Università degli Studi di Napoli *Federico II*

DOTTORATO DI RICERCA IN FISICA

Ciclo XXXVI

Coordinatore: prof. Vincenzo Canale

Search for Vector-Like Quark with Machine Learning techniques at the CMS experiment

Settore Scientifico Disciplinare FIS/04

Dottorando
Francesco Carnevali

Tutor
Prof. Alberto Orso Maria Iorio

Anni 2020/2024

Contents

Introduction

The Standard Model (SM) of particle physics is the most successful quantum field theory to date that describes three of the four fundamental forces of Nature. Precise predictions of the SM have been confirmed by several experiments over the course of the past century. However, the SM is widely regarded as an incomplete theory. It does for example not include the gravitation, nor does it provide an explanation for the existence of evidences of Dark Matter. It also exhibits unsatisfactory formal features, such as a large set of free parameters, that might hint to it being a low energy approximation of a more fundamental theory. Many new models that propose to solve some of these issues have supposed the existence of new particles, the Vector-Like Quarks (VLQs) with masses at the TeV scale [1]. The VLQs, differently from the SM quarks, are characterized by a chirality symmetry under the electroweak interactions. The Compact Muon Solenoid (CMS) experiment is one of the four experiments at the Large Hadron Collider (LHC). The latter, thanks to its high centre-of-mass energy in the multi-TeV range and its instantaneous luminosity of order of $10^{34}\ cm^{-2}\ s^{-1}$, would allow to find evidence of the existence of VLQs.
In this thesis, the search for a single produced VLQ T, decaying to a top quark and a Higgs boson or hypothetical new boson, labelled as A and with yet unknown mass, is presented. The considered final state is characterized by the hadronic decay of the boson and the leptonic decay of the top quark. This is the first time the semileptonic final state is investigated analyzing the Run 2 data, collected at CMS experiment. Moreover, this is the first search that takes into account the decay of the VLQ T into the hypothetical boson A. To improve the selection efficiency of the final state, a machine learning algorithm is used for the top quark candidate reconstruction. The reconstruction of top quark candidates with ML techniques has been performed in such searches with hadronic top quark final states [2], but this is the first time it is studied in depth for leptonic top quark final states in boosted topologies.
The data sample collected by CMS during Run 2 is considered, consisting of proton-proton collisions at $\sqrt{s} = 13\ TeV$ and corresponding to an integrated luminosity of $138 fb^{-1}$. The search is performed taking into account mass hypotheses of T between 0.6 and 3 TeV, as well as different mass hypotheses of A, between 25 and 250 GeV.
This thesis is organized as follows:

- Chapter 1 provides an introduction of the Standard Model and the unsolved problems.
- Chapter 2 describes the LHC accelerator and of the CMS experiment. The muon background studies in the CMS Muon system are presented, in particular focusing on the Resistive Plate Chambers background.
- Chapter 3 gives a brief description of the different theories that predicts the existence of the Vector-Like Quarks and of the boson A.
- Chapter 4 shows the selection of the objects in the final state, focusing on the top quark reconstruction and the description of the machine learning algorithm developed to improve the top quark selection.
- Chapter 5 reports on the description of the selection applied to data and simulation, the signal extraction procedure and its results.

Chapter 1

The Standard Model of particle physics

The Standard Model (SM) of particle physics is the quantum field theory that describes the elementary particles and three of the four fundamental forces, i.e. the electromagnetic, the weak, and the strong interactions. This theory provides precise predictions which have been confirmed by the experimental evidence collected throughout the past century in particle physics. Among its most recent additions we can mention the top quark discovery at TeVatron by the CDF and D0 collaborations [3, 4] in 1995 and the Higgs boson discovery in 2012 at LHC by the CMS and Atlas Collaborations [5, 6]. The SM, nevertheless, does not include the gravitation, it does not predict neutrino masses, that are confirmed by neutrino oscillation experiment [7], and the existence of dark matter and the dark energy, that is predicted by the ΛCDM cosmological model [8].

1.1 Standard Model overview

The SM describes the electromagnetic and the weak interactions with the unified Glashow, Weinberg, and Salam (GWS) theory [9, 10, 11], and the strong interaction through the Quantum Chromodynamics (QCD) [12]. The dynamics equations for the SM are obtained from a gauge principle, thus the free particle lagrangian $\mathcal{L}$ is requested to be invariant under a local transformation of the symmetry group:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

The three gauge groups are:

- the Unitary group, $U(1)_Y$, for which the transformation is represented by a unitary scalar complex operator multiplied by its quantum number called the weak hypercharge Y and the associated vector field is called B_μ ;
- the Special Unitary group with n = 2, $SU(2)_L$, whose three transformations are represented by the σ_i $(i = 1, 2, 3)$, the 2×2 Pauli matrices, multiplied by the third component of the weak isospin I_3 and the three vector field are $W_\mu^{1,2,3}$;
- the Special Unitary group with n = 3, $SU(3)_C$, whose eight transformations are represented by the λ_i $(i = 1, \ldots, 8)$, the 3×3 Gell-Mann matrices, multiplied by the colour charge $C = (r, g, b)$ and the eight vector field are $G_\mu^{1,\ldots,8}$;

A multiplet and an absolute conserved quantum number are associated to each interaction. The multiplet components are called bosons, since they obey the Bose-Einstein statistics and they have spin equals to one, and the number of generators of the symmetry group determines the number of mediator bosons of the interaction, for a total of 12 vector fields associated to the three gauge symmetries. A list of the physical bosons is reported in Table 1.1.

The SM predicts 12 fields with an half-integer spin, called fermions since they obey to the Fermi-Dirac statistics. Fermions are divided into lepton and quark fields, both of which are listed in three families or generations. Each generation consists of a doublet of particles associated to an isospin

Interaction	Mediators	Spin	Electric charge
Electromagnetic	γ	1	0
Weak	W^+, W^-, Z	1	$1, -1, 0$
Strong	8 gluons (g)	1	0

Table 1.1: Standard Model mediator bosons and fundamental interactions.

quantum number, and each particle carries electroweak charge. Only the quarks have strong charge. Each lepton and quark has an associated anti-particle with the same mass but opposite quantum numbers. In Table 1.2 a list of SM fermions is reported.

Particles			Spin	Charge
$\begin{pmatrix}\nu_e \\ e\end{pmatrix}$	$\begin{pmatrix}\nu_\mu \\ \mu\end{pmatrix}$	$\begin{pmatrix}\nu_\tau \\ \tau\end{pmatrix}$	1/2	$\begin{matrix}0 \\ -1\end{matrix}$
$\begin{pmatrix}\mathrm{u} \\ \mathrm{d}\end{pmatrix}$	$\begin{pmatrix}\mathrm{c} \\ \mathrm{s}\end{pmatrix}$	$\begin{pmatrix}\mathrm{t} \\ \mathrm{b}\end{pmatrix}$	1/2	$\begin{matrix}2/3 \\ -1/3\end{matrix}$

Table 1.2: Standard Model leptons and quarks.

In 1964 Peter Higgs predicted the existence of a complex scalar doublet field ϕ [13], named the Higgs field. Thanks to the Higgs mechanism the SM provides masses to particles. In Figure 1.1 the particles predicted by the SM are shown.

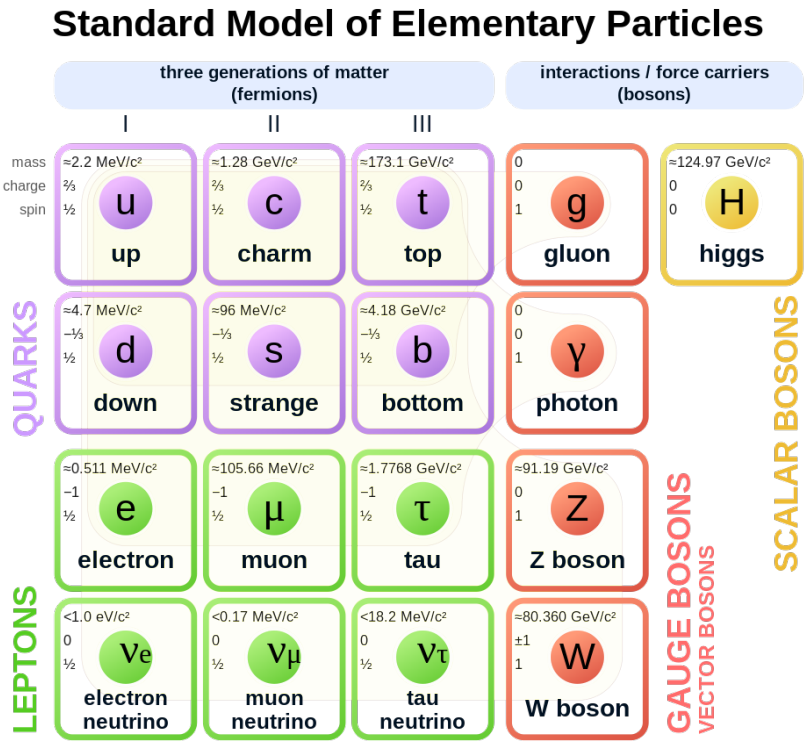

Figure 1.1: Standard Model particles.

1.2 Quantum electrodynamics

The dynamics between fermions in an electromagnetic field is described by the relativistic quantum field theory known as the Quantum Electrodynamics (QED). The abelian group $U(1)_q$ is the symmetry group associated to the electromagnetic interaction, and the charge of the particles q is the conserved quantum number.
The lagrangian density for the QED can be derived from the Dirac lagrangian density of a field ψ with mass m:

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

where γ^μ are the Dirac matrices and ψ and $\bar{\psi}$ are the 4-components spinor and its adjoint. The kinetic term is $\bar{\psi}\gamma^\mu\partial_\mu\psi$, while $m\bar{\psi}\psi$ is the mass term. The free photons propagation is included adding the term of the boson kinetic energy $\mathcal{L}_\gamma$ to the Equation 1.1, given by:

$$\mathcal{L}_\gamma = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{1.2}$$

where $F^{\mu\nu}$ is the field strength tensor. $\mathcal{L}_\gamma$ is invariant for local gauge transformation and $F^{\mu\nu}$ can be written in term of A^μ, which is the 4-vector electromagnetic field:

$$F^{\mu\nu} = -F^{\nu\mu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{1.3}$$

The lagrangian density $\mathcal{L}_{QED}$,

$$\mathcal{L}_{QED} = \mathcal{L}_D + \mathcal{L}_\gamma = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{1.4}$$

is globally invariant under a $U(1)_q$ transformation,

$$\psi \to \psi' = e^{i\theta}\psi, \tag{1.5}$$

where θ is an arbitrary constant. However, the fermionic kinetic term is not invariant under local gauge transformation:

$$\begin{gathered}
\psi \to \psi' = e^{i\theta(x)}\psi \\
\bar{\psi} \to \bar{\psi}' = e^{-i\theta(x)}\bar{\psi} \\
\partial_\mu\psi \to \partial_\mu\psi' = e^{iq\theta(x)}\partial_\mu\psi(x) + iqe^{iq\theta(x)}\psi(x)\partial_\mu\theta(x)
\end{gathered}$$

where $\theta(x)$ is a function of the space-time coordinates.
The covariant derivative D_μ is introduced to have an invariant lagrangian under local gauge transformation. For this purpose, this term undergoes the local gauge transformation as follows:

$$\begin{gathered}
\mathcal{D}_\mu\psi \to e^{iq\theta(x)}\mathcal{D}_\mu\psi \\
\mathcal{D}_\mu \equiv \partial_\mu + iqA_\mu(x),
\end{gathered} \tag{1.6}$$

where the vectorial field A_μ transforms like:

$$A_\mu(x) \to A_\mu(x) - \partial_\mu\theta(x). \tag{1.7}$$

The lagrangian density can be written as:

$$\mathcal{L}_D = i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - m\bar{\psi}\psi, \tag{1.8}$$

or:

$$\mathcal{L}_D = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu\psi A_\mu = \mathcal{L}_0 - J^\mu A_\mu \tag{1.9}$$

The term J^μ is the charged current, which is the probability current multiplied by the particle charge. The local invariant QED lagrangian for a Dirac particles in an electric field is given by adding the terms 1.9 and 1.2:

$$\mathcal{L}_{QED} = \mathcal{L}_D + \mathcal{L}_\gamma = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{1.10}$$

The Feynman rules [14] allow a diagrammatic representation of the processes amplitudes, these can be derived directly from the $\mathcal{L}_{QED}$ Lagrangian. A fundamental transition associated to the Lagrangian density in (1.10) is shown in Figure 1.2.

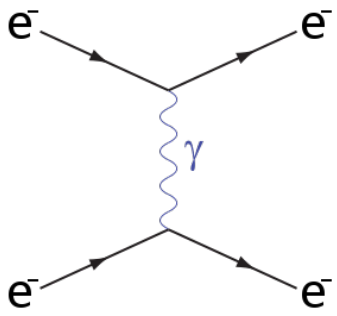

Figure 1.2: Basic vertex in Quantum Electrodynamics.

1.3 The Electroweak theory

The first phenomenological description of the weak interaction is given by the Enrico Fermi's article "Tentativo di una teoria dell'emissione dei raggi beta" [15]. The emission of β rays is described in this work and the weak interaction Lagrangian density is given by:

$$\mathcal{L}_F = \frac{G}{\sqrt{2}} J^{\dagger\mu}(x) J_\mu(x) \tag{1.11}$$

where G is the Fermi constant

$$G = 1.16638 \times 10^{-5}\ \text{GeV}^{-2}$$

and J^μ is the weak current. The interaction is determined by a current-current Lagrangian density, and the weak current is composed by the sum of a weak leptonic current, $l^\mu(x)$ and a weak hadronic current $h^\mu(x)$. However, depending on the interaction, the weak current can be classified as described in Tabble 1.3.

Type of process	Example	Weak Current
Leptonic	$\mu^- \to e^- + \bar{\nu_e} + \nu_\mu$	$\mathcal{L}_F = \frac{G}{\sqrt{2}} l^{\dagger\mu}(x) l_\mu(x)$
Semileptonic	$\mu^- \to e^- + \bar{\nu_e} + \nu_\mu$	$\mathcal{L}_F = \frac{G}{\sqrt{2}} [l^{\dagger\mu}(x) h_\mu(x) + h^{\dagger\mu}(x) l_\mu(x)]$
Hadronic	$\Lambda \to p + \pi^-$	$\mathcal{L}_F = \frac{G}{\sqrt{2}} h^{\dagger\mu}(x) h_\mu(x)$

Table 1.3: Weak currents for different processes.

In this model the involved particles are supposed to be in the same point of the space-time coordinates when the interaction occurs, thus all the terms described above represent contact interactions. The leptonic interactions involve only elementary particles, thus the vector and spin structure of the current is of the V-A type:

$$l^\mu(x) = \bar{e}(x)\gamma^\mu(1-\gamma^5)\nu_e(x) + \bar{\mu}(x)\gamma^\mu(1-\gamma^5)\nu_\mu(x). \tag{1.12}$$

Each Dirac field could be written as:

$$\psi = \psi_L + \psi_R \tag{1.13}$$

where ψ_L and ψ_R are, respectively, the left-chiral and the right-chiral Dirac field. The left-chiral projector, P_L, is proportional to $(1-\gamma^5)$, therefore only left-chiral particles are involved in the weak interaction.
However, this description leads to divergences, in particular the violation of the unitarity limit of the Fermi-like cross-section.

1.3.1 The GWS model

To solve the divergences problem, Glashow, Weingerg, and Salam[9, 10, 11] formulated the Electroweak theory in the 60's and 70's. This theory, also known as the GWS model, successfully unifies the weak and electromagnetic interactions, and has earned its authors the Nobel prize in Physics in 1979. The GWS model solves the issues of Fermi's theory by introducing an adimesional coupling constant and three intermediate vector bosons. The symmetry group for the GWS model is $G = SU(2)_L \otimes U(1)_Y$, where L means that only left-chiral particles are involved in the interaction and Y is the weak hypercharge associated to the particle. The invariance under the symmetry group $G = SU(2)_L \otimes U(1)_Y$ leads to the introduction of 4 gauge fields. Three of these are Yang-Mills fields, $W^{\mu}_{(i)}$, that are an isospin triplet, while the last one is an ispospin singlet B^{μ}, from $U(1)_Y$. Fermions are represented by doublets, eigenstates of the weak isospin I, since the three Pauli 2×2 matrices σ_i are the generators of the group $SU(2)_L$. For the leptons the doublets are:

$$I = 1/2 \qquad \begin{matrix} I_3 = +1/2 \\ I_3 = -1/2 \end{matrix} \qquad \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L ,$$

where I_3 is the observed third component of I.
A local gauge transformation $SU(2)_L$ could be written as:

$$\begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}'_L = e^{-i\vec{\alpha}(x)\cdot\vec{\tau}} \begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}_L , \tag{1.14}$$

where $\alpha_i(x)$ are three arbitrary functions, that depend on the space-time coordinates, and τ_i are the Pauli matrices divided by 2. Right-chiral components of charged leptons

$$I = 0 \qquad e^-_R, \mu^-_R, \tau^-_R .$$

are singlets of $SU(2)_L$ and do not take part to weak charged-current processes, but they are are involved in neutral-current processes. In the SM there are no right-chiral neutrinos. The weak hypercharge Y is the conserved quantum number associated to the symmetry group $U(1)_Y$, defined as:

$$Q = I_3 + \frac{Y}{2}. \tag{1.15}$$

Each $SU(2)_L$ multiplet member has the same hypercharge, which is given by:

$$Y_{multiplets} = 2\bar{Q}$$

where $\bar{Q}$ is the mean electric charge of the multiplet.

For quark families, similar relations as for the leptons hold true. The singlets and the doublets for the quarks are respectively:

$$I = 0 \qquad d_R, u_R, s_R, c_R, b_R, t_R$$

and

$$I = 1/2 \qquad \begin{matrix} I_3 = +1/2 \\ I_3 = -1/2 \end{matrix} \qquad \begin{pmatrix} \mathrm{u} \\ \mathrm{d}' \end{pmatrix}_L \begin{pmatrix} \mathrm{c} \\ \mathrm{s}' \end{pmatrix}_L \begin{pmatrix} \mathrm{t} \\ \mathrm{b}' \end{pmatrix}_L ,$$

where d', s' and b' are weak interaction eigenstates. One crucial aspect to note is that weak interaction eigenstates do not correspond to mass eigenstates, of which they form a linear combination. The flavour mixing is given by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [16, 17], that is an unitary complex matrix:

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = V_{CKM} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix},$$

where V_{CKM} can be written as:

$$V_{CKM} = \begin{pmatrix} \mathrm{V_{ud}} & \mathrm{V_{us}} & \mathrm{V_{ub}} \\ \mathrm{V_{cd}} & \mathrm{V_{cs}} & \mathrm{V_{cb}} \\ \mathrm{V_{td}} & \mathrm{V_{ts}} & \mathrm{V_{tb}} \end{pmatrix}.$$

The CKM matrix can be parametrized by 4 real parameters. A particularly convenient choice is one of the the so-called the Wolfenstein parametrization (λ, η, ρ, A), which provides:

$$V_{CKM} = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - \mathrm{i}\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1-\rho-\mathrm{i}\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathrm{O}(\lambda^4).$$

The fit for this parametrization [18] gives

$$\lambda = 0.22453 \pm 0.00044,$$

this means that the matrix is almost diagonal, and quarks are inclined to be associated with quarks to the same family in weak interactions. The complex terms lead the Electroweak Lagrangian not to be invariant under CP (Charge-coniguation and Parity) transformation. The CP symmetry violation is factor needed to understand the matter-antimatter asymmetry after the Big Bang, together with the Baryon number violating processes and the presence of non-equilibrium state during the universe expansion.
According to the gauge principle, as it has been done for the QED Lagrangian, the Electroweak Lagrangian can be obtained starting from the free fermionic Lagrangian by neglecting mass terms and by introducing a new covariant derivative:

$$D_\mu = \partial_\mu + ig\frac{\vec{\tau}}{2}W_\mu + ig'\frac{Y}{2}B_\mu, \tag{1.16}$$

where g and g' are the equivalent of the electromagnetic coupling constant e.
The Electroweak Lagrangian for fermions is

$$\mathcal{L}_{fermions} = \sum_f \bar{\psi}\gamma^\mu D_\mu \psi. \tag{1.17}$$

To complete the Lagrangian it needs to be added the dynamics terms of the gauge boson fields:

$$\mathcal{L}_{gauge} = -\frac{1}{4}W_i^{\mu\nu}W^i_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \tag{1.18}$$

with $W_i^{\mu\nu}$ and $B^{\mu\nu}$ the tensor fields:

$$\begin{aligned} W_i^{\mu\nu} &= \partial^\mu W_i^\nu - \partial^\nu W_i^\mu \\ B_i^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu. \end{aligned} \tag{1.19}$$

Therefore, the Electroweak Lagrangian, neglecting the mass term for the fermions and gauge bosons, can be written as:

$$\begin{aligned} \mathcal{L}_{EW} =& -i\psi_L\gamma^\mu\left(\partial_\mu + ig\frac{\vec{\tau}}{2}\cdot W_\mu + ig'YB_\mu\right)\psi_L + \\ & - i\psi_R\gamma^\mu\left(\partial_\mu + ig'YB_\mu\right)\psi_R + \\ & - \frac{1}{4}W_i^{\mu\nu}W^i_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \\ & + \frac{1}{2}g\,\epsilon_{ijk}W_i^{\mu\nu}W_{j\mu}W_{k\nu} + \frac{1}{4}g^2\,\epsilon_{ijk}\epsilon_{imn}W_{j\mu}W_{k\nu}W_m^\mu W_n^\nu, \end{aligned} \tag{1.20}$$

where the last two terms describe the cubic and quartic self interactions of the vector bosons, due to the non-Abelian nature of $SU(2)_L$.
The vector fields for the physical $W^\pm, Z$ bosons and for the photon can be obtained as a linear combination of the four gauge fields:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \\ Z_\mu &= -sin\theta_W B_\mu + {W_\mu}^3 cos\theta_W \\ A_\mu &= cos\theta_W B_\mu + {W_\mu}^3 sin\theta_W, \end{aligned} \tag{1.21}$$

In practice two of the gauge fields are combined together in order to give two vector bosons $W^{\pm}$, that are electrically charged and can induce transitions between the members of the weak isospin doublets. The third gauge boson of the weak triplet is electrically neutral, and is combined with the electromagnetic boson to give the Z boson and the photon. The parameter θ_W is called Weinberg angle, defined as:

$$cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}, \tag{1.22}$$

and its value is not predicted by the SM. The electromagnetic charge therefore is:

$$q = g' \cos\theta_W = g \sin\theta_W \tag{1.23}$$

In Figures 1.3 representative tree-level diagrams of charged- and neutral-current processes are shown.

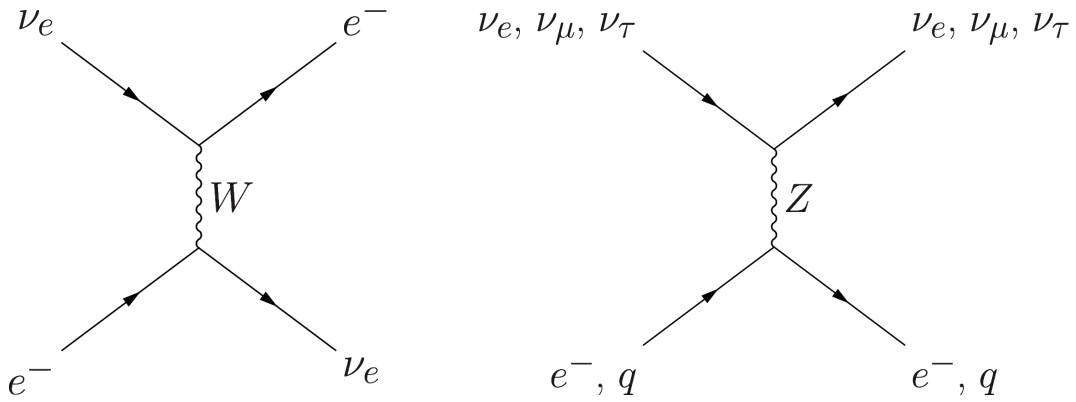

Figure 1.3: Charged current (left) and neutral current (right) processes in GWS model.

1.4 Quantum Chromodynamics

The strong interaction between quarks is described by the Quantum Chromodynamics (QCD). The $SU(3)_C$ symmetry group is associated to this quantum gauge field theory, where C is the color, i.e. the conserved charge in the interaction. Since the group is non-Abelian, the generators of the group do not commute between themselves and the QCD Lagrangian is made up of self-interaction terms between the gauge fields, called gluons, which therefore do carry color charge. This feature marks a stark contrast w.r.t. to their equivalent for weak interactions, the photons, that have no electric charge. The generators of the group are $\vec{T} = T^{\alpha}$, that are the Gell-Mann matrices divided by 2 and follow the commutation rules:

$$[T_\alpha, T_\beta] = i f_{\alpha\beta\gamma} T_\gamma, \tag{1.24}$$

where $f_{\alpha\beta\gamma}$ are the structure constants of the group $SU(3)_C$.
Under local transformation of $SU(3)_C$ the quarks fields change as:

$$\psi \to \psi' = e^{i g_s \vec{\theta}(x) \cdot \vec{T}} \psi, \tag{1.25}$$

where g_s is the strong coupling constant, $\vec{\theta}(x)$ are eight functions of the space-time coordinates, since $\vec{T}$ are represented by 3×3 matrices the fields ψ have three more additional freedom. The colour is the new degree of freedom and the three different possible state are labelled as red, green, and blue. The imposition of local invariance leads to introduce the covariant derivative:

$$D_\mu = \partial_\mu + i g_s T_\alpha G^\alpha_\mu, \tag{1.26}$$

where the G^α_μ terms are the 8 massless gauge fields, or gluon fields, that transform as:

$$G^\alpha_\mu \to G'^\alpha_\mu = G^\alpha_\mu + i g_s f^{\alpha\beta\gamma} \theta_\beta(x) G_{\gamma,\mu}. \tag{1.27}$$

To complete the Lagrangian density for the QCD it is necessary to add the contribution of the kinetic energy for each gluon:

$$\mathcal{L}_{QCD} = \bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi - ig_s\bar{\psi}\gamma^{\mu}\lambda_{\alpha}\psi G^{\alpha}_{\mu} - \frac{1}{4}G^{\mu\nu}_{\alpha}G^{\alpha}_{\mu\nu}, \tag{1.28}$$

where $G^{\mu\nu}_{\alpha}$ is the tensor field representing the gluon kinetic term, and is defined as:

$$G^{\mu\nu}_{\alpha} = \partial^{\mu}G^{\nu}_{\alpha} - \partial^{\nu}G^{\mu}_{\alpha} - g_s f_{\alpha\beta\gamma}G^{\beta,\mu}G^{\gamma,\nu}. \tag{1.29}$$

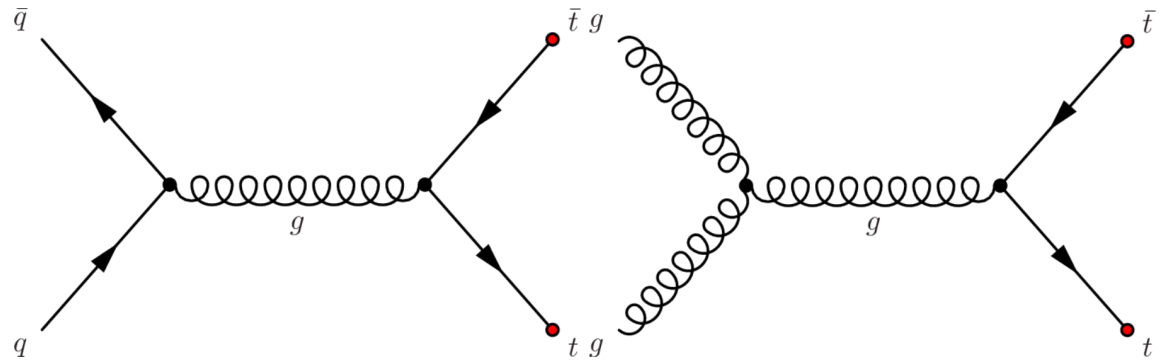

Figure 1.4: Leading order representative Feynman diagram for the production of a quark and antiquark pair via strong interactions.

In Figures 1.4 representative tree-level diagrams of QCD processes are shown, top quark pair production via quark-antiquark annihilation (left) and gluon-gluon fusion (right).
The QCD theory successfully explains two relevant properties, shown by experimental evidence long before its formulation: asymptotic freedom and colour confinement. The first feature refers to the observation that, the higher is the energy of the strongly interacting particles, the weaker becomes the interaction. In the SM, this can be explained by the the running of the strong coupling constant, which can be evaluated in perturbative calculations at high enough energy:

$$\alpha_s\left(|q^2|\right) = \frac{\alpha_s(\mu^2)}{\left[1 + \alpha_s(\mu^2)\frac{33-2N_f}{12\pi}\ln\frac{q^2}{\mu^2}\right]}, \tag{1.30}$$

where q^2 is the transferred 4-momentum, μ is the reference scale parameter for the strength of the coupling, and N_f is number of fermions capable of strong interactions at the scale considered. Equqation 1.30 shows that $\alpha_s(q^2)$ decreases as q^2 increases. For $|q| \sim 200$ MeV the value of α_s is large enough that any perturbative approach cannot be applied. In this region the calculations need to be carried on with other means, one of which is the QCD lattice approach [19].
The second property has been formulated since no coloured hadrons are observed in nature. Hadrons are interpreted as bound states of quarks in the QCD parton model and they are colour singlets.

1.5 Spontaneous symmetry breaking and Higgs mechanism

The Electroweak Lagrangian does not allow to introduce mass terms for the fermion, that would appear as:

$$-m\bar{\psi}\psi,$$

and for the bosons:

$$(1/2)m^2 B_\mu B^\mu,$$

as any of such mass terms would in fact break the gauge symmetry $SU(2)_L \otimes U(1)_Y$. To give mass to the intermediate bosons, the spontaneous symmetry breaking (SSB) and the Higgs mechanism provide the simplest and the most elegant explanation to the observed particles' masses. This mechanism introduces a new boson field ϕ, the Higgs field, which is an isospin doublet of complex scalar fields:

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.31}$$

The Lagrangian for the ϕ field is:

$$\begin{aligned} \mathcal{L}_H &= (D^\mu \phi)^\dagger D_\mu \phi - V(\phi) = \\ &= (D^\mu \phi)^\dagger D_\mu \phi - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{1}{4}\lambda(\phi^\dagger \phi)^2, \end{aligned} \tag{1.32}$$

where $V(\phi)$ is the Higgs potential and λ is assumed to be positive. It is possible to find the ground vacuum state ϕ_0 by minimizing the potential V. Adding the potential in Equation 1.32 to the Lagrangian of the gauge field sector results in:

$$\mathcal{L}_H = (D^\mu \phi)^\dagger D_\mu \phi - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{\lambda}{4}(\phi^\dagger \phi)^2 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \tag{1.33}$$

with:

$$D^\mu \phi = \left(\partial^\mu + ig\frac{\vec{\tau}}{2}W^\mu + ig'YB^\mu\right)\phi \tag{1.34}$$

$$F^{\mu\nu} = \partial^\mu W^\nu - \partial^\nu W^\mu - gW^\mu \times W^\nu \tag{1.35}$$

$$B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu. \tag{1.36}$$

As shown in Figure 1.5, for $\mu^2 > 0$ there is just one minimum at $\phi_0 = 0$ and therefore $V(\phi) = 0$; for $\mu^2 < 0$ instead V assumes a non-trivial minimum:

$$\phi_0^2 = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2},$$

which identifies a circumference in the complex plane.
Among all the possible ground states, a reference minimum is chosen, without losing generality:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}$$

Considering the fluctuations of the ϕ field around the vacuum, the Higgs field assumes the value:

$$\phi = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H(x)) \end{pmatrix}. \tag{1.37}$$

By substitution of Equation 1.37 in Equation 1.33, taking into account Equation 1.21 leads to:

$$\begin{aligned} \mathcal{L}_{G\Phi} = &\frac{1}{2}\partial_\mu H \partial^\mu H - \mu^2 H^2 + \\ &- \frac{1}{4}(\partial_\mu W^1_\nu - \partial_\nu W^1_\mu)(\partial^\mu W^{1\nu} - \partial^\nu W^{1\mu}) + \frac{1}{8}g^2 v^2 W^1_\nu W^{1\nu} \\ &- \frac{1}{4}(\partial_\mu W^2_\nu - \partial_\nu W^2_\mu)(\partial^\mu W^{2\nu} - \partial^\nu W^{2\mu}) + \frac{1}{8}g^2 v^2 W^2_\nu W^{2\nu} \\ &- \frac{1}{4}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu Z^\nu - \partial^\nu Z^\mu) + \frac{1}{8}(g^2 + g'^2)v^2 Z_\nu Z^\nu \\ &- \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \end{aligned} \tag{1.38}$$

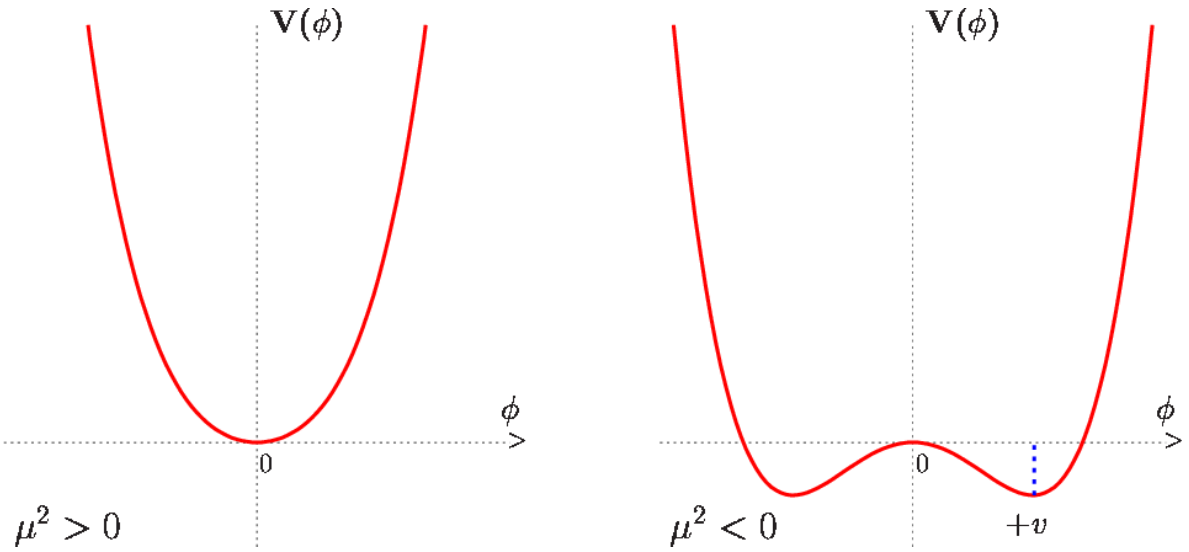

Figure 1.5: Shape of the Higgs potential $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$ for $\lambda > 0$ and $\mu^2 < 0$ and for $\mu^2 > 0$.

The first line of Equation 1.38 is the Lagrangian density of the Higgs scalar field, with mass $\sqrt{2}\mu$. The next two lines show that the components W_1^μ and W_2^μ of the triplet $\mathbf{W}^\mu$ acquire mass:

$$M_1 = M_2 = \frac{1}{2}gv \equiv M_W, \tag{1.39}$$

while the fourth line shows that the field Z^μ acquires a mass:

$$M_Z \equiv \frac{1}{2}v\sqrt{g^2 + g'^2} = \frac{M_W}{\cos\theta_W}. \tag{1.40}$$

Finally, the last line shows that the photon field A^μ has a zero mass.

1.5.1 Leptons masses

As said before, a mass term for fermion as $-m\bar{\psi}\psi$ is not invariant under $SU(2)_L \otimes U(1)_Y$, since the term can be written as $-m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$. It is possible to introduce a Yukawa coupling between the fermion and the Higgs fields resulting in a mass term of the form:

$$\mathcal{L}_Y = g_f(\bar{\psi}_L\phi\psi_R - \bar{\psi}_R\phi^\dagger\psi_L), \tag{1.41}$$

where g_f is the Yukawa coupling constant. By substituting 1.37 in 1.41 one obtains:

$$\begin{aligned}\mathcal{L}_Y =& \frac{g_f}{\sqrt{2}}\left[(\nu_\ell,\ \ell)\begin{pmatrix}0\\ v+H\end{pmatrix}\ell_R + \ell_R(0,\ v+H)\begin{pmatrix}\nu_\ell\\ \ell\end{pmatrix}\right] = \\ =& \frac{g_f}{\sqrt{2}}(v+H)(\ell_L\ell_R + \ell_R\ell_L).\end{aligned} \tag{1.42}$$

This part of the Lagrangian is therefore proportional to $(\ell_L\ell_R + \ell_R\ell_L)$, and is similar to a Dirac mass term that could not be introduced previously without violating the Electroweak symmetry. For each charged lepton the mass results to be:

$$m_l = \frac{v}{\sqrt{2}}g_l. \tag{1.43}$$

However, even having a mass term that does not break the symmetry of the group, the value of each g_ℓ is not predicted by the SM, and all of them can in fact assume different values, as it is shown by experimental evidence.

1.5.2 Quarks masses

In the quarks case the mechanism is quite similar, since it is possible to write:

$$\begin{aligned}\mathcal{L}_Y =& \frac{1}{\sqrt{2}}\left[g^d_{i,j}(u_{i,L},\ d_{i,L})\begin{pmatrix}0\\ v+H\end{pmatrix}d_{j,R} + g^u_{i,j}(u_{i,L},\ d_{i,L})\begin{pmatrix}-(v+H)^*\\ 0\end{pmatrix}u_{j,R} + h.c.\right] = \\ =& \frac{1}{\sqrt{2}}(v+H)[g^u_{ij}(u_{i,L}u_{j,R} + u_{j,R}u_{i,L}) + g^d_{ij}(d_{i,L}d_{j,R} + d_{j,R}d_{i,L}) + h.c.],\end{aligned} \tag{1.44}$$

with u_i = (u, c, t) and d_i = (d, s, b). Each mass term is equal to:

$$m^u_{ij} = -\frac{v}{\sqrt{2}}g^u_{ij} \quad m^d_{ij} = -\frac{v}{\sqrt{2}}g^d_{ij}, \tag{1.45}$$

the masses matrix is however not diagonal, but it can be made so with a series of transformations:

$$u_{\alpha,L} = (\mathcal{U}^u_L)_{\alpha i}u_{i,L} \quad u_{\alpha,R} = (\mathcal{U}^u_R)_{\alpha i}u_{i,R} \tag{1.46}$$

$$d_{\alpha,L} = (\mathcal{U}^d_L)_{\alpha i}d_{i,L} \quad d_{\alpha,R} = (\mathcal{U}^d_R)_{\alpha i}d_{i,R} \tag{1.47}$$

where α is the index in the mass diagonal basis and i is the index in the non-diagonal weak interaction basis.
The resulting part of Lagrangian is:

$$\mathcal{L}_Y = \frac{1}{\sqrt{2}}(v+H)\left[m^u \mathrm{u}\bar{\mathrm{u}} + m^d \mathrm{d}\bar{\mathrm{d}} + m^s \mathrm{s}\bar{\mathrm{s}} + m^c \mathrm{c}\bar{\mathrm{c}} + m^t \mathrm{t}\bar{\mathrm{t}} + m^b \mathrm{b}\bar{\mathrm{b}}\right]. \tag{1.48}$$

These transformations need to be applied also in the Electroweak interaction Lagrangian, therefore the charged current coupling term is:

$$\begin{aligned}\mathcal{L}_{CC} =& -\frac{g}{\sqrt{2}}(\bar{u}_{i,L},\ \bar{d}_{i,L})\gamma^\mu\tau_+ W^+_\mu\begin{pmatrix}u_{Li}\\ d_{Li}\end{pmatrix} + h.c. \\ =& -\frac{g}{\sqrt{2}}\bar{u}_{iL}\gamma^\mu d_{Li}W^+_\mu + h.c. \\ =& -\frac{g}{\sqrt{2}}\bar{u}_{\alpha_L}\left[(\mathcal{U}^u_L)_{\alpha i}(\mathcal{U}^d_L)^\dagger_{\beta i}\right]\gamma^\mu d_{L\beta}W^+_\mu + h.c.,\end{aligned} \tag{1.49}$$

where:

$$V_{\alpha\beta} = \left[\mathcal{U}^u_L\mathcal{U}^{d\dagger}_L\right]_{\alpha\beta} \tag{1.50}$$

is the CKM matrix, that emerges naturally in this approach, and that we already described in 1.3.1. The charged current Lagrangian can be written as:

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}}\bar{u}_{L\alpha}\gamma^\mu V_{\alpha\beta}d_{L\beta}W^+_\mu - \frac{g}{\sqrt{2}}\bar{d}_{L\alpha}\gamma^\mu V^\dagger_{\alpha\beta}d_{L\beta}W^-_\mu \tag{1.51}$$

1.6 Unsolved problems in the Standard Model

As we detailed in the previous sections, the Standard Model can explain three of four fundamental forces, while remaining a self-consistent quantum field theory. However, the SM does not give a complete depiction of Nature, as there are numerous fundamental physics phenomena it does not explain. Furthermore, the SM presents some aspects that suggests it might be a low-energy approximation of a more fundamental theory.
Some of the limitations of the SM are:

- **Large set of free parameters**: The SM doesn't predict the hierarchy of several of its parameter, i.e. the the particles masses, and their values:
  - 3 coupling constants: g, g' and g_s;
  - θ_W and also the GWS model can not be considered as a real unification theory, since the group of symmetry G in the tensorial product of two different groups;

- fermion masses, moreover the differences of magnitudes between fermions masses reach 5 order of magnitudes;
- mass of the Higgs boson and vacuum expectation value;
- CKM matrix elements;

- **Neutrinos masses**: in the SM neutrino are massless, however experimental evidence of neutrino oscillations can be explained assuming that neutrino masses are different from zero and that the Electroweak eigenstates are a mixing of mass eigenstates. The seesaw mechanism, indeed, incorporates neutrino masses into the SM by introducing heavy Majorana neutrinos, whose masses are inversely coupled to the light SM neutrino masses. The heavy neutrinos are still not found and even introducing them in the SM, the PMNS matrix elements, related to the neutrino oscillations, are not predicted by the theory.

- **Dark matter and dark energy**: the cosmological observation [20] of the galaxy rotation profiles is one of the evidences show that the Universe is made up of only for the 5% of ordinary matter. The rest of the Universe is composed partially by the so-called Dark Matter, which constitutes 24% of the universe, while the remaining 71% is ascribed to a constant vacuum energy called dark energy. The existence of the dark energy would account for the accelerating expansion of the Universe.

- **Gravity**: the failure of the Standard Model to introduce a quantum description for gravity is one of the most relevant open issues. The effects of gravity interactions are fully understood over large distances, while, at very short distances, there is not a satisfactory explanation. This is because the coupling strength of gravity is very weak if compared to other interactions.

- **Higgs mass fine tuning**: the tree-level (bare) Higgs mass receives corrections from fermion loop diagrams which are quadratically-divergent and that are not cancelled by the boson loop diagrams, seen in Figure 1.6. A precise tuning between fermions and scalars, known as fine tuning, is required in order to reduce or cancel this divergences.

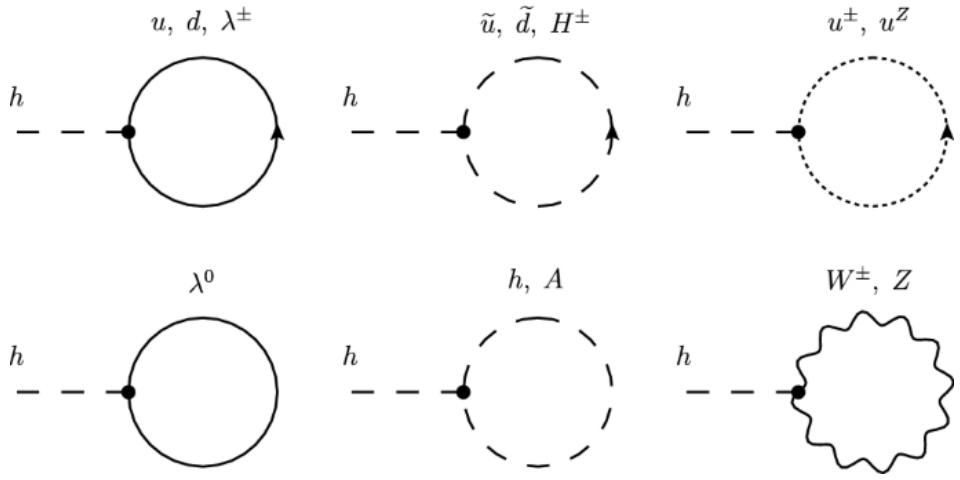

Figure 1.6: One-loop self-energy corrections to the Higgs mass.

- **Flavour Changing Neutral Currents (FCNC)**: as supposed by the Glashow, Iliopoulos, and Maiani (GIM) mechanism [21], are highly suppressed at three level and may occur in the SM only at next to the leading order. Also the presence of only three families of quarks and leptons is not predicted by the SM.

Chapter 2

The CMS Experiment at LHC

The Compact Muon Solenoid (CMS) [22] is one the main high energy physics experiments at CERN's Large Hadron Collider (LHC) [23]. The CERN, whose acronym derives from "*Conseil Européen pour la Recherche Nucléaire*", was established in 1954 in Geneva with 12 member states from the Western Europe. Nowadays, the acronym is still the same, but the name is changed to "*Organisation Européen pour la Recherche Nucléaire*" and it counts 23 member states. It is one the leading research institute for particle physics, hosting the biggest particle physics laboratory in the World. Several breakthrough discoveries have been made at CERN, such as the discovery of W and Z bosons at UA1 and UA2 experiments in 1983 [24, 25] or the already mentioned Higgs boson discovery in 2012 [5, 6]. Moreover, studies at CERN lead to improve technologies, helping us in our day-to-day lives like the World Wide Web in 1989.

2.1 The Large Hadron Collider

The Large Hadron Collider is the world's largest and most complex particles accelerator, it accelerates two high-energy beams travel at very high speed, then they are made to collide. The purpose is to study the interactions among colliding particles, probing the fundamental laws of nature. The beams are made up of protons and an hydrogen gas bottle is the source, whose electrons are stripped away thanks to an electric field. The LHC is the last element of an accelerator complex composed of multiple accelerator machines [26]: at each step, the respective accelerator machine increases the energy of the beams to its maximum achievable, then the beams are injected in the next step of the complex or are used for experiments conducted at that energy. The injection energy of beams at the LHC is 450 GeV. The accelerator complex is made up of:

- Linac2, a linear accelerator, that can accelerate protons up to 50 MeV;
- Proton Synchrotron Booster (PSB), which accelerates protons to 1.4 GeV;
- the Proton Synchrotron (PS), which increases the protons energy to 25 GeV;
- and, just before LHC, the Super Proton Synchrotron (SPS), that pushes the beam to 450 GeV.

In Figure 2.1 can be seen a schematic view of the accelerator complex.
The protons are injected in LHC in two different beam pipes, where the two beams travel in opposite directions. The LHC consists of a 27 Km ring, in which, thanks to a strong magnetic field realized with superconducting electromagnets, the beams are guided through the structure and are accelerated by radio frequency cavities.
During the Run 2 (2015-2018), the maximum beam energy is 6.5 TeV and the centre of mass energy at collision is 13 TeV. The data taking referred to as Run 3 has started in 2022, and LHC reached a centre of mass energy equal to 13.6 TeV. Also ions, from a source of a vaporised lead, are accelerated. The ions enter Linac3 and are injected in the Low Energy ion Ring, then follow the same route as the protons.
The high collision energy does allow to probe a wide energy range and affects the number of events produced in the collisions, but it is not the only figure of merit of the LHC accelerator. In fact, the

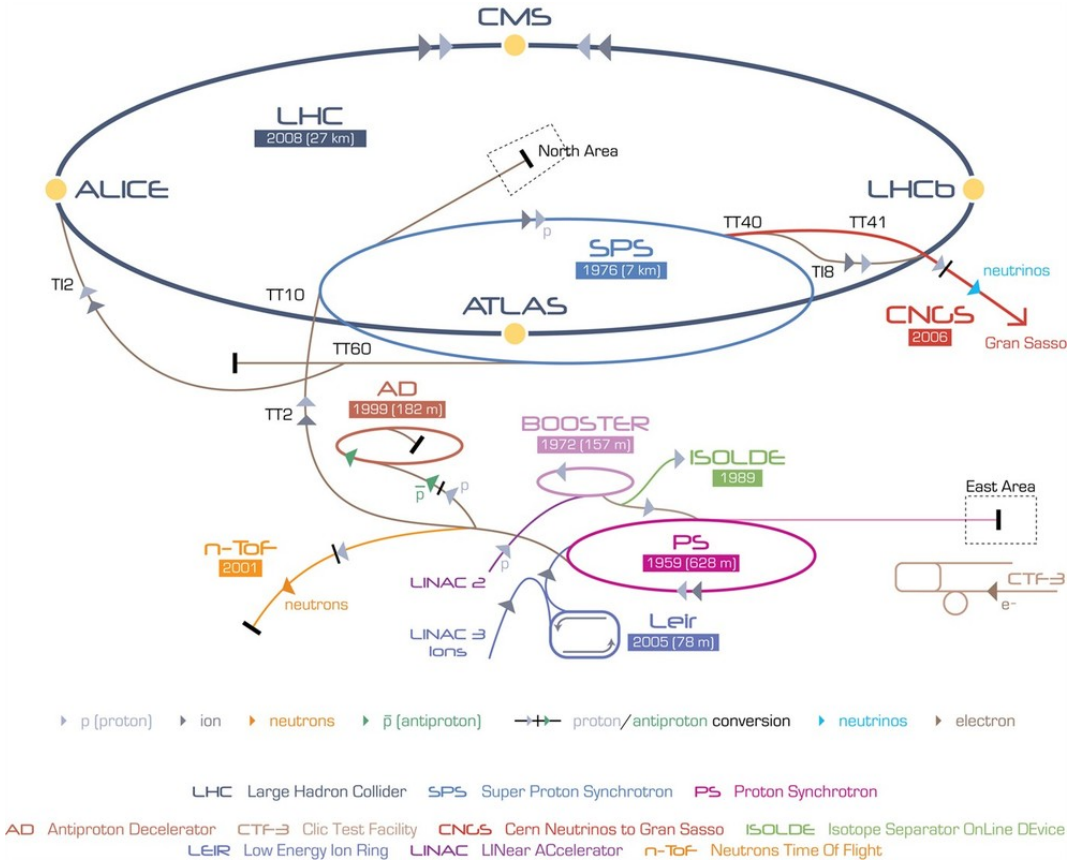

Figure 2.1: Schematic view of the accelerator complex at CERN.

rate of production is also affected by other constructive parameters of the machine. The rate R of a process with cross section σ is in particular equal to:

$$R = \mathcal{L}\sigma,$$

where $\mathcal{L}$ is the instantaneous luminosity. This parameter depends only on the beam parameters and, for a Gaussian profile beam, it can be written as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi\epsilon\beta^*} F, \tag{2.1}$$

where:

- N_b is the number of particles per bunch;
- n_b the number of bunches per beam;
- f_{rev} the revolution frequency;
- γ the relativistic Lorentz factor $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$;
- ϵ the normalized transverse beam emittance;
- β^* the beta function at the collision point, is a measure of how narrow the beam is at the interaction point, that could be seen as product of two separate function σ_x^* and σ_y^*, beam sizes in the horizontal and vertical plane divided by $\frac{\epsilon}{\gamma}$;

- F the geometric luminosity reduction factor due to the crossing angle at the interaction point, that can be written as follows:

$$F = \left(1 + \left(\frac{\theta_c \sigma_Z}{2\sigma^*}\right)^2\right)^{1/2}, \tag{2.2}$$

  where θ_c is the full crossing angle of the beams at the interaction point, σ_z is the bunch length, and σ^* is the transverse RMS beam size at the interaction point.

In Table 2.1 the value of the above parameters characteristic of the Run 2 Data taking is given [27]. The peak luminosity reached during Run 2 is equal to $2 \times 10^{34}\ cm^{-2}s^{-1}$ and Figure 2.2 shows the

Parameter	Value
N_b	1.10 - 1.15 $\times 10^{11}$
n_b	2220 - 2256
$\epsilon[\mu m]$	2.0 - 2.2
$\beta^*[cm]$	25 - 40
F	0.6

Table 2.1: The LHC machine parameters.

evolution of the LHC peak luminosity between 2011 and 2023.

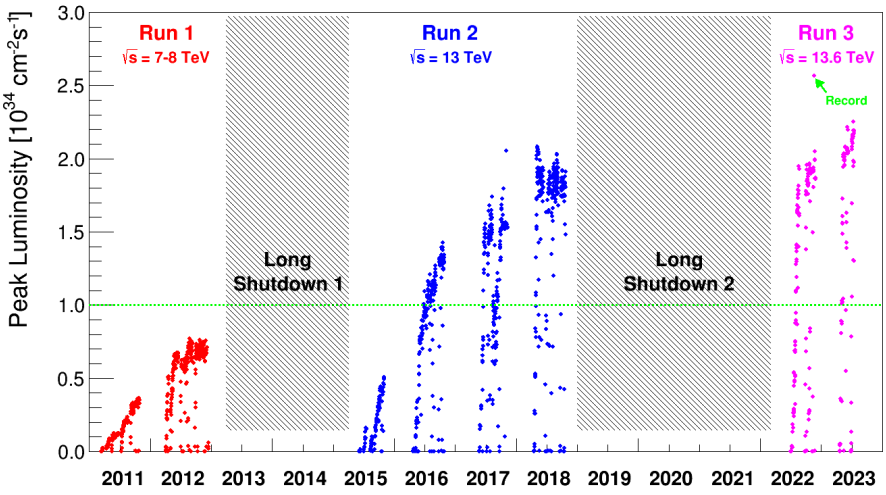

Figure 2.2: Evolution of the LHC peak luminosity between 2011 and 2023. [28]

The number of expected events can be written as:

$$N = L\sigma,$$

where, now, L is the integrated luminosity:

$$L = \int_0^{\Delta t} \mathcal{L} dt,$$

where Δt is the time considered.
After reaching the LHC target energy, beams are collided in four point where are the four main experiments:

- ALICE (A Large Ion Collider Experiment) [29] is studying the heavy-ion (Pb-Pb nuclei) collisions. The centre-of-mass energy in Run 2 per nucleon pair is 2.76 TeV, and it is equal to 5.36 TeV during Run 3. At this energy density and temperature are expected to be high enough to produce quark-gluon plasma and to recreate conditions, that are believed to have existed during the evolution of the Universe after the Big Bang.

- ATLAS (A Toroidal LHC ApparatuS) [30] is a general purpose detector, that has been designed to study a large variety of processes, that involves the production of the top and beauty quarks and the Higgs boson. Moreover its task is to study processes beyond the SM, as the existence of new particles, as the Vector-Like Quarks (VLQ), object of study in this thesis.
- CMS (Compact Muon Solenoid) [22] is a general purpose detector, and it will be described in detail in the next section.
- LHCb [31] is studying the properties of bottom quark, and searching answers to questions like the matter-antimatter asymmetry. The detector focuses on more forward decay products, aiming to study the violation of the CP symmetry in the decay of B mesons.

A schematic view of the LHC and the four experiments is shown in Figure 2.3.

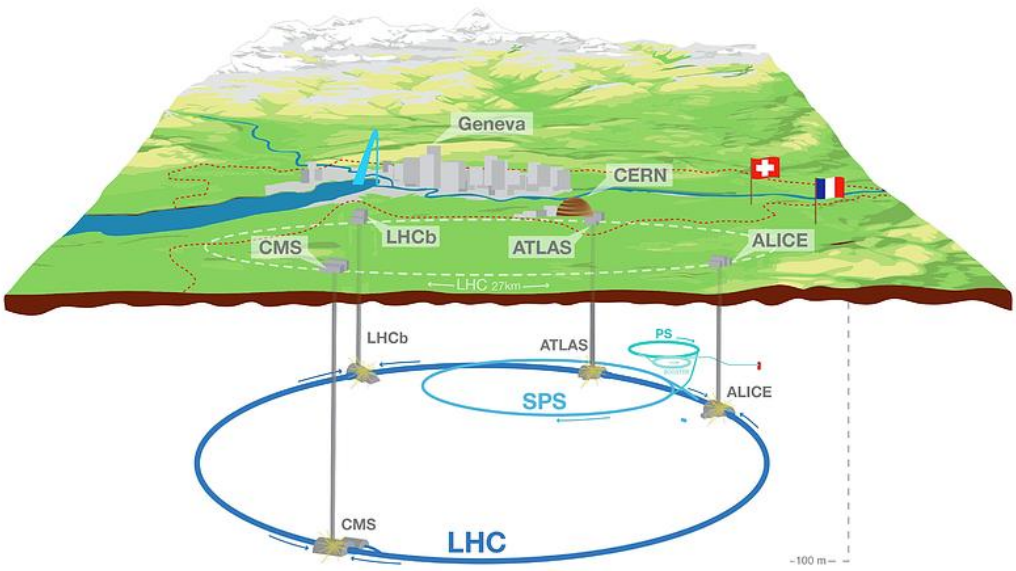

Figure 2.3: A schematic view of the four experiment at LHC.

2.2 The Compact Muon Solenoid Experiment

The CMS experiment is a general purpose detector, centered around a huge solenoid magnet, from which its name is derived. A magnetic field of 3.8 T is produced by a cylindrical coil of superconducting cable. The overall dimensions of the detector are 21 meters in length, 15 meters in width, and 15 meters in height. The main goal of CMS is to distinguish physical processes of interest from backgrounds, detecting the particles produced in the hadron collisions, and accurately measuring their energies, charges, masses, and momenta. To achieve this goal, the detector is made up of different subdetectors, each of them with high momentum or energy resolution in a range appropriate for particle coming from LHC collision events. The subdetectors are designed to work at high luminosities, therefore in a high radiation environment, maintaining good performances over the course of several years. It is also required an high spatial resolution, or high granularity, for the subdetectors, to reduce the contamination from the so-called pile-up. The pile-up is the overlapping between two particles of the same event or from different events in the same bunch-crossing, which are thus reconstructed as one

particle. If the time resolution is not good enough or the detector's response is slower than the time interval between two bunch-crossing, a similar pile-up phenomenon has to be taken into account. Also a wide angular coverage is required. In Figure 2.4 is shown a picture of CMS.

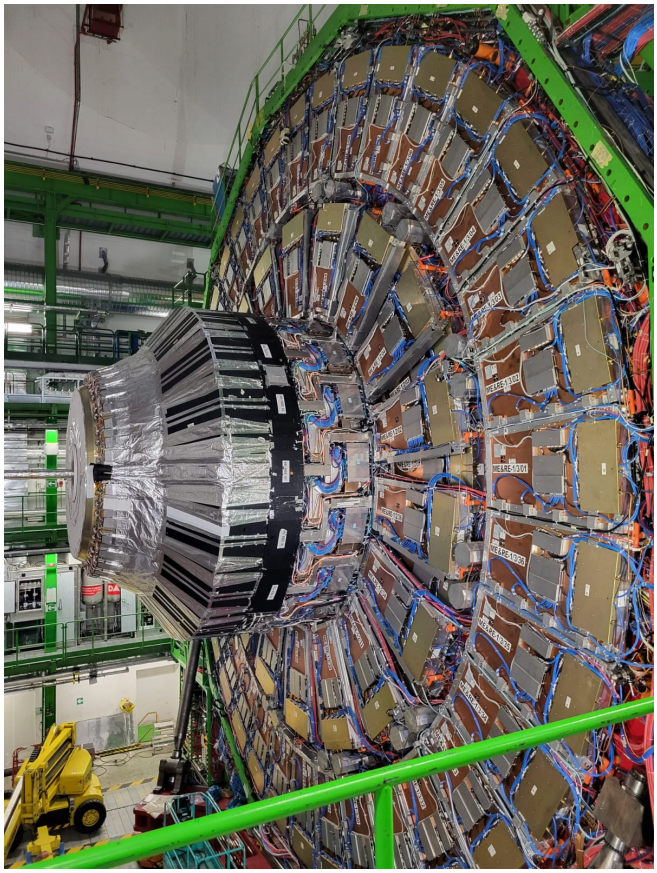

Figure 2.4: A CMS picture, taken in December 2023.

A right-handed Cartesian coordinate system is used to characterize the reconstructed objects in the detector. The coordinate system is centered in the nominal interaction point and oriented in the following way:

- x-axis points towards the center of the LHC ring and represents the horizontal coordinate;
- y-axis points upwards, perpendicular to the LHC plane;
- z-axis along the anticlockwise-beam direction.

However, the coordinate system chosen to describe physics objects is a cylindrical one, where coordinates are as follows:

- the radial distance r from the z-axis;
- the azimuth angle ϕ is the angle of rotation around the z-axis with origin identified on the x-axis and increasing clockwise looking in the positive direction of the z-axis;
- the polar angle θ is defined as the rotation angle around the x-axis with the origin on the $z-axis$ and increasing clockwise looking in the positive direction of the x-axis.

Instead of using θ, it is convenient to use the pseudorapidity η, which is defined as:

$$\eta = -ln\left[tg\left(\frac{\theta}{2}\right)\right], \tag{2.3}$$

because in the relativistic approximation it transforms linearly under boost along the z-axis. Moreover, under this kind of boosts, the difference in η between two relativistic particles ($\Delta\eta$) results to be invariant. As it can be seen in Figure 2.5, as the angle θ increases from zero, the pseudorapidity decreases from infinity. Thanks to these coordinates, the angular distances between two objects can

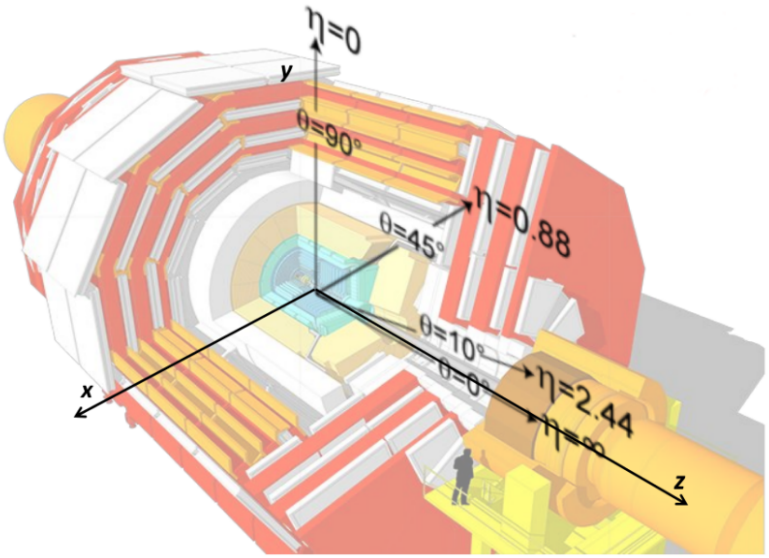

Figure 2.5: The coordinate system of CMS and the functional dependence of η by θ.

be written as a Lorentz invariant variable under boosts along the $z-axis$:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{2.4}$$

Two important variables are p_T and E_T, respectively the transverse momentum and the transverse energy, defined as:

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

$$E_T = E\sin\theta. \tag{2.6}$$

As previously said, CMS is composed by subdetectors as shown in Figure 2.6.
The subdetectors are:

- the tracking system;
- the electromagnetic calorimeter (ECAL);
- the hadronic calorimeter (HCAL);
- the magnet;
- the muon system.

Each subdetectos is made up of cylindrical layers coaxial to the beam pipe, called barrel layers, and two endcaps that ensure hermetical closure of the detector.
Of course, CMS is provided, also, of a trigger and data acquisition (DAQ) system.
In the next paragraphs the subdetectors will be described.

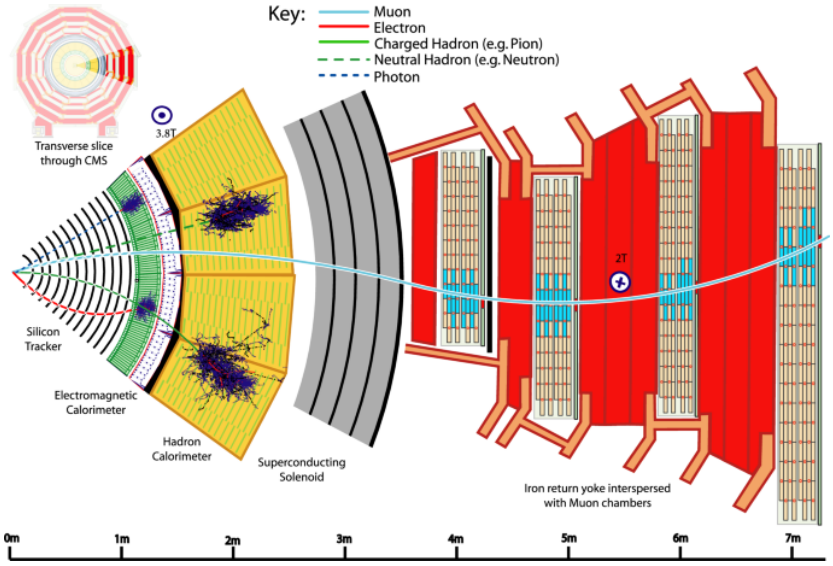

Figure 2.6: Schematic view of CMS subdetectors.

2.2.1 Tracking System

The inner tracking system [32] has been installed in the innermost part of the CMS experiment to provide a robust tracking and a detailed vertex reconstruction. Thanks to its high spatial resolution and fast response, it can distinguish very nearby tracks and ensure the correct working of the online trigger, despite the huge particles rate. The detector has a diameter of 2.5 m and a length of 5.8 m, covering up to $|\eta| < 2.5$. The total active area is equal to 200 m^2 and it is made up of pixel vertex detector and the silicon strip detector (SST). The pixel vertex detector is the closest one to the interaction point, it is composed by three barrel layer (BPIX) and two disk both in the positive and negative endcaps. With a total of over 65 million silicon pixels, it ensures an accuracy of 10 μm and of 20 μm respectively for radial and transverse position measurements. The external part of the pixel detector is covered by the silicon strip detector. It consists of 4 barrel of Silicon Strip layers parallel to the beam axis, Tracker Inner Barrel (TIB), and 3 disks of Silicon Strip at each end of the TIB, the Tracker Inner Disks, called TID. The outer volume, outside the TIB/TID, is also composed by SST and is called Tracker Outer Barrel, TOB. At both ends of the TOB are located other Tracker EndCaps named TEC+ and TEC-, whose signs indicate the location along the z-axis. It has a resolution from 35 μm to 52 μm in the radial direction and 530 μm in the transversal direction. Between 2016 and 2017 data taking the pixel detector has been upgraded [33], increasing the pixel detector layer from three to four and achieving a distance to the proton-proton interaction point of 29 mm. In Figure2.7 is shown a schematic view of the tracker system.

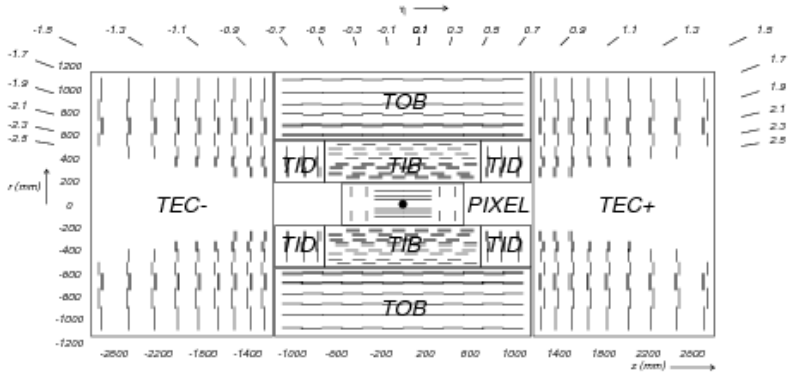

Figure 2.7: Schematic view of the tracker system. [32]

2.2.2 Electromagnetic Calorimiter

The energy of electrons and photons are measured by the electromagnetic calorimeter (ECAL) [34], which is a hermetic homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals. It is divided into a barrel section (EB) and two endcaps (EE+ and EE-), with a radius between 1.25 m and 1.8 m. The barrel section covers a range $|\eta| < 1.48$, while the two endcaps from 1.48 to 3. The lead tungstate is used because it assures an high density of crystals (8.28 g/cm^3), a small Molière radius (2.2 cm) and a short radiation length (0.89 cm). The photons are collected by photodetectors: in the EB avalanche photodiodes (APDs) are used, while in the EE vacuum phototriodes (VPTs) are used, since the radiation in this region results to be too high for APDs. To improve the photon-π^0 separation a preshower detector (ES) is placed in front of the endcap, it is based on lead absorber and silicon strip sensors. In Figure 2.8 a schematic view of the ECAL.

2.2.3 Hadronic Calorimeter

The energy of hadrons, such as protons, kaons, pions, and also neutrons, is measured in large part thanks to the hadronic calorimerter (HCAL) [35]. The hadron showers are generally bigger than electromagnetic showers, thus this is a sampling calorimeter, i.e. composed by layers of active material and absorber. The layers of active material are made up of fluorescent scintillators, while the absorber

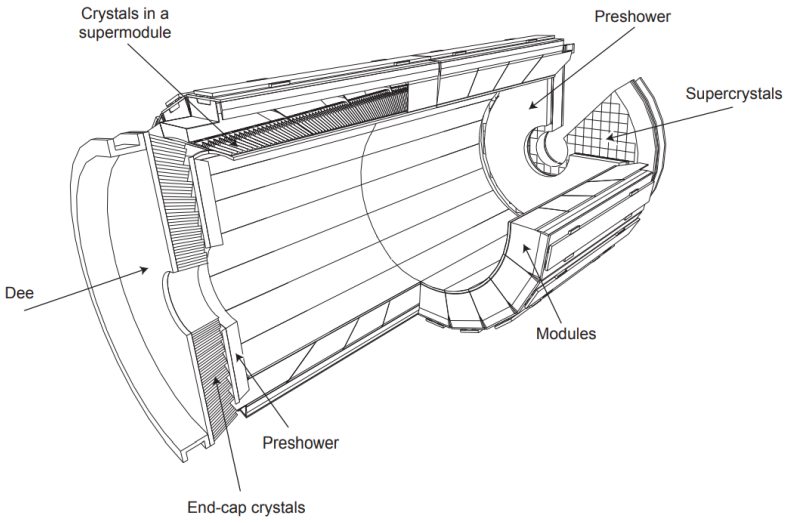

Figure 2.8: Schematic view of ECAL system.

used is cartridge brass (C26000). The latter, with a density of of 8.83 g/cm^3, provides a radiation length of X_0=1.49 cm and with a nuclear interaction length of λ_I=16.42 cm. It is composed by 70% of Cu and by 30% of Zn. The HCAL is used to measure the energy of hadrons such as protons, kaons, pions and also neutrons. To assure the containment of the entire hadron shower, it was realized as a sampling calorimeter. It is composed by layers of active material and of absorber. The different sections that compose the HCAL are: the Hadron Calorimeter Barrel (HB and HO), the Endcap (HE) and the Forward (HF) section. The HB and HE regions cover respectively the pseudorapidity range of $|\eta|$ <1.3 and of 1.3< $|\eta|$ <3. The Forward sections (HF) are located 11.2 m away from the interaction point, covering the pseudorapidity range 3< $|\eta|$ <5.2. It is made up of quartz fibres embedded within a 165 cm long steel absorber and uses a Cherenkov-based technology. The Hadron Calorimeter Outer(HO) is composed by additional scintillators that are placed outside the solenoid to ensure adequate sampling depth and to measure late shower development. In Figure 2.9 a schematic view of the HCAL is given.

The ECAL and HCAL energy resolution can be parametrized as:

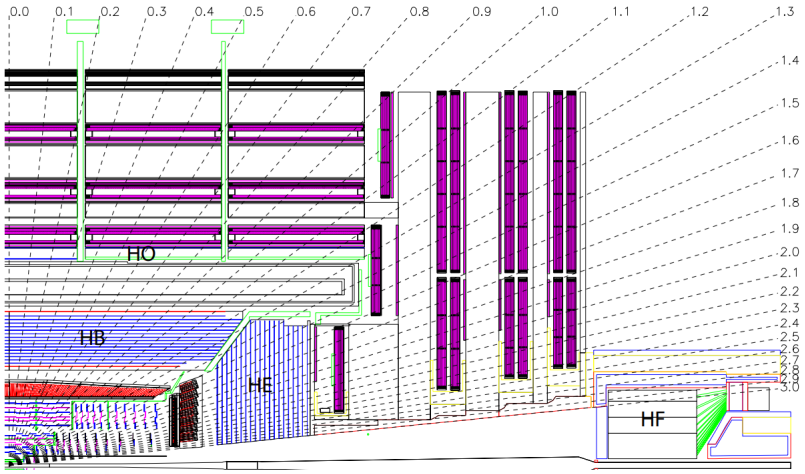

Figure 2.9: Schematic view of the HCAL.

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + (c)^2 \tag{2.7}$$

where:

- a is a stochastic term and it takes in account the statistical fluctuations of the signal in the shower containment, such as fluctuations in the number of primary particles and/or the number of photons which includes fluctuations in the shower containment collected by a photomultiplier;
- b is the noise term which contains the contributions from electronic noise and pile-up energy, negligible at low luminosity;
- c is the constant term which takes in account the fluctuations of the longitudinal leakage, of calibration errors and of leakage of energy from the back of the crystal.

At the CMS experiment, the energy resolution of the Electromagnetic Calorimeter is:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{0.0280}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.0030)^2\,, \tag{2.8}$$

where E is in GeV, while for the HCAL the energy resolution is:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{0.8470}{\sqrt{E}}\right)^2 + (0.0740)^2\,, \tag{2.9}$$

2.2.4 Magnet

The magnetic field of 3.8 T is provided by the magnet, installed outside the HCAL. It is made up of a superconducting coil and its main goal is to curve the particles produced in the collisions to measure their charge and momentum. The iron Return Yoke confines the magnetic field, introducing an approximately 1.8 T constant field outside the magnet and avoiding border effects. The yoke is made up of alternating layers interspersed with the muon detectors. Moreover, it stops all the remaining particles except for muons and neutrinos.

2.2.5 Muon System

Muon identification, momentum measurement, and triggering are provided by the CMS muon system [36]. Muons are objects that can be the present in the final stat of several physics processes, thus robust and precise muon reconstruction, identification, and measured properties can be used to distinguish processes of interest from backgrounds. A large detector surface is required since muons can cross several layer of material, arriving to the external paths of the detector. The total area covered by the muon system is equal to 25000 m^2. The detectors need to be reliable, robust, and of contained costs. To achieve these requests four different kinds of gas detector are used: Drift Tubes, Cathode Strip Chambers, Resistive Plate Chambers, and Gas Electron Multipliers.

- The Drift Tubes (DTs) are placed in the barrel and they cover a pseudorapidity range up to 1.2. The DTs are divided into 4 stations. In the first 3 stations, there are 2 layers of 4 chambers each, that measure the $r-\phi$ coordinate of the muon, and one layer of 4 chambers to measure the z coordinate. To avoid dead spots, each cell of the chamber is divided by an half-cell width with respect their neighbour.
- The Cathode Strip Chambers (CSCs) are placed in the two endcaps and cover the range $0.9 < |\eta| < 2.4$. The CSCs have a good segmentation, a fast response time, and an high radiation resistance, these properties are needed in the endcaps, where muon and background rates are higher and the magnetic field is not uniform. The cathode strips of each chamber provide position measurements in the $r-\phi$ plane. The anode wires allow to measure the pseudorapidity and the beam-crossing time of every muon. In each endcap there are four stations of CSCs.

- The Resistive Plate Chambers (RPCs) are placed both in the barrel and the endcaps. The RPCs provide a high time resolution, while the spatial resolution is lower than the DTs and CSCs. The RPCs are double-gap chambers, operated in avalanche mode to ensure robust operation at high rates. In the barrel muon system there are 6 layers of RPCs, while in the endcaps there are 4 layers, one for each of the four muon stations. In the first two stations of the barrel two layers of RPCs are located on the top and on the bottom of the DT chamber, while just one on the outermost side of the the last two stations of DTs. More details on the RPC detector are given in the next section.

- The triple Gas Electron Multipliers (triple-GEM) [37] have been installed during the Long Shutdown 2 (LS2) in the endcap region. The main goal of the GE1/1 detector is to improve the L1 trigger performances in the region $1.55 < |\eta| < 2.18$, adjacent to the CSCs. Each chamber consists of two PCBs enclosing the gas volume, between these two there are three foil of insulating polymer with copper conductors on the top and bottom surfaces. Microscopic holes are uniformly etched in a hexagonal pattern across the foil. Applying a potential difference across the foils creates strong electric fields within the holes. Electrons generated during gas ionization drift towards the foils and undergo multiplication within the holes. This multiplication process triggers a readout signal on the closely positioned strips.

Both DT and CSC subsystem can trigger on the muon p_T with high efficiency and good background rejection, independently from the rest of the detector. The Level-1 (L1) trigger p_T resolution is about 15% in the barrel and 25% in the endcap. Thanks to the fast response of the RPCs and GEMs, even low p_T tracks, that may stop before reaching the outer two stations, could be triggered.

2.2.6 Trigger and data acquisition system

The high instantaneous luminosity reached at LHC does not allow to save the entire set of data, thus only events with some properties are selected. The trigger system performs this event selection, reducing the rate by a factor 10^6 in two different steps: the Level-1 trigger and the High-Level Trigger (HLT) [38]. Figure 2.10 shows a scheme of the trigger system. 2.10.

The first step is provided by the fast and automatic event selection by the L1 trigger. Different components are taken into account: local, regional, and global components. The local components, called Trigger Primitive Generators (TGP), are based on energy deposits in calorimeter trigger towers and track segments or hit patterns in muon chambers. The regional triggers use a pattern logic to sort trigger objects like electron, photon or muon, in limited spatial regions. The global components, Global Calorimeter and Global Muon Triggers, determine the number of trigger objects across the entire experiment and transfer them to the Trigger Control System (TCS). This last module takes the final decision based on algorithm calculations and on the readiness of the subdetectors and the DAQ. The Level-1 Accept (L1A) decision is communicated to the subdetectors through the Timing, Trigger, and Control (TTC) system. The architecture of the L1 Trigger can be seen in Figure 2.11.

The HLT takes the final decision to accept or reject an event. A software filter system is used, accessing to the complete read-out data and performing an analysis based on the physical objects present in the event. For example events with bad vertex reconstruction or events with final objects characterized by a too low overall momentum are immediately refused by the HLT.

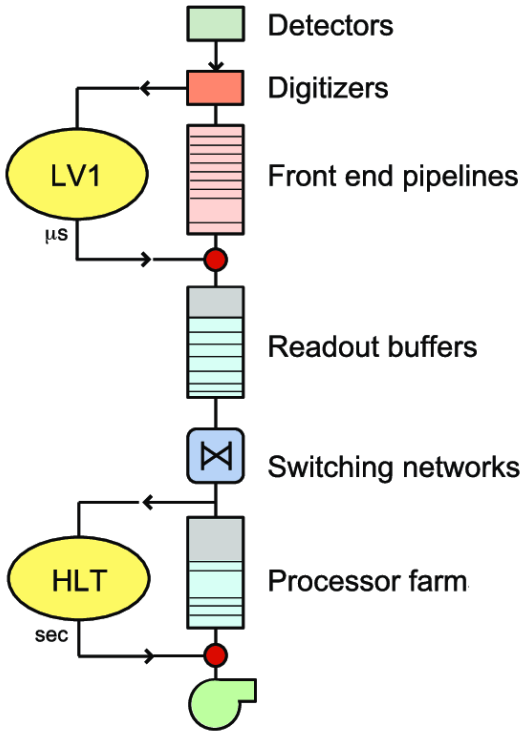

Figure 2.10: Trigger System at CMS.

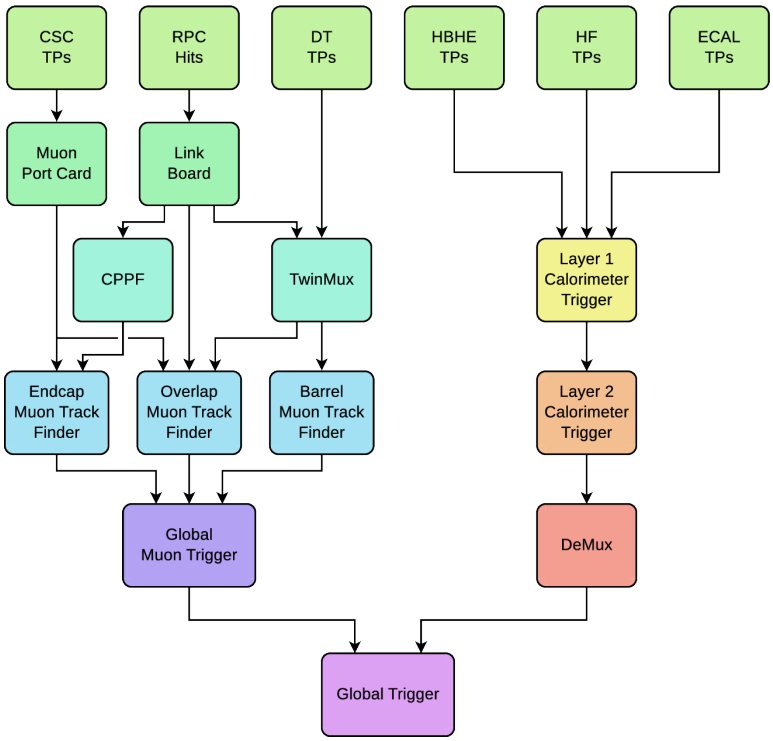

Figure 2.11: Architecture of Level1 Trigger.

2.3 CMS Muon Background studies

Background particles in the muon system could degrade the quality of muons triggering, identification, and reconstruction. To maintain robust operation and to choose developments for future upgrade, it is needed to understand and to analyse the background. Figure 2.12 shows a schematic view of the muon system, after the installation of the GEM detector. This section focuses on the background measured in

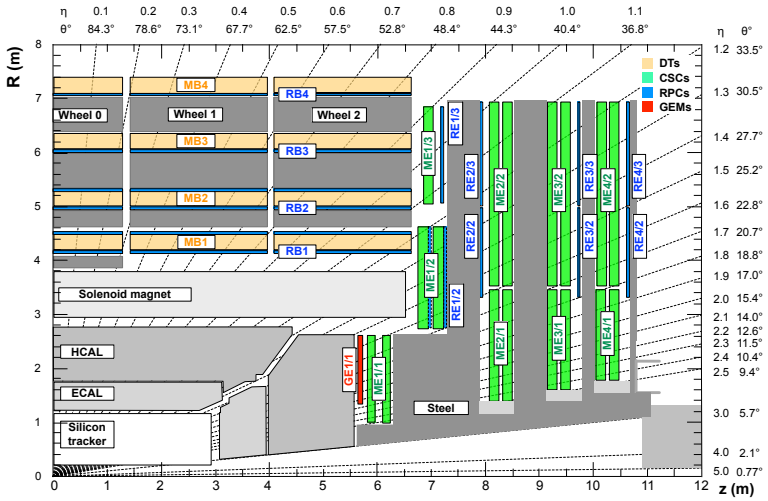

Figure 2.12: CMS Muon System schematic view. [39]

the RPCs. These double-gap chambers utilize Bakelite plates and operate in avalanche mode, offering an intrinsic time resolution of approximately 2 ns. This resolution significantly surpasses the 25 ns front-end electronic time window, allowing the precise assignment of time slots for detected particles. Consequently, these chambers serve mostly as trigger detectors in both the barrel and endcap [40]. Chambers in the barrel are labelled as *RB* followed by a number identifying the station, going from 1, the innermost station, i.e. closer to the beam, up to 4, the outermost station. Chambers in the endcap are labelled as *RE* followed by the number of the disk and the number of the ring. RPCs are installed up to $|\eta| < 1.9$, where η is the pseudorapidity. Particles are detected as "cluster" of adjacent hit strips, referred as reconstructed hit in the detector. The background comprises both long-lived background sources, such as neutron background in the CMS cavern, and promptly induced background during collisions. The background rate is defined as the number of "cluster" per time and area. It is expected

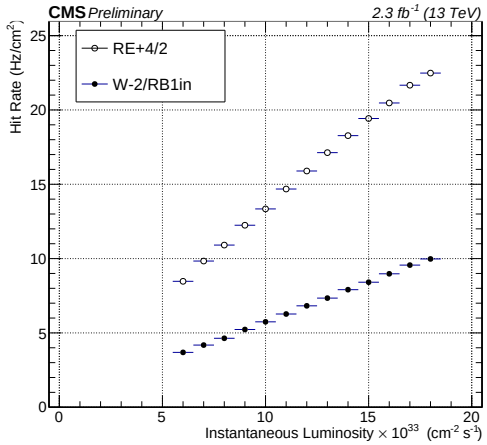

Figure 2.13: Background hit rate as a function of instantaneous luminosity for two different regions, the first layer of the innermost station of wheel -2 (W-2/RB1in) with full markers, and the second ring of the fourth positive endcap disk (RE+4/2). The rates show linear trends.

to increase with the instantaneous luminosity, since the number of simultaneous collisions increases. The analysis of data taken on 2018 (Run 2) shows in Figure. 2.13 a linear trend of the background hit rate as a function of the instantaneous luminosity for two different regions as example. The data analyzed are enriched in background, which is dominated by the one measured during collisions, since the number of triggered events is much larger than the one with no collisions, and as such is influenced by the LHC filling scheme. One LHC orbit consists of 3564 units, said buckets, spaced 25 ns, and assigned an identification number each, called BX Id, which refers to the number of bunch crossing in each train constituting the proton beam. The way those buckets are filled with a proton bunch or left vacant defines the filling scheme. The one adopted for this run period has 48 consecutive colliding bunches, spaced by small or large gaps where there are not collisions (7 or 31-35 consecutive non-colliding bunches). Figure 2.14 shows the rate as function of BX Id for the W-2/RB1in region at fixed valued of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ of instantaneous luminosity, the shaded area indicates positions along the LHC orbit without collisions and it shows lower rates as expected than the rates measured during collisions. The ensemble of all colliding BXs are referred to as *Colliding* region, to the *Non Colliding* region the whole of all gaps between two trains. The region before the first colliding bunch is labelled as *Pre-Beam* region, and the one after the last colliding bunch as *AbortGap* region.

The rate trend as function of instantaneous luminosity is studied separately for the four regions of the

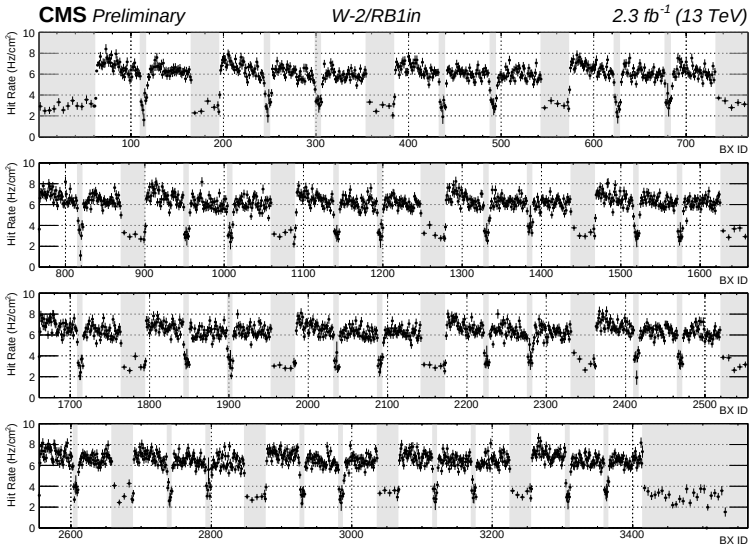

Figure 2.14: Background hit rate as function of BX Id for the W-2/RB1in region at fixed value of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ of instantaneous luminosity, the shaded area indicates positions along the LHC orbit without collisions.

filling scheme, i.e. *Colliding (C)*, *Non Colliding (NC)*, *Pre-Beam (PB)*, and *AbortGap (AG)* region. Figure 2.15 (left) shows the rate trend for the *Colliding (green)*, *Non Colliding (red)*, *Pre-Beam (cyan)*, and *AbortGap (blue)* region.

It is possible to define the inclusive (or total) background as:

$$B_{Tot} = \frac{N_C B_C + N_{NC} B_{NC} + N_{PB} B_{PB} + N_{AG} B_{AG}}{N_C + N_{NC} + N_{PB} + N_{AG}} \tag{2.10}$$

where B is the background measured and averaged in each region and N is the number of bunches of the region. The delayed (or secondary) background can be estimated as weighted average of the background measured in the three regions with no collisions:

$$B_{Sec} = \frac{N_{NC} B_{NC} + N_{PB} B_{PB} + N_{AG} B_{AG}}{N_{NC} + N_{PB} + N_{AG}}. \tag{2.11}$$

The prompt (or primary) background is estimated as the difference of the rate measured during collisions and the delayed background. Figure 2.15 (right) shows the inclusive (black), prompt (yellow), and delayed (magenta) background rates trend as function of instantaneous luminosity. All rates are

fitted as a linear function of the instantaneous luminosity, with the intercept representing the intrinsic RPC noise.

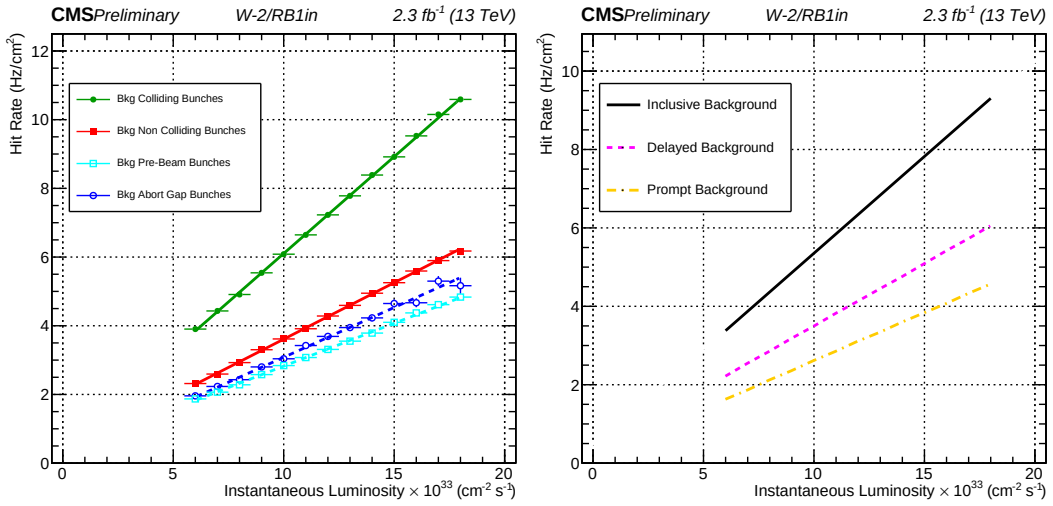

Figure 2.15: Background hit rate as function of instantaneous luminosity for the W-2/RB1in region in the *Colliding* (green), *Non Colliding* (red), *Pre-Beam* (cyan), and *AbortGap* (cyan) regions (left). Inclusive (black), delayed (magenta), and prompt (yellow) background hit rate (right).

2.3.1 Background Rate vs η

In order to have a clear overview of the rate in different detector regions, the background rate is evaluated as a function of η at fixed value of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ of instantaneous luminosity, extracted from a linear fit, as shown above.

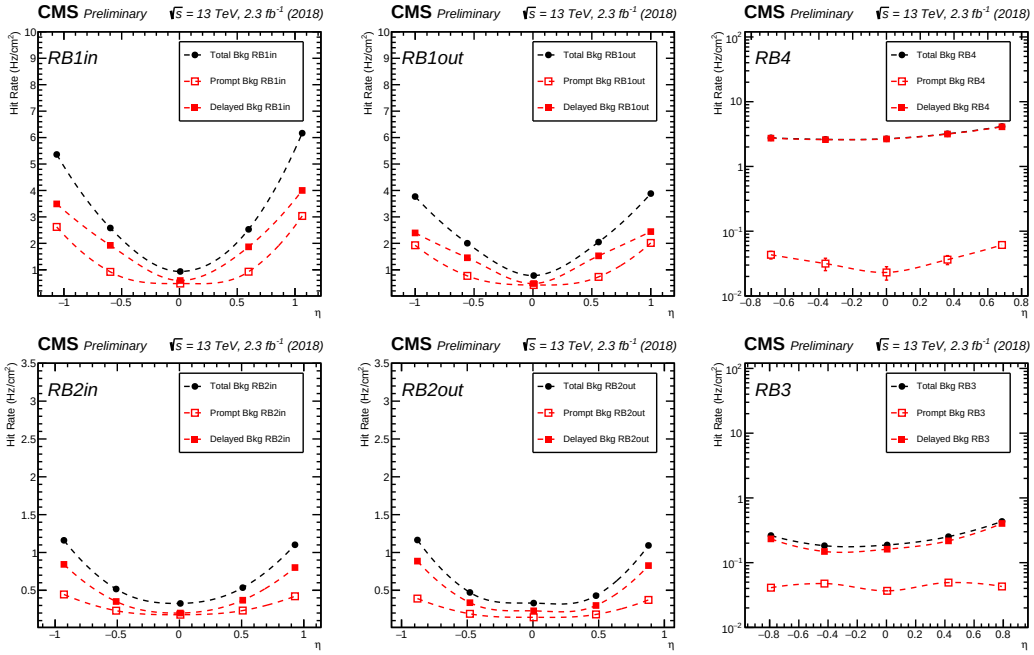

Figure 2.16: Total, prompt, and delayed background hit rate as function η at fixed value of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ instantaneous luminosity for the barrel. In the RB4 region the total background rate is dominated by the delayed component and the two distributions are overlayed.

Figure 2.16 shows the total, prompt, and delayed in the barrel region. The three rates increase with $|\eta|$. The prompt rate decrease with the increasing of R, i.e. the distance of the chambers to the beam axis, in the outermost region the delayed background dominates, associated to the slow neutron background of the cavern, thus the delayed and total rate distributions are overlayed.

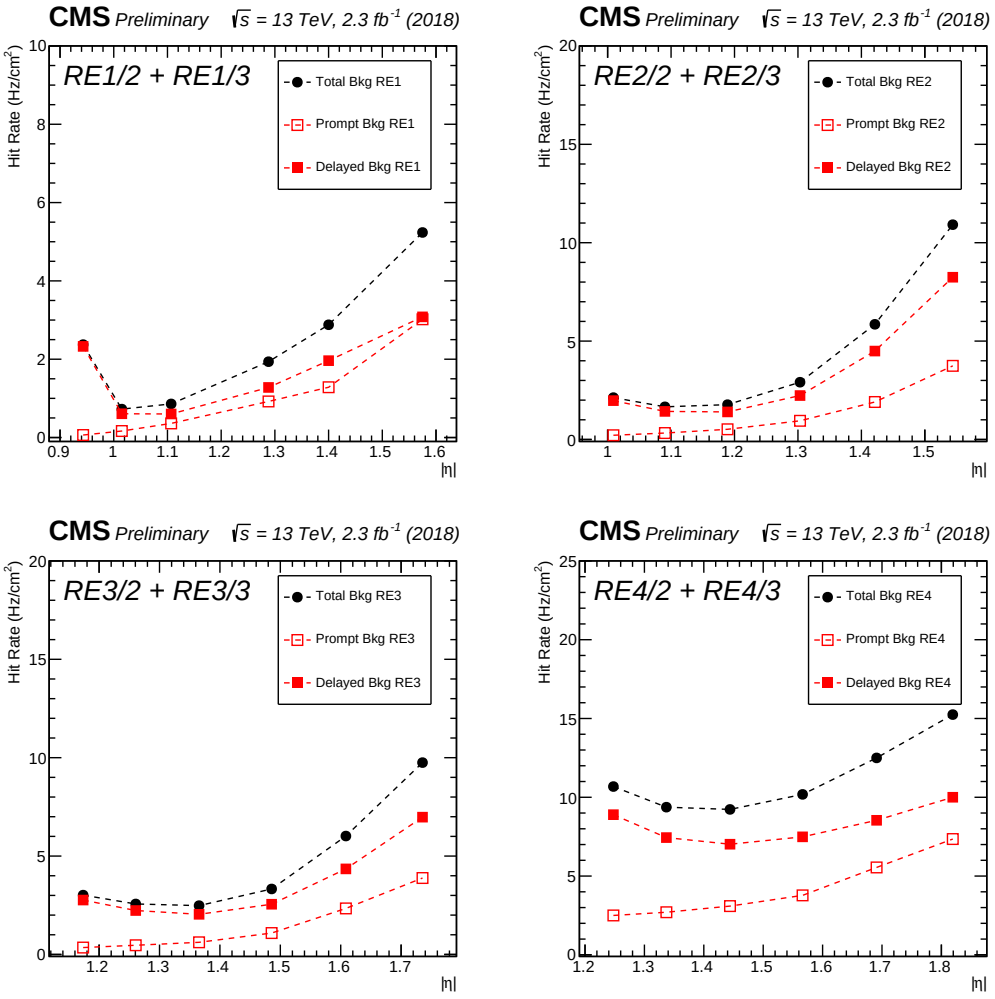

Figure 2.17: Total, prompt, and delayed background hit rate as function of $|\eta|$ at fixed value of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ instantaneous luminosity averaged for the positive and negative endcap regions.

Figure 2.17 shows the total, prompt, and delayed background rate in the endcap region, averaged between the positive and negative region. The prompt rate increases with the increasing of $|\eta|$, while the delayed background rate shows an initial decreasing and a consecutive increasing.

2.3.2 Background Rate vs Φ

Background studies have been both in barrel and endcap stations, considering the dependence by Φ. A granularity of 30 degrees is considered to have enough statistics and in the coordinate system Φ equal to 90 degrees will refer to the top sectors, while Φ equal to 270 degrees to the bottom sectors.

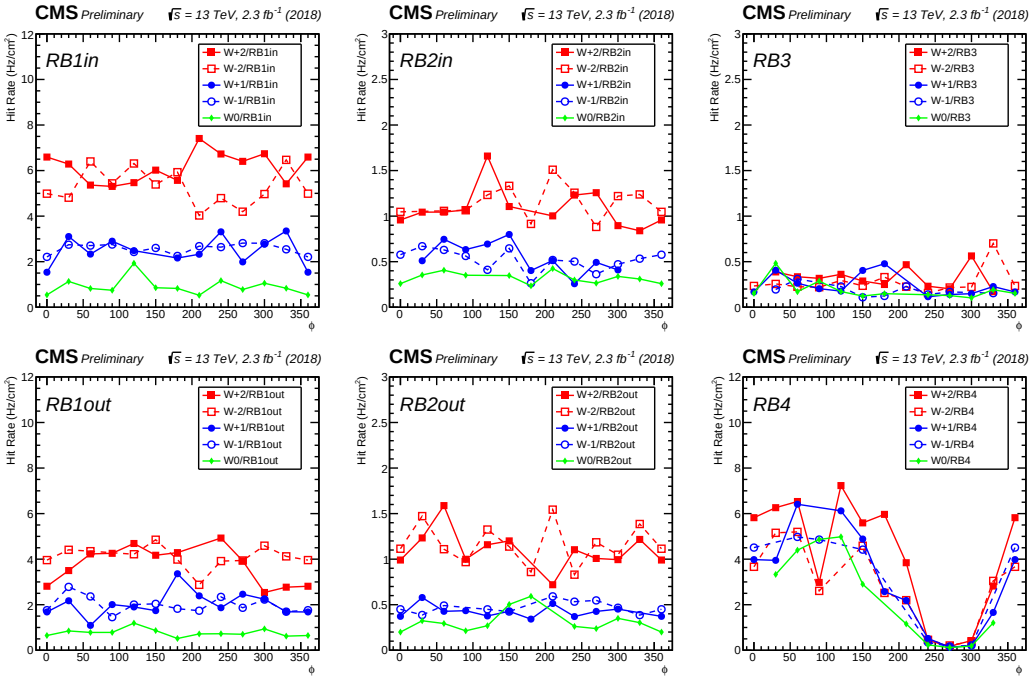

Figure 2.18: Total background hit rate as function of Φ at fixed value of $1.0 \cdot 10^{34}$ cm^{-2} s^{-1} instantaneous luminosity for the barrel regions.

In Figure 2.18 the total background as a function of Φ at fixed value of $1.0 \cdot 10^{34}$ cm^{-2} s^{-1} instantaneous luminosity for the barrel regions is shown. The rate is flat for all the region except for RB4. In particular, since the RB4 is dominated by the cavern background, the top sectors has larger background than the bottoms sectors. For the other regions fluctuation are given by noisy or dead strips.

Figure 2.19 shows the primary rate as a function of Φ. The primary rate decreases as the distance from the beam increases, as already shown, and it is flat for all the regions.

Figure 2.20 shows the secondary rate as a function of Φ. In the outermost region RB4 it increases, due to the cavern background contribution, which gives the modulation in Φ.

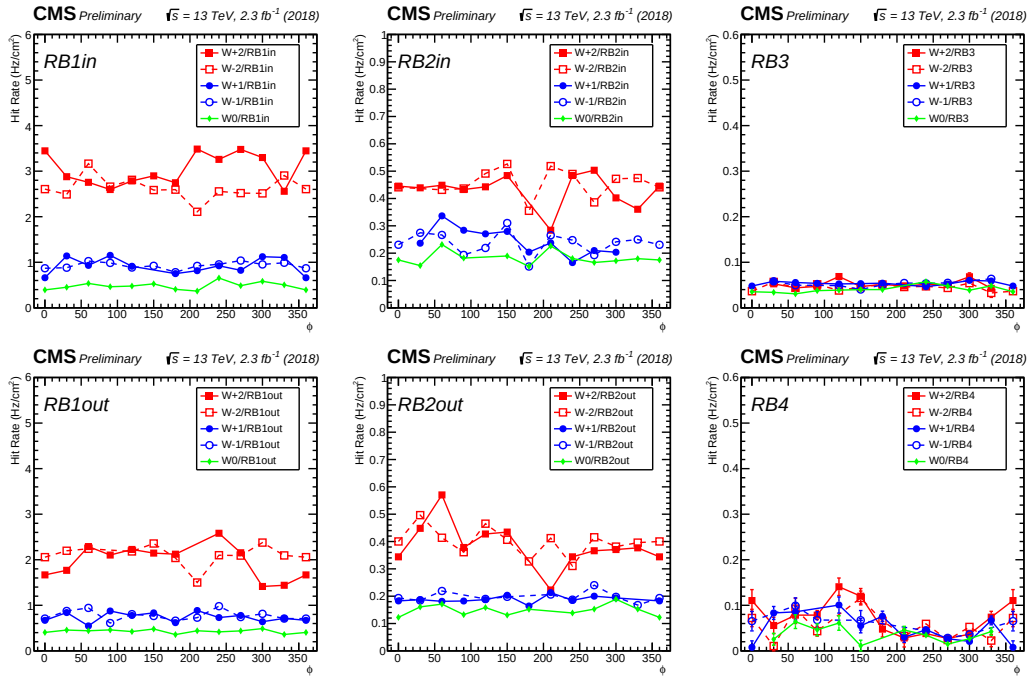

Figure 2.19: Prompt background hit rate as function of Φ at fixed value of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ instantaneous luminosity for the barrel regions.

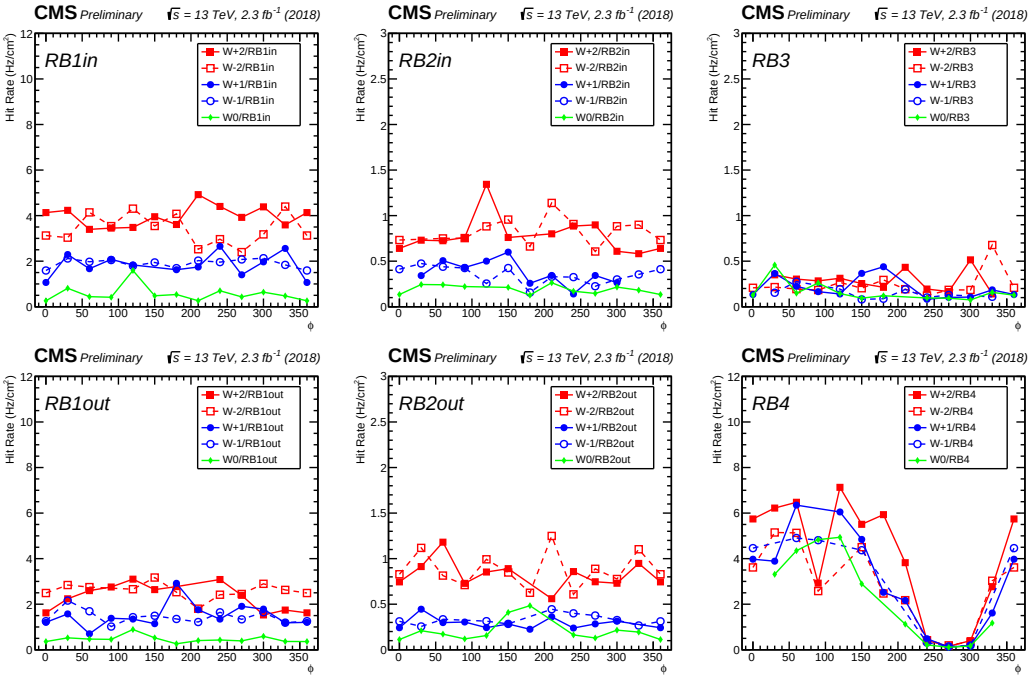

Figure 2.20: Delayed background hit rate as function of Φ at fixed value of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ instantaneous luminosity for the barrel regions.

The same studies are shown for the endcap regions in Figure 2.21, respectively the primary background on the left and the secondary background on the right. In the inner disks, i.e. the closest to the interaction point, the background is flat, while for the disk 3 and 4 a modulation is Φ is shown.

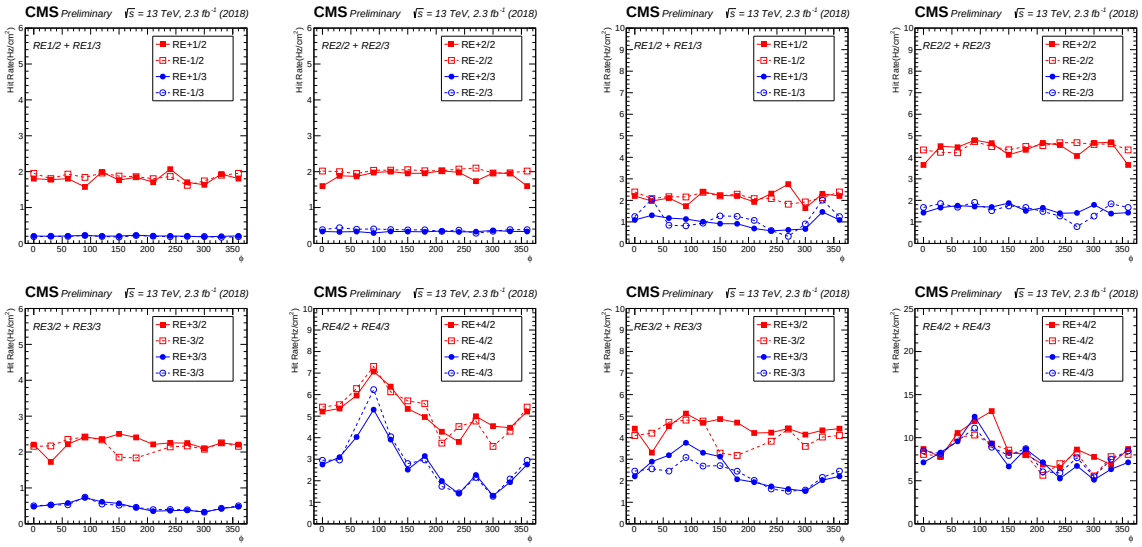

Figure 2.21: Primary (left) and Secondary (right) background hit rate as function of Φ at fixed value of $1.0 \cdot 10^{34}\ cm^{-2}\ s^{-1}$ instantaneous luminosity for the endcap regions.

2.3.3 Run 3 Results

The background studies have been performed on Run 3 data, taken on August 2022, at $\sqrt{s} = 13.6\ TeV$. The first results have been compared to the previous ones on Run 2. Figure 2.22 shows the background rate as a function of instantaneous luminosity for the Wheel+1 outermost region (W+1/RB4). As for this study the filling scheme is not taken into account and the colliding rate dominates, the two filling scheme are slightly different, 2400 colliding bunches in 2022 w.r.t. 2544 colliding bunches in 2018. Multiple mitigation actions, the most prominent being the installation of a set of shields during the Long Shutdown 2 (2019-2022), show a significant reduction of background in the outermost barrel region, improving the longevity and stability of the chambers. Understanding the background plays

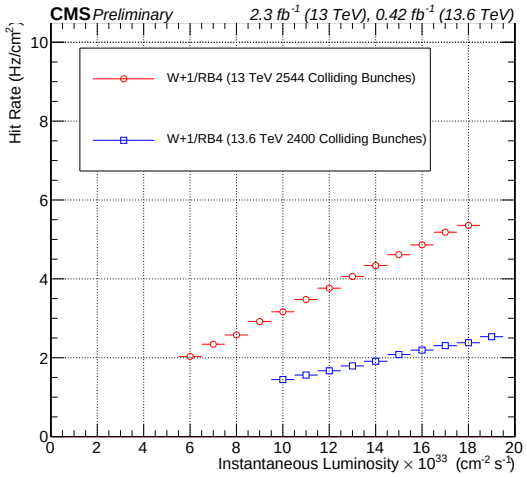

Figure 2.22: Background hit rate as a function of instantaneous luminosity for W+1/RB4 region. Run 2 measured rate (red) is higher than Run 3 measured one (blue).

a key role for choices for future upgrade, reducing it with shields or changes, i.e. in the beam pipe, could improve the detectors longevity and give the possibility to have good performances at higher luminosities.

Chapter 3

Beyond Standard Model physics and Vector-Like Quarks

As anticipated in Chapter 1, the SM does not explain all the observed phenomena in high energy physics. Different theories propose several solutions to address the unresolved issues of the SM. The high energy reached and luminosity of the LHC allow the experiments to explore physics Beyond the Standard Model (BSM), searching for new particles predicted by theories, validating or refusing them. Vector-Like Quarks (VLQs) are new particles, predicted by several theories, that aim to resolve the hierarchy problem and achieve naturalness. In this Chapter, an in depth description of three classes of BSM models will be provided, the Composite Higgs, the Extra Dimension models, and the Two-Higgs-Doublet models. The Composite Higgs models propose that the Higgs is a composite state resulting from a new strong interaction. The dimensionality of the Higgs mass operator could be greater than 4, without requiring fine-tuning for the measured value of the Higgs mass. On the other hand, Extra Dimension theories predict the existence of dimensions beyond the typical space-time (3+1), addressing both the hierarchy problem and the unification of fundamental interactions. In both scenarios, Vector-Like Quarks (VLQs) play an important role. These VLQs, also referred to as $top-partners$ or $beuty-partners$, are colored fermions designed, that avoid the divergences of the Higgs mass corrections. The Two-Higgs-Doublet models are an extension of the SM, introducing another Higgs doublet and allowing the existence of more than one Higgs boson and new particles such as the VLQs.

3.1 The Composite Theories

The Composite Higgs Model [41] has been proposed by Georgi and Kaplan in 1984. In this theory the Higgs boson is a composite pseudo Nambu-Goldston Boson (pNGB), therefore it is a bound state of a new strong interaction. At scale $f > v$, where v in the vacuum expectation value, the breaking $SO(5) \to SO(4)$ is associated the pNGB, that in this case is the Higgs particle [42, 43]. In this model, the vector Ψ_L of $SO(5)$ is considered as an extension of the SM left handed doublet q_L for the third generation of quarks and the complete fermionic spinor of the third quark generation is:

$$\Psi_L = \left(q = \begin{pmatrix} t \\ b \end{pmatrix}, X = \begin{pmatrix} X^{5/3} \\ X \end{pmatrix}, T\right)_L, \tag{3.1}$$

$$\Psi_R = \left(t, X = \begin{pmatrix} X^{5/3} \\ X \end{pmatrix}, T\right)_R. \tag{3.2}$$

The SM gauge group $G_{SM} = SU(2)_L \times U(1)$ is here given by $SU(2)_L$ and the σ_3 of $SU(2)_R$ of the subgroup $SO(4) = SU(2)_L \times SU(2)_R \subset SO(5)$, since $SO(5)$ breaks up as $(2,2) + 1$ under a $SU(2)_L \times SU(2)_R$ transformation. The right handed states in Equation 3.2 have been introduced to give mass to the new fermions and their hypercharges have been fixed to obtain the correct electric charges. The Yukawa Lagrangian of the fermion sector is made up of an $SO(5)$ symmetric mass term for the top and the most general gauge invariant mass terms for the VLQ X and T:

$$\mathcal{L}_{top} = \lambda_1 \bar{\psi}_L \phi t_R + \lambda_2 f \bar{T}_L T_R + \lambda_3 f \bar{T}_L t_R + M_X \bar{X}_L X_R + h.c. \tag{3.3}$$

where λ_i with $i = 1, 2, 3$ are the coupling constants, ϕ is the scalar quintuplet containing the Higgs Field and M_X is the mass of the heavy X quark. The absence of quadratic divergences to m_H is guaranteed by the symmetry of the mass term.
Since

$$\phi = \begin{pmatrix} H \\ H^c \end{pmatrix}, \tag{3.4}$$

the Equation 3.3 can be written as:

$$\mathcal{L}_{top} = \bar{q}_L H^c(\lambda_t t_R + \lambda_T T_R) + \bar{X}_L H(\lambda_t t_R + \lambda_T T_R) + M_T \bar{T}_L T_R + M_X \bar{X}_L X_R + h.c. \,. \tag{3.5}$$

The diagonalization of the mass matrix allow to obtain the physical field. The introduction of the VLQs ensure the absence of quadratically divergent terms.
Considering the potential

$$V = \lambda(\phi^2 - f^2)^2 - A^2 f^2 \vec{\phi}^2 + B f^2 \phi_5, \tag{3.6}$$

where $\vec{\phi}$ are the first four components of ϕ, the Higgs boson mass is controlled by the A parameter, that is by the $SO(5)$-breaking term, $m_H = 2v\sqrt{A}$ for big λ. The Higgs boson would be a massless Goldstone boson and the last relation is fine, since the symmetry under $SO(5)$ is not broken. Setting $v = 0$, the divergent part of the one loop correction to A could be written as:

$$\begin{aligned} \delta A &= -\frac{12 f^2}{64\pi^2}\lambda^2 \left(\frac{M_X^2}{f^2} - 4(\lambda_1 + \lambda_3) - 2\lambda_2^2 \right) log\lambda^2 = \\ &= -\frac{3}{16\pi^2 f^2}(\lambda_t^2 + \lambda_T^2)\left(M_X^2 + M_T^2 \left(\frac{2}{1 + \lambda_T^2/\lambda_t^2} \right) \right) log\lambda^2. \end{aligned} \tag{3.7}$$

Any quadratic divergence is present and the two masses of the VLQs, M_X and M_T, might be interpreted as the cut-off Λ in the original top-loop contribution. Nevertheless, these two masses should not significantly exceed 2 TeV. In this case it could lead to a logarithmic term generating a δm_H roughly equivalent to the weak-scale expectation value v, thereby failing to resolve the naturalness problem. This simplified model introduces the VLQs [44], but often models like this fail to explain the fermion masses origin or they don't pass the electroweak precision tests [45].
The prediction of Vector-Like Quarks (VLQs) in the Composite scenario [46] is not exclusive to Composite Higgs theories. Similarly, Composite Top theories propose that the top quark is a composite state. In these theories, Standard Model particles acquire their masses through mixing with composite states, thanks to a new strong sector. Given its large mass, the top quark may exhibit characteristics of compositeness, showing a significant mixture with the composite state. Electroweak precision tests also different models involving right-handed composite top quark scenarios.

3.2 The Extra Dimensions Theories

The Extra Dimension theories have been formulated to explain the gravitational interaction and achieve the unification of all the fundamental forces. The first theory that introduce a fifth dimension, extending beyond the four dimensions of space-time, to unify gravity and electromagnetism was developed by Kaluza and Klein. Kaluza extended classical general relativity to include five dimensions [47]. Afterwards, starting from Kaluza's work, Klein extended this theory into a quantum field theory [48]. The main idea in Klein's theory is that the fifth dimension is curled and its geometry has the form of a circle of 10^{-33} meters radius.
Due to the weakness of the gravity in relation to the other fundamental forces, Arkhani-Hamed, Dimopoulos, and Dvally (ADD) introduced an Extra Dimension theory [49]. According to this theory, the additional dimension should possess a scale ranging from a millimeter to approximately $\frac{1}{TeV}$ in natural units in which the gravity could propagate explaining its weakness.
Lisa Randall and Raman Sundrum, in 1999, proposed two distinct models, aiming at addressing the hierarchy problem of the Standard Model [50, 51]. In these models, the universe is postulated to have five dimensions with a warped geometry. In the first model (RS1), the extra dimension possesses a finite size, consisting of two branes connected to each other. The second model (RS2) foresees a similar spacetime structure, but one of the two branes is positioned significantly far from the other. By approximating it to be infinitely distant, the model effectively reduces to having only one remaining

brane. The two branes are the Planck brane and the TeV brane. Since in these models the spacetime is extremely warped, the Plank brane has positive brane energy, where the gravity is a relatively strong force, while the TeV brane has negative brane energy, where there are the SM particles.
Several theories based on these fundamental models have been developed: the theories which follow the ADD model are called "The Large Extra Dimension Theory"; the ones that get inspiration from the RS models are called "The Warped Dimensions Theory".

3.3 Two-Higgs-Doublet Theories

The Two-Higgs-Doublet Model (2HDM) [52] is an extension of the SM. The model proposes two distinct Higgs doublets, leading to the possibility of new particles, known as Higgs bosons. Different types of 2HDM, such as Type I, Type II, Type III, and Type IV, vary in the way particles interact with the Higgs doublets. The Higgs sector is much more complex than the SM one and it propose the existence of an additional neutral scalar H^0, a pseudoscalar A and a charged boson $H^{\pm}$. Some of these model introduce a Vector-Like top partner, allowing flavour-changing neutral currents [53]. Due to the mixing of the top quark and the VLQ, the mass of these particle can be very different and VLQ masses around 1 TeV are allowed.

3.4 Search for Vector-Like Quarks

The colored fermions, known as Vector-Like Quarks (VLQs), show the same transformation under $SU(2)_L$ for both left and right chiral components. This symmetry allows the introduction of a mass term in the lagrangian, such as $-m\bar{\psi}\psi$, without breaking the $SU(2)_L$ symmetry, given that the weak charged current is a vectorial current. Despite this theoretical framework, there is currently no experimental evidence supporting the existence of VLQs. As already explained, the introduction of Vector-Like Quarks can stabilize the mass of the Higgs boson. Additionally, due to sharing the same colored charge as Standard Model quarks, they represent the most straightforward introduction of new colored fermions beyond the SM. Despite this, their coupling with the Higgs field is not related with their masses, since the introduction of the mass term is different from the SM fermions. The mixing of VLQs with SM quarks modifies their coucoupling to the Z, W, and Higgs bosons and breaks the GIM mechanism. Flavour changing neutral currents are therefore allowed at the tree-level in the theories that predict their existence.
In the Table 3.1 different kind of VLQs are presented, depending on their electric charges. The VLQs

VLQ	**Electric charge**
X	+5/3
T	+2/3
B	-1/3
Y	-4/3

Table 3.1: Charge assignment for VLQs.

are usually grouped into multiplets of $SU(2)$, in particular singlets, doublets, or triplets as can be seen in the Table 3.2. Considering the different multiplets, the VLQs can be therefore represented as:

$$\begin{array}{ll} \textit{Singlets} & T^0_{L,R}, B^0_{L,R} \\ \textit{Doublets} & \left(XT^0\right)_{L,R}, \left(T^0B^0\right)_{L,R}, \left(B^0Y\right)_{L,R} \\ \textit{Triplets} & \left(XT^0B^0\right)_{L,R}, \left(T^0B^0Y\right)_{L,R}, \end{array}$$

where the subscripts L and R indicate the left- and right-handed components, respectively, and the superscript 0 denotes weak eigenstates. The weak and mass eigenstates for X and Y coincide due to their exotic charges, preventing them from mixing with SM quarks. By introducing the new fields $T^0_{L,R}$ to SM, the up-type eigenstates (u, c, t, T) can incorporate components of these fields, modifying the coupling of quarks to the Z boson. However, the degree of mixing is proportional to the ratio between the masses of the SM quarks and the VLQs. Consequently, VLQs are considered to mix with third-generation quarks, as the ratio m_Q/m_{VLQ} is negligible for the first two generations. The top quark,

Multiplet	Hypercharge
Singlets	
(T)	+2/3
(B)	-1/3
Doublets	
(X,T)	+7/6
(T,B)	+1/6
(B,Y)	-5/6
Triplets	
(X,T,B)	+2/3
(T,B,Y)	-1/3

Table 3.2: Hypercharge assignment for Vector-Like quarks in different $SU(2)$ representations.

due to its large Yukawa coupling and the these considerations, becomes directly linked to new physics associated with electroweak symmetry breaking and the fermion mass hierarchy. The signatures of VLQs have been analyzed both in model independent and in specific model-dependent scenarios. For the latter case, the possibility of flavour changing neutral currents processes for VLQs interactions lead to a wide range of possible final states, being sought for at the LHC. In the next sections it will be described the possible VLQs production mechanism, i.e. pair and single production, and their decay channels in details. Both the TeVatron and the LHC have conducted searches for VLQs. The primary investigations at the TeVatron took place during Run 2, operating at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV with a total integrated luminosity of 5.4 fb^{-1}, with the CDF and D0 experiments. At the LHC, studies have been done at a center-of-mass energy of $\sqrt{s} = 13$ TeV ($\sqrt{s} = 7 - 8$ TeV) by the ATLAS [54, 55] and CMS collaborations [56, 2], during Run 2 (Run 1). They will continue for Run 2, as well as for Run 3, with a $\sqrt{s} = 13.6$ TeV. The searches have been focused both on the pair production, which is model-independent, and on single production channels. The latter are model depended, but they allow to reach high sensitivity for larger VLQ masses hypotheses. The analysis strategies have necessitated the exploration of various final states, including all-hadronic searches, single- or multi-lepton final states, with or without transverse missing energy. In the Figure 1.1 are shown the results of the searches for the single and the pair production of VLQs by CMS in August 2023.

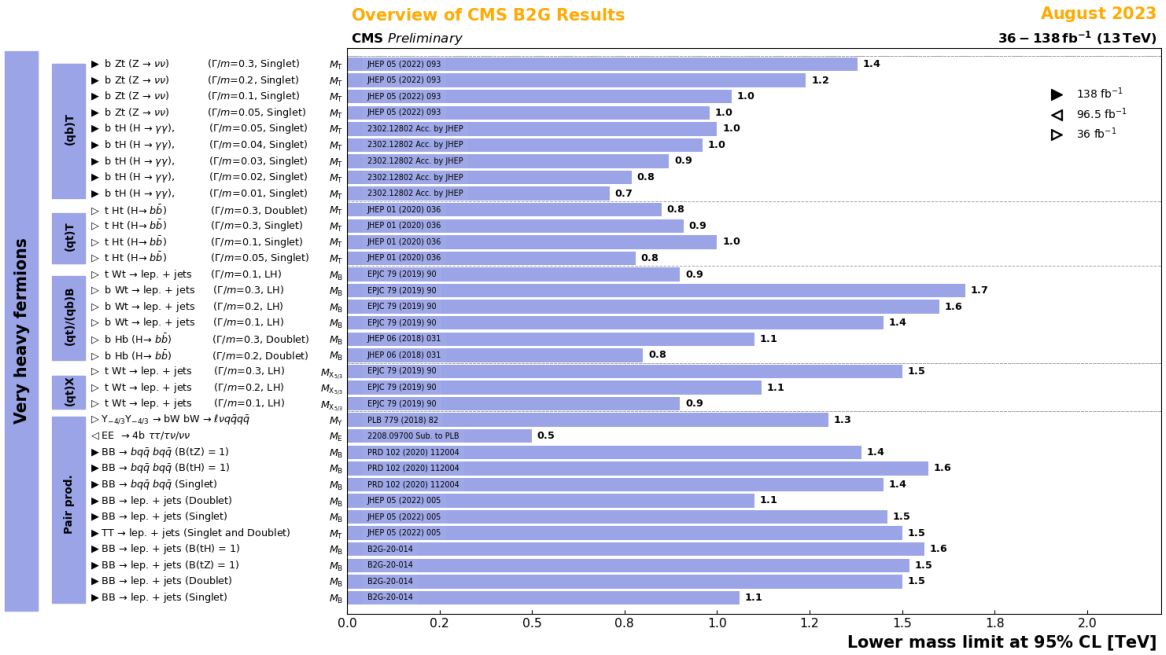

Figure 3.1: Summary of the CMS results of the VLQs.

3.4.1 VLQs production mechanisms

In pp collisions, the production mechanisms can be categorized into:

- the single production, via electroweak interaction processes. The single production process depends on the fermion mass, on the mixing parameters with SM particles and on the couplings between the new quarks and the W and Z bosons:

$$qq' \xrightarrow{V^*} qQ \quad V = W, Z$$

Due to the small masses of the light quarks, the contributions of the Higgs bosons are always suppressed.
In Figure 3.2 is shown the single electroweak interaction mediated by a vector boson in association with a SM quark and the one mediated by SM quark in association with a vector boson.

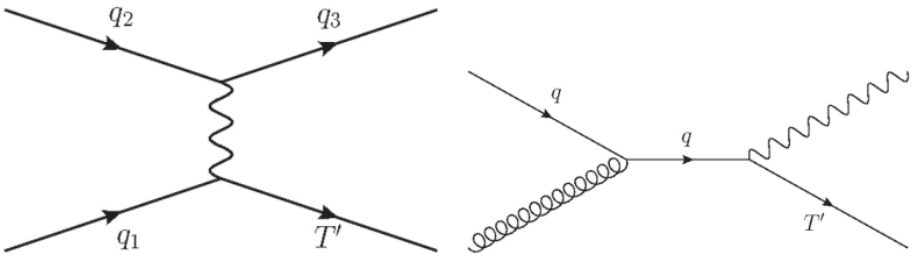

Figure 3.2: Feynman diagrams for single production of Vector-Like top quarks.

- The pair production is dominated by QCD processes. Due to the PDFs suppression at high energy, the cross section of this process steeply decreases for higher masses of the new fermion, with a much more significant slope with respect to the single-production case, as more energy is necessary to produce both particles. Moreover the cross section only depends on the mass of the VLQs, since it is model independent, and these processes are similar to the SM quarks pair

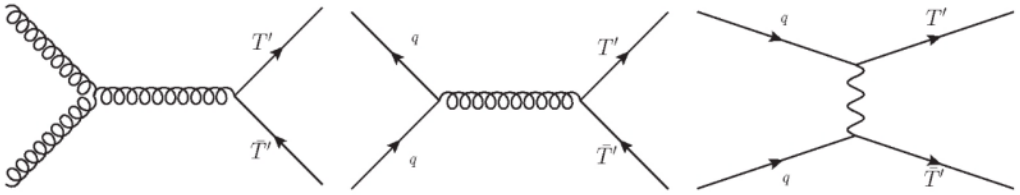

Figure 3.3: Feynman diagrams for pair production of vector-Like top quarks via gluon and W,Z or Higgs bosons.

production processes:

$$gg, q\bar{q} \to Q\bar{Q}$$

where $Q = T, B, X, Y$.
However other pair production processes are strongly suppressed, such as the electroweak charged and neutral processes:

$$\bar{q}q' \to W^+ \to \bar{T}X, \bar{B}T, \bar{Y}B \tag{3.8}$$

$$\bar{q}q' \to W^- \to T\bar{X}, B\bar{T}, Y\bar{B}, \tag{3.9}$$

$$\bar{q}q \to Z \to Q\bar{Q}, \tag{3.10}$$

or with the Higgs boson

$$\bar{q}q \to H \to Q\bar{Q}. \tag{3.11}$$

The cross section of these processes is model-dependent. In Figure are shown some Feynman diagrams for pair production.

The dependence of the pair production and the single production cross sections to the energy is illustrated in Figure 3.4, production cross sections decreas as the mass of the VLQ increase. However, the decreasing slope is different for the two production mechanisms and the plot shows that pair production processes, for the specific model considered, are the dominant processes for masses below $m_Q \sim 800 - 1000$ GeV, while they becomes less important for higher masses due to their phase-space suppression [57].

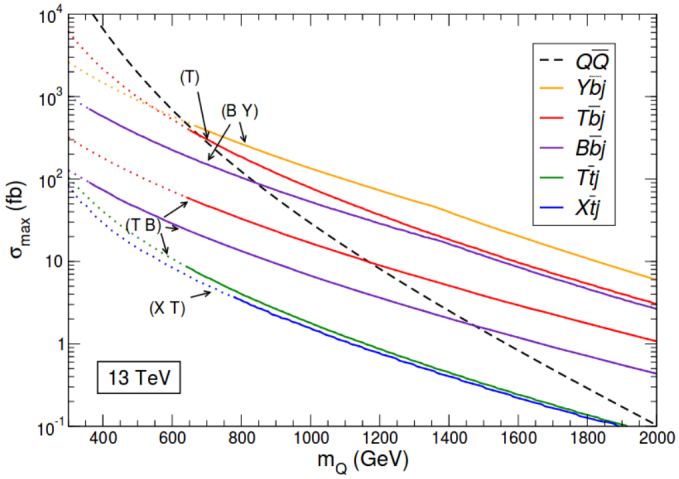

Figure 3.4: Production cross section for Vector-Like Quarks in pp collisions at $\sqrt{s} = 13$ TeV as a function of their mass, for pair production and for single production in different channels. The black dashed line represents VLQ pair production, while the colored lines represent the singly produced VLQs. The dashed coloured lines correspond to the values of cross section excluded by previous studies.

3.4.2 VLQs decay channels

Since the VLQs break the GIM mechanism, they can decay in electroweak charged and neutral interactions into SM quarks or other VLQs. Considering the decay channels into SM particles, the main allowed channels are:

$$T \to W^+b, Zt, Ht$$
$$B \to W^-t, Zb, Hb$$
$$X \to W^+t$$
$$Y \to W^-b.$$

However the branching ratio of T and B are different in the three scenarios of multiplets. For the isospin singlets T and B, all three decay are possible and the branching ratios depend on the VLQ mass and are not inter-depended:

$$Br\left(Q \to Wq'\right) + Br\left(Q \to Zq\right) + Br\left(Q \to Hq\right) = 1$$

with $(Q, q, q') = (T, t, b), (B, b, t)$.
In the case of doublets and triplets, the slight mass difference plays a role in suppressing the decay into other VLQs, thereby allowing decays into vector bosons and Higgs bosons with top or beauty quarks. The decay patterns of the doublet (T, B) are influenced by the mixing factor in the extended CKM matrix, denoted as V_{Tb} and V_{tB}. If $V_{Tb} \sim V_{tB}$ is true, it indicates that the T and B quarks exhibit the same decay characteristics as their corresponding singlets. Despite having different angular distributions due to only the right-handed component of (T, B) coupling to SM quarks, constraints on b quark mixing and mass hierarchy, where $m_t \gg m_b$ and $V_{Tb} \ll V_{tB}$, lead to much stronger mixing of the heavy quarks with the SM top quark. Consequently, the decays $T \to Wb$, $B \to Hb$, and $B \to Zb$ are suppressed. The possible decays of Vector-Likequarks are reported in Tables 3.3, 3.4 and 3.5. On the other hand, the branching ratios of the Vector-Like Quarks are model-dependent and they also depend on the heavy quark mass themselves.

Singlets	**Decay modes**
X	W^+t
T	W^+b, Ht, Zt
B	W^-t, Hb, Zb
Y	W^-b

Table 3.3: Allowed decay modes for Vector-Like singlets.

Doublets	**Decay modes**
$\begin{pmatrix} X \\ T \end{pmatrix}$	W^+t Ht, Zt
$\begin{pmatrix} T \\ B \end{pmatrix}$	Ht, Zt W^-t
$\begin{pmatrix} B \\ Y \end{pmatrix}$	Hb, Zb W^-b

Table 3.4: Allowed decay modes for Vector-Like doublets.

Triplets	Decay modes
$\begin{pmatrix} X \\ T \\ B \end{pmatrix}$	W^+t W^+b, Ht, Zt Hb, Zb
$\begin{pmatrix} T \\ B \\ Y \end{pmatrix}$	Ht, Zt W^-t, Hb, Zb W^-b

Table 3.5: Allowed decay modes for Vector-Like triplets.

Moreover, decay channels in non SM objects are allowed by different models. A new scalar or pseudoscalar boson A (denoted also as S) is predicted and its mass range goes from tens to hundreds of GeV. This 2HDM theory [53] allow the existence of a new particle A together with singlet representation of the SM electroweak group. In the simplified model the branching ratio of the non SM decay channel is almost 100% at scale of TeV, as it increases with the mass of the VLQ. Figure 3.5 shows the branching ratio of the VLQ T in function of the mass for the simplified model.

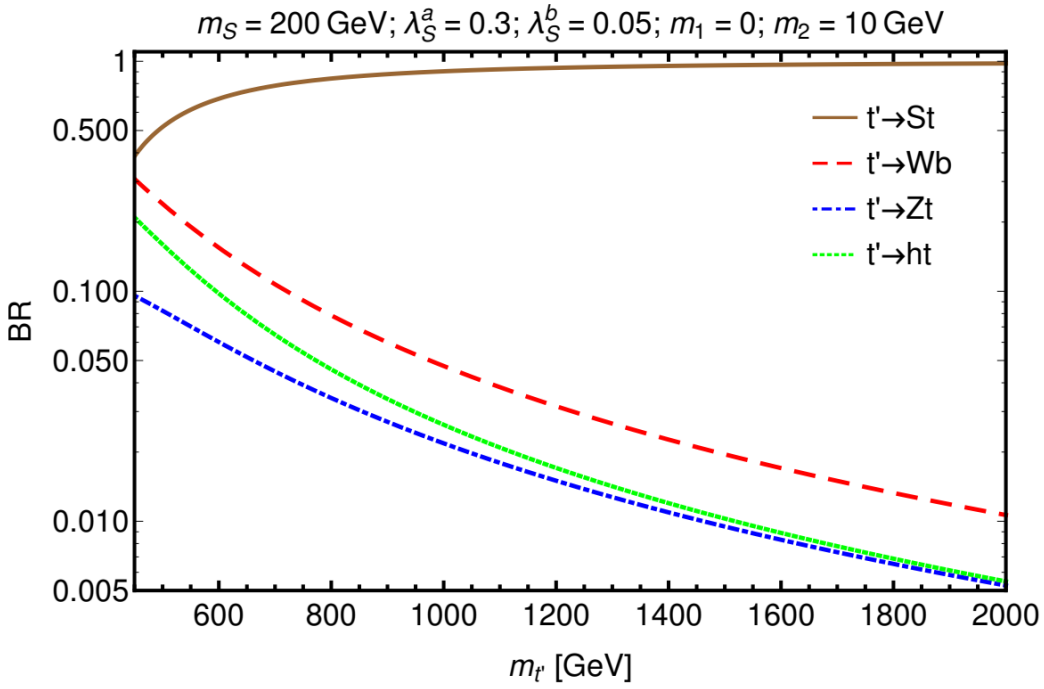

Figure 3.5: Branching ratios of VLQ T as a function of its mass. In this simplified model m_S is the mass of the scalar S (or A), while m_1 and m_2 are non-diagonal entries in the mass matrix, related to the top quark and VLQ T mixing.

The new boson A could decay in different way to SM object, partial widths of the different decay channel are shown in Figure 3.6.

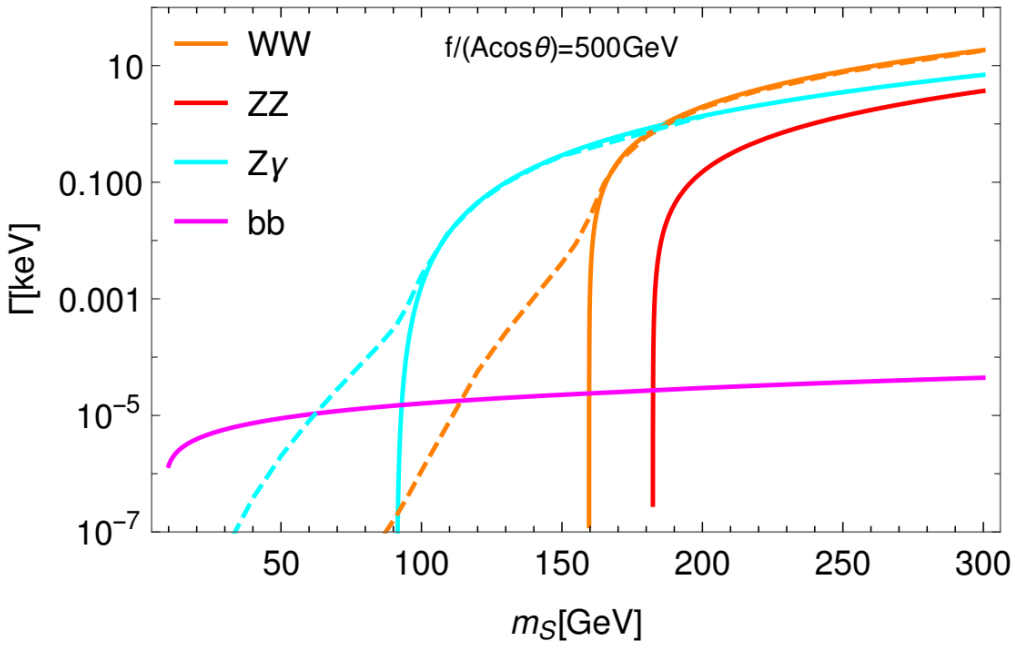

Figure 3.6: Partial width of A in the dominant decay channels for the composite Higgs model benchmark scenario.

Chapter 4

Physics object selection and reconstruction

The main goal of this work is to search for singly-produced VLQ T, decaying to a top quark and a Higgs boson or new scalar boson A, with yet unknown mass. It is shown in Chapter 3 that at low, yet unexplored masses, on the most important decay channel is through a beauty quark-antiquark pair. For this reason, this analysis focuses on the hadronic decay of the boson, into a beauty quark-antiquark pair, and on the leptonic decay of the top quark. The search is performed using the pp collision data collected by the CMS experiment during the Run 2 (2016-2018), with an integrated luminosity of 138 fb^{-1} and a centre-of-mass energy of 13 TeV. Figure 4.1 shows the feynman diagram of the process of interest. The final state is characterized by objects reconstructed in the CMS detector using the ParticleFlow algorithm.

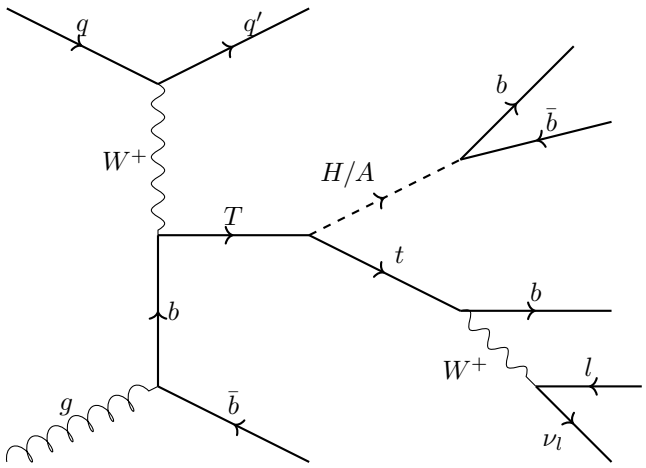

Figure 4.1: VLQ T single production, decaying to a top quark and Higgs or A boson in the semileptonic final state.

4.1 Physics Object Selection

As previously described, the CMS detector has a cylindrical symmetry along the beam axis,and has a layered structure of subsystems centered around the nominal interaction points, and interspersed so to be as hermetic as possible in the solid angle. This allows to get information about particles crossing the various subsystems, emerging from the nominal interaction point. These subsystems, using different diverse technologies with specific tasks and characteristics, contribute to the reconstruction strategy. Each subdetector is employed to achieve a complete and precise reconstruction of all particles. The tracker system, for instance, allows to reconstruct the charged-particle trajectories and interaction vertexes. Thanks to the magnetic field, it also determines charged-particle momenta by measuring the

bending of the trajectories in the x-y plane. The calorimeters, instead, focus on measuring particle energy. The ECAL primarily absorbs electrons and photons, while the HCAL is designed to absorb jets resulting from the hadronization of quarks, gluons, baryons, and mesons. The muon system is dedicated to measuring muon trajectories and momenta. Muons, passing through previously mentioned subdetectors without significant energy loss, need a specialized detection system. However, the neutrinos interact minimally with any subsystem and in the majority of cases escape undetected. To improve the event reconstruction, the ParticleFlow (PF) algorithm [58] links the basic elements, such as tracks and cluster, from all the subdetectors to identify each final state particle. By combining corresponding measurements, the particle properties are reconstructed based on this identification. The integration of measurements from the tracker, calorimeters, and the muon system allows for a global fit to physics objects, resulting in high efficiencies and resolutions. The final state objects, needed for our search, will be described in the next sections, as well as, the top quark reconstruction and selection.

4.2 Muons

The muon spectrometer efficiently identifies and reconstructs muon tracks with high efficiency across the entire detector acceptance. This results from the fact that the calorimeters absorb nearly all particles, excluding muons and neutrinos. Three distinct types of muon candidates can be classified based on their reconstruction method:

- **Standalone muon:** Initiated by hits from DT and CSC, tracks are reconstructed only using the muon spectrometer. This process collects hits from all muon subsystems through a Kalman filter (KF) procedure [59].
- **Tracker muon:** Tracks with transverse momentum greater than 0.5 GeV and a total momentum exceeding 2.5 GeV are extended to the muon system. If a match with muon system hits is identified, the track is categorized as a tracker muon.
- **Global muon:** A standalone muon is matched with information from the tracker, and a global fit to the resulting track is performed.

Global muons show the highest reconstruction efficiency when their momentum is greater than 10 GeV, indicating hits in at least two muon stations. Below this threshold, there is an increased likelihood of muons experiencing multiple scattering in the iron of the return yoke. Consequently, the tracker muon demonstrates higher efficiency. Standalone muons, reconstructed independently, may exhibit contamination from cosmic muons reaching the cavern. Additionally, charged hadrons can be misidentified as muons if their shower is not entirely blocked in the calorimeters. In the PF algorithm, muons are reconstructed by using information from muon energy deposits in ECAL and HCAL, to improve identification performance. Two different working points, Loose and Tight, are defined depending on additional quality criteria. These WPs are used to select final state muons of interest. The Loose WP only needs the muon to be reconstructed by the PF algorithm as a global muon. For the Tight WP, the muon is required to undergo reconstruction using the Particle-Flow (PF) algorithm as a global muon. The muon track must be fitted with a $\chi^2/\text{ndof} < 10$. Moreover, it must possess at least one muon-chamber hit included in the global-muon track fit and cross at least two muon stations. The tracker muon associated with the global muon is expected to have a transverse impact parameter $dxy < 2$ mm and a longitudinal impact parameter $dz < 5$ mm respect to the primary vertex of the event. These criteria aim to suppress cosmic muons, muons originating from decays in flight, and tracks from pileup (PU). Furthermore, the track is required to have at least one hit in the pixel detector and a minimum of 5 hits in the entire tracker system. In order to distinguish between prompt muons and the ones coming from secondary decay chains, usually included into a jet, an isolation algorithm is applied. A cone in ΔR is defined around the direction of the muon, and the isolation is defined as:

$$I^{\ell}_{rel} = \frac{\sum p_T(\text{ch.had from PV}) + max(0, \sum E_T(\text{neut.had}) + \sum E_T(\text{phot}) - 0.5 * \sum p_T(\text{ch.had from PU}))}{p_T(\ell)} \tag{4.1}$$

The sums in the above expression are performed by considering only PF candidates in the isolation cone. For standard isolation, the size of the cone is fixed to 0.4 for muons. A MiniIsolation algorithm

has been also defined to recover the efficiency when leptons are produced in the decay chain of a boosted object. As in the case of muons coming from top quark decay, in particular when the top quark Lorentz boost is sufficiently high, can overlap with a jet produced in the same decay chain. It is constructed by defining a cone whose radius varies with the lepton p_T. The radius of the cone is:

- 0.2 for leptons with $p_T \leq 50$ GeV,
- 10 GeV/p_T for leptons with $50 < p_T < 200$ GeV,
- 0.05 for leptons with $p_T \geq 200$ GeV.

Thus the isolation is derived by using a specific cone for each lepton, reducing the probability of the overlap for more boosted topologies. Both standard isolation and MiniIsolation are used, as described below, as input for the tog tagging machine learning algorithm.

4.3 Electrons

The PF algorithm reconstructs electrons by utilizing measurements from both the tracker and ECAL. As the electrons cross the tracker, they can emit bremsstrahlung photons, which subsequently transform into electron-positron pairs. Consequently, the fundamental properties and technical challenges for tracking and energy deposition patterns of electrons and photons shows similarities. When the energy radiated by the electron is minimal, the tracker can reconstruct its track with a good χ^2, extending the track to the inner ECAL surface for accurate alignment with the closest cluster. However, when the radiated energy is substantial, this becomes much more difficult, resulting in a reconstructed track with few hist and an elevated χ^2 from the track fitting. The tracks are preselected depending on the number of hits and χ^2 value from the fit and the selected tracks are fitted again with a Gaussian-sum filter (GSF) [60]. The latter shows better performances for electron reconstruction compared to the KF, taking into account the possibility of sudden and significant energy losses along the trajectory. The electron track candidate is then aligned with a cluster in the ECAL. After the electron reconstruction, all the matched tracks and clusters are masked to improve other object reconstruction. Additional quality criteria are used to select electron in the final state. Different boosted decision trees (BDT) algorithms have been trained with up to fourteen variables to discriminate prompt electrons from background. The BDTs are trained separately in ECAL barrel and endcap acceptance, and for low and high p_T electrons (with a threshold of 10 GeV) for a total of six trainings. Two different version of the algorithm are developed including in the input features isolation variable (IsoID) or not including it (NoIsoID). In this analysis are used both Iso and NoIso IDs, in particular the working points WP80 and WP90, respectively with 80% and 90% signal efficiency. Both IsoID and NoIsoID will be used in the analysis strategy, as described in the next chapter.

4.4 Jets

Quarks and gluons, generated in the final state of intense proton-proton interactions at the LHC, undergo the process of hadronization before they can directly interact with the detectors. They can be indirectly detected through the formation of jets, which are clusters of color-neutral hadrons. The interaction between the partons and the showering can be described by perturbative theory and hadronization models. In the CMS detector, the jets are reconstructed starting from the PF particles with the anit-k_T algorithm [61], which is infrared and collinear safe avoiding discrepancy between data and MonteCarlo simulation due to the jet reconstruction. This algorithm is the default choice at LHC, thanks of its robustness against pile up and underlying event contribution, as well as providing conical jets. The two distances, d_{ij} and d_{iB}, are considered, being the first the distance between the particle i and the jet candidate j, while the second between the particle i and the beam B. These two are defined as:

$$d_{ij} = min(p_{Ti,}^{-22}, p_{T,j}^{-2}) \frac{(\Delta R_{ij})^2}{R^2}, d_{iB} = p_{T,i}^{-2} \tag{4.2}$$

where $\Delta R_{i,j} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$ and $p_{T,i}$, η_i, and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i. If d_{ij} is less than d_{iB}, the two object are combined together, otherwise i is called jet. The distances are recalculated and the procedure repeated until no entities

are left and the final jet is reconstructed. The R parameter defines the size of the jet in the (η, ϕ) space, in CMS, two distinct type of jet are reconstructed, the "narrow jet" and the "fat jet", respectively with R equal to 0.4 and 0.8. These jets are also called AK4 and AK8. Reconstructed PF charged hadrons, associated to pile up interactions, are removed to the list of PF candidates for jet clustering, this is called pile up charged-hadron subtraction or CHS. This procedure is used to cluster the AK4 jets. For the AK8 jets, instead, is used the Pile Up Per Particle Identification (PUPPI) technique [62]. Jet Energy Corrections (JECs) and Jet Energy Resolutions (JERs) are applied to the transverse momentum in order to relate the measured energy to the true energy of the particles originating the jet. In our search, both AK4 and AK8 are required to pass a tight ID criteria [63] and $|\eta| < 2.4$, in order to be used in the physics analysis. Moreover, only AK4 and AK8 with p_T respectively greater than 30 GeV and 500 GeV are selected. To reduce the contribution of jets coming from pile-up, AK4 jets with p_T lower than 50 GeV are required to pass a pile-up tight ID. In next paragraph, it will be described briefly two algorithms: the DeepJet and the ParticleNet. The first one is used to discriminate AK4 jet containing the decay of b hadron (b jets), originated by a beauty quark. This have a key role as it is used as input variables for the machine learning technique developed for top tagging. The second one is used to tag AK8 originated from the decay of the A or Higgs boson into a quark-antiquark beauty pair.

4.4.1 The DeepJet Tagger

The b jet are characterized to have a features that make it possible to distinguish them from jet produced by hadronization of light quarks and gluons. The b jets have longer lifetime, large masses and high track multiplicity. Due to the long lifetime a Secondary Vertex (SV) can be reconstructed and the Impact Parameter (IP) is defined to measure the distance between the primary vertex and tracks of the jet constituents. The DeepJet algorithm [64] makes use fo a wide set of jet information, such as the SV features, global variables of the jet, and charged and neutral PF candidates, used for jet clustering, features. The algorithm uses a deep neural network architecture with convolutional layers and fully connected layers. For each candidate it provides the probability for the jet to contain a (or multiple) B-hadron, one C-hadron, or coming from light quark or gluon.

4.4.2 The ParticleNet Tagger

The ParticleNet algorithm [65] is a dynamic graph convolutional neural network. The AK8 jet are represented as an unordered set of particles, permutation invariant. The jet is seen as a particle cloud, linking together all the advantages of particle based representation and point cloud representation of 3D shapes. It gives the probability of belonging to a specific category of AK8 jet, such as jets originated from QCD processes, hadronically decaying top quark, or from a generic heavy boson into a beauty quark-antiquark pair. Different version have been trained, in this work it is used the $Mass\ Decorellated$ one, to be large decorrelated with the mass of the jet. In particular the $XbbVsQCD$ tagger is a good discriminant to tag AK8 jet of interest from the QCD processes background jets. The tagger is defined as:

$$XbbVsQCD = \frac{P(Xbb)}{P(Xbb) + P(QCD)}, \tag{4.3}$$

where P(bb) is the probability of coming from a particle X decaying into a quark-antiquark beauty pair, while P(QCD) coming from a light quark or gluon.

4.5 Missing Transverse Energy

At the LHC the Z-component of the colliding proton momentum is known, while the X and Y components are zero. Momentum conservation in the X-Y plane enables the determination of the momentum lost due to particles escaping the detector. Along the Z-axis, instead,the fraction of momentum lost due to proton fragments ending up in the beam pipe is not measurable. The missing transverse energy (MET) is reconstructed by the particle flow algorithm, utilizing the vectorial sum of the transverse momentum of all PF candidates:

$$\vec{p}_T^{\,miss} = -\sum_{i}^{all\ PF} \vec{p}_{T,i} \tag{4.4}$$

However, MET is highly sensitive to detector effects impacting kinematic closure, such as miscalibrations or inactive/damaged detector regions. Similar to jets, energy corrections, notably the JEC, are applied to MET. In this thesis, to account for the presence of neutrinos in top quark decay, a requirement is set that MET must be greater than 25 GeV. Moreover, MET is a crucial variable for exploring Beyond Standard Model physics, as many models predict particles that leave no trace in the detector. The presence of non-interacting particles in the detector can be quantified through MET, as they result in an energy imbalance.

4.6 Top Reconstruction and selection with Machine Learning techniques

For each event, top quark candidates are reconstructed with leptons, jets and the MET. However, even for signal events combinatorial background could be high enough. In this section is described the top quark candidates reconstruction and the top quarks selection.

4.6.1 Top quark reconstruction

A top quark candidate is reconstructed from a lepton, the MET, and an AK4 jet. The longitudinal component of the neutrino four momentum ($p_{z,\nu}$) is estimated by imposing the W candidate mass equal to its nominal value of $m_W = 80.4\ GeV$, assuming that the x and y components of the MET are entirely associated to the escaping neutrino:

$$p_{x,\nu} = p_x^{miss}, p_{y,\nu} = p_y^{miss} \tag{4.5}$$

Solving the equation

$$m_W^2 = \left(E_\ell + \sqrt{{p_T^{miss}}^2 + p_{z,\nu}^2}\right)^2 - (\vec{p}_{T,\ell} + \vec{p}_T^{miss})^2 - (p_{z,\ell} + p_{z,\nu})^2, \tag{4.6}$$

imposing m_ℓ =0 GeV, there are in general two solutions:

$$p_{z,\nu} = \frac{\Lambda \cdot p_{z,\ell}}{p_{T,\ell}^2} \pm \sqrt{\frac{\Lambda^2 \cdot p_{z,\ell}^2}{p_{T,\ell}^4} - \frac{E_\ell^2 \cdot {p_T^{miss}}^2 - \Lambda^2}{p_{T,\ell}^2}}\ , \tag{4.7}$$

$$\Lambda = \frac{m_W^2}{2} + \vec{p_{T,\ell}} \cdot \vec{p_T^{miss}}. \tag{4.8}$$

In the analysis presented in this thesis, in case of two real solution for $p_{z,\nu}$, we choose the solution that minimizes the χ^2 of the reconstructed invariant mass of the W candidate.

If the m_W^T is larger than m_W the radicant in 4.7 is negative and the solutions have an imaginary component. The latter is eliminated by modifying $p_{x,\nu}$ and $p_{y,\nu}$ in order to have $m_W^T = m_W$. The couple $(p'_{x,\nu}, p'_{y,\nu})$ is chosen by minimizing the distance between $p'_{T,\nu}$ and p_T^{miss}, and then $p_{z,\nu}$ is obtained from 4.7.

A top quark candidate is selected only if the angular separation $\Delta R(l_{top}, jet_{top})$ between the lepton and the jet associated to the candidate is less than 2. Events with no top quark candidates satisfying this requirement are rejected.

The selected top quark candidates are then assigned to two different categories depending on the value of $\Delta R(l_{top}, jet_{top})$:

- if $\Delta R(l_{top}, jet_{top}) > 0.4$ the lepton is not inside the jet, and the top quark candidate is in the **Resolved configuration**, shown in Figure 4.2.

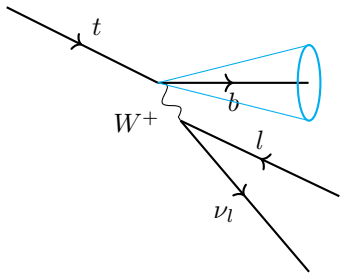

Figure 4.2: Top Resolved configuration.

- if $\Delta R(l_{top}, jet_{top}) < 0.4$ the lepton is inside the jet, and the top quark candidate is in the **Merged configuration**, shown in Figure 4.3.

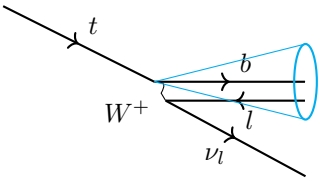

Figure 4.3: Top Merged configuration.

Since in the Merged configuration the lepton is likely to be included already in the jet constituents, the lepton four momentum is subtracted by the jet four momentum. In this configuration muons and electrons have been taken into account in the jet clustering respectively for 94% a 98% of the top quark candidates.

4.6.2 Top quark candidates classification

The result of top quark reconstruction is a collection of top quark candidates for each event, each referring to a lepton-jet pair. A top quark identification algorithm based on ML techniques is implemented in order to choose the top candidate in the event correctly associated to the top quark decay, and distinguishing it from the combinatorial background.

Different types of combinatorial background are identified by studying the MC truth matching of the lepton-jet pair of the top quark candidate. Thus, the reconstructed top quark candidates can be classified as:

- **True Top**, if both the lepton and the jet match with the MC truth lepton and b-quark from a top quark decay as shown in Figure 4.4.
- **Top-Jet Match**, if only the jet matches the MC truth b-quark from a top quark decay. This kind of combinatorial background is expected to be particularly important, especially in the Merged configuration, where the lepton from the top quark could have similar properties to the leptons coming from pions and kaons decay from B mesons decay chains, naturally present in all b-originated jets. In Figure 4.5 this kind of combinatorial background is shown.
- **Top-Lep Match**, if only the lepton matches the MC truth, and the jet could be a mistagged jet or a real b-jet coming from other processes, e.g., for signal events, it could be the other jet coming from $g \to b\bar{b}$ as shown in Figure 4.6.
- **False Top**, if neither the lepton nor the jet match the MC truth as shown in Figure 4.7.

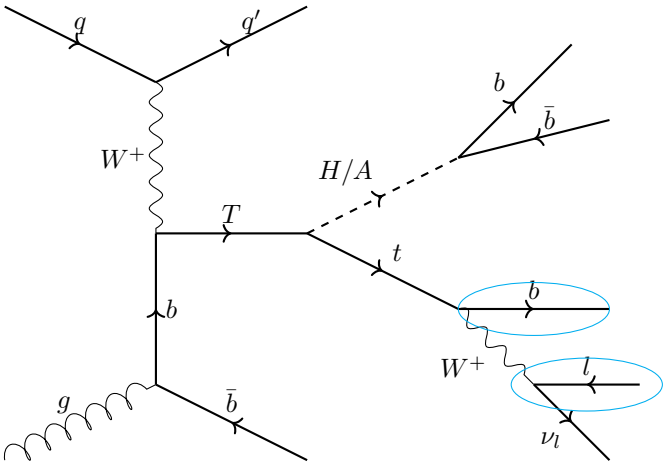

Figure 4.4: True Top configuration.

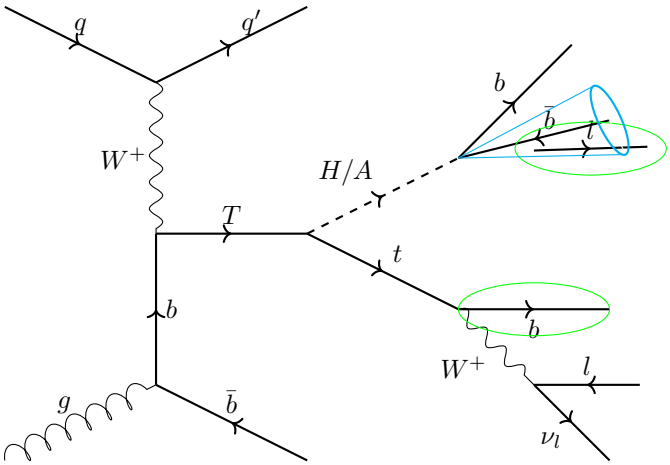

Figure 4.5: Top-Jet Match configuration.

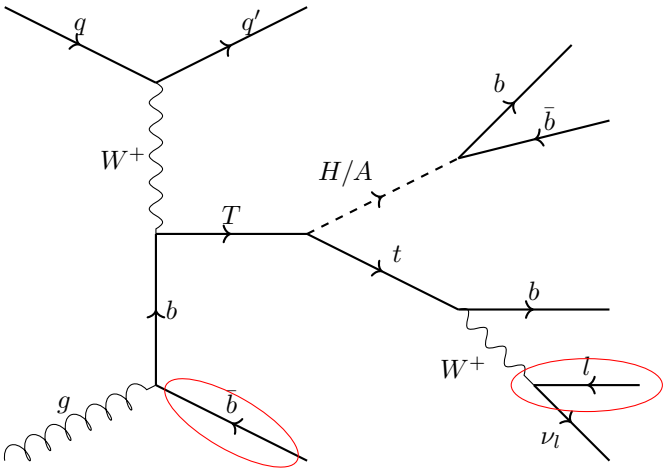

Figure 4.6: Top-Lep Match configuration.

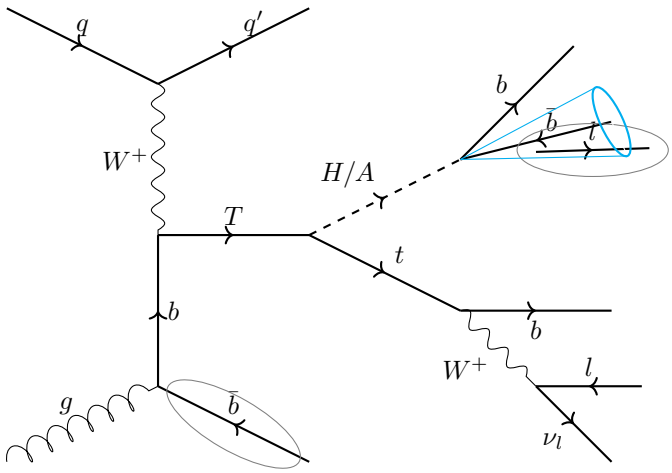

Figure 4.7: False Top configuration.

- **Top Comb**, if both the lepton and the jet match the MC truth, but they came from two different top quarks. This configuration is not expected in signal events but it is expected in background events such as $t\bar{t}$, as shown in Figure 4.8.

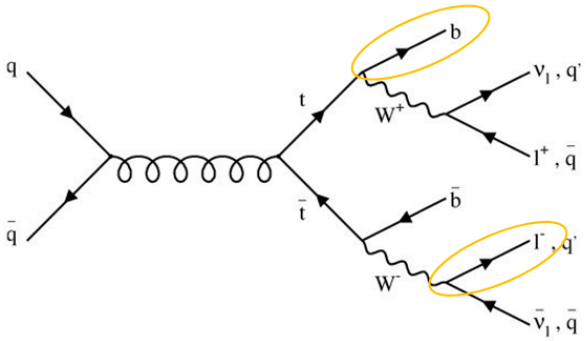

Figure 4.8: Top Comb configuration.

Three different categories of top quark candidates are defined:

- **True top**;
- **QCD-like top**, i.e. combinatorial background with unmatched lepton and unmatched jet, which is expected mainly from QCD multijet processes;
- **Other**, i.e. all other combinatorial backgrounds.

Figure 4.9 shows the reconstructed top quark mass distribution in the merged configuration, with an electron (right) or a muon (left). Distributions have been split for True (red), Other (green), and QCD-like (blue) top quark candidate. Figure 4.10 shows the reconstructed top quark mass distribution in the resolved configuration, with an electron (right) or a muon (left). Distributions have been split for True (red), Other (green), and QCD-like (blue) top quark candidate.

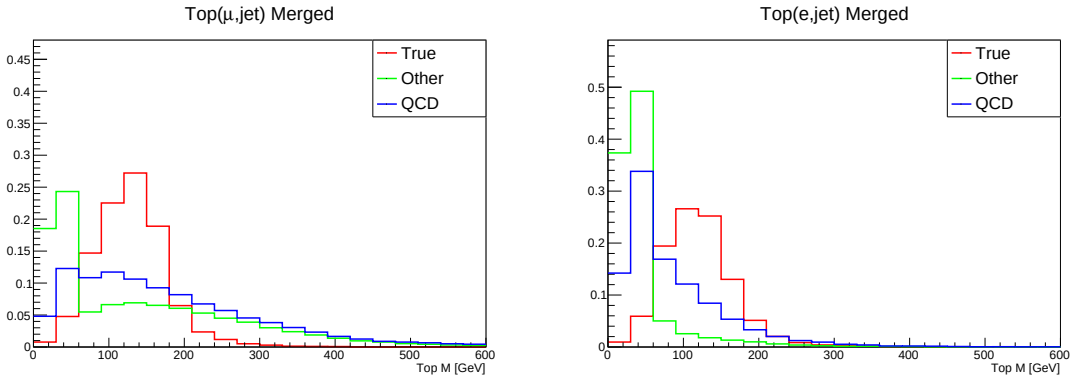

Figure 4.9: Merged top quark reconstructed mass distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

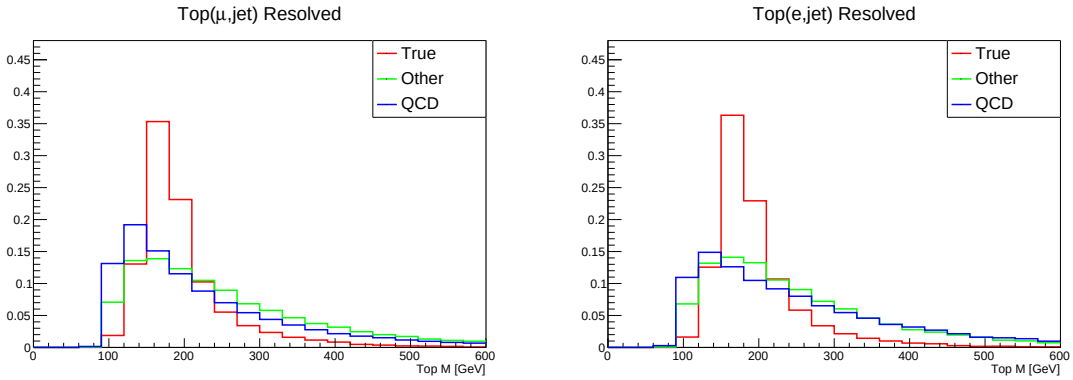

Figure 4.10: Resolved top quark reconstructed mass distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

4.6.3 Top quark tagging with ML techniques

A MultiClass Boosted Decision Trees (BDT) algorithm has been implemented to discriminate the True Top candidates against the combinatorial background. In order to maximize the discrimination power and the corresponding selection performances, separate trainings of the BDT have been performed according to the different features of the candidate, in particular: the Merged or Resolved configuration of the top candidate, the lepton flavor (muon or electron), and the top candidate momentum, split in two categories below or above the $p_T = 500\ GeV$ threshold. The top lepton flavour splitting ensure to take into account different reconstruction features for the muons and the electrons. A top candidate with $p_T > 500\ GeV$ is defined as a high p_T top candidate, otherwise it is a low p_T top candidate. In total, there are eight different trainings and each of them uses a different set of variables, as reported in Table 4.1. Feature selection has been performed in order to choose the best smaller discriminating set of variable, minimizing computational costs and maximizing the signal selection efficiency. In Appendix A distributions of the most discriminating variables are shown.

Top Lepton	Electron				Muon			
Top Configuration	Resolved		Merged		Resolved		Merged	
Top p$_T$	Low p$_T$	High p$_T$	Low p$_T$	High p$_T$	Low p$_T$	High p$_T$	Low p$_T$	High p$_T$
Jet_Mass	✓	✓	✓	✓	✓	✓	✓	✓
Jet_p$_T$	✓	✓	✓		✓			
Jet_DeepFlavB	✓	✓	✓	✓	✓	✓	✓	✓
Lepton_p$_T$	✓							
Lepton_Dxy	✓	✓	✓	✓	✓	✓	✓	✓
Lepton_DxyErr	✓				✓			
Lepton_ Dz	✓	✓	✓	✓	✓		✓	✓
Lepton_DzErr	✓				✓			
Lepton_Iso	✓	✓	✓	✓	✓	✓	✓	✓
Lepton_MiniIso	✓	✓	✓	✓	✓	✓	✓	✓
$\frac{Lepton\_p_T}{Jet\_p_T}$	✓		✓	✓			✓	
Top_Mass	✓	✓	✓	✓	✓	✓	✓	✓
Top_p$_T$	✓		✓		✓		✓	✓
Top_Mass (No MET)	✓	✓			✓	✓		
Top_p$_T$ (No MET)	✓							
Top_m$_T$	✓	✓	✓	✓		✓		✓
Top_Rel_p$_T$	✓	✓	✓	✓		✓	✓	✓
Top_cosθ	✓		✓		✓			✓
$\Delta R(l_{top}, jet_{top})$	✓		✓	✓	✓	✓		

Table 4.1: List of variables used in the trainings of the multiclass BDT algorithm.

The XGBoost BDT algorithm [66] has been trained on 2018 MC samples, in particular on signal samples T $\rightarrow$ tH with T mass equal to 600 GeV, 1200 GeV, and 1800 GeV, and on a $t\bar{t}$ background sample with M($t\bar{t}$) $> 1000\ GeV$, with a total number of top quark candidates equal to 929566. The training splitting in Top p_T allow us to improve the top quark selection and to remove the kinematics bias of the event. However the SFs are measured for each year separately.
The BDT algorithm is trained with a sample of n example top quark candidates, each with m features, the training dataset D is made by:

$$D = \{(\mathbf{x}_i, y_i)\}, |D| = n, \mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}, \tag{4.9}$$

where y_i is the truth label of the $i-$th entry.
The prediction $\hat{y}_i$ is function Φ of the input features, with given by a set of K additive functions:

$$\hat{y}_i = \Phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \tag{4.10}$$

where $f_k \in \mathcal{F}$, the space of regression tree:

$$\mathcal{F} = \{f(x) = w_q(x)\}, q : \mathbb{R}^n, w \in \mathbb{R}^T, \tag{4.11}$$

each function f_k corresponds to an independent tree, with T leaves in each tree, and q representing the structure of the tree, mapping the example to the leaf index. The weight of each leaves w_i is a continuous score and the decision rules of the trees are given by q. Then the weights are summed in order to predict the outcome of the input example.

In each iteration is computed the loss function to minimize:

$$\mathbb{L}(\Phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{4.12}$$

the second term is a penalty term, in our case defined as:

$$\sum_{k=1}^{K} \Omega(f_k) = \alpha \sum_{j=1}^{T} |w_j| \tag{4.13}$$

in order to regularize the training and avoid overfitting.
The gradient tree boosting is provided in an additive manner, recovering information from the previous steps. Suppose $\hat{y}_i$ is the prediction at the $t-$th iteration the loss at this step is:

$$\mathbb{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t), \tag{4.14}$$

approximating at the second order:

$$\mathbb{L}^{(t)} = \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \Omega(f_t), \tag{4.15}$$

with $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$. Defining $I_j = i | q(\mathbf{x}_i) = j$ as the instance set of a leaf j, for a fixed structure $q(\mathbf{x})$ the optimal weight w_j^* of the $j-$th leaf to be subtracted to the original one is (taking into account the learning rate η):

$$w_j^* = -\frac{\sum_{i \in I_j} g_i + \frac{\alpha}{\eta}}{\eta \sum_{i \in I_j} h_i}. \tag{4.16}$$

The BDT is a multiclassifier algorithm provides three scores, which are the probabilities of belonging to each of the three categories: True, Other, and QCD-like top. Hyperparameters optimization have been performed with a train and test data splitting of 30%, maximizing the accuracy of the training and minimizing the overtraining. The eight trainings are performed with the same sample splitting. The top quark candidates selection is done using two different taggers:

- **Top_TvsQCD**, used to reject the most abundant QCD-like combinatorial background; it is defined as:
$$\text{Top\_TvsQCD} = \frac{P(True)}{P(True) + P(QCD)} = \frac{P(True)}{1 - P(Other)} \tag{4.17}$$
- **Top_TvsOther**, used to select the best top candidate of the event and reject other combinatorial backgrounds; it is defined as:
$$\text{Top\_TvsOther} = \frac{P(True)}{P(True) + P(Other)} = \frac{P(True)}{1 - P(QCD)} \tag{4.18}$$

The top quark candidates must first pass a requirement on the Top_TvsQCD tagger, defined to provide a signal efficiency of 99%. The selected top quark candidates are then required to pass the so-called Loose or the Tight working points (WPs) of the Top_TvsOther tagger: the Loose WP is defined to provide 90% of background rejection while the Tight WP is defined to provide 99% of background rejection (the background rejection is evaluated on the top quark candidates passing the cut on the Top_TvsQCD tagger). The cuts on the Top_TvsQCD and Top_TvsOther taggers are different for each of the eight trainings and are listed in Table 4.2.

In Figures-4.11,-4.12,-4.13,-4.14 the ROC curves for the TvsQCD (left) and TvsOth (right) output of the BDT for muons (top) and electrons (bottom). The red line in the TvsQCD plot is the preselction cut, the blue and orange lines are respectively the Loose and Tight WP cuts for the TvsOth tagger.

Top Lepton	Electron				Muon			
Top Configuration	Resolved		Merged		Resolved		Merged	
Top p_T	Low p_T	High p_T	Low p_T	High p_T	Low p_T	High p_T	Low p_T	High p_T
TvsQCD	0.09	0.17	0.035	0.09	0.09	0.17	0.05	0.06
TvsOth (Loose WP)	0.648	0.903	0.648	0.867	0.6	0.88	0.157	0.688
TvsOth(Tight WP)	0.962	0.985	0.962	0.987	0.964	0.99	0.808	0.99

Table 4.2: Selection requirements on the Top_TvsQCD tagger, and Tight/Loose working points for the Top_TvsOth tagger for the eight different trainings.

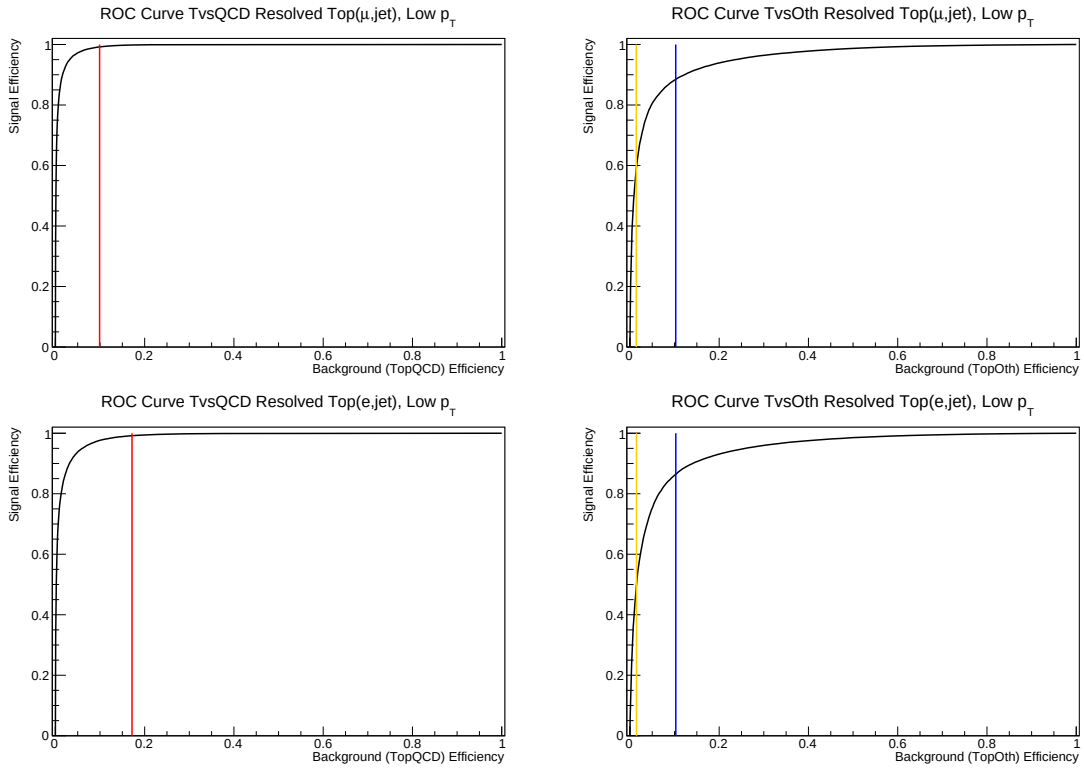

Figure 4.11: ROC curves for the TvsQCD (left) and TvsOth (right) output of the BDT for the Top Resolved Low p_T category for muons (top) and electrons (bottom). The red line in the TvsQCD plot is the preselction cut, the blue and orange lines are respectively the Loose and Tight WP cuts for the TvsOth tagger.

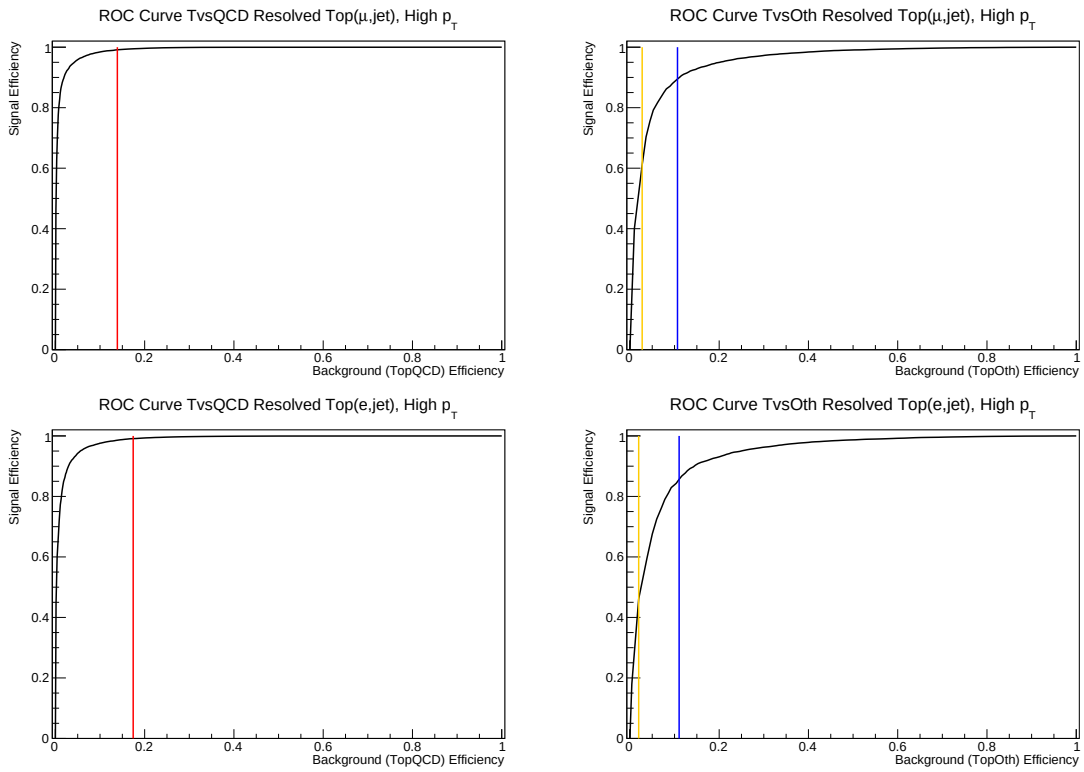

Figure 4.12: ROC curves for the TvsQCD (left) and TvsOth (right) output of the BDT for the Top Resolved High p_T category for muons (top) and electrons (bottom). The red line in the TvsQCD plot is the preselction cut, the blue and orange lines are respectively the Loose and Tight WP cuts for the TvsOth tagger.

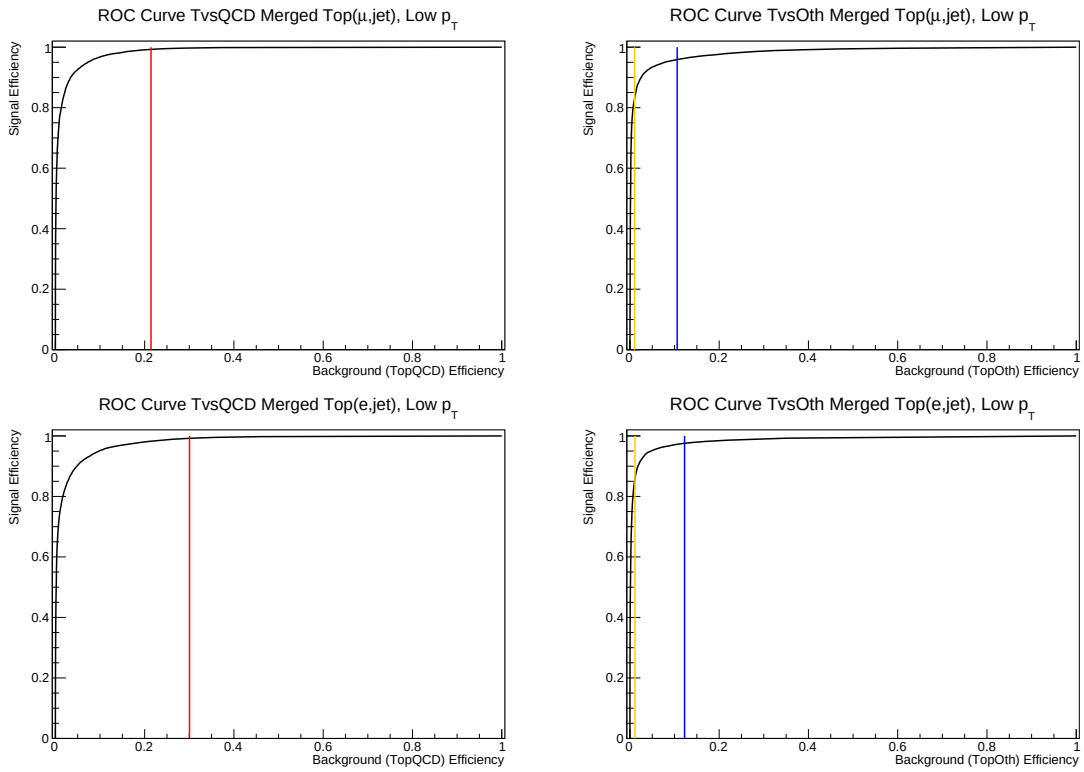

Figure 4.13: ROC curves for the TvsQCD (left) and TvsOth (right) output of the BDT for the Top Merged Low p_T category for muons (top) and electrons (bottom). The red line in the TvsQCD plot is the preselction cut, the blue and orange lines are respectively the Loose and Tight WP cuts for the TvsOth tagger.

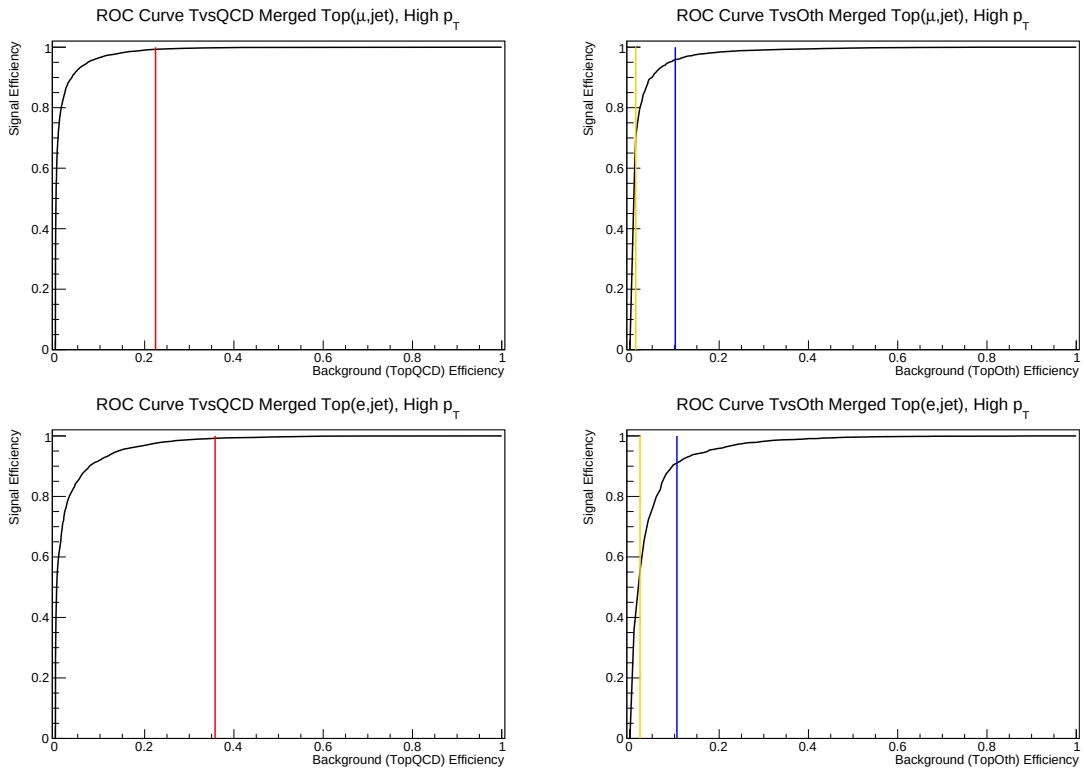

Figure 4.14: ROC curves for the TvsQCD (left) and TvsOth (right) output of the BDT for the Top Merged High p_T category for muons (top) and electrons (bottom). The red line in the TvsQCD plot is the preselction cut, the blue and orange lines are respectively the Loose and Tight WP cuts for the TvsOth tagger.

In Figures 4.15-4.22 the taggers' output score distributions for the respective training sample (line) and the test sample (marker) are shown.

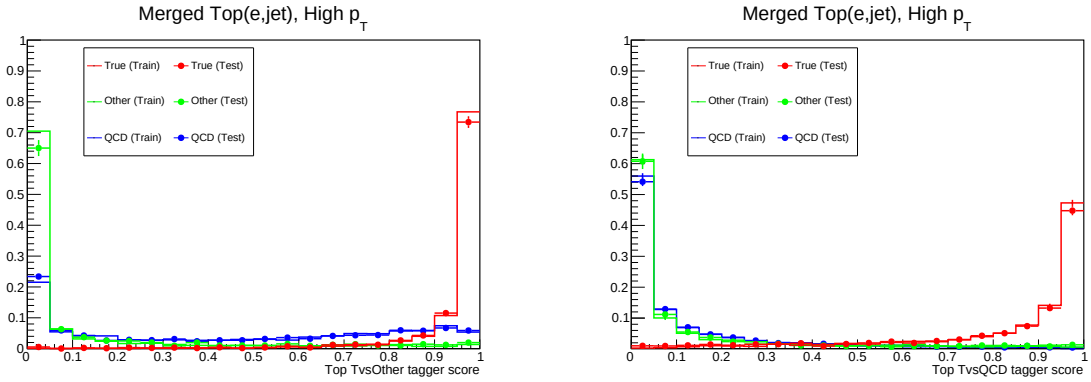

Figure 4.15: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with an electron, in Merged configuration and with $p_T \geq 500\ GeV$

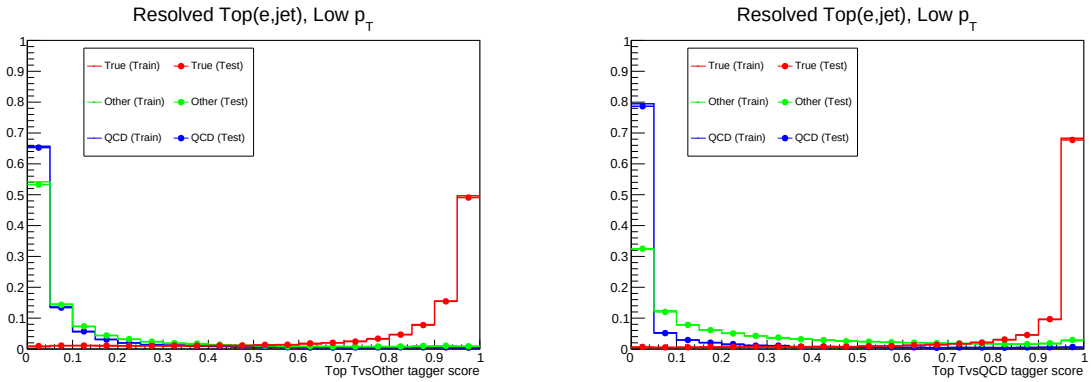

Figure 4.16: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with an electron, in Resolved configuration and with $p_T < 500\ GeV$

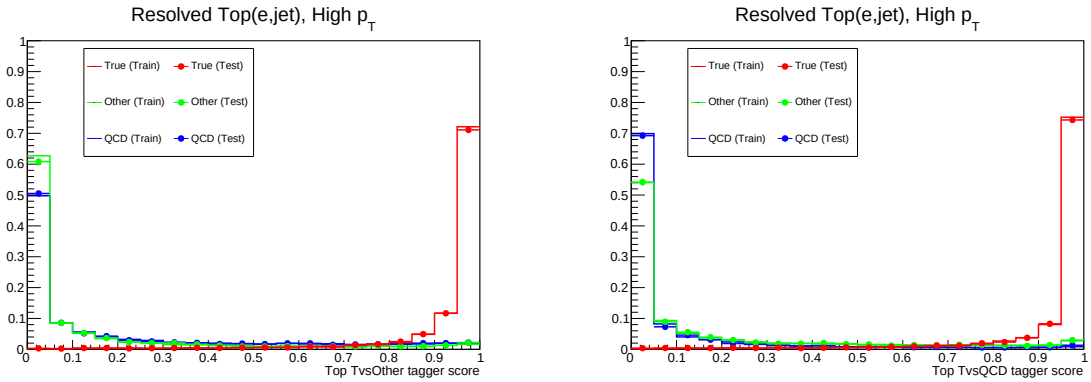

Figure 4.17: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with an electron, in Resolved configuration and with $p_T \geq 500\ GeV$

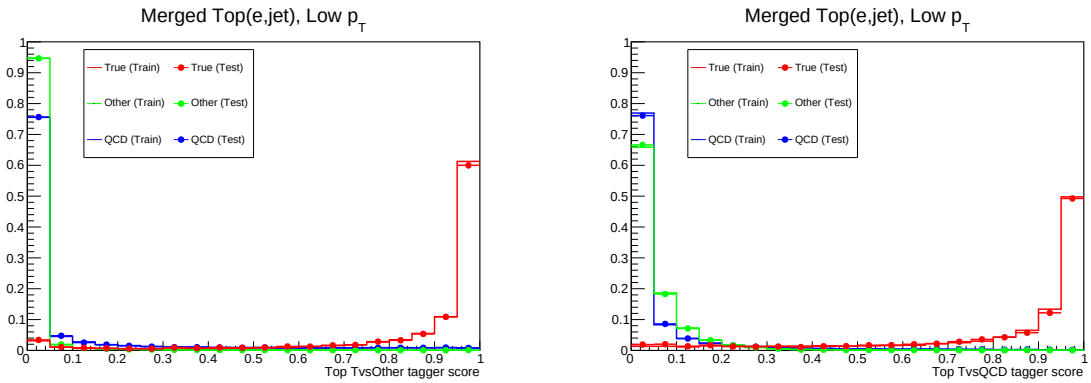

Figure 4.18: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with an electron, in Merged configuration and with $p_T < 500\ GeV$

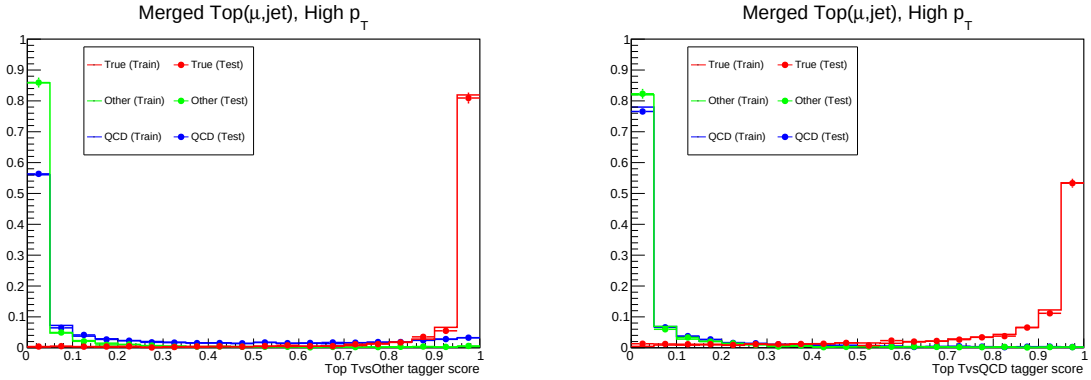

Figure 4.19: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with a muon, in Merged configuration and with $p_T \geq 500\ GeV$

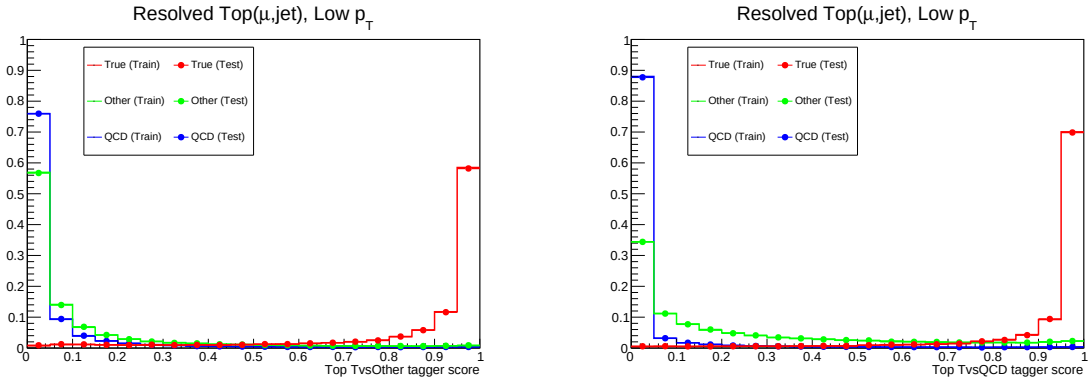

Figure 4.20: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with a muon, in Resolved configuration and with $p_T < 500\ GeV$

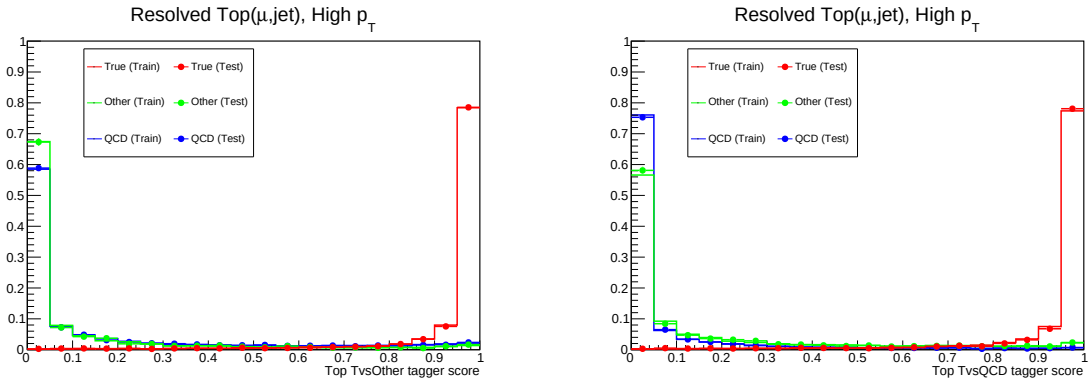

Figure 4.21: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with a muon, in Resolved configuration and with $p_T \geq 500\ GeV$

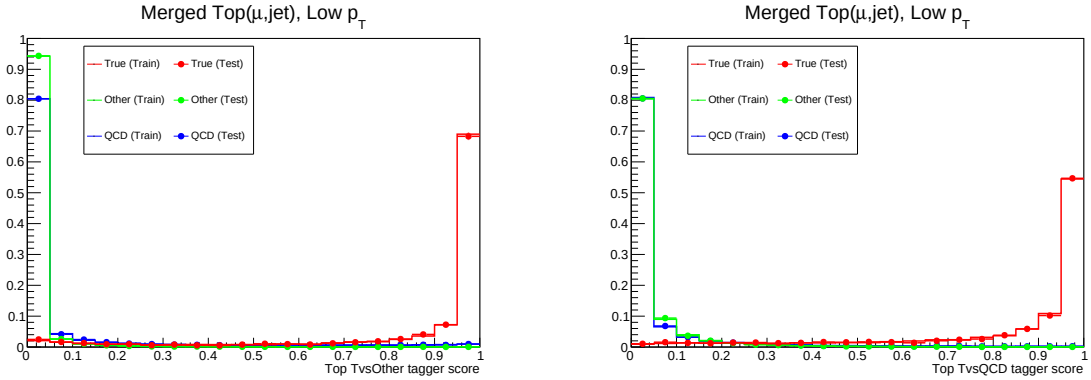

Figure 4.22: Distributions for the top taggers TvsOth (left) and TvsQCD (right) for top candidates with a muon, in Merged configuration and with $p_T < 500\ GeV$

In order to evaluate the overtraining, the score distribution in train and test datasets have been compared via χ^2 test. In Figure 4.23 the reduced χ^2 distribution performed on the train and test output of each training, for each category and both taggers, to evaluate the absence of overtraining. It shows that most of them are close to one, around 38% has a reduced χ^2 greater than one, and 4% greater than 2. These results are compatible with one and two sigma standard deviation of a reduced χ^2 distribution.

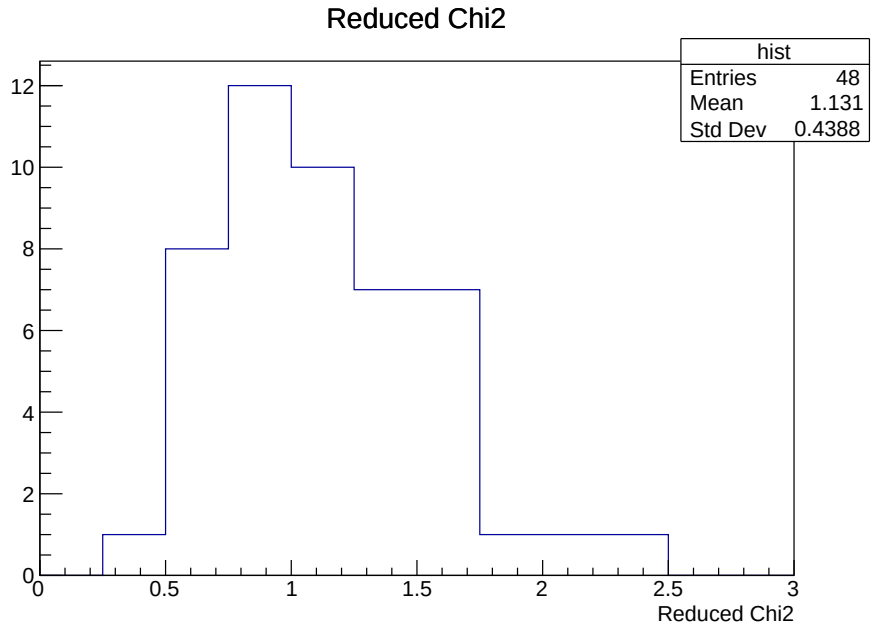

Figure 4.23: Distribution of the reduced χ^2, the χ^2 test has been performed to evaluate the absence of overtraining on the test and train samples.

Chapter 5

Analysis strategy

In this chapter will be described the analysis strategy used to extract the number of signal events and the signal cross section production, as well as the discovery potential for the process T to tH/A, and the exclusion upper limits. The analyzed data have been collected by the CMS experiment with an integrated luminosity of 138 fb^{-1}. Monte Carlo simulations have been used to estimate the signal and SM background contributions. Different T mass hypotheses have been taken into account, in particular for the T $\rightarrow$ tH search between 600 GeV and 1800 GeV, with steps of 100 GeV. Considering the search for T $\rightarrow$ tA, it was considered different hypotheses of couple (T,A). The T mass would be considered between 600 GeV and 3000 GeV, while A masses hypotheses between a minimum of 25 GeV and a maximum of 250 GeV. The 4 momentum of the T candidate is obtained by summing together the top quark 4 momentum, reconstructed as described in the previous chapter, and the A/H-tagged jet 4 momentum. The T reconstructed mass is used as discriminant variable to perform a maximum likelihood fit in the signal regions, in order to extract signal event yield.

5.1 Data and simulated samples

The analysis is performed using data collected by the CMS experiment during the so-called Run 2 of LHC (2016-2018), with a $\sqrt{s} = 13$ TeV and an integrated luminosity of $138 fb^{-1}$. Data are collected with triggers requiring either one muon, one electron, or one AK8 jet in the final state. Different primary dataset, i.e. a dataset collected requiring a specific kind of triggers in OR, are used, in order to maximize the signal efficiency. In Tables 5.1 the list of primary dataset employed is shown. The 2016 dataset have been split in two different parts, labelled as "preVFP" and "postVFP". The two eras will be treated separately, since substantial changes in the detector. The "2016postVFP" era is characterized by an higher signal to noise ratio and much bigger number of hits in the strip tracker, reducing the saturation effect in the readout chip after changing the voltage feedback preamplifier (VFP) [67].

5.1.1 Signal description

The search for singly-produced VLQ T is performed considering its decay to a top quark and a boson. The semileptonic final state is considered and, as shown in Figure 5.1, it is characterized by the presence of MET, a lepton, a jet, and a large-radius jet, from the clustering of the beauty quark-antiquark pair.

2016preVFP Dataset	Run	Integrated luminosity
SingleMuon(Electron/Photon), JetHT	B	5.8 fb^{-1}
SingleMuon(Electron/Photon), JetHT	C	2.6 fb^{-1}
SingleMuon(Electron/Photon), JetHT	D	4.2 fb^{-1}
SingleMuon(Electron/Photon), JetHT	E	4.0 fb^{-1}
SingleMuon(Electron/Photon), JetHT	F	2.9 fb^{-1}
2016postVFP Dataset	Run	Integrated luminosity
SingleMuon(Electron/Photon), JetHT	F	0.8 fb^{-1}
SingleMuon(Electron/Photon), JetHT	G	7.6 fb^{-1}
SingleMuon(Electron/Photon), JetHT	H	8.6 fb^{-1}
2017 Dataset	Run	Integrated luminosity
SingleMuon(Electron/Photon), JetHT	B	4.8 fb^{-1}
SingleMuon(Electron/Photon), JetHT	C	9.6 fb^{-1}
SingleMuon(Electron/Photon), JetHT	D	4.3 fb^{-1}
SingleMuon(Electron/Photon), JetHT	F	9.3 fb^{-1}
SingleMuon(Electron/Photon), JetHT	G	13.5 fb^{-1}
2018 Dataset	Run	Integrated luminosity
SingleMuon(Electron/Photon), JetHT	A	14.0 fb^{-1}
SingleMuon(Electron/Photon), JetHT	B	7.1 fb^{-1}
SingleMuon(Electron/Photon), JetHT	C	6.9 fb^{-1}
SingleMuon(Electron/Photon), JetHT	D	31.9 fb^{-1}

Table 5.1: List of primary datasets of pp collision data produced at $\sqrt{s} = 13\ TeV$ and collected by CMS during Run 2 employed in the analysis. Collected data correspond to an integrated luminosity of 138 fb^{-1}.

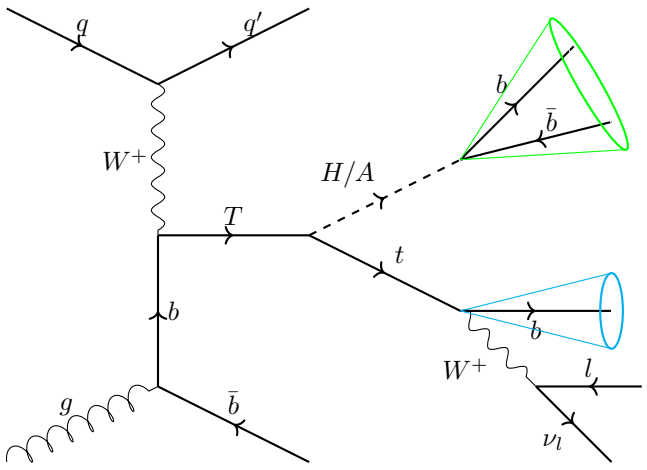

Figure 5.1: Feynman diagrams for singly-produced VLQ T decaying to a top quark and Higgs or new boson A. The final state consists of a lepton, MET, a jet, and a large-radious jet.

5.1.2 Background description

Different SM processes can mimick the final state we search for. The main SM backgrounds are :

- $t\bar{t}$, where a top quark-antiquark pair is produced as shown in Figure 5.2. This process is the main background, especially the semileptonic decay, where one of the two top quarks decays leptonically and the other one decays hadronically through $t \to Wb \to b\ q\bar{q}$. The W boson in latter case can fake the H/A boson decay.

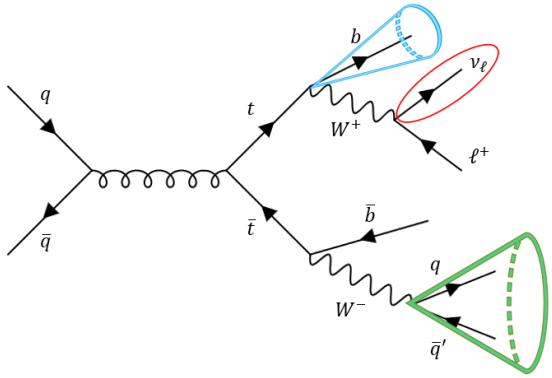

Figure 5.2: Feynman diagrams of the $t\bar{t}$ background.

- Single Top (ST) processes shown in Figure 5.3 can reproduce in some cases the signal topology if, the top quark decays leptonically and one of the additional quarks is reconstructed as a A/H-jet. For example, in the $t + W$ case, an hadronically decaying W boson can produce a very similar signature to the one of a Higgs/A boson.

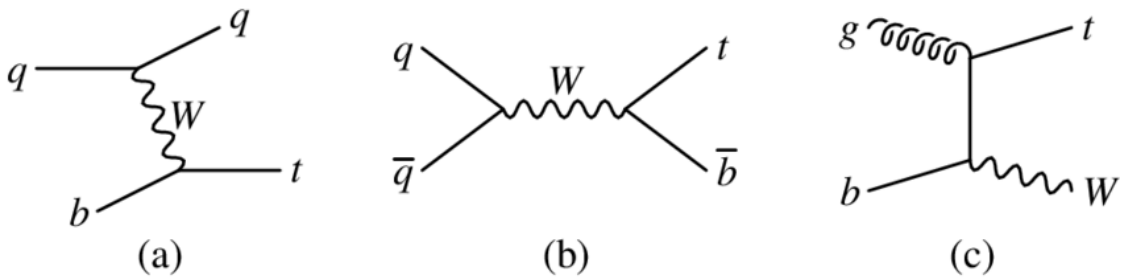

Figure 5.3: Feynman diagrams of the ST background.

- The $W + Jets$ processes can fake the signal topology if the W boson decays leptonically. The presence of jets from gluon splitting to b-quark pairs could fake the boson and the top quark, if another jet is close to the muon, could be reconstructed. The process is shown in Figure 5.4.

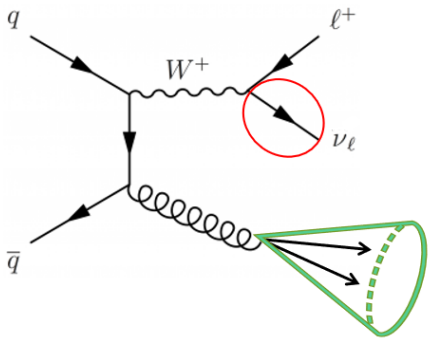

Figure 5.4: Feynman diagrams of the $W + Jets$ background.

- The QCD multijet processes, shown in Figure 5.5, could fake the signal topology, especially in the Merged configuration. However this background is reduced considerably thanks to the requirements on MET and to the ML top reconstruction.

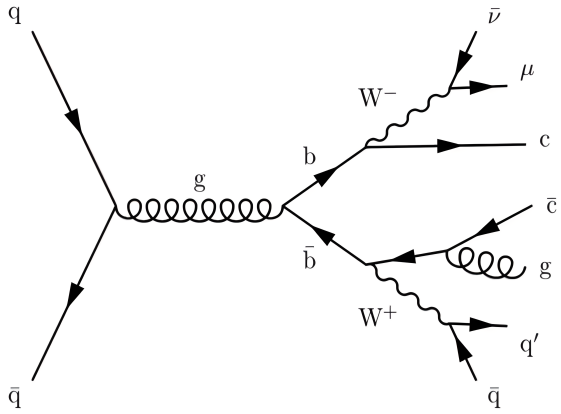

Figure 5.5: Feynman diagrams of the QCD background.

5.1.3 Signal and background simulation

Monte Carlo event generators are used to simulate signal and background samples. All the simulations are generated using different software frameworks. The matrix elements are generated either at leading order or at next-to-leading order with *Pythia* generates and simulates the hadronization of the particle produced in the pp collisions, while *GEANT* 4 [68] is used to simulate the particle interaction with the different CMS sub detectors. Signal event are generated at leading-order with *Madgraph* [69] and *Pythia* [70]. Inclusive samples are used to search for the VLQ decaying to a Higgs boson, thus all the SM decays of the top quark and Higgs boson have been simulated. Considering the decay to a new boson A, it has been simulated events concerning the desired final state, i.e. $T \to t(bl\nu)A(b\bar{b})$. Table 5.2 shows the list of simulated samples with relative cross sections times branching ratio.

The $t\bar{t}$ background processes are generated with *POWHEG*2.0[71, 72, 73] at the next-to-leading order (NLO) in perturbative QCD. Four different samples are used, considering the semileptonic final state, the dileptonic final state, and all inclusive samples in bin of the mass of the top quark-antiquark pair. In order to avoid the double counting, events from the first two samples are considered if the mass of the pair $t\bar{t}$ is lower than 700 GeV. Simulated QCD multijet events are generated at LO with *Pythia*, considering different bin of the total hadronic transverse energy (HT). Also for the W+Jets background have been considered different bin of HT, simulating events at NLO with *Madgraph* with up to two additional partons emitted in the hard scattering. The single top production is considered both in the t-channel and s-channel. Moreover, the production of a top quark in association with a W boson is considered. In Table 5.1.3 background simulated samples and cross sections.

T' mass [GeV]	A boson mass [GeV]												H
	25	50	75	100	125	150	175	200	250	350	450	500	
600	0.2903	0.6855	1.038	1.359	1.647	1.89	2.103	2.245	2.365	1.81	0.01525	8.808e-05	74.53
700	0.2274	0.5375	0.8201	1.083	1.327	1.547	1.74	1.907	2.135	2.099	1.408	0.7664	59.55
800	0.1764	0.4184	0.6423	0.8548	1.053	1.241	1.409	1.558	1.809	1.99	1.746	1.46	47.82
900	0.1424	0.3344	0.5126	0.6827	0.8462	0.9996	1.149	1.273	1.497	1.763	1.734	1.604	38.1
1000	0.1116	0.2655	0.4078	0.5434	0.6745	0.7994	0.9197	1.031	1.227	1.513	1.575	1.544	30.37
1100	0.08915	0.2115	0.3253	0.4356	0.5412	0.6435	0.7414	0.8344	1.003	1.263	1.382	1.396	24.4
1200	0.07205	0.1711	0.2633	0.3525	0.4401	0.5238	0.6072	0.6849	0.8286	1.061	1.187	1.218	19.73
1300	0.00836	0.1387	0.2136	0.2862	0.3573	0.4262	0.4933	0.5579	0.678	0.8797	1.014	1.051	15.96
1400	0.00705	0.1119	0.1724	0.2309	0.2881	0.3442	0.4004	0.4536	0.5538	0.7273	0.8543	0.8969	13.16
1500	0.03827	0.09102	0.1403	0.1883	0.2353	0.2808	0.3257	0.3694	0.4541	0.6032	0.7173	0.7597	10.59
1600	0.03119	0.07417	0.1144	0.1536	0.1921	0.2298	0.2668	0.3029	0.3721	0.4956	0.5936	0.6336	8.746
1700	0.00541	0.06043	0.09323	0.1252	0.1567	0.1876	0.2178	0.2476	0.3049	0.4092	0.4959	0.5317	7.071
1800	0.00076	0.04931	0.00608	0.1022	0.128	0.1533	0.1782	0.2026	0.25	0.3371	0.4113	0.4428	5.829
1900	0.00696	0.04037	0.0623	0.08373	0.1048	0.1257	0.1461	0.1662	0.2061	0.2791	0.341	0.3716	
2000	0.0139	0.03308	0.00107	0.06868	0.08602	0.1031	0.12	0.1367	0.169	0.2297	0.2834	0.3069	
2100	0.00143	0.00715	0.04193	0.05637	0.07061	0.08467	0.09855	0.1123	0.1395	0.1904	0.2359	0.2563	
2200	0.009434	0.00246	0.03467	0.04672	0.00855	0.07027	0.0819	0.09334	0.1157	0.1579	0.1955	0.2133	
2300	0.007755	0.01848	0.00853	0.00837	0.0481	0.00757	0.06703	0.07641	0.09486	0.1299	0.1618	0.1764	
2400	0.006361	0.01515	0.0034	0.03148	0.03946	0.00736	0.00519	0.06293	0.07813	0.1075	0.1343	0.1467	
2500	0.000237	0.01247	0.01926	0.0059	0.03251	0.03901	0.04547	0.00186	0.06452	0.08859	0.1111	0.1215	
2600	0.004255	0.01015	0.01568	0.0011	0.00646	0.03171	0.03697	0.04218	0.00243	0.07213	0.09067	0.09929	
2700	0.003545	0.008424	0.01301	0.00751	0.00196	0.00643	0.03084	0.03519	0.04376	0.0603	0.00577	0.0831	
2800	0.000937	0.006968	0.01076	0.00448	0.00817	0.00177	0.00538	0.00896	0.036	0.04967	0.06256	0.06865	
2900	0.000403	0.000726	0.008842	0.0019	0.00493	0.00793	0.00091	0.00387	0.00971	0.04106	0.00175	0.0069	
3000	0.001972	0.004696	0.000255	0.009765	0.00225	0.00471	0.0072	0.01964	0.00446	0.00381	0.0027	0.04694	

Table 5.2: Cross section times branching ratio in femtobarn (fb).

Sample	Cross section × BR [pb]
$t\bar{t}$ ($SemiLeptonic$)	364.35
$t\bar{t}$ ($DiLeptonic$)	87.31
$t\bar{t}$ (700-1000)	80.5
$t\bar{t}$ (1000-Inf)	21.3
QCD (HT100-200)	27990×10^3
QCD (HT200-300)	1712×10^3
QCD (HT300-500)	347.7×10^3
QCD (HT500-700)	32.1×10^3
QCD (HT700-1000)	6831
QCD (HT1000-1500)	1207
QCD (HT1500-2000)	119.9
QCD (HT2000-Inf)	25.24
W+Jets (lν,HT 100-200)	1340×1.26
W+Jets (lν,HT 200-400)	359.7×1.48
W+Jets (lν,HT 400-600)	48.91×1.26
W+Jets (lν,HT 600-800)	12.05×1.03
W+Jets (lν,HT 800-1200)	5.501×1.05
W+Jets (lν,HT 1200-2500)	1.329×0.77
W+Jets (lν, HT 2500-Inf)	0.03216×0.77
Single top ($\bar{t}$, $t-channel$)	80.95
Single top (t, $t-channel$)	136.02
Single top (tW,$t-channel$)	35.85
Single top ($\bar{t}$W,$t-channel$)	35.85
Single top ($s-channel$)	10.32

Table 5.3: SM background samples and their cross-sections. The W+Jets samples cross sections is multiplied by the scale factor obtained from the ratio NLO/LO.

5.2 Event selection

5.2.1 Trigger

Events are required to pass a combination of muon, electron, photon, or hadronic triggers. Trigger requirements have been chosen to exploit the presence of a muon reconstructed with a transverse momentum greater than 50 GeV or meeting isolation requirements with lower threshold of 24 GeV (27 GeV for 2017). Electron triggers require a threshold in p_T and passing the Tight working point for their reconstruction. Moreover, high-p_T electron can be reconstructed as photon by the algorithm used online at HLT and the photon trigger path have been included in the selection. The jet triggers require a reconstructed AK8 jet in the final state passing a p_T threshold different for each year, choosing the lower p_T threshold available for each year. For 2016 and 2018 this threshold is lower but the jet is required to have a trimmed mass greater than 30 GeV. In Table 5.4 the trigger paths used.

Purpose	Path	year
High p_T electron	HLT_Photon175	2016
	HLT_Photon200	2017, 2018
Low p_T electron	HLT_Ele27_WPTight_Gsf	2016
	HLT_Ele32_WPTight_Gsf	2017,2018
Muon	HLT_Mu50 or HLT_IsoMu24 or HLT_TkMu50 or HLT_IsoTkMu24	2016
	HLT_Mu50 or HLT_IsoMu27 or HLT_TkMu100 or HLT_OldMu100	2017
	HLT_Mu50 or HLT_IsoMu24 or HLT_TkMu100 or HLT_OldMu100	2018
Jet	HLT_PFJet360_TrimMass30	2016
	HLT_PFJet500	2017
	HLT_PFJet400_TrimMass30	2018

Table 5.4: Trigger paths required in OR.

5.2.2 Large-radius jet

Jets reconstructed with a cone parameter R=0.8 (AK8 jets or *FatJets*) are employed to search for the boosted hadronic decay of the heavy bosons into a quark-antiquark pair. Only AK8 jets with p_T greater than 500 GeV and $|\eta| < 2.4$ are considered. The AK8 jets are required to pass the working point Loose of the ParticleNet XbbVsQCD mass decorrelated tagger. Events are selected if at least one jet meets this condition. However, events are selected as well if at least one fatjet has its ParticleNet mass [74] between 60 GeV and 220 GeV.

5.2.3 Jets

As described above the AK4 jets are used to reconstruct the top quark candidate. All the jet are required to pass the Tight identification criteria [63] and to have a transverse momentum greater than 30 GeV. Forward jet are defined if reconstructed in the forward region of the detector, i.e. $2.4 < |\eta| < 4$. Only jets reconstructed in the barrel region ($|\eta| < 2.4$) are used to reconstruct the top quark candidates. Data-MC corrections are applied in order to take into account discrepancy in the pile-up identification efficiency.

5.2.4 Muons

Muons are required to have p_T greater than 30 GeV and $|\eta| < 2.4$, they are used to reconstruct the top quark if passing the Loose identification criteria. These muons will be labelled as Loose muons. The Tight muons will meet the same requirements in η and p_T, however these one pass the Tight identification criteria and the isolation in a cone with R equal to 0.4 is less than 0.15. Veto muons are defined as loose muons with an isolation less than 0.25. Data-MC corrections are applied in order to take into account discrepancy in the reconstruction, identification, and isolation.

5.2.5 Electron

Electrons are required to have p_T greater than 35 GeV and $|\eta| < 2.5$, they are used to reconstruct the top quark if passing the "mvaFall17V2noIso_WP90" identification criteria (see also Sec. 4.3). These electrons will be labelled as Loose electrons. The Tight electrons will meet the same requirements in η and p_T, however these one pass the "mvaFall17V2Iso_WP80" identification criteria, while the Veto electron passing the "mvaFall17V2Iso_WP90" identification criteria. Data-MC corrections are applied in order to take into account discrepancy in the reconstruction and identification.

5.2.6 Missing transverse energy

In order to take into account the presence of of the neutrino from the decay chain of the top quark, events are required to have missing transverse energy (MET) greater than 25 GeV.

5.2.7 Top quark selection

In every event all the possible top quark candidates are reconstructed, considering all the possible combination of the loose leptons and the AK4 jets. A procedure is then defined to select the best top quark candidate in the event, and to select and assign the events to mutually exclusive categories, which will be used to define signal and control regions. An event is assigned to the **TopTight** category if there is at least one top quark candidate passing the Tight WP of the TvsOth tagger. If there is more than one top quark candidate passing the Tight WP, they are prioritized as follows:

1. Top Merged High p_T,
2. Top Resolved Low p_T,
3. Top Resolved High p_T,
4. Top Merged Low p_T.

The Top Merged High p_T candidates are expected to be present in the final state for the boosted regime of top quarks, i.e. for high VLQ T mass hypotheses, while Top Resolved Low p_T candidates are expected for low VLQ T mass. The Top Resolved High p_T, and even more the Top Merged Low p_T candidates, have a low probability to be found the final state, given that for both categories the p_T regime would prefer the other configuration. A different prioritization choice, selecting first the Top Resolved Low p_T and then the Top Merged High p_T to be more sensitive for low VLQ T mass hypotheses, has been done, leading to the same final result. Top quark candidates with the same priority are ordered by choosing first the top quark candidates with muons and then with electrons, since muon reconstruction could be better thanks to the muon system and it can improve also the top quark reconstruction. Finally, top candidates with the same lepton are ordered by descending order of the TvsOth score. Only the best top candidate in the event, according to the above ordering, is chosen.

5.3 Event categorization

Events are preselected for the final analysis by requiring:

- at least one top quark candidate passing the TvsQCD cut and the Loose WP of the TvsOth tagger or, alternatively, at least one tight lepton as defined in Sec. 4.2.
- at least one AK8 jet passing the ParticleNet Loose WP or, alternatively, with a ParticleNet mass between 60 and 220 GeV (see Sec. 5.2.2).

If there are no top quark candidates passing the Tight WP but there is at least one top candidate passing the Loose WP of the TvsOth tagger, the event is assigned to the **TopLoose** category. The same procedure described above is applied in order to prioritize multiple top candidate and to select the best top candidate in the event.

The remaining events, which are not in the TopTight or TopLoose category, have at least one tight lepton and are assigned to the **TightLepton** category. In this case, the leading tight lepton will be referred in the following as the tight lepton of the event.

In order to reduce the $t\bar{t}$ and QCD-multijet background, the following additional requirements are applied, which depend on the event category.

For events in the TopTight or TopLoose category:

- the top candidate must have a TvsQCD score greater than 0.6;
- no additional veto muons or electrons (as defined in Sec. 4.2), beyond the lepton of the top candidate, must be present.

For events in the TightLepton category:

- no additional veto muons or electrons, beyond the tight lepton, must be present.

The AK8 jets in the event are ordered by descending order of the ParticleNet XbbVsQCD score, but only the jets passing the following ΔR requirements are selected.

For events in the TopTight or TopLoose category:

- $\Delta R(\text{AK8jet}, \text{Top}) > 1.2$;
- $\Delta R(\text{AK8jet}, \text{Top-jet}) > 1.2$;
- $\Delta R(\text{AK8jet}, \text{Top-lepton}) > 0.8$.

For events in the TightLepton category:

- $\Delta R(\text{AK8jet}, \text{Tight lepton}) > 0.8$.

If there are no AK8 jets passing the above ΔR cuts, the event is discarded. Otherwise, the best AK8 jet is the one with the largest XbbVsQCD score.

The selected events are then further classified in three categories depending on the XbbVsQCD score of the best AK8 jet and its ParticleNet mass:

- if XbbVsQCD > 0.98 the event is assigned to the **AK8Tight** category;
- if $0.8 <$ XbbVsQCD < 0.98 the event is assigned to the **AK8Loose** category;
- if XbbVsQCD < 0.8 and the ParticleNet mass is between 60 and 220 GeV the event is assigned to the **AK8Veto** category.

A schematic view of the regions is shown in Figure 5.6.

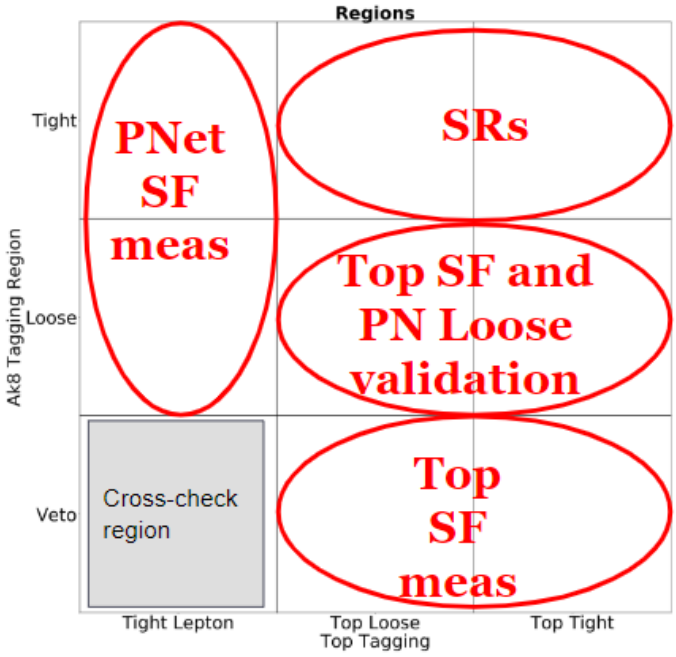

Figure 5.6: Schematic view of the signal and control regions defined in the analysis.

Different signal and control regions are defined, which depend on the top quark tagging category the and AK8 category the event belongs to. The control regions are employed to measure the XbbVsQCD and the TopVsOther scale factors, or as cross-check regions. They are defined as follows:

- **(TopTight, AK8Tight)**: signal region;
- **(TopLoose, AK8Tight)**: signal region;
- **(TopTight, AK8Loose)**: validation region;
- **(TopLoose, AK8Loose)**: validation region;
- **(TightLepton, AK8Tight)**: control region for AK8Tight SF measurement;
- **(TightLepton, AK8Loose)**: control region for AK8Loose SF measurement;
- **(TopTight, AK8Veto)**: control region for TopTight SF measurement;
- **(TopLoose, AK8Veto)**: control region for TopLoose SF measurement;
- **(TightLepton, AK8Veto)**: cross-check region.

In Figure 5.7 the reconstructed top quark candidates mass distribution in the control region (TopTight and TopLoose/AK8Veto), for 2016preVFP(top), 2016postVFP(mod-top) 2017(mid-bottom), and 2018(bottom) with a muon(left) or an electron(right) in the final state. The distribution peaks around the top quark mass, showing good agreement between data and MC simulations.

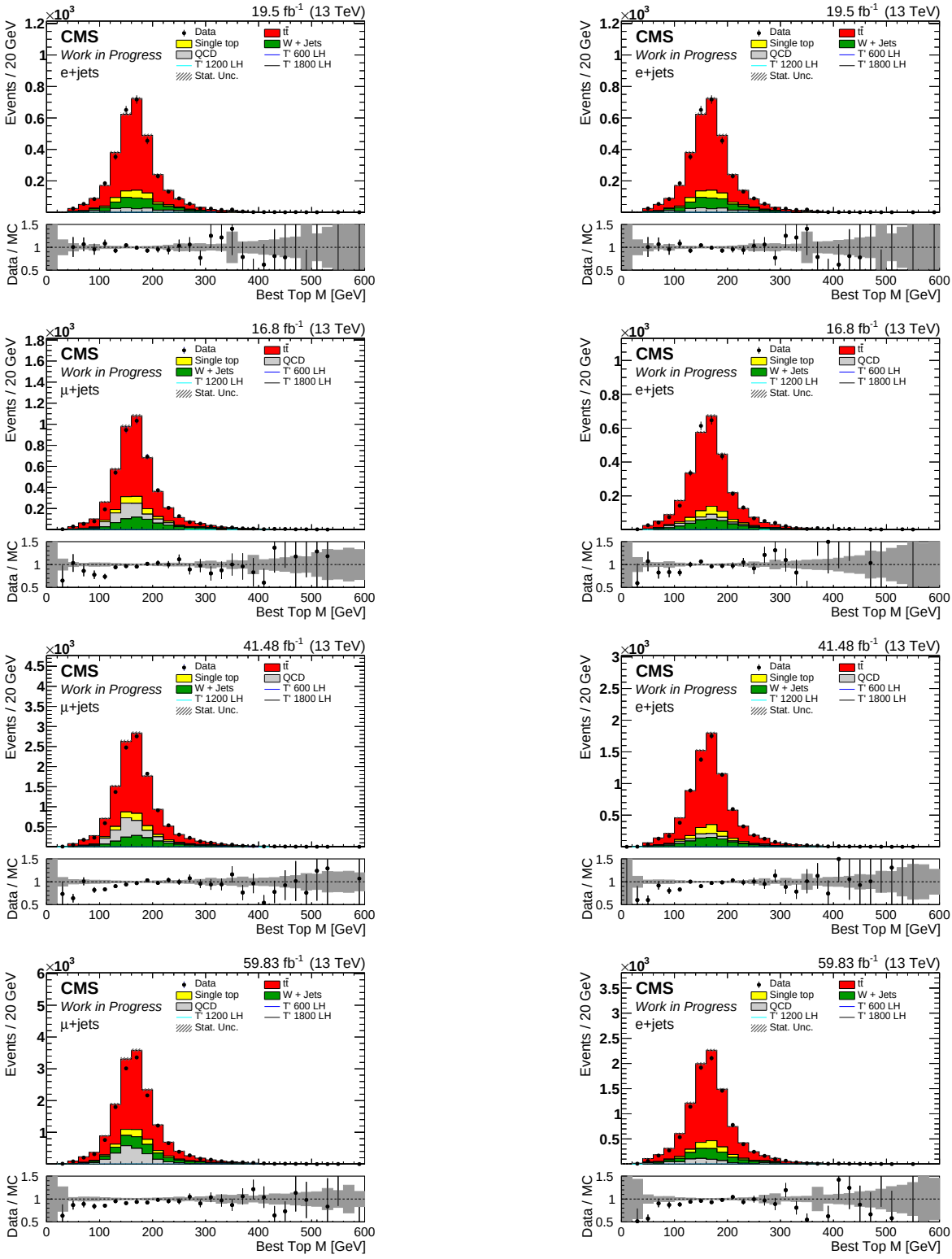

Figure 5.7: The reconstructed top quark candidates mass distribution in the control region (TopTight and TopLoose/AK8Veto), for 2016preVFP(top), 2016postVFP(mod-top) 2017(mid-bottom), and 2018(bottom) with a muon(left) or an electron(right) in the final state.

5.4 Data-MC corrections

Simulation is corrected to take into account several experimental effects which change year-by-year, due to the altered experimental conditions. In the following, we describe the corrections applied to our MC samples that haven't been described previously. For some of the corrections, scale factors are provided by the CMS collaboration. Other corrections, such as trigger and taggers efficiencies corrections have been measured for this analysis.

5.4.1 Pile-up reweighting

The pile-up (PU) distribution in simulation does not reproduce the one observed in data. The number of pile-up interactions in simulation is therefore reweighted to obtain a better agreement with the one observed in data. PU weights are obtained from the ratio of distributions of primary vertices in minimum bias data over that obtained from the true primary vertices in the MC simulation. The recommended cross-section for minimum bias in the Run 2 is 69.2 mb, with an uncertainty equal to 4.6%, that is treated as a systematic uncertainty.

5.4.2 Top p_T reweighting

A scale factor is applied to the generated top quark p_T spectrum to correct the discrepancy between data and $t\bar{t}$ simulation. The CMS collaboration provides scale factors, based on dedicated measurements [75, 76].

5.4.3 Pre-firing

In 2016 and 2017 a fraction of high η Trigger Primitives in the ECAL was associated with the wrong bunch crossing. Because of trigger rules vetoing subsequent events, this could cause a physics event of interest to self-veto in case a trigger in the interested region was fired. This effect is not described in simulation. The probabilities for a photon or a jet to pre-fire have been measured in data. By combining the probabilities for all the photon and jets in a simulated event, it is possible to compute the probability of the event not to pre-fire, that can be applied as an event weight.

5.4.4 HEM15/16 issue in 2018

During 2018 Runs C and D, two different sector of the HCAL in the endcap have stopped to work, after collecting $\sim 21\ fb^{-1}$. The afflicted region has $-3.2 < \eta < -1.3$ and $-1.57 < \phi < -0.87$ To take into account this issue, events with an electron or a jet in the region have been vetoed for these runs. Moreover, simulated events with 2018 MC where an electron or jet falls in this region are weighted by the luminosity corresponding to the data period up to the stop of the sectors. This corresponds approximately to 35% of the total 2018 sample.

5.4.5 DeepJet b-tagging discriminant shape

The CMS collaboration provides scale factors in order to correct the DeepJet discriminant shape, taking into account the discrepancy between data and MC of the algorithm performances. Since this discriminant has been used as input in the top tagger algorithm, events in the top loose and tight categories have been reweighted depending on DeepJet score of the final state AK4 jet, used to reconstructed the top quark selected candidate.

5.4.6 Trigger efficiency and scale factors measurements

The lepton trigger efficiency has been measured by considering an orthogonal sample with respect to the one used for this analysis.

A selection requiring:

- 1 tight lepton **AND** 1 additional veto lepton, in the Tight Lepton region
- 1 Loose or Tight Top **AND** 1 additional veto lepton, in the Top Loose and Top Tight regions

Moreover requirements on top-lepton p_T have been released to 10 GeV, in order to study the trigger efficiency as a function of a wider p_T range. The additional veto lepton is required to be an electron (muon), if the tight lepton or the top-lepton is a muon (electron). This selection is applied to data samples described in Sec. 5.1, as well as on di-lepton $t\bar{t}$ samples. The efficiency is then derived as:

$$\epsilon = \frac{N_{pass}}{N_{total}}, \tag{5.1}$$

where for:

- Muon trigger efficiency: N_{total} is defined as the number of events in which the electron or photon triggers are fired and N_{pass} is found by additionally requiring the muon is fired too;
- Electron trigger efficiency: N_{total} is defined as the number of events in which the muon trigger is fired and N_{pass} is found by additionally requiring the electron or photon are fired too.

For the measurement samples of dileptonic $t\bar{t}$ MC, reported in Table 5.1.3 have been used, while for Data, the same samples used for the analysis have been considered. Trigger efficiencies are shown in Figure 5.8, 5.9, 5.10, and 5.11 as a function of lepton p_T. Loose muons are required to have pt greater than 30 GeV, while Loose electrons are required to have pt greater than 35 GeV.

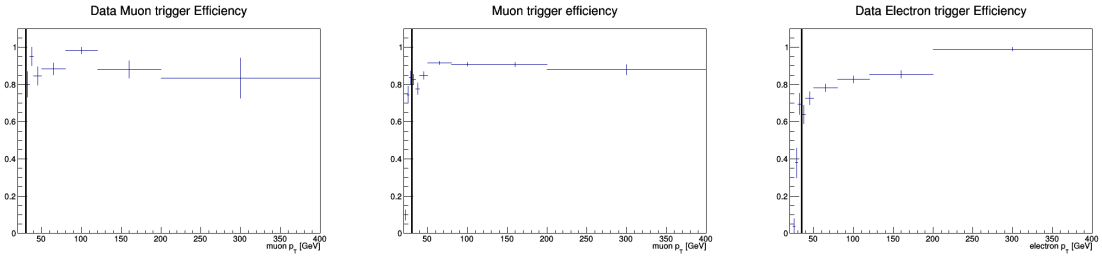

Figure 5.8: Trigger efficiency for data in 2016preVFP for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

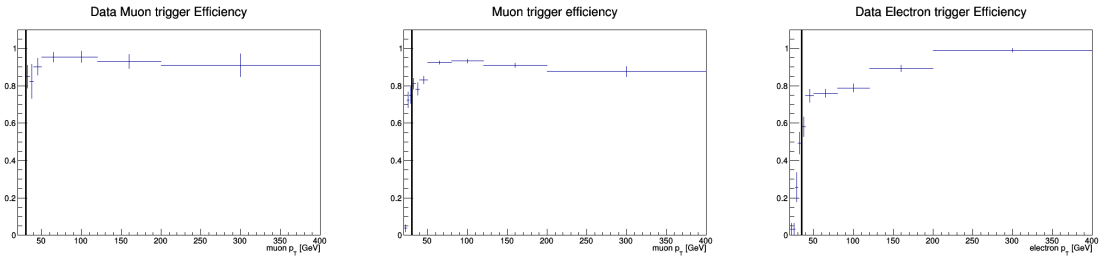

Figure 5.9: Trigger efficiency for data in 2016postVFP for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

In order to have the most precise tuning of MC samples to reproduce the data, the trigger efficiency has been derived with p_T-$|\eta|$ dependence. Moreover for the muons tight lepton region and Loose/Tight Top regions are split, allowed by the grater statistics in the dataset of events with a muon in the final state. Trigger efficiencies of data and MC, and scale factors are shown in Figure 5.12, 5.13, 5.14, and 5.15.

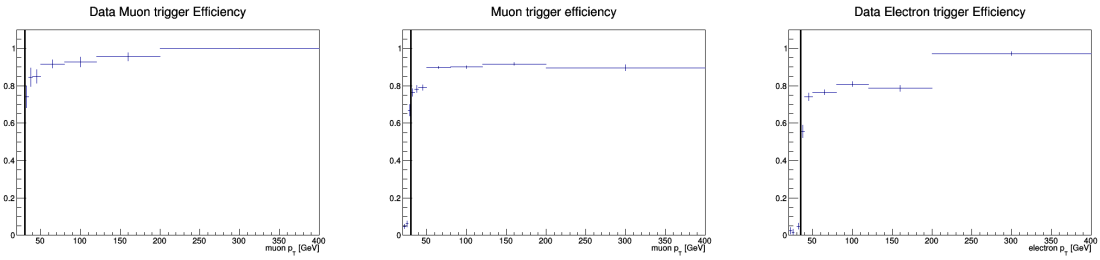

Figure 5.10: Trigger efficiency for data in 2017 for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

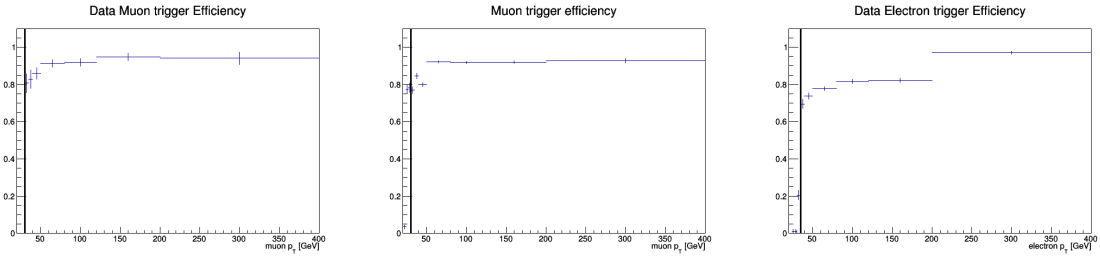

Figure 5.11: Trigger efficiency for data in 2018 for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

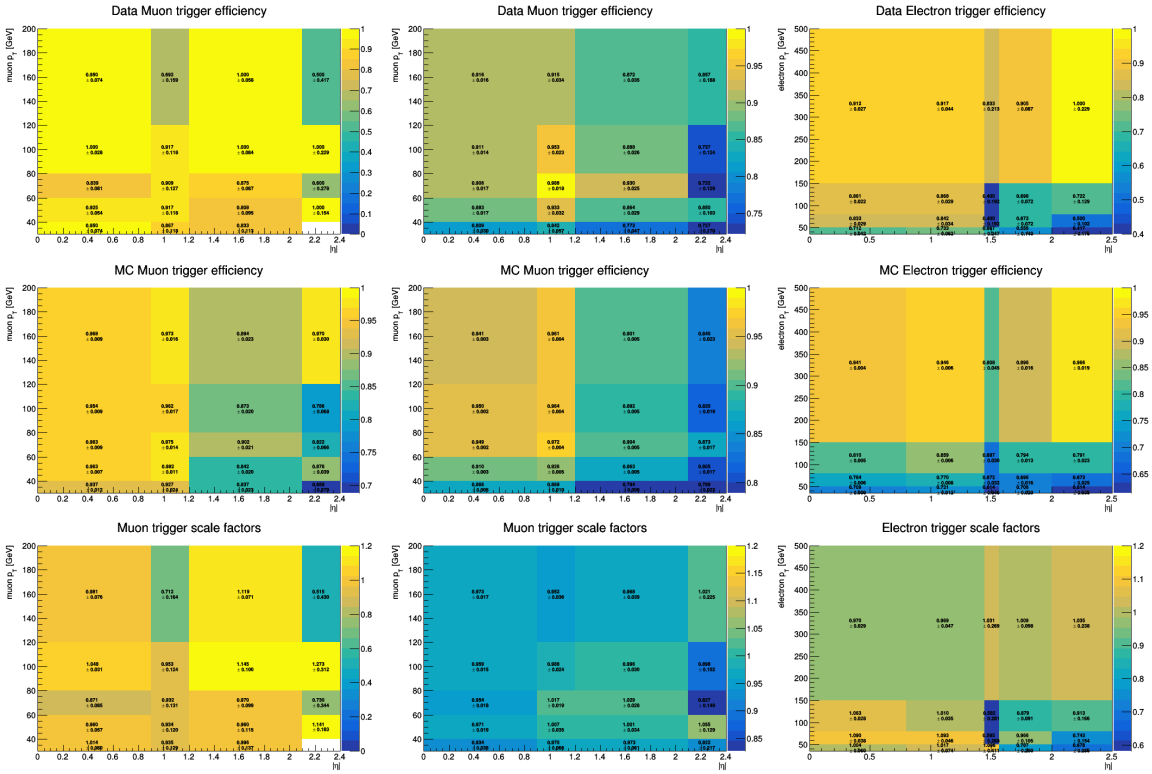

Figure 5.12: Trigger efficiency for data (top), $t\bar{t}$ (mid), and scale factor (bottom) in 2016preVFP for for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

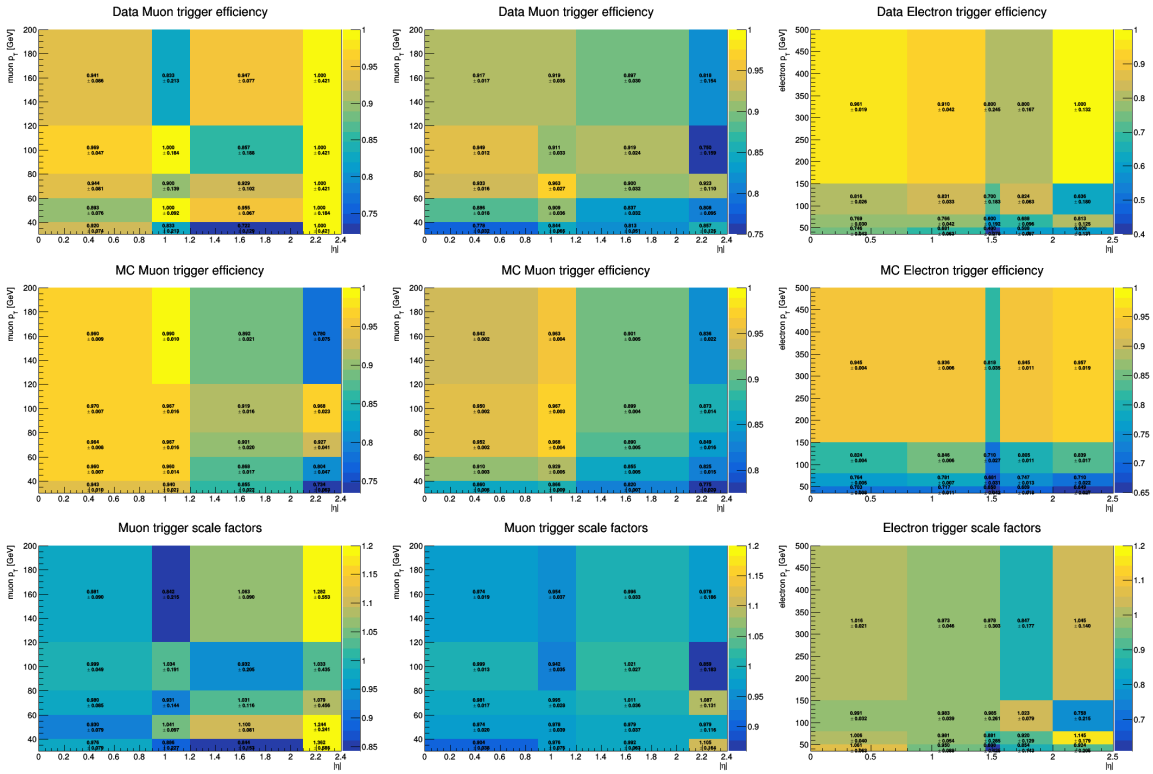

Figure 5.13: Trigger efficiency for data (top), $t\bar{t}$ (mid), and scale factor (bottom) in 2016postVFP for for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

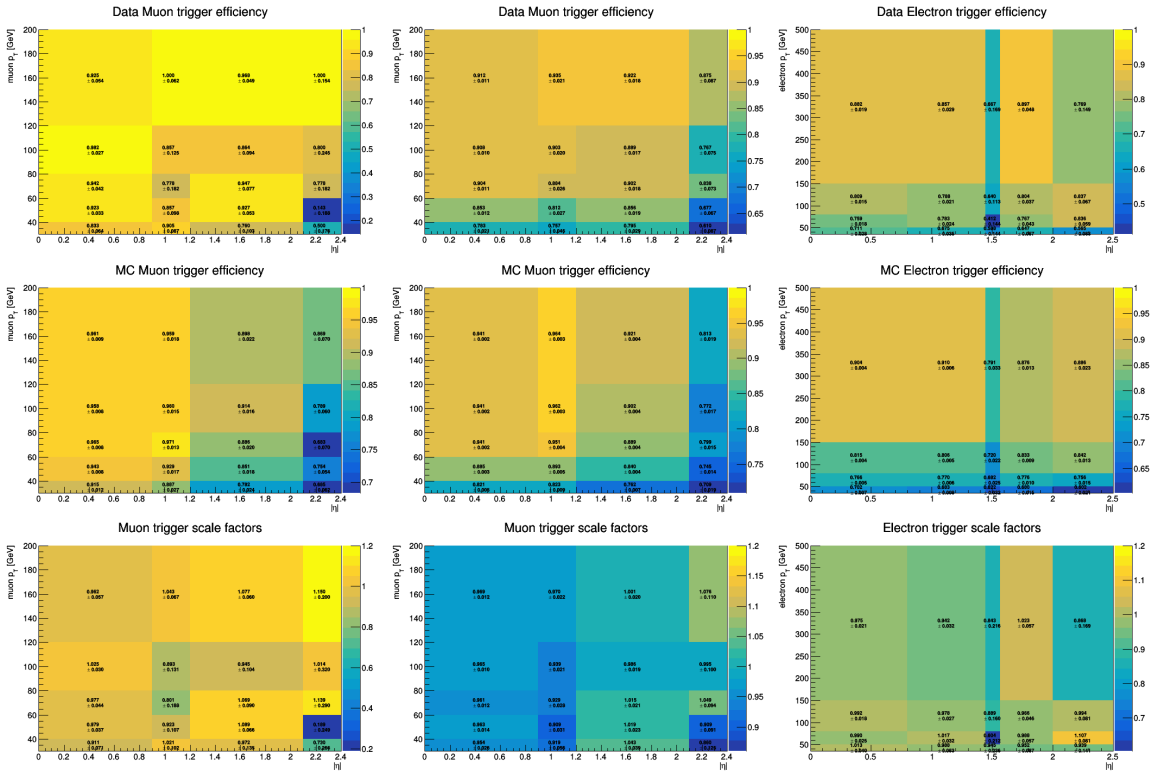

Figure 5.14: Trigger efficiency for data (top), $t\bar{t}$ (mid), and scale factor (bottom) in 2017 for for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

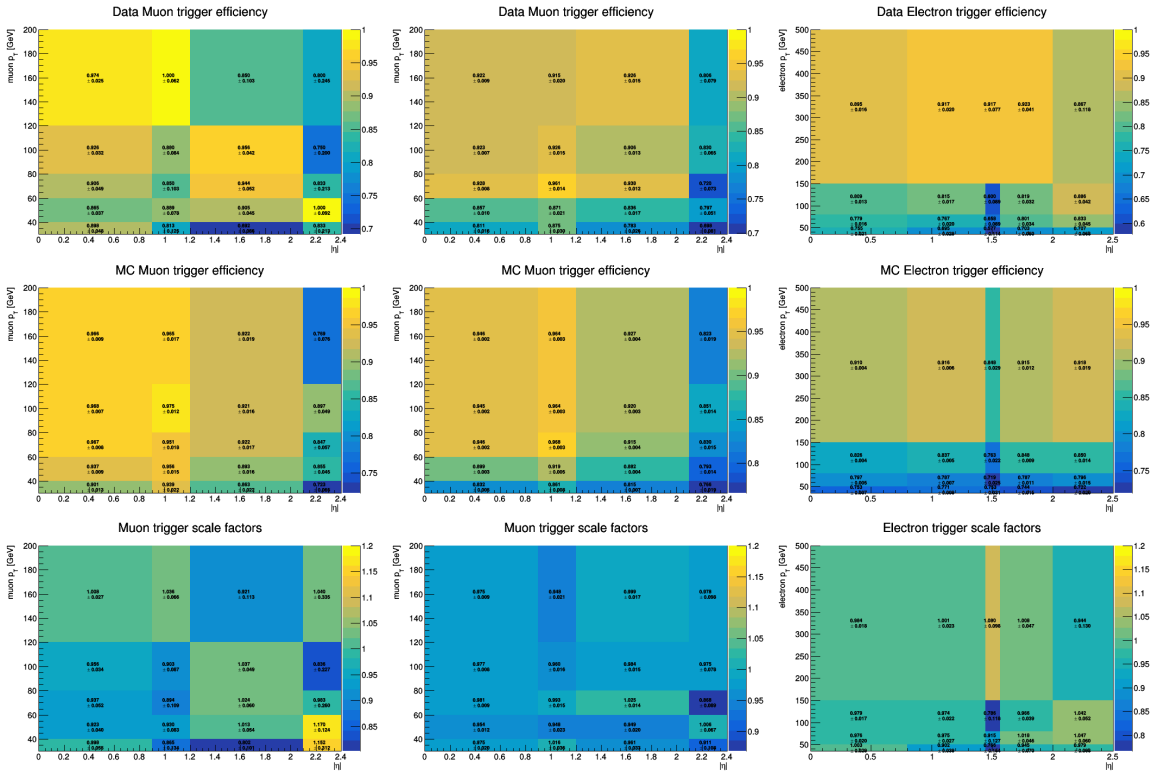

Figure 5.15: Trigger efficiency for data (top), $t\bar{t}$ (mid), and scale factor (bottom) in 2018 for for muons, in Tight Lepton region (left) and Top Loose/Tight (mid), and electrons (right).

The Jet trigger efficiency and scale factors are measured in the cross-check region, AK8 Veto - Tight Lepton, with the standard selection, i.e.with no additional Veto leptons. This selection is applied to data, as well as all MC samples described in Sec. 5.1.3.

Following the same approach of the lepton triggers, the efficiency is given by the ration of N_{total} and N_{pass}, where N_{total} is defined as the number of events in which the electron, photon, or muon triggers are fired and N_{pass} is found by additionally requiring the Jet trigger is fired too.

The jet trigger efficiency has been derived as a function of FatJet pt and soft drop mass. The fatjet pt is required to be greater than 500 GeV. Trigger efficiencies and scale factors are shown in Figure 5.16, 5.17, 5.18, and 5.19.

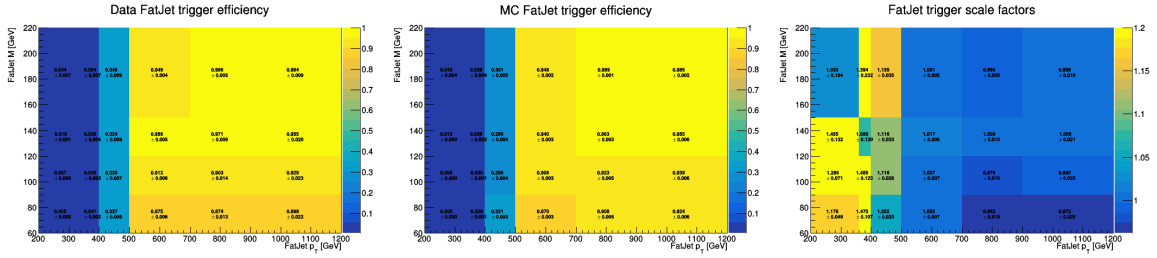

Figure 5.16: Trigger efficiency for data (left), MC (mid), and scale factor (right) in 2016preVFP for jets.

The selected events are required to have a triggerable, i.e. passing the p_T threshold, lepton as well as a triggerable jet and the total trigger scale factor is a combination of the lepton trigger and jet trigger scale factors.

The total number of the events in data is N_{Data}, give by:

$$N_{Data} = N^{l}_{Data} + N^{jet}_{Data} - N^{l\&jet}_{Data} \tag{5.2}$$

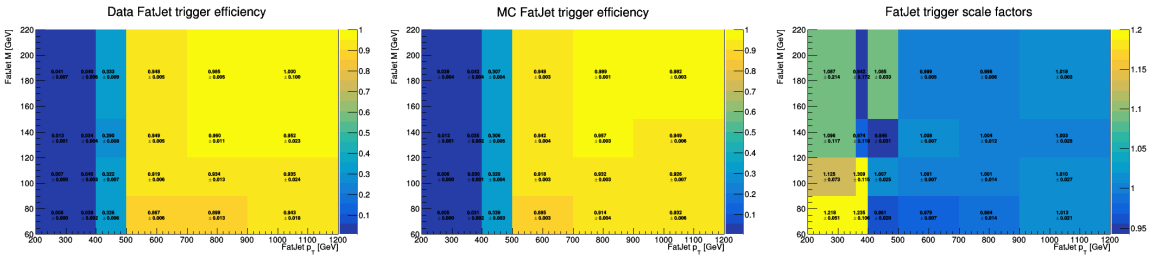

Figure 5.17: Trigger efficiency for data (left), MC (mid), and scale factor (right) in 2016postVFP for jets.

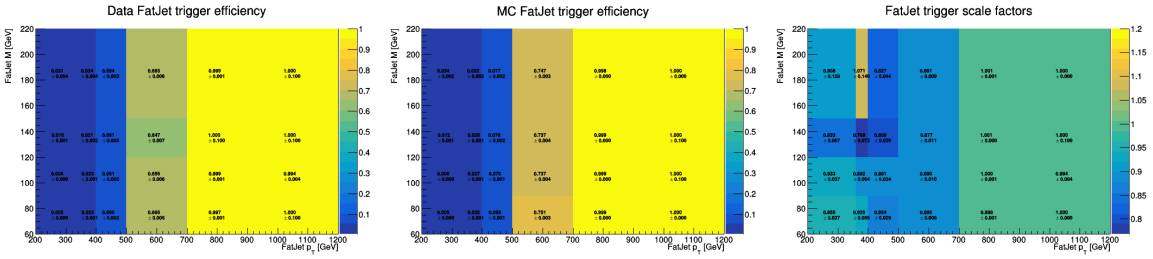

Figure 5.18: Trigger efficiency for data (left), MC (mid), and scale factor (right) in 2017 for jets.

where N^{l}_{Data} are the events with a triggerable lepton, N^{jet}_{Data} are the events with a triggerable jet, and $N^{l\&jet}_{Data}$ are the events with a triggerable lepton and jet, subtracted to remove the double counting of events that could happen in data due to an event passing through two trigger streams.

The total trigger efficiency in the data is:

$$\epsilon^{Data}_{tot} = \frac{\epsilon^{Data}_{l} \cdot N^{l}_{Data} + \epsilon^{Data}_{jet} \cdot N^{jet}_{Data} - \epsilon^{Data}_{l} \cdot \epsilon^{Data}_{j} \cdot N^{l\&jet}_{Data}}{N_{Data}}. \tag{5.3}$$

Since there is always a triggerable jet and lepton:

$$N^{l}_{Data} = N^{jet}_{Data} = N^{l\&jet}_{Data} = N_{Data} \tag{5.4}$$

and therefore

$$\epsilon^{Data}_{tot} = \epsilon^{Data}_{l} + \epsilon^{Data}_{jet} - \epsilon^{Data}_{l} \cdot \epsilon^{Data}_{j}. \tag{5.5}$$

The same approach is valid for MC sample and the total trigger scale factor is given by:

$$SF_{tot} = \frac{\epsilon^{Data}_{tot}}{\epsilon^{Data}_{MC}}. \tag{5.6}$$

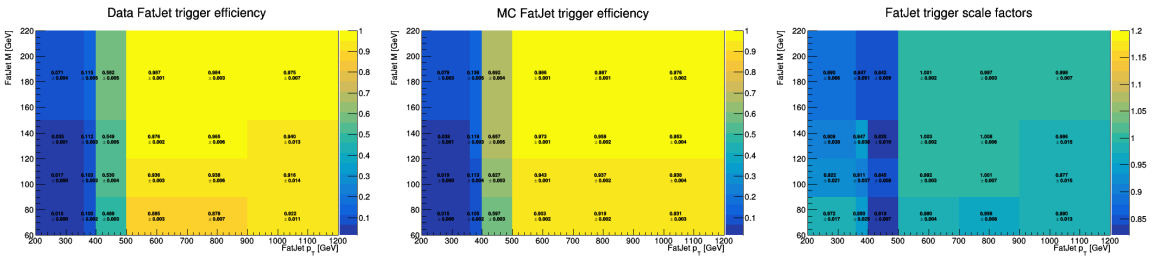

Figure 5.19: Trigger efficiency for data (left), MC (mid), and scale factor (right) in 2018 for jets.

5.5 Systematic uncertainties

All the systematic effects are included as nuisance parameters in the final fit for signal extraction. The final fit is performed on the invariant mass distribution of the reconstructed T quark candidates and the systematic uncertainties are considered to affect both the shape and the yield of the distributions.

5.5.1 Luminosity

According to the CMS collaboration measurements, uncertainties of 0.6%, 2.0%, and 1.5% are assigned corresponding to to the central values of the integrated luminosity in the years 2016, 2017, and 2018, respectively. This uncertainty is treated as affecting only the yields of the distributions and is uncorrelated across year. Moreover, an additional correlated component of the uncertainty is assigned to each year, amounting of 0.6%, 0.9%, and 2.0%for the years 2016, 2017, and 2018 respectively. Finally, an additional correlated component, specific to years 2017 and 2018 is added, amounting to 0.6% and 0.2% for the two years, respectively .

5.5.2 Factorization and renormalization scales

Uncertainties coming from the selection of factorization and renormalization scales are determined by considering variations where alternative scales are set to double or half of their nominal values. The maximum deviation among all variations is regarded as the uncertainty. These uncertainties are assessed separately for each process.

5.5.3 Parton Distribution Functions

The uncertainty attributed to Parton Distribution Functions (PDFs) is evaluated by utilizing reweighted distributions derived from all PDF sets, following the guidelines of PDF4LHC [77].

5.5.4 Top p_T reweighting

For the $t\bar{t}$ background an uncertainty related to the top p_T reweighting process is attributed, by shifting of a $\pm 50\%$ the value of the correction parameters.

5.5.5 Pile-up modelling

Systematic uncertainties related to pile-up modelling, as mentioned in Sec. 5.4.1, are taken into account by varying by $\pm 4.6\%$ the minimum bias cross section of 69.2 mb, used to calculate the data pile-up distributions.

5.5.6 Simulation statistics

The uncertainty due to the limited number of events in the MC samples is evaluated by allowing each bin of the T quark invariant mass distribution to fluctuate independently according to the statistical uncertainty of the bin content. This uncertainty is uncorrelated across all bins, therefore also across all regions and years, according to the Barlow-Beeston "light" method [78].

5.5.7 Lepton scale factor

The lepton reconstruction, identification, and isolation scale factors are applied to MC samples. The systematic uncertainty due to the lepton scale factors is obtained by shifting the nominal value of the scale factors of $\pm 1\sigma$, where σ is the scale factor uncertainty. Total statistical uncertainty is uncorrelated between all years, while systematic uncertainties on the reconstruction, identification, and isolation scale factors are correlated between years.

5.5.8 Jet pile-up identification scale factor

The AK4 pile-up identification scale factors are applied to MC samples. The systematic uncertainty due to the jet pile-up identification scale factors is obtained by shifting the nominal value of the scale factors of $\pm 1\sigma$, where σ is the scale factor uncertainty. Total statistical uncertainty is uncorrelated between all years.

5.5.9 Trigger scale factor

The trigger efficiency scale factors have been measured as described in detail in Sec. 5.4.6. The systematic uncertainty due to the trigger efficiency is obtained by shifting the nominal value of the trigger scale factors of $\pm 1\sigma$,where σ is the scale factor uncertainty. It is treated uncorrelated for each year and separately for events with a muon or an electron in the final state.

5.5.10 Pre-firing

As stated in Sec. 5.4.3, simulated events have been corrected for the pre-firing effect by applying a per-event weight, which is computed from the pre-firing probabilities of each jet and photon in the event. The uncertainties on the pre-firing probabilities have been propagated to derive the uncertainty on the weight.

5.5.11 Jet energy scale and resolution

All reconstructed jet four-momenta in simulated events are simultaneously varied according to the η- and p_T-dependent uncertainties in the jet energy scale. In order to evaluate the systematic effect due to differences in the jet energy resolution between data and MC, a smearing is also applied to MC events by increasing or decreasing the jet resolutions by their uncertainties. Jec and jer uncertainties are considered as uncorrelated source for each year.

5.6 Top tagger scale factors measurements

Scale factors related to the TvsOth tagger are measured and applied to simulation. The scale factors are measured by performing an extended maximum likelihood fit to the DeepJet discriminant of the AK4, used to reconstruct the best top candidate of the event. This results to be the most powerful discriminating variable. The fit is performed in the TopLoose/TopTight, AK8Veto control regions, considering both electrons and muons in the final state.

The fit has been performed assuming only background components. The systematics already described have been used also in this fit. The TvOth taggers SFs are obtained as rate parameters considering 3 different components of backgrounds:

- $t\bar{t}$ and ST, with a True Top,
- WJets, $t\bar{t}$, and ST, when the selected top doesn't match the MC truth,
- QCD.

The TvsOth scale factors are different for top quarks reconstructed with electrons and muons. To take into account transition from the Top Loose region to Top Tight region and vice-versa two additional parameters are used β_μ, β_{el}.

Due to the TopLoose $\longleftrightarrow$ TopTight transition, the number of the events in the TopTight region with a muon is given β_μ. The same is done for the electron case, but considering β_{el} in the final state is $N^T_{T\longleftrightarrow L}$:

$$N^T_{T\longleftrightarrow L} = \beta_\mu \cdot N^T_0, \tag{5.7}$$

where N^T_0 is the number of event in the TopTight region before the transition. The difference ΔN^T is given by:

$$\Delta N^T = N^T_{T\longleftrightarrow L} - N^T_0 = (\beta_\mu - 1) \cdot N^T_0. \tag{5.8}$$

The same approach is valid for the TopLoose regions, with a parameter $\beta_\mu^{'}$. However, ΔN^T and ΔN^L are not independent, since the vent migrating from the TopLoose region would enrich the TopTight region of the same number of events:

$$\Delta N^T = -\Delta N^L. \tag{5.9}$$

In the fit is added just the parameter β_μ, since $\beta_\mu^{'}$ is dependent by the first one being:

$$\beta_\mu^{'} = 1 - \frac{N_0^T}{N_0^L} \cdot (\beta_\mu - 1). \tag{5.10}$$

Moreover, due to the low contribution of QCD background in the regions with an electron in the final state, the scale factor of the QCD is set to unity. The QCD scale factor for the event with a muon in the final state is unique for the Top Loose and Top Tight regions. A systematic uncertainty, uncorrelated across years, is taken into account by varying the BDT SFs of $\pm 1\sigma$.

In Figure 5.20 the distribution of the DeepJet discriminant of the best top AK4 jet in the control region AK8 Veto, Top Loose(left) and Top Tight(right), with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.5 the top tagger SFs and their uncertainties for the 2016preVFP era.

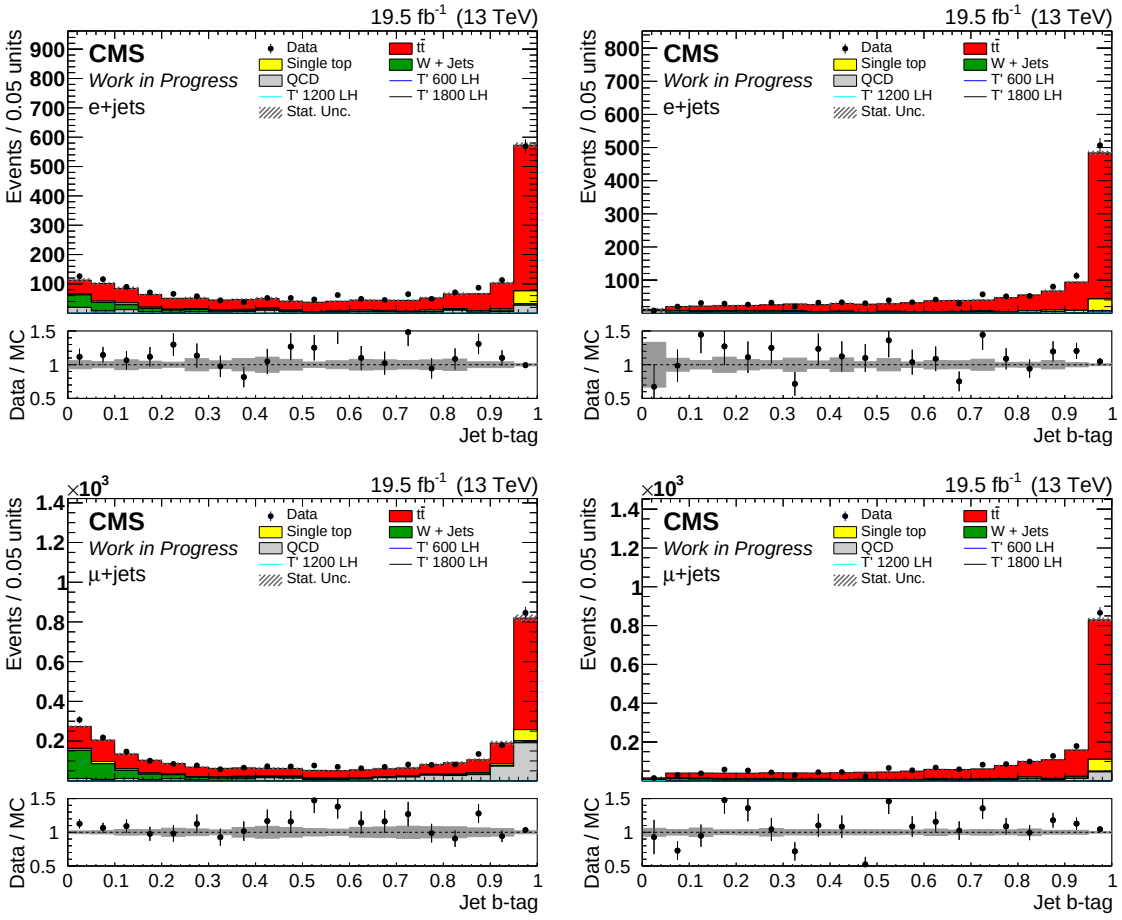

Figure 5.20: Distribution of the DeepJet discriminant of the best top AK4 jet after applying the SFs in the 2016preVFP era.

	SF	Δ SF	Backgrounds
Top True (el)	1.09	0.07	$t\bar{t}$, ST
Top Other (el)	0.50	0.12	$t\bar{t}$, ST, Wjets
Top True (μ)	1.00	0.12	$t\bar{t}$, ST
Top Other (μ)	0.61	0.09	$t\bar{t}$, ST, Wjets
QCD (el)	1	0	QCD
QCD (μ)	0.77	0.25	QCD
β_{el}	0.99	0.03	All
β_{μ}	0.94	0.05	All

Table 5.5: Top tagger scale factors and related uncertainties determined from the fit in the control regions in 2016preVFP

In Figure 5.21 the distribution of the DeepJet discriminant of the best top AK4 jet in the control region AK8 Veto, Top Loose(left) and Top Tight(right), with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.6 the top tagger SFs and their uncertainties for the 2016postVFP era.

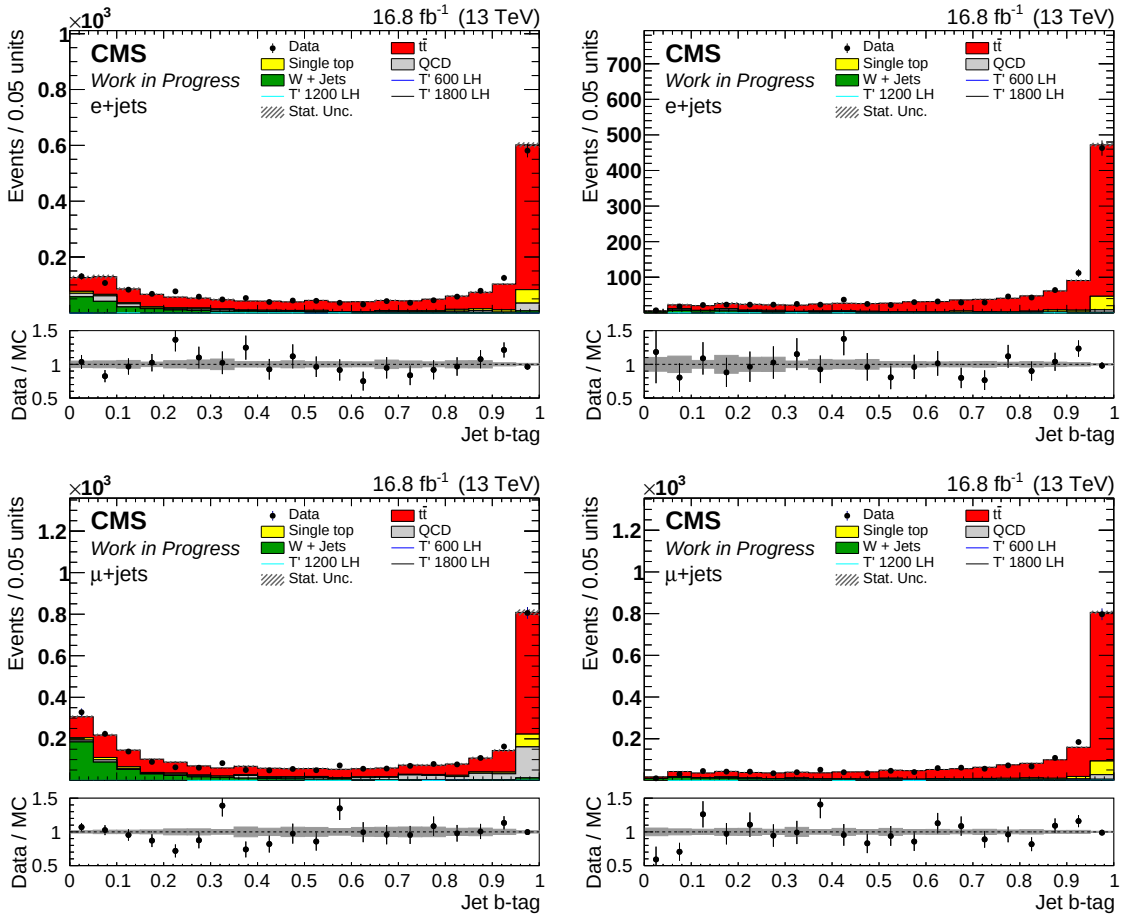

Figure 5.21: Distribution of the DeepJet discriminant of the best top AK4 jet after applying the SFs in the 2016postVFP era.

	SF	Δ SF	Backgrounds
Top True (el)	1.13	0.07	$t\bar{t}$, ST
Top Other (el)	0.68	0.12	$t\bar{t}$, ST, Wjets
Top True (μ)	1.06	0.10	$t\bar{t}$, ST
Top Other (μ)	0.88	0.10	$t\bar{t}$, ST, Wjets
QCD (el)	1	0	QCD
QCD (μ)	0.70	0.25	QCD
β_{el}	0.91	0.03	All
β_{μ}	0.92	0.05	All

Table 5.6: Top tagger scale factors and related uncertainties determined from the fit in the control regions in 2016postVFP

In Figure 5.22 the distribution of the DeepJet discriminant of the best top AK4 jet in the control region AK8 Veto, Top Loose(left) and Top Tight(right), with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.7 the top tagger SFs and their uncertainties for the 2017 era.

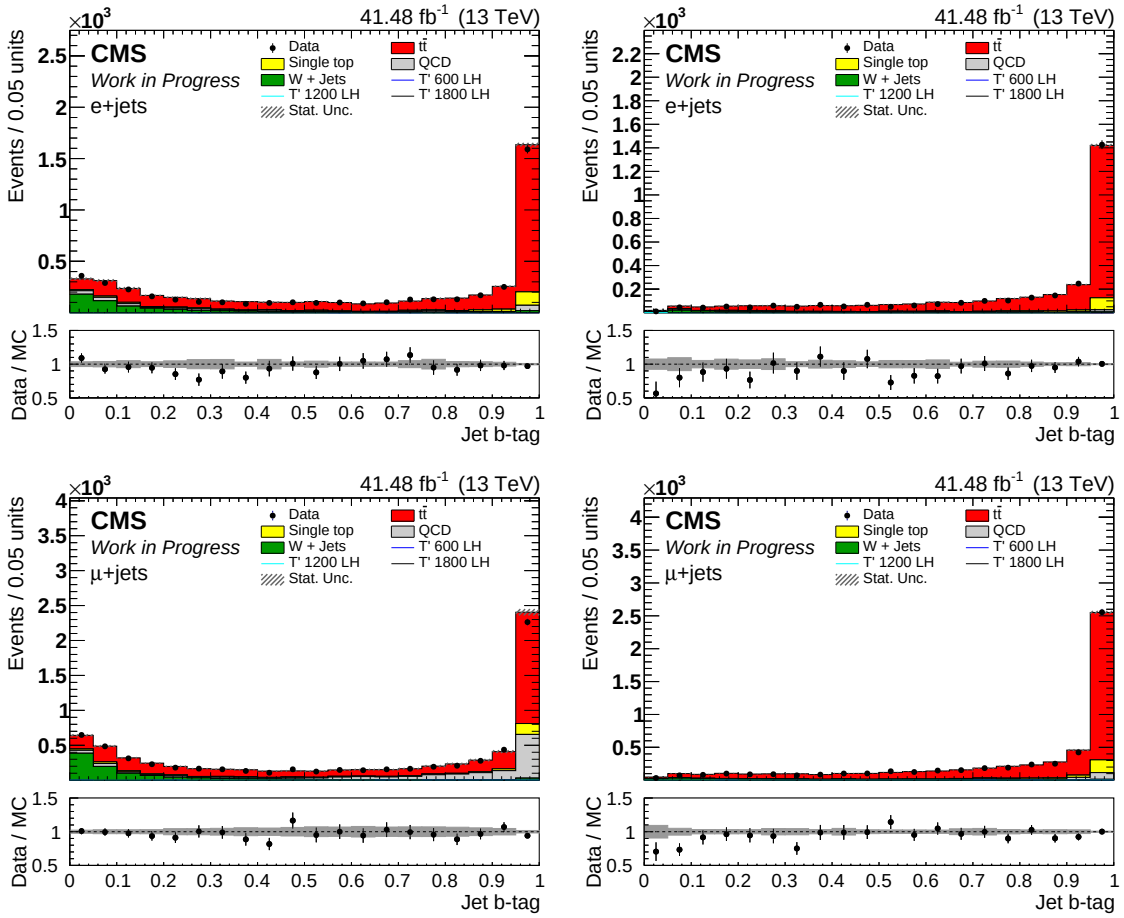

Figure 5.22: Distribution of the DeepJet discriminant of the best top AK4 jet after applying the SFs in the 2017 era.

	SF	Δ SF	Backgrounds
Top True (el)	1.06	0.06	$t\bar{t}$, ST
Top Other (el)	0.82	0.07	$t\bar{t}$, ST, Wjets
Top True (μ)	1.07	0.08	$t\bar{t}$, ST
Top Other (μ)	0.84	0.07	$t\bar{t}$, ST, Wjets
QCD (el)	1	0	QCD
QCD (μ)	0.94	0.18	QCD
β_{el}	0.91	0.02	All
β_{μ}	1.01	0.04	All

Table 5.7: Top tagger scale factors and related uncertainties determined from the fit in the control regions in 2017

In Figure 5.23 the distribution of the DeepJet discriminant of the best top AK4 jet in the control region AK8 Veto, Top Loose(left) and Top Tight(right), with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.8 the top tagger SFs and their uncertainties for the 2018 era.

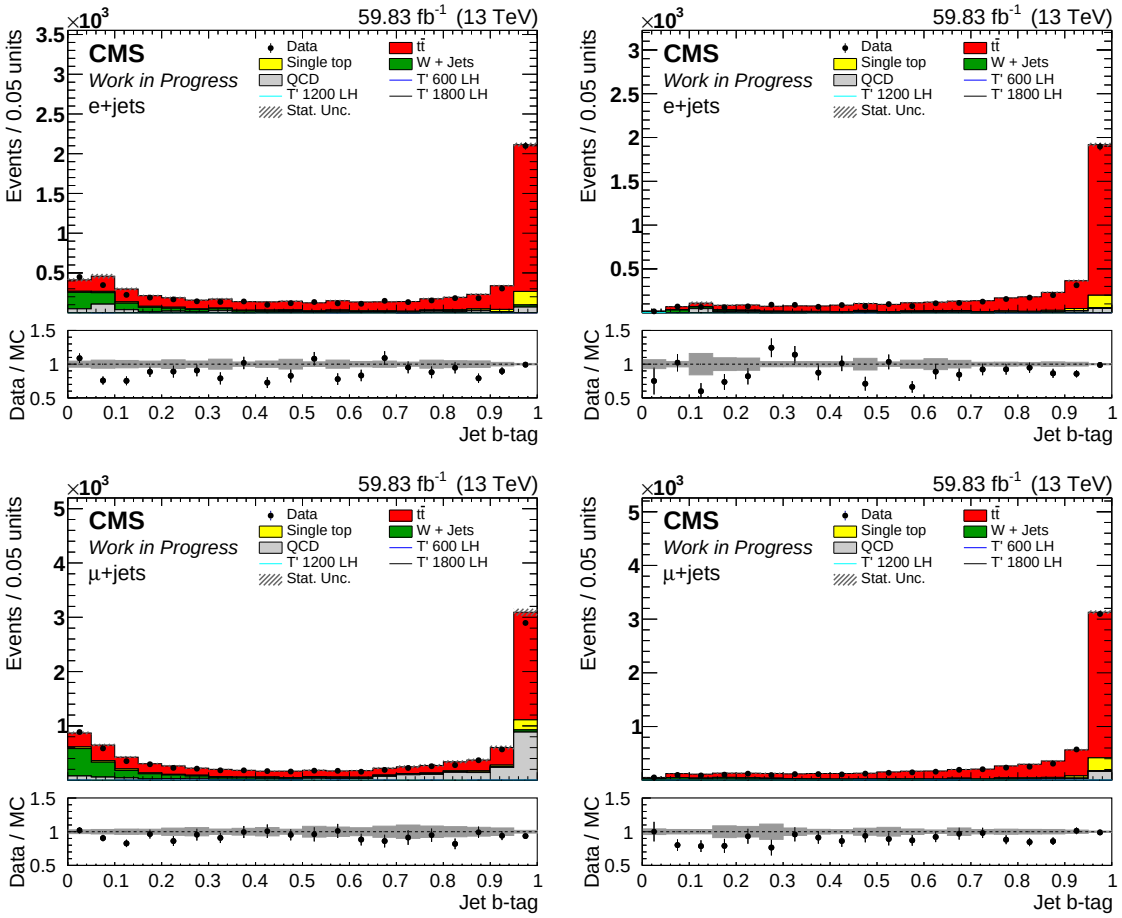

Figure 5.23: Distribution of the DeepJet discriminant of the best top AK4 jet after applying the SFs in the 2018 era.

	SF	Δ SF	Backgrounds
Top True (el)	1.15	0.06	$t\bar{t}$, ST
Top Other (el)	0.74	0.08	$t\bar{t}$, ST, Wjets
Top True (μ)	1.05	0.07	$t\bar{t}$, ST
Top Other (μ)	0.83	0.06	$t\bar{t}$, ST, Wjets
QCD (el)	1	0	QCD
QCD (μ)	1.05	0.16	QCD
β_{el}	0.93	0.02	All
β_{μ}	0.98	0.03	All

Table 5.8: Top tagger scale factors and related uncertainties determined from the fit in the control regions in 2018

5.7 ParticleNet tagger scale factors measurements

The ParticleNet scale factors are determined by fitting to the distributions of the ParticleNet mass of the best AK8 jet for events in the (TightLepton, AK8Loose/AK8Tight) regions with a ParticleNet mass greater than 110 GeV. Two different scale factors are measured, considering separate component of background:

- **TT(2q1b) component**:the AK8 jet matches with 2 light quarks and 1 beauty quark; in this case the ParticleNet mass distribution peaks at the top quark mass being hadronically decaying top quark passing the ParticleNet Loose WP,
- **Other(non resonant) component**: comprises all the other cases.

A different choice of background splitting, considering separate components for AK8 jets matching with 2 light quark or 1 light quark and 1 beauty quark, has been performed, leading to the same result.

To take into account transition from the AK8 Loose region to the AK8 Tight region and vice-versa the additional parameter α is defined and used in the same way described above for $beta_{\mu}$. The scale factors are measured in the high mass region, but they are applied in the validation and control region to the full ParticleNet mass spectrum. A systematic uncertainty, uncorrelated across years, is taken into account by varying the ParticleNet SFs of $\pm 1\sigma$.

In Figure 5.24 the distribution of the AK8 tagged jet ParticleNet mass in the control region AK8 Loose (left), AK8 Tight (right) T and Top Veto, with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.9 the top tagger SFs and their uncertainties for the 2016preVFP era.

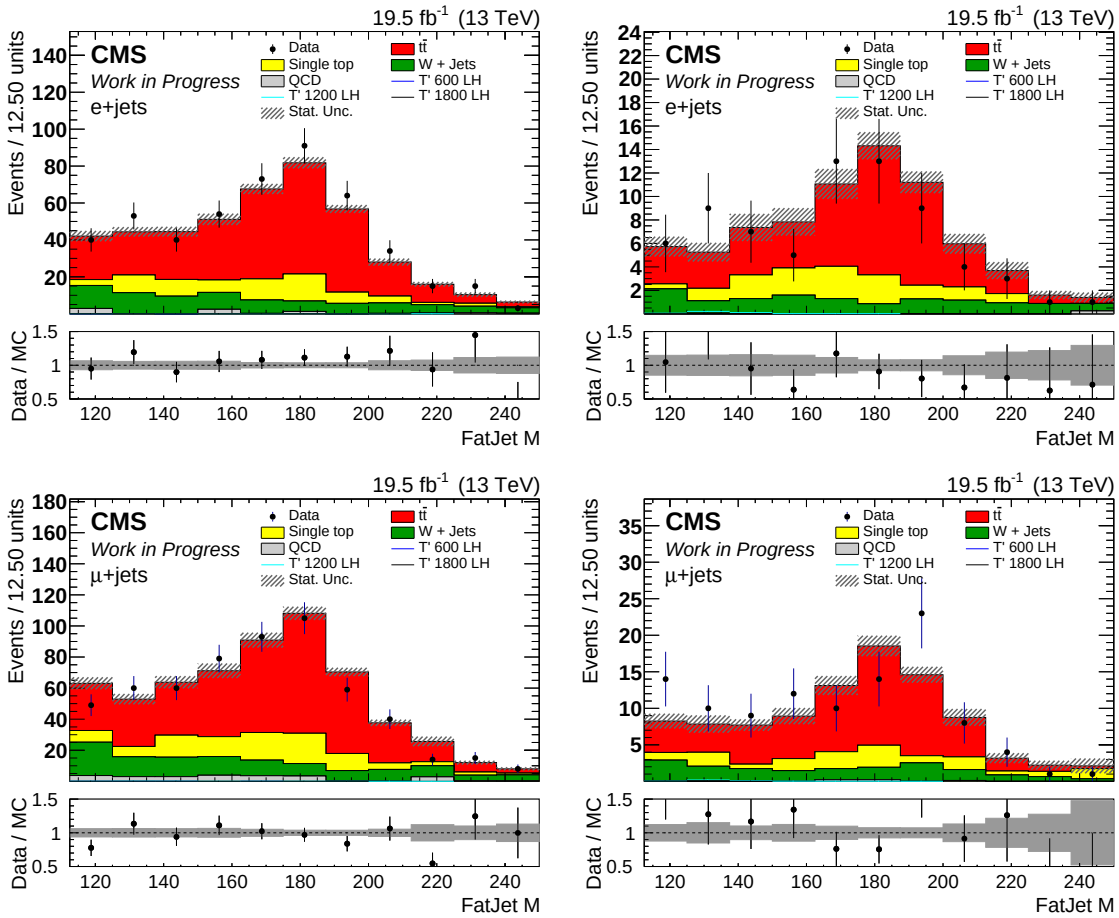

Figure 5.24: Distribution of the ParticleNet mass of the best the AK8 jet after applying the SFs in the 2016preVFP era.

	SF	Δ SF	Backgrounds
FatJet2q1b (pass)	0.83	0.08	$t\bar{t}$, ST
FatJetOther (pass)	1.14	0.17	All
α	1.00	0.11	All

Table 5.9: ParticleNet tagger scale factors and related uncertainties determined from the fit in the control regions in 2016preVFP

In Figure 5.25 the distribution of the AK8 tagged jet ParticleNet mass in the control region AK8 Loose(left), AK8 Tight(right) T and Top Veto, with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.10 the top tagger SFs and their uncertainties for the 2016postVFP era.

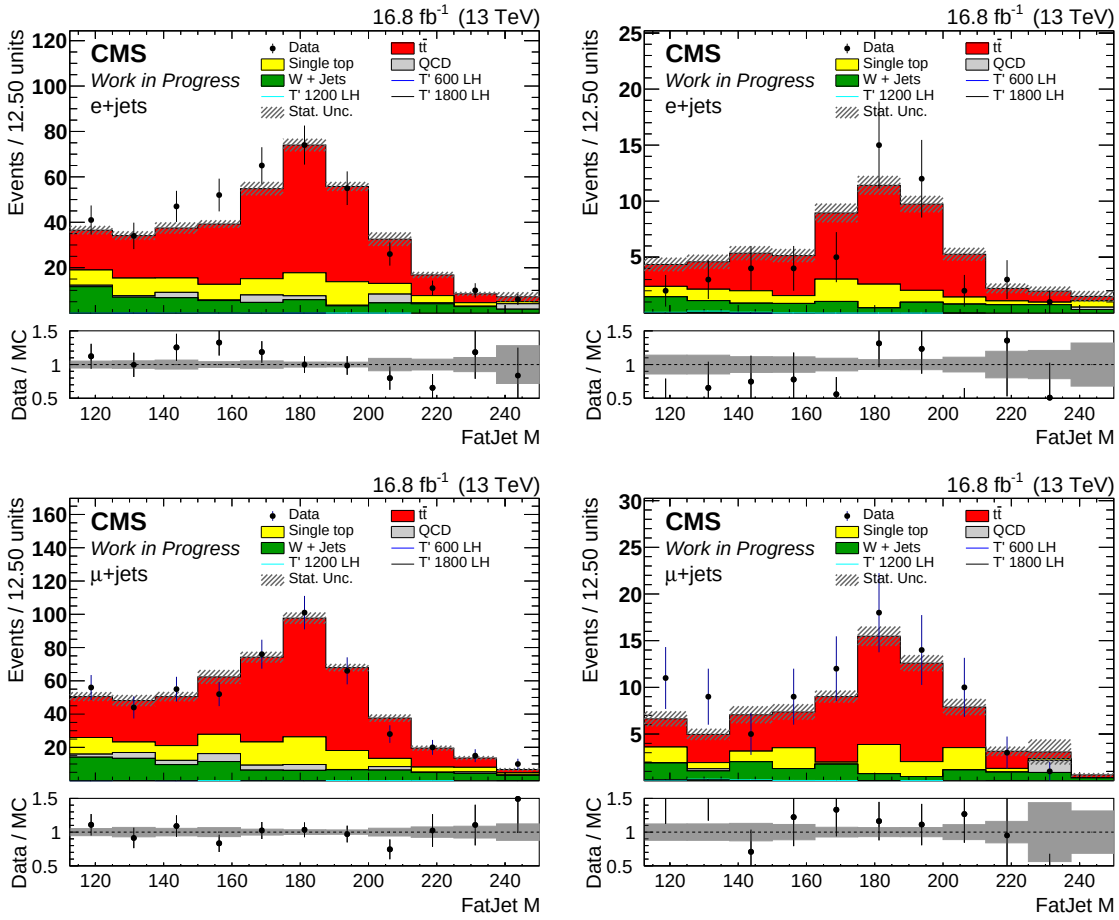

Figure 5.25: Distribution of the ParticleNet mass of the best the AK8 jet after applying the SFs in the 2016postVFP era.

In Figure 5.26 the distribution of the AK8 tagged jet ParticleNet mass in the control region AK8 Loose(left), AK8 Tight(right) T and Top Veto, with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.11 the top tagger SFs and their uncertainties for the 2017 era.

	SF	Δ SF	Backgrounds
FatJet2q1b (pass)	0.88	0.07	$t\bar{t}$, ST
FatJetOther (pass)	1.02	0.12	All
α	1.01	0.08	All

Table 5.10: ParticleNet tagger scale factors and related uncertainties determined from the fit in the control regions in 2016postVFP

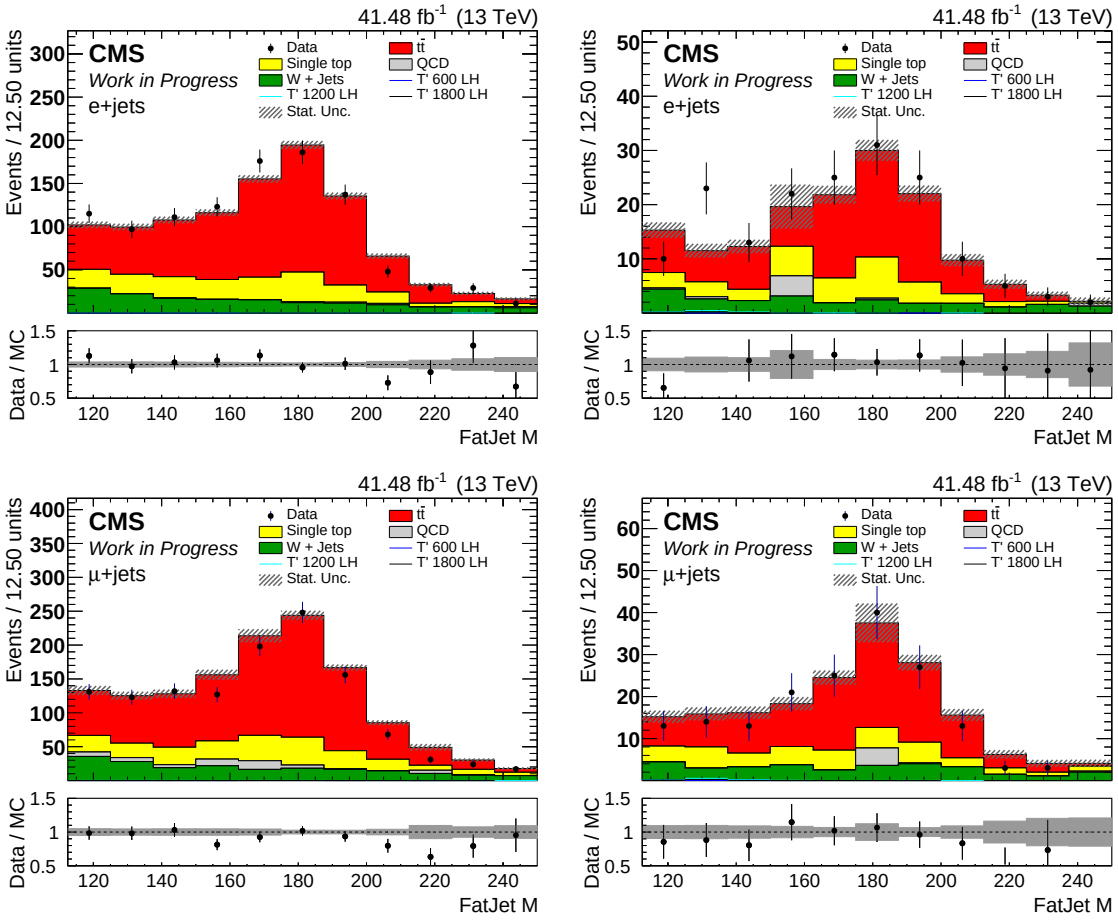

Figure 5.26: Distribution of the ParticleNet mass of the best the AK8 jet after applying the SFs in the 2017 era.

	SF	Δ SF	Backgrounds
FatJet2q1b (pass)	1.05	0.08	$t\bar{t}$, ST
FatJetOther (pass)	1.23	0.15	All
α	1.07	0.09	All

Table 5.11: ParticleNet tagger scale factors and related uncertainties determined from the fit in the control regions in 2017

In Figure 5.27 the distribution of the AK8 tagged jet ParticleNet mass in the control region AK8 Loose(left), AK8 Tight(right) T and Top Veto, with a muon (bottom) or an electron (up) in the finals state. The distribution is shown after the fit procedure. In Table 5.12 the top tagger SFs and their uncertainties for the 2018 era.

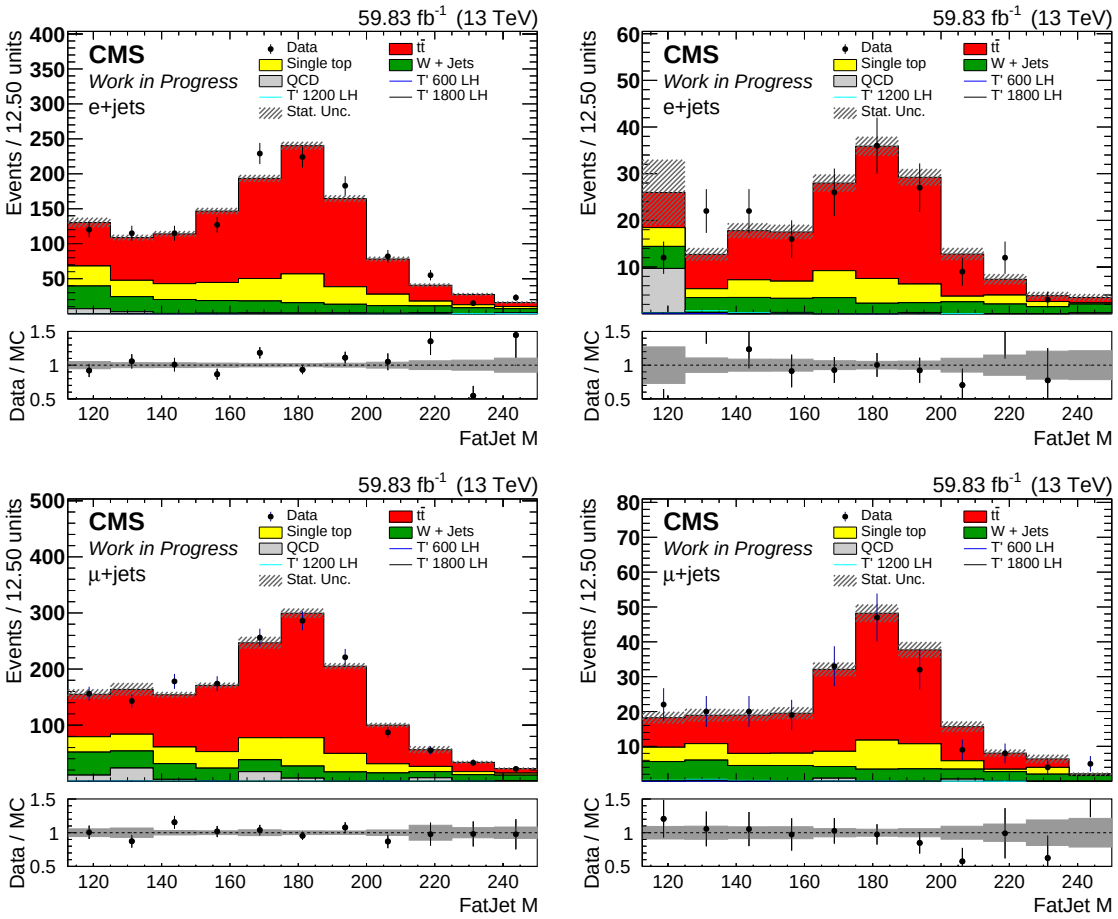

Figure 5.27: Distribution of the ParticleNet mass of the best the AK8 jet after applying the SFs in the 2018 era.

	SF	Δ SF	Backgrounds
FatJet2q1b (pass)	0.98	0.06	$t\bar{t}$, ST
FatJetOther (pass)	1.04	0.10	All
α	0.96	0.06	All

Table 5.12: ParticleNet tagger scale factors and related uncertainties determined from the fit in the control regions in 2018

5.8 Summary on systematic uncertainties

All the corrections described above are applied to MC simulation. In Table 5.13 the list of systematic uncertainties and their correlation across years.

In Figures 5.28, 5.29, 5.30, and 5.31 the ParticleNet mass distribution of the best AK8 tagged jet in the validation regions (TopLoose/Tight,AK8Loose) is shown.

Source	Correlation between years	Process affected
Luminosity	Correlated, Uncorrelated	All
Pile-up	Correlated	All
Simulation statistics	Uncorrelated	All
Factorization and renormalization scales	Correlated	All
Parton Distribution Functions	Correlated	All
Top p_T reweighting	Correlated	All
Pre-firing (ECAL L1 trigger inefficiency)	Correlated	All
Jet energy scale	Uncorrelated	All
Jet energy resolution	Uncorrelated	All
Jet Pile-up Identification	Correlated	All
Muon SF	Correlated, Uncorrelated	All
Electrons SF	Correlated, Uncorrelated	All
Muon Trigger SF	Uncorrelated	All
Electrons Trigger SF	Uncorrelated	All
Top candidate tagging	Uncorrelated	All
β	Uncorrelated	All
AK8 jet tagging	Uncorrelated	All
α	Uncorrelated	All

Table 5.13: List of systematic uncertainties and affected processes.

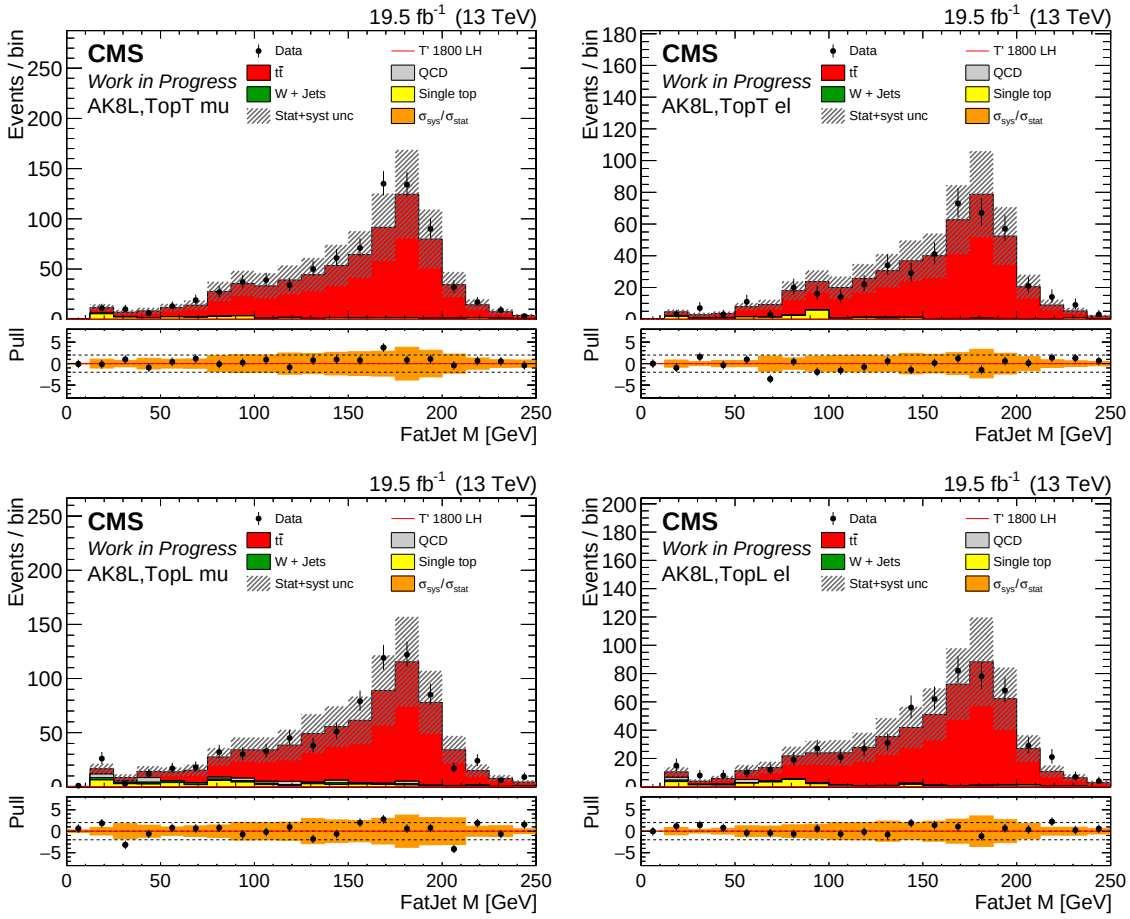

Figure 5.28: Distribution of the ParticleNet mass of the AK8 tagged jet after applying the BDT and ParticleNet SFs in the validation region (TopTight, AK8Loose) (top) and (TopLoose, AK8Loose) (bottom), for muons (left) and electron (right) in 2016preVFP.

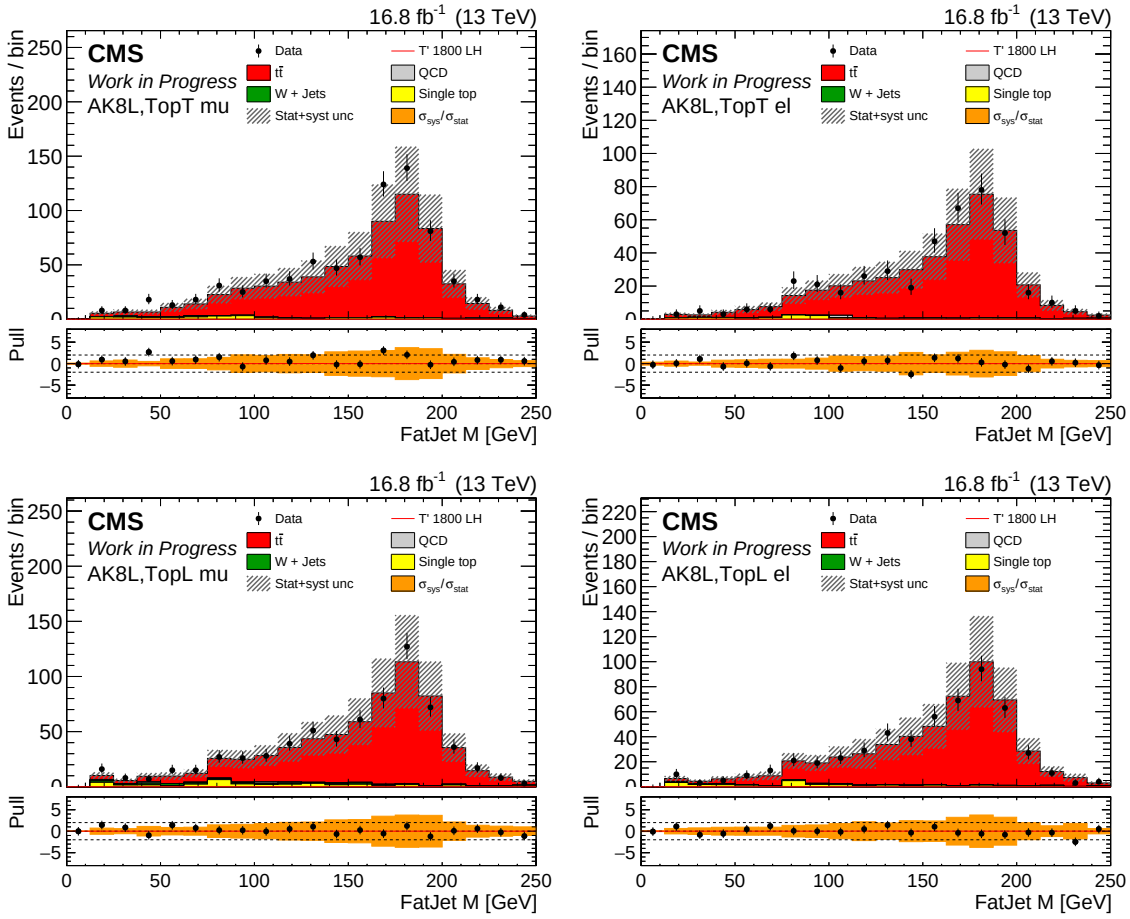

Figure 5.29: Distribution of the ParticleNet mass of the AK8 tagged jet after applying the BDT and ParticleNet SFs in the validation region (TopTight, AK8Loose) (top) and (TopLoose, AK8Loose) (bottom), for muons (left) and electron (right) in 2016postVFP.

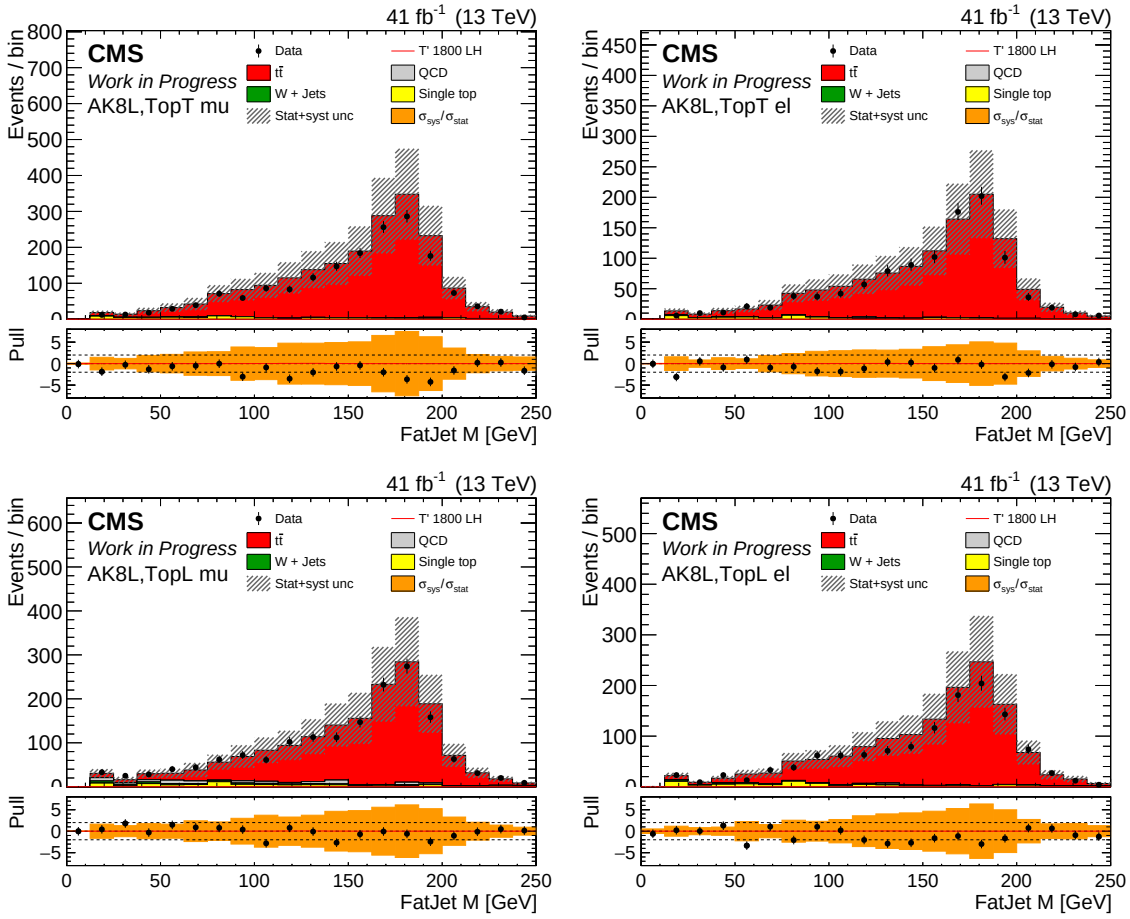

Figure 5.30: Distribution of the ParticleNet mass of the AK8 tagged jet after applying the BDT and ParticleNet SFs in the validation region (TopTight, AK8Loose) (top) and (TopLoose, AK8Loose) (bottom), for muons (left) and electron (right) in 2017.

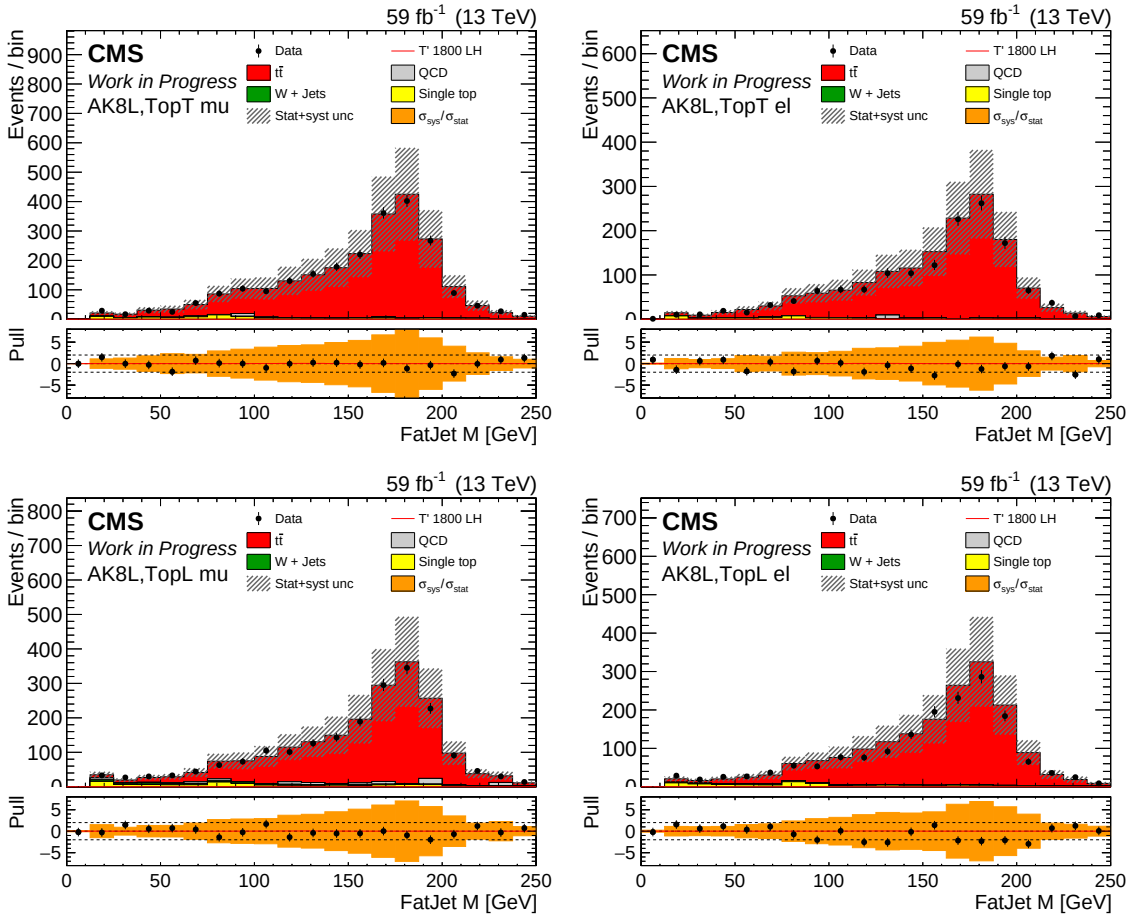

Figure 5.31: Distribution of the ParticleNet mass of the AK8 tagged jet after applying the BDT and ParticleNet SFs in the validation region (TopTight, AK8Loose) (top) and (TopLoose, AK8Loose) (bottom), for muons (left) and electron (right) in 2018.

5.9 Fit Strategy and Results

The invariant mass of the reconstructed T quark candidates shows good discriminating power between signal and backgrounds, motivated by the presence of a resonance in case a new particle is present, that would manifest on top of a smoothly falling distribution for SM processes. For the signal extraction, a simultaneous maximum likelihood fit to the T candidates mass distribution is performed [79, 80]. The regions included in the fit include the presence of either one TopLoose or TopTight, and one AK8Loose or AK8Tight, and are labelled: (TopLoose, AK8Tight), (TopTight, AK8Tight), (TopLoose, AK8Loose), and (TopTight, AK8Loose). The latter two regions are part of the validation region, and are used to control the top quark-antiquark background. Events are further divided considering the presence or not of at least one forward AK4 jet. Events are also split depending on whether a muon or an electron is present. A total of 16 regions is included in the fit.

In Figure 5.32 a scheme of the regions considered for the final fit is shown.

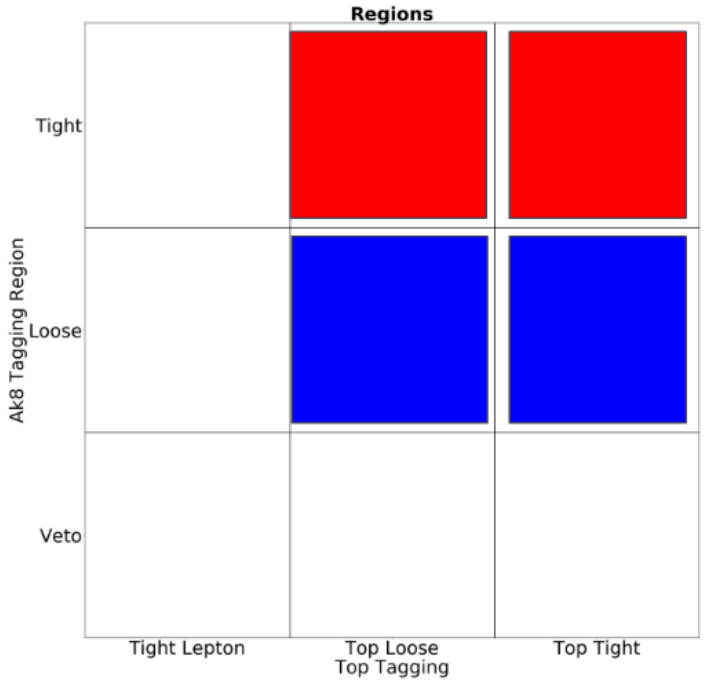

Figure 5.32: Signal(red) and validation(blue) regions used to perform the fit for signal extraction.

Moreover, for the T $\rightarrow$ tH channel, the ParticleNet mass of the best AK8 jet is required to be between 110 and 140 GeV.

Simulated Monte Carlo samples are used for signal and background modelling. All fits are implemented making use of the CMS Combine package [81].

For the fit, two hypotheses have been considered H_0, i.e. absence of new physics, and H_1, i.e. the signal produced by new physics have been detected. Since the number of total events per region is a random variable, thus it follows the Poisson probability distribution, the extended likelihood function is defined as:

$$\begin{aligned}\mathbb{L}(\vec{m}_{T'}; r, \theta) = \prod_{k}^{regions} \mathcal{P}(N_k; \nu_k(\theta)) \prod_{i=1}^{N_k} f_k(m_{T'}; \nu_k, \theta) = \\ = \prod_{k}^{regions} e^{-\nu_k(\theta)} \frac{{\nu_k}^{N_k}(\theta)}{N_k!} \prod_{i_k=1}^{N_k} f_k(m_{T'}; \theta),\end{aligned} \tag{5.11}$$

where k identifies the fitted region, $\mathcal{P}$ is the poisson distribution, $f_k(m'_T; \theta)$ is the probability density function modelled after simulation, ν_k, is the expected number of events in the region K, and the

index i_k runs over the events in the region. The number of events ν_k is given by:

$$\nu_k(\theta) = r \cdot s_k(\theta) + b_k(\theta), \tag{5.12}$$

where $b_k(\theta)$ and $s_k(\theta)$ are respectively the expected number of background and signal events in the region k. A fully floating parameter r represents the signal strength for the considered signal process with the hypothesized masses. In case of a binned maximum likelihood function it is possible to write the likelihood as:

$$\mathbb{L}_{binned}(\vec{m}_{T'}; r, \theta) = \prod_{k}^{regions} \prod_{i=1}^{j_k} \mathcal{P}(N_{i,k}; \nu_{i,k}(\theta)), \tag{5.13}$$

Where j_k is the number of bins of the k-th region, $N_{i,k}$ and $\nu_{i,k}$ are the observed and expected number of events in the i-th bin of the k-th region. As for the total number of events, $\nu_{i,k}$ can be decomposed as:

$$\nu_{i,k}(\theta) = r \cdot s_{i,k}(\theta) + b_{i,k}(\theta). \tag{5.14}$$

All systematic uncertainties described in Sec. 5.5 and the ones related to the Top and ParticleNet taggers are treated as nuisance parameters in the fit. Appropriate nuisance parameters, henceforth referred to as "rate parameters", are introduced as multiplicative factors for the for the $t\bar{t}$ number of events, depending on the Top tagging region, i.e. Loose or Tight, on the lepton in the final state, and on the presence of the forward jet. Each of the rate parameters will act simultaneously on the AK8Loose and AK8Tight regions. If the fit is compatible with the H_0 hypothesis, it is possible to extract the production cross sections upper limit estimation. The limits for the VLQ T production are estimated using the modified frequentist approach called CL_s method [82, 83]. The test statistic used is:

$$\lambda = \frac{\mathcal{L}(m|H_1)}{\mathcal{L}(m|H_0)}, \tag{5.15}$$

that, according to the Neyman-Pearson lemma, is the most powerful discriminator among the two hypotheses, minimizing the error of type-II at a significance level α. The error of type-II represents the probability if H_0 is true of not rejecting H_1. To consider the nuisance parameters of the two hypotheses, the profile likelihood ratio was used, defined as:

$$\lambda(r_{sig}) = \frac{\mathcal{L}(m|r_{sig}, \hat{\hat{\theta}}(r_{sig}))}{\mathcal{L}(m|\hat{r}_{sig}, \hat{\theta})} \tag{5.16}$$

where $\hat{r}_{sig}$ and $\hat{\theta}$ are the best fit values observed from the data sample, while $\hat{\hat{\theta}}$ are the best fit values of the nuisance parameters obtained for a fixed value r_{sig}.
The test statistic, therefore, is:

$$q_r = -2ln(\lambda(r_{sig})), \tag{5.17}$$

if q_r is higher than 0 the events appear to be more compatible with the H_0 hypothesis, while if it's lower than 0, they are more compatible with the H_1 hypothesis. The confidence level for the signal is defined as the ration between the confidence level observed for the H_1 hypothesis, C_{s+b}, and the confidence level observed, CL_b, for the H_0 hypothesis:

$$CL_s = \frac{C_{s+b}}{C_b}. \tag{5.18}$$

CL_b is the probability to have for a given value of r_{sig} a value of the test statistics equal or larger than the value observed in the experiment, under the hypothesis H_0:

$$CL_b = P_b\left(q_r \geq q_r^{obs}\right), \tag{5.19}$$

and CL_{s+b} is defined as the probability to have, for a given value of r, a value of the test statistics equal or larger than the value observed in the experiment, under the hypothesis of signal+background H_1:

$$CL_{s+b} = P_{s+b}\left(q_r \geq q_r^{obs}\right). \tag{5.20}$$

The limits have been computed at 95% CL. Considering s_{obs} the observed signal yield and s_{th} the expected one, using the MC samples, r_{sig} can be seen as:

$$r_{sig} = \frac{s_{obs}}{s_{th}} = \frac{\sigma_{obs}}{\sigma_{th}}, \tag{5.21}$$

therefore the upper limits on r_{sig} can be seen as the upper limits on the ratio $\frac{\sigma_{obs}}{\sigma_{th}}$.

In Figure 5.33 the invariant mass distributions of the reconstructed T quark candidates in the validation regions (TopTight/AK8Loose) with at least one forward jet are shown. In Figure 5.34 the invariant mass distributions of the reconstructed T quark candidates in the signal regions (TopTight/AK8Tight) with at least one forward jet are shown. In Figure 5.35 the invariant mass distributions of the reconstructed T quark candidates in the validation regions (TopTight/AK8Tight) with no forward jet are shown. In Figure 5.36 the invariant mass distributions of the reconstructed T quark candidates in the signal regions (TopTight/AK8Tight) with no forward jet are shown. In these distributions the widths of bins have been chosen in order to ensure enough statistics in each bin.

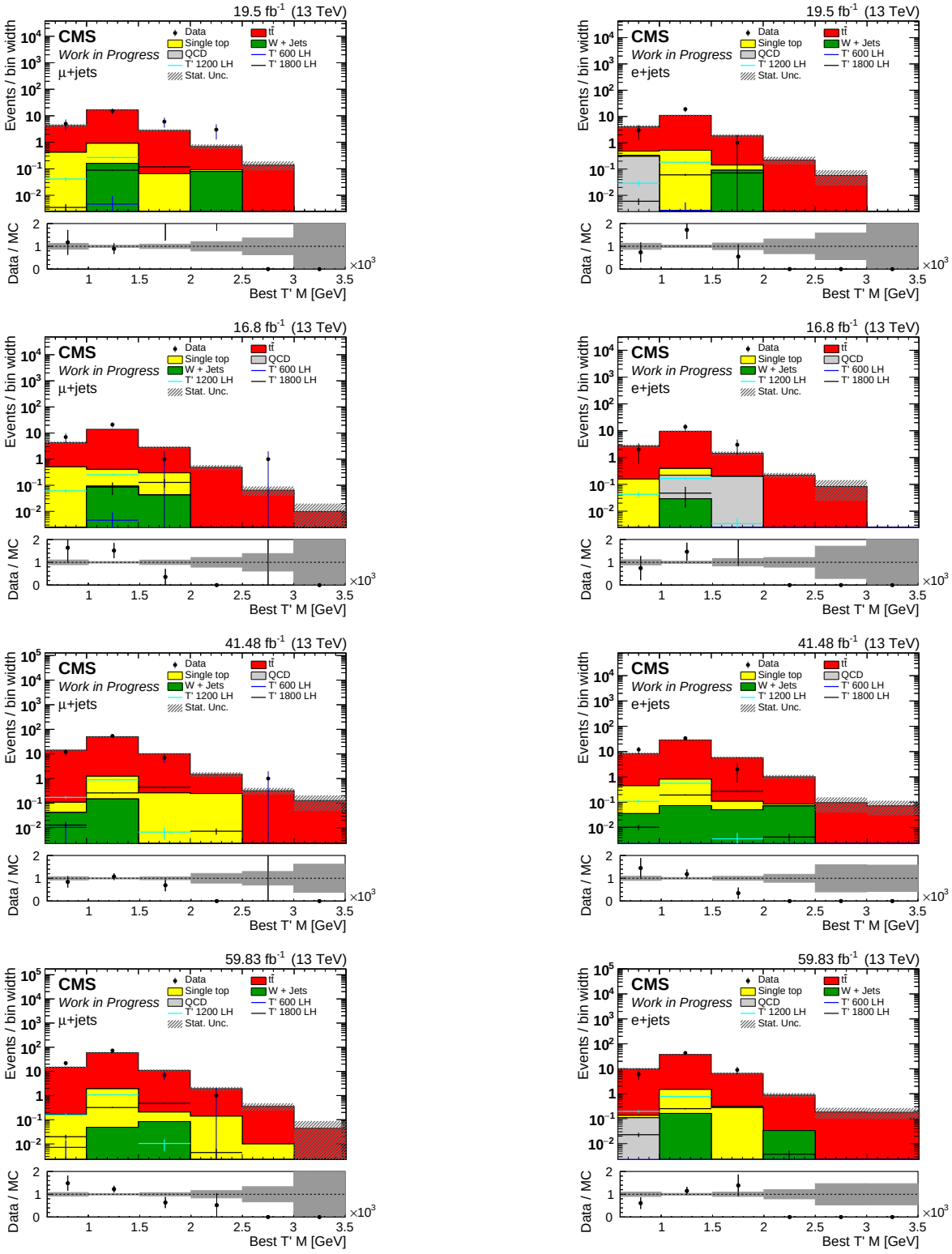

Figure 5.33: The reconstructed T quark candidates mass distribution in the validation region (TopTight/AK8Loose) with at least one forward jet, for 2016(top), 2017(mid), and 2018(bottom) with a muon(left) or an electron(right) in the final state.

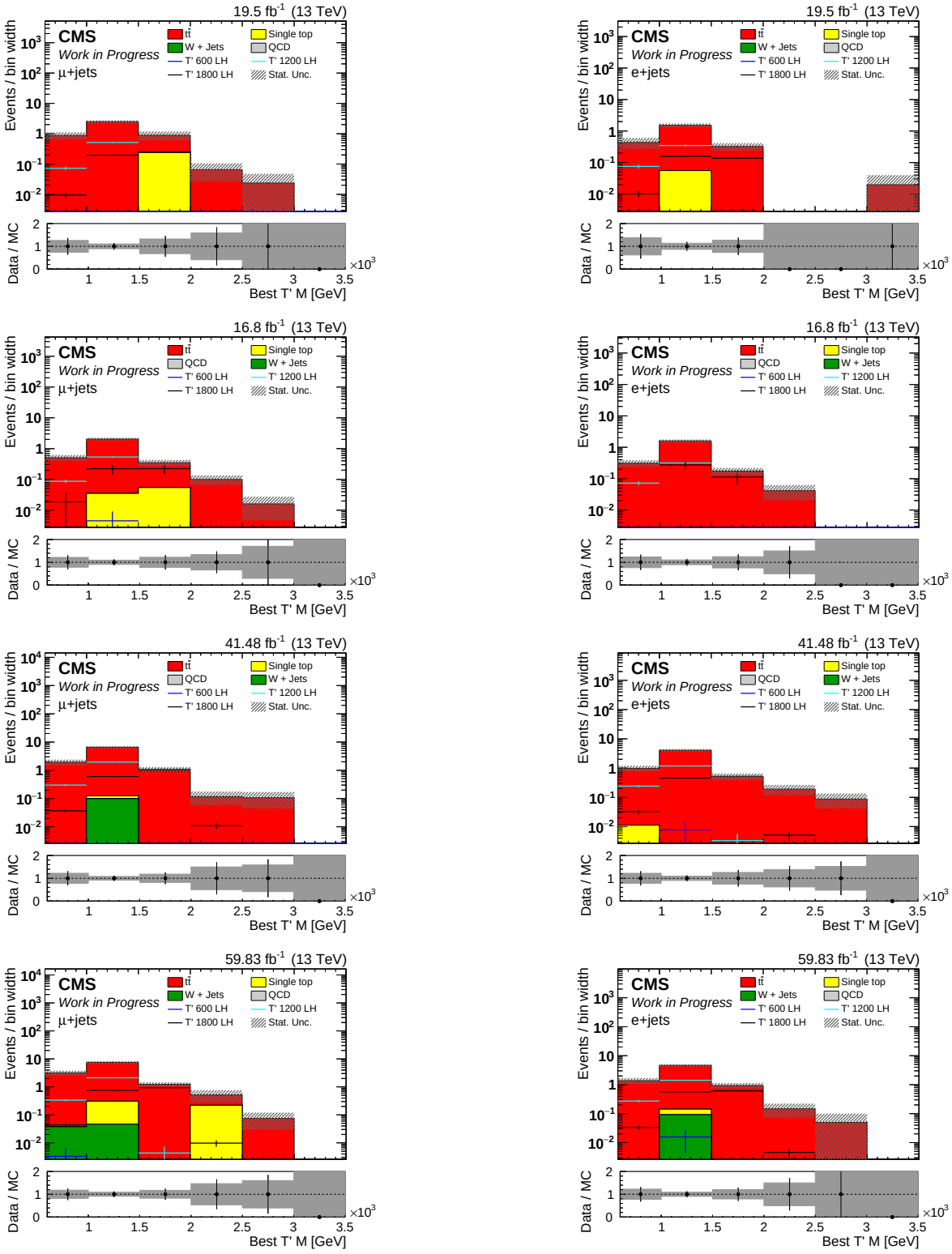

Figure 5.34: The reconstructed T quark candidates mass distribution in the signal region (TopTight/AK8Tight) with at least one forward jet, for 2016preVFP(top), 2016postVFP(mod-top) 2017(mid-bottom), with a muon(left) or an electron(right) in the final state.

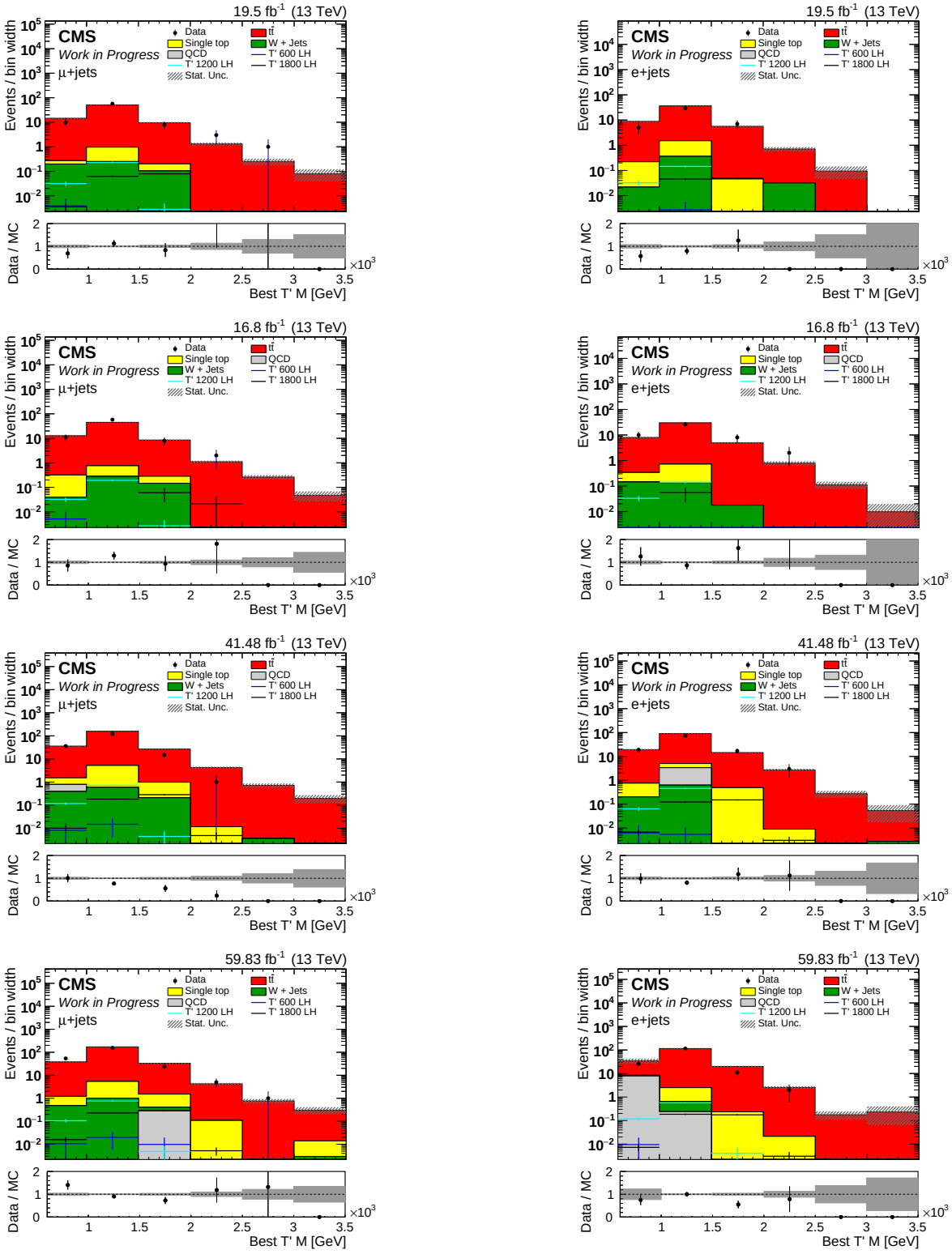

Figure 5.35: The reconstructed T quark candidates mass distribution in the validation region (TopTight/AK8Loose) with no forward jet, for 2016preVFP(top), 2016postVFP(mod-top) 2017(mid-bottom), with a muon(left) or an electron(right) in the final state.

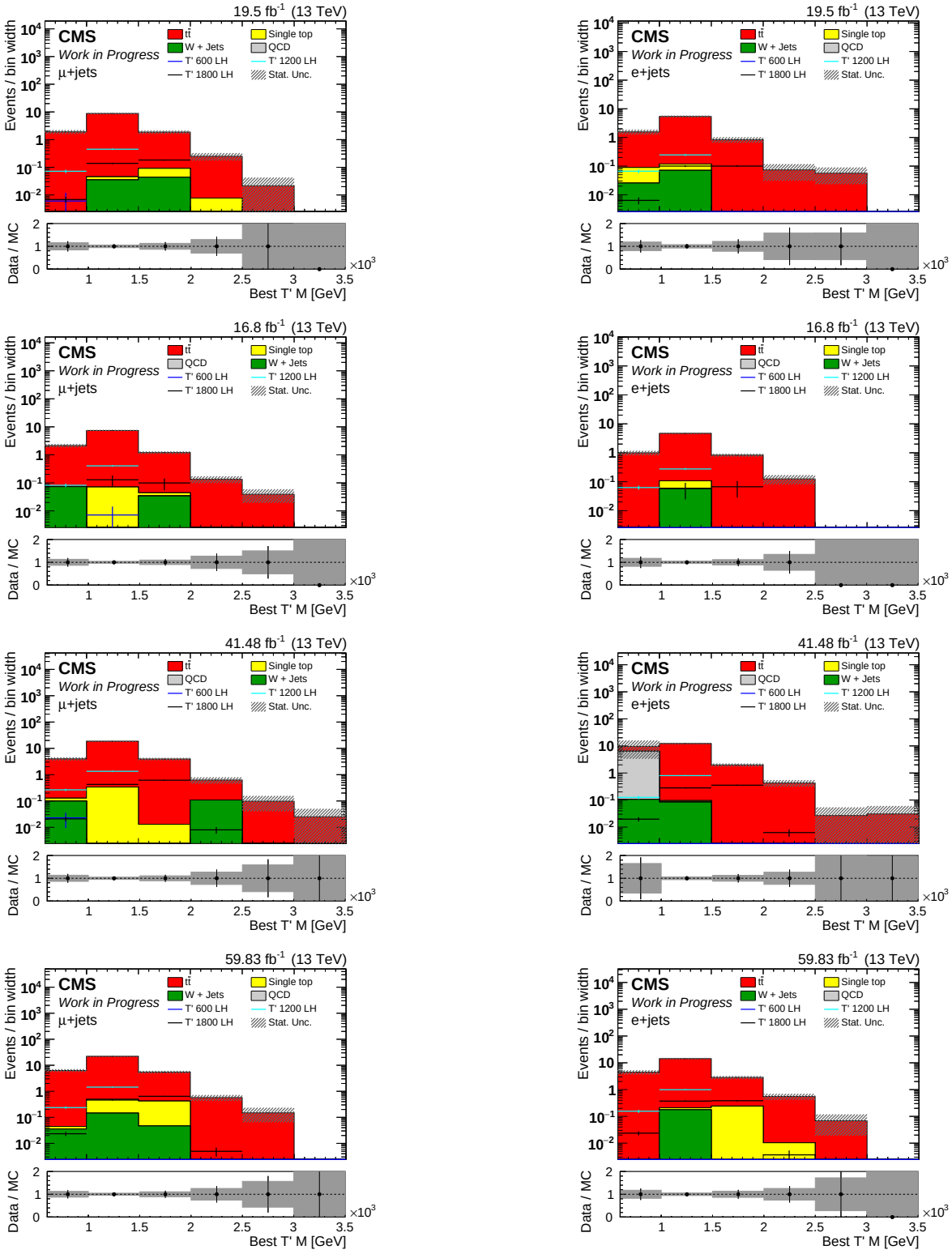

Figure 5.36: The reconstructed T quark candidates mass distribution in the signal region (TopTight/AK8Tight) with no forward jet, for 2016preVFP(top), 2016postVFP(mod-top) 2017(mid-bottom), with a muon(left) or an electron(right) in the final state.

5.9.1 Expected upper limit

The 95 % CL upper limit on the signal cross section times the branching fraction of the T → tH decay channel are estimated. Figure 5.37 shows the 95 % CL upper limit as a function of the T quark mass for the T → tH decay channel. The expected upper limit is really close to the theory cross section for mass hypotheses greater than 1.2 TeV, the observed limit could probe the exclusion of this range of mass if lower than the theory cross section.

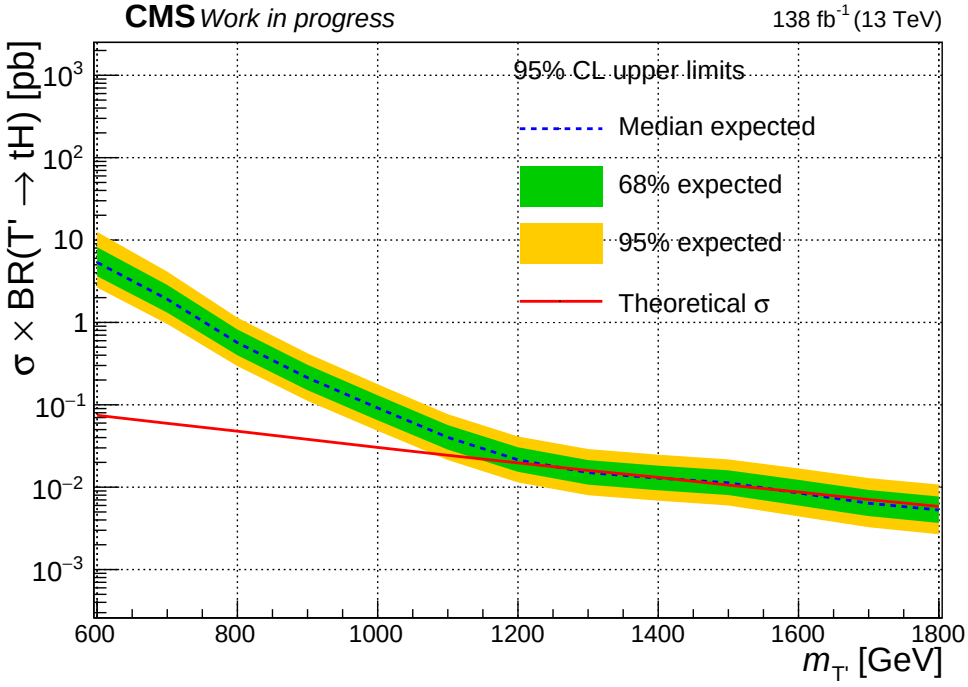

Figure 5.37: 95 % CL upper limits on the cross section times branching fraction for the T → tH channel as function of the T quark mass.

This is the first time this channel is investigated analyzing the Run 2 data, collected at the CMS experiment. The analysis optimization for the boosted regime has allowed greater sensitivities w.r.t. the previous searches. In order to evaluate the robustness of the fit procedure, the impact of the systematic uncertainties are determined. The impacts are evaluated by performing the full fit removing one by one the uncertainties, and evaluating the measured signal strength. The pull gives the constraint on that particular nuisance from data. In Figure 5.38 an example of impact plot obtained fitting a VLQ T mass hypothesis equal to 1.8 TeV.

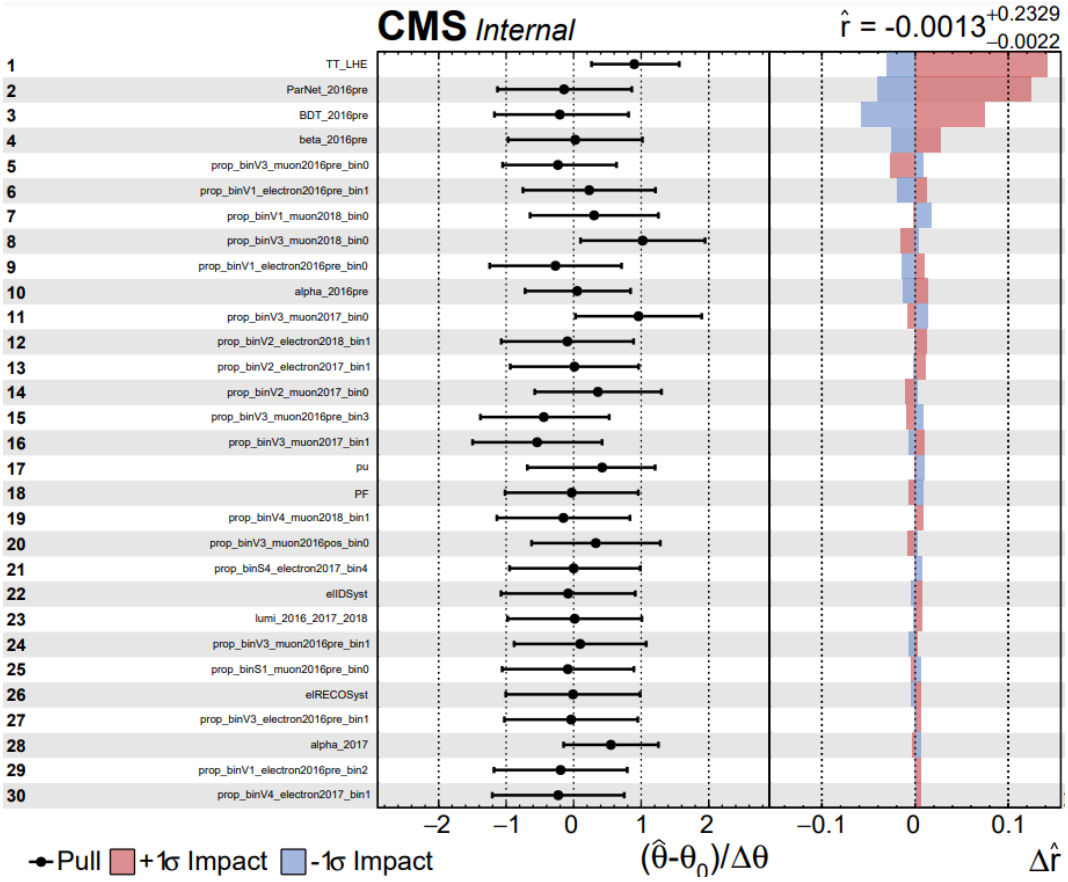

Figure 5.38: An example of impact plot obtained fitting a VLQ T mass hypothesis equal to 1.8 TeV.

For the T $\rightarrow$ tA channel, the fit strategy is the same. However, ParticleNet mass windows are optimized in order to select the 95% of the signal depending on the A mass hypothesis. The mass windows for each A mass hypothesis are shown in Table 5.14

Nominal A boson mass $[GeV]$	Mass window identified $[GeV]$
25	$[20.3, 34.3]$
50	$[40.6, 73.0]$
75	$[59.5, 100.8]$
100	$[77.5, 128.0]$
125	$[97.0, 157.8]$
150	$[119.5, 186.0]$
175	$[139.5, 210.3]$
200	$[159.5, 228.0]$
250	$[194.6, -\infty]$

Table 5.14: Mass windows for each nominal A boson mass hypothesis corresponding to the 95% of selection efficiency: the best AK8 candidate reconstructing the A boson is selected if its mass is in these ranges.

In Figure 5.39 the cross section production times branching ratio expected upper limits as a function of VLQ T mass for different fixed values of A mass hypotheses.

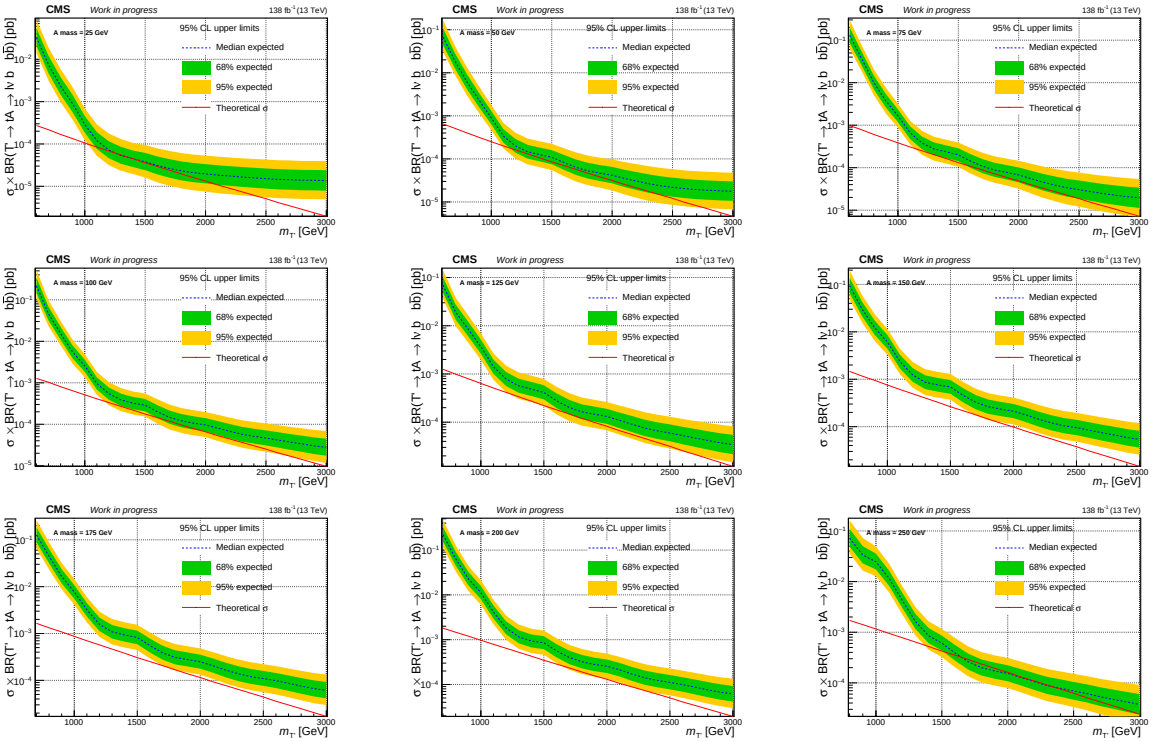

Figure 5.39: Cross section production times branching ration expected upper limits as a function of VLQ T mass for different fixed values of A mass hypotheses.

In Figure 5.40 the cross section production times branching ration expected upper limits as a function of VLQ T mass and boson A mass. This is the first time the VLQ T BSM decay channel to a new hypothetical boson A and top quark has been studied, setting upper limits with high sensitivity for a wide range of A mass hypotheses.

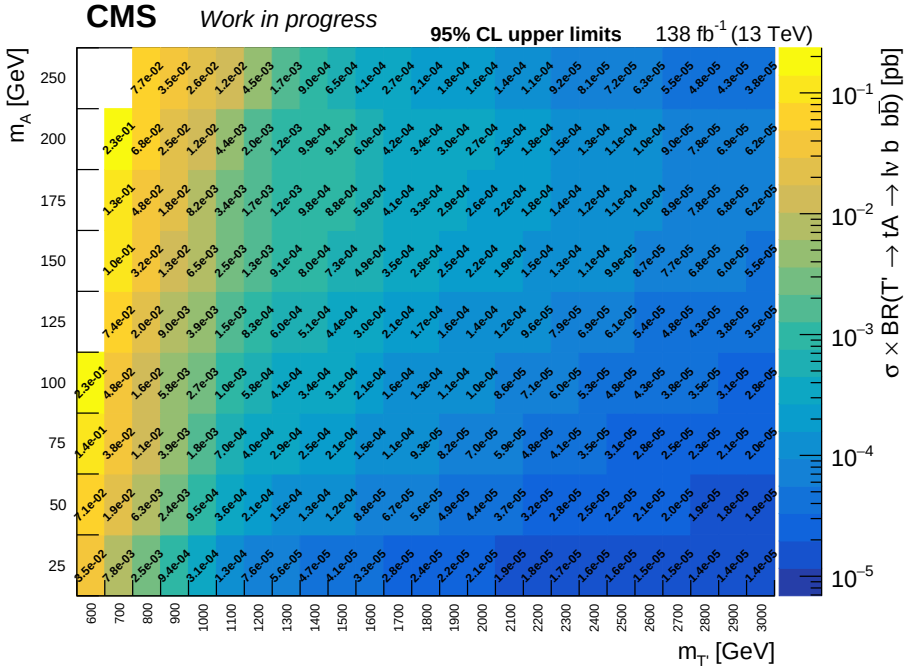

Figure 5.40: Cross section production times branching ration expected upper limits as a function of VLQ T mass and boson A mass.

Conclusions

In this thesis, a search for the singly produced Vector-Like Quark T decaying to top quark and a Higgs boson or new boson, labelled as A, with yet unknown mass, is presented. It has been performed using the data collected by the CMS experiment at LHC during Run 2, with $\sqrt{s} = 13\ TeV$ and an integrated luminosity of 138 fb^{-1}. The final state considered is characterized by the hadronic decay of the boson and the leptonic decay of the top quark. A new identification criteria for the leptonically decaying top quarks has been defined by making use of an appropriately tuned multiclass Machine Learning algorithm. To this end two different configurations have been taken into account depending on angular separation between the lepton and the jet: the Top Merged configuration is characterized by a smaller angular distance than the b-jet radius, otherwise the top quark candidate is classified as Top Resolved. The T candidate was reconstructed with the top quark candidate and the H/A-tagged jet 4-momenta, with the purpose of using the corresponding mass M_T as discriminating variable to find evidence of a potential signal. The signal regions have been chosen considering the number of forward jets, which is expected be equal or greater to one for the signal, the H/A-tagging region, as well as the top quark tagging region. An extended maximum likelihood fit has been performed simultaneously for the distribution of M_T in all the regions and for all the different masses hypothesis of the VLQ T. The considered masses, in natural units, start from 600 GeV up to 1800 GeV, in steps of 100 GeV for the T→tH decay mode. For the other channel VLQ T mass hypotheses are considered up to 3 TeV, in addition several mass hypotheses of the boson A are considered starting from 25 GeV up to 250 GeV. Upper limits on the production cross sections have been estimated at 95% CL. The expected limits for the only background hypothesis are close to the theory cross sections, for the T→tH channel, for mass hypotheses greater than 1.2 TeV. The analysis shows therefore good sensitivity in such mass range, and even in case no signal was observed, upper limit could therefore exclude a wide range of mass for the T. For the T→tA channel, expected cross section upper limits are extracted as function of the VLQ T and the boson A masses for the first time.

Appendix A

Main discriminating variables distributions used in the Top tagging algorithm

Different variables are used as input for the Multiclass BDT algorithm for Top tagging. These are related to the reonctructed top quark candidate and its constituents, i.e. the lepton and the jets. The most discriminating variable distribution are reported here. Distributions have been split for True (red), Other (green), and QCD-like (blue) top quark candidate.

Figure A.1 shows the reconstructed top quark transverse momentum distribution in the merged (top) and resolved (bottom) configuration, with an electron (right) or a muon (left).

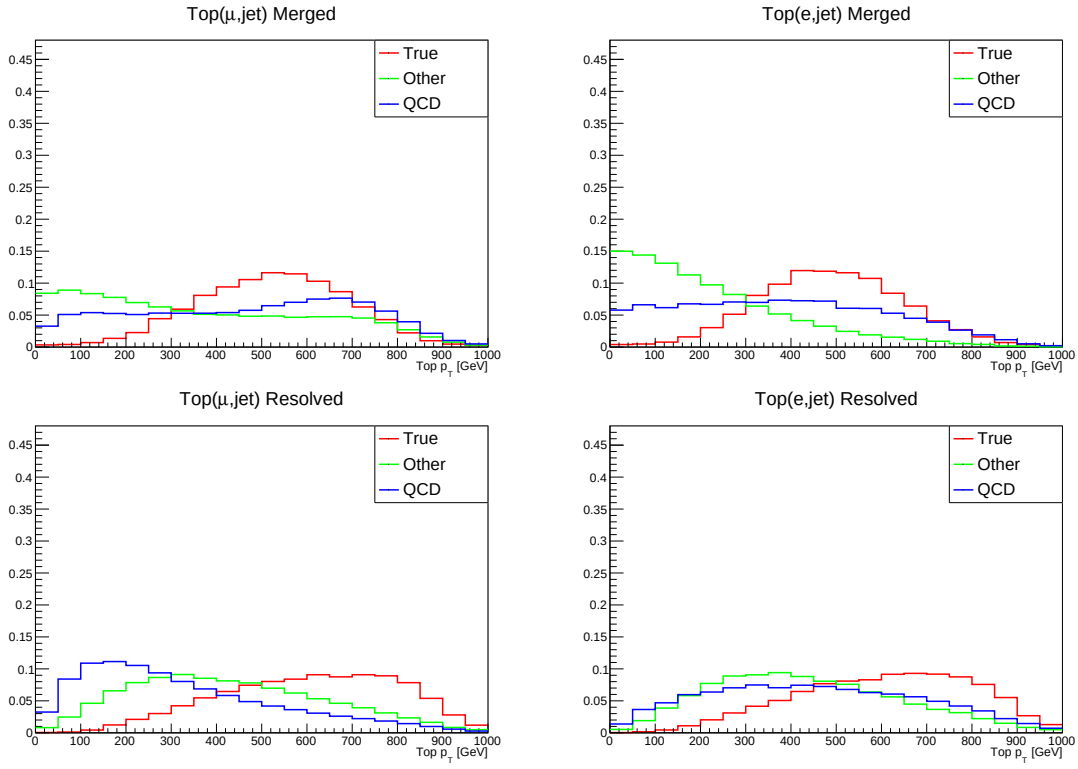

Figure A.1: Merged (top) and Resolved (bottom) top quark transverse momentum distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

Figure A.2 shows the reconstructed top quark mass without MET distribution in the merged (top) and resolved (bottom) configuration, with an electron (right) or a muon (left).

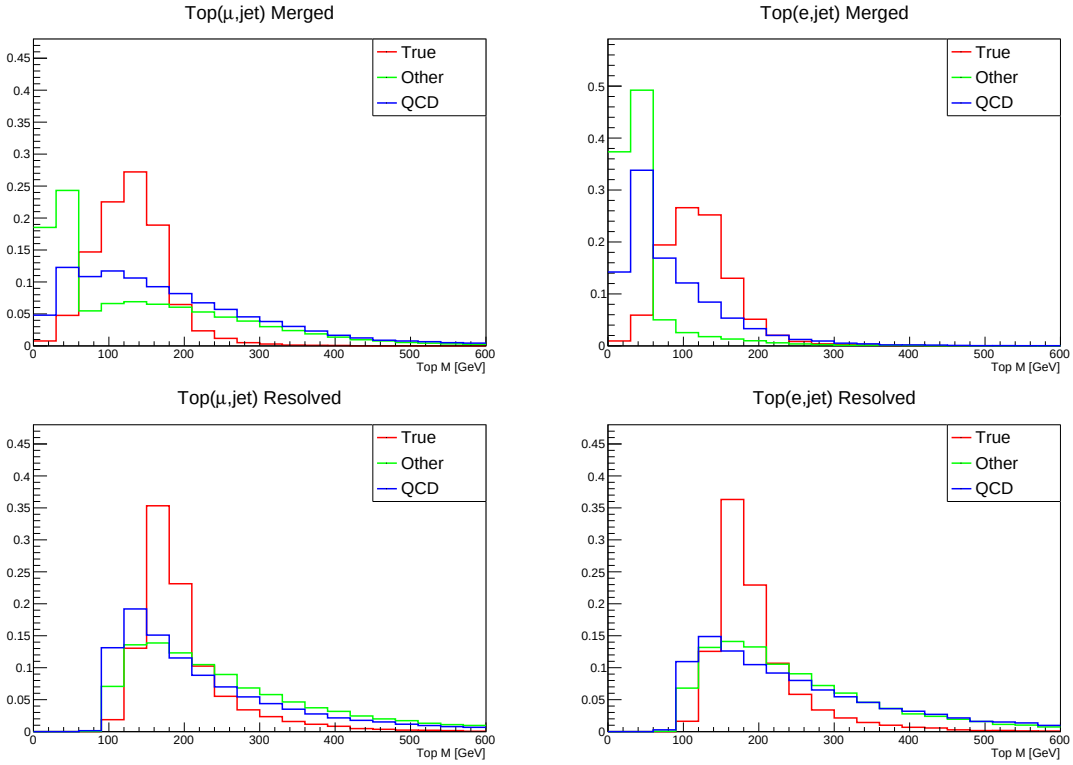

Figure A.2: Merged (top) and Resolved (bottom) top quark reconstructed without MET mass distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

Figure A.3 shows the reconstructed top quark transverse momentum distribution in the merged (top) and resolved (bottom) configuration, with an electron (right) or a muon (left).

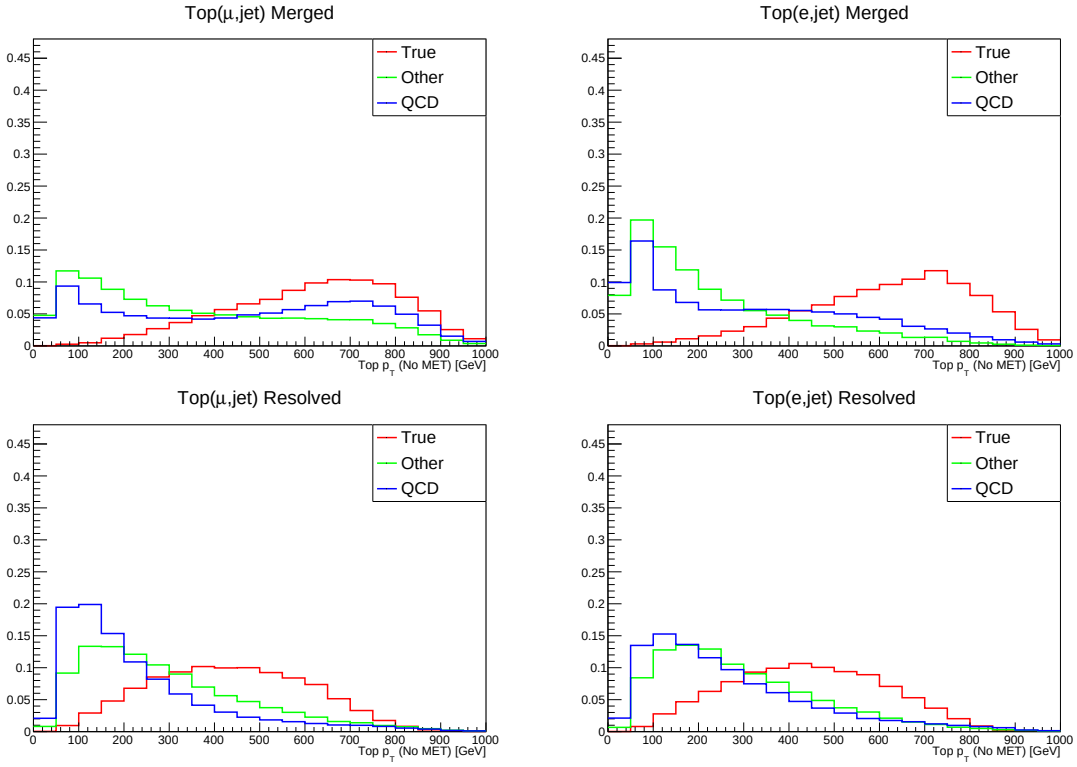

Figure A.3: Merged (top) and Resolved (bottom) top quark transverse momentum distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

Figure A.4 shows the transverse momentum distribution of the top AK4 jet in the merged (top) and resolved (bottom) configuration, with an electron (right) or a muon (left).

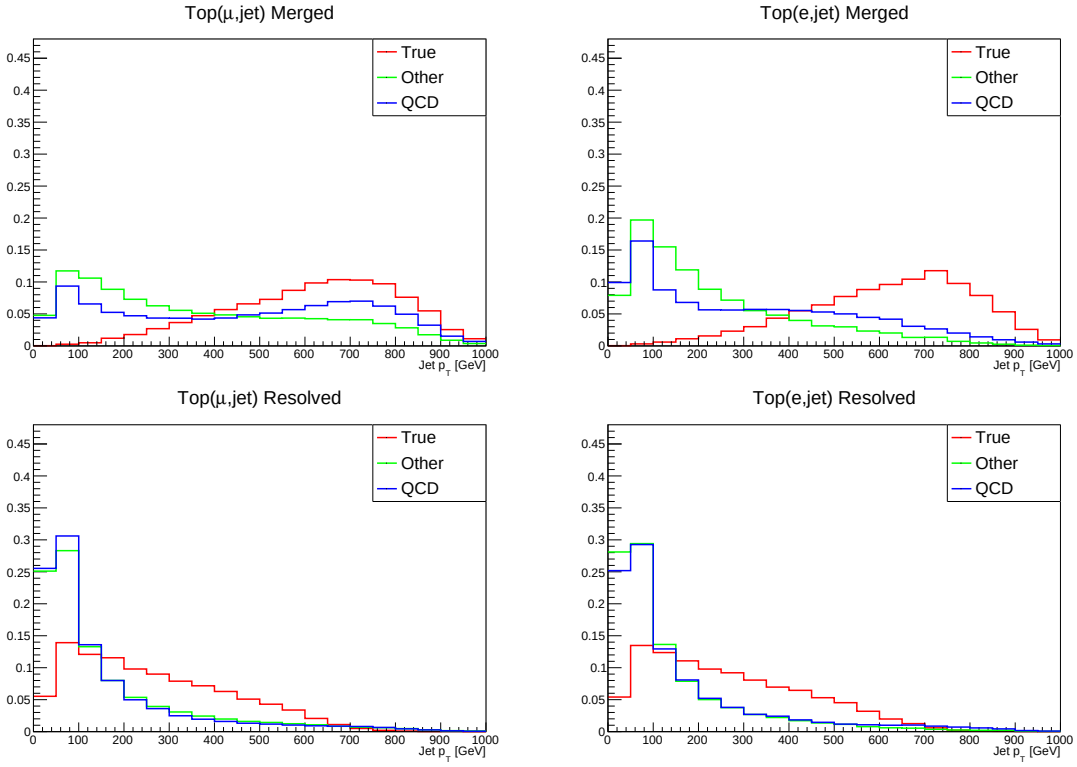

Figure A.4: Merged (top) and Resolved (bottom) transverse momentum of the top AK4 jet distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

Figure A.5 shows the mass distribution of the top AK4 jet in the merged (top) and resolved (bottom) configuration, with an electron (right) or a muon (left).

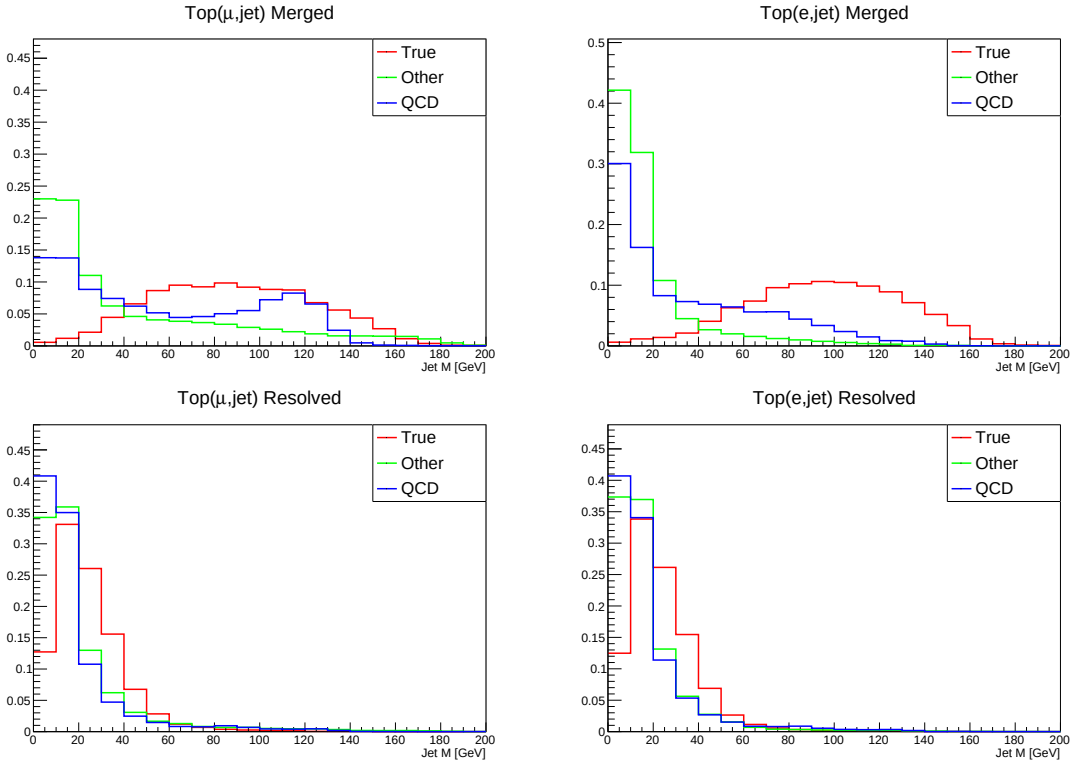

Figure A.5: Merged (top) and Resolved (bottom) mass of the top AK4 jet distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

Figure A.6 shows the DeepJet score distribution of the top AK4 jet in the merged (top) and resolved (bottom) configuration, with an electron (right) or a muon (left).

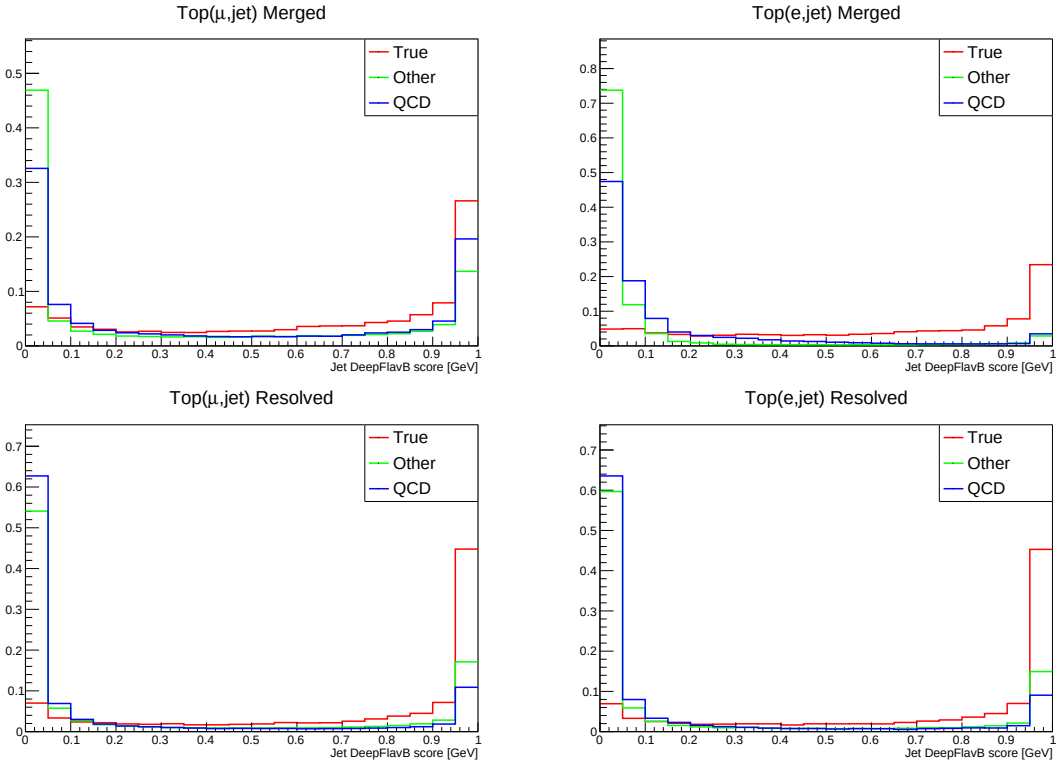

Figure A.6: Merged (top) and Resolved (bottom) DeepJet score of the top AK4 jet distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

Figure A.7 shows the transverse momentum distribution of the top lepton in the merged (top) and resolved (bottom) configuration, with an electron (right) or a muon (left).

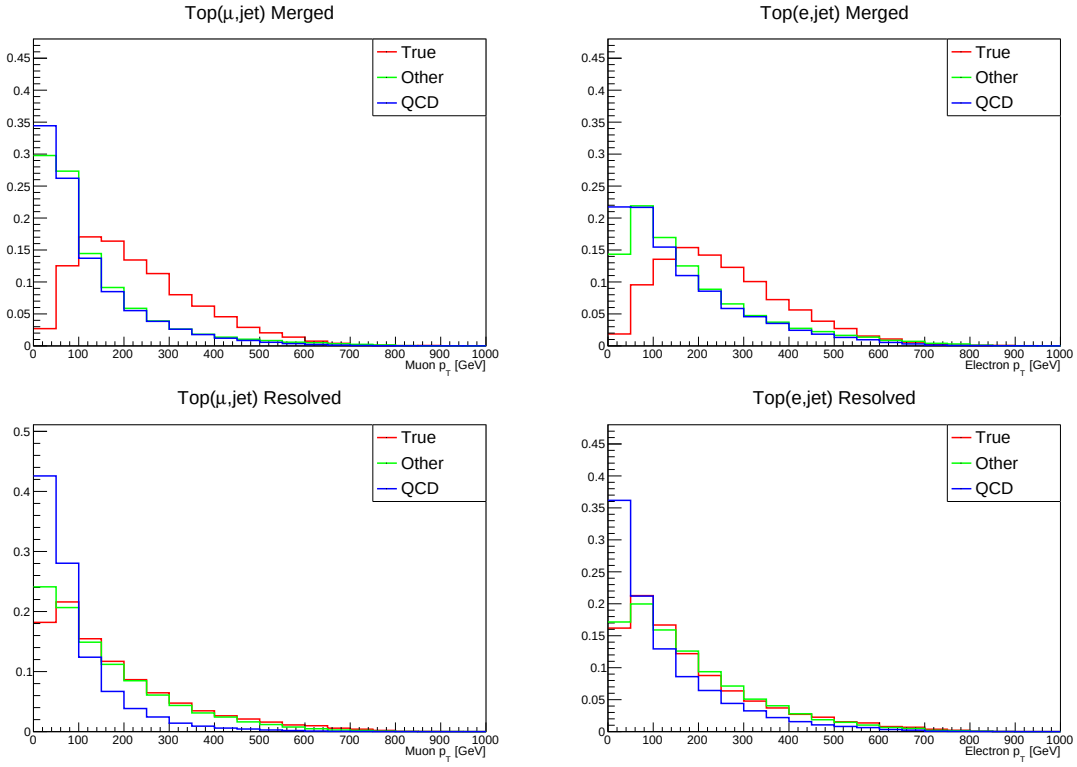

Figure A.7: Merged (top) and Resolved (bottom) transverse momentum of the top lepton jet distribution for True (red), Other (green), or QCD-like (blue) top, with a muon (left) or an electron (right).

Bibliography

[1] Juan Antonio Aguilar-Saavedra et al. "Handbook of vectorlike quarks: Mixing and single production". In: *Physical Review D* 88.9 (2013), p. 094010.

[2] Armen Tumasyan et al. "Search for pair production of vector-like quarks in leptonic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 07 (2023), p. 020. DOI: 10.1007/JHEP07(2023)020. arXiv: 2209.07327 [hep-ex].

[3] S. at al. Abachi. "Observation of the Top Quark". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[4] F. at al. Abe. "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[5] Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[6] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[7] Tomasz Wachala. "T2K Experiment Neutrino Oscillation Results". In: *Acta Phys. Polon. B* 50 (2019), pp. 1757–1764. DOI: 10.5506/APhysPolB.50.1757.

[8] Michael S Turner. "The case for LambdaCDM". In: *arXiv preprint astro-ph/9703161* (1997).

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

[10] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[11] Abdus Salam and John Clive Ward. "Gauge theory of elementary interactions". In: *Phys. Rev.* 136 (1964), B763–B768. DOI: 10.1103/PhysRev.136.B763.

[12] George Zweig. *An SU _3 model for strong interaction symmetry and its breaking.* Tech. rep. CM-P00042884, 1964.

[13] Peter Ware Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133.

[14] Richard Feynman. *Feynman lectures on gravitation.* CRC Press, 2018.

[15] E FERMI. "Tentativo di una teoria dell'emissione dei raggi «Beta». ricerca scientifica, v. 4". In: (1933).

[16] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[17] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[18] Kenzo Nakamura. "Review of particle physics". In: *Journal of Physics G: Nuclear and Particle Physics* 37.7 A (2010).

[19] Rajan Gupta. "Introduction to lattice QCD: Course". In: *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*. July 1997, pp. 83–219. arXiv: hep-lat/9807028.

[20] Vera C. Rubin and W. Kent Ford Jr. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *Astrophys. J.* 159 (1970), pp. 379–403. DOI: 10.1086/150317.

[21] SL Glashow, J Iliopoulos, and I Maiani. "GLASHOW 1970". In: *Phys. Rev* 2 (1970), p. 1285.

[22] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[23] "LHC Machine". In: *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[24] G. Arnison et al. "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider". In: *Phys. Lett. B* 126 (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0.

[25] G. Arnison et al. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV". In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.

[26] Esma Mobs. "The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019". In: (2019). General Photo. URL: https://cds.cern.ch/record/2684277.

[27] Jorg Wenninger. "SISSA: LHC status and performance". In: *PoS* (2019), p. 001.

[28] CERN. *LHC Operation section - operating the LHC at peak energy and luminosity*. 2024. URL: https://be-dep-op-lhc.web.cern.ch/.

[29] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[30] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[31] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[32] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[33] W. Adam et al. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16.02 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304 [physics.ins-det].

[34] Cms Collaboration et al. "The CMS electromagnetic calorimeter project: technical design report". In: *Technical Design Report CMS. CERN, Geneva* 47 (1997).

[35] "The CMS hadron calorimeter project: Technical Design Report". In: (1997).

[36] JG Layter et al. "The CMS muon project: Technical design report". In: *Technical design report. CMS. CERN, Geneva* 38 (1997).

[37] M. A. Akl et al. "CMS Technical Design Report for the Muon Endcap GEM Upgrade". In: (June 2015). Ed. by A. Colaleo et al.

[38] Attila Rácz and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger*. Tech. rep. CMS-TDR-006, 2002.

[39] M. Abbas et al. "Quality control of mass-produced GEM detectors for the CMS GE1/1 muon upgrade". In: *Nucl. Instrum. Meth. A* 1034 (2022), p. 166716. DOI: 10.1016/j.nima.2022.166716. arXiv: 2203.12037 [physics.ins-det].

[40] M. Abbrescia et al. "Beam test results on double-gap resistive plate chambers proposed for CMS experiment". In: *Nucl. Instrum. Meth. A* 414 (1998), pp. 135–148. DOI: 10.1016/S0168-9002(98)00571-3.

[41] David B Kaplan, Howard Georgi, and Savas Dimopoulos. "Composite higgs scalars". In: *Physics Letters B* 136.3 (1984), pp. 187–190.

[42] Kaustubh Agashe, Roberto Contino, and Alex Pomarol. "The minimal composite Higgs model". In: *Nuclear Physics B* 719.1-2 (2005), pp. 165–187.

[43] Giuliano Panico, Andrea Wulzer, et al. *The composite nambu-goldstone higgs*. Vol. 913. Springer, 2016.

[44] Paolo Lodone. "Vector-like quarks in a"composite"Higgs model". In: *Journal of High Energy Physics* 2008.12 (2008), p. 029.

[45] ALEPH collaboration et al. "Precision electroweak measurements on the Z resonance". In: *arXiv preprint hep-ex/0509008* (2005).

[46] Avik Banerjee et al. "Phenomenological aspects of composite Higgs scenarios: exotic scalars and vector-like quarks". In: (Mar. 2022). arXiv: 2203.07270 [hep-ph].

[47] Th Kaluza. "Zum unitätsproblem der physik". In: *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* 1921.arXiv: 1803.08616 (1921), pp. 966–972.

[48] Oskar Klein. "Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English)". In: *Z. Phys.* 37 (1926). Ed. by J. C. Taylor, pp. 895–906. DOI: 10.1007/BF01397481.

[49] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. "The Hierarchy problem and new dimensions at a millimeter". In: *Phys. Lett. B* 429 (1998), pp. 263–272. DOI: 10.1016/S0370-2693(98)00466-3. arXiv: hep-ph/9803315.

[50] Lisa Randall and Raman Sundrum. "Large mass hierarchy from a small extra dimension". In: *Physical review letters* 83.17 (1999), p. 3370.

[51] Lisa Randall and Raman Sundrum. "An alternative to compactification". In: *Physical Review Letters* 83.23 (1999), p. 4690.

[52] G.C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Physics Reports* 516.1 (2012). Theory and phenomenology of two-Higgs-doublet models, pp. 1–102. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2012.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0370157312000695.

[53] Rachid Benbrik et al. "Signatures of vector-like top partners decaying into new neutral scalar or pseudoscalar bosons". In: *JHEP* 05 (2020), p. 028. DOI: 10.1007/JHEP05(2020)028. arXiv: 1907.05929 [hep-ph].

[54] Georges Aad et al. "Search for single production of vector-like T quarks decaying into Ht or Zt in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 08 (2023), p. 153. DOI: 10.1007/JHEP08(2023)153. arXiv: 2305.03401 [hep-ex].

[55] Georges Aad et al. "Search for pair-produced vector-like top and bottom partners in events with large missing transverse momentum in pp collisions with the ATLAS detector". In: *Eur. Phys. J. C* 83.8 (2023), p. 719. DOI: 10.1140/epjc/s10052-023-11790-7. arXiv: 2212.05263 [hep-ex].

[56] Armen Tumasyan et al. "Search for a vector-like quark $T' \to tH$ via the diphoton decay mode of the Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 09 (2023), p. 057. DOI: 10.1007/JHEP09(2023)057. arXiv: 2302.12802 [hep-ex].

[57] J. A. Aguilar-Saavedra et al. "Handbook of vectorlike quarks: Mixing and single production". In: *Phys. Rev. D* 88.9 (2013), p. 094010. DOI: 10.1103/PhysRevD.88.094010. arXiv: 1306.0572 [hep-ph].

[58] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[59] Rudolph Emil Kalman. "A new approach to linear filtering and prediction problems". In: (1960).

[60] W. Adam et al. "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC". In: *eConf* C0303241 (2003), TULT009. DOI: 10.1088/0954-3899/31/9/N01. arXiv: physics/0306087.

[61] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[62] Daniele Bertolini et al. "Pileup per particle identification". In: *Journal of High Energy Physics* 2014.10 (2014), pp. 1–22.

[63] *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256875.

[64] Emil Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (2020), P12012.

[65] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph].

[66] Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[67] CMS Collaboration et al. *Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data*. Tech. rep. CMS Detector Performance Note CMS-DP-2020-045, 2020, URL https://cds.cern ..., 2020.

[68] Sea Agostinelli et al. "GEANT4—a simulation toolkit". In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303.

[69] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[70] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[71] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[72] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[73] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[74] "Mass regression of highly-boosted jets using graph neural networks". In: (2021). URL: https://cds.cern.ch/record/2777006.

[75] Albert M Sirunyan et al. "Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons". In: *JHEP* 02 (2019), p. 149. DOI: 10.1007/JHEP02(2019)149. arXiv: 1811.06625 [hep-ex].

[76] Vardan Khachatryan et al. "Measurement of differential cross sections for top quark pair production using the lepton+ jets final state in proton-proton collisions at 13 TeV". In: *Physical Review D* 95.9 (2017), p. 092001.

[77] Jon Butterworth et al. "PDF4LHC recommendations for LHC run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (2016), p. 023001.

[78] R. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), p. 219. DOI: 10.1016/0010-4655(93)90005-W.

[79] J. S. Conway. "Incorporating nuisance parameters in likelihoods for multisource spectra". In: *Proc. workshop on statistical issues related to discovery claims in search experiments and unfolding (PHYSTAT 2011)*. 2011, p. 115. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[80] Luca Lista. *Statistical Methods for Data Analysis*. Lecture Notes in Physics. Switzerland: Springer, 2023. DOI: 10.1007/978-3-031-19934-9.

[81] *The CMS Combine Package*. URL: https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/.

[82] Alexander L. Read. "Presentation of search results: The CL_s technique". In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[83] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].