UNIVERSITÀ DEGLI STUDI DI NAPOLI
"FEDERICO II"

Scuola Politecnica e delle Scienze di Base

Area Didattica di Scienze Matematiche Fisiche e Naturali

Dipartimento di Fisica "Ettore Pancini"

Laurea Magistrale in Fisica

Search for single vector like quark T at CMS with machine learning techniques.

Relatore:
Dr. Alberto Orso Maria Iorio
Correlatore:
Prof. Guglielmo De Nardo

Candidato:
Francesco Carnevali
Matr. N94000484

Anno Accademico 2018/2019

Ai miei genitori.

Contents

Introduction

The Standard Model (SM) of particle physics is the most successful quantum field theory 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. The SM however appears as an incomplete theory, as it does not include the gravitation, the existence of Dark Matter and a large set of parameter is not postulated by the theory. In order to solve some issues, like the so-called hierarchy problem, several theories have supposed the existence of new particles, the Vector-Like Quarks (VLQs) with masses at the TeV scale. 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), that thanks to its high centre-of-mass energy of $\sqrt{s} = 13\ TeV$ and the instantaneous luminosity of $10^{34}\ cm^{-2}\ s^{-1}$, could find evidences for VLQs.
In this thesis is presented the search for a singly produced VLQ T, decaying to a top quark and a Z boson. The considered final state is characterized by the hadronic decay of the Z boson and the leptonic decay of the top quark, $t \to \mu\nu_\mu b$. To improve the selection 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, but this is the first time it is studied in depth for leptonic top quark final states.
To perform the analysis the data sample collected by CMS in 2018 is considered, and the simulations of the main background from the SM and of the searched signal, considering different mass hypotheses.
This thesis is organized as follow:

- Chapter 1, an introduction of the Standard Model and the unsolved problems.
- Chapter 2, the description of the LHC accelerator and of the CMS experiment.
- Chapter 3, a brief description of the different theories that predicts the existence of the VLQs and of the different searches for the VLQs in progress.

- Chapter 4, the selection of the objects in the final state, focusing on the top quark reconstruction.
- Chapter 5, 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 describes three of the four fundamental forces, the electromagnetic, weak, and strong interactions. This quantum field theory provides precise predictions of the fundamental particles and their interactions, and has been confirmed by the experimental evidence collected throughout the past century in particle physics.
The SM, nevertheless, doesn't include the gravitation, doesn't predict neutrino masses, that are confirmed by neutrino oscillation experiment [47], and the existence of dark matter and the dark energy,that is predicted by the ΛCDM cosmological model [46].

1.1 Standard Model overview

The SM is the quantum field theory that describes the electromagnetic and the weak interactions, unified in the GWS theory, and the strong interaction, throught the Quantum Chromodynamics (QCD). The SM provides masses thanks to the the Higgs mechanism.
The dynamics equations for the SM are obtained from a gauge principle: a 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 $U(1)_Y$, 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 $SU(2)_L$, the 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 $SU(3)_C$, the 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 $G_\mu^{1,\ldots,8}$;

To each interaction an absolute conserved quantum number and a multiplet are associated. The components of such multiplet are called bosons because they obey to Bose-Einstein statistics and they have spin equals to one. The number of the mediator bosons is determined by the number of the generator of the symmetry group associated to each interaction. The SM has 12 vector fields associated to the three gauge symmetries. A list of the physical bosons is reported in Table 1.1.

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.

The SM predicts 12 fields with an half-integer spin, that obey to the Fermi-Dirac statistics and they are called fermions. Fermions are divided into lepton and quark fields, both lepton and quark fields are listed in three families or generations. Each generation is a doublet of particles associated to an isospin quantum number and each particle carries electroweak charge, but only the quarks have strong charge. In Table 1.2 is reported a list of SM leptons and quarks.

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.

Each lepton and quark has an associated anti-particle with the same mass but opposite quantum numbers.
The last field in the Standard Model is a complex scalar doublet field ϕ, named the Higgs field by the theorists who predicted its existence in 1964 [23]. In Figure 1.1 the particles predicted by the SM are shown.

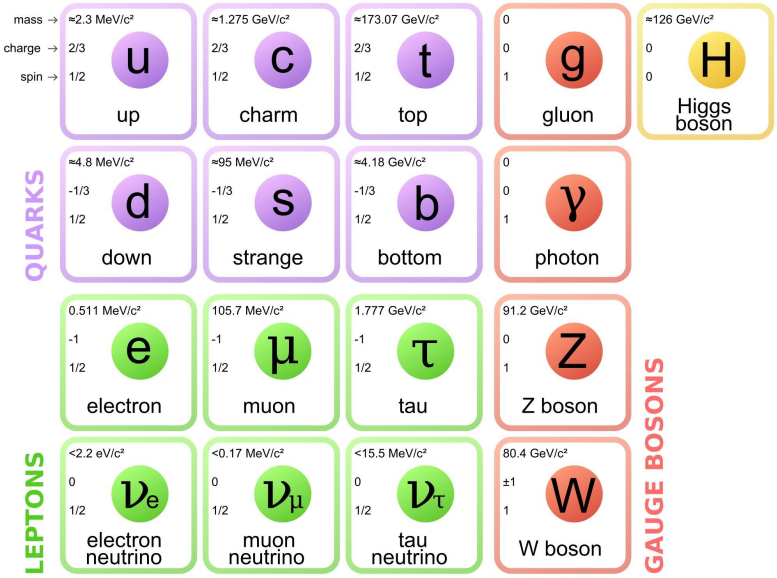


Figure 1.1: Standard Model particles.

1.2 Quantum electrodynamics

The Quantum Electrodynamics (QED) is the relativistic quantum field theory that describes the dynamics between fermions in an electromagnetic field. The symmetry group associated to the electromagnetic interaction is the abelian group $U(1)_q$, where q is the conserved quantum number, that is the charge of the particles.
The QED lagrangian density is obtainable 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.
To include the free photons propagation, the term to the boson kinetic energy is added to the equation 1.1:

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

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 new 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. The fermionic kinetic term is not invariant under local gauge transformation:

$$\psi \to \psi' = e^{i\theta(x)}\psi$$

$$\bar{\psi} \to \bar{\psi}' = e^{-i\theta(x)}\bar{\psi}$$

$$\partial_\mu\psi \to \partial_\mu\psi' = e^{iq\theta(x)}\partial_\mu\psi(x) + iqe^{iq\theta(x)}\psi(x)\partial_\mu\theta(x)$$

where $\theta(x)$ is a function of the space-time coordinates.
To have an invariant lagrangian under local gauge transformation, must be introduced the covariant derivative D_μ, that undergoes the same phase transformation of the field:

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

knowing that the magnetic vectorial field A_μ transforms like:

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

The new 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 charge current, which is the probability current multiplied by the particle charge.
By adding 1.9 and 1.2 is obtained the local invariant QED lagrangian for a Dirac particle in an electromagnetic field:

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

From the $\mathcal{L}_{QED}$ lagrangian is possible to extract the Feynman rules which allow a diagrammatic representation of the process amplitudes. In Figure 1.2 a fundamental transition associated to the lagrangian density in (1.10) can be seen.

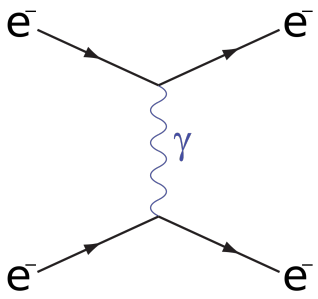


Figure 1.2: Basic vertex in Quantum Electrodynamics.

1.3 The Electroweak theory

The Enrico Fermi's article "Tentativo di una teoria dell'emissione dei raggi beta"[21] and following developments led to the first phenomenological description of the weak interaction, determined by a current-current lagrangian density:

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

where G is the Fermi constant

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

and J^μ is the weak current composed by the sum of a weak leptonic current, $l^\mu(x)$ and a weak hadronic current $h^\mu(x)$. The interactions could be:

- leptonic(e.g. $\mu^- \to e^- + \bar{\nu_e} + \nu_\mu$)

$$\mathcal{L}_F = \frac{G}{\sqrt{2}} l^{\dagger\mu}(x) l_\mu(x) \tag{1.12}$$

- semileptonic(e.g. $n \to e^- + \bar{\nu_e} + p$)

$$\mathcal{L}_F = \frac{G}{\sqrt{2}} [l^{\dagger\mu}(x) h_\mu(x) + h^{\dagger\mu}(x) l_\mu(x)] \tag{1.13}$$

- hadronic(e.g. $\Lambda \to p + \pi^-$)

$$\mathcal{L}_F = \frac{G}{\sqrt{2}} h^{\dagger\mu}(x) h_\mu(x). \tag{1.14}$$

All the above terms represent contact interactions, meaning that in this model the involved particles are supposed to be in the same point of the space-time coordinates, when the interaction occurs.
For leptonic interactions, which involve only elementary particles, the vector and spin

structure of the current is of the V-A type:

$$l^{\mu}(x) = \bar{e}(x)\gamma^{\mu}(1-\gamma^{5})\nu_{e}(x) + \bar{\mu}(x)\gamma^{\mu}(1-\gamma^{5})\nu_{\mu}(x). \tag{1.15}$$

Each Dirac field could be written as:

$$\psi = \psi_L + \psi_R \tag{1.16}$$

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

In 60's and 70's Glashow, Weinberg, and Salam theorized the Electroweak theory, to solve the divergences problem, unifying the electromagnetic and the weak interactions. This theory is also known as the GWS model and Glashow, Weinberg, and Salam won the Nobel prize in Physics in 1979. The GWS model is a development of Fermi's theory, where has been introduced an adimensional 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.
Since the three Pauli 2x2 matrices σ_i are the generators of the group $SU(2)_L$, fermions could be presented as doublets, eigenstates of the weak isospin I. 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:

$$\begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}'_L = e^{-i\vec{\alpha}(x)\cdot\vec{\tau}} \begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}_L , \tag{1.17}$$

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 are singlet of $SU(2)_L$ and don't take part to weak charged-current processes, but the right-chiral components of charged fermions:

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

are involved in neutral-current processes. In the SM there aren't right-chiral neutrinos. The symmetry group $U(1)_Y$ is associated to the weak hypercharge Y, defined as:

$$Q = I_3 + \frac{Y}{2}. \tag{1.18}$$

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

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

where $\bar{Q}$ is the mean electric charge of the multiplets.
The request of invariance under $G = SU(2)_L \otimes U(1)_Y$ introduces 4 gauge fields, 3 Yang-Mills fields $W^{\mu}_{(i)}$, that are an isospin triplet, from $SU(2)_L$, and 1 singlet B^{μ}, from $U(1)_Y$. 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 that are a linear combination between the mass eigenstates. The flavour mixing is given by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, 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 parameter, the Wolfenstein parametrization (λ, η, ρ, A) 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 [45] gives

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

this means that the matrix is almost diagonal and quark are inclined to be associated with quarks to the same family in the interactions.
According to the gauge principle, as it has been done for the QED lagrangian, the Electroweak lagrangian can be obtainable starting from the free fermionic lagrangian, by neglecting mass terms, and by introducing a new covariant derivative:

$$D_\mu = \partial_\mu + ig\frac{\vec{\tau}}{2}W_\mu + ig'\frac{Y}{2}B_\mu, \tag{1.19}$$

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

$$\mathcal{L}_{fermions} = \sum_f \bar{\psi}\gamma^\mu D_\mu \psi, \tag{1.20}$$

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

$$\mathcal{L}_{gauge} = -\frac{1}{4}W_i^{\mu\nu}W^i_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{1.21}$$

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

$$\begin{aligned} W_i^{\mu\nu} &= \partial^\mu W_i^\nu - \partial^\nu W_i^\mu \\ B_i^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu. \end{aligned} \tag{1.22}$$

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

$$\begin{aligned} \mathcal{L}_{EW} =& -i\psi_L\gamma^\mu\left(\partial_\mu + ig\frac{\vec{\tau}}{2}\cdot W_\mu + ig'YB_\mu\right)\psi_L + \\ & - i\psi_R\gamma^\mu\left(\partial_\mu + ig'YB_\mu\right)\psi_R + \\ & - \frac{1}{4}W_i^{\mu\nu}W^i_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \\ & + \frac{1}{2}g\,\epsilon_{ijk}W_i^{\mu\nu}W_{j\mu}W_{k\nu} + \frac{1}{4}g^2\,\epsilon_{ijk}\epsilon_{imn}W_{j\mu}W_{k\nu}W_m^\mu W_n^\nu, \end{aligned} \tag{1.23}$$

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

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \\ Z_\mu &= -sin\theta_W B_\mu + {W_\mu}^3 cos\theta_W \\ A_\mu &= cos\theta_W B_\mu + {W_\mu}^3 sin\theta_W, \end{aligned} \tag{1.24}$$

The parameter θ_W is called Weinberg angle, defined as:

$$cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}, \tag{1.25}$$

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

$$q = g' \cos\theta_W = g \sin\theta_W \tag{1.26}$$

Two of these 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 triplet is electrically neutral. In Figures 1.3 and 1.4 representative tree-level diagrams of charged- and neutral-current processes are shown.

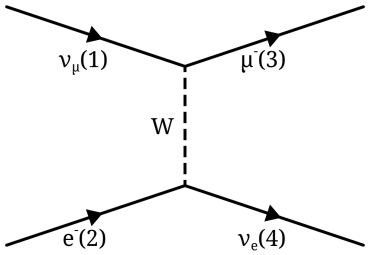


Figure 1.3: Charged current process in GWS model.

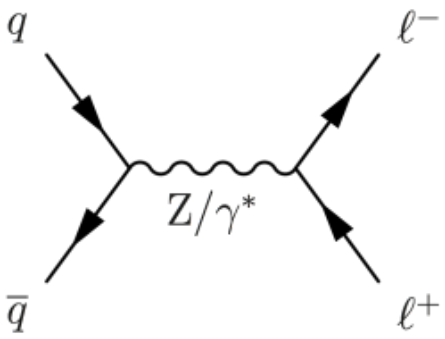


Figure 1.4: Neutral current process in GWS model.

1.4 Quantum Chromodynamics

The Quantum Chromodynamics (QCD) is the quantum gauge field theory, that describes the strong interaction between quarks. The symmetry group for the QCD is $SU(3)_C$, where C is the charge associated to this symmetry, the colour. QCD is similar to QED, but due to the non-Abelian nature of $SU(3)_C$, the generators of the group do not commute between themselves, this means that in the QCD lagrangian there would be self-interaction terms between the gauge fields. The gauge fields, the gluons, therefore have a colour charge, differently from 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:

$$[T_\alpha, T_\beta] = i f_{\alpha\beta\gamma} T_\gamma, \tag{1.27}$$

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:

$$\psi \to \psi' = e^{i g_s \vec{\theta}(x)\cdot\vec{T}}\psi, \tag{1.28}$$

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.This new degree of freedom, colour, has three possible states labelled as red, green and blue. The imposition of local invariance leads to introduce the covariant derivative:

$$D_\mu = \partial_\mu + i g_s T_\alpha G^\alpha_\mu, \tag{1.29}$$

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

$$G^\alpha_\mu \to G'^\alpha_\mu = G^\alpha_\mu + i g_s f^{\alpha\beta\gamma}\theta_\beta(x) G_{\gamma,\mu}. \tag{1.30}$$

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

$$\mathcal{L}_{QCD} = \bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - i g_s \bar{\psi}\gamma^\mu\lambda_\alpha\psi G^\alpha_\mu - \frac{1}{4} G^{\mu\nu}_\alpha G^\alpha_{\mu\nu} \tag{1.31}$$

where $G^{\mu\nu}_\alpha$ is the tensor field that is defined as

$$G^{\mu\nu}_\alpha = \partial^\mu G^\nu_\alpha - \partial^\nu G^\mu_\alpha - g_s f_{\alpha\beta\gamma} G^{\beta,\mu} G^{\gamma,\nu}. \tag{1.32}$$

In Figures 1.5 and 1.6 representative tree-level diagrams of QCD processes are shown. From experimental evidence the strong interaction shows two relevant properties, successfully explained in the SM by the QCD theory: asymptotic freedom and colour con-

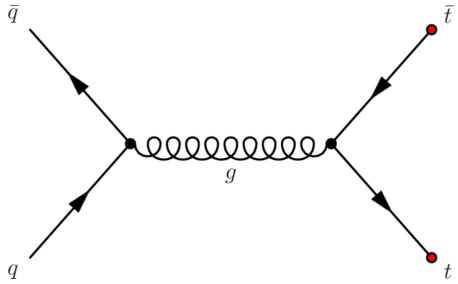


Figure 1.5: Basic vertex in Quantum Chromodynamics.

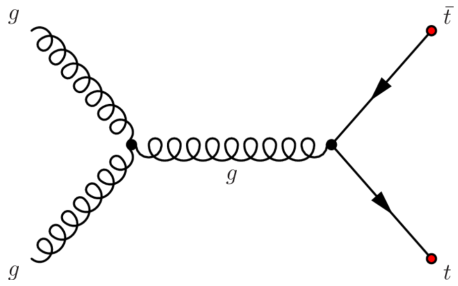


Figure 1.6: Quantum Chromodynamics process with a self interaction vertex between glouns.

finement.

Asymptotic freedom refers to the feature that strong interactions between quarks become weaker the higher the energy of the particle. In the SM, this allows for perturbative calculation. This is due to the the running of the strong coupling constant:

$$\alpha_s\left(|q^2|\right) = \frac{\alpha_s(\mu^2)}{\left[1 + \alpha_s(\mu^2)\frac{33-2N_f}{12\pi}\ln\frac{q^2}{\mu^2}\right]}, \tag{1.33}$$

where q^2 is 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. Eq. 1.33 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.

The second property has been formulated since no coloured hadrons are observed in nature. Hadron 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 doesn't allow to introduce mass terms for the fermion, that would appear as:

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

and for the bosons:

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

or they would 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. This mechanism introduces a new boson field ϕ, the Higgs field, which is an isospin double of complex scalar field:

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.34}$$

The lagrangian for ϕ field is:

$$\begin{aligned} \mathcal{L}_H &= (D^\mu \phi)^\dagger D_\mu \phi - V(\phi) = \\ &= (D^\mu \phi)^\dagger D_\mu \phi - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{1}{4}\lambda(\phi^\dagger \phi)^2, \end{aligned} \tag{1.35}$$

where $V(\phi)$ is the Higgs potential and λ is assumed to be positive. Minimizing the potential V, it is possible to find the ground vacuum state ϕ_0. Adding the potential in Equation (1.35) to the lagrangian of gauge field sector, when the covariant derivative acts, it has:

$$\mathcal{L}_H = (D^\mu \phi)^\dagger D_\mu \phi - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{\lambda}{4}(\phi^\dagger \phi)^2 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \tag{1.36}$$

with:

$$D^\mu \phi = \left(\partial^\mu + ig\frac{\vec{\tau}}{2}W^\mu + ig'YB^\mu\right)\phi \tag{1.37}$$

$$F^{\mu\nu} = \partial^\mu W^\nu - \partial^\nu W^\mu - gW^\mu \times W^\nu \tag{1.38}$$

$$B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu. \tag{1.39}$$

As shown in Fig 1.7, for $\mu^2 > 0$ there is just one minimum at $\phi_0 = 0$ and therefore $V(\phi) = 0$, while for $\mu^2 < 0$ 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

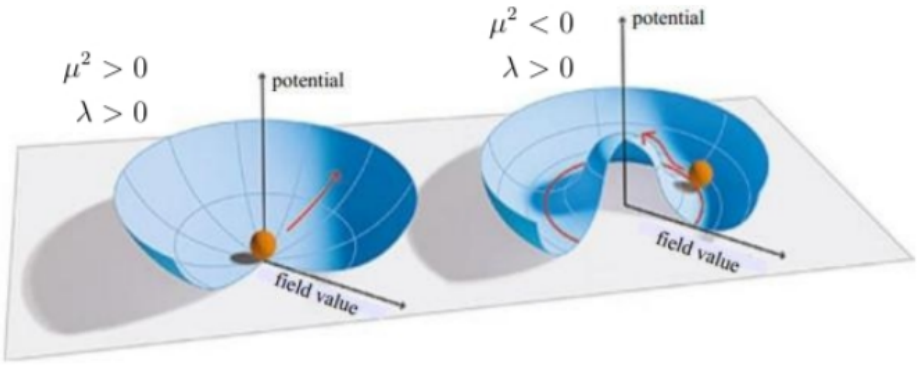


Figure 1.7: Shape of the Higgs potential $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$ for $\lambda > 0$ and $\mu^2 < 0$ and for $\mu^2 > 0$.

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:

$$\phi = \begin{pmatrix}0\\ \frac{1}{\sqrt{2}}\left(v + H(x)\right)\end{pmatrix}. \tag{1.40}$$

By substitution of the Equation 1.40 in Equation 1.36, taking into account the Eq. 1.24, leads to:

$$\begin{aligned}\mathcal{L}_{G\Phi} = &\frac{1}{2}\partial_\mu H\partial^\mu H - \mu^2 H^2 + \\ &-\frac{1}{4}(\partial_\mu W^1_\nu - \partial_\nu W^1_\mu)(\partial^\mu W^{1\nu} - \partial^\nu W^{1\mu}) + \frac{1}{8}g^2v^2W^1_\nu W^{1\nu} \\ &-\frac{1}{4}(\partial_\mu W^2_\nu - \partial_\nu W^2_\mu)(\partial^\mu W^{2\nu} - \partial^\nu W^{2\mu}) + \frac{1}{8}g^2v^2W^2_\nu W^{2\nu} \\ &-\frac{1}{4}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu Z^\nu - \partial^\nu Z^\mu) + \frac{1}{8}(g^2 + g'^2)v^2Z_\nu Z^\nu \\ &-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}.\end{aligned} \tag{1.41}$$

The first line of 1.41 is the lagrangian density of the Higgs scalar field, with mass $\sqrt{2}\mu$. The next two lines show that the components W^μ_1 and W^μ_2 of the triplet $\mathbf{W}^\mu$ acquire mass:

$$M_1 = M_2 = \frac{1}{2}gv \equiv M_W. \tag{1.42}$$

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

$$M_Z \equiv \frac{1}{2}v\sqrt{g^2 + g'^2} = \frac{M_W}{\cos\theta_W}. \tag{1.43}$$

In the end 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, this mass term is:

$$\mathcal{L}_Y = g_f(\bar{\psi}_L\phi\psi_R - \bar{\psi}_R\phi^\dagger\psi_L), \tag{1.44}$$

where g_f is the Yukawa coupling constant. By substituting 1.40 in 1.44 one obtains:

$$\begin{aligned}\mathcal{L}_Y =& \frac{g_f}{\sqrt{2}}\left[(\nu_\ell,\ \ell)\begin{pmatrix}0\\ v+H\end{pmatrix}\ell_R + \ell_R(0,\ v+H)\begin{pmatrix}\nu_\ell\\ \ell\end{pmatrix}\right] = \\ =& \frac{g_f}{\sqrt{2}}(v+H)(\ell_L\ell_R + \ell_R\ell_L)\end{aligned} \tag{1.45}$$

The part of lagrangian is proportional to $(\ell_L\ell_R + \ell_R\ell_L)$ and, therefore, the first term is similar to the Dirac mass term that could not be introduce by hand. For each charged leptons the mass results to be:

$$m_l = \frac{v}{\sqrt{2}}g_l. \tag{1.46}$$

However, even having a mass term that doesn't 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, it is possible to write:

$$\begin{aligned}\mathcal{L}_Y =& \frac{1}{\sqrt{2}}\left[g^d_{i,j}(u_{i,L},\ d_{i,L})\begin{pmatrix}0\\ v+H\end{pmatrix}d_{j,R} + g^u_{i,j}(u_{i,L},\ d_{i,L})\begin{pmatrix}-(v+H)^*\\ 0\end{pmatrix}u_{j,R} + h.c.\right] = \\ =& \frac{1}{\sqrt{2}}(v+H)[g^u_{ij}(u_{i,L}u_{j,R} + u_{j,R}u_{i,L}) + g^d_{ij}(d_{i,L}d_{j,R} + d_{j,R}d_{i,L}) + h.c.]\end{aligned} \tag{1.47}$$

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

$$m^u_{ij} = -\frac{v}{\sqrt{2}}g^u_{ij} \quad m^d_{ij} = -\frac{v}{\sqrt{2}}g^d_{ij}, \tag{1.48}$$

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

$$u_{\alpha,L} = (\mathcal{U}^u_L)_{\alpha i} u_{i,L} \quad u_{\alpha,R} = (\mathcal{U}^u_R)_{\alpha i} u_{i,R} \tag{1.49}$$

$$d_{\alpha,L} = (\mathcal{U}^d_L)_{\alpha i} d_{i,L} \quad d_{\alpha,R} = (\mathcal{U}^d_R)_{\alpha i} d_{i,R} \tag{1.50}$$

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:

$$\mathcal{L}_Y = \frac{1}{\sqrt{2}}(v+H)\,[m^u \mathrm{u\bar{u}} + m^d \mathrm{d\bar{d}} + m^s \mathrm{s\bar{s}} + m^c \mathrm{c\bar{c}} + m^t \mathrm{t\bar{t}} + m^b \mathrm{b\bar{b}}], \tag{1.51}$$

but these transformations need to be applied also in the Electroweak interaction lagrangian, therefore the charged current coupling term is:

$$\begin{aligned}
\mathcal{L}_{CC} &= -\frac{g}{\sqrt{2}}(\bar{u}_{i,L},\ \bar{d}_{i,L})\gamma^\mu \tau_+ W^+_\mu \begin{pmatrix} u_{Li} \\ d_{Li} \end{pmatrix} + h.c. \\
&= -\frac{g}{\sqrt{2}}\bar{u}_{iL}\gamma^\mu d_{Li} W^+_\mu + h.c. \\
&= -\frac{g}{\sqrt{2}}\bar{u}_{\alpha_L}\left[(\mathcal{U}^u_L)_{\alpha i}(\mathcal{U}^d_L)^\dagger_{\beta i}\right]\gamma^\mu d_{L\beta} W^+_\mu + h.c.,
\end{aligned} \tag{1.52}$$

where:

$$V_{\alpha\beta} = \left[\mathcal{U}^u_L \mathcal{U}^{d\dagger}_L\right]_{\alpha\beta} \tag{1.53}$$

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:

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}}\bar{u}_{L\alpha}\gamma^\mu V_{\alpha\beta} d_{L\beta} W^+_\mu - \frac{g}{\sqrt{2}}\bar{d}_{L\alpha}\gamma^\mu V^\dagger_{\alpha\beta} d_{L\beta} W^-_\mu \tag{1.54}$$

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, being a coherent 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 some of its parameters, e.g. 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;
- **Neutrino masses**: in the SM neutrinos 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 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 quantistic description for gravity 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 1.8. A precise tuning between fermions and scalars, known as fine tuning, is required in order to reduce or cancel these divergencies.

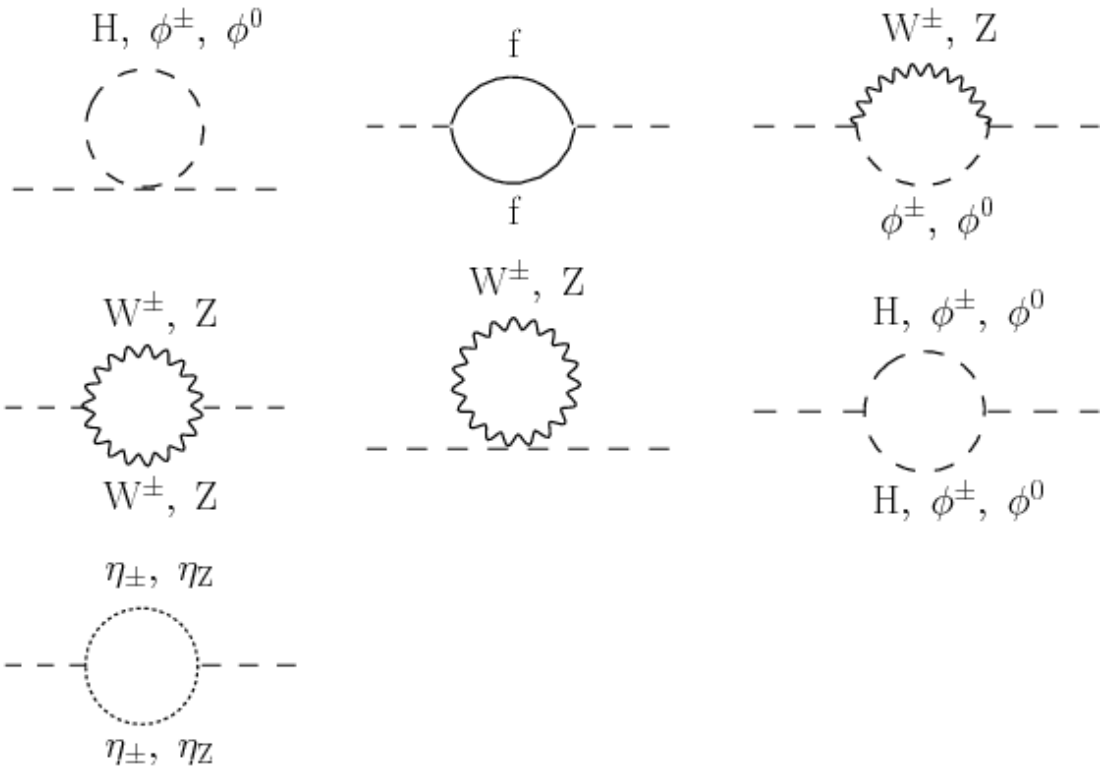


Figure 1.8: One-loop self-energy corrections to the Higgs mass.

- **Flavour Changing Neutral Currents (FCNC)**: as supposed by the Glashow, Iliopoulos and Maiani (GIM) mechanism, are highly suppressed at three level and may occur in the SM only 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

The Compact Muon Solenoid(CMS) is one of the main experiments at the CERN's Large Hadron Collider. The acronym CERN derives from "*Conseil européen pour la recherche nucléaire*" and it is one of the leading research institute in the Europe, whose main site is located in Geneva and hosts the biggest particle physics laboratory in the world. It was established in 1954, when the member states were only 12, all of them in the Western Europe. Today it counts 23 member states and, even if has been decided that the acronym still be CERN, the name changed in "*Organisation Européenne pour la Recherche Nucléaire*". At CERN, breakthrough discoveries have been made, such as the discovery of W and Z bosons at UA1 and UA2 experiments in 1983 [15, 14] or the discovery of Higgs boson in 2012 [18].

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest and most complex accelerator, where two high-energy beams travel at very high speed, before they are made to collide. The collision is made up to understand how particles interact and to understand the nature law. LHC is the last elements of an accelerator complex, in which each machine increases the energy of the beams, and, reached the wanted energy, the beams are injected in the next step of the complex or are used for lower energy experiments. The injection energy of beams at the LHC is 450 GeV. The particles used in the beams are protons and the source used is an hydrogen gas bottle. To strip hydrogen atoms of their electrons thus yielding protons to be accelerate, an electric field is used. 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, where the two beams travel in opposite direction, in

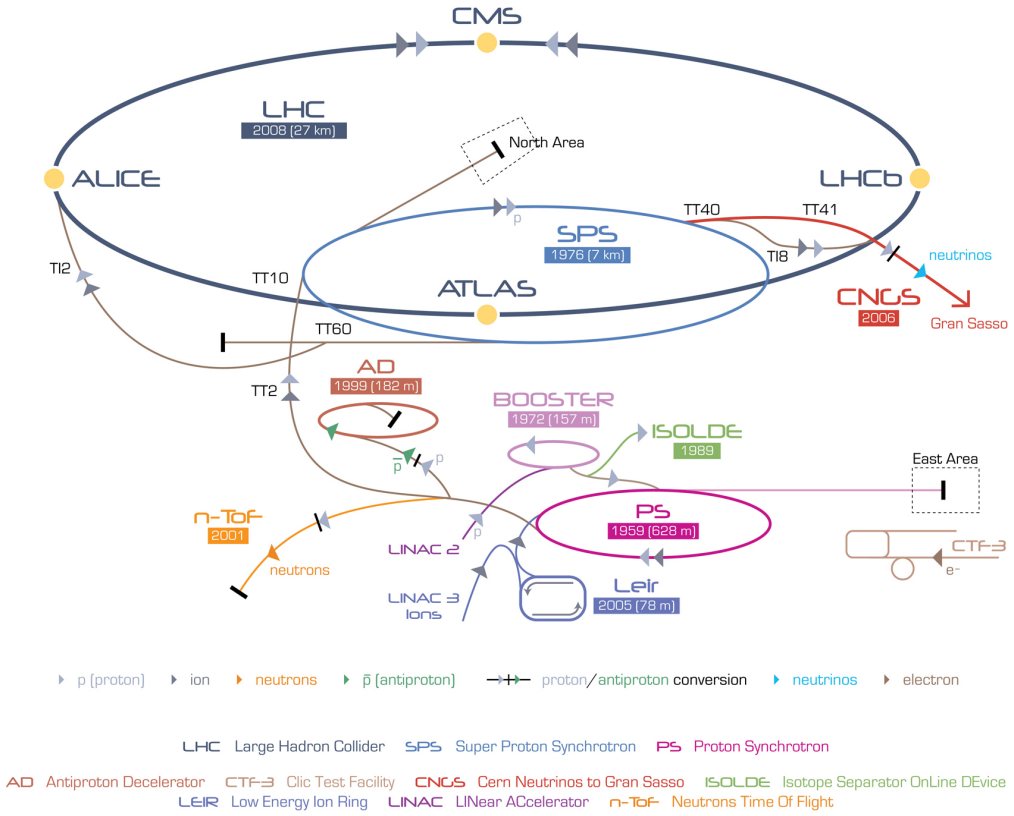


Figure 2.1: Schematic view of the accelerator complex at CERN.

two different beam pipes. 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.
The maximum energy is 6.5 TeV and the centre of mass energy at collision is 13 TeV, reached by thousand of magnets of different varieties and sizes. 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 is not the only one property that characterizes the number of expected events. The rate R of a process with cross section σ is 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 follow:

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

  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 of the above parameters characteristic of the 2018 Data taking [48]. The peak luminosity in 2018 reached the value of $2 \times 10^{34}\ cm^{-2}s^{-1}$ and in Figure

Parameter	Value
N_b	1.15×10^{11}
n_b	2256
$\epsilon[\mu m]$	2
$\beta^*[cm]$	30 - 25
F	0.6

Table 2.1: The LHC machine parameters.

2.2 the evolution of the LHC peak luminosity between 2011 and 2018.
The number of expected events can be written as:

$$N = L\sigma,$$

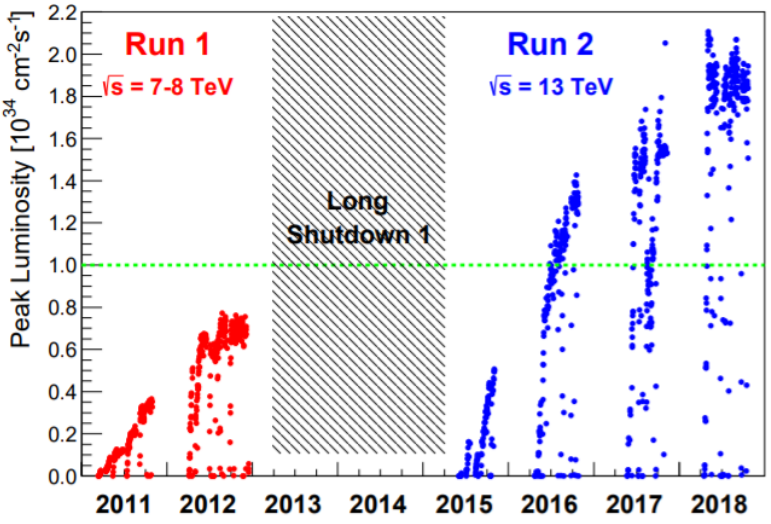


Figure 2.2: Evolution of the LHC peak luminosity between 2011 and 2018.

where, now, L is the integrated luminosity:

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

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

- ALICE (A Large Ion Collider Experiment) is studying the heavy-ion (Pb-Pb nuclei) collisions. The energy in center-of-mass per nucleon pair is 2.76 TeV, 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 just after the Big Bang.

- ATLAS (A Toroidal LHC ApparatuS) is a general purpose detector, that has been designed to study a large variety of processes, that include the top and beauty quarks and the Higgs boson. Moreover its task is to study processes beyond the SM, as the existence the existence of new particles, like Vector Like Vector Like Quark (VLQ), this last will be described in the next chapter.

- CMS (Compact Muon Solenoid) is a general purpose detector and it will be detailed discussed in the next section.

- LHCb is studying the properties of b type quark, and searching answers to questions like the matter-antimatter asymmetry. At the LHCb interaction points, two beams are collided at different energies: one has en energy of 7 TeV and the other

has an energy of 450 GeV. These conditions allow to provide generated particles with a significant Lorentz boost along the direction of the higher energy beam.

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

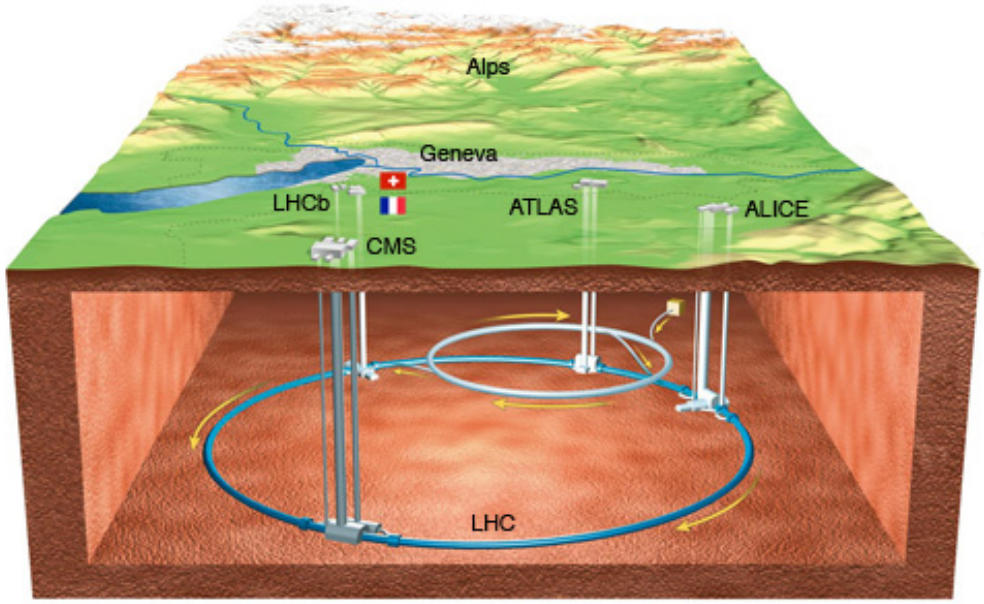


Figure 2.3: A schematic view of the four experiments at LHC.

2.2 Compact Muon Solenoid (CMS)

CMS is a general purpose detector, it is built around a huge solenoid magnet, from which it takes the name. The magnetic field, of 3.8 T, is generated by a cylindrical coil of superconducting cable and it is confined by a steel yoke. The complete detector is 21 metres long, 15 metres wide and 15 metres high.
The main goal of CMS is to distinguish physical processes of interest from backgrounds, detecting particles, produced in the hadron collisions, and measuring their energies, charges, masses and momenta. The detector, therefore, is composed by different subdetectors, with high momentum resolution in a large range of energy. Each subdetector is 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 sub-detectors, to reduce the pile-up. The pile-up is the overlapping between two particles of the same event or from different events in the same bunch-crossing, reconstructed as one particle. It would have same problem if the time resolution is not good and the detector's response is slower than the time interval between two bunch-crossing. Also a wide angular coverage is required. In Figure 2.4 is shown a picture of CMS.

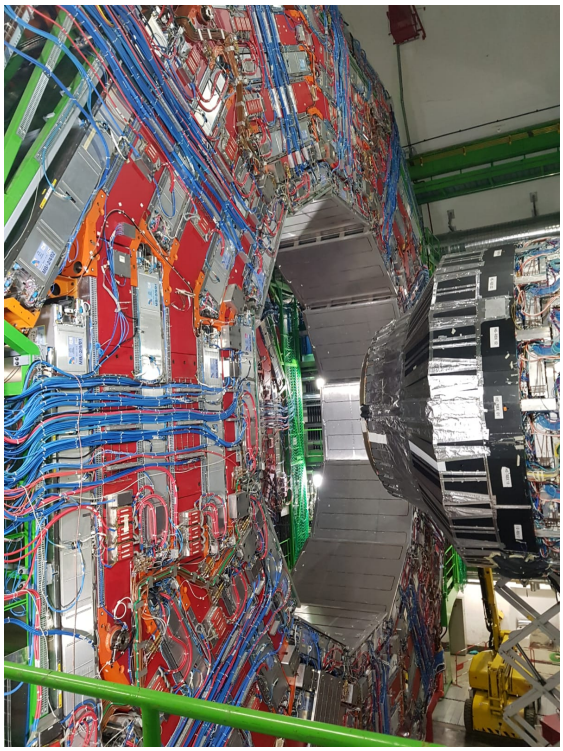


Figure 2.4: A CMS picture, taken in December 2019.

To characterize the reconstructed object in the detector, it is used a right-handed Cartesian coordinate system, 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.

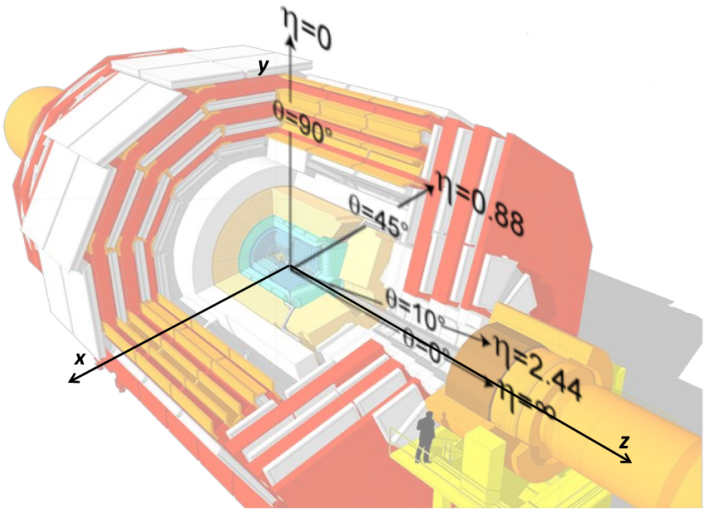


Figure 2.5: The coordinate system of CMS and the functional dependence of η by *theta*.

Using these coordinates, the angular distances between two objects can be written as a Lorentz invariant variable, under boosts along the $z-axis$:

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

Usually two important variables are p_T and E_T, referred to the Cartesian system, they are respectively the transverse momentum and the transverse energy, defined as:

$$\vec{p}_T = \sqrt{\vec{p}_x^{\,2} + \vec{p}_y^{\,2}} \tag{2.5}$$

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

2.2.1 The subdetector system of CMS

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.

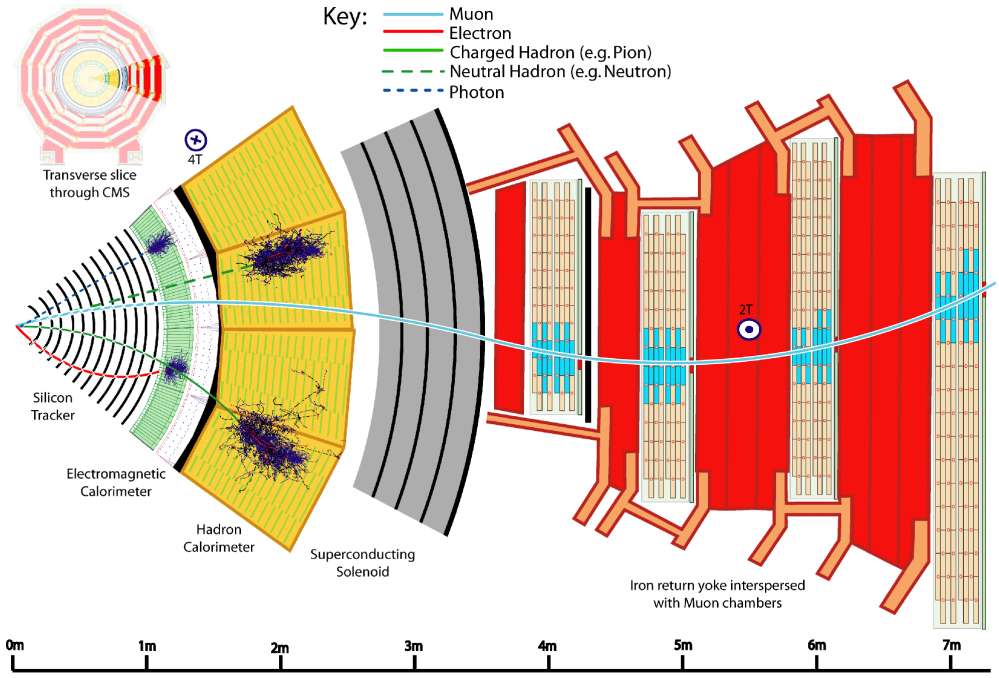


Figure 2.6: Schematic view of CMS subdetectors.

Tracking system The inner tracking [26] [3] system is designed to provide a robust tracking and a detailed vertex reconstruction. This detector has a diameter of 2.5 m and a length od 5.8 m, with an acceptance of $|\eta| < 2.5$. Due to the huge rate of particles, it must have high spatial resolution, to have a good distinction also of very nearby tracks,

and a fast response, to ensure the correct working of the online trigger. It has an active area of 200 m^2 and it is composed by the pixel vertex detector and the silicon strip detector (SST).
The pixel vertex detector is the closest one to the interaction point. It provides an accuracy of 10 μm and of 20 μm respectively for radial and transverse position measurements, thanks to over 65 million silicon pixels. It composed by three barrel layers (BPIX) and two disks both forward and backward (FPIX) as endcaps.
The silicon strip detector covers the external part of the pixel detector. It has a resolution from 35 μm to 52 μm in the radial direction and 530 μm in the transversal direction. 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 $z - axis$.
In Figure 2.7 is shown a schematic view of the tracker system.

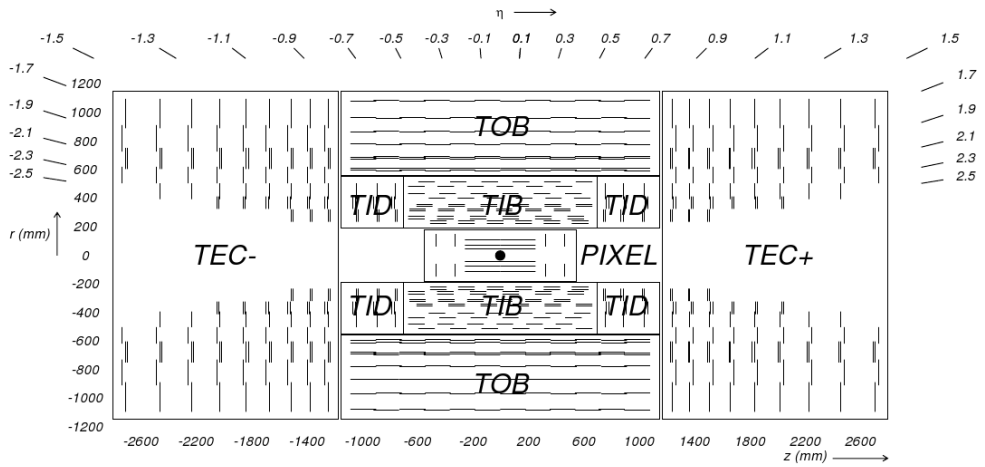


Figure 2.7: Schematic view of the tracker system.

Electromagnetic calorimeter (ECAL) The electromagnetic calorimeter (ECAL) [1] is a hermetic homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals. Its main goal is to measure the energy released from photons and electrons in it. It has a radius between 1.25 m and 1.8 m and it is divided into a barrel section (EB) and two endcaps (EE+ and EE-). The barrel section covers a range $|\eta| < 1.48$, while the two endcaps from 1.48 to 3, as can be seen in Figure 2.8. 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. In Figure 2.8 a schematic view of the ECAL.

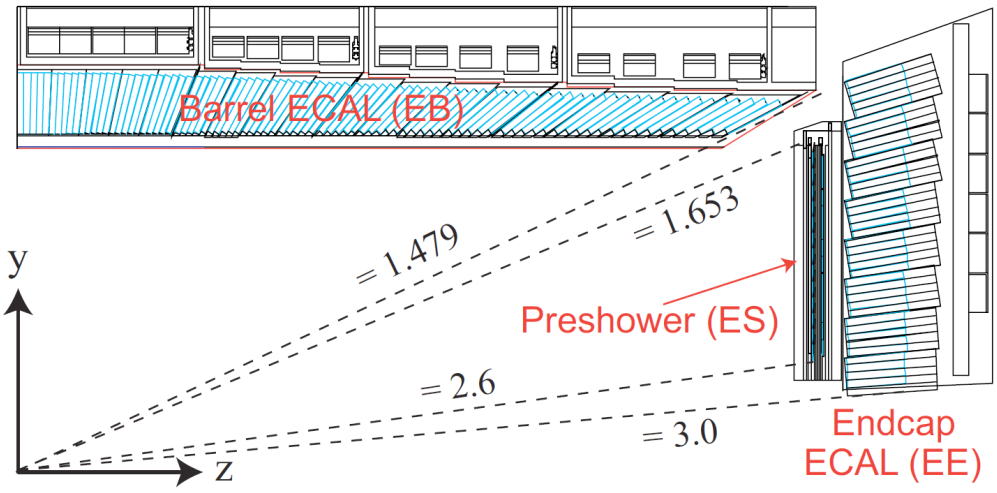


Figure 2.8: Schematic view of ECAL system.

Hadronic calorimeter (HCAL) The HCAL [2] 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 layers of active material are made up of fluorescent scintillators, while the absorber used is cartridge brass (C26000). This cartridge brass is composed by 70% of Cu and by 30% of Zn, that, with a density 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.
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.

One of the most important properties of the ECAL and the HCAL is the energy

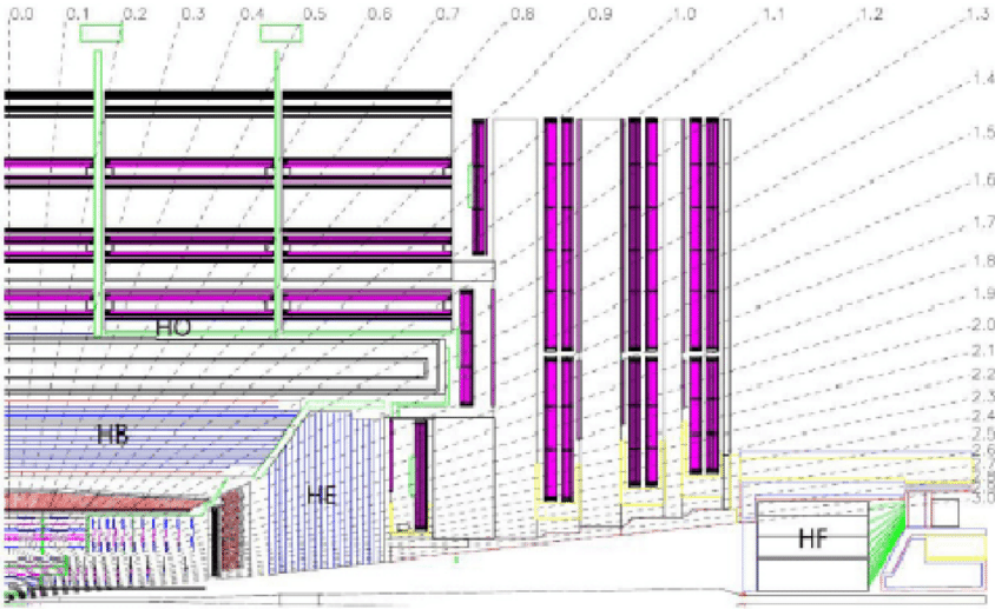


Figure 2.9: Schematic view of the HCAL.

resolution, that can be parametrized as:

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

where:

- a represents the stochastic term: it takes in account the statistical fluctuations of the signal in the shower containments, such as fluctuations in the number of primary particles and/or the number of photons which includes fluctuations in the shower containments 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:

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

Magnet The magnet embraces the tracker detector and the calorimeters. It provides a magnetic field of 3.8 T and it is composed by a superconducting coil. The magnet's goal is to curve the particles produced in the collisions to know their charge and momentum. Thanks to the iron Return Yoke, border effect are avoided. Moreover the yoke introduces a 1.8 T constant field outside the magnet. The yoke is made up of alternating layers interspersed with the muon detectors.

Muon system The CMS muon system [29] is used for muon identification, momentum measurement and triggering. In many physics processes muons result to be object reconstructable in the final, therefore precise and robust muon measurement could be used to distinguish interesting processes from backgrounds. The muons can cross several layer of material and they arrive to the external parts of the detector, so a large surface detector is required. The muon system covers an area of about 25000 m^2, therefore the detectors of choice have to be reliable, robust and of contained costs.
To manage these requests three different kinds of gas detector are used: Drift Tubes, Cathode Strip Chambers, and Resistive Plate Chambers.

- 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 an high time resolution, while the spatial resolution is lower than th DTs and CSCs. The RPCs are double-gap chambers, operated in avalanche mode to ensure good operation at high rates. In the barrel muon system there are 6 layer of RPCs, while in the endcaps there is one layer for each of the four stations. In the first two stations of the barrel two layers of RPCs are located, while just one in the last two station of DTs.

A characteristic of the DT and CSC subsystems is that they can each trigger on the p_T of muons with good efficiency and high background rejection, independent of the rest of

the detector.
The Level-1 trigger p_T resolution is about 15% in the barrel and 25% in the endcap. Due to the fast response of the RPCs, even low p_T tracks, that may stop before reaching the outer two stations, could be triggered.

Trigger and data acquisition (DAQ) system The trigger system [16, 19] performs the events selection. Due to the high instant luminosity at LHC, the entire set of data can be acquired and analyzed, therefore only the events that have some proprieties are selected.
The trigger system reduces the rate by a factor 10^6 and it consists in two steps: the Level-1 Trigger (L1) and the High-Level Trigger (HLT), shown in Figure 2.10.

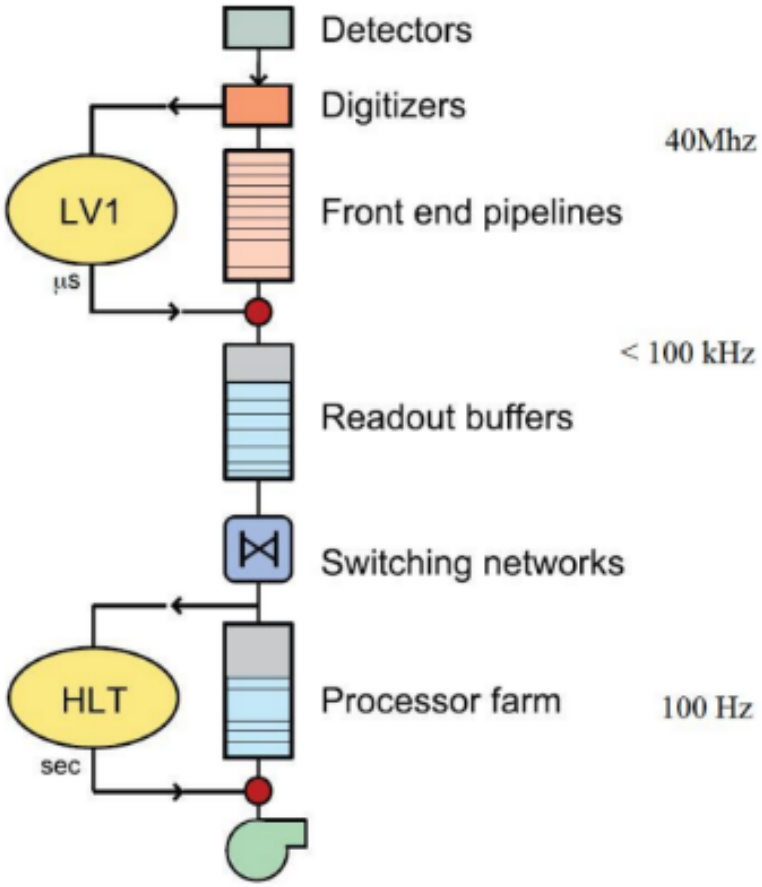


Figure 2.10: Trigger System at CMS.

The Level-1 Trigger provides a fast and an automatic event selection, it is organized

in 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 sub-detectors and the DAQ. The Level-1 Accept (L1A) decision is communicated to the sub-detectors through the Timing, Trigger and Control (TTC) system. The architecture of the L1 Trigger can be seen in Figure 2.11.

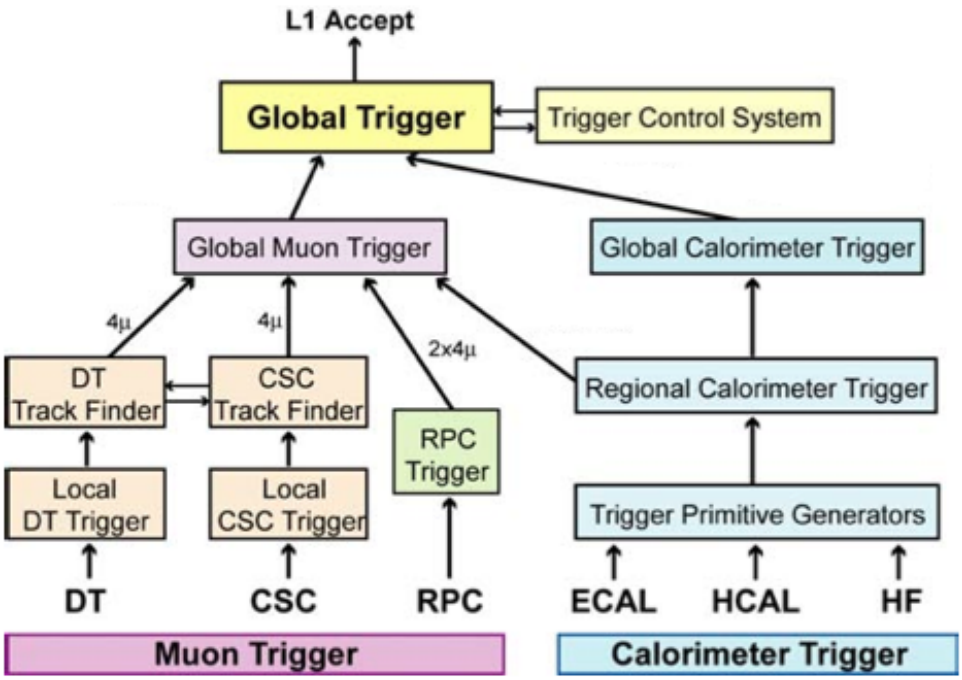


Figure 2.11

The final decision to reject or to accept an event is taken by the HLT, this last uses a software filter system. It has the access to the complete read-out data and it perform an analysis based on the physical objects present in the event. For example events with bad vertices reconstruction or events with final objects characterized by a too low momenta are immediately refuses by the HLT.

Chapter 3

Physics BSM and Vector Like Quark T

As previously discussed, the SM doesn't explain all observed phenomena in high energy physics. Many theories provide different solutions for the unsolved issues of the SM. Thanks to the high energy reach and luminosity of the LHC machine, the experiments can search for new particles are predicted by several Beyond the Standard Model theories, allowing to either confirm, or refute them. Many theories that that aim at solving the hierachy problem and realizing the naturalness, predict the existence of new particles, the Vector Like Quarks (VLQs). In this Chapter will be described the Composite Higgs and the Extra Dimension models.
The Composite Higgs models suppose that the Higgs is a composite state of a new strong interaction. The Higgs mass operator's dimensionality could be higher than 4, and therefore there are no issues arise from the low measured value of the Higgs mass.
The Extra Dimension theories predict the existence of extra dimensions beyond the usual space-time (3+1) in order to solve also the hierarchy problem and the unification of the fundamental interactions.
In both scenarios the VLQs have an important role. The VLQs are supposed to be coloured fermions, also called $top - partners$, that avoid the divergences of the Higgs mass corrections.

3.1 The Composite theories

In 1984 Georgi and Kaplan proposed the Composite Higgs Model [25] , in which 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 $SO(5) \rightarrow SO(4)$ is associated the pNGB, that in this case in the Higgs particle [10] [32].

Considering the vector Ψ_L of $SO(5)$ as an extension of the SM left handed doublet q_L for the third generation of quarks, it is possible to report a general model. 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 Eq 3.2 have been introduced to give mass to the new fermions and them 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 symmetry of the mass term guarantees the absence of quadratic divergences to m_H.
Since

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

the Eq 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 physical field can be obtained with the diagonalization of the mass matrix. The introduction of the VLQs ensure the absence of quadratically divergent terms.
One can consider 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, and it can be shown that 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 on loop correction to A could be written as:

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

It is shown that there is no quadratic divergence and the two mass of the VLQs, M_X and M_T, could be seen as the cut-off Λ in the original top-loop contribution. However the two masses can not be too much above 2 TeV, or the logarithmic term could produce a δm_H of the same order of the weak-scale expectation value v, and therefore the naturalness problem is not solved.
This simplified model introduces the VLQs [31], but often models like it fail to explain the fermion masses origin or they don't pass the electroweak precision tests.
The Composite Higgs are not only ones theories that predict the VLQs in the Composite scenario. The Composite Top theories, in a similar way, suppose that the top quark is a composite state. The SM particles, thanks to a new strong sector, get their masses by mixing themselves with composite states. The top quark, due to its large mass, could show the properties of Compositeness, having a sizeable admixture of the composite state. The electroweak precision tests further models of right-handed composite top quark models.

3.2 The Extra Dimension theories

The Extra Dimension theories have been developed to explain the gravitational interaction ad the unification between all the fundamental forces. The first theory to introduce a fifth dimension beyond the four of the space time to unify the gravity and the electromagnetism, was developed by Kaluza and Klein. Kaluza extended the classical general relativity to five dimensions [24] and Klein, starting from this work, has exported this theory into a quantum field theory [27]. The main idea of the Klein's theory is that the fifth dimension is curled and its geometry has the form of a circle of 10^{-33} radius.
Considering the weakness of the gravity compared to the other fundamental forces, Arkhani-Hamed, Dimopoulos and Dvally (ADD) proposed an Extra Dimension theory [**?**]. This extra dimension should have a size between a millimetre and $\approx \frac{1}{TeV}$, in natural units, in which the gravity could propagate explaining its weakness.
Moreover Lisa Randall and Raman Sundrum have proposed two different models in 1999 to solve the hierarchy problem of the SM [33][34]. The proposed model hypothesize that the Universe has 5 dimension and a warped-geometry. According to the first model (RS1), the extra dimension has a finite size and it is composed by two branes linked

each other; the second model (RS2), predicts a similar space-time structure, but one of the two branes is placed very far away from the other,, and approximating it to be infinitely distant there is effectively one brane left in the model.
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, that is our home with 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 The Vector Like Quarks

The Vector Like Quarks (VLQs) are coloured fermions and the left and right chiral components have the same transformation under $SU(2)_L$. This allows to introduce in the lagrangian mass terms like $-m\bar{\psi}\psi$, since the $SU(2)_L$ symmetry is not broken by this term, being the weak charged current a vectorial current. There is still no evidence of their existence.
As previously described, the introduction of VLQs stabilizes the Higgs boson mass. Moreover, having the same coloured charge of the SM quarks, they are the simplest new coloured fermions that would appear beyond the SM. However, 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 VLQs can mix with the SM quarks, modifying their coupling to the Z, W and the Higgs boson and breaking the GIM mechanism. Therefore flavour changing neutral currents are allowed at the tree-level in the theories that predict them existence.
In the Table 3.1 different kind of VLQs are presented, depending on their electric charges. The VLQs could be grouped into multiplets of $SU(2)$, in particular singlets, doublets or

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

Table 3.1: Charge assignment for VLQs.

triplets as can be seen in the Table 3.2. Considering the different multiplets, the VLQs

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.

can be 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 T_L and T_R are the left- and right- handed components, while the superscript 0 indicates weak eigenstates. The weak and the mass eigenstates for X and Y coincide, since they cannot mix with the SM quarks due to their exotic charges.
Adding the new fields $T^0_{L,R}$ to the SM, the up-type eigenstates (u, c, t, T) can contain components of the new fields, changing the coupling of the quarks to the Z boson. However the mixing is proportional to the ratio between the mass of the SM and the VLQ, and this allow to consider the VLQs mixing only with the third generation quarks, being m_Q/m_{VLQ} is negligible for the first two generations. The top quark, thanks to these considerations and to its large Yukawa coupling, is linked to new physics, related 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 after at the LHC.

3.3.1 VLQs production mechanism

In the pp collisions the VLQs production cross section depends on the mixing with the SM quarks, and therefore on the square of the couplings to W and Z bosons. The production mechanism could be divided 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 3.1 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.

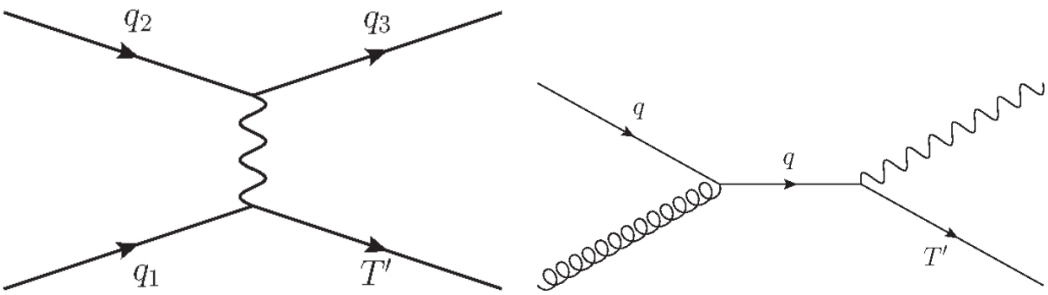


Figure 3.1: Feynman diagrams for single production of Vector-like top quarks.

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

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

  where $Q = T, B, X, Y$.
  However other pair production processes are strongly suppressed, such as the elec-

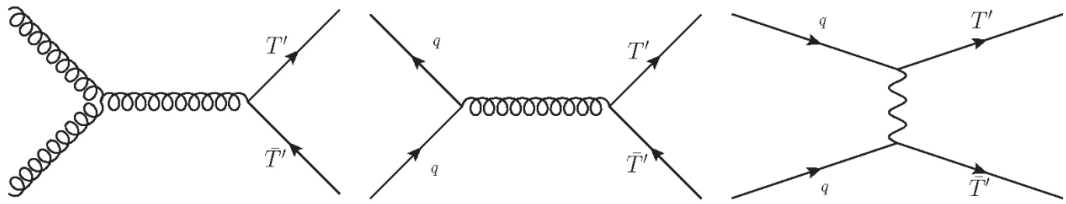


Figure 3.2: Feynman diagrams for pair production of vector-like top quarks via gluon and W,Z or Higgs bosons.

troweak 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 3.3. The plot shows that pair production processes 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 [12].

3.3.2 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 the decay channels into SM particles, the main allowed channels are:

$$T \to W^+b, Zt, Ht$$

$$B \to W^-t, Zb, Hb$$

$$X \to W^+t$$

$$Y \to W^-b.$$

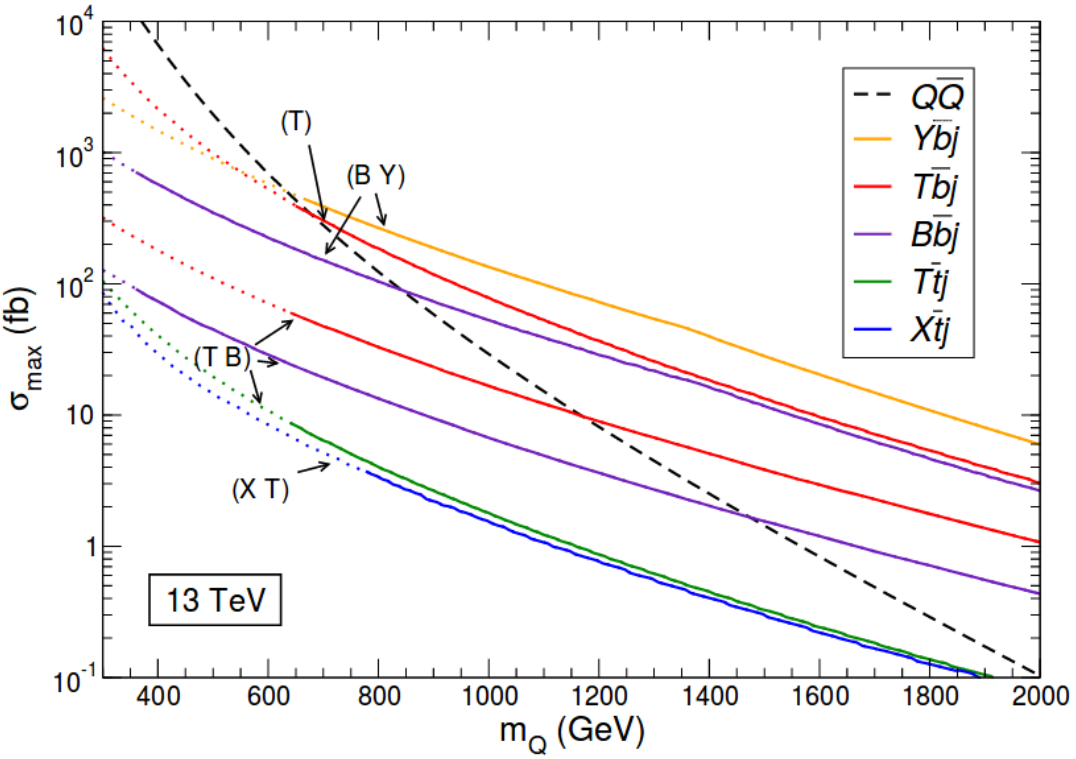


Figure 3.3: Production cross section for Vector-Like Quarks in pp collisions at $\sqrt{s} = 13$ TeV as a function of their mass, for pair production and for single production in different channels. The black dashed line represents VLQ pair production, while the colored lines represent the singly produced VLQs. The dashed coloured lines correspond to the values of cross section excluded by previous studies.

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 $(Q, q, q') = (T, t, b), (B, b, t)$.
For doublets and triplets, the small mass difference suppresses the decay into other VLQs, allowing decays into vector bosons and Higgs boson with top or beauty quark. The decays of the doublet (T, B) depend on the mixing factor of the extended CKM matrix V_{Tb} and V_{tB}, if $V_{Tb} \sim V_{tB}$ implies that the T and B quarks have the same decays as the corresponding singlets. They have different angular distributions since only the right-handed component of (T, B) couples to the SM quarks. However, due to constraints on the b quark mixing and by mass hierarchy, $m_t \gg m_b$ where $V_{Tb} \ll V_{tB}$, the mixing of the heavy quarks with the SM top quark is much stronger. Therefore the $T \to Wb$, $B \to Hb$ and $B \to Zb$ decays are suppressed.

The possible decays of vector-like quarks 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. In Figure 3.4 are illustrated the branching ratios of the decays of T and B in the case of a VLQ coming from the singlet or doublet of $SU(2)$.

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

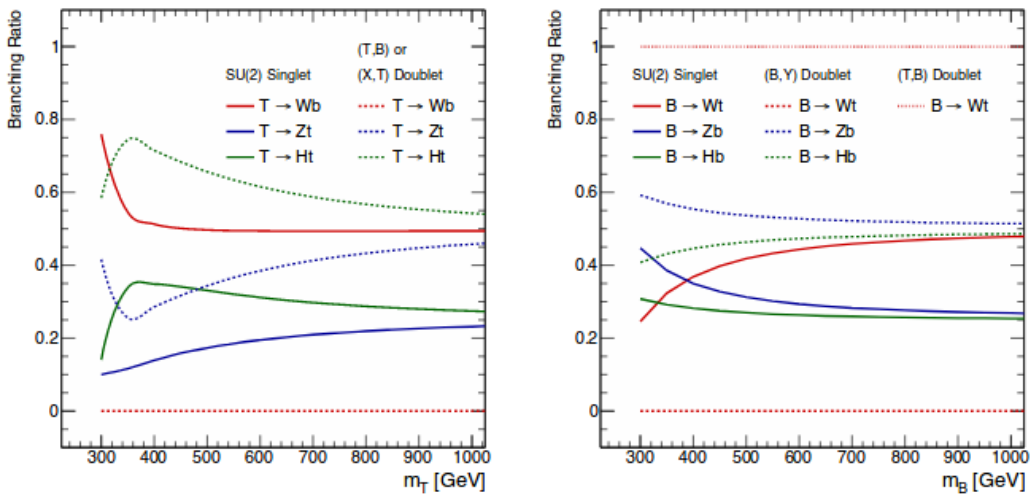
Table 3.4: Allowed decay modes for Vector-like doublets.

3.4 Search for VLQs

Both at Tevatron and LHC searches have been conducted for VLQs. The main studies at Tevatron have been done during the Run II, with a $\sqrt{s} = 1.96\ TeV$ and a luminosity of 5.4 fb^{-1} by the CDF and D0 experiments. At LHC with a $\sqrt{s} = 13\ TeV$ and a instantaneous luminosity of $10^{34}\ cm^{-2}s^{-1}$, studies by ATLAS and CMS have been done and are still ongoing. The main searches are based on QCD pair production since the cross section doesn't depend on the model. However, single production channels can prove to be complementary thanks to the higher mass reach and potential sensitivity to wider width samples. The analysis strategies have required a lot of different final states:

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.



(a) (b)

Figure 3.4: Branching ratio of vector-like top (a) and bottom (b) partners as a function of the heavy quark mass m_T and m_B respectively for isosinglets and isodoublets.

all-hadronic searches, single- or multi-lepton final state, with or without transverse missing energy.

The possible decays that have been studied for the $T\bar{T}$ pair production are:

- The decay of at least one T in Ht or a T in tZ, with $H \to bb$ and $Z \to \nu\nu$ respectively.It was required in the final states at least one lepton, from the top decay, multi-jets and missing transverse energy [9].
- The decay of $T\bar{T}$ both in the Wb channel, where one W decays to leptons and one decays to quarks, assuming a $Br(T \to Wb) = 100\%$ [7].
- The study of the $Zt + X$ final state with exactly one charged lepton and $Z \to \nu\nu$. The analysis has provided an upper limits on the T mass of 0.85 (1.05) TeV, considering the weak-isospin singlet (doublet) model; while an upper limit of $m_T \leq$ 1.16 TeV for the pure Zt decay mode [8].
- The decay of the T pair to $bWbW \to bl\nu\bar{b}qq'$. In the final state are required just one charged lepton, at least 4 jets and a boosted W-tagged. The assuming a branching fraction $Br(T \to bW) = 100\%$ and the reconstructing the mass of the T quark, thanks to a kinematic fit, have provided an upper limit on the T quark mass of 1295 GeV [38].
- The pair production of T or B in fully hadronic final states, using a multiclassification algorithm to improve the reconstruction of the events [42].
- The decay of a T pair, where one decays via $T \to tZ$ and the other one via $T \to tZ/bW/tH$ with two oppositely charged leptons in the final state coming from the Z boson [43]. This final state is studied also for the B pair production.
- The decay of a T pair considering three channels, corresponding to final states with a single lepton, two leptons with the same sign of the electric charge, or at least three leptons [40].

The searches for $B\bar{B}$ pair production have studied the decays into $tW/bZ/bH$. The analysis strategies follow the same of the $T\bar{T}$ searches.
The studies of single production of VLQs have analyzed the following channels:

- production of a T quark decaying to tZ with $Z \to ll$ and $t \to hadrons$ [39];
- production of $B \to bH$ with $H \to bb$ [4].
- production of a T quark decaying to tZ or tH in fully hadronic final states [41].

In the case of the searches for the pair production of charge 5/3 VLQs have been studied the decays into $WtWt$, the analysis covers the single lepton and dilepton same-sign channels [5][6].
In the Figures 3.5 and 3.6 are shown the results of the searches for the single and the pair production of VLQs by CMS in July 2019.

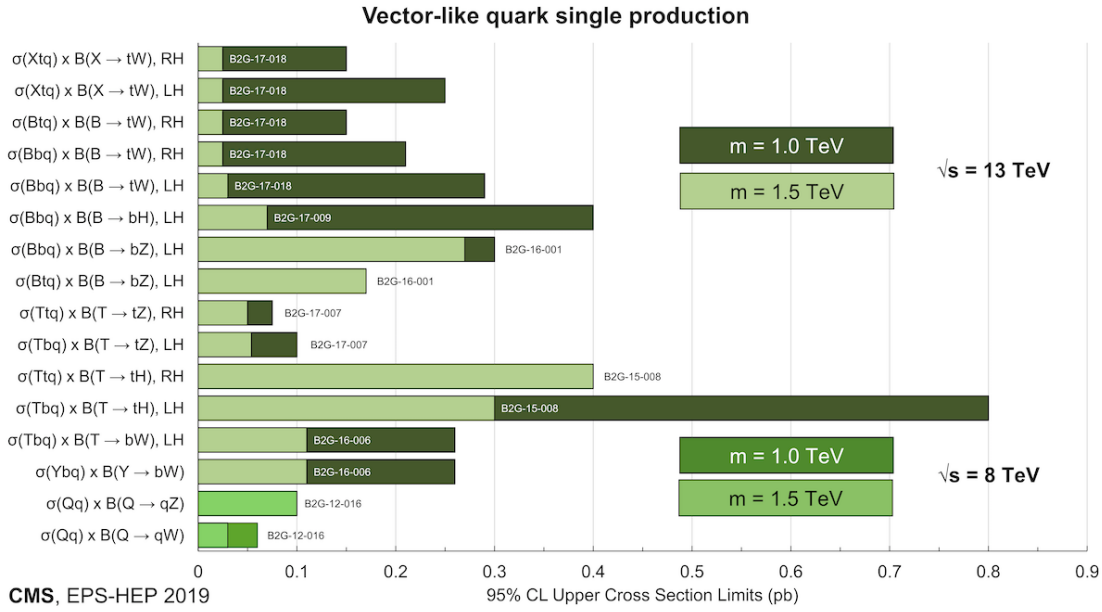


Figure 3.5: Summary of the results of the VLQs single production.

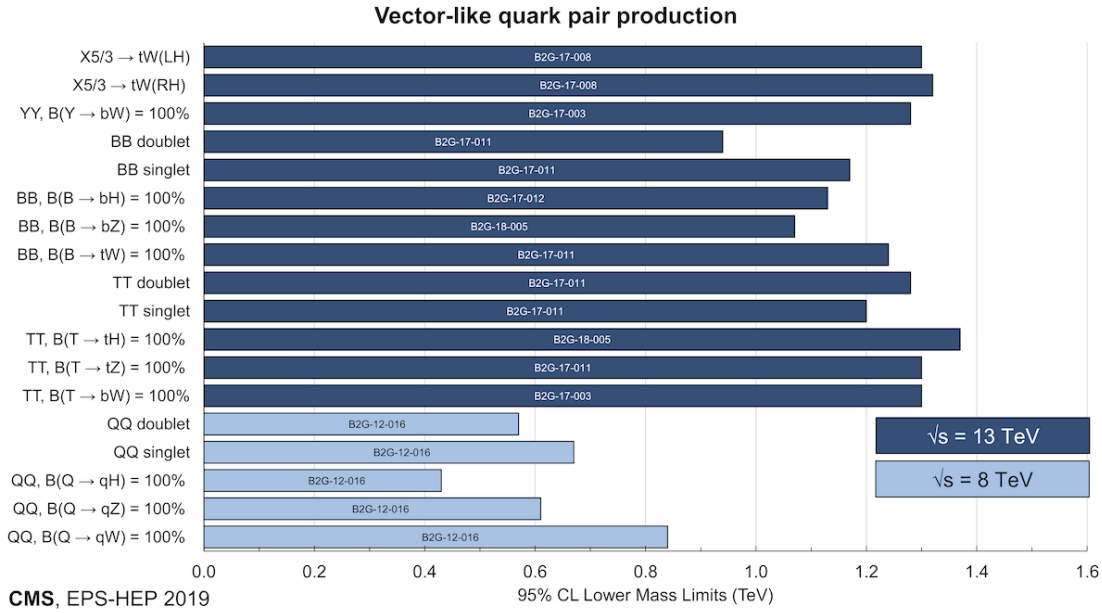


Figure 3.6: Summary of the result of the VLQs pair production.

Chapter 4

Physics object selection and reconstruction

The main goal of this thesis is the search for a singly-produced VLQ T. The decay channel analyzed is $T \to Zt$, in particular in cases with the hadronic decay of the Z boson and in the leptonic decay of the top quark. In Figure 4.1 it is shown the Feynman diagram of the process searched for.

The search has been done using pp collision data collected by the CMS experiment

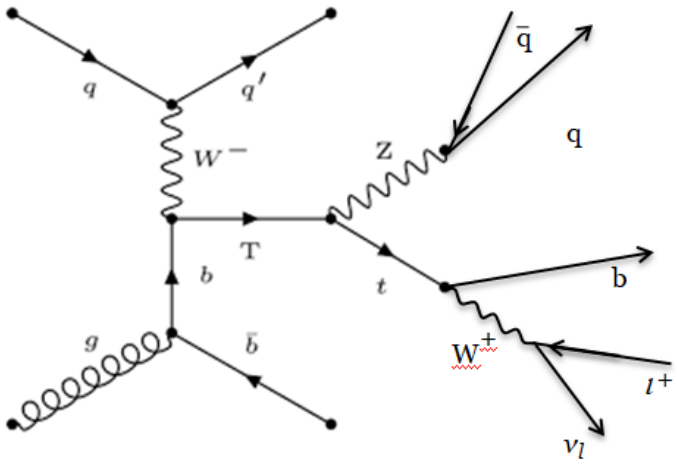


Figure 4.1: Feynman diagram of the searched process.

in 2018, in the pp collision at LHC with an integrated luminosity of 49.25 fb^{-1} and a centre-of-mass energy of $\sqrt{s} = 13\ TeV$.

This channel has been chosen due to the high branching ratio of the $Z \to \bar{q}q$ channel and the low contamination from QCD multijet background, obtained by means of requiring

the leptonic decay of the top quark. Moreover the products of the decay Z and t are expected to be "boosted", i.e. to have a non-negligible Lorentz boost in the laboratory centre-of-mass. This would allow us to study also the properties of this collimated final states. Two different methods have been used in order to improve the reconstruction of the top quark. In the first method several requirements have applied on different discriminant variables to provide an high signal efficiency ϵ_s, a low efficiency to mistakenly reconstruct top quarks from the backgrounds, ϵ_b, and an high signal accuracy (SA) defined as:

$$SA = \frac{N^{Sel}_{Real\ Top}}{N^{Sel}_{Top}}, \tag{4.1}$$

where $N^{Sel}_{Real\ Top}$ is the number of the top quark reconstructed matching the MC truth after the requirements and N^{Sel}_{Top} is the total number of reconstructed top quark.
The second approach to improve the selection of top quarks involves making use of machine learning algorithms. Both approaches are detailed in this Chapter.

4.1 Physics objects selection

The search for VLQ T is performed by looking for b-jets, muons, missing transverse energy, and Z-tagged jets. The stable particles produced in the pp collision are reconstructed and identified by the Particle Flow (PF) algorithm [35]. This algorithm combines the information from all the CMS sub detector to identify the particles and also to reconstruct their 4-momenta. The reconstruction is done in the following order, considering a list of PF candidates:

- muons: a track reconstructed from the hits in the tracker and in the muon system, is identified as a muon, then then the track is removed by the list of objects;
- electrons: a charged particle track is associated with one or more ECAL cluster is identified as an electron. Both the track and the clusters are removed from the list;
- charged hadron: the remaining tracks that are linked to ECAL and HCAL clusters are used to reconstruct charged hadrons. The momentum is directly evaluated by the tracker and then corrected using the information coming from the calorimeters;
- photons and neutral hadrons: cluster in the ECAL and HCAL unmatched with charged tracks are identified respectively as photons and neutral hadrons.

The information given by the PF algorithm are used to reconstruct high-level objects, such as the missing transverse energy, and the jets.

4.2 Jets

The quarks and the gluons, produced in the final state of hard pp interactions at the LHC, do undergo the hadronisation process before being able to directly interact with the detectors, and they can be indirectly observed through "jets", i.e. clusters of color-neutral hadrons. Perturbative theory and hadronization models can describe the interaction between the partons and the showering.
CMS provides different algorithms for the jet reconstruction, in this thesis only the jets reconstructed with the anti-k_T algorithm [17] have been used. This algorithm is the default choice at LHC, because of its robustness against pile up and underlying event contributions, its feature of providing conical jets, besides being infrared and collinear safe. This algorithm considers all the PF candidates for the clustering. It considers two distances d_{ij}, which is the distance between the particle of the PF candidate i and the particle j, defined as:

$$d_{ij} = min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{(\Delta R_{ij})^2}{R^2} \tag{4.2}$$

where R is a parameter in the (η, ϕ) metrics, and $(\Delta R_{ij})^2$ is

$$(\Delta R_{ij})^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2,$$

and the distance between the particle i and the beam:

$$d_{i,B} = \frac{1}{p_{T,i}^2}. \tag{4.3}$$

If $d_{ij} < d_{iB}$ the entities i and j are recombined, if it is not i is called jet and it j is removed from the list of entities. The distances are recalculated and the procedure repeated until no entities are left and the final jet is reconstructed.
Two different kinds of jets have been used for the analysis: jets with $R = 0.4$ are the AK4 jets, also referred as "narrow jets", and jets with $R = 0.8$ are the AK8, or "fat jets". A set of identification criteria ("loose" working point), provided by the CMS dedicated group that works on Jets and MET, is applied on AK4 jets collection. Moreover, different corrections are applied while calculating the energy of the jets at various levels. Jet Energy Corrections (JEC) are used to scale the jet energy by a factor that describes the detector response depending on the transverse energy and the pseudorapidity of the jet. The correction factor Jet Energy Scale (JES) is calibrated on the η, p_T, energy density, and area of the jet; the Jet Energy Resolution (JER) for the simulated jets is degraded to reproduce the resolution observed in data. In Figure 4.2 is shown a schematic view of jet evolution.

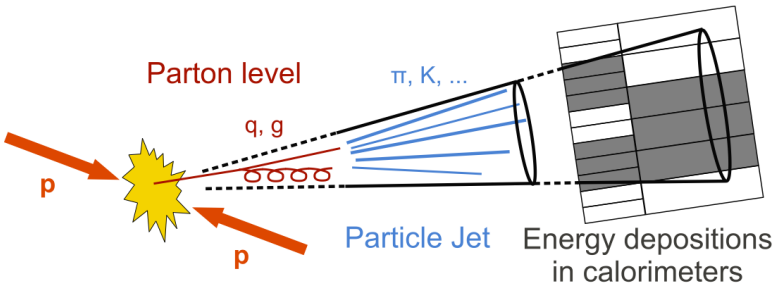


Figure 4.2: Schematic view of jet evolution.

4.2.1 Z tagging

The events selection starts with the Z tagging of an AK8 jet, which allows to reconstruct its hadronic decays to a pair of quark-antiquark. It was required for the jet:

- $60 < M_{SD} < 105\ GeV$, where M_{SD} is the mass of the jet calculated with the Soft Drop algorithm [28];
- $p_T > 200\ GeV$;
- $\frac{\tau_2}{\tau_1} < 0.45$, where τ_N is the N-subjettiness of the jet.

The N-subjettiness is defined as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} min\{\Delta R_{1,k}, \ldots, \Delta R_{N,k}\}, \tag{4.4}$$

where d_0 is a normalization constant, and $\Delta R_{i,j}$ are the distances between a particle j clustered in the jet and a main axis i. The N-subjettiness indicate the proximity of the jet to a particular hypothesis on the number of sub-jets of which it is composed, and which are therefore less when the jet is more compatible with hypothesis N. Thanks to the presence of two sub-jets, directed along the momenta of the two quarks, in the Z-jets it is expected $\tau_2 < \tau_1$ while in a QCD jet it is expected the opposite as shown in Figure 4.3.
This is a "loose" working point already used in the previously analysis at CMS.

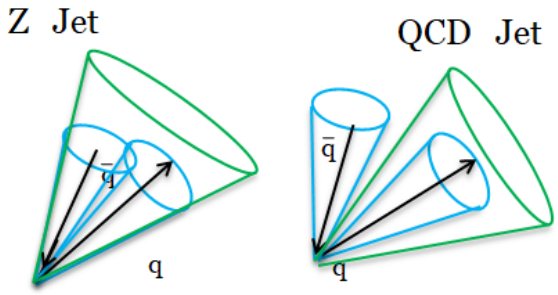


Figure 4.3: Main axis in a AK8 Z-Jet and in a QCD Jet.

4.2.2 b-tagging

To reconstruct the top quark leptonic decays it is required a jet AK4 b-tagged. To identify the jets from a b quark, CMS provides the Deep Combined Secondary Vertex (DeepCSV) algorithm [36]. It uses deep machine learning algorithm and the discriminating variables exploit the fact that long living particles, such as B-hadrons, travel a considerable distance from the primary vertex before their decay happens. The Impact Parameter (IP), Figure 4.4, is the variable used to define the distance between the two vertices. It is a Lorentz invariant and so it is also invariant with respect to changes of the long lived particle kinetic energy. The typical value for the B-hadrons corresponds to $c\tau \sim 450\ \mu m$ that, in CMS, can be measured with precision between 30 μm and hundreds μm.

The AK4 jets are considered coming from b-qaurk if they pass a given threshold on the

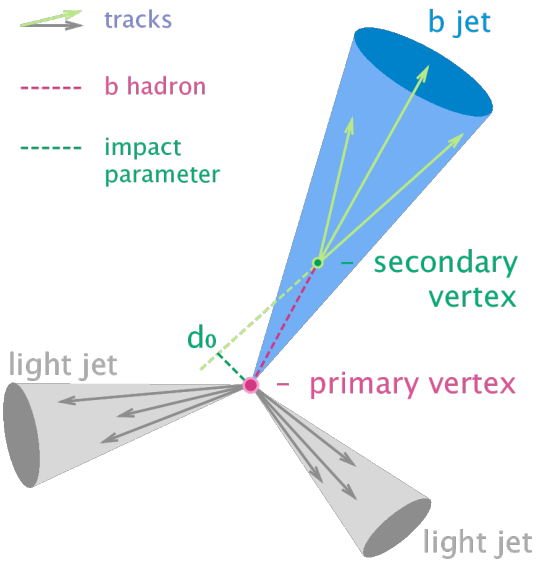


Figure 4.4: Impact Parameter.

value of DeepCSV, which is defined by the CMS group working on b-tagging based on the selection efficiency of b-originated jets and the mis-tagging efficiency of light quark-originated jets. The chosen working is "Loose", meaning it provides a misidentification probability less than 10%. Moreover,in order to avoid that one or more b-jets coming from the possible decay channel $Z \to b\bar{b}$ were reconstructed as narrow jet, the separation distance between b-tagged jet and Z

$$\Delta R(bjet, Z) = \sqrt{(\eta_b - \eta_Z)^2 + (\phi_b - \phi_Z)^2}$$

is required to be greater than 1.2.

4.3 MET

At LHC, as for any hadron collider, the Z-component of the colliding proton momentum is known, the X and Y components are zero. Momentum conservation in the X-Y plane allows to infer the momentum lost due to particles escaping the detector. Along the Z-axis on the other hand, the portion of momentum lost due to proton fragment ending up in the beam pipe is not measurable. The missing transverse energy (MET) is reconstructed by the particle flow algorithm using the vectorial sum of the transverse momentum of all PF candidates. However the MET is very sensitive to detector effects that impact the kinematic closure, like mis-calibrations or parts of the detector that are inactive or damaged. As done for jets, energy corrections are propagated to the MET, in particular the JEC. In this thesis to take in account the presence of the neutrino in the top quark decay the MET is required to be greater than 70 GeV. During the data taking in 2018, there was a malfunction in a detector region, to take this into consideration some events have been rejected and the MET has been reweighted.
Moreover it is also a crucial variable to search for Beyond Standard Model physics, since many models predicts particles that would not leave any trace in the detector. The presence of particles that do not interact in the detector can be measured with MET since they create an energy imbalance.

4.4 Muons

The muons are detected both in the tracking system and in the muon system. In the first detector they are reconstructed using a technique based on the Kalman filter algorithm, in the second one the hits from DTs, CSCs and RPCs are fitted reconstructing the *Stand Alone Muons*.
The muons could be reconstructed as *Tracker Muon* if the muons are reconstructed in the tracker system, and the track is matched with a single hit in the muon system. The

Global Muon reconstruction is given by the refitting of two tracks: a *Tracker Muon*, and *Stand Alone Muon*, that are matched together. This different kind of reconstruction defined identification criteria used in CMS, together with other variables such as the number of pixels hit. In this thesis to reconstruct the muons form the top quark decay are also used muons with "*tight*" selection requirement [37] and always with p_T greater than 10 GeV. The isolation requirement will be described in the next section, since they depend on the ΔR between the muon and the b-tagged jets used for the top quark reconstruction.

4.5 Top Quark Reconstruction

The top quark is reconstructed as composed by a muon, the b-jet and the MET. The three 4-momenta are added together to evaluate the top quark 4-momentum. Moreover, since the neutrino momentum along the beam axis is not computable, it has been chosen to estimate it imposing $\sqrt{s}(\mu,\nu) = 80,4\ GeV$, which is the W boson mass. Here we do choose the approximation that the W boson width is negligible with respect to the experimental resolution effects involved in the top quark reconstruction. This leads to the follow equation:

$$p_{z,\nu} = \frac{\Lambda p_{z,l}}{p_{T,l}^2} \pm \frac{1}{p_{T,l}^2}\sqrt{\Lambda^2 p_{T,l}^2 - p_{T,l}^2(E_l^2\ \not{E}_T^2 - \Lambda^2)}, \tag{4.5}$$

where:

$$\Lambda = \frac{m_W^2}{2} + p_{T,l}^{\vec{}} \cdot \not{p}_T^{\vec{}}. \tag{4.6}$$

m_W is the mass of the W boson, E_l and $p_{T,l}$ are the energy and the transverse momentum of the lepton and $\not{E}_T$ is the MET. The discriminant in Equation 4.5 is usually positive and the solution with the smallest absolute value is chosen. In the other cases, the imaginary component is eliminated by imposing that the discriminant, as the square-root term, is null behaving a quadratic relation between $p_{x,\nu}$ and $p_{y,\nu}$, with two possible solution and one remaining degree of freedom. The solution that leads to a minimum vectorial distance between $\vec{p_{T_\nu}}$ and $\not{\vec{p}}_T$ is chosen.
Moreover the reconstructed top quark candidate is classified as:

- Top Resolved, if $\Delta R(\mu, b-jet)$ is higher than 0.4 and less than 2, therefore the muon is not included in the AK4 jet's cone as shown in Figure 4.5.
- Top Merged, if $\Delta R(\mu, b-jet)$ smaller than 0.4 so the muon is included in the jet's cone as shown in Figure 4.6. It is expected this configuration for boosted top quarks, from high mass VLQ T decaying.

The muons selection have been improved separately in the two configurations, to take in account the differences between them. For both resolved and merged, a selection has

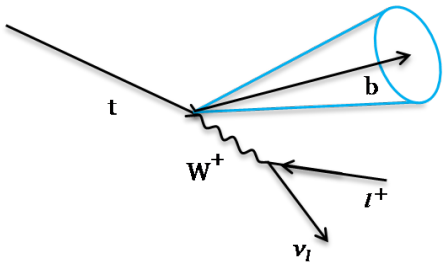


Figure 4.5: Top Resolved configuration.

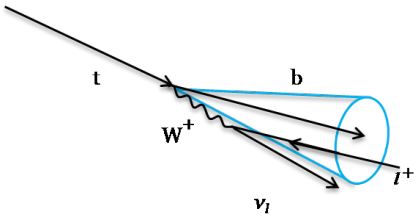


Figure 4.6: Top Merged configuration.

been defined in order to reject top quarks combinatorial backgrounds. In the sample used, reconstructed muons and b-jets are matched with their counterparts produced in the hard scattering by requiring an angular separation less than 0.4 in the ΔR metrics. Different sources of combinatorial background are identified, and the reconstructed top quark candidates are divided into:

- Top True, if both the muon and the b-tagged jet match with the MC truth muon and b-quark from a top quark decay as shown in Figure 4.7.

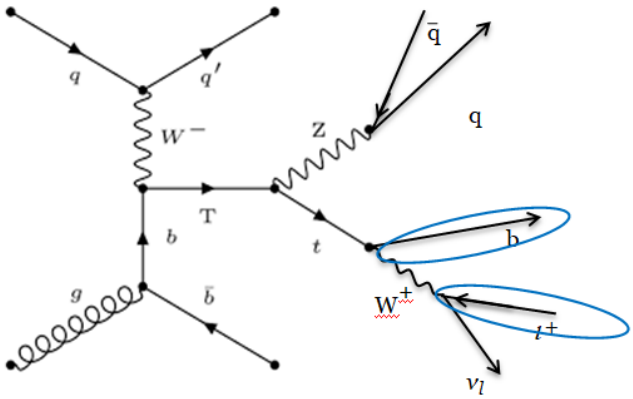


Figure 4.7: Top True configuration.

- Top Jet Match, if only the b-tagged 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 muon from the top quark could have similar properties to the muons coming from pions and kaons decay from B mesons decay chains, naturally present in all b-originated jets. In Figure 4.8 is shown this kind of combinatorial background.

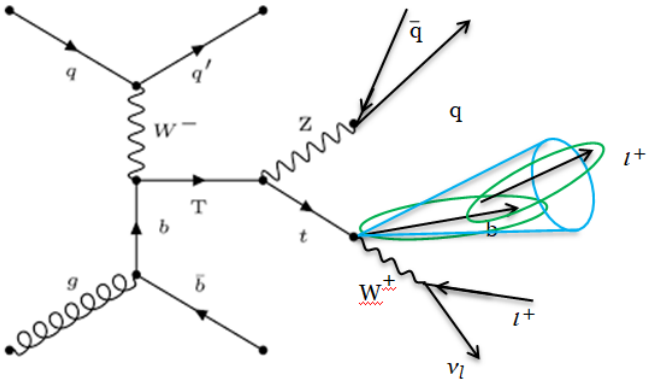


Figure 4.8: Top Jet Match configuration.

- Top Muon Match, if only the muon matches the MC truth, and the b-tagged jet could be a mistagged jet or a real b-jet coming from other processes, e.g., for signal events, the other jet coming from $g \to b\bar{b}$ as shown in Figure 4.9.

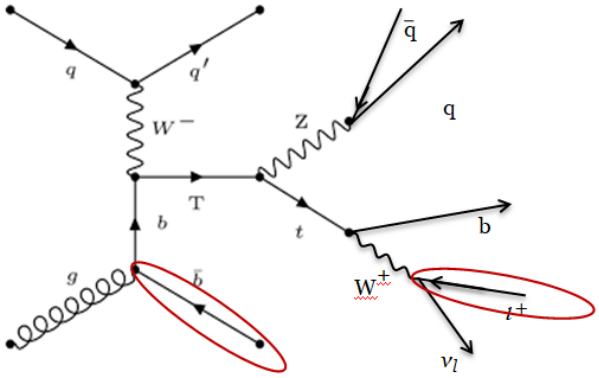


Figure 4.9: Top Muon Match configuration.

- Top False, if both don't match the MC truth as shown in Figure 4.10.
- Top Comb, if both match the MC truth, but they came from two different top quarks. It is not expected in the signal, but in other background events such as the $t\bar{t}$ events.

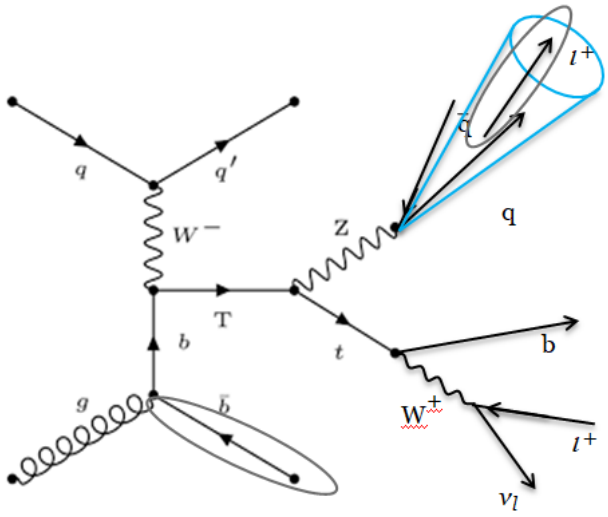


Figure 4.10: Top False configuration.

Therefore the selections is optimized separately in the two cases, Merged and Resolved, in order to select Top True and to reject the combinatorial backgrounds. The studies have been done on $t\bar{t}$ events, focusing on events with the invariant mass of the $t\bar{t}$ pair greater than 1000 GeV, which have a similar kinematic range as the signal ones.

4.5.1 Top Resolved selection

For the Top Resolved configuration the kinematic variables of the muons have been studied, as well as the so-called "isolation variables", which gauge the energy deposited around the muon by jet-originated particles, and that play an important role in the discrimination. The isolation variable Iso04, where 04 refers to a cone of $R = 0.4$ in the metric (η, ϕ) around the muon momentum direction, is defined as:

$$\sum \frac{p_{T_{\gamma+Had}}}{p_{T_\mu}} \tag{4.7}$$

where the sum is done for all the charged hadrons, photons and neutral hadrons, if the distances between them and the muon is less than R. Since the muons are outside the jet, this variable is expected to assume much lower values for the so-called *prompt muons*, i.e. muons originated from the primary interactions, than for $non-prompt$ ones, i.e. originated from secondary decays from hadrons.
The relative transverse momentum is also used, defined as:

$$\frac{\|\vec{p}_\mu \times \vec{p}_{Jet}\|}{\|\vec{p}_{Jet}\|}. \tag{4.8}$$

The selection has been done choosing couples that have:

- *Tight* ID for the muon;
- $\sum \frac{p_{T_{\gamma+Had}}}{p_{T_\mu}} < 0.1$;
- $\frac{\|\vec{p}_\mu \times \vec{p}_{Jet}\|}{\|\vec{p}_{Jet}\|} > 20\ GeV$;
- $p_{T\,Jet} > 30\ GeV$.

The result of this selection is shown in Figure 4.11. These requirements lead to a ϵ_s

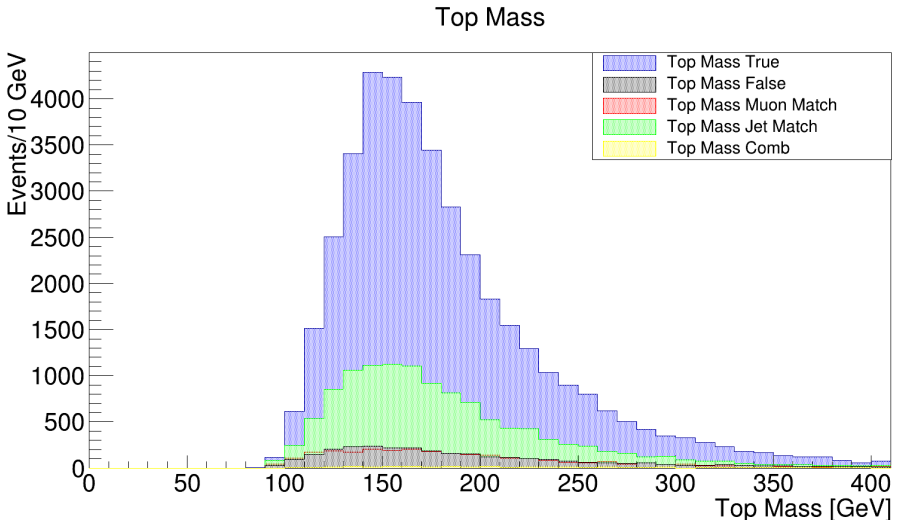


Figure 4.11: Top Mass distribution in the region with just one reconstructed top quark and in resolved configuration.

equal to 68%, ϵ_b equal to 9% and a SA equal to 62%.

4.5.2 Top Merged selection

The Top Merged selection has also been performed using kinematic variables of the muon and the jet, isolation variables and the impact parameters of the muon. Instead of using the isolation in a $R = 0.4$ cone, it has been used the MiniIso, defined in the same way of the Iso04, except that R is varying as a function of the muon p_T, in particular $R/propto\frac{1}{p_{T_\mu}}$. The impact parameters used are d_{xy} and d_B, which are both the transverse impact parameters but computed with different tools. Moreover the computation of d_B allow to know also the error of this parameter, which is the d_{Berr}. However in Figure 4.12 is shown that distribution of the ΔR between the reconstructed muon and the b-tagged jet, after matching the MC truth, is different from the distribution of ΔR computed using the MC information of the particles produced (i.e. the p_T or the pseudorapidity).

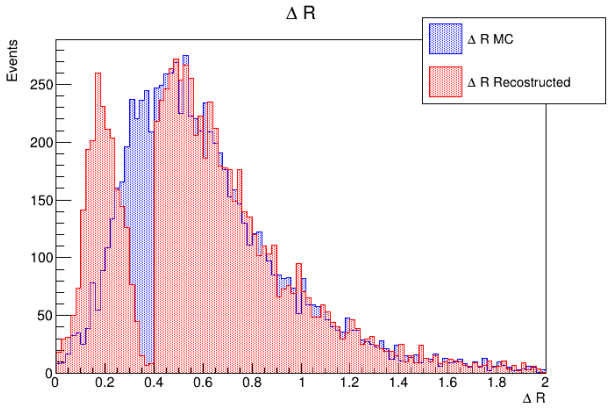


Figure 4.12: Comparison between $\Delta R(\mu, b-jet)$ after matching the MC truth with the reconstructed objects in red and the ΔR computed using the MC information in blue.

The discrepancy between the MC ΔR and the reconstructed one starts at 0.4, therefore it is a big problem in the Top Merged configuration. In the Top Merged configuration the jet clustering algorithm also considered the muon from the W boson. To better clarify this, one can consider as an example the limit scenario where the distance between muon and the main axis of the b-originated shower is exactly $\Delta R = 0.4$. In this case, the the jet clustering algorithm, including the high momentum muon from the W boson in the reconstruction, will set a new axis in the middle of the two directions and this explain the peak at $\Delta R = 0.2$ in the plot shown. To correct this effect, at the first order, it is possible to subtract the muon 4-momenta from the b-tagged jet 4-momenta, as shown in Figure 4.13 the ΔR should be closer to the one computed with the MC truth information.

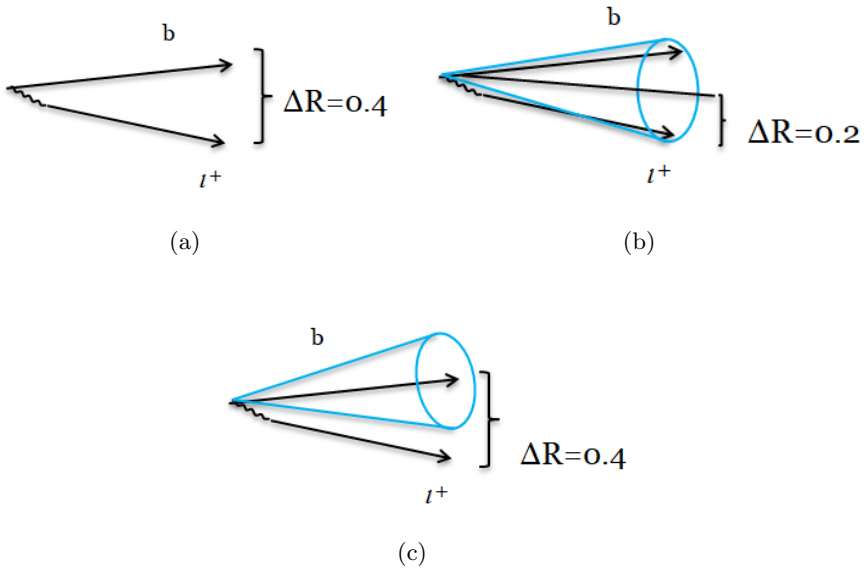


Figure 4.13: Scheme of the ΔR computed with the MC information (a), with the objects reconstructed (b) and with the objects reconstructed after the correction.

The comparison between the distance corrected and the MC one is shown in Figure 4.14, and it is possible to note a better agreement. A further improvement could be the reclustering of the jet with the AK4 algorithm not considering the muon in the list of the objects, but this goes beyond the aims of this thesis.

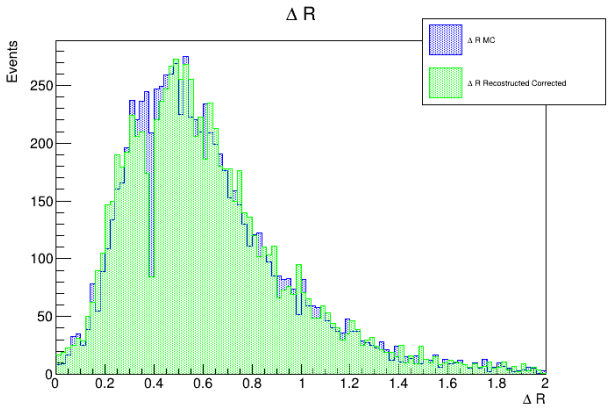


Figure 4.14: Comparison between $\Delta R(\mu, b-jet)$ after matching the MC truth and after the correction of the objects reconstructed in green and the ΔR computed using the MC information in blue.

In this configuration the selection has been done requiring:

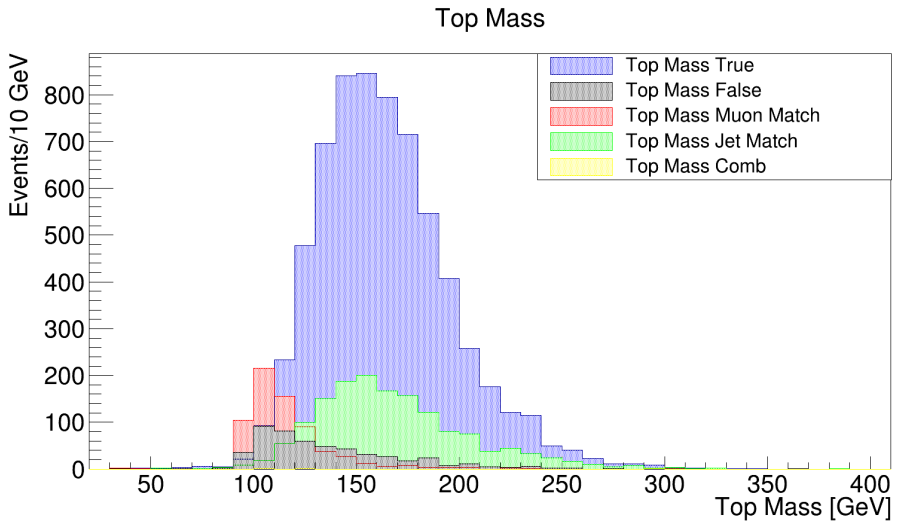


Figure 4.15: Top Mass distribution in the region with just one reconstructed top quark and in merged configuration.

- $\left|\frac{d_B}{d_{Berr}}\right| < 2.5$;
- $|d_{xy}| < 0.002\ cm$;
- $\frac{p_{T\,\mu}}{p_{T\,Jet}} > 0.1$;
- $\frac{\|\vec{p}_\mu \times \vec{p}_{Jet}\|}{\|\vec{p}_{Jet}\|} > 5\ GeV$;
- $\sum \frac{p_{T_{\gamma+Had}}}{p_{T_\mu}} < 1$, in a cone with $R \propto \frac{1}{p_{T\,\mu}}$

In Figure 4.15 is shown the distribution of the top quark reconstructed mass.

This requirements lead to a ϵ_s equal to 72%, ϵ_b equal to $2,7\%$ and a SA equal to 68%.

4.6 Boosted Decision Tree

To improve the top quark reconstruction, a machine learning algorithm was used. The idea is to have an higher signal efficiency and a lower background efficiency than the previous selection method. The algorithm used is a Boosted Decision Tree [30]. The algorithm receives as input a set of variables, in order to exploit simultaneously their single information, as well as the correlations amongst them. Then the algorithm computes an output, combining the input variables. Performing a requirement on the output, whose value is between 1 and 0, it is possible to have an improved signal selection. This is a supervised machine learning algorithm. The first step is the training, where a sequence of selection requirements are applied; each requirement splits the sample into nodes, each of which correspond to a given number of observations classified as signal or as background. If in the nodes signal or background dominant they are classified as leaf and no more requirements are applied, or they could be splitted again with another requirement. Each branch of the tree represents on sequence of requirements as shown in Figure 4.16.

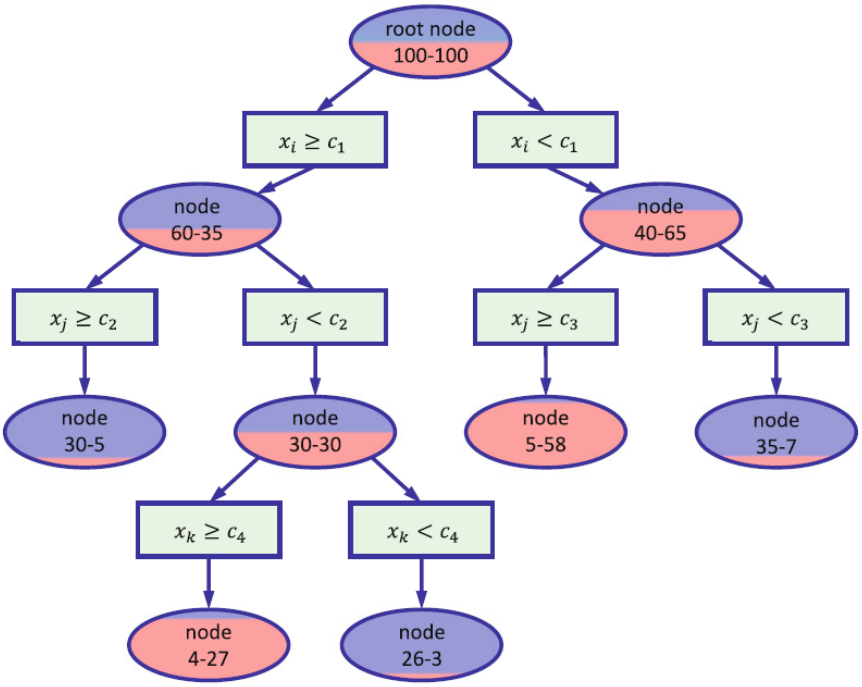


Figure 4.16: Schematic example of a tree [30].

To understand which sequence provides the best selection it is used a metric, the

log loss defined as:

$$E = -(p * ln(p) + (1 - p) * ln(1 - p)), \tag{4.9}$$

where p is the purity if the node, in this binary case the fraction of signal events. To evaluate the gain of the splitting of the node A into the nodes B and C, it could be computed the ΔE:

$$\Delta E = E(A) - E(B) - E(C), \tag{4.10}$$

maximizing it the purity of the nodes C and B increases.
To improve the selection a boosting procedure to the tree is applied. The boosted procedure consists in different steps:

- the training observations are reweighted using the results of the previous trees.
- A new tree is built with the reweighted sample as training.
- To each tree is given a score w.
- The final output is the weighted average of the all single tree outputs:

$$y(\vec{x}) = \sum_{k=1}^{N_{trees}} w_k C^k(\vec{x}), \tag{4.11}$$

  where $y(\vec{x})$ is the final output and $C^k(\vec{x})$ is the output of the k-tree.

The boosted procedure used is called the "*XGBoost*", the different boosting depend on the different methods used to compute the weight w.
To improve the efficiency of the BDT algorithm the follow parameters have been set: maximum depth of a tree equal to 4, the minimum sum of instance weight needed in a child equal to 4, a learning rate equal to 0.2 and number of trees equal to 100. Moreover the setting of this parameters help to avoid the *overtraining*, it happens when the output corresponds too closely or exactly to a particular set used for the training as shown in Figure 4.17.

4.7 ML Top Quark Reconstruction

As previously mentioned the BDT algorithm has been used to improve the top quark reconstruction, using $t\bar{t}$ background sample and different VLQ T left-handed samples for the trainings. The training sessions have been separate for the two different configurations: Merged and Resolved. In both categories a set of variable as input has been chosen to maximize the accuracy as reported in Table 4.1, where $\theta^*_{\mu Jet}$ is the angle between the muon and the b-jet in the top quark reconstructed frame. Moreover to take

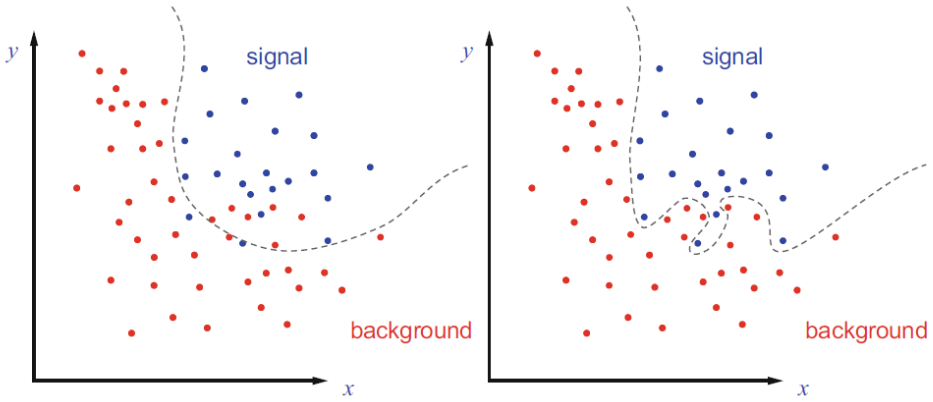


Figure 4.17: On the left smooth selection based on a right algorithm while on the right an example of overtraining [30].

Top Resolved	Top Merged
Iso04 of the muon	MiniIso of the muon
Impact parameters of the muon	Impact parameters of the muon
Top Quark reconstructed mass with and without MET	Top Quark reconstructed mass with and without MET
$p_{T_\mu}, p_{T_{Jet}}$	Top Quark reconstructed 4-momentum
Relative transverse momentum	Relative transverse momentum
$cos\theta^*_{\mu Jet}$	$cos\theta^*_{\mu Jet}$

Table 4.1: Table of the variables used in the trainings.

in account p_T dependence, the data were analyzed separately in three different bins of the transverse momentum of the top quark reconstructed:

- Low, $p_T < 200\ GeV$;
- Medium $200 < p_T < 500\ GeV$;
- High $p_T > 500\ GeV$.

Due to the large combinatorial background, different preselection requirements were performed in the two configurations.

4.7.1 Top Resolved

For the Top Resolved configuration it was required muons with:

- *Tight* ID;
- $\sum \frac{p_{T_{\gamma+Had}}}{p_{T_\mu}} < 0.2$;
- $\frac{\|\vec{p}_\mu \times \vec{p}_{Jet}\|}{\|\vec{p}_{Jet}\|} > 10\ GeV$,

it can be noted that this requirements are looser than the selection performed in the previous case. Muons used in this training are therefore a subset of the ones previously described.
In the low p_T region the accuracy, choosing an output higher than 0.4, is equal to 60%. In Figure 4.18 is shown the distribution of the BDT output.

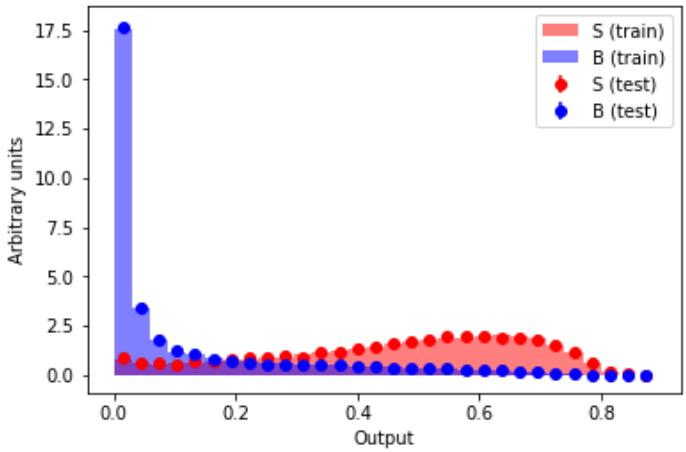


Figure 4.18: BDT output in the low region for the Resolved configuration.

The accuracy in the medium p_T region is equal to 76%, it could be seen that a better selection could be done in this region, having chosen the requirement on the output $label > 0.6$. In Figure 4.19 is shown the distribution of the BDT output.

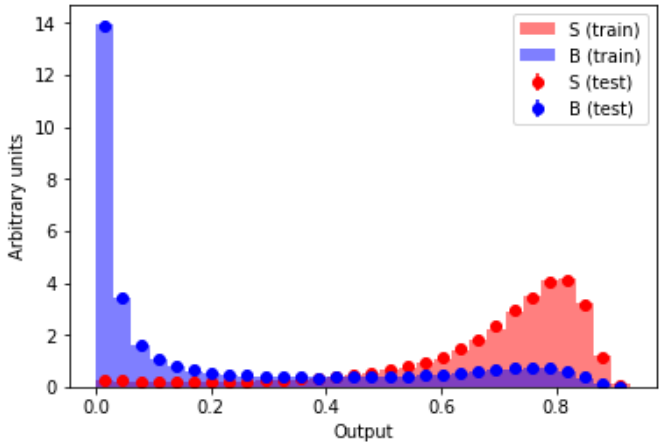


Figure 4.19: BDT output in the medium region for the Resolved configuration.

In the end the high p_T were performed choosing a value of the *label* higher than 0.6. In Figure 4.20 is shown the distribution of the BDT output.

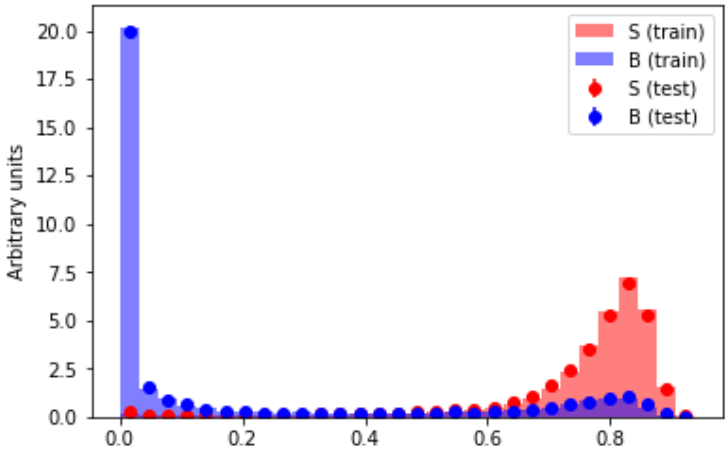


Figure 4.20: BDT output in the high region for the Resolved configuration.

The combination of all these selections leads to a signal efficiency equal to 65%, while the background efficiency now is 4.4%. In Figure 4.21 the top quark reconstructed mass is shown. In Table 4.2 the comparison between the first selection and the selection using the ML algorithm. It can be noted that the ϵ_s is smaller but this second selection allow to have a better SA and a smaller ϵ_b.

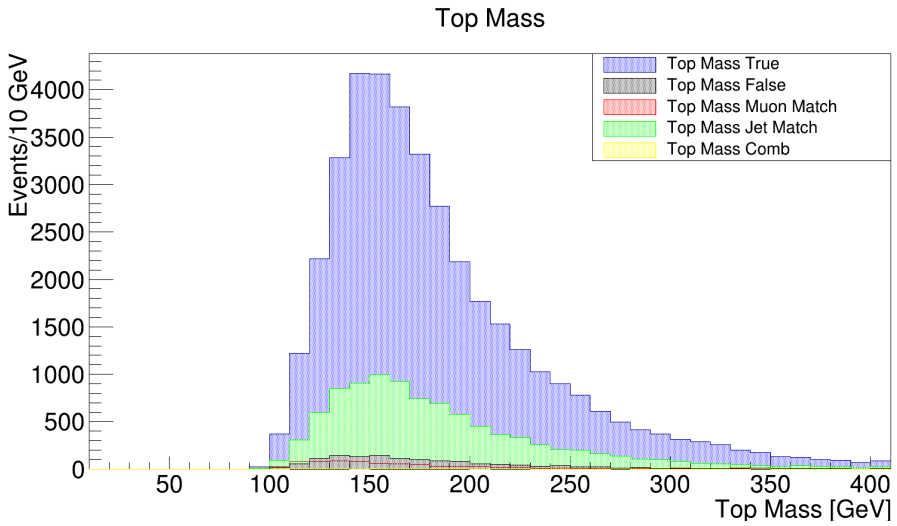


Figure 4.21: Top Mass distribution after the ML selection.

	Old Resolved Selection	ML Resolved Selection
ϵ_s	68%	65%
ϵ_b	9%	4.4%
SA	62%	75%

Table 4.2: Comparison between the old selection and the ML selection with just one reconstructed top quark and in resolved configuration.

4.7.2 Top Merged

The muons selected for the Top Merged trainings were required to have:

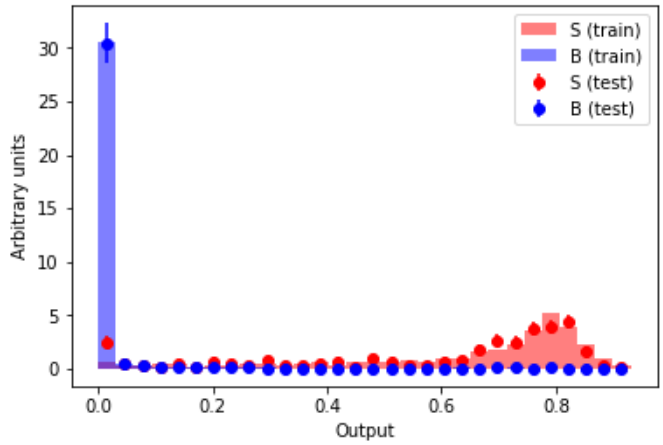
- $|d_{xy}| < 0.01\ cm$;
- $\sum \frac{p_{T_{\gamma+Had}}}{p_{T_\mu}} < 6$, in a cone with $R \propto \frac{1}{p_{T_\mu}}$.

In addition, two different training sessions were done:

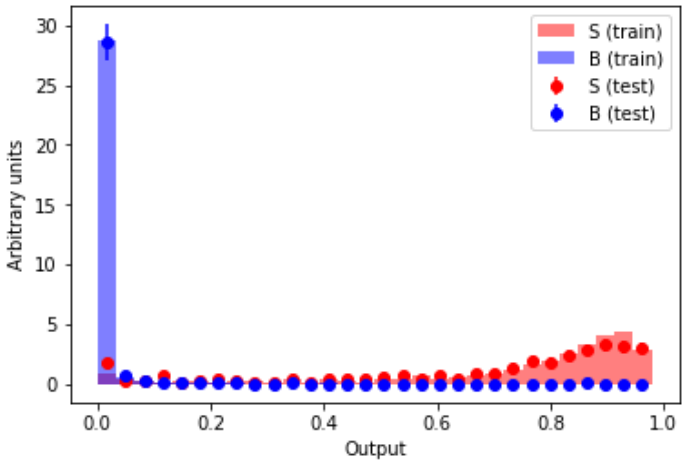
- True vs Jet Match;
- True vs Muon Match.

In the case of True vs Jet match, three separate trainings were done per p_T bins. In the case of True vs Muon match, only one training was performed for the second and the third bin of transverse momentum to avoid overtraining. Only the reconstructed top quarks that passes the signal selection in both categories have been selected. It is expected that with this two trainings also the number of Top False reconstructed decreases, because the first training let us know the properties of the *prompt muons*, while the second one the properties of the right b-Jets.

In the low p_T region the accuracy, choosing the output of the *True vs Jet Match* training higher than 0.6 and of the *True vs Muon Match* training higher than 0.7, is equal respectively to 75% and 83%. In Figure 4.22 are shown the distributions of the BDT outputs. It can be noted that the output is not very performing, being the merged configurations expected for very boosted top quarks.



(a)

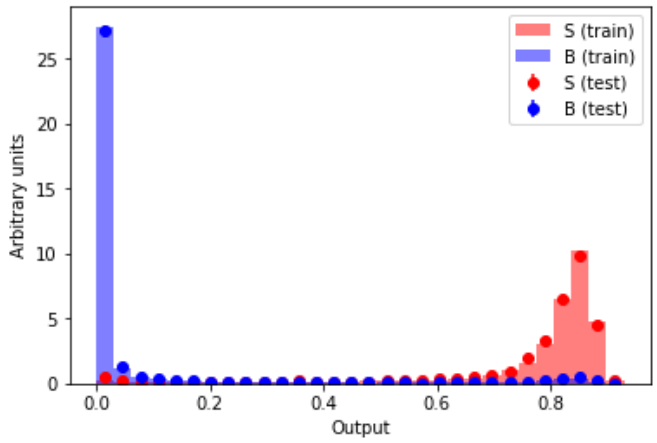


(b)

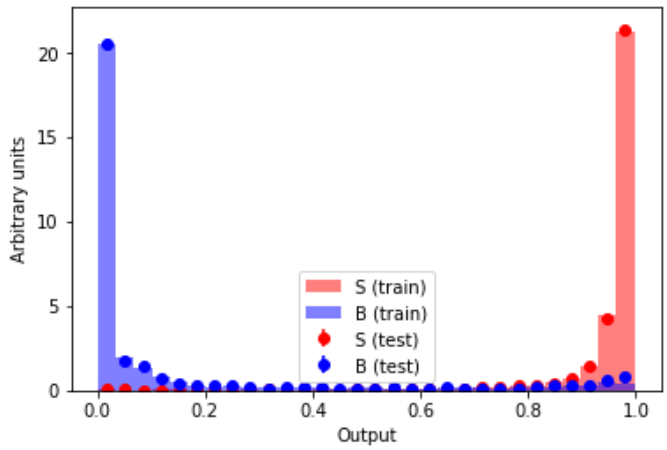
Figure 4.22: BDT output in the low region for the Merged configuration trainings, (a) *True vs Jet Match* and (b) *True vs Muon Match*.

The accuracies in the medium p_T region are equal to 83% and 97%, respectively for the *True vs Jet Match* and *True vs Muon Match* trainings. It could be seen that a better selection could be done in this region, having chosen the requirement on the output $label[Jet\ Match] > 0.75$ and $label[Muon\ Match] > 0.8$. In Figure 4.23 is shown

the distribution of the BDT outputs.



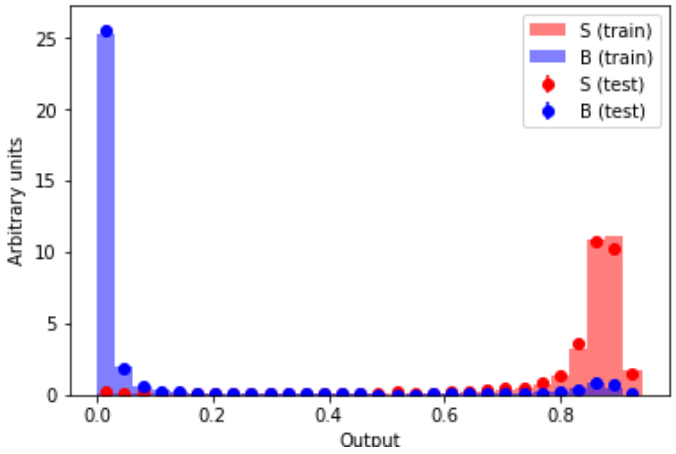
(a)



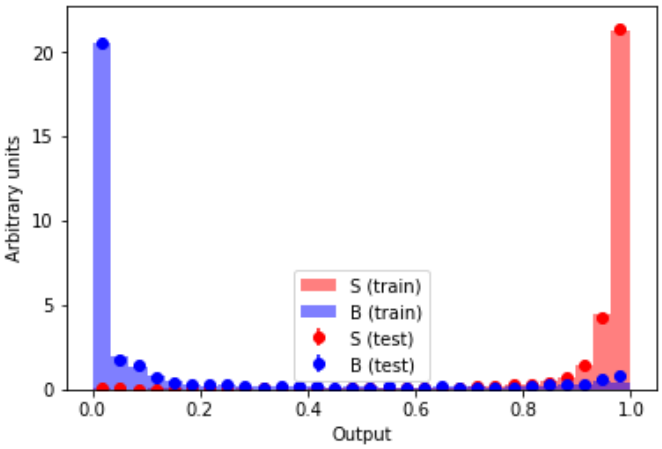
(b)

Figure 4.23: BDT output in the medium region for the Merged configuration trainings, (a) *True vs Jet Match* and (b) *True vs Muon Match*.

In the end the high p_T were performed choosing a value of the $label[Jet\ Match]$ higher than 0.75 and the $label[Muon\ Match] > 0.8$ in order to have the efficiencies respectively equal to 86% and 97%. In Figure 4.24 are shown the distributions of the BDT outputs.



(a)



(b)

Figure 4.24: BDT output in the high region for the Merged configuration trainings, (a) *True vs Jet Match* and (b) *True vs Muon Match*.

The combination of all these selections lead to a signal efficiency equal to 72%, while the background efficiency now is 1.5%. In Figure 4.25 the top quark reconstructed mass to compare the old selection and the ML selection.

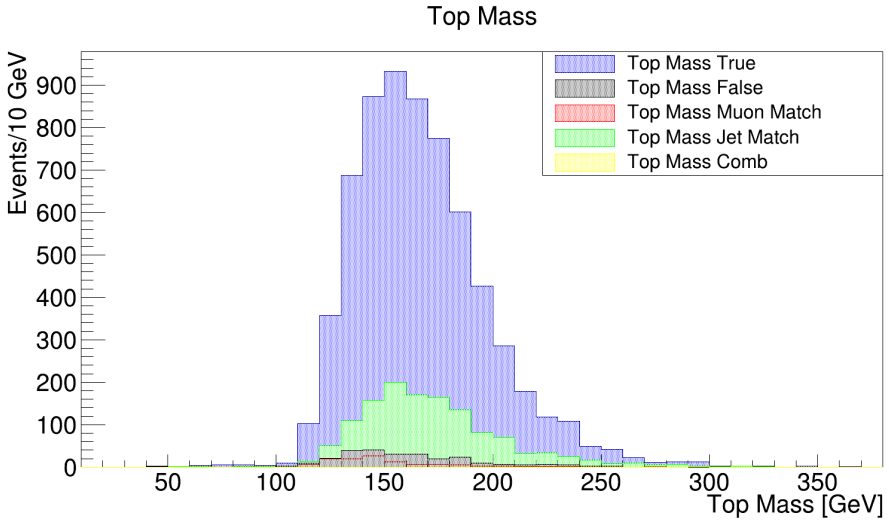


Figure 4.25: Top Mass distribution after the ML selection with just one reconstructed top quark and in merged configuration.

In Table 4.3 the comparison between the first selection and the selection using the ML algorithm. It can be noted that the ϵ_s is similar while the background is reduced of $\approx 50\%$, with an increase of the SA.

	Old Merged Selection	ML Merged Selection
ϵ_s	72%	71%
ϵ_b	2.7%	1.5%
SA	68%	80%

Table 4.3: Comparison between the old selection and the ML selection.

Chapter 5

Analysis strategy

In this Chapter will be described the analysis strategy used to extract the number of signal events, ultimately deriving a limit on the cross section of the process $T \to tZ$. The data sample used has been collected by CMS in 2018 with an integrate luminosity of 49.2 fb^{-1}, while Monte Carlo samples have been used to estimate the background and the signal contribution. For the signal several simulated samples have been used to probe different mass hypotheses, in particular the Left-Handed chirality VLQ T with mass between 700 GeV and 1800 GeV, with steps of 100 GeV.
The 4 momentum of the T candidate is obtained by summing together the top quark 4 momentum, reconstructed as described in the Chapter 4, and the Z-tagged jets 4 momentum. The T reconstructed mass is then used as discriminant to preform a maximum likelihood fit, in the signal regions, in order to extract signal event yield.

5.1 Data and simulation samples

The analysis is performed using data collected by CMS in 2018, with a $\sqrt{s} = 13\ TeV$ and an integrated luminosity of 49.2 fb^{-1}.
The lists of data sets used is shown in Table 5.1, where $SingleMu$ refers to the trigger used to select the data.

Data set	Integrated luminosity $[fb^{-1}]$
$SingleMu\_Run2017A$	11.73
$SingleMu\_Run2017B$	6.05
$SingleMu\_Run2017C$	6.39
$SingleMu\_Run2017D$	25.08

Table 5.1: List of pp collsion data sets produced at $\sqrt{s} = 13$ TeV and collected by CMS in 2018, used in this analysis.

5.1.1 Background description

The main backgrounds that can mimick the final state we search for are:

- $t\bar{t}$, where a top quark-antiquark pair is produced as shown in Figure Figure5.1. This process is the main background, especially the semi-leptonic 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 Z boson decay.

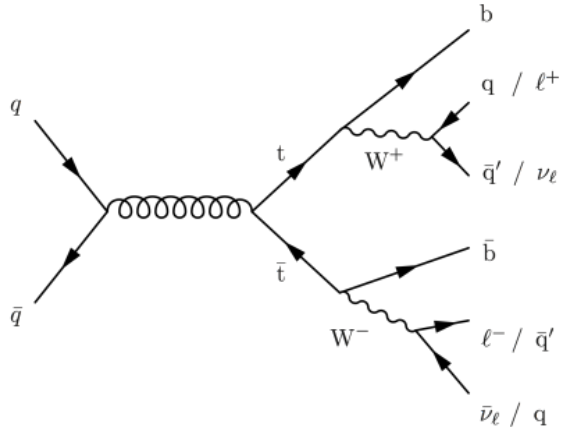


Figure 5.1: Feynman diagrams of the $t\bar{t}$ background.

- Single Top (ST) processes shown in Figure 5.2 can reproduce in some cases the signal topology if, the top quark decays leptonically and one of the additional quarks is reconstructed as a Z-jet. For example, in the $t+W$ case, an hadronically decaying W boson can produce a very similar signature to the one of a Z boson.

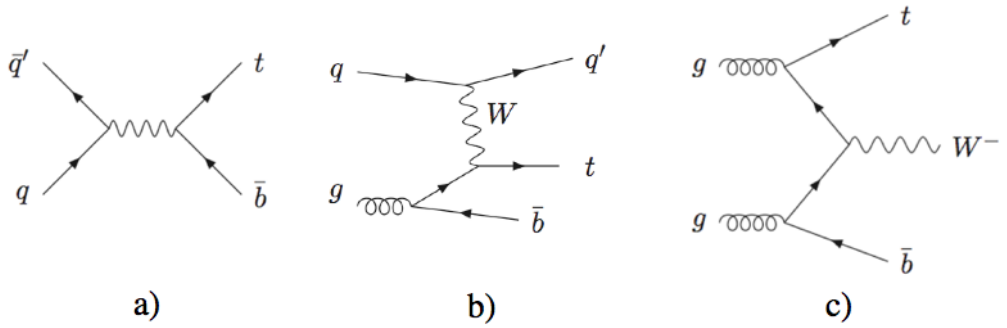


Figure 5.2: Feynman diagrams of the ST background.

- 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 Z boson and the top quark, if a b-tagged jet is close to the muon, could be reconstructed. The process is shown in Figure 5.3.

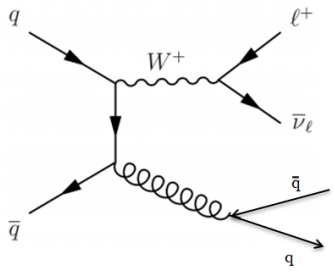


Figure 5.3: Feynman diagrams of the $W + Jets$ background.

- The QCD multijet processes, shown in Figure 5.4, 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.

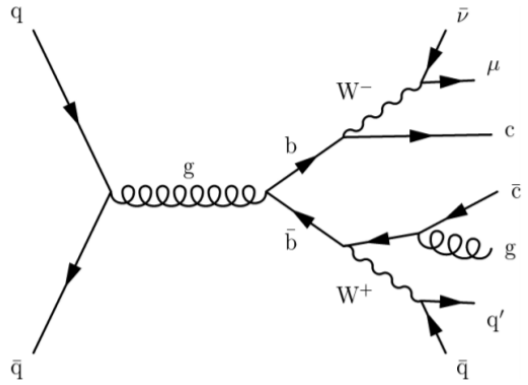


Figure 5.4: Feynman diagrams of the QCD background.

5.1.2 Signal and background simulations

The Monte Carlo simulations of the main background and signal events are performed by the CMS Generator Group. The simulations are generated using different software frameworks:

- *Madgraph* or *POWHEG* [22, 13], to generate matrix elements either at leading order (LO) or at next-to-leading order corrections (NLO);
- *Pythia* [44], to generate and simulate the hadronization of the particle produced in the pp collision;

- *GEANT* 4 [11] to simulate particle interaction with CMS sub-detectors.

The signal event samples are generated using *Madgraph* and *Pythia*, in order to have processes at leading-order (LO). Different event samples have been used in the analysis, in particular they are simulated with different VLQ T mass, from 700 GeV to 1800 GeV in steps of 100 GeV. In Table 5.2 are shown the LO cross-sections of the signal event samples.

The $t\bar{t}(SemiLeptonic)$ and $t\bar{t}(DiLeptonic)$ pair production processes are generated with

Mass [GeV]	Cross section $[fb]$
700	78.04
900	23.34
1000	13.62
1100	8.23
1300	3.25
1400	2.12
1500	1.41
1600	0.94
1700	0.64
1800	0.44

Table 5.2: Simulated $Tb \to tZ$ samples employed in the analysis with their corresponding production cross sections.

POWHEG, evaluating their cross sections at the next-to-next-to leading order (NNLO) in perturbative QCD. Two additional simulated samples are generated to increase the $t\bar{t}$ process statistics in the signal region, by generating events where the mass of the $t\bar{t}$ pair is greater 700 GeV, and the corresponding cross sections are evaluated at NLO. The single top quark events are simulated using a match of *POWHEG* and Pythia as well. Multijet QCD production and $W + Jets$ samples are generated with the *Madgraph* tree-level matrix-element generator matched to *Pythia* for the parton-shower simulation, their cross section are computed at leading order (LO).

In Table the main background samples used and their cross-sections. To avoid the double-counting in the samples $t\bar{t}(SemiLeptonic)$ and $t\bar{t}(DiLeptonic)$, events with $M(t\bar{t}) > 700\ GeV$ have been removed.

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

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.

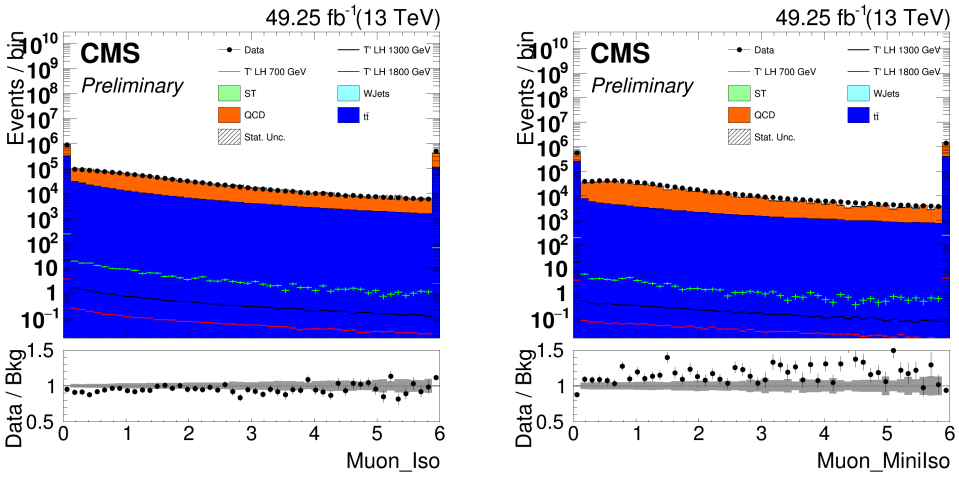
5.2 Preselection

In order to reduce the background sources that are the least likely to mimic the signal final state, and to apply the top quark tagging procedure, a first set of requirements, named preselection, is defined. The presence of a muon in the final state of the signal events leads to consider the events that satisfy specific trigger conditions based on muon reconstruction. The trigger conditions are provided by CMS and for the analysis it was required the logical "or" between the triggers:

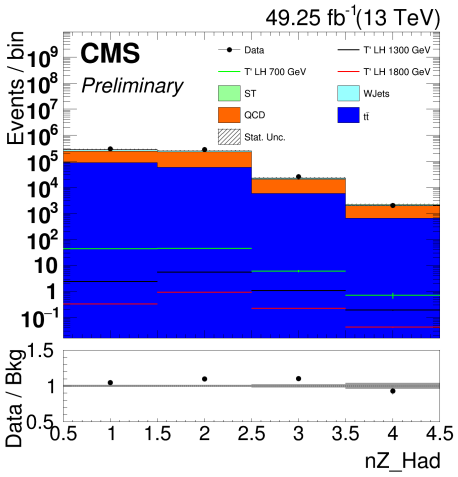
- HLT_Mu50, HLT_TkMu50;
- $HLT\_IsoMu24$, $HLT\_IsoMu27$;

The first two provide the selection of events with at least a muon or a $Tracker$ muon with $p_T > 50\ GeV$, while the second ones select events with at least one muon in the final state with isolation requirements and p_T higher than 24 GeV or 27 GeV respectively.

To further reduce the QCD background, it was required that the leading muon has a p_T higher than 50 GeV and the number of Z-tagged jets (nZ_Had) is higher than 0. In



(a) (b)



(c)

Figure 5.5: Distribution of the isolation of the muon in a cone with $R = 0.4$ (a), in a cone with $R \propto \frac{1}{p_{T\,\mu}}$ (b), and number of Z-tagged jets (c) after the preselection.

Figure 5.5 such variables are shown, and a good agreement between simulation and data can be observed.

5.3 T Mass reconstruction

After the preselection, the VLQ T candidates masses have been reconstructed for the events in the regions:

- 1 *Top Merged*, selecting events with just one top quark candidate and in the Merged configuration;
- 1 *Top Resolved*, selecting events with just one top quark candidate and in the Resolved configuration.

The T 4 momentum is obtained by summing together the top quark 4 momentum, reconstructed as described in the Chapter 4, and the Z-tagged jets 4 momentum. However, in the analysis only VLQ T reconstructed with a mass higher than 650 GeV have been included. The invariant mass is computed as:

$$M_T = \sqrt{(E_t + E_Z)^2 - \|\vec{p_t} + \vec{p_Z}\|^2}, \tag{5.1}$$

where E and $\vec{p}$, are the energy and the 3-momentum of the particles.
In Figure 5.6 is shown the T mass distribution after the preselection in the region 1 *Top Merged*.

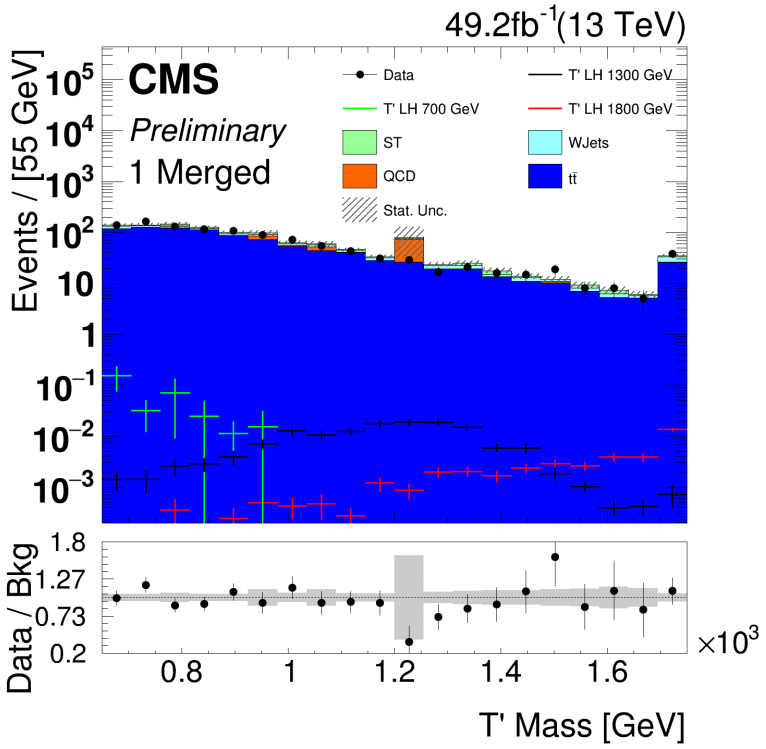


Figure 5.6: VLQ T mass distribution after the preselection in the region 1 *Top Merged*.

In Figure 5.7 is shown the T mass distribution after the preselection in the region 1 *Top Resolved*.

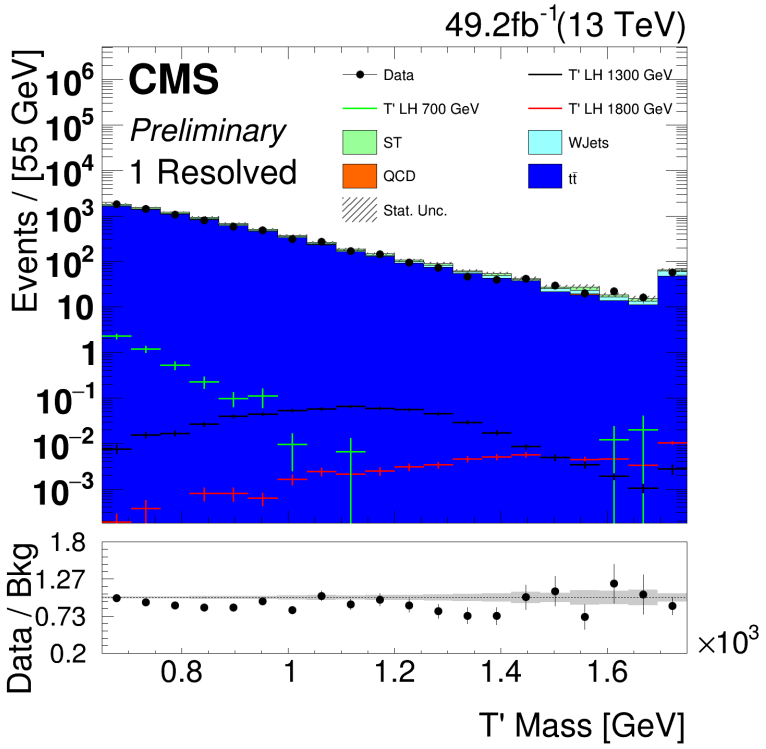


Figure 5.7: VLQ T mass distribution after the preselection in the region 1 *Top Resolved*.

5.4 Signal and Control regions

The two regions taken into account, 1 *Top Resolved* and 1 *Top Merged*, have been further divided in::

- *signal regions*(SRs), enriched in signal events,
- *control regions*(CRs), enriched in background events.

To characterize the regions, two variables have been used:

- the number of forward recoil jets (N_Fwd), that is supposed for the signal greater than one as shown in Figure 5.8. To tag the recoil jet it was required:
  - $p_{T\,Jet} > 40\ GeV$;
  - "Loose" ID;
  - $|\eta| > 2.5$
- The number of b-tagged subjet of the Z-tagged fatjet (Z_sub_b), to avoid the Z-mistag due to the hadronically decay of the W boson. For the signal is expected to be equal to 0 or 2.

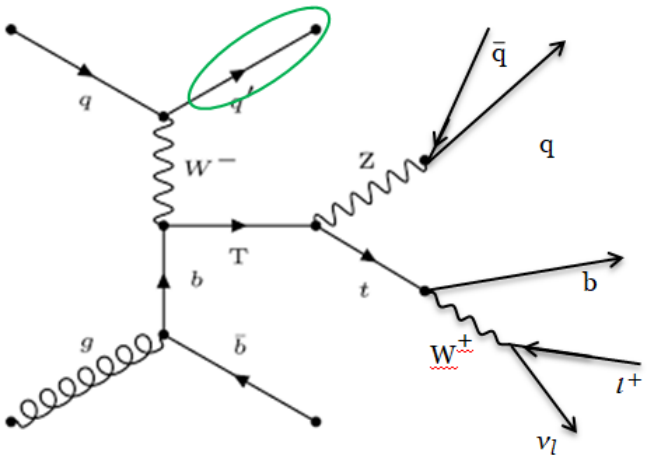


Figure 5.8: Feynman diagram of the signal, focusing on the recoil forward jet.

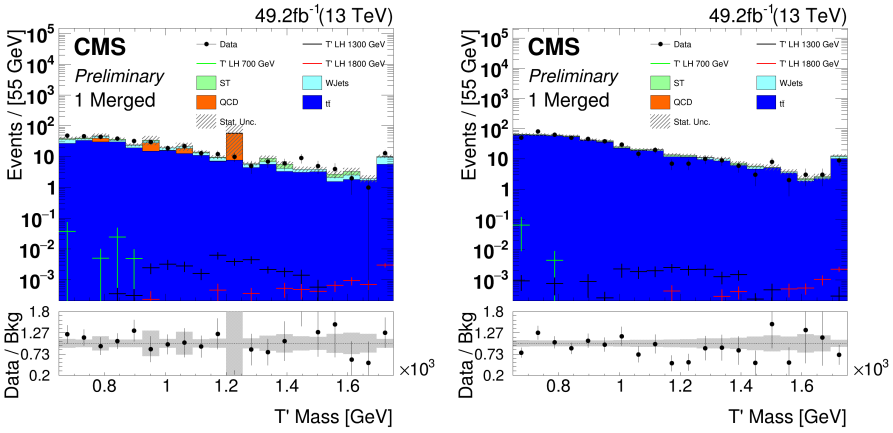
Due to the inefficiencies of the selection requirements for the forward recoil jet and b-jet, the signal regions are not only the $N\_Fwd > 0$ and $Z\_sub\_b = 2$ or 0.
In Table 5.4 signal regions and control regions are listed. The regions have been chosen after studying which ones did impact the least the signal extraction performances. The signal regions have been used to perform the fit of the M_T reconstructed, while the

control regions could be used in future to estimate the background in the signal regions through data driven methods.

	1 Top Merged		1 Top Resolved	
	N_Fwd = 0	N_Fwd > 0	N_Fwd = 0	N_Fwd > 0
Z_sub_b = 0	CR	SR	CR	SR
Z_sub_b = 1	CR	SR	CR	SR
Z_sub_b = 2	SR	SR	SR	SR

Table 5.4: List of signal and control regions.

In Figure 5.9 the T mass distributions in the CRs 1 *Top Merged* are shown, while in Figure 5.10 the ones in the SRs are shown.



(a) (b)

Figure 5.9: M_T distribution in the control regions 1 *Top Merged*, $N\_Fwd = 0$ and $Z\_sub\_b = 0$ (a),$N\_Fwd = 0$ and $Z\_sub\_b = 1$ (b).

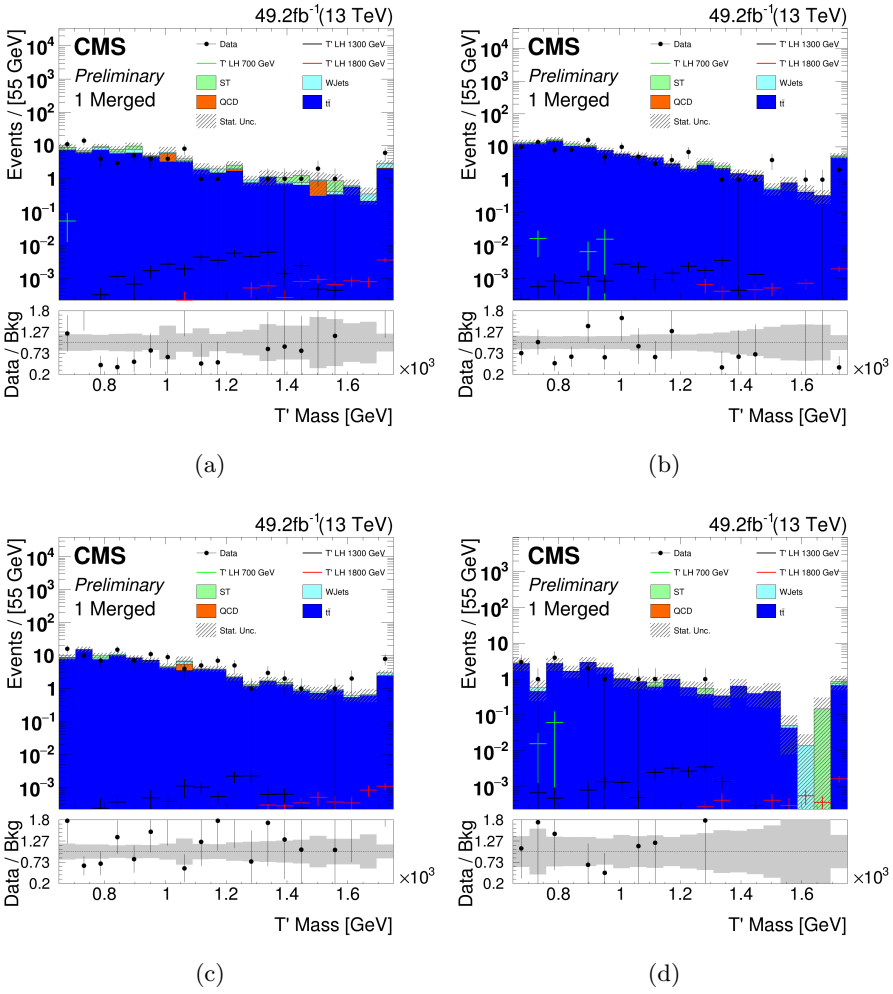
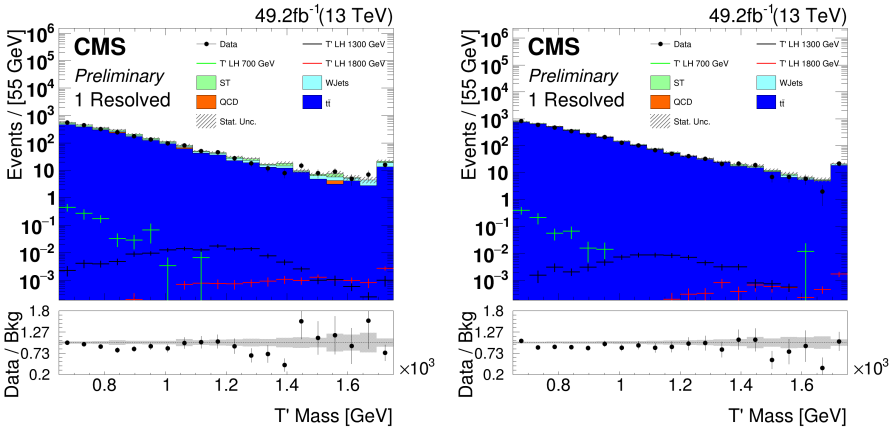


Figure 5.10: M_T distribution in the signal regions 1 *Top Merged*, $N\_Fwd > 0$ and $Z\_sub\_b = 0$ (a),$N\_Fwd > 0$ and $Z\_sub\_b = 1$ (b),$N\_Fwd = 0$ and $Z\_sub\_b = 2$ (c), $N\_Fwd > 0$ and $Z\_sub\_b = 2$ (d).

In Figure 5.11 the T mass distributions in the CRs 1 *Top Resolved* are shown, while in Figure 5.12 the ones in the SRs are shown.



(a) (b)

Figure 5.11: M_T distribution in the control regions 1 *Top Resolved*, $N\_Fwd = 0$ and $Z\_sub\_b = 0$ (a),$N\_Fwd = 0$ and $Z\_sub\_b = 1$ (b).

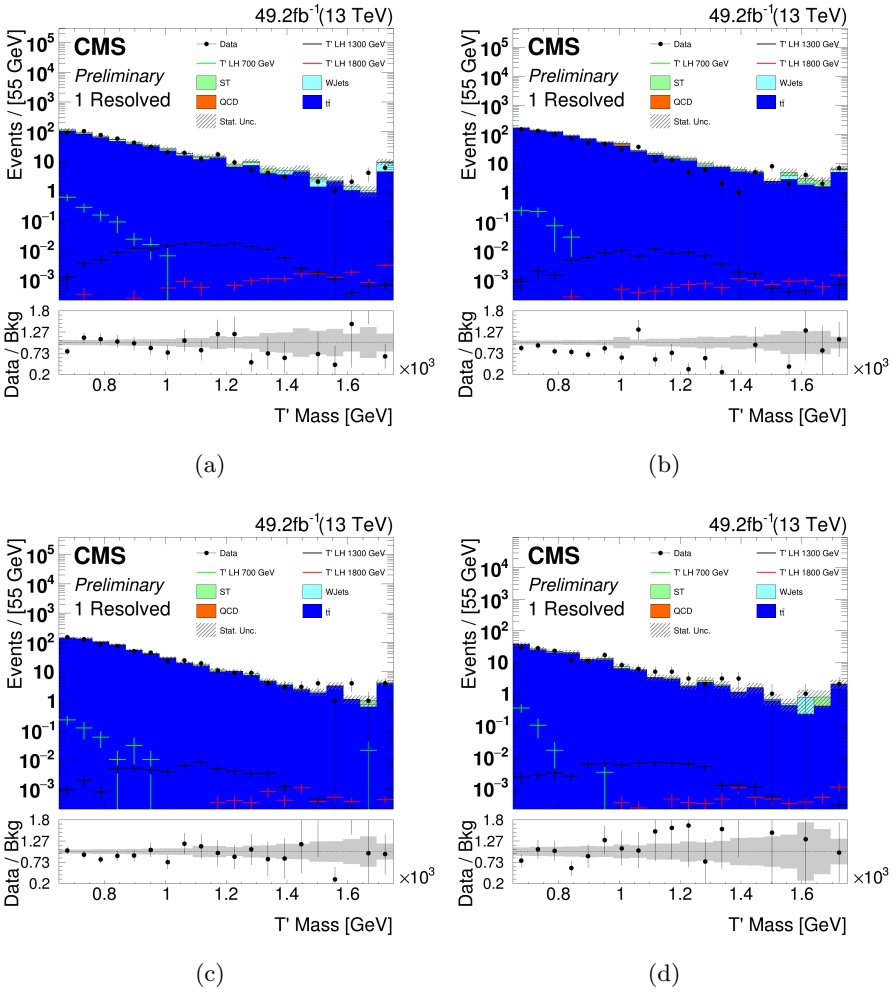


Figure 5.12: M_T distribution in the signal regions 1 *Top Resolved*, $N\_Fwd > 0$ and $Z\_sub\_b = 0$ (a),$N\_Fwd > 0$ and $Z\_sub\_b = 1$ (b),$N\_Fwd = 0$ and $Z\_sub\_b = 2$ (c), $N\_Fwd > 0$ and $Z\_sub\_b = 2$ (d).

5.5 Systematic uncertainties

In the extraction of the signal, a set of major systematic uncertainties has been taken into account. Systematic uncertainties affect the background and the signal predictions, they could be of two different types:

- Yield effect, these systematics change the distributions of the variables in a flat way, modifying just the integral of the distribution.
- Yield and shape effect, these systematics change the shape of the distribution and the its integral.

In this analysis the following systematic uncertainties have been taken into account:

Uncertainty	Yield effect	Yield and shape effect
Luminosity	×	
Simulation statistics		×
MET reweighting		×

Table 5.5: List of systematic uncertainties taken into account in this analysis and their effects on the variables distribution.

Luminosity An uncertainty of $2,5\%$ on the integrated luminosity of 49,25 fb^{-1} at 13 TeV has been taken in account.

Simulation statistics Due to the limited number of events, the simulated samples are affect by a poisson uncertainty for each bin of the M_T distribution. Each bin can fluctuate up and down in two different scenarios, according to an upwards or downwards variations with respect to the nominal value, both of the size of the statistical uncertainty from simulation in that bin.

MET reweighting As said in the previous Chapter, during the data taking in 2018, there was a malfunction in a detector region, to take this into consideration some events have been rejected and the MET has been reweighted, after the preselection of the events.
Moreover, the up and down scenarios have been considered to take into account the fluctuation of the weight used for each bin of the distribution around the nominal value. The up and down scenarios are characterized by using weight raised to 1.2th power, and 0.8th power respectively.

5.6 Fit procedure

The hypothesis of existence of the VLQ T single production has been tested.
Two hypotheses have been considered:

- H_0, absence of new physics, therefore the signal is absent or too little to be detected;
- H_1, the signal produced by new physics have been detected.

An extended binned Maximum Likelihood fit [30] has been performed for M_T in the 8 SRs chosen, considering 12 points of T mass from 700 GeV to 1800 GeV as said before. The likelihood is the product of a Poissonian distribution multiplied by a term that

describes the signal and background models. It is, therefore, defined as:

$$\mathcal{L}(m|\mu_{sig},\theta) = Poisson(n_{obs},\lambda)\prod_{k=1}^{N_{events}} f(m_k|\mu_{sig},\theta), \tag{5.2}$$

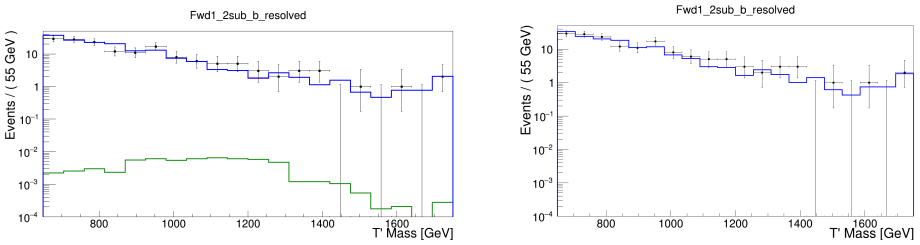
where λ is defined as

$$\lambda = \mu_{sig}\cdot s(\theta) + b(\theta), \tag{5.3}$$

with $s(\theta)$ and $b(\theta)$ are the expected yields of the signal and background, that depend on the set θ of nuisance parameters, which represents the systematics uncertainties. The *signal strength* μ_{sig} is expected to be 1 if H_0 is false, while it's equal to 0 if H_1 is false. The function $f(m_k|\mu_{sig},\theta)$ is a combination of Probability Densitiy Functions (PDFs) for the signal and for the background. In this analysis have been introduced rate parameters for the background to take into account the fluctuations of the main background $t\bar{t}$ and for the QCD multijet that is not well reproduced by the simulations.
The function is, therefore, defined as:

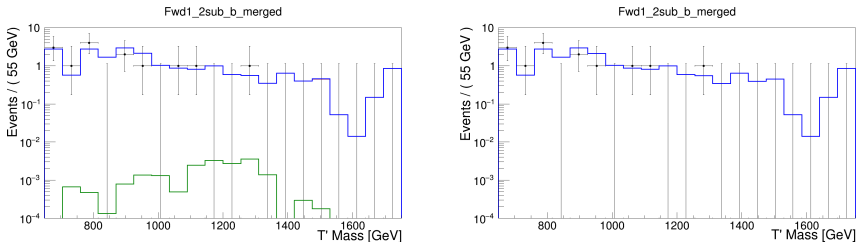
$$\begin{aligned} f(m_k|\mu_{sig},\mu_{t\bar{t}},\mu_{QCD},\theta) = \frac{\mu_{sig}\cdot s(\theta)}{\lambda} f_s(m_k|\theta) + \sum_i \frac{\mu^i_{t\bar{t}}\cdot b^i_{t\bar{t}}(\theta)}{\lambda} f^i_{b_{t\bar{t}}}(m_k|\theta) + \\ + \frac{\mu^i_{QCD}\cdot b^i_{QCD}(\theta)}{\lambda} f^i_{b_{QCD}}(m_k|\theta), \end{aligned} \tag{5.4}$$

where i is equal to $Merged$ and $Resolved$, in order to have that $\mu^{merged}_{background}$ and $\mu^{resolved}_{background}$ are independent.
The fit procedure has been performed for each of the 12 masses supposed for the VLQ T. In Figure and are shown the distribution of the T mass before the fit (a) and after the fit in the H_1 hypothesis, respectively in the SRs 1 *Top Resolved* and 1 *Top Merged* with $N\_Fwd > 0$ and $Z\_sub\_b = 2$, that are the regions in which there is the highest ratio signal over background for the two configurations. It can be noted that no signal evidence is present. This particular fit is obtained for the signal mass hypothesis of 1300 GeV, but similar results are obtained in all the regions after the fit, for all the mass considered.



(a) (b)

Figure 5.13: M_T distribution in the signal region $N\_Fwd > 0$ and $Z\_sub\_b = 2$, 1 *Top Resolved*, before the fit (a) and after the fit (b). In blue the background, while in green the signal, which in this case is the sample with mass of 1300 GeV.



(a) (b)

Figure 5.14: M_T distribution in the signal region $N\_Fwd > 0$ and $Z\_sub\_b = 2$, 1 *Top Merged*, before the fit (a) and after the fit (b). In blue the background, while in green the signal, which in this case is the sample with mass of 1300 GeV.

5.6.1 Upper limits

Since no excess related to the production of a new particle is observed, and the data are compatible with the H_0 hypothesis of only background, the upper limits are estimated on the production cross sections and the branching ratio of the process $T \to tZ$.
Thanks to the *Combine tool* dedicated software [20], it has been possible to evaluate the limits in order to measure the level of incompatibility of data with the H_1 hypothesis. To measure the expected and observed upper limits, the modified frequentist approach [30]. The test statistic used is:

$$\lambda = \frac{\mathcal{L}(m|H_1)}{\mathcal{L}(m|H_0)}, \tag{5.5}$$

that, according to the Neyman-Pearson lemma, is the most powerful discriminator, 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:

$$\lambda(\mu_{sig}) = \frac{\mathcal{L}(m|\mu_{sig}, \hat{\hat{\theta}}(\mu_{sig}))}{\mathcal{L}(m|\hat{\mu}_{sig}, \hat{\theta})} \tag{5.6}$$

where $\hat{\mu}_{sig}$ 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 μ_{sig}.
The test statistic, therefore, is:

$$q_\mu = -2ln(\lambda(\mu_{sig})), \tag{5.7}$$

if q_μ is higher than 0 the events appear to be under the H_0 hypothesis, while if it's lower than 0, they are more compatible with the H_1 hypothesis. The limits for the VLQ T production are estimated using the Confidence Level method. The confidence level for the signal is defined as the ration between the confidence level observed for the H_1 hypothesis, CL_{s+b}, and the confidence level observed, CL_b, for the H_0 hypothesis:

$$CL_s = \frac{CL_{s+b}}{CL_b}. \tag{5.8}$$

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

$$CL_b = P_b\left(q_\mu \geq q_\mu^{obs}\right), \tag{5.9}$$

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

$$CL_{s+b} = P_{s+b}\left(q_\mu \geq q_\mu^{obs}\right). \tag{5.10}$$

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

$$\mu_{sig} = \frac{s_{obs}}{s_{th}} = \frac{\sigma_{obs}}{\sigma_{th}}, \tag{5.11}$$

therefore the upper limits on μ_{sig} can be seen as the upper limits on the ratio $\frac{\sigma_{obs}}{\sigma_{th}}$. In Figure 5.15 is shown the upper limits on μ_{sig}, it can be noted that the observed values are compatible within 2σ with the expected one, for all the different masses considered.
Moreover in Figure 5.16 is shown the upper limits on $\sigma \cdot Br(T \to tZ)$, in order to be independent by the σ_{th} of the model considered.

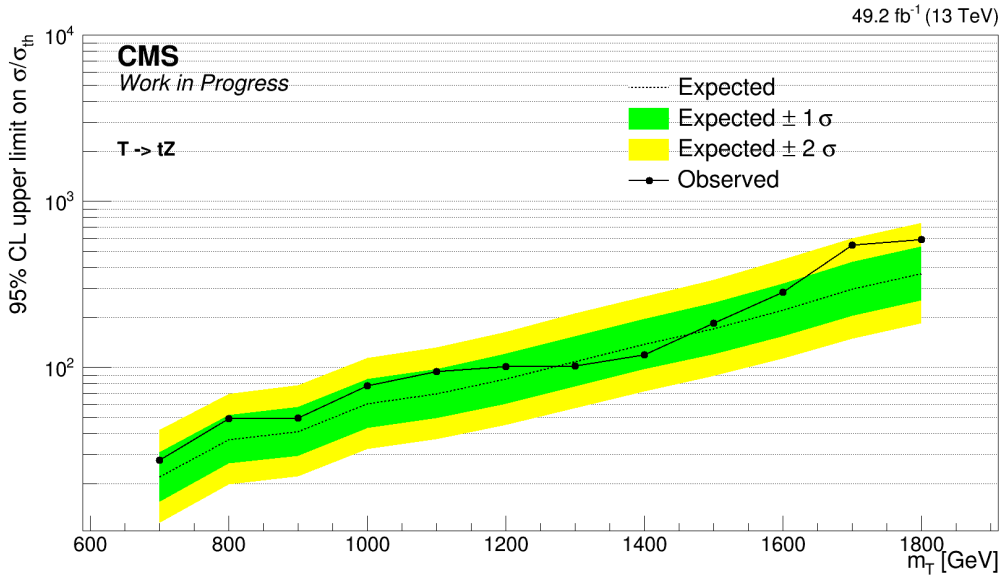


Figure 5.15

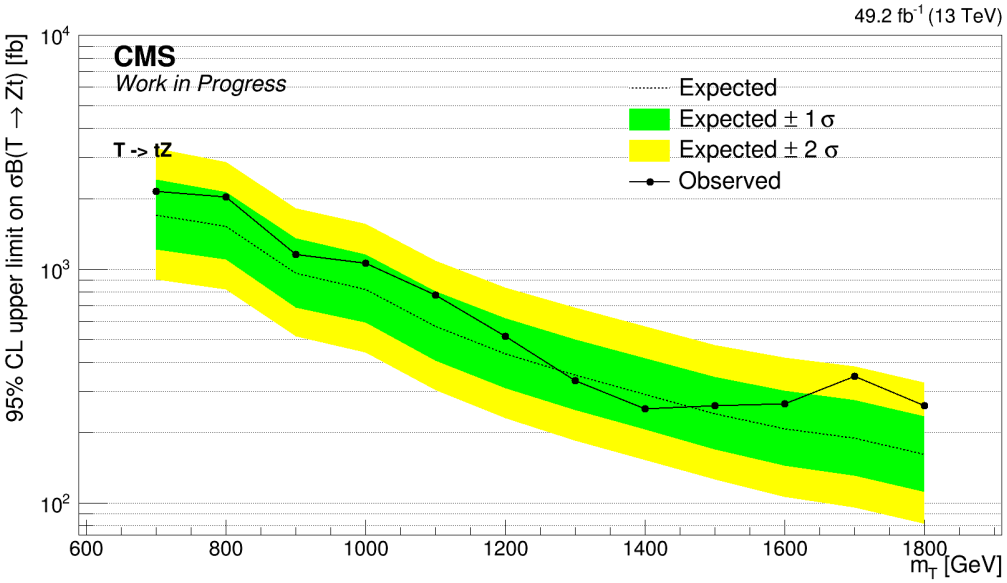


Figure 5.16

Conclusions

In this thesis it is presented a search for the singly produced Vector-Like Quark T decaying to a Z boson and a top quark. It has been performed using the data collected by the CMS experiment at LHC in 2018, with $\sqrt{s} = 13\ TeV$ and an integrated luminosity of 49.25 fb^{-1}.
The final state considered is characterized by the hadronic decay of the Z boson and the leptonic decay of the top quark, in particular $t \to \mu\nu_{\mu}b$. The goal was to define a new identification criteria for the leptonically decaying top quarks. Two different configurations have been taken into account depending on angular separation. The $Top\ Merged$ configuration is characterized by a smaller angular distance than the b-jet radius, otherwise the top quark candidate is classified as $Top\ Resolved$.
Two different methods have been used in order to improve the reconstruction of the top quark, in the first one several requirements have been applied on the most important different discriminating variables against fake or non-prompt muons, while in the second approach the selection of top quarks involves making use of machine learning algorithms. The second selection was found to be more accurate, and events with just one top quark candidate and in the merged configuration or in the resolved configuration have been taken into account. This is the first time reconstruction of top quark candidates with ML techniques is studied in depth for leptonic top quark final states.
The mass M_T was reconstructed with the top quark candidate and the Z-tagged jet 4-momenta. The signal regions have been chosen considering the number of forward jets, which is expected be equal or greater than one for the signal, and the number of narrow-subjet of the Z-tagged jet b-tagged, which is expected to be equal to zero or two for the signal. The regions have been considered for both categories: 1 $Top\ Resolved$ and 1 $Top\ Merged$.
An extended maximum likelihood fit has been performed simultaneously for the distribution of M_T in all the signal regions and for all the different masses hypothesis of the VLQ T in the LH coupling scenario. The considered masses, in natural units, start from 700 GeV up to 1800 GeV, in steps of 100 GeV.
Upper limits on the production cross sections have been estimated at 95% C.L., since no excess related to the production of a new particle has been observed. The observed

limits and the expected limits for the only background hypothesis are compatible within statistical and systematic uncertainties, for all the supposed masses.
Further developments for the analysis include adding other systematic uncertainties, using the data collected in 2016 and in 2017 by CMS. A further extension consists of using the same ML approach to electrons, which would improve the upper limit.

Ringraziamenti

Questo lavoro di tesi chiude un lungo percorso che mi ha insegnato tanto, sia nell'ambito didattico ma soprattutto nella mia crescita personale. Sento di ringraziare in primis tutta la mia famiglia, in particolare i miei Genitori, a cui dedico questo lavoro, per il supporto ricevuto in questi anni e per la comprensione nei momenti più difficili.
Ringrazio il mio relatore il Dr. Iorio per l'immenso aiuto ricevuto in questi mesi e la disponibilità nei miei confronti, a partire dal primo anno di studi con Laboratorio I fino ad oggi. Ringrazio il Prof. Guglielmo De Nardo per il supporto che ho ricevuto. Ringrazio il gruppo di CMS di Napoli che da subito mi ha fatto sentire parte di un bellissimo progetto.
Ringrazio Daniele e Renato, con i quali ho affrontato questo percorso a partire dal primo anno, supportandoci l'un l'altro. Abbiamo superato molti ostacoli ma sapevamo anche quando fosse il momento di staccare tutto per evitare ulteriori danni. Ringrazio i ragazzi che ho conosciuto all'Università, tra cui Lollo, Reb, Giorgia(La Garbata), Sasy, Alessio, Lilla, Claudio, Ludovica e Gege, con cui ho creato un bellissimo legame, che nonostante la lontananza di alcuni di loro continua ad esserci. Ringrazio Maria Grazia ('Orm) sempre pronta ad aiutarmi.
Ringrazio Manuela.
Ringrazio gli amici che ormai conosco da anni, Manuela, Mariangela, Lello, Carlo, Matteo, Ylenia, Attilio, Sissi, Sara, Luigi, Pietro, Domenico e Amalia, perchè nonostante i mille impegni di ognuno di noi, alla fine riusciamo sempre ad esserci.
Ringrazio Fabio il mio amico di sempre. Bastano due minuti insieme e sembra non essere mai cresciuti davvero.
Ringrazio Valeria, con la quale ho vissuto tantissime esperienze in questi ultimi anni, tra risate e sconforti. Ringrazio i ragazzi conosciuti al FermiLab in particolare Roberto, Cecilia, Pica, Paolo, Bastiano, Carlo ed Emanuele, perchè mi hanno fatto sentire a casa fin da subito.
Ringrazio Francesca..che mi ha definito il suo punto di riferimento, probabilmente senza vedere quanto lei abbia aiutato me e mi abbia guidato in molte scelte.
Ringrazio tutti coloro che non ho menzionato esplicitamente, ma che mi sono stati accanto in questi anni.

Infine ringrazio Giovanni, Raffaele, Mery e Martina...e si ovviamente anche Anna, per essere la mia quotidianità, perchè sanno bene quando c'è bisogno del loro aiuto, senza che io dica nulla. Probabilmente ci vorrebbero altre due pagine per parlare di loro singolarmente, ma purtroppo non c'è più tempo...e li ringrazio perchè so benissimo che ancora una volta me la faranno passare liscia.

Bibliography

[1] *The CMS electromagnetic calorimeter project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[2] *The CMS hadron calorimeter project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[3] *The CMS tracker: addendum to the Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 2000.

[4] Search for a singly produced vector-like quark B decaying to a b quark and a Higgs boson in a fully hadronic final state using boosted topologies. Technical Report CMS-PAS-B2G-17-009, CERN, Geneva, 2017.

[5] Search for heavy vector-like quarks decaying to same-sign dileptons. Technical Report CMS-PAS-B2G-16-019, CERN, Geneva, 2017.

[6] Search for top quark partners with charge 5/3 in the single-lepton final state at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-B2G-17-008, CERN, Geneva, 2017.

[7] M. Aaboud et al. Search for pair production of heavy vector-like quarks decaying to high-p$_T$ W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:141, 2017.

[8] M. Aaboud et al. Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *JHEP*, 08:052, 2017.

[9] M. Aaboud et al. Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b-jets with the ATLAS detector. *JHEP*, 07:089, 2018.

[10] K. Agashe, R. Contino, and A. Pomarol. The Minimal composite Higgs model. *Nucl. Phys.*, B719:165–187, 2005.

[11] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. . Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[12] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria. Handbook of vectorlike quarks: Mixing and single production. *Phys. Rev.*, D88(9):094010, 2013.

[13] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[14] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV. *Phys. Lett.*, 122B:103–116, 1983. [,611(1983)].

[15] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett.*, 126B:398–410, 1983. [,7.55(1983)].

[16] G. L. Bayatyan, N. Grigorian, V. G. Khachatrian, A. T. Margarian, and e. a. Sirunyan.

[17] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[18] S. Chatrchyan et al. A New Boson with a Mass of 125 GeV Observed with the CMS Experiment at the Large Hadron Collider. *Science*, 338:1569–1575, 2012.

[19] S. Cittolin, A. Rácz, and P. Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. CERN, Geneva, 2002.

[20] C. Collaboration et al. Documentation of the roostats-based statistics tools for higgs pag (rev. 106). *CMS twiki page*, 2013.

[21] E. Fermi. Tentativo di una teoria dell'emissione dei raggi beta. *Ric. Sci.*, 4:491–495, 1933.

[22] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[23] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[24] T. Kaluza. Zum Unitätsproblem der Physik. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1921:966–972, 1921. [Int. J. Mod. Phys.D27,no.14,1870001(2018)].

[25] D. B. Kaplan, H. Georgi, and S. Dimopoulos. Composite Higgs Scalars. *Phys. Lett.*, 136B:187–190, 1984.

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

[27] O. Klein. Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English). *Z. Phys.*, 37:895–906, 1926. [,76(1926)].

[28] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler. Soft Drop. *JHEP*, 05:146, 2014.

[29] J. G. Layter. *The CMS muon project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[30] L. Lista. *Statistical methods for data analysis in particle physics.* Springer, 2016.

[31] P. Lodone. Vector-like quarks in a 'composite' Higgs model. *JHEP*, 12:029, 2008.

[32] G. Panico and A. Wulzer. The Composite Nambu-Goldstone Higgs. *Lect. Notes Phys.*, 913:pp.1–316, 2016.

[33] L. Randall and R. Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

[34] L. Randall and R. Sundrum. An Alternative to compactification. *Phys. Rev. Lett.*, 83:4690–4693, 1999.

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

[36] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.

[37] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018.

[38] A. M. Sirunyan et al. Search for pair production of vector-like quarks in the $\mathrm{bW\bar{b}W}$ channel from proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B779:82–106, 2018.

[39] A. M. Sirunyan et al. Search for single production of a vector-like T quark decaying to a Z boson and a top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B781:574–600, 2018.

[40] A. M. Sirunyan et al. Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV. *JHEP*, 08:177, 2018.

[41] A. M. Sirunyan et al. Search for electroweak production of a vector-like T quark using fully hadronic final states. 2019.

[42] A. M. Sirunyan et al. Search for pair production of vectorlike quarks in the fully hadronic final state. *Phys. Rev.*, D100(7):072001, 2019.

[43] A. M. Sirunyan et al. Search for vector-like quarks in events with two oppositely charged leptons and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C79(4):364, 2019.

[44] T. Sjöstrand, S. Mrenna, and P. Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[45] M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka, K. Agashe, G. Aielli, C. Amsler, and M. e. a. Antonelli. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[46] M. S. Turner. The Case for ΛCDM. 1997.

[47] T. Wachala. T2K Experiment Neutrino Oscillation Results. *Acta Phys. Polon.*, B50:1757, 2019.

[48] J. Wenninger. Lhc status and performance. *PoS*, page 001, 2018.