Planck scale. For energies, it is at the order of 10^{19} GeV and, for distances, 10^{-35} m. These are unachievable energies with the current accelerators.

For example, the Large Hadron Collider operates at 13 TeV of energy in the centre of mass, which makes, at the present time (and near future), impossible to explore the Planck energies. In a modern field theory context, the SM is seen as an effective field theory approximation at low energies of a more general theory with gravity.

The hierarchy problem: In the SM the Higgs mass acquires very large quantum corrections due to the presence of virtual particles, mostly top quarks. Figure 1.10 represents the corrections to the propagator of the Higgs boson. These diagrams will contribute transforming $-\mu^2 \rightarrow -\mu^2 + \delta\mu^2$ where μ is the parameter of the Higgs potential of Eq. 1.5. The problem is that these Feynman diagrams are quadratically divergent with the cutoff scale, in other words, $\delta\mu^2 \sim \Lambda^2$ where Λ is the cutoff of the theory. If $\Lambda \gg$ TeV then $\delta\mu^2 \gg \mu^2_{\text{physical}}$ and the bare potential must be fine tuned to get the correct physical mass. This is the so-called hierarchy problem and it gets worse the higher the cutoff. Small cutoffs give rise to the little hierarchy problem and cutoffs near the Planck scale $\Lambda \sim M_{Pl} \sim 10^{19}$ GeV give rise to the big hierarchy problem. In summary, the Hierarchy problem is the issue that the Higgs mass m_h is quadratically sensitive to any high scale in the theory.

Physics beyond the SM could solve this problem introducing more contributions to the virtual loops that can cancel the top quark loops without the necessity of fine-tuning. Naively one would expect that m_h should be on the order of the scale of new physics. This is the main reason for which most particle physicist believe that new physics should be close to the TeV scale where the experiments are now looking.

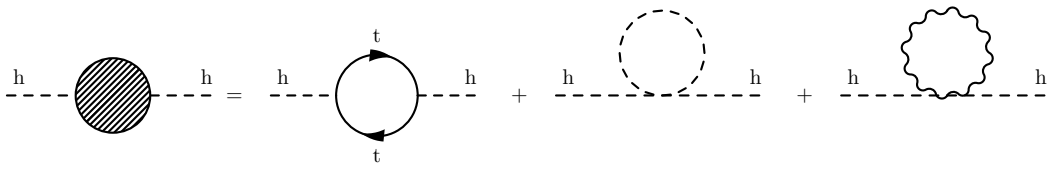

Figure 1.10: One loop corrections to the Higgs mass parameter. All three diagrams are quadratically divergent.

Neutrino masses: Neutrinos in the SM do not interact with the Higgs boson and, therefore, remain massless. However, neutrino oscillations, that were observed by the Sudbury Neutrino Observatory, imply that they have mass. This fact indicates that the SM has to be modified to include the neutrinos masses. Currently, there is no agreement in how to include these masses in the SM.

Dark matter: Astronomical observation have shown that another type of matter that only interacts gravitationally is present in the Universe. Its abundance relative to the

total matter and energy content of the Universe is 26%. In the SM there is no possible candidate to the dark matter particle.

Several theoretical models have been proposed to explain these phenomena. Some of them are: Supersymmetry (SUSY), String theory or extra dimensions. Although they will not be covered in this Master's thesis, lots of searches have been performed and, more are planned in the LHC physics programme.

Chapter 2

Experimental setup

2.1 Large Hadron Collider

The LHC is the world's largest and most powerful particle accelerator. It is located beneath the French-Swiss border near Genève in a 27 km tunnel at around 100 m of depth. It was built by the European Organisation for Nuclear Research (CERN) between 1998 and 2008. The LHC primarily collides proton beams, but it can also produce heavy-ion collisions like: lead-proton or lead-lead collisions. In Fig. 2.1 is represented the CERN accelerator complex where the LHC is the dark blue circle on the top of the figure. It is the last step of a larger accelerator chain that accelerates particles to the energy required for the collisions. In this thesis only proton-proton collisions will be studied, thus, we focus only on describing these ones. First, protons are obtained from a hydrogen bottle and injected into the Linac 2 accelerator, followed by the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) to finally end in the LHC. In this process the protons are divided in two beams that travel in opposite directions in the LHC ring. Each beam is divided in various proton bunches of 10^{11} protons separated 25 ns. There are four collision points in the LHC (marked by the yellow dots in Fig. 2.1) where experiments are located to detect the particles produced in the collisions. These are: CMS, ATLAS, ALICE and LHCb. Different collaborations operate each detector and have different physics programmes. For this thesis, data recorded with the CMS experiment is used.

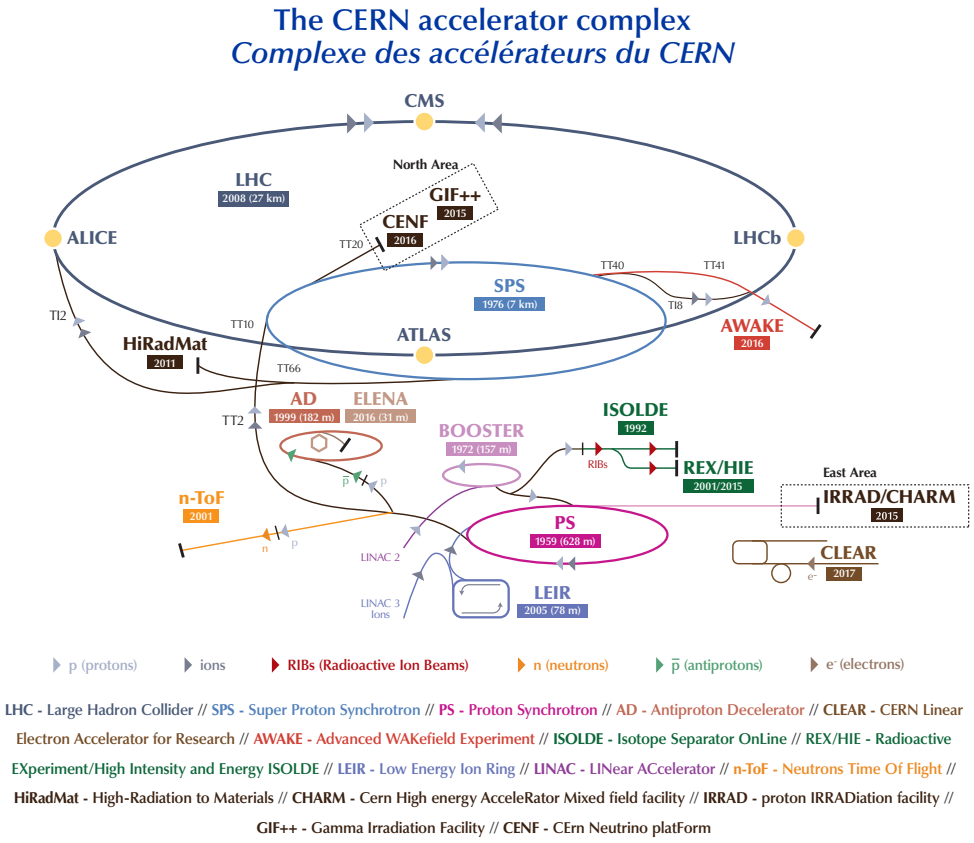

Figure 2.1: The CERN accelerator complex. The LHC is the last ring (dark blue line) in a complex chain of smaller accelerators [20].

2.1.1 Parameters of an accelerator

There are two basic parameters that characterise a particle collider: the centre of mass energy and the luminosity.

Centre of mass energy

The centre of mass energy in the LHC is equal to the sum of energies of each beam at the laboratory frame. The LHC reaches a centre of mass energy $\sqrt{s}$ of 13 TeV, which corresponds to 6.5 TeV per beam. This quantity measures the energy put into play in a collision to produce new particles. In the future, it is expected to reach 14 TeV, which is the design energy of the LHC.

Luminosity

The luminosity can be expressed in terms of the instantaneous luminosity or, integrating over time, in terms of the integrated luminosity. The instantaneous luminosity is a measure of the number of collisions produced per unit of area and

time, more luminosity implies more collisions. It is represented by the letter $\mathcal{L}$ and it is usually expressed in $\text{cm}^{-2}\text{s}^{-1}$. Currently, the LHC reaches values of $2 \cdot 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$. If we assume Gaussian proton bunches, with σ_x and σ_y widths in the x and y directions, colliding head-on at near the speed of light, we can obtain the analytic expression for the luminosity of Eq. 2.1.

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \tag{2.1}$$

Where N_1 and N_2 are the number of protons per bunch, f the revolution frequency of the protons and N_b the number of bunches in each beam.

Integrating over time the instantaneous luminosity, we obtain the integrated luminosity

$$L = \int_0^T \mathcal{L}(t')dt'. \tag{2.2}$$

It has units of inverse of area which are usually expressed in fb^{-1} where $1\text{b} = 10^{-28}$ m^2. In this thesis, data corresponding to an integrated luminosity of 137.2 fb^{-1} will be analysed.

The golden formula in particle physics relates the integrated luminosity L and the cross section σ, with the total number of expected events N as

$$N = L \cdot \sigma. \tag{2.3}$$

For comparison, the total inelastic cross section of a proton-proton collision is ~ 80 mb while for the tW process it is 72 pb (both at 13 TeV in the centre of mass). In other words, only one of 10^9 inelastic collisions is a tW process.

A direct consequence of the high instantaneous luminosity is the **pileup**. This term refers to the contamination of multiple low energy interactions between protons in the same bunch crossing. It is also possible to have contamination from previous o later bunch crossings and they are known as out-of-time pileup. This affects the measurements as it becomes more difficult to recognise which particles belong to the collision of interest. The average pileup for the data of the Run 2 is represented in Fig. 2.2.

2.1.2 Main components of the LHC

Simplifying the description of a particle collider, there are two main functionalities that it needs to control: the particles energy and its trajectory. In the LHC particles are accelerated using radiofrequence cavities. On the other hand, dipolar magnets are used

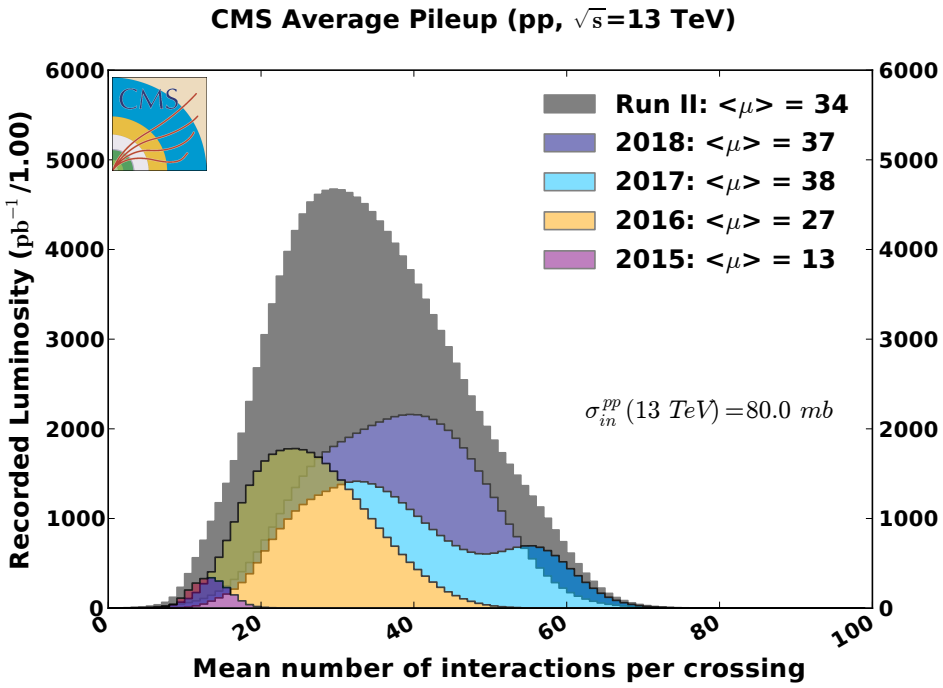

Figure 2.2: Average pileup interactions for each bunch crossing for the years 2015 to 2018 [21].

to keep the circular trajectory of the beam. In this section the components of the LHC are briefly described. However, a detailed description of the LHC and its components can be found in [22].

As mentioned before, radiofrequence cavities are used to accelerate particles in the LHC. There are 16 in the LHC and they are able to increase the energy from the 450 GeV, provided by the previous accelerators, up to 6.5 TeV. This cavities are grouped 4 by 4 and require very low temperatures to behave as superconductors. To achieve this, they are placed inside 4 cylindrical criomodules. They need 20 minutes to reach the maximum energy, time during which protons are able to do 10 millions of laps. Radiofrequence cavities are synchronised with the beam pass and oscillate at 400 MHz. This causes that particles with higher or lower energy than desired are decelerated or accelerated respectively, homogenising the beam energy. Figure 2.3 shows one of the radiofrequence cavities installed at the LHC.

Another important component of the LHC are superconducting magnets. There are several types of magnets, each with a specific function such as keeping the circular trajectory, correct the shape of the bunch which can be affected by the repulsive forces inside the bunch and magnetic field imperfections, or focus the beam.

The magnets with the function of bending the beam along the circular trajectory of the LHC are the dipole magnets. There are 1232 in total, their length is 15 m,

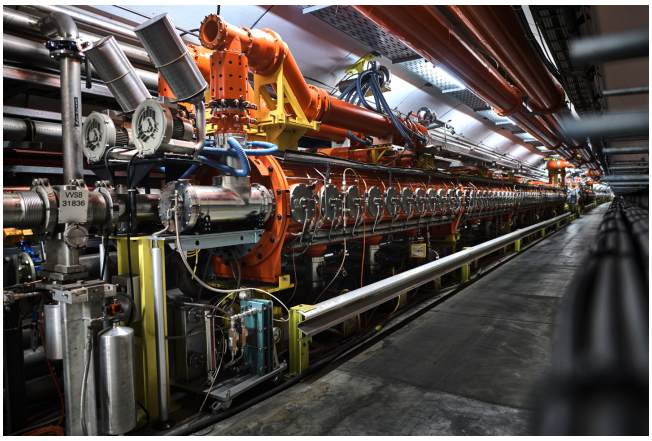

Figure 2.3: One of the radiofrequence cavities of the LHC [23].

weight 35 tons and reach a magnetic field of 8.3 T [24]. Figure 2.4 shows the internal structure of these magnets. The magnetic field generated is along the vertical axis, in the plane perpendicular to the LHC. However, this magnetic field is not completely homogeneous and it shows imperfections, specially in the magnets ends. To fix this, sextupole, octapole and decapole magnets are used.

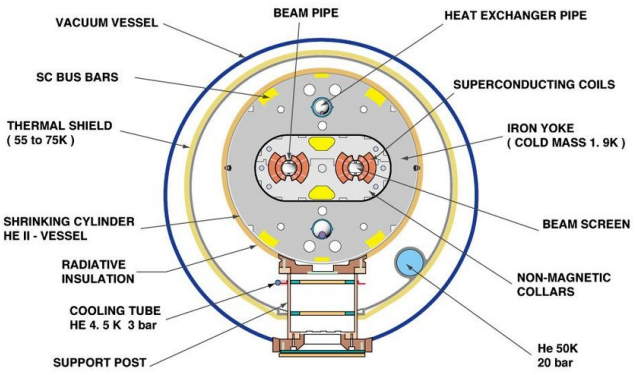

Figure 2.4: Cross section of a LHC dipole magnet [25].

The other mentioned functionality is focusing the beam. This is an important aspect in such a way that it substantially improves the number of collisions. If the beam dimensions σ_x and σ_y are decreased, luminosity increases as can be seen in Eq. 2.1. This task is carried out by the so called quadrupole magnets. They have four poles placed symmetrically around the beam pipe, and can exert compression forces over the bunches. They are especially used near the detectors to drive protons coming from opposite directions.

After each bunch crossing only few protons will have interacted. There is another set of magnets that allow to focus and reuse the beam. To get rid of the beam when it is no longer useful, or in case some problem is detected, it is directed to a graphite block

where it is absorbed.

2.2 The CMS detector

CMS is a general purpose detector designed to detect a great variety of processes produced at the energy scales of the LHC. Its physics programme goes from the study of the SM processes to BSM physics, like dark matter or extra dimensions searches.

CMS is a cylindrical detector placed around the beam pipe where previously accelerated hadrons in the LHC collide. It is composed of various subdetectors, each one specialised in the detection of one or various kinds of particles. A representation of the detector is showed in Fig. 2.5. Geometrically, the detector is divided in two regions: the barrel which is the lateral side of the cylinder and the endcaps. The whole detector weights 14000 tons, it is 15 m tall and 28.7 m long. It has the worlds' most powerful superconducting solenoid, that reaches a magnetic field of 3.8 T.

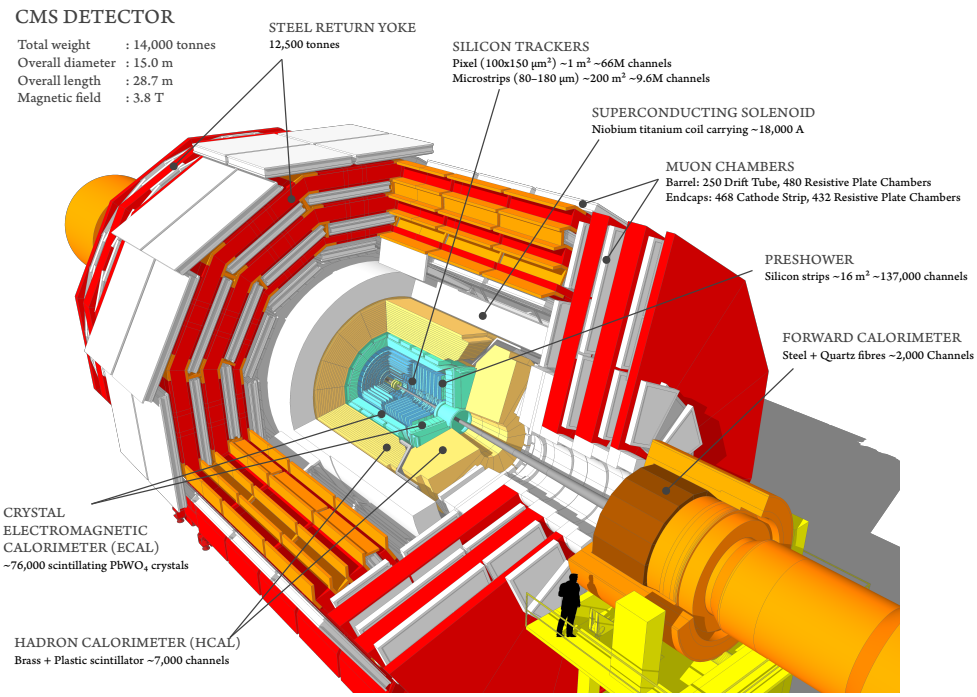

Figure 2.5: Structure of the CMS detector [26].

The subdetectors that make up CMS are: the tracking detector, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the muon chamber. Due to the importance of every subdetector to measure a collision a specific section will be dedicated to each one.

A complete description of CMS can be found in [27].

2.2.1 Coordinate system

The coordinate system used in CMS has the origin in the collision point. The z axis is oriented along the beam line and pointing to the west. The XY plane is commonly referred as the transverse plane and it is defined with the x axis pointing the centre of the LHC circumference.

In Fig. 2.6 is represented the CMS coordinate system. Due to the CMS symmetry, cartesian coordinates are less useful and instead the coordinates ϕ, η and r are used. The azimuthal angle $\phi \in [0, 2\pi)$, is measured with respect to the x axis in the XY plane. The radial coordinate r is also measured in this plane. The pseudorapidity $\eta \in (-\infty, \infty)$ is defined in terms of the angle $\theta \in [0, \pi]$ which is measured from the z axis in the YZ plane. A value of $\eta = 0$ indicates that the particle is moving in the transverse plane. The pseudorapidity is defined as

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

It is a Lorentz invariant quantity for massless particles and under approximations also for ultrarelativistic particles.

Using these coordinates the transverse momentum $\vec{p}_T$ is defined as the projection of the linear momentum in the XY plane. In this thesis I will denote by p_T the module of $\vec{p}_T$

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{2.5}$$

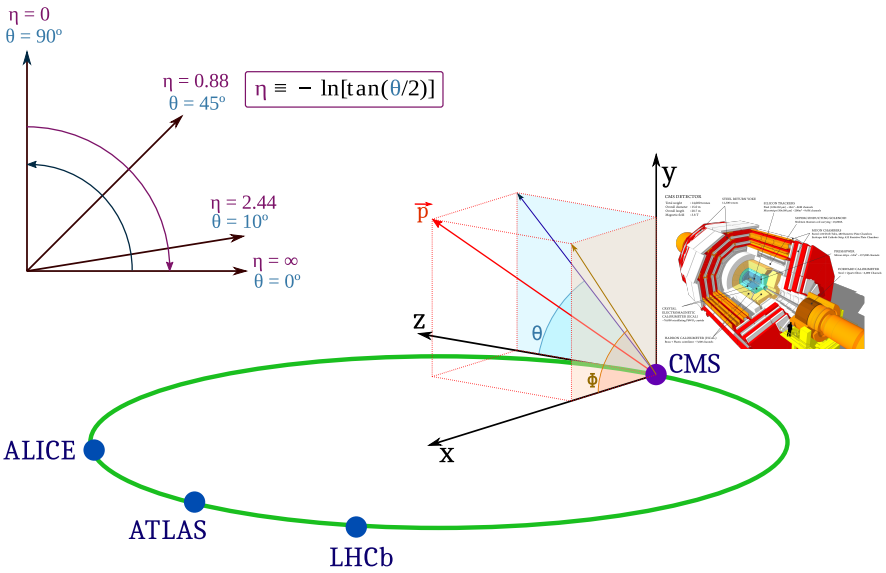

Figure 2.6: CMS coordinate system [28].

In a hadronic head-on collision, the total transverse momentum is conserved and its

0. Thus, if we add all the transverse momentum of all the particles produced in the collision the result should be

$$\sum_{i}^{\text{all part.}} \vec{p}_T^{\,i} = 0. \tag{2.6}$$

If some of the particles produced in the collision are not detected, it would be a momentum imbalance that could be associated to the non detected particles. This is commonly called missing transverse momentum $\vec{p}_T^{\,\text{miss}}$ (Eq. 2.7). Its modulus is represented by p_T^{miss}.

$$\vec{p}_T^{\,\text{miss}} = -\sum_{i}^{\text{detected part.}} \vec{p}_T^{\,i} \tag{2.7}$$

2.2.2 Tracking detector

The tracker is the innermost subdetector in CMS [29]. It allows to measure the trajectory of charged particles produced in the collisions. Due to the magnetic field created by the superconducting solenoid, charged particles describe helical trajectories. Measuring the curvature of the trajectory, the momentum p can be estimated using

$$p[\text{GeV}] = 0.3 \cdot B[\text{T}] \cdot R[\text{m}]. \tag{2.8}$$

Where B is the magnetic field and R the radius of the trajectory.

Extrapolating the trajectories to the beam line, the interaction vertex can be reconstructed. The high instantaneous luminosity of the LHC introduces a large number of pileup interactions. A graphic representation of a collision with a high number of pileup interactions is shown in Fig. 2.7, where there are 78 reconstructed interaction vertexes. The primary vertex is defined as the vertex with the higher total momentum of the tracks associated to it. It is the one with the higher momentum transfer and, hence, where the most interesting physic processes may appear. It is necessary to be able to discriminate the origin of each track in order to subtract it from the primary collision. The high granularity and efficiency of the tracker allows to distinguish them, reaching a spatial resolution of 10 μm [30]. This resolution also allows to measure secondary vertices originated from particles produced at the primary vertex which travel a distance before decaying. Besides, it interacts little with particles so subsequent measurements of energy in the calorimeters will not loose too much performance.

The tracking detector is formed by the silicon pixels which are located in the innermost part of the detector, and the silicon strips. The pixel detector has 124 millions of

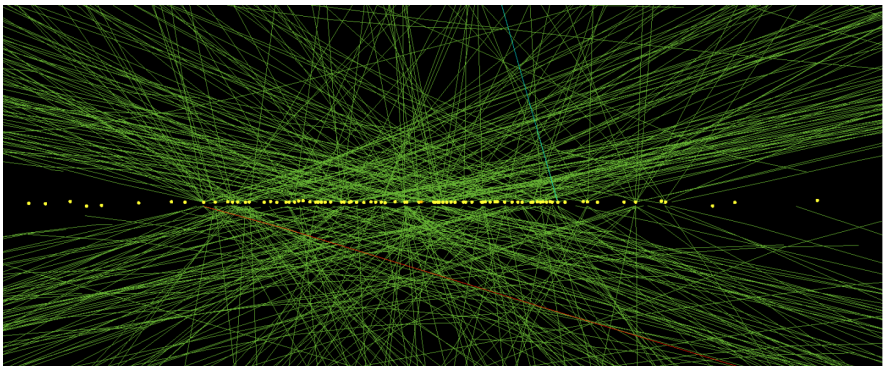

Figure 2.7: 78 reconstructed vertices in a high pileup event [31]. The green lines correspond to the charged tracks and the yellow dots with the interaction vertices.

channels in total and 6000 connections per cm^2. It is divided in 4 cylindrical layers at 2.9 cm, 6.8 cm, 10.9 cm and 16.0 cm of distance to the beam pipe and 3 outer and inner rings (before 2017 the pixel detector only had 3 layers and 2 rings) that extend the tracker coverage up to $|\eta| = 2.5$ [32]. Around 10 millions of particles traverse the tracker per second and square centimetre, so the tracker is built to be radiation resistant. When a charged particle travels through the tracker, it ionises the medium and generates free electrons. These free electrons are collected and amplified to obtain a more intense electric signal.

The silicon strips are located in the most external part of the tracker. It is divided in 10 concentric cylindrical layers up to a radius of 116 cm. Four of these layers form the inner barrel region and the other six the outer region. Both regions are closed with 12 disks in the endcaps. Figure 2.8 shows a sketch of the tracking system after the upgrade (Phase-1) of the detector in 2017.

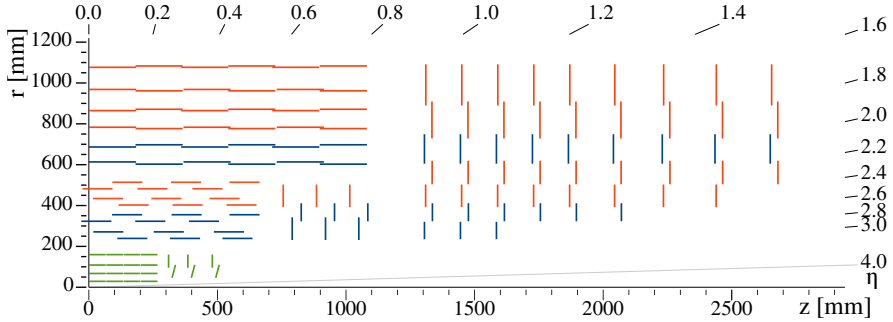

Figure 2.8: Sketch of one quarter of the Phase-1 CMS tracking system in $r - z$ view. The pixel detector is shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively [33].

The momentum resolution is better in the transverse region than in the z direction.

For the transverse region it is expressed as

$$\frac{\sigma(p_T)}{p_T} = \frac{0.015\% \cdot p_T}{\text{GeV}} \oplus 0.5\%. \tag{2.9}$$

Particles with more momentum are less bent by the magnetic field and it worsens the resolution, this is represented by the first term. The effects of multiple scattering are represented by the second term.

2.2.3 Electromagnetic calorimeter

The ECAL measures the energy and direction of photons and electrons by total absorption of the particle. It is an homogeneous scintillator calorimeter, where all the material contribute to the signal. The other calorimeter in CMS, the HCAL, is, on the other hand, a sampling calorimeter. The ECAL is formed by near 76000 lead tungstate crystals ($PbWO_4$) [34], which is a very dense material (8.28 g/cm^3) that scintillates when electrons and photons pass through it. Such a dense material with a Molière radius[1] of 2.2 cm and a radiation length[2] of 0.89 cm, allows for a very compact calorimeter. The lead tungstate crystals are $25.8X_0$ long in the barrel and $24.7X_0$ long in the endcaps, where X_0 is one radiation length. The amount of emitted light is proportional to the energy of the particles. Photodetectors are glued onto the back of the crystals to detect the scintillation light and convert it to an electrical signal that is amplified. The ECAL is designed to have a very fast response (25 ns) and to be radiation tolerant.

The ECAL is made up of a barrel and two endcaps. The crystals are distributed in groups called supermodules. There are 36 supermodules in the barrel with 1700 crystals each. Besides, the ECAL has installed preshower detectors that sits in front of the endcaps. These increase the spatial resolution allowing to distinguish between high energy photons and close pairs of low energy photons.

The ECAL barrel coverage is up to $|\eta| = 1.48$ and the endcaps extend the coverage to $|\eta| = 3.0$. The preshower detector fiducial area is approximately $1.65 < |\eta| < 2.6$. A representation of the CMS ECAL is showed in Fig. 2.9.

[1] A Molière radius is by definition the radius of a cylinder containing 90% of the energy of the electromagnetic shower.

[2] One radiation length is the amount of material needed in the longitudinal direction to absorb almost 2/3 of the energy of an electron or, 7/9 of the mean free path for pair production by a high-energy photon.

The energy resolution of the CMS ECAL [35] is given by

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E/\mathrm{GeV}}} \oplus \frac{12\%}{E/GeV} \oplus 0.3\%. \tag{2.10}$$

The first term is the stochastic term, the second one is due to electronic noise and the last one accounts for the non-uniformity of the detector and calibration uncertainties. Because of the very small stochastic term inherent to homogeneous calorimeters, the photon energy resolution is excellent in the 1–50 GeV range typical of photons in jets.

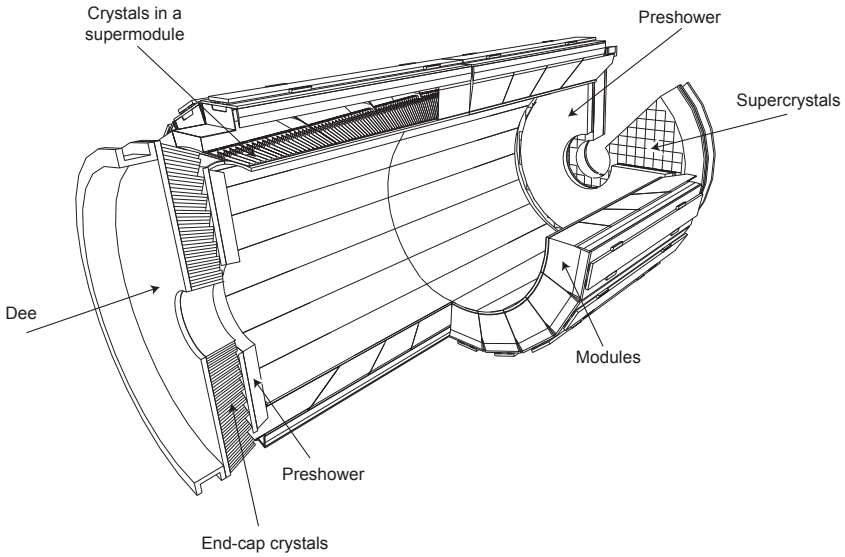

Figure 2.9: Layout of the CMS ECAL, showing the barrel supermodules, the two endcaps and the preshower detectors [36].

2.2.4 Hadronic calorimeter

The next subdetector is the hadronic calorimeter or HCAL. It measures the energy from charged and neutral hadrons. It is a sampling calorimeter which consist of a metallic absorber sandwiched or (threaded) with an active material which generates signal. The parameter that characterise the size of an hadronic calorimeter is the interaction length λ_I, which for most materials its greater than the radiation length. Thus, the size of the HCAL is greater than the ECAL. The HCAL is composed by alternating layers of plastic scintillators and non-magnetic brass used as absorbing material with $\lambda_I = 16.4$ cm.

The hadronic calorimeter was designed to be as hermetic as possible to provide a good estimation of $\vec{p}_T^{\,\mathrm{miss}}$. The HCAL is made up of a barrel region (HB) that covers up to $|\eta| < 1.2$, two endcaps (HE) that extend the calorimiter up to $|\eta| < 3$, two

forward calorimeters (HF) that covers $3 < |\eta| < 5.2$ and are placed at $|z| = 11.15$ m from the centre of the detector and, an hadron outer calorimeter (HO) mounted outside the magnetic coil in the barrel region [37]. The forward calorimeter is made up from different materials, steel is used as absorber material and quartz fibbers are used as active medium. A schematic view of one quarter of the CMS HCAL is shown in Fig. 2.10.

The combined energy resolution of the CMS HCAL and ECAL [35] is given by

$$\frac{\sigma(E)}{E} = \frac{110\%}{\sqrt{E/\text{GeV}}} \oplus 9\%. \tag{2.11}$$

Where the first term is the stochastic term and the second accounts for calibration uncertainties.

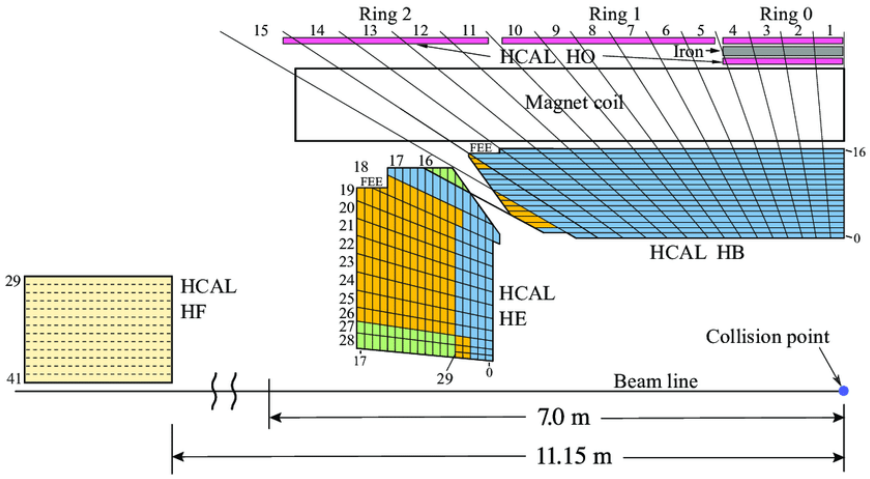

Figure 2.10: A schematic view of one quarter of the CMS HCAL, showing the positions of its four major components: the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeters [37].

2.2.5 Superconducting solenoid

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. A powerful magnet is needed to bend charged particles as they fly outwards from the collision point. As explained before this allows to measure the momentum of the particles and its electric charge. In Fig. 2.11 is represented the magnetic field strength $|\vec{B}|$ and the magnetic field lines on a longitudinal section of the CMS detector. Outside the solenoid the magnetic field is less intense and homogeneous. The direction of the magnetic field is reversed on the outside part of the detector.

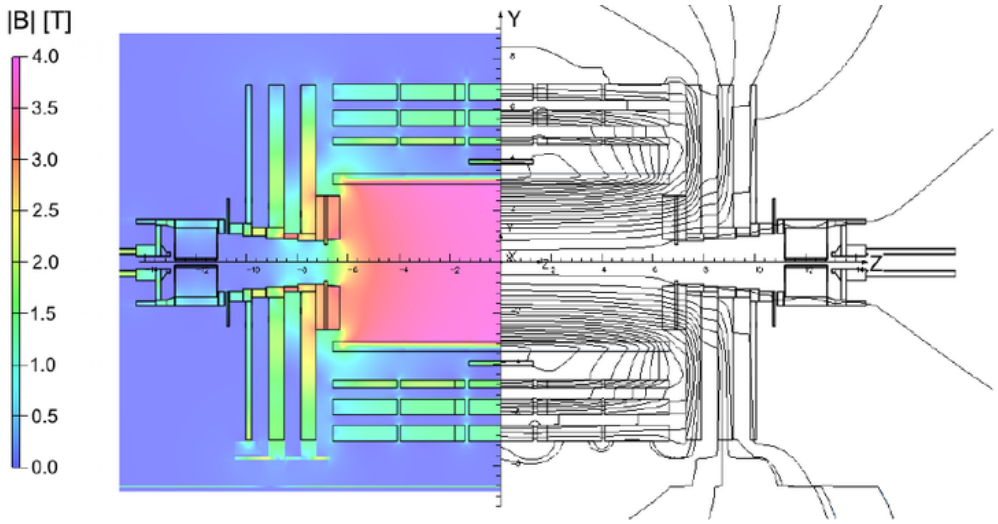

Figure 2.11: Value of $|\vec{B}|$ (left) and field lines (right) predicted on a longitudinal section of the CMS detector at a central magnetic flux density of 3.8 T [38].

The solenoid magnet is formed by a cylindrical coil of superconducting fibres over which circulates a current of 18500 A without resistance. The superconducting state is achieved cooling the solenoid down to 4.65 K. A steel yoke is used to confine the magnetic field into the volume of the detector. The magnet coils and its return yoke are the heaviest parts of CMS.

2.2.6 Muon chambers

The outermost and biggest subdetector in CMS are the muon chambers. They are placed outside the superconducting solenoid, interleaved with the iron return yoke plates and, they are used to measure the trajectory and momentum of the muons. Unlike most particles, muons are not stopped by the calorimeters. The muon chambers are gas detectors which detect particles through ionization of the gas. These hits in the chambers are used to reconstruct the full trajectory of the muons in the chambers and, combining the information with the tracker tracks, is possible to reconstruct the full trajectory along CMS.

Figure 2.12 shows one quadrant of CMS with the muon detectors in colour. There are three detector technologies that compose the muon system: the drift tubes (DTs) in the barrel region, the resistive plate chambers (RPCs) in the barrel and endcaps and the cathode strip chambers (CSCs) in the endcaps. In total, CMS has 250 DTs [39], 1056 RPCs [40] and 540 CSCs [39] that provide a pseudorapidity coverage up to $|\eta| < 2.4$.

The DTs system is divided in 5 wheels along the z axis, 12 sectors along the ϕ angle

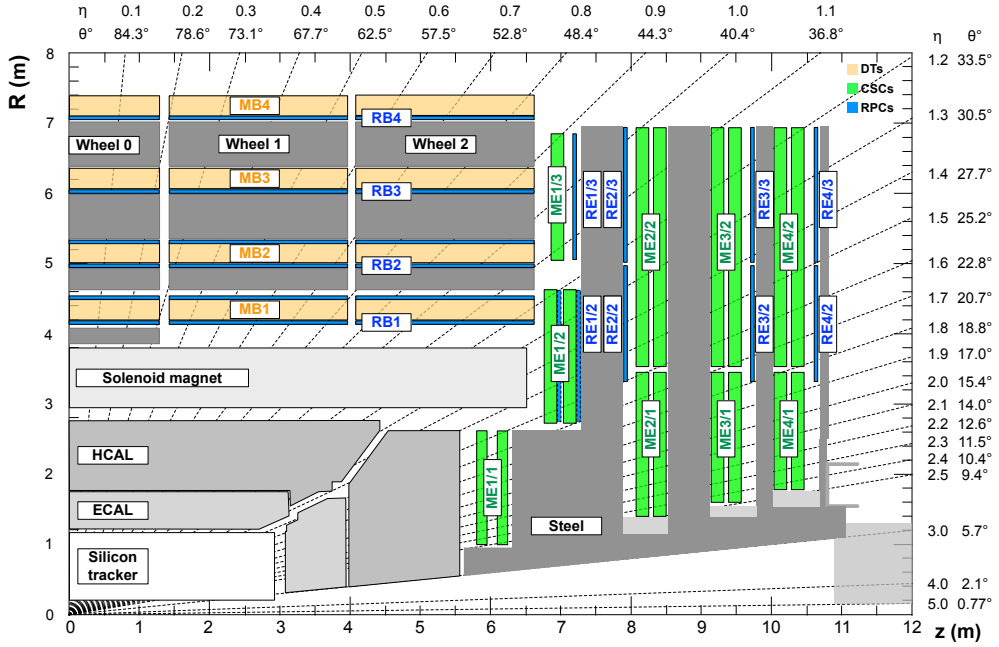

Figure 2.12: One quadrant of the CMS detector in its Run II configuration (from 2015), with the muon detectors in colour [41].

and each sector has 4 drift tubes detectors. In Fig. 2.13 is represented a DT chamber on the left. Each DT chamber is made of small groups of cells, represented on the right in Fig. 2.13, filled with a mixture of the gases Ar (85%) and CO_2 (15%). Each DT cell is 2.1 m long, 42 mm wide and 13mm tall, it contains an anode wire in the centre and two electrode plates. A high voltage of 1800 V is kept between the anode wire and the cathode strips. Lastly, the DT cells are grouped in layers and these in superlayers. One DT chamber has three or two superlayers (depending on the case) and each superlayer 4 layers. The DTs system cover the barrel region up to $|\eta| < 1.2$.

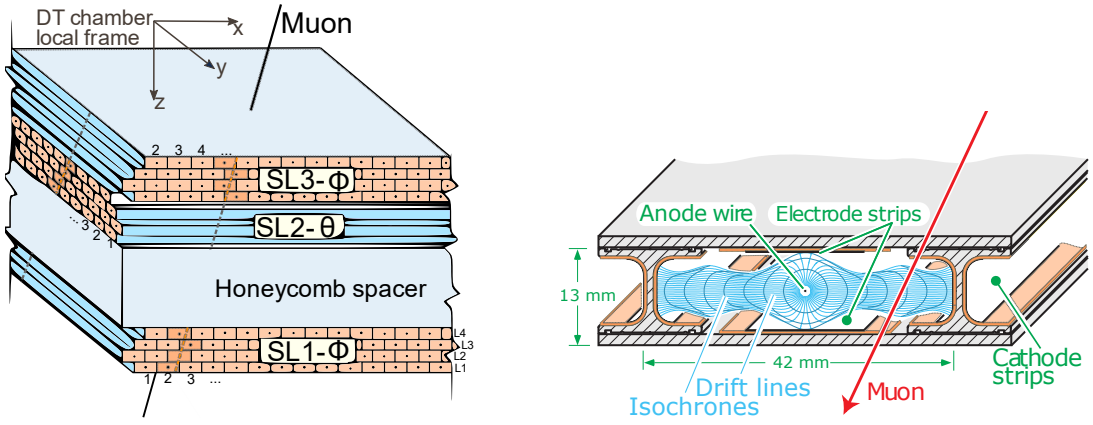

Figure 2.13: Left: Structure of a DT chamber with three superlayers composed of 4 layers of drift cells each. Right: Layout of a drift cell showing the electric field lines in the gas volume [42].

The CSCs are placed between $0.9 < |\eta| < 2.4$ in the endcap region. They consist of

positively charged anode wires crossed with negatively charged cathode strips within a gas volume. The gas composition is: 40% Ar, 50% CO_2 and 10% CF_4. They are designed to be radiation resistant because they are subjected to a high flux of neutrons. When a muon passes, it ionizes the gas atoms. Electrons move towards the wires and ions towards strips. Each CSC module contains six layers of strips and wires and are mounted in disks perpendicular to the beam line.

Lastly, RPCs are mounted together with the DTs and CSCs up to $|\eta| < 1.9$ [40]. Its spatial resolution is less than in the DTs and CSCs but they have more temporal resolution (1-2 ns). They complement the measurements of the other chambers and improve the precision in the trajectory of the muons. RPCs are parallel plate gaseous detectors made up of two gas gaps, with a strip readout plane in between. The gas mixture is: 95.2% $C_2H_2F_4$, 4.5% i-C_4H_{10} and 0.3% SF_6 [40]. The detector operational principle is based on an avalanche mode.

2.2.7 Trigger system

Due to the high luminosity of the LHC a large amount of collisions occur per second. It is not possible to store the information from all the collisions for later processing. The LHC produces collisions at a rate of 40 MHz but only a small amount of them are interesting from the physics and research point of view. As can be seen in Fig. 1.5, most of the collisions are $b\bar{b}$ events or QCD processes that were studied in previous accelerators. To select low cross section events like $W^{\pm}$, Z^0 or top quark production, a trigger system is required. An optimal performance of the trigger is necessary to reduce the collision rate down to 1 kHz. All unsaved collisions are lost forever.

The CMS trigger system selects collisions that contain certain signatures, mostly related to the presence of muons or high energy depositions on the calorimeters. The CMS trigger is divided in two stages: the Level-1 trigger (L1), based on hardware, with an output rate up to 100 kHz, and the High-Level trigger (HLT), based on software.

The L1 trigger uses information from the muon chambers and the calorimeters, leaving for the next level the tracker information. It is composed by a farm of custom hardware processors that select events with a latency of 4 μs. This level is divided into local, regional and global trigger. The L1 first creates the seeds for the objects using local information from the subdetectors. This information is combined in the regional trigger to create electron, muon and jet L1 objects. Finally the global trigger combine these objects to reject or pass the event to the HLT.

Event reconstruction at the HLT is seeded by a positive decision of one or more L1

triggers. The HLT can access information from the whole detector and perform a similar reconstruction to the one done offline. At the HLT, events are reconstructed using a computing farm of commercial computers. This HLT reconstruction is optimised to reduce the computing time for each event. The typical HLT processing time depends on the running conditions, such as pileup, but it is around 300 ms. Finally, selected events are classified and stored in different HLT paths, each one with some specific characteristics of the event.

Chapter 3

Event reconstruction and simulation

3.1 Particle Flow

In the previous chapter was described the subdetectors of CMS, each one designed to measure specific particles produced in the collisions. Event reconstruction aims to identify these collision products and measure their kinematic properties, to ultimately construct the relevant physical observables. The Particle Flow (PF) algorithm [35] performs this task by combining the information from all the subdetectors to reconstruct individual particles. These particles are: photons, charged and neutral hadrons, electrons, and muons. This list is then used in higher level reconstruction algorithms like: reconstruction of jets, calculation of the p_T^{miss} or identification of b jets among others. In Fig. 3.1 is represented a transverse slide of the CMS detector with examples of trajectories of several particles.

The inputs to the PF algorithm are sets of hits measured in the tracking system, the energy depositions in the calorimeters, as well as information of the muon system. Information on the electron and muon reconstruction is also introduced as an input [44].

The first tasks that PF does is track reconstruction of charged particles in the inner detectors and, calorimeter clustering. The starting point of the track reconstruction are hits in the tracking system and, by means of an iterative algorithm, tracks are reconstructed joining those hits. Track reconstruction consists in three different steps: initial seed generation with a few hits compatible with a charged-particle trajectory,

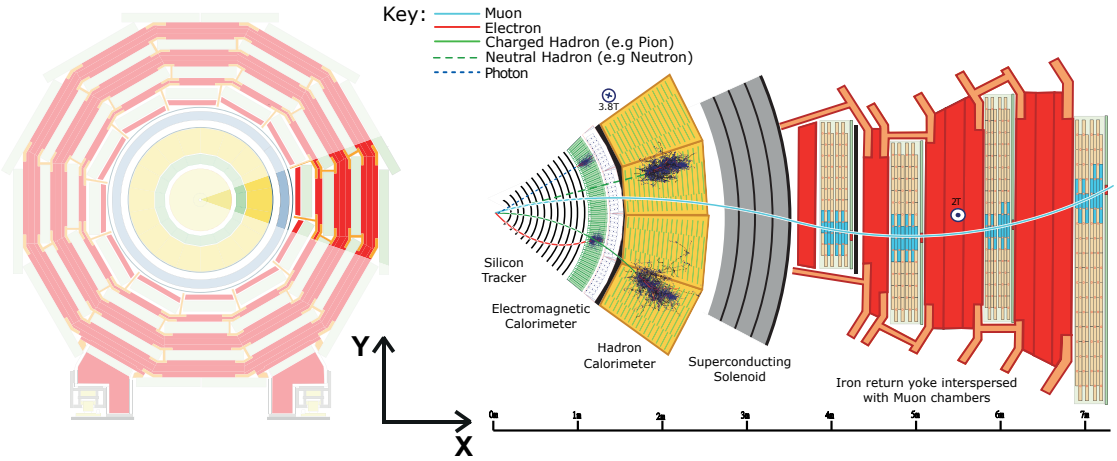

Figure 3.1: Left: transverse view of the CMS detector. Right: transverse slide of the CMS detector together with trajectories of several particles inside the detector. The muon and the charged pion are positively charged, and the electron is negatively charged [43].

pattern recognition to gather hits from all tracker layers along this charged-particle trajectory and final fit to determine the charged particle properties: origin, transverse momentum, and direction. Hits associated to a track in a given iteration are then masked for the following iterations to avoid them being associated to a different track. Pattern recognition is based on a combinatorial Kalman filter method [45]. The interaction vertex can be reconstructed extrapolating the trajectories to the beam line to see where they intersect. Due to pileup, there will be more than one interaction vertex. The primary vertex is defined to be the vertex with the highest p_T sum of the tracks associated. Calorimeter clustering starts from energy deposits in the calorimeters and joins all the deposits that are likely to come from the same particle.

Once it has been reconstructed all the information in each subdetector, PF gathers these inputs and use a linking algorithm to connect them. This step is important to identify particles based on the signatures left in the detector. Muons are identified by inner tracks associated with tracks in the muon system and, electrons, by inner tracks associated with clusters in the ECAL. Charged hadrons are detected by the inner tracker and by the HCAL deposits while neutral hadrons only by clusters in the HCAL with no tracks associated. Finally, photons only leave energy deposits in the ECAL without any track associated.

3.1.1 Muons

Muon tracking [46] is not specific to PF reconstruction. It relies on the muon system and the tracker detector. The muon system allows to identify muons with a high

efficiency over the detector acceptance. On the other hand, the inner tracker provides a precise measurement of the momentum of the muons.

Starting with the information from the CSCs and DTs (hits), a local reconstruction is performed joining the hits to form segments in each chamber. These segments are used as seeds to build a muon candidate using a Kalman filter that takes into account all the muon subdetectors: DTs, CSCs and RPCs. These muon candidates are called standalone-muon tracks.

Tracker muons are built extrapolating the inner tracks, with p_T larger than 0.5 GeV and total momentum greater than 2.5 GeV, to the muon system. If at least one muon segment matches the extrapolated track, the tracker muon is accepted. On the other hand, the reconstruction can be performed outside-in to form the so called global muons. Starting with a standalone muon, its track is propagated to the tracker detector and matched with an inner track. At large transverse momentum, $p_T > 200$ GeV, the global muon fit improves the momentum resolution with respect to the tracker only fit. For momenta below 10 GeV, tracker muon reconstruction is more efficient. About 99% of the muons produced within the geometrical acceptance of the muon system are reconstructed either as global or tracker muon and very often as both. If tracker and global muons share the same inner track, they are merged into a single category. Muons reconstructed only as standalone muons have worse momentum resolution and higher admixture of cosmic muons. Charged hadrons may be misreconstructed as muons if they reach the muon system (punch-through). Identification criteria can be applied to the muon tracks in order to obtain the desired balance between identification efficiency and purity.

After that, muons are used as inputs to the PF algorithm where more quality criteria are applied and, using the information available, isolation criteria are applied. Isolation allows to distinguish between prompt muons and non-prompt muons. Muons coming from decays at the primary vertex, generally decays of massive bosons like W or Z are called prompt muons. They are in general more isolated than non-prompt muons. The latter are wrongly identified muons, for example hadrons that reach the muon chambers.

The isolation is evaluated estimating the total transverse momentum of the particles emitted along the direction of the muon. It is computed using Eq. 3.1, where the sum is extended to all charged hadrons ($h^{\pm}$), photons (γ) and neutral hadrons (h^0) inside a cone $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ around the muon with transverse momentum p_T. A

working point is defined based on this to consider if a muon is isolated or not.

$$I = \frac{1}{p_T}\left(\sum_{h^\pm} p_T^{h^\pm} + \sum_{\gamma} p_T^{\gamma} + \sum_{h^0} p_T^{h^0}\right) \tag{3.1}$$

Several quality criteria are imposed to select muons at the analysis level defining working points to target different levels of purity and efficiency. Some examples of these categories are: loose muon and tight muon (their definition can be found in [47]). The loose muon identification criteria aims to select muons produced in prompt decays in the primary vertex, but also those produced in light and heavy flavour decays, with a very high efficiency. This working point efficiently reject charged hadrons that are reconstructed as muons. The tight muon selection aims to reject muons from decays-in-flight and hadronic punch-through. The tight selection criteria is less efficient but more pure.

3.1.2 Electrons

Electrons (and positrons) deposit almost all of their energy in the ECAL leaving also hits in the tracker layers. As an electron propagates through the material in front of the ECAL, it interacts with it. The main process is emission of photons via bremsstrahlung and the photons then converting to e^+e^-. At the end, when an electron reaches the ECAL, it is no longer a single particle but rather a shower of multiple electrons and photons. A dedicated algorithm is used to combine clusters from individual particles into a single object associated to the primary electron. Additionally, to estimate the track parameters of an electron, an algorithm based on a Gaussian sum filter (GSF) [48] is used to account for the changes in curvature due to the bremsstrahlung emission.

Electron reconstruction in CMS [49] is fully integrated into the Particle Flow framework. The reconstruction workflow starts by the formation of clusters in the ECAL by grouping crystals with energies exceeding a predefined threshold. Then, the ECAL clusters are grouped in superclusters to include photon conversions and bremsstrahlung. Tracker seeds compatible with the supercluster position are used in the GSF tracking step that is then used by the PF algorithm to link all the elements into blocks of particles. These blocks are then resolved into electron and photon objects. Electrons are identified by superclusters with a GSF track associated and photons by unlinked superclusters.

To select well reconstructed electrons (prompt electrons) additional isolation and

quality criteria are applied. The quantity

$$\left|\frac{1}{E} - \frac{1}{p}\right|, \tag{3.2}$$

where E is the energy of the electron from the supercluster and p its momentum from the track, is usually used to discriminate between prompt and non-prompt electrons. This observable tends to 0 for prompt electrons.

The isolation variables are obtained by summing the transverse momenta of charged hadrons (I_{ch}), photons (I_γ), and neutral hadrons (I_{n}), inside an isolation cone of ΔR with respect to the electron direction. The energy spread in the subdetectors increases with the energy of the incoming electron. For this reason, the thresholds applied on the isolation quantities are frequently parametrised as a function of the electron E_T. Also, the isolation variables are corrected to mitigate the contribution from pileup I_{PU}. Then, the combined PF isolation [50] is defined as

$$I_{\text{combined}} = I_{\text{ch}} + \max(0, I_{\text{n}} + I_\gamma - I_{\text{PU}}). \tag{3.3}$$

3.1.3 Jets

The experimental signature of quarks and gluons in a high energy collision are jets. Jets are groups of particles produced by the hadronisation of quarks and gluons which normally are clustered in a cone-shaped volume.

The reconstruction of jets at CMS is performed using the anti-κ_{T} algorithm [51]. This algorithm cluster PF candidates from the event into cone-shaped jets with a given angular width ΔR. For this analysis, jets are clustered with a distance parameter of 0.4. Jet momentum is determined as the vectorial sum of all particle momenta in the jet. The differences between this momentum and the true value at particle level is, on average, within 5% to 10%. Jet energy corrections (JEC) are derived from simulation so that the average measured energy of jets becomes identical to that of particle level jets. These corrections take into account several factors such as: pileup, underlying event particles and detector effects. Finally, the differences between data and simulation are also corrected.

The uncertainties associated to the JECs are usually one of the most important uncertainty sources in many analysis. Figure 3.2 gives an idea of their magnitude in terms of the jet p_T and η.

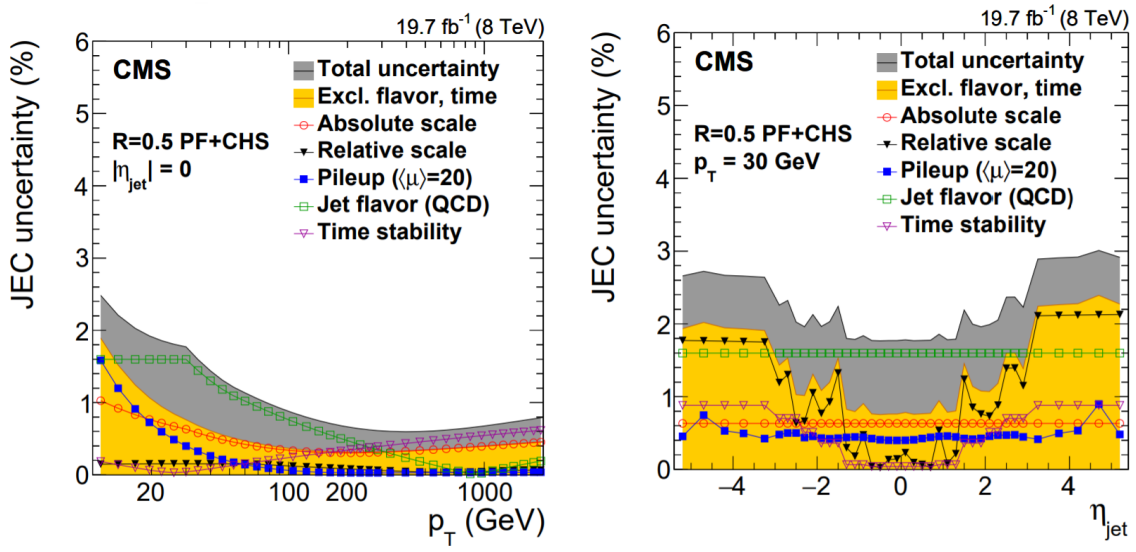

Figure 3.2: Summary of JEC uncertainties as a function of jet p_T (left) and of η_{jet} (right) at $\sqrt{s} = 8$ TeV. The markers show the single effect of different sources, the gray dark band the cumulative total uncertainty. The total uncertainty, when excluding the effects of time dependence and flavor, is also shown in yellow light. [52].

Identification of b jets

Jets which are originated from a b quark are called b jets. The hadrons formed by the hadronisation of a b quark (B^0, $B^{\pm}$ and B_s^0 among others) can travel a measurable distance from the primary vertex forming secondary vertices. This is shown in Fig. 3.3 where is represented a b jet and two light jets, produced by the hadronisation of the lightest quarks and gluons.

The b tagging algorithms exploits the characteristics of the b jets to discriminate them from the light jets. These algorithms are normally based on multivariate techniques (MVA) that takes into account all the input variables simultaneously to give a probability as an output. This probability indicates whether the input jet is originated from a b quark or not. Several working points are defined in an equilibrium of efficiency and purity. These categories are: loose, medium and tight which are selected to have a misidentification rate of 10%, 1% and 0.1% respectively. The most recent algorithms are based on neural networks, for example: CSVv2 [53], DeepCSV [53] and Deep Jet [54].

The identification of b jets is crucial in studies involving quarks top. Quarks top decay almost always into b quarks and W bosons, identifying those b quarks is a good discriminator of the background processes.

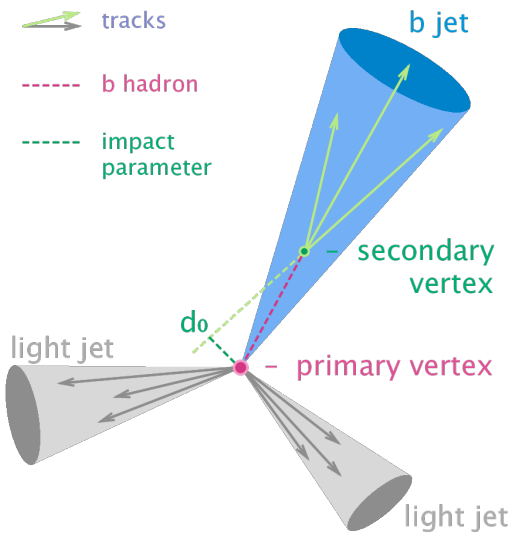

Figure 3.3: Representation of two light jets in grey and one b jet in blue.

3.1.4 Missing transverse momentum

The missing transverse momentum p_T^{miss} was defined in section 2.2.1 as the momentum imbalance in a collision. This imbalance could be produced by neutrinos, which are essentially invisible to the detectors, or, undetected particles that escape the detector due to the inefficiencies or through regions without instrumentation. It is calculated using both clustered and unclustered PF candidates.

This observable is widely used in BSM studies and also in processes with W bosons and quarks top.

3.2 Simulation of events

The purpose of simulated events in CMS is twofold. On the one hand, it provides a way to compare experimental data with the theory predictions, in this case, the SM. On the other hand, simulations are also useful to understand the detector performance and reconstruction and trigger efficiencies.

The complexity of the processes that occur in the LHC makes infeasible to obtain analytic results, instead, Monte Carlo methods are used to simulate the collisions. This technique essentially consists in generating pseudorandom experiments that reproduce the processes that occur in the LHC.

Event simulation consists in three steps. First, the collision and the processes that took place are simulated in a step called generation. Then, the second step is the simulation of particle interactions with the detector materials and all the processes involved like

multiple scattering and bremsstrahlung among others. Finally, the third step is the digitalization, where the electronic signals of the detectors produced by the interactions of particles with the active materials is simulated.

The reconstruction procedure for the Monte Carlo simulations is the same than in the real data. This is done to reproduce the results in a more reliable way.

3.2.1 Generation

In this step all the possible Feynman diagrams are generated. As there is an infinite number of them, only the most relevant ones are generated. In this sense, generators are classified in terms of the degree of approximation that they are able to simulate: LO, NLO, NNLO and so on. The expansion is normally made in powers of α_S, but also in EWK powers. Examples of generators are: Madgraph [55] at LO and PowHeg [56] and aMC@NLO [57] at NLO. These generators allow to compute the total cross section of the process, but also its differential cross section, particles involved and their distribution of energy and momentum. The generation consists of several stages, an illustration of all of them is showed in Fig. 3.4.

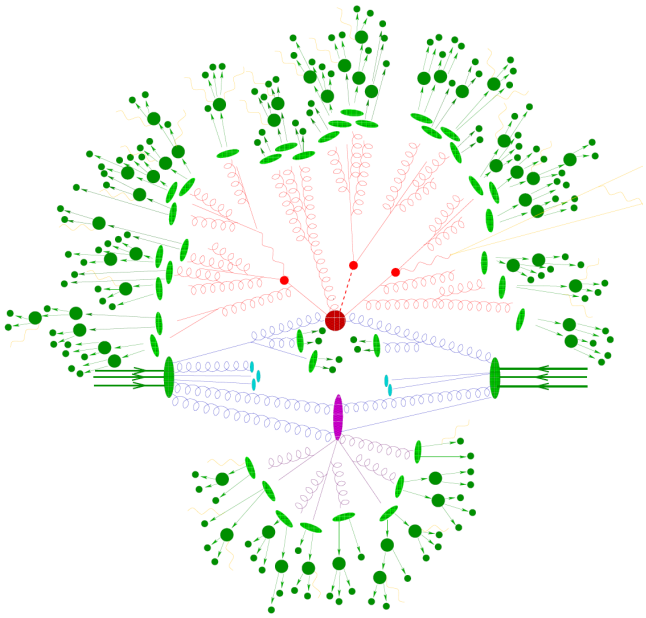

Figure 3.4: Representation of an event as produced by an event generator. The hard interaction (big red blob) is followed by additional hard QCD radiation (red). Underlying event takes place (purple blob) before the final state partons hadronise (light green blobs) and hadrons decay (dark green blobs). Photon radiation occurs at any stage (yellow).

In the first stage the hard-scattering is simulated from the PDFs of the protons. In

this stage all the Feynman diagrams up to a certain order are simulated and also the matrix elements. The second stage is the parton shower that simulates the gluon and quark emission processes from the coloured particles in the event. This can occurs before the hard-scattering and it is called Initial State Radiation (ISR) or, after the hard-scattering and is called Final State Radiation (FSR).

The partons that do not participate in the principal process can also interact. These processes are less energetic and they are called underlying-event. They belong to the non-perturbative regime of QCD and their simulation rely on phenomenological models with certain parameters that have to be tuned from data.

The next step is the hadronisation of all coloured particles in the event to form hadrons. This is also under the non-perturbative regime of QCD where α_S is near 1. The simulation of the parton showering, underlying event and hadronisation is usually done with Pythia [58, 59].

Finally, all unstable particles in the event decay into more stable particles that can travel from the beam spot and then being detected by CMS. These particles are essentially: $e^{\pm}$, $\mu^{\pm}$, γ, $\pi^{\pm}$, $K^{\pm}$, K^0, $p^{\pm}$ and n.

The effect of the pileup in the events should also be taken into account in the simulations. This task is also done by Pythia adding this effect by taking into account the luminosity profiles of each year.

At the end, the result from the generation is a set of particles with their energy and momentum which offer a very complete description of the collision. However, to compare with data is necessary the event description at detector level where the particles have interacted with the detector. So the next step is the simulation of the interactions of stable particles with the detector.

3.2.2 Simulation and digitalization

After the generation stage, the interaction of stable particles with the detector materials is simulated. A simulation of the whole detector is implemented using the Geant4 [60, 61, 62] simulation toolkit. This takes into account all the detector characteristics, materials and geometry. The results are the energy deposits and electronic signals from each subdetector. The simulated electronic signals can be used to reconstruct the event as in real data. These simulations are very precise and computationally expensive.

Chapter 4

Study of the tW process

In this Master's Thesis I perform a measurement of the tW process in the dilepton channel using data collected by the CMS detector in proton-proton collisions at $\sqrt{s} =$ 13 TeV. The objective is to measure the inclusive and differential cross sections of tW and study the uncertainty sources that most affect the measurement.

The content of this Master's Thesis is planned to be published in an article where a measurement of the tW process using the full Run 2 data will be presented. At the time of writing this thesis this article is under internal revision by the CMS collaboration. The blinding policies of CMS forbid a this moment including all Run 2 data to avoid any bias in the measurements. For this reason, only expected values, predicted with the Monte Carlo simulations, of the inclusive and differential cross sections will be given in the results using the full Run 2 dataset (Chapter 5). However, measurements using only 2016 data are allowed since they were published in the past [1, 2].

4.1 Signal and backgrounds

The selected dilepton final state for the signal process tW corresponds to the decay chain: $tW \rightarrow bW^+ + W^- \rightarrow b + e^{\pm}\nu_e + \mu^{\mp}\nu_{\mu}$. The experimental signature will be two oppositely-charged different flavour leptons $e^{\pm}\mu^{\mp}$, one jet resulting from the fragmentation of the b quark and missing transverse momentum from the neutrinos. This final state is also possible for the $t\bar{t}$ process as shown in Fig. 1.8. Due to its similarity in the final states with tW and its high cross section, $t\bar{t}$ is going to be the main background of the analysis.

Other background sources are: Drell-Yan (DY), VV+$t\bar{t}$V (where V refers to the vector

bosons W, Z and γ) and non-W/Z. The DY process shown in Fig. 1.3 takes place when a quark and an antiquark annihilate producing a Z boson or a virtual photon that then decays into a pair of oppositely-charged same flavour leptons. If these leptons are a $\tau^+\tau^-$ pair, one can decay in an electron and the other in a muon. This will give the same final state of tW without a b jet. Requiring the presence of one identified b jet in the final state will reduce this uncertainty source. However, due to its large cross section ($\sim$ 6000 pb) it will contribute in the measurement.

The VV+t$\bar{\text{t}}$V background refers to a combination of the diboson processes WZ, ZZ and WW (VV), the triboson processes WWW, WZγ, WWγ, WWZ, WZZ, ZZZ (VVV) and, the production of t$\bar{\text{t}}$ in association with a gauge boson t$\bar{\text{t}}$Z, t$\bar{\text{t}}$W and t$\bar{\text{t}}\gamma$ (t$\bar{\text{t}}$V).

Lastly, the non-W/Z background gives a small contribution with events were a non-prompt lepton is identified as a prompt lepton. This category groups two processes: W+jets and semileptonic t$\bar{\text{t}}$ where one W boson decays leptonically and the other in quarks. Due to detector inefficiencies one jet can be identified as a lepton and give a similar final state as the selected for tW. In general, this only occurs a small percentage of all the events but, due to the huge cross sections of W+jets (61527 pb) and semileptonic t$\bar{\text{t}}$ production (365 pb), they yield a sizeable contribution. In Fig. 4.1 is represented some Feynman diagrams for VV+t$\bar{\text{t}}$V and non-W/Z processes.

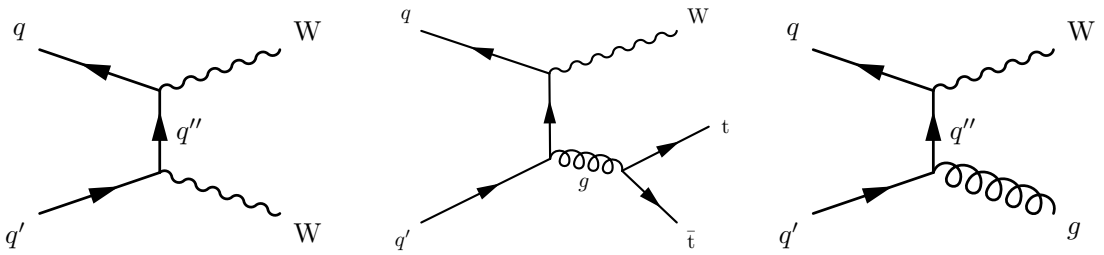

Figure 4.1: Leading order Feynman diagrams for the WW, t$\bar{\text{t}}$W and W+jets production.

The election of the $e^{\pm}\mu^{\mp}$ final state relies in the discrimination power and detection efficiency of leptons in CMS. Selecting final states with leptons allows to discriminate most of the QCD backgrounds present in an hadronic collider. Besides, selecting different flavour pairs of leptons reduces reduces resonant backgrounds such as DY which its main contribution is in the e^+e^- and $\mu^+\mu^-$ channels,

4.2 Signal and background simulation

Both signal and background processes are estimated using Monte Carlo simulations.

The signal process tW is simulated at NLO using the POWHEG v1 [63] generator with the DR method to prevent double counting with the $t\bar{t}$ samples.

The $t\bar{t}$ process is simulated at NLO in QCD using POWHEG (v2) [64]. Samples to calculate the dependency of the $t\bar{t}$ acceptance on m_{top} and on the factorisation and renormalisation scales (μ_F and μ_R respectively) are also simulated. The DY process is generated at LO and $t\bar{t}V$ processes at NLO using MG5_AMC@NLO (v2.2.2) [65] generator. The contributions from WW, WZ and ZZ are simulated at LO using PYTHIA v8.205 [66].

Parton showering and hadronisation are handled by PYTHIA using the underlying event tune CP5 [67] for all processes. The response of the CMS detector is simulated with the Geant4 [60, 61, 62] package.

Simulated events are normalised according to the integrated luminosity and the theoretical cross section of each process using Eq. 2.3.

4.3 Corrections on the efficiency for MC samples

When comparing efficiencies and the detector response between data and MC, differences are observed. These differences, due to imperfections in the simulation, have to be taken into account to correctly model the data. For this purpose, correction factors, commonly referred as scale factors, are applied to correct the difference in efficiencies in the reconstruction and identification of the events and the efficiency on simulation. In this analysis the following scale factors are applied:

Trigger scale factors To correct the differences between trigger efficiencies in data and MC, trigger SF are applied. They are represented as a function of the p_T of the two leptons in the event. As an example, in Fig. 4.2 is represented the trigger scale factors for each year.

Lepton scale factors These scale factors depend on the definition of muons and electrons and corrects the differences in the identification and isolation. They depend on the lepton p_T and η and are different for each year.

Jet energy scale and resolution These scale factors correct the differences in the jet energy scale and resolution between data and MC. They are parametrised as a function of the jet p_T and η.

B tagging scale factors B tagging scale factors are applied to correct the differences in b tagging performance between data and MC. They are parametrised as a

function of the jet flavour (three categories: jets coming from b quarks, c quarks or light quarks), p_T and η.

Pileup reweighting The simulated pileup in MC samples is corrected by applying weights to simulated events so the distribution of the number of interaction vertices matches the observed distribution.

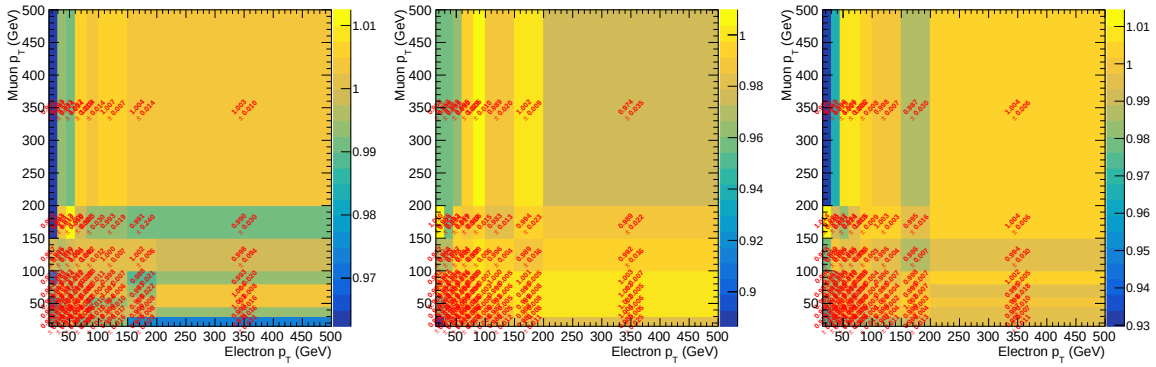

Figure 4.2: Trigger scale factors for 2016, 2017 and 2018 (from left to right) for the $e^{\pm}\mu^{\mp}$ channel. The total uncertainty is shown in each bin.

4.4 Object identification

After the event reconstruction performed by PF, a more constrained identification is done in PF candidates to select electrons, muons and jets more suitable for the analysis.

For electrons the selection criteria are:

- $p_T > 20$ GeV and $|\eta| < 2.4$.
- Veto of the region $1.4442 < |\eta| < 1.5660$.
- Relative electron isolation (REI), based on PF candidates, defined as I_{combined}/E_T where I_{combined} is defined in Eq. 3.3 and is computed inside a cone of $\Delta R = 0.3$ around the electron direction, and E_T is the transverse energy of the electron. Two different cuts are applied on the REI based on the electron η:

$$\begin{aligned} &\text{REI} < 0.0287 + 0.506/E_T; \text{ for electrons with } \eta < 1.479, \\ &\text{REI} < 0.0445 + 0.963/E_T; \text{ for electrons with } \eta > 1.479. \end{aligned} \tag{4.1}$$

- For the variable defined in Eq. 3.2

$$\begin{aligned} \left|\frac{1}{E}-\frac{1}{p}\right| &< 0.159 \text{ GeV}^{-1}; \text{ for electrons with } \eta < 1.479, \\ \left|\frac{1}{E}-\frac{1}{p}\right| &< 0.0197 \text{ GeV}^{-1}; \text{ for electrons with } \eta > 1.479. \end{aligned} \tag{4.2}$$

- Barrel: $d_z \leq 0.10$ cm, $d_{xy} \leq 0.05$ cm. Endcap: $d_z \leq 0.20$ cm, $d_{xy} \leq 0.10$ cm. Where d_z and d_{xy} are the z and transverse component of the impact parameter respectively.

The identification criteria for muon candidates from PF are the following:

- $p_T > 20$ GeV and $|\eta| < 2.4$.
- The candidate is reconstructed as a global muon.
- Relative muon isolation (RMI) < 0.15, based on PF candidates, defined as Eq. 3.1. Each component is computed in a cone of $\Delta R = 0.4$ around the muon direction, where charged PF candidates from PU events are removed (PF charged subtraction).

As mentioned in Section 3.1.3 jets are reconstructed using PF candidates with the anti-κ_{T} algorithm with an angle ΔR of 0.4. The selection criteria for jets are the following:

- $p_T > 30$ GeV and $|\eta| < 2.4$.
- Exclude jets that overlap with identified leptons (electrons or muons) if $\Delta R(\text{jet}, \ell) < 0.4$.

Another category of jets called loose jets is also defined with a relaxed p_T requirement of 20 GeV to 30 GeV. This category is useful for the analysis because of the similarity between signal and the $\mathrm{t\bar{t}}$ background. It is expected that the tW process at LO will only have one jet in the dileptonic channel, whereas $\mathrm{t\bar{t}}$ will have two. It is possible to have $\mathrm{t\bar{t}}$ events with only one jet identified in the first category and the other in the loose category. The loose jet category can be used to discriminate between $\mathrm{t\bar{t}}$ and tW events where no loose jets are expected.

The Deep Jet tagger algorithm [54] is used to identify b jets. This is useful to discriminate the tW signal from backgrounds like Drell-Yan where no b jets are expected. The medium working point is used in this analysis as it provides a good balance between efficiency (80%) and misidentification rate (1%).

4.5 Event selection

To record the events, high level trigger paths designed to select events in the dilepton channel complemented with single lepton triggers have been considered.

The event selection uses the object identification criteria defined in the previous section. First, events are selected if the two most energetic leptons (leading leptons) are an electron and a muon with opposite charge ($e^{\pm}\mu^{\mp}$ final state). This includes electrons and muons from a τ lepton with a non-hadronic decay. Then, the leading lepton is required to have $p_T > 25$ GeV. Finally, to reduce the contamination from low mass resonances, the minimum of the invariant mass of all identified lepton pairs is required to be greater than 20 GeV. This defines the baseline selection.

The targeted final state of tW suggest that the number of jets and b tagged jets could be useful to discriminate between processes. To show this, the number of jets and b tagged jets (N(jet,b tag) distribution) is represented in Fig. 4.3 (left). Based on this distribution different regions are defined according to each of its bins. There are three regions that are used in this analysis, which are denoted by: 1j1b, 2j1b and 2j2b. The 1j1b region is the region with only 1 b tagged jet (third bin of the distribution). The 2j1b region is the region with only 2 jets and one of them identified as a b jet (fifth bin of the distribution). And the 2j2b region is the region with only two b tagged jets (sixth bin of the distribution).

Notice that, as expected, the first bin corresponding to 0 identified jets is dominated by the DY process. Requiring one jet reduces this background, which is further reduced if this jet is identified as a b jet. This is the 1j1b region which has the best signal to background ratio. In the inclusive measurement this region is combined with the 2j1b and 2j2b to improve the measurement. The 2j1b region also contains a good amount of tW processes and the 2j2b region is a pure $t\bar{t}$ region which is used to constrain the uncertainties arising from this background.

Regarding the differential measurement, only the 1j1b region with the additional requirement of having 0 identified loose jets is used. This corresponds to the first bin of Fig. 4.3 (right) where is represented the number of loose jets in the 1j1b region. As mentioned before, this variable discriminates between $t\bar{t}$ and tW events.

4.5.1 Yields

The expected yields and observed data after applying the object and event selection of the regions used in both inclusive and differential measurements described above, are

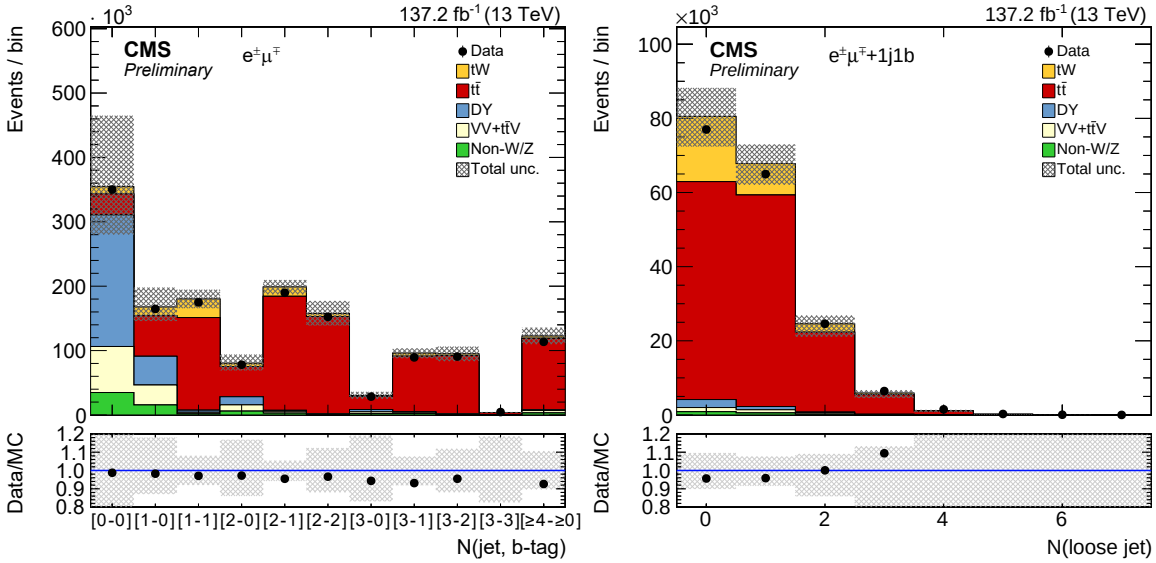

Figure 4.3: Data-MC comparisons of the number of jets and b tagged jets distribution with the baseline selection (left) and the number of loose jets in the 1j1b region (right). In the y axis is represented the total number of events. All the contributions from the processes are stacked to be able to compare with the data points. The hatched bands include the statistical and all systematic uncertainties. The bottom of each panel shows the ratios of data to the sum of the expected yields.

shown in Table 4.1.

Process	1j1b	Differential	2j1b	2j2b
tW	29000±2000	17600±1700	14500±1400	4800±700
$t\bar{t}$	144000±12000	59000±7000	177000±12000	151000±18000
Drell-Yan	3300±1400	2200±900	1500±700	210±100
VV+$t\bar{t}$V	2400±1000	1100±500	2900±1200	1400±600
Non-W/Z	1700±700	900±400	2800±1200	250±120
Total	180000±13000	81000±7000	199000±12000	158000±19000
Data	174800±400	77000±300	189800±400	152300±400

Table 4.1: Total number of events observed in data and the number of signal and background events expected from simulation in the regions considered in the analysis. The uncertainties include both statistic and systematic sources.

4.6 Uncertainty sources

In this section, the uncertainty sources that affect the measurement of the tW inclusive and differential cross sections are described. Depending on its nature uncertainties are classified as statistical or systematical.

4.6.1 Statistical uncertainties

The statistical uncertainties arise from the fact that we are performing counting experiments of processes naturally random. If N is a random variable that follows a Poisson distribution, then, the statistical uncertainty associated is $\sqrt{N}$. In this case, N would be the number of observed events.

4.6.2 Systematic uncertainties

Apart from statistical sources of uncertainties, the measurement of the tW cross section is affected by systematic uncertainties that originate from both detector effects and theoretical assumptions. Each source of systematic uncertainty is assessed individually by suitable variations of the MC simulations or by variations of parameter values in the analysis within their estimated uncertainties. In the description of each uncertainty source, the correlation between years will be indicated. Also, the correlation between different processes will be indicated for the modelling uncertainties.

In this explanation, the systematic uncertainties are divided in three groups: experimental, modelling and background normalisation sources.

Experimental sources

Jet energy scale and resolution When jets are reconstructed, the value of their energy carry an uncertainty which is divided in two parts: its scale and its resolution. The uncertainty due to the limited knowledge of the jet energy scale (JES) and jet energy resolution (JER) is determined by varying them within the uncertainties in bins of p_T and η, typically by a few percent [68]. The JES uncertainties are propagated to the $\vec{p}_T^{\,\text{miss}}$ and are separated in multiple components which can be correlated or uncorrelated across years, whereas the uncertainty in the resolution is uncorrelated per year.

B tagging efficiency The uncertainties resulting from the b tagging efficiency and misidentification rate are assessed by varying, within their uncertainties, the b tagging scale factors of the b jets and the light-flavor jets, respectively. These uncertainties vary with the p_T and η of the jet and amount to approximately 2% for b jets and 10% for mistagged jets [69], as determined in simulated $\mathrm{t\bar{t}}$ events. They are splitted into one correlated source across years, and another source uncorrelated per year.

Trigger and lepton identification The uncertainties in the trigger and

lepton identification efficiencies in simulation are estimated by varying data-to-simulation scale factors by their uncertainties. These are about 0.7% and 1.5%, respectively, with some dependence on the lepton p_T and η. For the case of muons, an additional uncertainty of 0.5% for the relative isolation is added quadratically. The trigger sources are uncorrelated across years, the same way as the statistical part of the lepton identification uncertainties. In the case of the systematic part of these ones, it is correlated through years.

Muon energy scales To account for the uncertainties in the muon energy scales, the momentum of the muons is varied by their uncertainties, taken from the muon scale corrections. They are uncorrelated across years.

Pileup The uncertainty assigned to the number of pileup events in simulation is obtained by changing the inelastic pp cross section, which is used to estimate the pileup in data, within its uncertainty of ±4.6% [70]. This uncertainty is fully correlated across the three years.

Luminosity The uncertainty on the integrated luminosity is estimated to be 2.3% for 2017 [71] and 2.5% for 2016 and 2018. The total uncertainty is split in different sources partially correlated across years.

Modelling sources

To estimate uncertainties related with the modelling, more Monte Carlo samples are used. They are determined by repeating the analysis and replacing the standard POWHEG+PYTHIA $t\bar{t}$ or tW simulation by dedicated simulation samples with altered parameters. The difference with the nominal is taken as an uncertainty.

Matrix element (ME) scale The uncertainty in the modeling of the hard-production process is assessed by changing independently μ_R and μ_F in the POWHEG sample by factors of 2 and 0.5 relative to their common nominal value. This variation is correlated through years, but performed separately for $t\bar{t}$ and tW events.

Parton shower In order to take into account parton-shower (PS) uncertainties, different effects are studied:

- Underlying event: The underlying event is modelled using phenomenological models with parameters that are tuned to reproduce real data. The uncertainty in those parameters is then propagated through the analysis to see the impact in the measurement. This variation is correlated through

years and between $t\bar{t}$ and tW events.

- ME/PS matching: The uncertainty in the combination of the ME calculation with the parton shower is taken correlated through years and only considered for $t\bar{t}$ events.

- Initial- (final-) state radiation scale: The PS scale used for the simulation of the initial- (final-) state radiation is varied up and down by a factor of two. These variations are motivated by the uncertainties in the PS tuning [72]. This variation is correlated through years, but performed separately for $t\bar{t}$ and tW events.

- Colour reconnection: The effect of multiple parton interactions and the parameterisation of colour reconnection have been studied in Ref. [73] and are varied accordingly in simulated $t\bar{t}$ and tW events. In addition, a simulation including colour reconnection of early resonant decays is used. The uncertainties that arise from ambiguities in modeling colour-reconnection effects are estimated by comparing the default model in Pythia with two alternative models of colour reconnection, a model with string formation beyond leading colour [74] and a model in which the gluons can be moved to another string [75]. All models are tuned to measurements of the underlying event [73, 72]. The largest variation in each bin with respect to the nominal yield is taken as the systematic uncertainty. This variation is correlated through years and between $t\bar{t}$ and tW events.

PDF and α_S The uncertainty from the choice of PDFs is determined by reweighting the sample of simulated $t\bar{t}$ and tW events according to the 100 NNPDF3.1 replicas [76]. As they represent the contents of a diagonalised Hessian matrix, the variations are summed quadratically. Then, the uncertainty in α_S is added in quadrature. This uncertainty is correlated through years and between $t\bar{t}$ and tW events.

Top quark mass The nominal top quark mass is modified by twice the measured uncertainty. The difference with respect to the nominal results is taken as the corresponding uncertainty. This variation is correlated through years and between $t\bar{t}$ and tW events.

Top quark p_T Measurements of the differential cross section for top quark pair production [77] have shown that the transverse momentum of the top quark is softer than predicted by the PowHeg simulation. The effect of this mismodelling

of the p_T spectrum is corrected by reweighting the $t\bar{t}$ simulation to that measured in the data. The uncertainty associated to the mismodeling of the momentum of the top quark is taken as the difference with respect to the uncorrected shapes, and symmetrised. This variation is correlated through years.

DR/DS different methods The difference between the two schemes defining the signal, DR and DS is taken into account as a systematic uncertainty. This variation is fully correlated through years.

Background normalisation sources

A normalisation uncertainty of 4% [78] is used for $t\bar{t}$ events. For VV+$t\bar{t}$V, DY and non-W/Z background contributions, a normalisation uncertainty of ±50% is assumed, as done in Ref. [1].

MC sample size

Lastly, the simulated samples have a statistical uncertainty associated to the total number of events simulated. I will refer to this uncertainty in the results as MC finite sample size.

4.7 Strategy for the inclusive cross section measurement

Based on the N(jet,b tag) distribution of Fig. 4.3 (left), the region with the biggest amount of signal events is 1j1b. A counting experiment could be performed to extract the signal contribution. However, the overwhelming presence of $t\bar{t}$ events makes this approach less precise than other strategies. In this analysis, to discriminate between $t\bar{t}$ and tW events, two BDTs are trained, one in the 1j1b region and the other in the 2j1b region. Then, a maximum-likelihood (ML) fit is performed to extract the signal contribution using the output distributions of the BDTs and the subleading jet p_T distribution in the 2j2b region.

4.7.1 Distributions considered

As there is no single observable that clearly discriminates between $t\bar{t}$ and tW events, two independent BDTs [79], one for the 1j1b region and the other for the 2j1b region, are trained to discriminate between tW signal and $t\bar{t}$ background. BDTs outperform

single tree classifiers [80] by training a set of trees (forest) and taking their weighted vote as the prediction. Each tree is derived from the same training ensemble by reweighing their events (boosting) to mitigate the statistical fluctuations and increase the overall stability. In this analysis, the BDT implementation is provided by the TMVA package [79].

The BDTs are trained and tested using different Monte Carlo samples from the ones used for the signal extraction. From the samples used in the BDT, 70% is used for the training stage, and the last 30% for testing. As signal, the tW inclusive (and $\bar{\text{t}}$W) samples are used, while the dileptonic $\text{t}\bar{\text{t}}$ samples are used for the background.

1 jet, 1 b tag

The input variables used for the training of the BDT for the 1j1b region are:

- $p_T(e^{\pm}, \mu^{\mp}, j)$: transverse momentum of the dilepton and jet system.
- $C(e^{\pm}, \mu^{\mp}, j)$: centrality (ratio between the transverse energy $E_T(e^{\pm}, \mu^{\mp}, j)$ and the total energy $E(e^{\pm}, \mu^{\mp}, j)$) of the jet and the two leptons. Where E_T and E are calculated over the vector sum of the four-momenta of the same system.
- $m(e^{\pm}, \mu^{\mp}, j, p_T^{miss})$: invariant mass of the combination of the leptons, jet, and p_T^{miss}.
- Leading loose jet p_T: p_T of the most energetic loose jet in the event. If the event doesn't have any loose jet this variable is set to 0.
- Jet p_T.
- N(loose jet): number of loose jets in the event.

The training is performed using the $e^{\pm}\mu^{\mp}$ channel and events with only one jet identified as a b jet. To achieve the best performance I optimise the hyperparameters using as a figure-of-merit the integral under the ROC curve (the area-under-the-curve or AUC figure of merit). In this case, with the BDT setup defined by the hyperparameters in Table 4.2, the measured area under the ROC curve is 0.646. In Table 4.3 the number of signal and background events in the training and test sets is included.

Hyperparameter	Value
Number of trees	2000
Shrinkage (learning rate)	0.01
Maximum tree depth	4

Table 4.2: Hyperparameters of the BDT used in the 1j1b region.

1j1b	Training	Test
Signal	134262	57541
Background	963396	412885

Table 4.3: Signal and background events in the train and test sets for the BDT in the 1j1b region.

Regarding the hyperparameter optimisation, I have considered the following configurations. For the number of trees I explored the range $100 - 2000$, for the learning rate $0.001 - 0.1$ and, for the maximum tree depth $2 - 10$. In these different settings, the AUC showed little variation and its value was between 0.640 and 0.646.

Figures 4.4 and 4.5 show the data-MC and signal-background comparison, respectively, of the input variables used for the training in the 1j1b region. The BDT discriminant for the training and testing samples of both signal and background is shown in Fig. 4.6 and, in Fig. 4.10, the expected distribution of the BDT discriminant. Table 4.4 shows the relative importance of the input variables in the BDT.

The order of importance is determined by counting how often each variable is used to split decision tree nodes. The counts are weighted by the separation gain squared achieved by the variable and by the number of events in the node.

Variable	Relative importance
$p_T(e^{\pm}, \mu^{\mp}, j)$	0.187
$C(e^{\pm}, \mu^{\mp}, j)$	0.177
$m(e^{\pm}, \mu^{\mp}, j, p_T^{miss})$	0.177
Leading loose jet p_T	0.173
Jet p_T	0.172
N(loose jet)	0.115

Table 4.4: Relative importance of the input variables of the BDT used in the 1j1b region.

Overtraining assessment To study the overtraining of the BDT in the 1j1b region I looked at various figures of merit. In Fig. 4.6 the Kolmogorov-Smirnov (KS) goodness-of-fit (GOF) test p-values are shown. The p-values in the signal and background distributions are low, suggesting a possible overtraining in the model, although at first glance the shapes of the distributions seem very similar. However, the results from the K-S test must be interpreted carefully. First of all, because this test is intended for continuous distributions and not for binned distributions as in this case. In addition, enough events in all bins are required for it to work as expected. To cross-check, I have also done a χ^2 GOF test. Finally, the values of the AUC for the training and test sets are also compared.

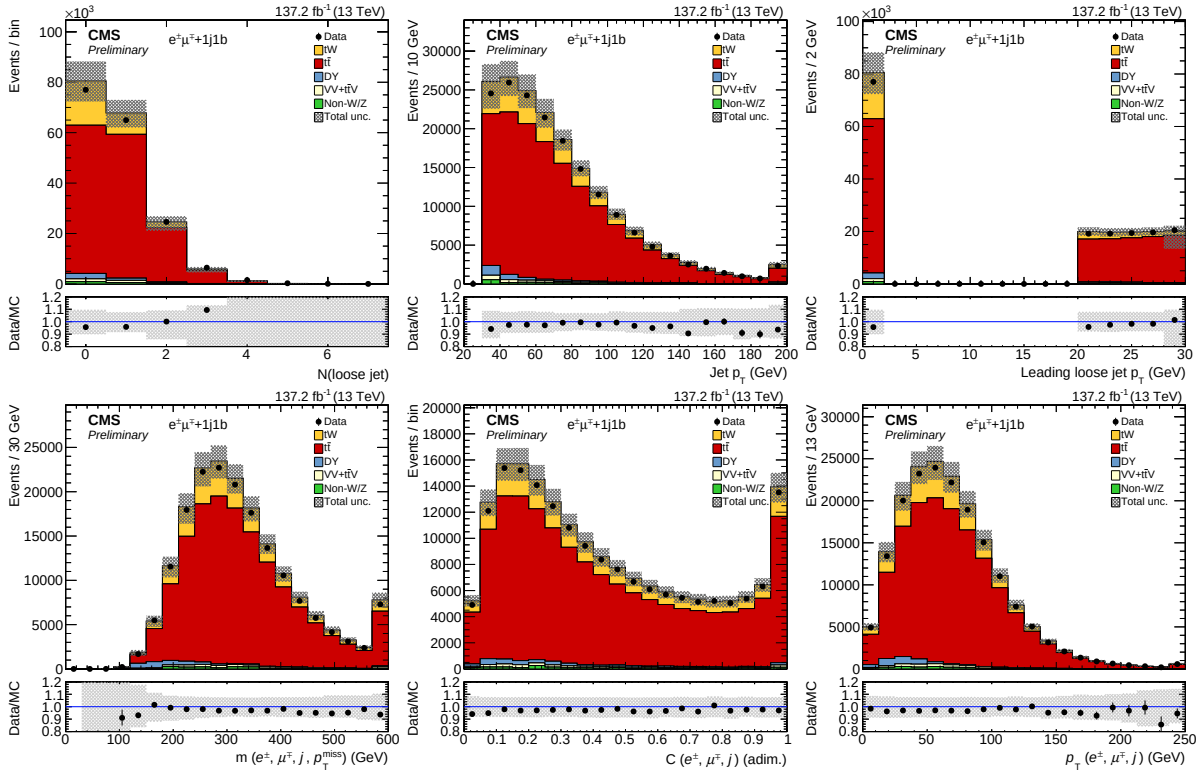

Figure 4.4: Data-MC comparison of the input variables used for the BDT trained in the 1j1b region. From top left to bottom right: N(loose jet), jet p_T, leading loose jet p_T, $m(e^{\pm},\mu^{\mp},j,p_T^{miss})$, $C(e^{\pm},\mu^{\mp},j)$ and $p_T(e^{\pm},\mu^{\mp},j)$.

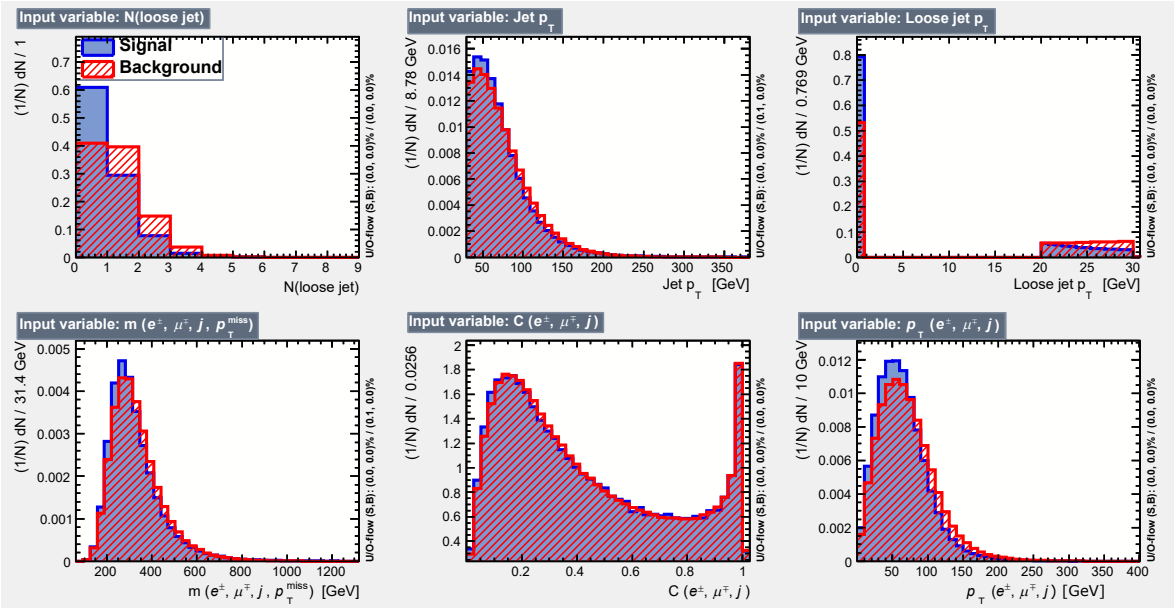

Figure 4.5: Signal-background comparison of the input variables used for the BDT trained in the 1j1b region. From top left to bottom right: N(loose jet), jet p_T, leading loose jet p_T, $m(e^{\pm},\mu^{\mp},j,p_T^{miss})$, $C(e^{\pm},\mu^{\mp},j)$ and $p_T(e^{\pm},\mu^{\mp},j)$.

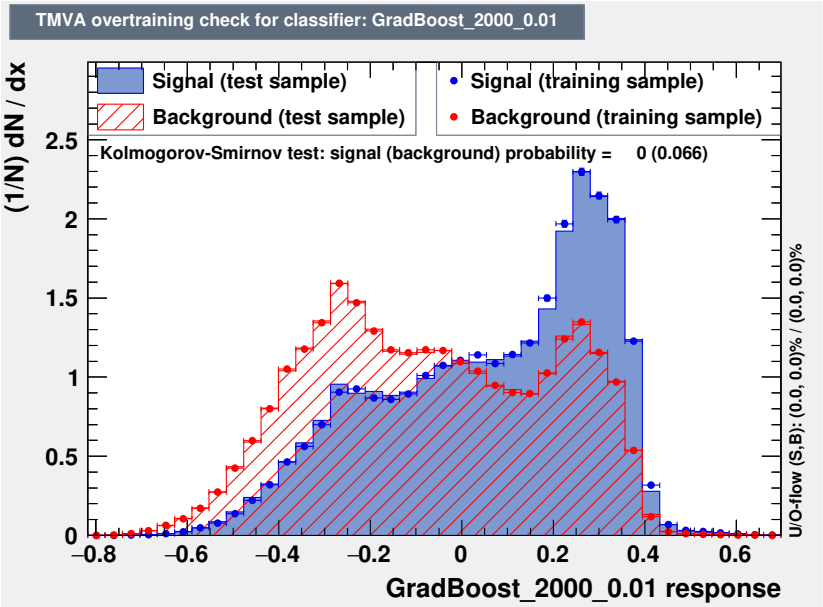

Figure 4.6: Signal-background comparison of the BDT output in the 1j1b region for the training and test samples.

The values of the χ^2 test and also the K-S test are included in Table 4.5, and, in Table 4.6, the AUC for training and testing samples. The relative difference between the training and testing AUC is 0.02%. This very low difference (and the larger p-values obtained with the χ^2 GOF test) suggest that the model is not overtrained, or that the amount of overtraining is very low.

1j1b	K-S	χ^2
Signal	0.000	0.355
Background	0.066	0.102

Table 4.5: Kolmogorov-Smirnov and χ^2 goodness-of-fit tests for signal and background distributions of the BDT in the 1j1b region.

1j1b	AUC
Train	0.6464
Test	0.6463

Table 4.6: Area-under-the-curve (AUC) of the BDT in the 1j1b region for training and testing samples.

2 jet, 1 b tag

The input variables used for the training of the BDT for the 2j1b region are:

- $\Delta R(\ell_1,\ jet_1)$: separation in the ϕ - η space between the leading lepton and the leading jet.

- $\Delta R(\ell_{12},\ jet_{12})$: separation in the ϕ - η space between the dilepton and dijet systems.
- Subleading jet p_T: p_T of the second most energetic (in terms of p_T) jet in the event.

The BDT for this region is trained using the same samples as in the 1j1b region but changing the requirement in the number of jets to have exactly 2 jets and one of them identified as a b jet. In Table 4.7 the number of signal and background events in the training and test sets in this region is included. They are different from the ones in the 1j1b region because of the new selection criteria. Tuning the hyperparameters again to achieve the best performance, the best value found for the AUC was 0.597. The ranges of the hyperparameters that I have explored are: $50 - 800$ for the number of trees, $0.001 - 0.5$ for the learning rate and $2 - 10$ for the maximum tree depth. For these ranges of the hyperparameters, the AUC value was in the $0.584 - 0.597$ interval. The hyperparameters chosen for this BDT are in table 4.8.

2j1b	Training	Test
Signal	65156	27924
Background	1102516	472507

Table 4.7: Signal and background events in the train and test sets for the BDT in the 2j1b region.

Figures 4.7 and 4.8 show the data-MC and signal-background comparisons, respectively, for the input variables. The relative importance of these variables in the BDT are in Table 4.9. The signal-background comparison of the BDT output in the 2j1b region is shown in Fig. 4.9 and, the expected distribution in MC simulation, in Fig. 4.10.

Hyperparameter	Value
Number of trees	200
Shrinkage (learning rate)	0.05
Maximum tree depth	4

Table 4.8: Hyperparameters of the BDT used in the 2j1b region.

Variable	Relative importance
$\Delta R(\ell_1,\ jet_1)$	0.340
$\Delta R(\ell_{12},\ jet_{12})$	0.337
Subleading jet p_T	0.323

Table 4.9: Relative importance of the input variables used in the BDT used in the 2j1b region.

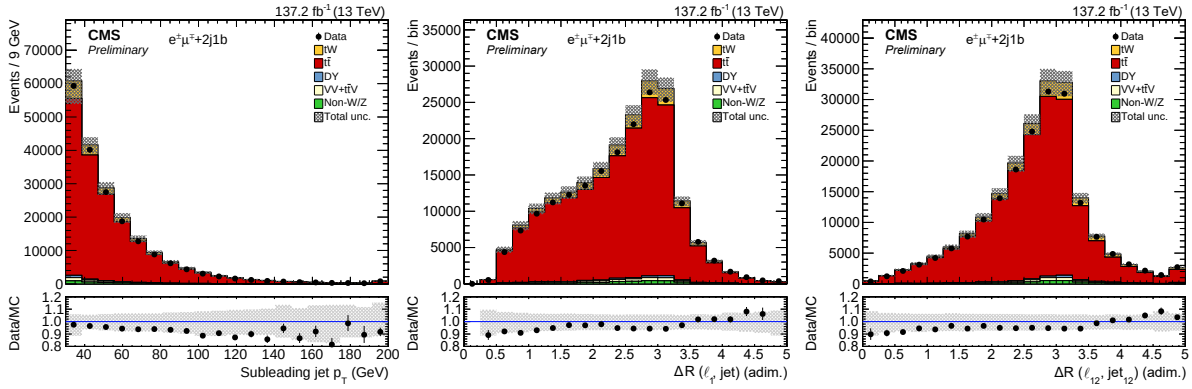

Figure 4.7: Data-MC comparison of the input variables used for the BDT trained in the 2j1b region. From left to right: subleading jet p_T, $\Delta R(\ell_1,\ jet_1)$ and $\Delta R(\ell_{12},\ jet_{12})$.

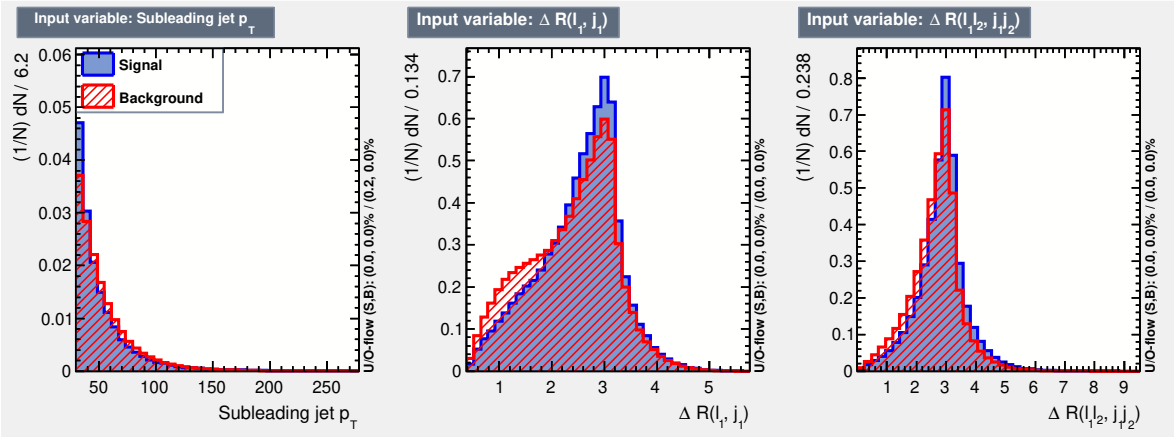

Figure 4.8: Signal-background comparison of the input variables used for the BDT trained in the 2j1b region. From left to right: subleading jet p_T, $\Delta R(\ell_1,\ jet_1)$ and $\Delta R(\ell_{12},\ jet_{12})$.

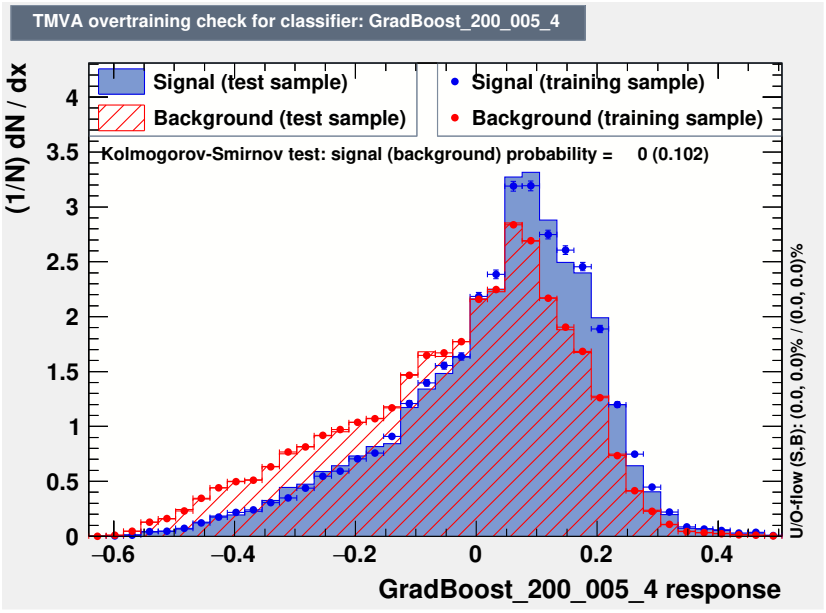

Figure 4.9: Signal-background comparison of the BDT output in the 2j1b region for the training and test samples.

Overtraining assessment The same analysis as in the 1j1b region for the overtraining of the BDT in the 2j1b region is performed. Table 4.10 shows the results of the K-S and χ^2 GOF tests on signal and background distributions and Table 4.11, the AUC for the training and test sets. The relative difference between the train and test AUC is 0.05%, very low as in the case for the 1j1b region. In addition, the p-values of the χ^2 GOF tests are also higher than its counterparts for the K-S. With all this, it has been proven that the amount of overtraining, if any, is negligible or very low.

2j1b	K-S	χ^2
Signal	0.000	0.018
Background	0.102	0.353

Table 4.10: Kolmogorov-Smirnov and χ^2 goodness-of-fit tests for the signal and background distributions of the BDT in the 2j1b region.

2j1b	AUC
Train	0.5975
Test	0.5972

Table 4.11: Area-under-the-curve (AUC) of the BDT in the 2j1b region for training and testing samples.

2 jet, 2 b tag

The 2j2b control region is highly enriched with $t\bar{t}$ events and is used to constrain the uncertainties associated to this background using the p_T distribution of the subleading jet (Fig. 5.1).

4.7.2 Signal extraction

To tackle the overwhelming presence of $t\bar{t}$ in the signal region, the analysis employs the different variables and regions described in Section 4.7.1 in a maximum-likelihood fit to extract the signal contribution. The binning of the BDT is chosen such that each bin contains the same number of $t\bar{t}$ events. This selection of binning ensures that enough background events populate all the bins of the distribution, helping to constrain the systematic uncertainties and preventing statistical fluctuations coming from uncertainty sources estimated from varied MC samples. An optimisation of the number of bins was performed in order to increase the sensitivity of the measurement. This optimisation is described in Section 4.7.3. The uncertainties are handled using different nuisance parameters, one for each systematic and for all regions, that parameterise the effect of the nuisance on the background and signal expected yields.

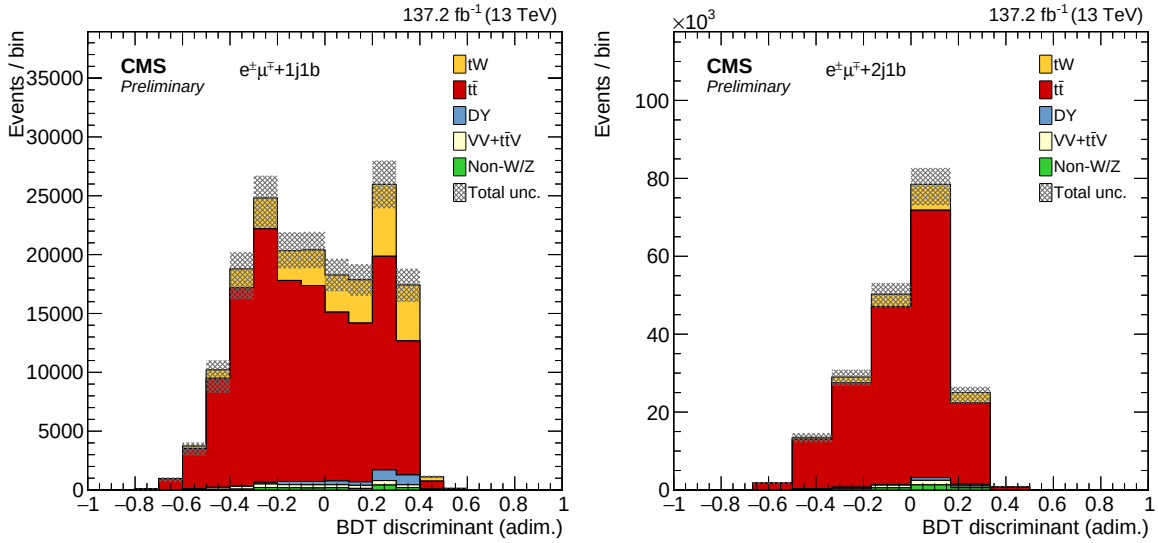

Figure 4.10: Expected distributions of the BDT output in the 1j1b (left) and 2j1b (right) regions.

The likelihood used in this ML-fit, $\mathcal{L}(\vec{n}|r,\vec{\theta})$, is a function of the signal strength, defined as the ratio of measured and expected cross sections $r = \sigma_{\mathrm{tW}}/\sigma_{\mathrm{tW}}^{SM}$, and a set of nuisance parameters $\vec{\theta}$ that parametrise the systematic uncertainties. It is constructed as the product of Poisson probabilities (Eq. 4.4) corresponding to the number of observed events in each bin of the distributions. The likelihood function is expressed as:

$$\mathcal{L}(\vec{n}|r,\vec{\theta}) = \prod_{i=1}^{N_{\mathrm{bins}}} \mathrm{Pois}(n_i|r\cdot s_i(\vec{\theta}) + b_i(\vec{\theta})) \prod_{j=1}^{N_{\mathrm{uncs.}}} p_j(\tilde{\theta}_j|\theta_j). \tag{4.3}$$

$$\mathrm{Pois}(n_i|r\cdot s_i(\vec{\theta}) + b_i(\vec{\theta})) = \frac{e^{-r\cdot s_i(\vec{\theta})-b_i(\vec{\theta})}(r\cdot s_i(\vec{\theta}) + b_i(\vec{\theta}))^{n_i}}{n_i!} \tag{4.4}$$

Where n_i is the total number of observed events in bin i and, $s_i(\vec{\theta})$ and $b_i(\vec{\theta})$, the expected numbers of signal and background events in each bin of the distributions respectively. They are obtained using normalised distributions (templates) from simulation, and are a function of $\vec{\theta}$. Additionally, the systematic uncertainties are introduced in the likelihood by multiplying it by the prior of each nuisance parameter $p_j(\tilde{\theta}_j|\theta_j)$. In this analysis, for nuisances affecting the normalisation of processes, a log-normal probability density function is used and, for shape uncertainties, a Gaussian distribution. Lastly, $N_{\mathrm{uncs.}}$ is the total number of nuisance parameters and, N_{bins}, the total number of bins of the distributions considered in the likelihood. In this case, 20 for the BDT in the 1j1b region, 12 for the BDT in the 2j1b region and 16 for the subleading jet p_T in the 2j2b region.

The objective is to find the parameters r and $\vec{\theta}$ of the model that maximise the likelihood for the observed data. Multiplying r by the expected signal yields, the signal can be extracted from the data.

4.7.3 Binning optimisation for the signal extraction

In order to improve the precision in the tW inclusive cross section measurement, the number of bins in the three fitted distributions (the two BDTs in the 1j1b and 2j1b regions, and the subleading jet p_T in the 2j2b region) is optimised. Increasing the number of bins indefinitely could produce statistical fluctuations that could lead to unphysical constraints in the nuisance parameters. This is specially observed for uncertainties estimated by varied MC samples. Statistical fluctuations also appear for low presence processes like Drell-Yan and non-W/Z background. The optimal configuration for the fit is chosen looking for a compromise where sensitivity is improved, which is done increasing the number of bins, without having significant statistical fluctuations.

The initial configuration is chosen to be the same as in the previous tW measurement [1] where only 2016 data was available. Denoting with N_x the number of bins in the distribution used for the fit in the region x, the initial configuration is $N_{\text{1j1b}} = 10$, $N_{\text{2j1b}} = 6$ and $N_{\text{2j2b}} = 8$. This must be optimised for this analysis where more data is available.

The optimisation is summarised in Table 4.12, where the expected uncertainty of the signal strength in terms of the number of bins in the three regions is shown.

$(N_{\text{1j1b}},N_{\text{2j1b}},N_{\text{2j2b}})$	$r = 1+$	$r = 1-$
(10,6,8)	0.102	0.112
(20,6,8)	0.093	0.100
(20,12,8)	0.091	0.098
(20,12,16)	**0.089**	**0.096**
(20,18,16)	0.088	0.093
(20,18,24)	0.087	0.092
(30,18,24)	0.085	0.090
(30,24,24)	0.084	0.088

Table 4.12: Expected uncertainty of different binning configurations.

The chosen binning for the three distributions is $N_{\text{1j1b}} = 20$, $N_{\text{2j1b}} = 12$ and $N_{\text{2j2b}} = 16$. This doubles the number of bins used in [1] and reduce the expected uncertainty by 1%. With this configuration no significant statistical fluctuations are observed.

4.8 Strategy for the differential cross section measurement

In the last section, the measurement of the inclusive cross section only involved the calculation of the total number of signal events. To measure the differential cross section with respect to some variable, the information that is needed is the number of signal events in each bin of that variable. For the differential measurement, the signal extraction is going to be performed by subtracting the background from data in each bin. Once the signal is extracted, it is necessary to remove the detector effects to obtain the real distribution at particle level. This step is called unfolding and allows to obtain the real distribution from the measured one. This corrects detector effects such as: particles that are not detected and as consequence some signal events are not observed (normalisation), and migration effects, for example an electron of 30 GeV that is detected as an electron of 40 GeV. Also, the unfolding corrects the statistical fluctuations that arise from taking a sampling distribution which is not exactly the real one. In this Master's Thesis, the measured distribution is referred as the detector level distribution (folded) and the real one as the particle level distribution (unfolded). Particle level is defined by the obtained particles after the generation of the strong process and after parton shower simulations. These objects are constructed over stable (lifetime larger than 30 ps) generated particles following the conventions given in Ref. [81]. In the following, the unfolding method is described.

4.8.1 Unfolding

The problem of unfolding may be written as:

$$y_i = A_{ij} x_j, \tag{4.5}$$

where x_j is the true distribution and y_i the measured one. The matrix A_{ij} parametrise the detector effects and is called response matrix. It is obtained through simulations that only contain signal events in the fiducial phase space. Table 4.13 summarises the object selection and Table 4.14 the event selection at particle level for the fiducial region. Using the information at detector level in the signal region defined in Section 4.5 for the differential measurement, the response matrix is defined as

$$A_{ij} = \frac{n_{ij}}{n_i}. \tag{4.6}$$

Where n_{ij} is the number of reconstructed events whose value of the measured variable fell in the bin j that had the simulated value of the variable in the bin i, and n_i is the number of events whose simulated value of the variable fell in the bin i. The number of bins of the variable at detector level does not have to be the same as in the particle level: actually, it can be seen that the choice of the discretisation that is used to define Eq. 4.5 can be relevant for the numerical convergence of the method. A non-optimal binning can yield to numerical instabilities that will result in large uncertainties. In this case, the chosen binning is in the relation 2:1 (detector:particle).

Muons		Electrons		Jets		Loose jets	
p_T (GeV)	$\lvert\eta\rvert$	p_T (GeV)	$\lvert\eta\rvert$	p_T (GeV)	$\lvert\eta\rvert$	p_T (GeV)	$\lvert\eta\rvert$
> 20	< 2.4	> 20	< 2.4, veto $\lvert\eta\rvert \in (1.4442, 1.5660)$	> 30	< 2.4	> 20 , < 30	< 2.4

Table 4.13: Selection requirements of particle-level objects.

Number of leptons	≥ 2
$p_T(\ell_1)$	> 25 GeV
$m_{e\mu}$	> 20 GeV
Number of jets	1
Number of loose jets	0
Number of b jets	1

Table 4.14: Definition of the fiducial region.

To choose an optimal binning there are two quantities that can be studied: stability of a bin i in the unfolded space and purity of a bin j in the folded space. They are defined as

$$s_i := \frac{\sum_{j=1}^{N_{\text{bins}}^{\text{fol.}}} n_{ij}}{n_i} \qquad p_j := \frac{\sum_{i=1}^{N_{\text{bins}}^{\text{unf.}}} n_{ij}}{n_j^R}, \tag{4.7}$$

where n_{ij} and n_i are the same as before and n_j^R is the amount of simulated events in the folded space bin j. Both purity and stability, when maximised simultaneously in all the spectre of the distribution, enhance the diagonality of the response matrices, making the unfolding problem easier.

The problem of unfolding could be uniquely solved by inverting the response matrix. However, the response matrix can have large dimension or be highly non-diagonal, so its diagonalisation can be numerically complex. Several approaches can be followed to tackle this problem. In this work, the `TUnfold` [82] implementation of Tikhonov regularisation is used.

The unfolding algorithm, as implemented in `TUnfold`, returns an estimator $\vec{x}$ of the truth distribution given a set of observations $\vec{y}$. This is done by determining the

stationary point of

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 \\ &= (\vec{y} - A\vec{x})^T V_y (\vec{y} - A\vec{x}) \\ &+ \tau^2 (\vec{x} - f_b \vec{x_0})^T (L^T L)(\vec{x} - f_b \vec{x_0}) \\ &+ \lambda (Y - e^T \vec{x}),\end{aligned} \tag{4.8}$$

where $Y = \sum_i y_i$ and $e_j = \sum_i A_{ij}$. V_y is the covariance matrix of $\vec{y}$.

The first term of Eq. 4.8 $\mathcal{L}_1$, is the expected from a least square minimisation. The term $\mathcal{L}_2$ describes the regularisation, which damps fluctuations in $\vec{x}$. Such fluctuations originates from the statistical fluctuations of $\vec{y}$ which are amplified when determining the stationary point of Eq. 4.8. The strenght of the regularisation is controlled by the parameter τ. This parameter is considered as a constant in the calculation of the stationary point of $\mathcal{L}$. One of the ways of choosing this parameter is by the L-curve method. In this method, a scan for values of τ is performed. For each value of τ, the stationary point of $\mathcal{L}$ is obtained and the values $L_x = \log \mathcal{L}_1$ and $L_y = \log \mathcal{L}_2/\tau^2$ are evaluated. This curve often shows a kink (L-shape), the optimal τ is chosen to be at that point. Finally, the term $\mathcal{L}_3$ is an optional area constraint that is necessary to account for some problems that can arise when bins with a low amount of events are present. If that happen, the count of those bins would follow a Poisson distribution, not a Gaussian, which is a necessary condition for the method.

4.8.2 Distributions considered

For the differential analysis, the following physical observables to measure the differential cross section of the process are considered:

- p_T of the leading lepton.
- p_T of the jet.
- The invariant mass of the system formed by the electron, the muon and the jet, $m(e^{\pm}, \mu^{\mp}, j)$.

The first and second variables provide information regarding the kinematic properties of the events. And the last variable $m(e^{\pm}, \mu^{\mp}, j)$, allows to study properties of the whole tW system. The data/Monte Carlo comparison plots, using the full Run 2 data, for these three variables in the signal region for the differential analysis can be seen in

Fig. 4.11. For completeness, these variables using only 2016 data are shown in Fig. 4.12 because they will be used in the unblind measurement of the differential cross section.

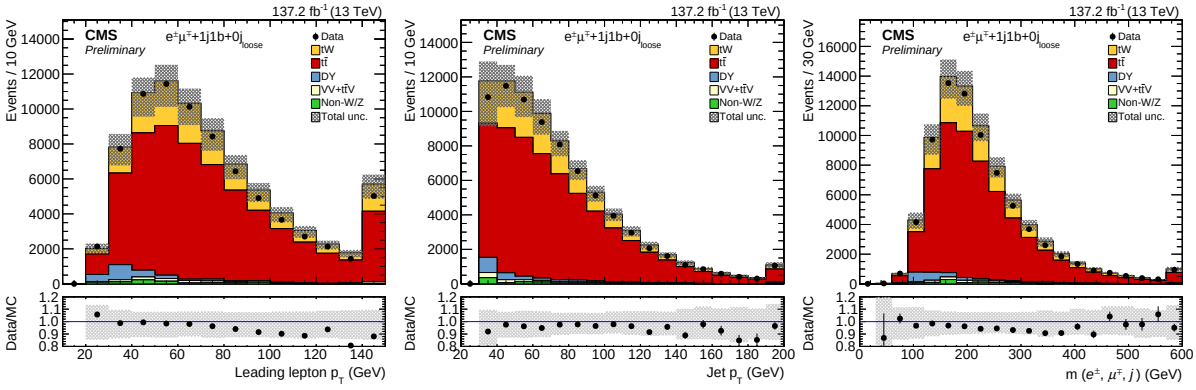

Figure 4.11: Data-MC comparison of the chosen variables for the differential measurement. From left to right: leading lepton p_T, jet p_T and $m(e^{\pm}, \mu^{\mp}, j)$. The hatched bands include the statistical and all systematic uncertainties. The last bin of each contribution contains overflow events. The bottom of each panel shows the ratios of data to the sum of the expected yields.

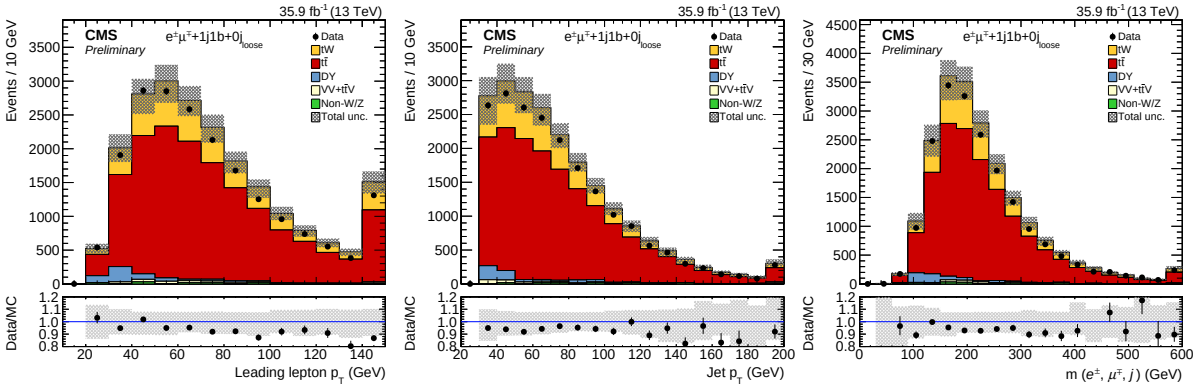

Figure 4.12: Data-MC comparison of the chosen variables for the differential measurement with 2016 data. From left to right: leading lepton p_T, jet p_T and $m(e^{\pm}, \mu^{\mp}, j)$.

The choice of the number of bins used for each distribution at particle level is the following:

- p_T of the leading lepton: 8 bins.
- p_T of the jet: 5 bins.
- $m(e^{\pm}, \mu^{\mp}, j)$: 6 bins.

Distributions of the stability and purity are shown in Fig. 4.13 and the response matrices in Fig. 4.14. As can be seen, the response matrices are quite diagonal and the values of stability and purity are reasonably high. This suggest that the additional terms in Eq. 4.8 ($\mathcal{L}_2$ and $\mathcal{L}_3$) are not necessary. This was checked by comparing the

results of performing and not performing regularisation. First the best value of τ is obtained by the L-curve method. The values of τ for the variables chosen are

- $\tau = 3 \cdot 10^{-10}$ for the p_T of the leading lepton,
- $\tau = 1 \cdot 10^{-10}$ for the p_T of the jet,
- $\tau = 0.0$ for $m(e^{\pm}, \mu^{\mp}, j)$.

Notice the low values of τ in all the cases, it is even 0 for $m(e^{\pm}, \mu^{\mp}, j)$. Then, the unfolded distributions obtained with $\tau = 0$ (no regularisation) and τ at the kink point are divided, giving results very near to one in all the bins. The same is checked for the area constrain term giving identical results when it is excluded. As a consequence, regularisation and area constrain will not be included in this unfolding.

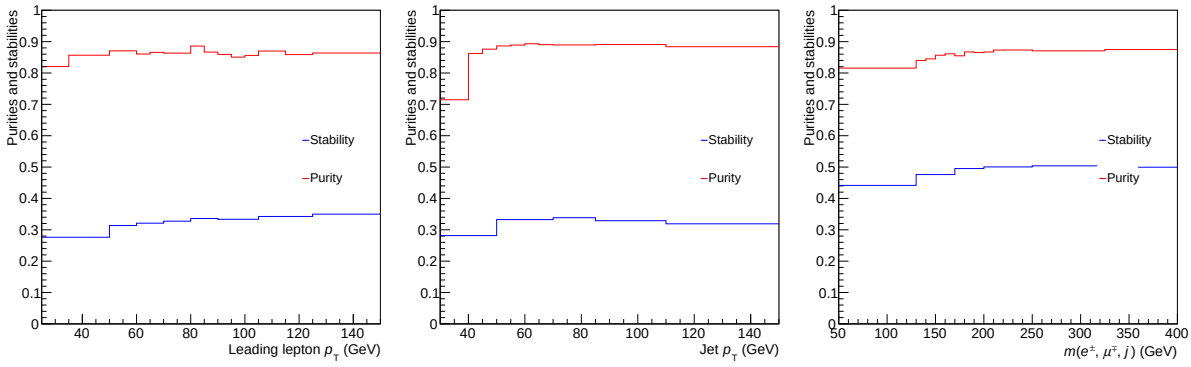

Figure 4.13: Graphs showing the purities and stabilities of each bin of the folded and unfolded (respectively) spaces for all the distributions chosen for the differential measurement.

4.8.3 Uncertainty propagation and normalisation to the fiducial cross section

The uncertainty propagation is done by first computing all the response matrices with each systematic variated $\pm 1\sigma$. Then, the signal extraction is performed and the unfolding is done with each response matrix. The propagated uncertainty is the difference between the nominal value and the variated result. Then, all the uncertainty contributions are added in quadrature to give the total uncertainty.

After performing the unfolding, the measured differential cross sections are normalised to the fiducial cross section and to the bin width. This allows to compare the shape of the distributions and to reduce the uncertainty. The uncertainties are reduced because some cancellations between the uncertainties of the fiducial and differential

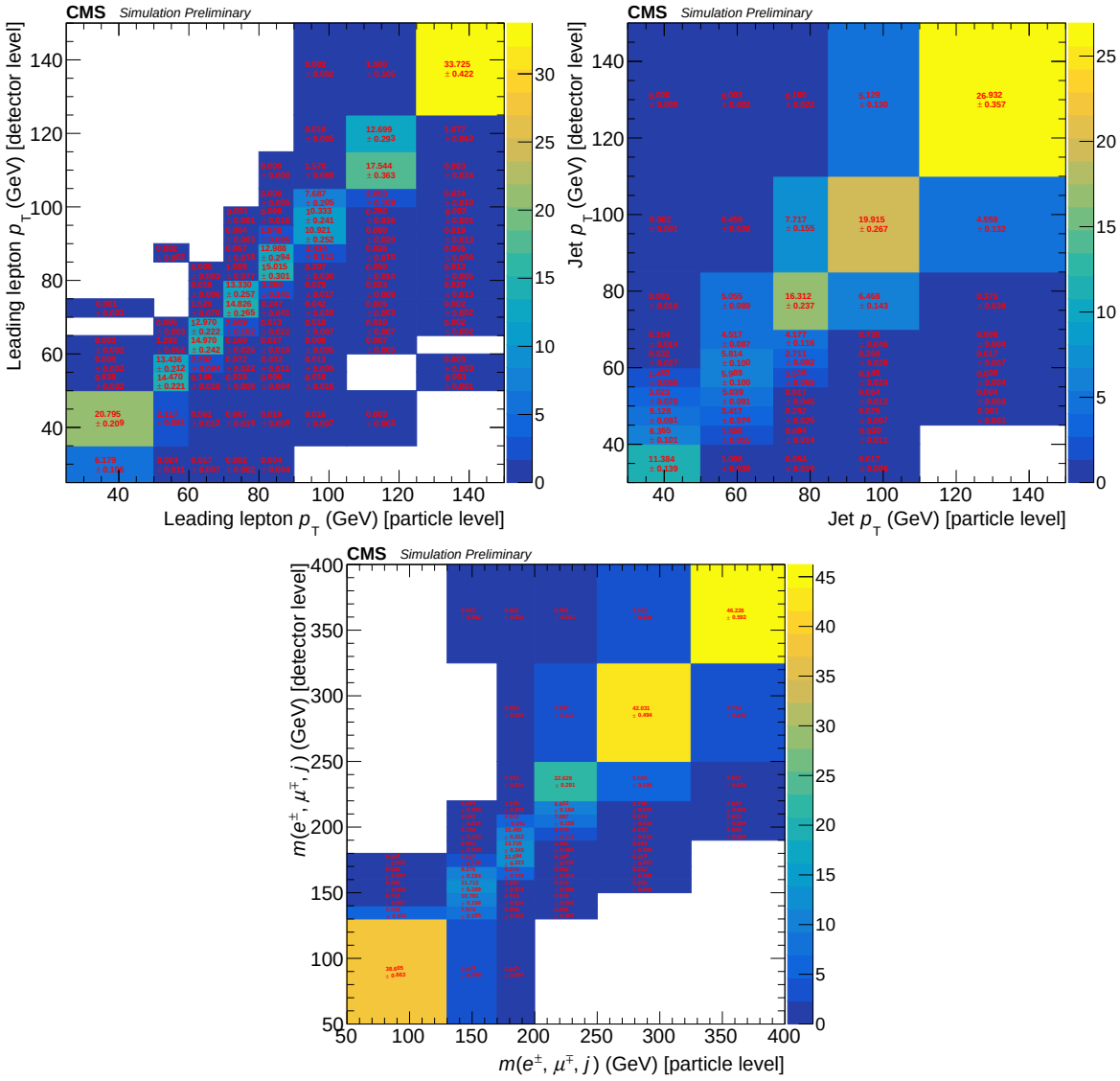

Figure 4.14: Response matrices of the chosen variables for the differential measurement for the full Run 2. They are scaled to 100 (all contents multiplied by it).

cross sections may happen. This is due to the existing correlations between the two values. The uncertainty in each bin of the differential cross section normalised to the fiducial cross section is

$$Var(d\sigma_{fid.}(i)) = \frac{Var(d\sigma(i))}{\sigma^2_{fid.}} + \frac{Var(\sigma_{fid.}) \cdot (d\sigma(i))^2}{\sigma^4_{fid.}} - \frac{2(d\sigma(i))}{\sigma^3_{fid.}} \cdot Cov(d\sigma(i), \sigma_{fid.}), \quad (4.9)$$

where i represents one bin of the differential cross section, $Var(x)$ the variance of the random variable x, $Cov(x, y)$ the covariance between two random variables x and y, $d\sigma_{fid.}(i)$ the differential cross section normalised to the fiducial cross section in the bin i, $d\sigma(i)$ the differential cross section in the bin i and $\sigma_{fid.}$ the fiducial cross section.

Chapter 5

Results

5.1 Inclusive cross section measurement

The best value for the signal strength r of tW is obtained by maximising the likelihood function (Eq. 4.3) with respect to all its parameters. The fit is performed in the three distributions of Fig. 5.1. These are, the BDT discriminants in the 1j1b and 2j1b regions and the subleading jet p_T in the 2j2b region. At this moment the analysis with the full Run 2 is still blind to the data. This means that only expected values for the signal strength and its uncertainty can be given. The result is

$$r = 1 \pm^{0.09}_{0.10} . \tag{5.1}$$

To illustrate the effect of each region in the fit, the expected signal strength computed using each region separately is:

$$r_{\text{1j1b}} = 1 \pm^{0.16}_{0.15}, \quad r_{\text{2j1b}} = 1 \pm^{0.47}_{0.43}, \quad r_{\text{2j2b}} = 1 \pm^{1}_{1} . \tag{5.2}$$

The uncertainties of the expected signal strength include the statistical and all systematic sources. The impact of each uncertainty to the total is summarised in Table 5.1 for the result using the three regions (Eq. 5.1). To compute the values of the table the procedure is the following:

1. Perform the fit including all the uncertainties. This will give the total uncertainty Δr_{total}.

2. Remove one uncertainty source "u" and compute the fit again to obtain the uncertainty $\Delta r_{\text{total}-u}$.

3. The impact of this uncertainty source will be: $\Delta r_u = \sqrt{\Delta r^2_{\text{total}} - \Delta r^2_{\text{total}-u}}$.

4. Then, include the removed uncertainty and repeat the process with all the uncertainties.

From all these results, it is clear that combining the three regions reduces clearly the expected uncertainty on the signal strength. In addition, the measurement is unmistakingly dominated by systematical uncertainty sources, being the leading ones the normalisation of the Drell-Yan background, the final state radiation and pileup reweighting.

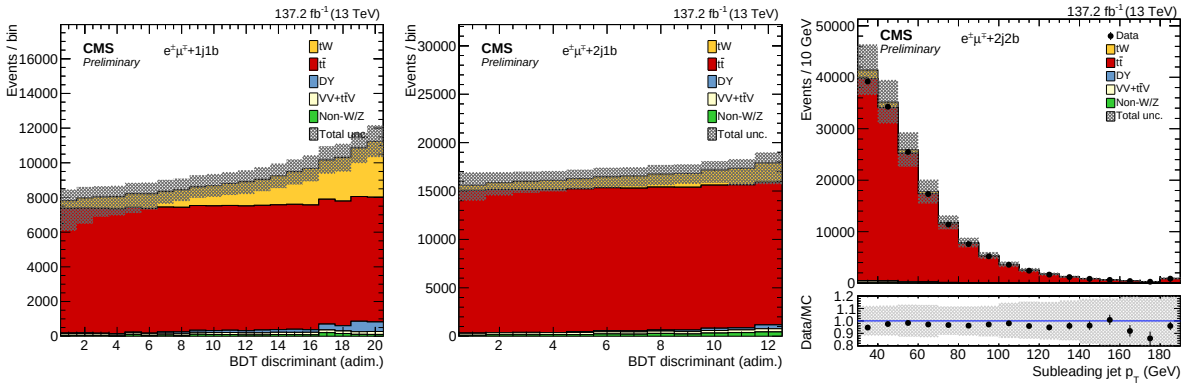

Figure 5.1: Expected distribution of the BDT output in the 1j1b (left) and 2j1b (middle) regions and the data-MC comparison of the subleading jet p_T in the 2j2b region (right). The uncertainty band includes the statistical and systematic sources.

5.1.1 Measurement using 2016 data

In this section the measured signal strength and the inclusive cross section of tW using 2016 data is presented. The strategy is the same as in the full Run 2 measurement performing the fit simultaneously in the distributions of Fig. 5.2. The result for the signal strength is

$$r_{2016} = 0.91 \pm 0.03(\text{stat.}) \pm^{0.13}_{0.14} (\text{syst.}) \pm^{0.018}_{0.016} (\text{lumi.}) = 0.91 \pm^{0.13}_{0.14} . \tag{5.3}$$

And for the inclusive cross section:

$$\sigma_{\text{tW, 2016}} = 65 \pm 1.8(\text{stat.}) \pm^{9}_{10} (\text{syst.}) \pm^{1.3}_{1.1} (\text{lumi.}) = 65 \pm^{9}_{10} \text{ pb}. \tag{5.4}$$

Source	Average unc. (%)
Systematic	
Experimental	
Trigger efficiencies	0.6
Electron efficiencies	2.2
Muon eff. and energy scales	0.5
Jet energy scale	2.4
Jet energy resolution	1.1
b tagging efficiencies	2.4
Mistagging efficiencies	0.7
Pileup	3.6
2016 and 2017 prefiring detector fix	0.4
Modelling	
$t\bar{t}$ μ_R and μ_F scales	0.3
tW μ_R and μ_F scales	1.5
Initial state radiation	1.5
Final state radiation	5.4
Underlying event	1.9
Matrix element/PS matching	1.5
Top p_T reweighting	1.5
Colour reconnection	2.5
PDFs	1.0
DR-DS	0.7
Top quark mass	0.5
Background normalisation	
$t\bar{t}$	2.7
VV+$t\bar{t}$V	1.3
Drell-Yan	4.6
Non-W/Z	2.0
MC finite sample size	3.1
Integrated luminosity	1.3
Statistical	0.9

Table 5.1: Estimation of the expected impact of each systematic uncertainty to total uncertainty of the tW signal strength of Eq. 5.1.

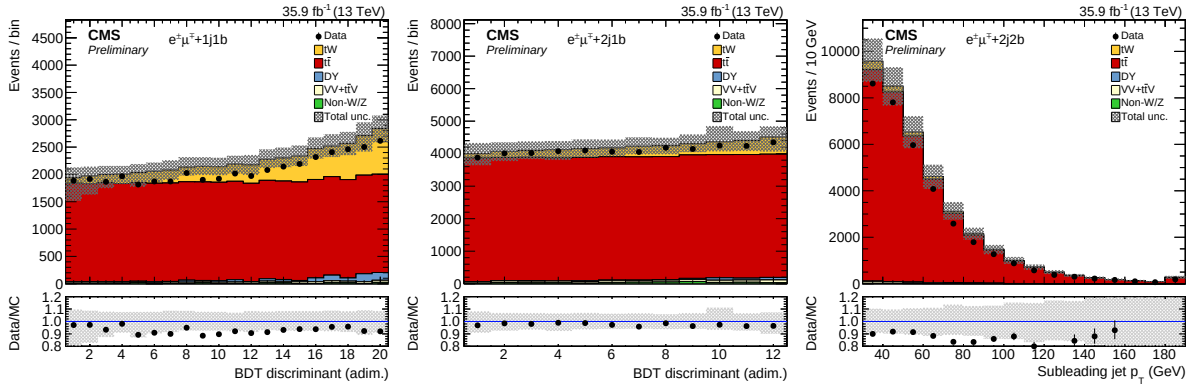

Figure 5.2: Data-MC comparison of the BDT output in the 1j1b (left) and 2j1b (middle) regions, and the subleading jet p_T in the 2j2b region (right) with 2016 data. The uncertainty band includes the statistical and systematic sources.

This result is compatible with the SM expectation, $\sigma_{tW} = 71.7 \pm 1.8(\text{scale}) \pm 3.4(\text{PDF})$ pb [17] at NNLO approximation, within the uncertainties.

5.2 Differential cross section measurement

In this section the results for the differential cross sections of the three variables chosen is presented. The results are normalised to the fiducial cross section and to the bin width and compared with the predictions from: POWHEG DR + PYTHIA 8 (sample used in the whole analysis), POWHEG DS + PYTHIA 8 to study the other scheme to define the signal, and POWHEG DR + HERWIG 7 to study other generator. As in the inclusive cross section measurement, only expected results for the differential cross sections using the full Run 2 are presented. To simulate data points the Asimov dataset is used. This dataset is constructed using the simulated results from the Monte Carlo samples.

In Fig. 5.3 the differential cross sections for the chosen variables are shown (plots on the left) together with the relative uncertainty in each bin (plots on the right). The first thing to notice is that the Asimov data points agree with the predicted values from POWHEG DR + PYTHIA 8 as it should, because the Asimov data is constructed with these simulations. The other predictions also agree within the uncertainties. In this measurement, the leading uncertainties are: top quark mass, top quark p_T, b tagging efficiency and Drell-Yan normalisation.

5.2.1 Measurement using 2016 data

Lastly, the results using the measured 2016 data are presented in Fig. 5.4. In general, the predictions from the different generators agree with the measured differential cross sections within the uncertainties. The resolution of these measurements does not allow to distinguish between the DR and DS schemes. The total uncertainty is bigger than in the full Run 2 measurement with some bins with near 1.0 relative uncertainty. This is probably due to statistical fluctuations of the systematic uncertainties estimated with variated Monte Carlo samples. For example, in these bins with near 1.0 relative uncertainty, the leading uncertainties are top quark mass and colour reconnection, which are estimated with variated samples.

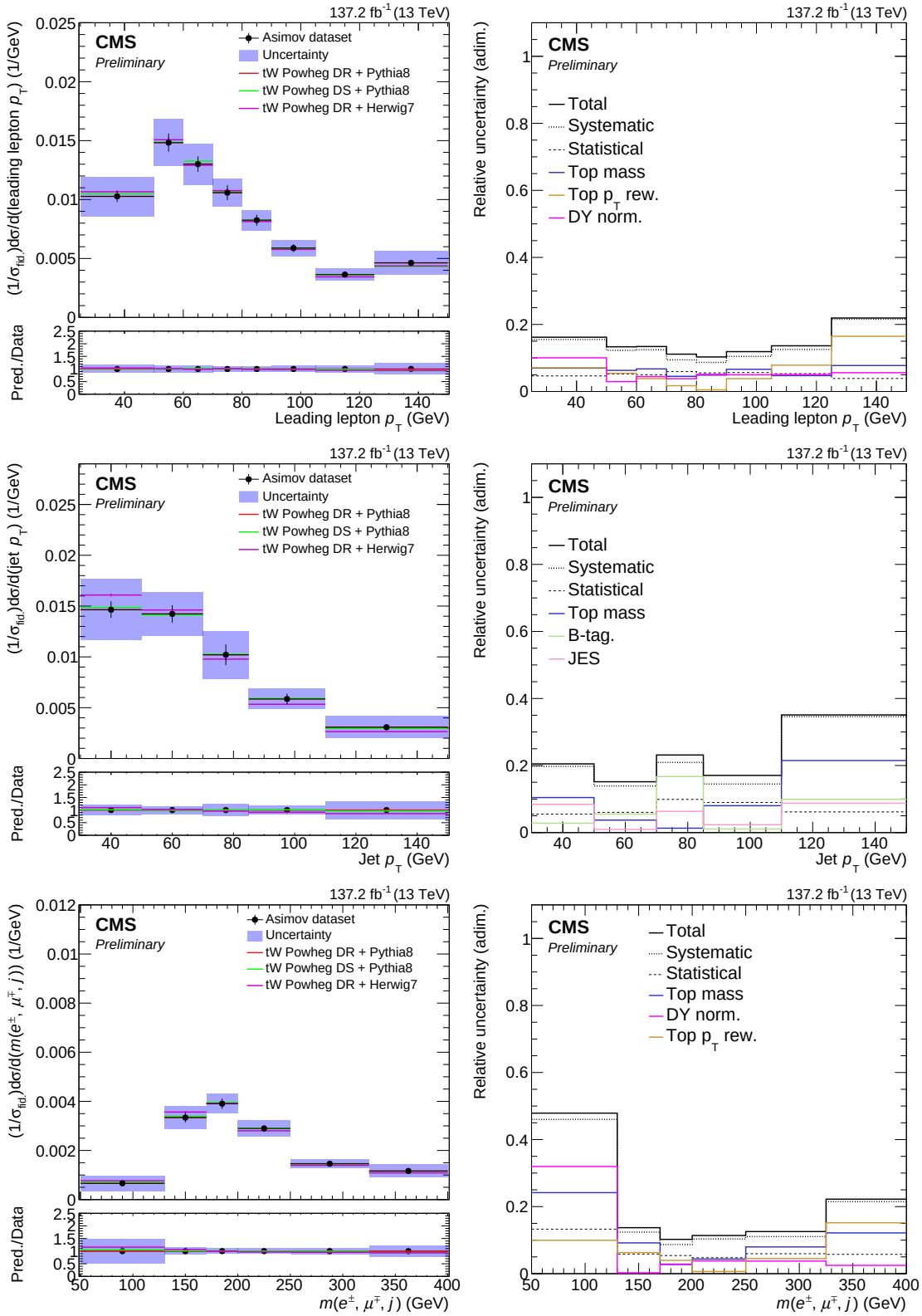

Figure 5.3: Left: normalised expected differential tW production cross section as a function of the p_T of the leading lepton (top), p_T of the jet (middle) and $m(e^{\pm}, \mu^{\mp}, j)$ (bottom) using Run 2 Asimov dataset. The solid band represents the total uncertainty. Predictions from POWHEG + PYTHIA 8 DR/DS, and POWHEG + HERWIG 7 DR are also shown. In the bottom panel, the ratio between predictions and Asimov data is shown. Right: total, systematic, statistical, and the three leading uncertainties (averaging over all bins).

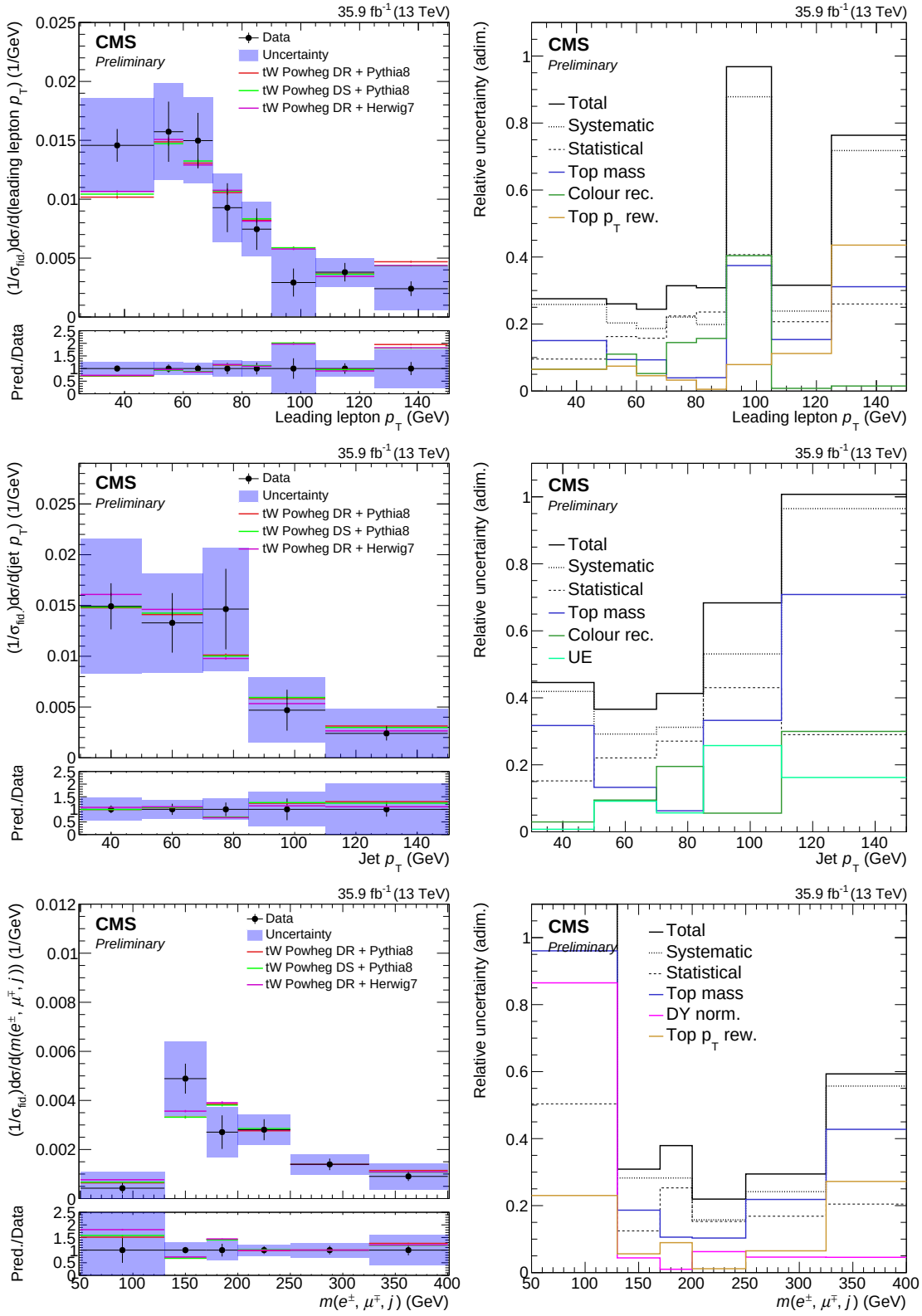

Figure 5.4: Left: normalised differential tW production cross section as a function of the p_T of the leading lepton (top), p_T of the jet (middle) and $m(e^{\pm}, \mu^{\mp}, j)$ (bottom) using the 2016 data. The solid band represents the total uncertainty. Predictions from POWHEG + PYTHIA 8 DR/DS, and POWHEG + HERWIG 7 DR are also shown. In the bottom panel, the ratio between predictions and data is shown. Right: total, systematic, statistical, and the three leading uncertainties (averaging over all bins).

Chapter 6

Conclusions

In this work, the inclusive and differential cross sections measurements for the tW process in the $e^{\pm}\mu^{\mp}$ channel are presented. The analysed dataset corresponds to that collected by the CMS detector during 2016, 2017 and 2018 in proton-proton collisions at 13 TeV of energy in the centre-of-mass.

For the inclusive cross section measurement, a likelihood fit is used to extract the signal contribution from the dominant $t\bar{t}$ background. This likelihood fit employs three different regions defined in terms of the number of jets and b tagged jets: 1j1b, 2j1b and 2j2b. For the 1j1b and 2j1b regions two BDTs are trained to discriminate between tW and $t\bar{t}$ events. The output of the BDTs is used in the fit together with the subleading jet p_T of the 2j2b region. The result for the expected signal strength from the full Run 2 data is

$$r = 1 \pm^{0.09}_{0.10} . \tag{6.1}$$

And the measurement with 2016 data

$$\sigma_{\mathrm{tW,\ 2016}} = 65 \pm 1.8(\mathrm{stat.}) \pm^{9}_{10} (\mathrm{syst.}) \pm^{1.3}_{1.1} (\mathrm{lumi.}) = 65 \pm^{9}_{10} \ \mathrm{pb}. \tag{6.2}$$

This result is compatible with the SM expectation of $\sigma_{\mathrm{tW}} = 71.7 \pm 1.8(\mathrm{scale}) \pm 3.4(\mathrm{PDF})$ pb [17] at NNLO approximation, within the uncertainties. This measurement is limited by the systematic uncertainties, being the most important: the normalisation of the Drell-Yan background, the final state radiation and pileup reweighting.

On the other hand, for the differential cross section measurement only the 1j1b region with the additional veto on the loose jets is considered. The signal extraction is performed by subtracting background from data. The measurements are made as

a function of various properties of the event: the transverse momentum of the leading lepton, the transverse momentum of the jet and the invariant mass of the system formed by the electron, the muon and the jet. The results are first normalised to the fiducial cross section and to the bin width, and then compared with the SM predictions. The expected Run 2 result (Fig. 5.3) shows that the predictions from POWHEG + PYTHIA 8 DR/DS, and POWHEG + HERWIG 7 DR, give compatible results within the uncertainties. The expected leading uncertainties are: top quark mass, top quark p_T, b tagging efficiency and Drell-Yan normalisation. Lastly, the 2016 measurement of the differential cross section, which is represented in Fig. 5.4, shows compatible results between the measured values and the SM expectations.

Bibliography

[1] Albert M Sirunyan et al. Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 10:117, 2018.

[2] Measurement of differential cross sections for single top quark production in association with a W boson at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2020.

[3] César Tomé. Del modelo estándar. *Cultura Científica*, Jun 2014. https://culturacientifica.com/2014/06/17/del-modelo-estandar/.

[4] P.A.M Dirac. The Quantum Theory of the Electron. *Royal Society*, pages 610–624, 1928. https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1928.0023.

[5] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.

[6] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[7] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966.

[8] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[9] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, Sep 2012. http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[10] G. Aad, T. Abajyan, et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics*

Letters B, 716(1):1 – 29, 2012. http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[11] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[12] Mcfm - monte carlo for femtobarn processes. https://mcfm.fnal.gov/.

[13] V. M. Abazov, B. Abbott, M. Abolins, B. S. Acharya, M. Adams, T. Adams, E. Aguilo, M. Ahsan, G. D. Alexeev, G. Alkhazov, and et al. Observation of single top-quark production. *Physical Review Letters*, 103(9), Aug 2009.

[14] T. Aaltonen, J. Adelman, T. Akimoto, B. Álvarez González, S. Amerio, D. Amidei, A. Anastassov, A. Annovi, J. Antos, G. Apollinari, and et al. Observation of electroweak single top-quark production. *Physical Review Letters*, 103(9), Aug 2009.

[15] M. Beneke, P. Falgari, S. Klein, and C. Schwinn. Hadronic top-quark pair production with nnll threshold resummation. *Nuclear Physics B*, 855(3):695–741, Feb 2012.

[16] Tim M. P. Tait and C.-P. Yuan. Single top quark production as a window to physics beyond the standard model. *Physical Review D*, 63(1), Dec 2000.

[17] Nikolaos Kidonakis. Theoretical results for electroweak-boson and single-top production, 2015.

[18] S. Chatrchyan et al. Observation of the associated production of a single top quark and a w boson in pp collisions at $\sqrt{s}$=8 tev. *Physical Review Letters*, 112(23), Jun 2014.

[19] G. Aad et al. Measurement of the production cross-section of a single top quark in association with a w boson at 8 tev with the atlas experiment. *Journal of High Energy Physics*, 2016(1), Jan 2016.

[20] Esma Mobs. The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018. Aug 2018. General Photo.

[21] Public cms luminosity information. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[22] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[23] Accelerating: Radiofrequency cavities. *CERN*. https://home.cern/science/engineering/accelerating-radiofrequency-cavities.

[24] CERN. Pulling together: Superconducting electromagnets. https://home.cern/science/engineering/pulling-together-superconducting-electromagnets.

[25] Stepan Kunc. *Study of the Magnetically Induced QED Birefringence of the Vacuum in experiment OSQAR*. PhD thesis, abr 2018. https://www.researchgate.net/figure/Cross-section-of-LHC-dipole-superconducting-magnet-62_fig14_324588058.

[26] CMS Document 11514-v5. 3d sketchup images of the CMS detector. https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=11514.

[27] G. L. Bayatian et al. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. 2006.

[28] Mario González. http://www.hep.uniovi.es/wiki/.

[29] V Karimäki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudenreich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[30] CMS Collaboration. Tracking. http://cms.cern/detector/identifying-tracks.

[31] CMS Collaboration and Holzner. A. 78 reconstructed vertices in event from high-pileup run 198609. https://cds.cern.ch/record/1479324, CMS Collection, Sep 2012.

[32] Martin Lipinski. The Phase-1 Upgrade of the CMS Pixel Detector. Technical Report CMS-CR-2017-135. 06, CERN, Geneva, May 2017. https://cds.cern.ch/record/2265423.

[33] CMS tracker detector performance results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK.

[34] López, Silvia Goy. Cms detector performance. *EPJ Web Conf.*, 182:02076, 2018.

[35] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[36] Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ tev. https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResultsPASEGM11001.

[37] A.M. Sirunyan et al. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 15(05):P05002–P05002, may 2020.

[38] Sergey Chatrchyan et al. Precise mapping of the magnetic field in the cms barrel yoke using cosmic rays. *Journal of Instrumentation*, 5:T03021, 03 2010.

[39] CMS Collaboration. Detecting muons. http://cms.cern/detector/detecting-muons.

[40] Priyanka Kumari et al. Improved-RPC for the CMS muon system upgrade for the HL-LHC. *JINST*, 15(11):C11012, 2020.

[41] Performance of the CMS Muon Detectors in early 2016 collision runs. https://twiki.cern.ch/twiki/bin/view/CMSPublic/MuonDPGPublic160729.

[42] G. Abbiendi et al. Study of the effects of radiation on the CMS Drift Tubes Muon Detector for the HL-LHC. Technical report, CERN, Geneva, Dec 2019.

[43] CMS Document 5697-v2. CMS slice image with transverse/longitudinal/3-D views. https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=5697.

[44] Particle Flow. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideParticleFlow.

[45] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960. https://www.cs.unc.edu/~welch/kalman/media/pdf/Kalman1960.pdf.

[46] Serguei Chatrchyan et al. The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8:P11002, 2013.

[47] A.M Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06), 2018. https://inspirehep.net/literature/1667449.

[48] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9–N20, Jul 2005.

[49] Albert M Sirunyan et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. 12 2020.

[50] Albert M Sirunyan et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16(05):P05014, 2021.

[51] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.

[52] V. Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, feb 2017.

[53] A.M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al. Identification of heavy-flavour jets with the cms detector in pp collisions at 13 tev. *Journal of Instrumentation*, 13(05):P05011–P05011, May 2018.

[54] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia. Jet flavour classification using deepjet. *Journal of Instrumentation*, 15(12):P12012–P12012, Dec 2020.

[55] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), Jun 2011.

[56] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *Journal of High Energy Physics*, 2007(11):070–070, Nov 2007.

[57] Stefano Frixione and Bryan R Webber. Matching nlo qcd computations and parton shower simulations. *Journal of High Energy Physics*, 2002(06):029–029, Jun 2002.

[58] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026–026, May 2006.

[59] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852–867, Jun 2008.

[60] S. Agostinelli, J. Allison, K. Amako, et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[61] J. Allison, K. Amako, J. Apostolakis, et al. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1):270–278, 2006.

[62] J. Allison, K. Amako, J. Apostolakis, et al. Recent developments in geant4. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835:186–225, 2016.

[63] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, 71:1547, 2011.

[64] Simone Alioli et al. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[65] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[66] Torbjorn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015.

[67] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, and et al. Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1), Jan 2020.

[68] Jet algorithms performance in 13 TeV data. CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, 2017.

[69] Albert M Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13:P05011, 2018.

[70] Morad Aaboud et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC. *Phys. Rev. Lett.*, 117:182002, 2016.

[71] CMS luminosity measurements for the 2016 data taking period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017.

[72] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, 74:3024, 2014.

[73] 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. CMS Physics Analysis Summary CMS-PAS-TOP-16-021, CERN, 2016.

[74] Jesper R. Christiansen and Peter Z. Skands. String formation beyond leading colour. *JHEP*, 08:003, 2015.

[75] Spyros Argyropoulos and Torbjörn Sjöstrand. Effects of color reconnection on $t\bar{t}$ final states at the LHC. *JHEP*, 11:043, 2014.

[76] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Patrick Groth-Merrild, Alberto Guffanti, Nathan P. Hartland, Zahari Kassabov, José I. Latorre, and et al. Parton distributions from high-precision collider data. *The European Physical Journal C*, 77(10), Oct 2017.

[77] Measurement of the differential cross section for $t\bar{t}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-TOP-16-011, CERN, Geneva, 2016.

[78] Albert M Sirunyan et al. Measurement of the $t\bar{t}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79:368, 2019.

[79] Helge Voss, Andreas Höcker, Jörg Stelzer, and Frederick Tegenfeldt. TMVA, the toolkit for multivariate data analysis with ROOT. In *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, page 040. SISSA, 2007. PoS(ACAT2007)040.

[80] Leo Breiman, Jerome Friedman, Charles J. Stone, and R. A. Olshen. *Classification and Regression Trees*. Chapman and Hall, 1984.

[81] CMS Collaboration. Object definitions for top quark analyses at the particle level. Technical Report CMS-NOTE-2017-004, CERN, Geneva, 6 2017.

[82] Stefan Schmitt. Tunfold: an algorithm for correcting migration effects in high energy physics. 2012.

[83] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[84] CMS Collaboration. Combined measurements of the Higgs boson's couplings at $\sqrt{s} = 13$ TeV. http://cds.cern.ch/record/2308127/files/Figure_010-b.pdf.