Universidad de Antioquia
Instituto de Física
Facultad de Ciencias Exactas y Naturales
Grupo de Fenomenología en interacciones fundamentales

Search for new physics in the final state b, τ, ν in pp collisions in the LHC

M.Sc. Thesis

TOMAS ATEHORTUA GARCÉS

Advisor: José David Ruiz Álvarez

Medellín, 2022

Universidad de Antioquia

Universidad de Antioquia
Instituto de Física
Facultad de Ciencias Exactas y Naturales
Grupo de Fenomenología en interacciones fundamentales

Search for new physics in the final state b, τ, ν in pp collisions in the LHC

M.Sc. Thesis

TOMAS ATEHORTUA GARCÉS

Advisor: José David Ruiz Álvarez

Approved by Date:.

(Signature) *(Signature)* *(Signature)*

....

José David Ruiz Álvarez

Medellín, 2022

Abstract

The standard model of particle physics is one of the most successful theories developed by the human being. It is a quantum theory of fields that explains the matter contained in the universe and its interactions via the interchange of mediators. The standard model predicts the same B-meson decay rates ($\Gamma_i/\Gamma \approx 10.6\%$ [1]) for all generations of charged leptons. However, experimental observations of semi-leptonic decays of B mesons into τ-leptons are enhanced in comparison with the other charged leptons. The results where B mesons decay to D^* mesons are referred to as the R_{D^*} anomalies, and new theoretical frameworks have been considered in order to explain the deviations between the observation and the theory.

Different experiments have already searches for unknown mediators to explain these anomalies, utilizing different final states, and have restricted the parameter space in certain models, depending on the particular final state.

This thesis considers a methodology to discover signs of new physics beyond the standard model, using a final state with a b-quark jet, a τ lepton which decays to hadrons, and missing momentum from proton-proton collisions at the Large Hadron Collider (LHC). This thesis proposes a phenomenological new physics model to probe the $R_{D^{(*)}}$ anomalies using the LHCs Run 2 data collected with $\sqrt{s} = 13\,\text{TeV}$ and $L = 150\,\text{fb}^{-1}$ at the compact muon solenoid experiment.

The results of this research are presented as mass spectrum distributions, comparing the contributions from standard model processes with the new physics model, motivated by the $R_{D^{(*)}}$ anomaly.

Keywords

Leptons, quarks, hadrons, detector, collider, decays, anomaly, events, mass, Monte Carlo.

Resumen

El Modelo Estándar de la física de partículas es una de las teorías más exitosas desarrolladas por el ser humano. Es una teoría cuántica de campos que explica la materia contenida en el universo y sus interacciones a través del intercambio de mediadores. La anomalía $R_{D^{(*)}}$ es una mancha en el modelo estándar, que predice las mismas tasas de interacción para los leptones cargados. A medida que las observaciones de las desintegraciones semi-leptónicas de mesones B en τ-leptones aumentan en comparación con los otros leptones cargados, se deben considerar nuevos marcos teóricos para explicar las desviaciones entre la observación y la teoría.

Diferentes experimentos ya han buscado mediadores desconocidos en diferentes estados finales, y han restringido el espacio de parámetros en ciertos modelos, para estados finales aunque diferentes, relacionados.

Esta tesis considera un jet b, así como un τ hadrónico y energía faltante en estado final de colisiones protón protón, y se buscan indicios de diferentes modelos de física más allá del Modelo Estándar. De esta forma es propuesto un diseño fenomenológico para buscar estas señales en experimentos con colisionadores, especialmente considerando el run 2 del LHC ($\sqrt{s} = 13\,\text{TeV}$ y $L = 150\ \text{fb}^{-1}$) y el experimento CMS.

Los resultados de esta investigación se presentan como distribuciones espectrales de masas, y son la comparación entre diferentes contribuciones del Modelo Estándar y de física más allá de los modelo estándar, motivadas por la anomalía $R_{D^{(*)}}$.

Para Lupe,
porque siempre e incondicionalmente cuando necesité una mano amiga... recibí una pata.
Gracias por todo.

Contents

Chapter 1

Theoretical Framework

In this chapter are discussed the necessary basic concepts and a historical context related to the standard model, it's properties and some measurement problems. By this point is important to explain that the standard model (SM) is a quantum theory of fields that explain the interactions between matter and energy in the universe. The construction of this theory will be presented in a historical context, and will cover since the last years from the XIX century and goes until 2012 with the Higgs discovery. Nevertheless other discoveries related with the SM had been made since then.

1.1 Historical context

Since the early days of the humanity, cultures as the Greeks asked themselves for the origin of the nature. Some of them proposed the existence of a minimal structure that could not be divided by any agent or force, and they called them "atoms" for the greek *átomon* which mean without division. Nevertheless, it was not until the XIX century when the chemist Mendeléyev [2] proposed the periodic table, that the real search for the atoms really started. In the beginning, the chemists thought that the indivisible particles that composed our universe were the ones that determined the chemistry of the elements. However in the late XIX century, J.J. Thomson showed that cathode rays were composed of previously unknown negatively charged particles (now called electrons) [3]. He also calculated that they must have a size much smaller than atoms and a very large charge-to-mass ratio, from that moment on, the particles began to proliferate, with the discovery of the proton [4], then with the neutron [5] and many others. The chemical atoms kept the name, but the non-divisible particles that should be the smallest bricks were not discovered yet.

On the other hand, in the latest days of the 1920 decade, Paul A. Dirac in an attempt to combine the quantum mechanics and the special relativity postulated his equation [6]

$$i\hbar\gamma^{\mu}\partial_{\mu}\psi - mc\psi = 0, \tag{1.1}$$

where $i = \sqrt{-1}$, $\hbar$ is the Planck's constant, γ^{μ} is the $\mu - th$ gamma matrix, ψ is the wave function, m is the particle's mass and c is the speed of light. This equation could explain the existence of

antiparticles as positrons if we interpret correctly the negative energy solutions.

With these discoveries and descriptions, by the early days of the 50's years, the Λ^0, $\mu^{\pm}$, $e^{\pm}$, P, N were well known by the scientific community. By the same time, Richard Feynman, Julian Schwinger and Shin'ichir Tomonaga [7,8] started to work in a description of the electrodynamics as a relativistic quantum theory.

The results that Tomonaga, Feynman and Schwinger achieved were the firsts bricks of a huge building. By combining the Dirac's Lagrangian and the electromagnetic Lagrangian Schwinger, Feynman and Tomonaga reach the Quantum Electro Dynamics (QED) Lagrangian. The first foundations of the SM were now built.

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi + e\bar{\psi}\gamma^{\mu}A_{\mu}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.2}$$

In the last equation ψ are interpreted as the fermions fields, e is the electromagnetic coupling, A_{μ} is the $4^{\text{th}}-$potential field and $F_{\mu\nu}$ is the Faraday's electromagnetic tensor. Here are some details that were clues for the development of the theory:

- The fields ψ, A_{μ} are quantum operators.
- The Lagrangian is invariant under local U(1) transformation.

Years after the release of their research, Chen Ning Yang, and Robert Mills extended the concept of gauge theory from abelian groups to non-abelian groups, allowing to provide an explanation for strong interactions [9]. With the invention of the bubble chambers in the 50's, a large number of particles called hadrons began to proliferate. Those particles were classified according to their electric charge, and to their iso-spin, a property that was proposed by Heisenberg to explain the difference between nucleons taking in account that proton's mass is almost equal to the neutron's mass [10].

In 1953 Murray Gell-Mann introduced the strangeness concept to explain the fact that certain particles as kaons or the Σ and Λ were easily created in particle collisions but decayed much slower than it were expected due to their mass and the production cross section. When Gell-Mann sorted hadrons into groups having similar properties as masses and strangeness in 1961 [11], he proposed that the structure of these groups could be explained with the existence of three flavors of smaller particles inside the hadrons: The quarks. Gell-Mann also briefly discussed a field theory model in which quarks interact with gluons [12].

The evidence for the quarks to be the fundamental constituents of hadrons came in the 1970's with the help of the *Standford Linear Accelerator Center* (SLAC) and the same happened with gluons with the *Positron-Electron Tandem Ring Accelerator* (PETRA), complementing the SM with what we know today as the Quantum Chromo Dynamics or QCD.

In the other hand, Enrico Fermi had already proposed the Fermi's interaction where he achieved to explain the beta decay with a four fermion interaction, involving a contact force with no range [13]. In the 1950's were made different attempts to formulate a theory similar to the quantum electrodynamics, with the pourpose to explain the weak nuclear force, some of

them predicting the existence of the $W^{\pm}$ and Z bosons. However in 1960's Glashow, Salam and Weinberg unified the electromagnetic force and the weak interaction by showing them to be two aspects of a single force, now called the electroweak force.

The discovery of the $W^{\pm}$ and Z bosons had to wait until 1980's when the experiments UA(1) and UA(2) in the *Super Proton Synchrotron* [14].

By the end of the 70's the particle physicist could explain theoretically 3 out of 4 forces in nature as quantum interactions. Beyond that, those theories were approximately-invariant under certain symmetries, The Electro-Weak (EW) theory was invariant under SU(2) × U(1) while the Quantum Chromo Dynamics (QCD) were invariant under SU(3). These symmetries were approximate because the mass terms from the fermions and bosons violate these symmetries.

Many others contributions have been made to the SM, but maybe the most important was the Higgs mechanism. The whole lagrangian of the standard model is invariant under $\mathrm{SU}_C(3) \times \mathrm{SU}_L(2) \times \mathrm{U}_Y(1)$ if the mass terms are neglected. However the mass is a measurable characteristic of many particles, so it should be explained. Then, the explanation came from François Englert and Peter Higgs in the 1964 [15, 16]. They proposed that a scalar field could interact with the different fields in different ways, providing a high mass the fields that interact "strongly" with this scalar field. The process describes a field with a null vacuum expectation value (VEV) in the beginning and then acquires a non-null VEV. This process consist in a spontaneous symmetry breaking and is called the Higgs mechanism. The discovery of the Higgs boson, the particle associated with the Higgs field was made in 2012 in the ATLAS and CMS [17, 18] and was the one of the big last achievements for the particle physicist in the LHC related with the SM.

1.2 Theoretical Context

In this section there must discussed certain properties from the standard model, which will be important for the research's motivations, or well for the justification of the final state. Particularly, there are discussed the Lepton Flavor Universality and the Crossing Symmetry.

1.2.1 Lepton Flavor Universality

The standard model was built to be invariant under $\mathrm{SU}_C(3) \times \mathrm{SU}_L(2) \times \mathrm{U}_Y(1)$, as it was said in the last section. However it respects other symmetries and have other "accidental" properties. It means that those properties were not conceived at the moment the theory was made. Nevertheless the theory predicts those properties and symmetries, and the nature respects these symmetries and preserves these properties. One of these properties is the Lepton Flavor Universality [13] (LFU).

The LFU is a property that predicts that the gauge bosons couples the same way to every charged lepton. In that sense, in a process that involves a W^-, it's as probable to obtain a $\tau^-, \bar{\nu}_\tau$ as it is to obtain $\mu^-, \bar{\nu}_\mu$ or a $e^-, \bar{\nu}_e$ [1] as it can be seen in the Table 1.1.

W^+ Decays	Fraction (Γ_i/Γ)
$e^+\nu$	$(10.71 \pm 0.16)\%$
$\mu^+\nu$	$(10.63 \pm 0.15)\%$
$\tau^+\nu$	$(11.38 \pm 0.21)\%$
Hadrons	$(67.41 \pm 0.27)\%$

Table 1.1: W^+ *decay modes.*

The same situation occurs for the Z decays or productions from a off-shell photon, or Drell-Yang (DY) processes.

1.2.2 The Crossing Symmetry

In this section a little example will help to understand the situation. The process of $e^-\mu^- \to e^-\mu^-$ can be mediated by a neutral boson as a γ (or a Z), and will have a specific cross section. One can also think about a similar process: pair production and annihilation $\mu^+\mu^- \to e^+e^-$. Beyond that, one can suspect that there is a deep relation between these two processes since a process can be diagrammed as a rotation from the other-one without making any other change in the diagram.

If we make a more quantitative approach we shall note that the Scattering Matrix $\mathcal{M}(e^-\mu^- \to e^-\mu^-) = \mathcal{M}(\mu^+\mu^- \to e^+e^-)$. In fact, always the process respect kinematics, we can say

$$\mathcal{M}(X(k) + ... \to ...) = \mathcal{M}(... \to \bar{X}(-k) + ...) \tag{1.3}$$

This conservation of the scattering matrix is called the *crossing symmetry* and allow us to calculating different cross sections without calculating them directly [19].

1.3 $R_{D^{(*)}}$ Anomaly

As it's been said, the SM has the property of the LFU, which means that in a process involving a charged lepton in a final state, it's as probable to have an electron, as it is to have a muon or a tau. However, in the recent years, many experimental measurements related with B mesons (hadrons made from a quark-anti quark, in which one of those is a b or a $\bar{b}$) had been made, showing interesting deviations from their SM expectations. One of the most interesting deviations has a significance of 3.1σ and involves semi-leptonic decays of these mesons, and suggests a violation of LFU. This deviation is known as the $R_{D^{(*)}}$ Anomaly and it's defined as

$$R_{D^{(*)}} = \frac{\Gamma(B \to D^{(*)}\tau\nu)}{\Gamma(B \to D^{(*)}l\nu)} \tag{1.4}$$

These quotient measures the ratio between the Branching fraction of B mesons decaying into D mesons (or excited D mesons) and a τ and ν respect with the branching ratio of the B mesons decaying into D mesons (or excited D mesons) and a l and ν. This measure is enhanced with respect to the SM prediction by roughly 30% [20, 21].

1.4 Physics Beyond the Standard Model

The SM is not a theory of everything, and that's something that the physicist already know. Observations as the neutrino's masses [22], the baryon-asymmetry [23], Dark-Matter [24], among others confirms that there is more fundamental physics beyond what we already know. In order to find a more complete theory, many theoretical models had been proposed during the recent years, each one differing in couplings, symmetries and interactions. However they are classified by the type of interactions they have, i.e. the Lagrangian. In this section we will discuss 3 different models that will help us to study the final state $b + \tau_H + p_T^{\text{miss}}$.

1.4.1 Sequential Standard Model

In the recent years, specially since 1980, many theoretical models with extra gauge-bosons had been proposed. Barger [25] proposed lighter gauge bosons than the ones in the SM. Most of these models consider extra symmetries such as SU(2)′ which predicts the existence of both W' and Z' different to the SM W and Z. In the simplest case, these bosons have the same chiral couplings, and couples the same way to the SM fermions and the interaction with other gauge or scalar-bosons are omitted. These particular model is called the Sequential Standard Model. These phenomenology is described by the lagrangian

$$\mathcal{L}_{\text{CC}}^{W'} = \frac{g_2'}{\sqrt{2}} \left[\bar{u}_i \gamma^\mu \left(\left[\kappa_{q,L}^{W'}\right]_{ij} P_L + \left[\kappa_{q,R}^{W'}\right]_{ij} P_R \right) d_j + \bar{\nu}_i \gamma^\mu \left(\left[\kappa_{\ell,L}^{W'}\right]_{ij} P_L + \left[\kappa_{\ell,R}^{W'}\right]_{ij} P_R \right) \ell_j \right] W'_\mu + \text{h.c.} \tag{1.5}$$

$$\mathcal{L}_{\text{NC}}^{Z'} = \frac{g_2'}{\cos\theta_W} \left[\bar{q}_i \gamma^\mu \left(\left[\kappa_{q,L}^{Z'}\right]_{ij} P_L + \left[\kappa_{q,R}^{Z'}\right]_{ij} P_R \right) q_j + \bar{\ell}_i \gamma^\mu \left(\left[\kappa_{\ell,L}^{Z'}\right]_{ij} P_L + \left[\kappa_{\ell,R}^{Z'}\right]_{ij} P_R \right) \ell_j \right] Z'_\mu + \text{h.c.} \tag{1.6}$$

In the equations 1.5 and 1.6 CC and NC mean *Charged Current* and *Neutral Current* respectively, the g_2' is the SU(2)′ coupling constant. The parameters κ are complex values of 3×3 matrices in the flavour space. Note that the index in the quarks and leptons flavour is summed, and if we replace the conditions :

$$\left[\kappa_{q,L}^{W'}\right]_{ij} = V_{ij}^{\text{CKM}}, \quad \left[\kappa_{\ell,L}^{W'}\right]_{ij} = V_{ij}^{\text{PMNS}}, \quad \kappa_{q(\ell),R}^{W'} = 0 \tag{1.7}$$

$$\left[\kappa_{f,L}^{Z'}\right]_{ij} = \left(T_L^{3,f} - Q_f \sin^2\theta_W\right)\delta_{ij}, \quad \left[\kappa_{f,R}^{Z'}\right]_{ij} = \left(-Q_f \sin^2\theta_W\right)\delta_{ij} \tag{1.8}$$

the Standard Model is recovered [26]. In this model as it's been described the W' and the Z' masses are free parameters. From now and on, every time we name SSM we'll be talking about the sequential standard model, specifically to the W′.

1.4.2 Effective Field Theory

An EFT is a way of approximation that allows the model to include degrees of freedom, in such a way to describe phenomena occurring at a certain energy scale, without taking into account the substructure and the higher energies degrees of freedom [27]. That means that if there is any kind of massive mediator, heavy enough to be impossible to produce it on the LHC, or at least

to have a low probability to be produced, the EFT is a way to explore the phenomenology of the interaction, without the need of compromising with the whole substructure of the nature. That is because it only takes into account relevant degrees of freedom i.e. those states with $m \ll \Lambda$ while the heavier excitations with $M \gg \Lambda$ are integrated out from the action.

So addressing the EFT means not compromising with an special underlying physics for the phenomenology. Then analyzing the process supposing an EFT is helpful now that the underlying fundamental theory is unknown. A Lagrangian with these characteristics and thought for the b, τ, ν_τ is the Eq 1.9:

$$\begin{aligned}\mathcal{L}_{\text{eff}} \supset -\frac{2V_{ib}}{v^2}\Big[\Big(1+\epsilon_L^{ib}\Big)\left(\bar{\tau}\gamma_\mu P_L\nu_\tau\right)\left(\bar{u}_i\gamma^\mu P_L b\right) + \epsilon_R^{ib}\left(\bar{\tau}\gamma_\mu P_L\nu_\tau\right)\left(\bar{u}_i\gamma^\mu P_R b\right) \\ + \epsilon_T^{ib}\left(\bar{\tau}\sigma_{\mu\nu}P_L\nu_\tau\right)\left(\bar{u}_i\sigma^{\mu\nu}P_L b\right) + \epsilon_{S_L}^{ib}\left(\bar{\tau}P_L\nu_\tau\right)\left(\bar{u}_i P_L b\right)\Big] + \text{ H.c.}\end{aligned} \tag{1.9}$$

In the equation 1.9 V_{ij} are the Cabibbo-Kobayashi-Maskawa (CKM) matrix elements, u_i are the i^{th} "up-type" quark, b is the *beauty* or *bottom* quark, ν_τ is the τ neutrino, γ^μ is the μ^{th} Dirac's matrix, $P_{R,L}$ are the chiral projectors, $\sigma^{\mu\nu} = i/2[\gamma^\mu, \gamma^\nu]$, and $v \approx 246\,\text{GeV}$ as the electroweak symmetry breaking scale, and the ϵ_Γ are the Wilson Coefficients (WC) [20].

1.4.3 Lepto-Quark

In the SM, the quarks and leptons do not interact directly, but through a scalar o a vector boson. Nevertheless, different theoretical models had proposed a brand new and undiscovered interaction where a new mediator couples both quarks an leptons. The different models had been considered, between different symmetries for the Lagrangian, different type of mediators and different modes that the leptons couples with the quarks [28–31].

In these case we have considered a model were the *up*-type quarks couples with the neutral leptons and the *down*-type quarks couples to the charged leptons with a scalar mediator under a $U(1)$ symmetry [20]. Theoretical details of this model can be reviewed in [28].

Chapter 2

LHC and the CMS experiment

The CMS experiment is the second biggest experiment placed in the LHC, which nowadays is the main collider managed by the CERN. This last institution is the biggest organization for the research in high energy physics in the world. It was founded in 1954 by 12 european countries. Right now it counts with 21 countries and other 28 non-member countries with around 220 institutes and universities all around the world [32].

The CERN began gathering achievements in 1984 with Carlo Rubbia and Simon van der Meer when they won the nobel prize for the discovery of the W and the Z bosons.

In this chapter different aspects of the LHC and the CMS experiment will be presented. A particular emphasis in the experiment is made in order to approach the objects related to the main subject of this work.

2.1 The LHC

The Large Hadron Collider, or LHC, is a machine that accelerate and collide protons. Nowadays is the most powerful particle collider in the world achieving an energy of 13 TeV in the center of mass of the collision. It's a ring with 27 km of circumference. It's placed between France and Switzerland near the city of Geneva. The tunnel of the main ring it's around 100 m under the ground and 170 m near the lake Léman. The LHC uses as much as possible the Linear Electron-Positron Collider (LEP), that was built between 1984 and 1989. It began to operate in 2008, but the initial tests officially began in 2009. During the run 1, the LHC reached 4 TeV per beam in the center of mass, with an integrated luminosity of 20 fb^{-1}. Since then the LHC has reached 140 fb^{-1} and 13 TeV [33].
The LHC can be thought as 2 complementary parts: the main ring and the injection chain.

2.1.1 Injection Chain

The injection chain begins with a proton source. An hydrogen gas is ionized in the duoplasmoatron proton ion source. these ions are the first proton bunch structure, which is accelerated up to 50 MeV in the linear accelerator Linac2. It continues to the Proton Synchrotron Booster (PSB), where they are accelerated until they reach 1.4 GeV, then they arrive to Proton Synchrotron (PS) where they are accelerated to 28 GeV. Here the initial protons bunches are splited in 72 bunches, spaced by 25 ns. Finally they reach the Super Proton Synchrotron where the bunches are finally accelerated up to 450 GeV right before they get into the main ring [34].

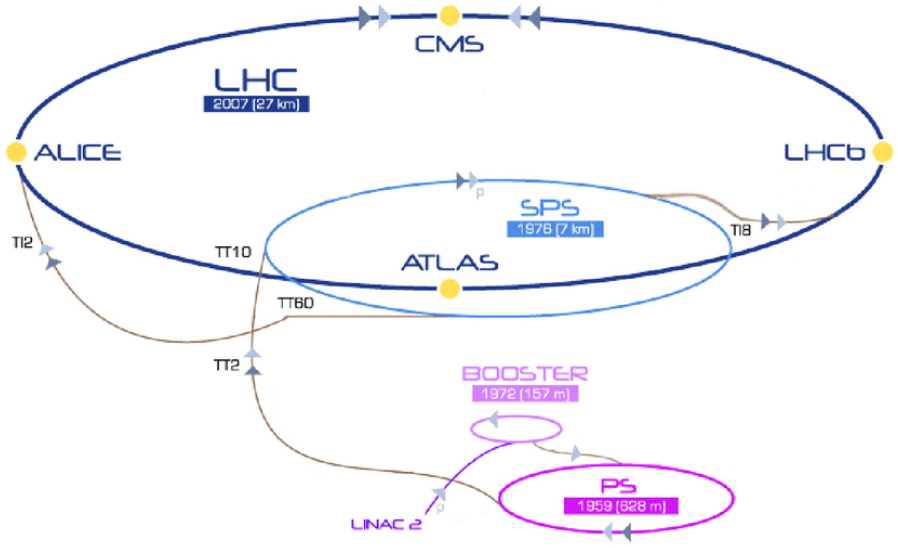


Figure 2.1: *LHC injector chain and main ring.*

2.1.2 The main ring

Once the bunches reach the main ring they are inserted in two rings that accelerate the protons in the opposite directions, one clockwise and other counter-wise. These rings intersect each other in different points to collide the protons, in these intersection can be found the 4 main experiments of the LHC:

1. ***LHCb:*** The LHCb is dedicated mainly to the study of the physics of the b-quark. It's hosted in the point 8 of the main ring. It's conically shaped. It's very precise in identifying different hadrons [35].

2. ***ALICE:*** For its letters A Large Ion Collider Experiment. Hosted in the point 2 of the main ring is dedicated to the studies of the quark-gluon plasma, designed for heavy ion physiscs, it's very precise detecting a high number of tracks per event [36].

3. ***ATLAS:*** A Toroidal LHC ApparatuS, just as the CMS experiment, is dedicated to the search of new physics, it's located in the point 1 of LHC and it's the biggest experiment in the ring [37].

An illustrative images of the injection chain and the main ring can be found in Figure 2.1 and Figure 2.2.
Finally there is the ***CMS*** experiment which we will discus in the next section.

2.2 The Compact Muon Solenoid (CMS) Experiment

The CMS experiment is hosted at point 5 in the main ring. It is the second biggest experiment in the LHC, with a cylindrical shape, has 15 m of diameter and 28.7 m of length, and is the heaviest experiment in the LHC, weighting around the 14 kTons, In the Figure 2.3 an illustrative image can be found [38]. It is built concentrically organized around the beam line. Its name is due to its main characteristics:

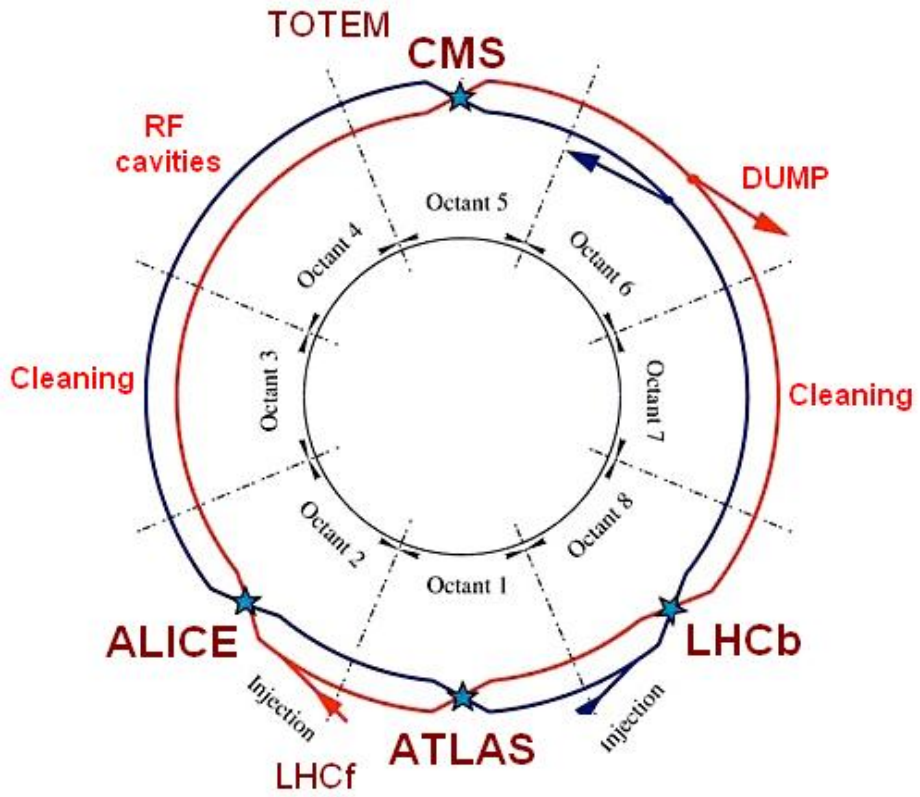


Figure 2.2: *LHC Main ring separation.*

- Compact because it's calorimeters are located inside the magnet.
- Solenoid is because the super conductor magnet is a solenoid.
- Muon is due to it's high precision in the muon detection.

The CMS collaboration is formed by around 4000 scientist in over 200 institutes and universities located in more than 40 countries. It's design was thought for the precise identification of different particles and measurements of their properties. The CMS experiment has a powerful solenoid magnet capable of producing a 3.8 T magnetic field, that allows to bend the very energetic particles. Besides that, the calorimeters allow to measure with a very high precision the energy from hadrons, electrons and photons. In it's most external layer are placed the muon chambers, responsible for the detection, recognition of the mouns and the measurement of their properties. In the innermost part, the tracking system, reconstructs the collision points and the charged particles track.

2.2.1 CMS experiment parts

The CMS experiment is a multi-purpose engineering masterpiece, designed mainly bot not only for the precise study of the physics beyond the standard model. It is composed of several parts, the main ones will be listed and enumerated below:

- **The superconducting magnet:** In principle, it's designed to reach 3.8 T field, has 6 m of diameter and a length of 12.5 m and it's capable to store an energy of 2.6 GJ. It's crucial to identify the charged particles, because it bends the path that they will follow [39].

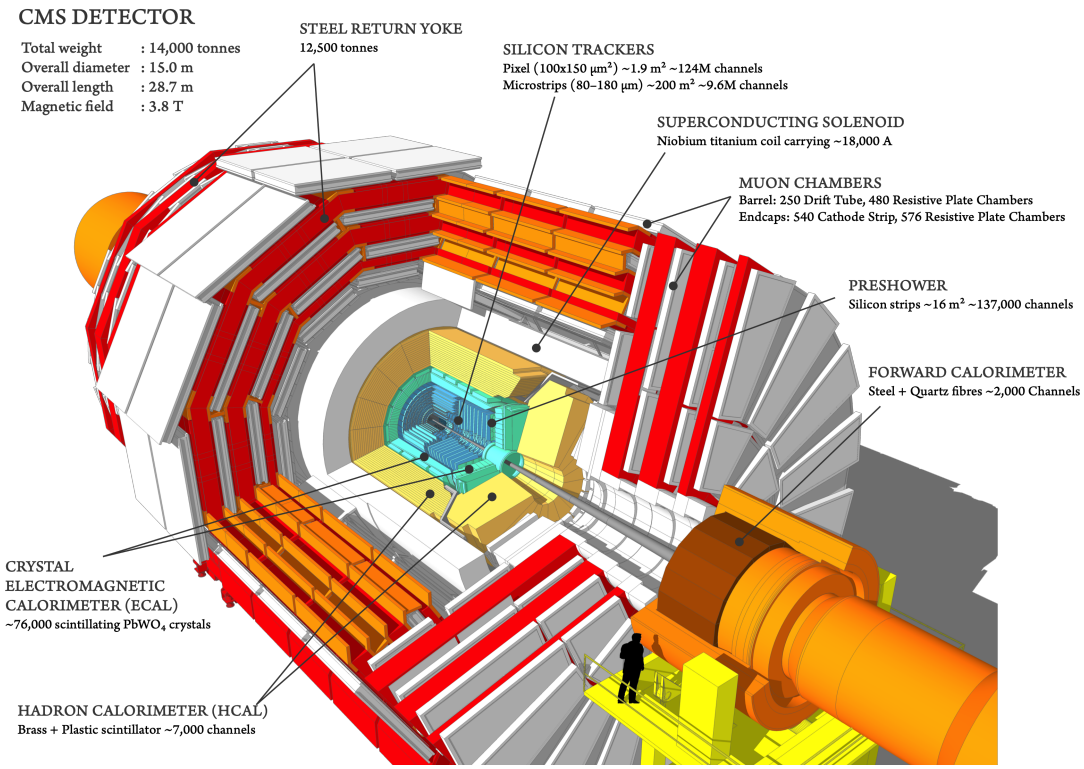


Figure 2.3: *The CMS experiment.*

- **The inner tracker:** It's designed to provide a high precision and an efficient measurement for the charged particles trajectories. It's placed arround the collision point with diameter of 2.5 m and a length of 5.8 m. It's composed of a pixel detector with three barrel layers and a silicon strip tracker that has ten barrel detection layers [40].

- **The electromagnetic calorimeter:** Also known as ECAL is built with 61200 lead tungstate crystals mounted in the central barrel part, made by 7324 crystals in each of the two endcaps. These crystals make the calorimeter radiation resistant. It's main purpose is the measurement of the energy of photons and electrons [41].

- **The hadronic calorimeter:** Just like the electromagnetic caloritmeter, the purpose of this calorimeter is to measure the energy of the hadrons. It's composed by the barrel, the endcaps, the outer and the forward in order to cover different sections of the calorimeter [42].

- **The muon chambers:** It's located at the most external layer of the CMS experiment due to the high penetration power of these type of particles. Muons and neutrinos cannot be stopped by the calorimeters, and easily can scape from the detector. The muon chambers are placed around the hadron calorimeter outsider and in the endcaps, and are there to provide a high efficiency in muon identification and precision measurement of the muon's momentum [43].

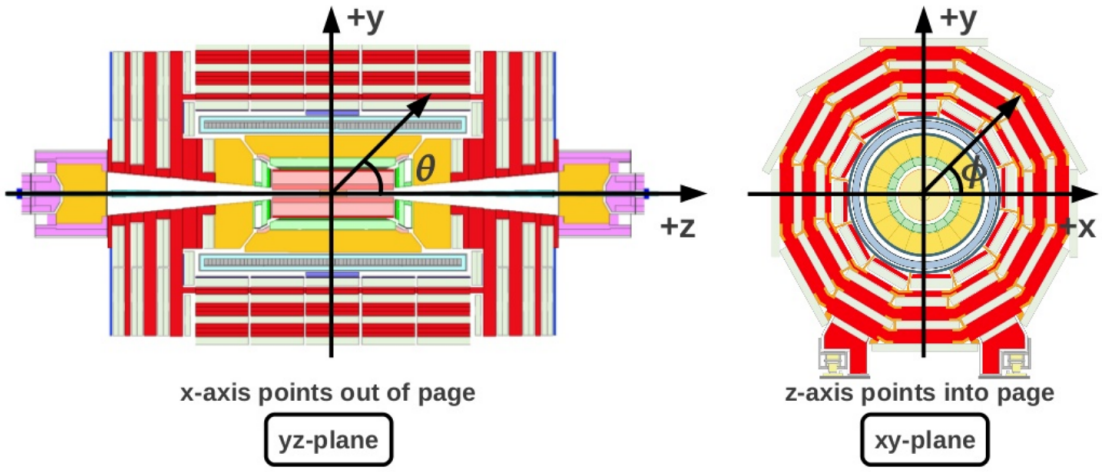


Figure 2.4: *The CMS experiment coordinate system.*

2.2.2 The coordinate system

To understand how the different objects are identified and reconstructed, the coordinate system must be settle. The origin of this coordinate system is the center of the detector. This point is denominated the collision point or interaction point. Then the z axis is defined from this point along the beam pipe line towards the Jura's mountain. The x axis goes to the center of the LHC ring and the y axis in the zenit direction. An illustrative image can be seen in the Figure 2.4. However due to the LHC plane's inclination, this coordinate system is slightly tilted compared with the true vertical.

In experimental particle physics, specially in collider experiments, is very useful to define two angles: the ϕ angle in the $x-y$ plane from the x axis to the y axis, this plane is often called the transverse plane, and is quiet important in the description of high energy experimental physics, variables as the transverse momenta p_T and missing energy p_T^{miss} are completely embedded in this plane. The other angle is the θ angle in the $z-y$ plane from the z axis to the y axis. Now at these amounts of energy is necessary to work with relativistic invariant quantities, then instead of the θ angle there must be defined the pseudorapidity η:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{2.1}$$

The radial quantities in the $x-y$ plane are known as the transverse component. In that sense, one can define the p_T or the transverse momentum. The other component of the moment will be called as the longitudinal component p_L.

The previous information about the CMS experiment parts was taken from [44, 45].

2.3 CMS objects

Different particles are detected in diverse ways in a detector experiment. As we have seen briefly in the past sections, some parts of the detector are specialized to detect specific particles, experiment measures energies, hits and charges. In order to transform these variables to detection

and recognition of particles, the information from each subsystem must be combined through the Particle Flow algorithm.

2.3.1 Particle Flow (PF) algorithm

CMS experiment uses an specialized algorithm dedicated to reconstruct the particles with the information of each subsystem: charged and neutral hadrons, electrons, muons and photons. The analysis performance is affected by quality of the reconstructed objects, then it has to be optimized in terms of minimal fake rate, exact energy and momentum reconstruction. As an iterative process that looks for the best combination of the information coming from an event to reconstruct different objects [46], those objects will be described next.

Tracks and vertex

It can be said that the tracks are the path that leave the particles when they pass through the detector. Then the vertices are defined as the point where several tracks are originated. The reconstruction of these objects is crucial to the event reconstruction and are decisive to the reconstruction and identification of other objects.

The tracks reconstruction starts with the clustering of pixel [47] and silicon strips [48] signals into hits. Then with these hits the trajectory of the charged particles are reconstructed. At least two or three hits and a vertex constraint are needed to reconstruct trajectories, so the reconstruction proceeds from the propagation of the hits in different layers. Such propagation is done in an iterative way that considers the whole possible combination of tracks. This could result in multiple possible tracks. Then, in order to discard fake tracks and choose the most probable tracks, only the ones compatibles with the vertices and number of hits are used. The chosen tracks are the ones that achieves the highest number of hits. Finally a χ^2 fitting is used to discriminate between the tracks [49].

Calorimeter clustering

The calorimeter clusters are defined as a group of deposits in a specific part of the calorimeter. Their identification is a crucial step in the reconstruction of particles in CMS experiment. That reconstruction starts in the cell energy maxima, then after determining a threshold to avoid adding noise, the adjacent deposits are associated to the cell if they reach an energy above the threshold. Those thresholds are 80 MeV for the ECAL barrel, 300 MeV for the ECAL endcaps and 800 MeV for the HCAL. The PF algorithm uses the clusters to reconstruct objects as jets. This procedure needs both HCAL and ECAL. Calorimeter clusters are used to identify, for example, neutral hadrons and reconstruct the energy of charged hadrons outside the tracker acceptance [41, 42].

2.3.2 Jets

As we are interested in τ_h and b, it's important to describe how hadrons are recognized by the experiment. So in the next subsection jets will be discussed.

Due to the color confinement [50] the quarks and gluons cannot live alone by themselves in nature, instead they need to make composed particles with no color charge. The process in which

Decay	Probability
$\tau \to e\,\nu_e\nu_\tau$	17.8%
$\tau \to \mu\,\nu_\mu\nu_\tau$	17.4%
$\tau \to \pi^\pm\nu_\tau$	11.4 %
$\tau \to \pi^0\pi^\pm\nu_\tau$	25.4%
$\tau \to \pi^0\pi^\pm\pi^0\nu_\tau$	9.19%
$\tau \to \pi^0\pi^0\pi^0\pi^\pm\nu_\tau$	1.08%
$\tau \to \pi^\pm\pi^\pm\pi^\pm\nu_\tau$	8.98%
$\tau \to \pi^0\pi^\pm\pi^\pm\pi^\pm\nu_\tau$	4.30%
$\tau \to \pi^0\pi^0\pi^\pm\pi^\pm\pi^\pm\nu_\tau$	0.5%
$\tau \to \pi^0\pi^0\pi^0\pi^\pm\pi^\pm\pi^\pm\nu_\tau$	0.11%
$\tau \to K^\pm X\nu_\tau$	3.74 %
$\tau \to (\pi^0)\pi^\pm\pi^\pm\pi^\pm\pi^\pm\pi^\pm\nu_\tau$	0.1%
others	0.03%

Table 2.1: τ *decays*

a quark or a gluon form a particle with other quark and gluon is called hadronization. A hadron is a colorless particle composed by quarks. As QCD processes are very radiative, while these hadrons travel through the detector they radiate a lot of other hadrons in a process called the hadronization cascade. All the particles produced from an initial quark or gluon define a jet [51].

The jet energy corrections relate the energy of the reconstructed jets to the true particle level energy, which is independent of the detector response. The precise understanding of the jet energy scale is crucial for many physical analysis.

At next the discussion will be focused in two special types of jets: the $b-$jets and the $\tau-$jets.

2.3.3 τ-Jets

A very good identification of the leptons is a basic ingredient of many analyses at the LHC. In particular τ leptons, which are the most difficult leptons to identify, are expected to be produced by the decay of several interesting physic channels.

The τ lepton decays hadronically with a probability of $\sim 65\%$ [1,52], in that sense, high energy experiments as ATLAS or CMS the τ is considered a jet. When the momentum of the τ is large compared to the mass, a very collimated jet is produced. For example for a transverse momentum $p_T > 50$ GeV/c, 90% of the energy is contained in a cone of radius $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2$ [53]. In Table 2.1 the main τ decays branching ratios are shown.

Reconstruction and identification of τ_h

The reconstruction of the τ_h is based on the Hadron-Plus-Strips (HPS) algorithm which will be discussed at next, however, the τ_h candidates must be discriminated from common jets, so specific isolation requirements will be specified after the HPS subsection.

The hadrons-plus-strips algorithm

The main idea of the HPS algorithm is to reconstruct the different decays of the τ lepton into hadrons. The final state could include charged and neutral hadrons as it can be seen in

the Table 2.1. The neutral pion π^0 mesons promptly decay into two photons, which have a high probability to convert into e^+, e^- pairs. These e^+, e^- travel along the tracker material and then the magnet split these pair in the (ϕ, η) plane. To reconstruct the full energy of the neutral pions, the electrons and photon candidates in a certain region of $\Delta\eta \times \Delta\phi$ are clustered together. The resulting is called a "strip". The strip momentum is defined by the vectorial sum of all its constituent momenta.

In the case of the charged particles, they require to have $p_T > 0.5\,\text{GeV}$ and must be compatible with originating from the primary vertex of the event, where the criterion on the transverse impact parameter is not highly restrictive, to minimize the rejection of genuine τ leptons with long life times. The requirement in the p_T guarantees that the corresponding track has enough quality, and pass the minimal required number of layers with hits in the tracking detector.

Based on the set of charged particles and strips contained in a jet, the HPS algorithm generates all possible combinations of hadrons for the following decay modes: $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$, and $h^\pm h^\mp h^\pm$. The combinations of charged particles and strips considered by the HPS algorithm represent all the hadronic τ lepton decay modes in Table 2.1, except for $\tau \to \pi^0\pi^\pm\pi^\pm\pi^\pm\nu_\tau$ with $\mathcal{B} = 4.3\%$. The τ_h candidates of charge other than ± 1 are rejected, as are those with charged particles or strips outside the signal cone, define by $R_{\text{sig}} = 3.0\,\text{GeV}/p_T$, where the p_T is that of the hadronic system, with cone size limited to range $0.05 - 0.10$. Finally, only the τ_h candidte with largest p_T is kept for further analysis, resulting in a single τ_h candidate per event.

Discrimination of τ_h candidates against light-jets

Requiring τ_h candidates to pass certain specific isolation requirements provides a strong handle for reducing the light-jet $\to \tau_h$ misidentification probability. A cone with $\Delta R = 0.5$ was originally used in the definition of isolation for all event types. However, in processes with a high number of final state objects, such as for Higgs boson production in association with top quarks ($t\bar{t}H$) the isolation is affected by the presence of nearby objects.

The L1, the first level trigger, composed of custom made hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of $\approx 100\,\text{kHz}$, went through a series of upgrades [65] in 2015 and 2016, and it is now based on more powerful, fully-programmable FPGA processors and μTCA logic boards. This allows more sophisticated τ_h reconstruction and isolation algorithms at L1. The performance of this reconstruction can be found in [54].

The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event before data storage. It reconstructs software, optimize the fast processing, and reduces the event rate to 1 kHz. This system uses the full-granularity information of all CMS sub-detectors, and runs a version of the CMS reconstruction that is slightly different than that used offline, as the HLT decision is made within 150 ms, on average, a factor of 100 faster than offline reconstruction. This is achieved using specialized, different versions of reconstruction algorithms, designed to reduce the number of events processed by more complex, and therefore more time consuming subsequent steps. Both methods are exploited in the τ_h reconstruction at the HLT. The previous methods are discussed in detail in [55].

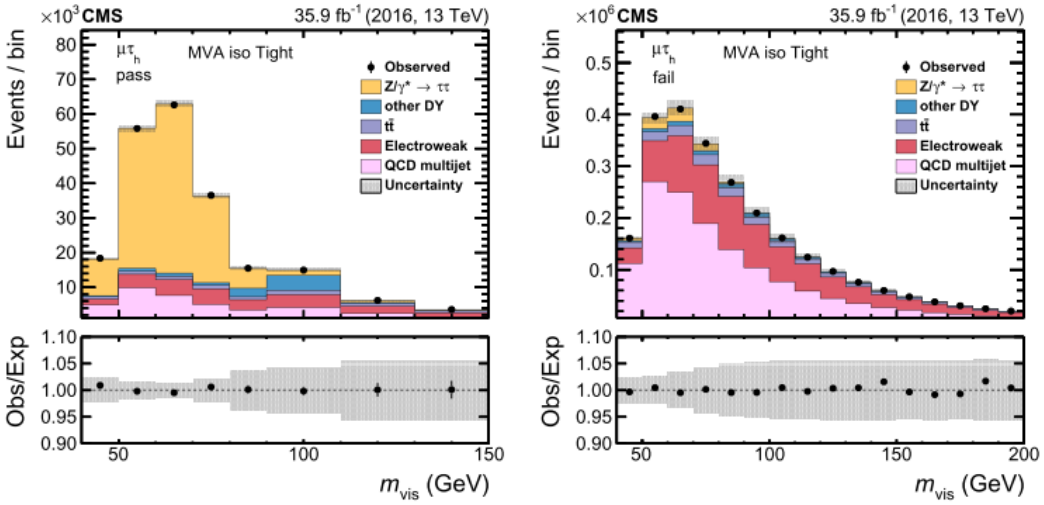


Figure 2.5: *The fitted distribution in m_{vis} in the passing (left) and failing (right) categories for the tight WP of the MVA-based isolation, plot taken from [56].*

Efficiency in τ_h identification

The efficiency is measured in different $p_T^{\tau_h}$ regions: small $p_T^{\tau_h}$ between 20 and $\approx 60\,\text{GeV}$, using $\mu\tau_h$ final state of $Z/\gamma^* \to \tau\tau$, intermediate $p_T^{\tau_h}$ of up to $\approx 100\,\text{GeV}$ using $\mu\tau_h$ final states in $t\bar{t}$ events, and high for $p_T^{\tau_h} > 100\,\text{GeV}$ using a selection of highly virtual W bosons ($m_W > 200\,\text{GeV}$) decaying into τ leptons. The data-to-simulation scale factors obtained through these measurements are combined to extrapolate to higher $p_T^{\tau_h}$ regions not covered by these measurements. Finally, the identification efficiency for τ_h candidates reconstructed using the algorithm dedicated to highly boosted τ lepton pairs is measured using the tag-and-probe method.

Tag-and-probe method in Z/γ^* events

The τ_h identification efficiency for $p_T^{\tau_h}$ up to $\approx 60\,\text{GeV}$ is estimated in $\mu\tau_h$ final state of Z/γ^*. The events are subdivided into passing (pass region) and failing (fail region) categories, depending on whether the τ_h candidate passes or fails the appropriate working point of the τ_h isolation discriminant. The data-to-simulation scale factor for the τ_h identification efficiency is extracted from a maximum likelihood fit of the invariant mass distribution of the reconstructed $\mu\tau_h$ system, referred to as m_{vis}. The SM predictions contributing to the distribution of the m_{vis} are a sample of $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events, where the reconstructed τ_h candidate is required to be matched to the generated one, and a set of backgrounds. All background events, except for QCD multijet production, rely on simulated m_{vis}. The results are presented in the Figure 2.5.

$t\bar{t}$ events

A sample of $t\bar{t}$ events with a muon and a τ_h in the final state is used to measure the τ_h identification efficiency for $p_T^{\tau_h}$ up to 100 GeV. The selected τ_h candidate must be accepted using the appropriate working point of the τ_h isolation discriminant. The distribution in m_T of the muon and the p_T^{miss} is used to determinate data-to-simulations scale factors.

Contributions to m_T distribution from $Z/\gamma^* \to \tau\tau$ single top quark, diboson, and W+jets

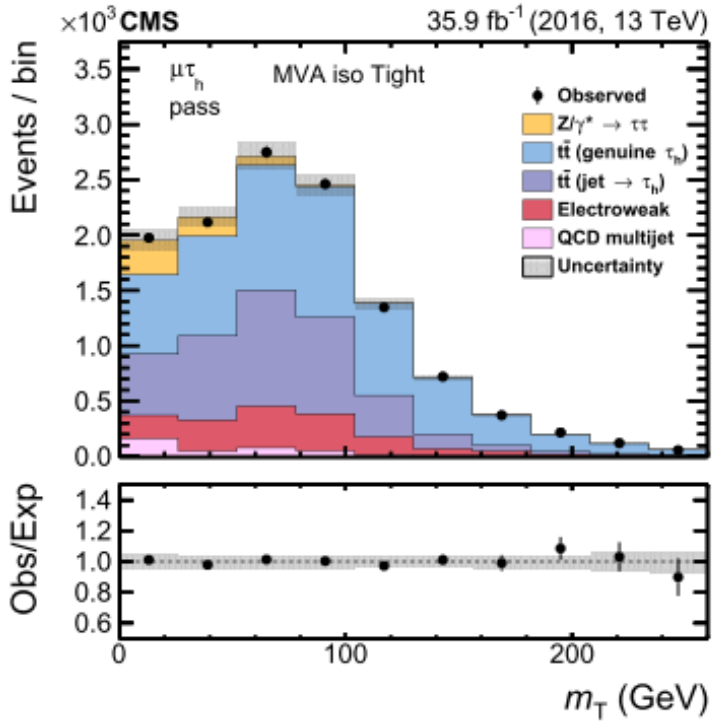


Figure 2.6: *Fitted distributions for the signal and backgrounds, plot taken from [56].*

events are modeled using simulations normalized to theoretical cross sections. The major background contribution is from $t\bar{t}$ events where a jet is misidentified as a τ_h candidate. The distribution is taken from simulation and a dedicated sample of events is selected to constrain the normalization of this background, as well as the probability of a jet to be misidentified as a τ_h candidate. The results are shown in the Figure 2.6.

Using off-shell $W \to \tau\nu$ events

The identification efficiency for τ_h leptons with $p_T > 100\,\text{GeV}$ is measured using a sample of events in which a highly virtual W boson ($m_W > 200\,\text{GeV}$) is produced at small p_T, and decays into a τ and ν_τ. The signature for those events consists of a single τ_h decay and p_T^{miss} balanced by the $p_T^{\tau_h}$.

A large fraction of events selected in this channel originate from processes where a jet is misidentified as a τ_h candidate. The main processes contributing to this kind of background are QCD multijet, $Z/\gamma^* \to \nu\bar{\nu}+$ jets and $W \to l\nu+$ jet events. The results are shown in Figure 2.7.

These results and discussion were taken from [55, 56].

2.3.4 b-jets

Jets that come from a bottom quark hadronization are called $b-$jets. The ability to identify $b-$jets is crucial in reducing the overwhelming background to these channels from processes involving jets that arise from gluons (g) and light-flavour quarks (u, d, s) and from c-quark fragmentation: jets that come from a charmed quark hadronization.

The properties of the bottom and, to a lesser extent, the charm hadrons can be used to identify

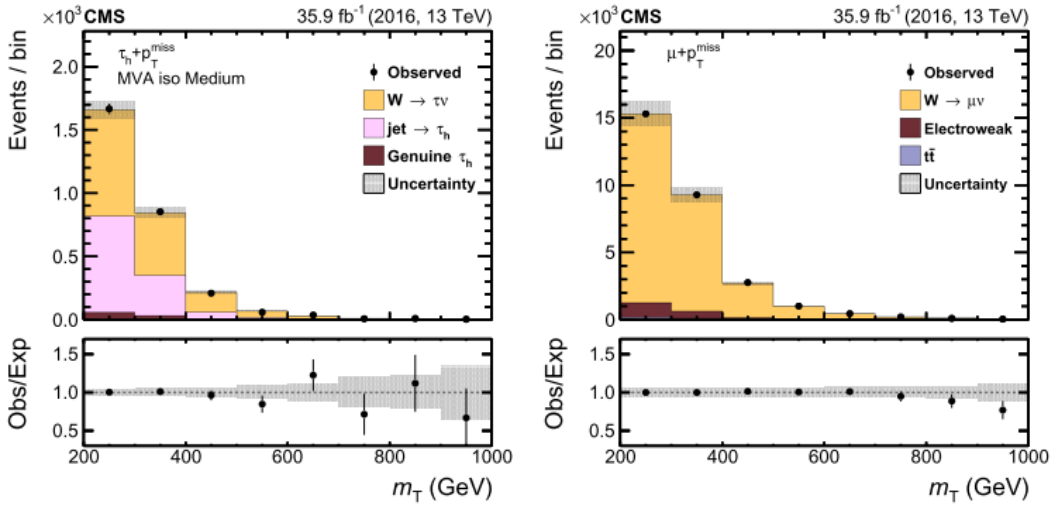


Figure 2.7: *The m_T distribution for selected $W \to \tau\nu$ (left) and $W \to \mu\nu$ (right) events after the maximum likelihood fit, plot taken from [56].*

the hadronic jets into which the b and c quarks fragment. These hadrons have relatively large masses, long lifetimes and daughter particles with hard momentum spectra. Their semileptonic decays can be exploited as well. The CMS experiment, with its precise charged-particle tracking and robust lepton identification systems, is well matched to the task of b-jet identification (b-jet tagging).

Algorithms for b-jet identification

Algorithms for b-jet identification need a variety of different objects as tracks, vertices and reconstructed leptons to discriminate between b and light-parton jets. Several algorithms have been developed, some of them use just one single observable, others are more complex and combine several of these observables to achieve a higher discrimination power. These algorithms yield a discrimination value for each jet. The minimum thresholds on this discriminator define the loose ("L"), medium ("M"), and tight ("T") operating points.

Identification using track impact parameter

The impact parameter (IP) is defined as the perpendicular distance between the path of a projectile and the center of a field. This field is generated by an object that the projectile is approaching towards. The IP can be used to distinguish the decay products of a b hadron from its tracks. The IP is calculated in three dimensions by taking advantage of the excellent resolution of the pixel detector along the z axis. Tracks originated from the decay of particles travelling along the jet axis usually have positive IP values. In contrast, the impact parameters of prompt tracks can have positive or negative IP values. The impact parameter significance SIP, defined as the ratio of the IP to its estimated uncertainty, is used as an observable.

Nevertheless the impact parameter significance by itself has no discriminating power between the decay products of b and non-b jets. So the *Track Counting* (TC) algorithm sorts tracks in a jet by decreasing values of the IP significance. Despite the algorithm bias the values for the first track to high positive significances, the light parton jets have low probability to have several

tracks with high positive values. Therefore the two different versions of the algorithm use the IP significance of the second and third ranked track as the discriminator value the *Track Counting High Efficiency* (TCHE) and *Track Counting High Purity* (TCHP).

There is a natural extension of the TC algorithms which is a combinatoin of the IP of several tracks in a jet. It's called the *Jet Probability* (JP) and uses an estimate of the likelihood that all tracks associated to the jet come from the primary vertex. Also the *Jet B Probability* (JBP) that gives more weight to the tracks with the highest IP significance, up to a maximum of four such tracks.

Identification using secondary vertices

The kinematic variables related to the secondary vertex can be used to discriminate between $b-$jets and non-$b-$jets. Two of the variables are the distance and the direction between the primary and the secondary vertices. Secondary-vertex candidates must satisfy the following requirements to enhance the b purity:

1. The vertices must share less than 65% of their associated tracks with the primary vertex and the significance of the radial distance between the two vertices has to exceed 3σ.

2. The secondary vertex candidates with a radial distance of more than 2.5 cm with respect to the primary vertex, with masses compatible with the mass of K^0 or exceeding $6.5\,\mathrm{GeV}/c^2$ are rejected, reducing the contamination by vertices corresponding to the interactions of particles with the detector material and by decays of long-lived mesons.

3. The flight direction of each candidate has to be within a cone of $\Delta R < 0.5$ around the jet direction

The *Simple Secondary Vertex* (SSV) uses the significance of the flight distance as the discriminating variable. Exist 2 types of algoritm just as in the case of TC; the *High Efficiency* (SSVHE) that uses vertices with at least two associated tracks, and the *High Purity* (SSVHP) where at least three tracks are needed.

A more complex approach involves the presence of a secondary vertex. Together with the track-based lifetime information, these algorithms are called the *Combined Secondary Vertex* and provide discrimination even in cases when the secondary vertices cannot be found.

The performance of the algorithms is specified in the Figure 2.8.

Finally in the Figure 2.9 and 2.10. The efficiencies and misidentification probabilities as a function of the jet p_T and the pseudorapidity are presented for the JPL and the CSVM taggers. There are used two different simulation samples: a QCD miltijet sample with p_T trigger threshold of $60\,\mathrm{GeV}/c$ applied only to the leading jet, and a $t\bar{t}$ sample. Only candidates with $p_T > 30\mathrm{GeV}/c$ and $|\eta| < 2.4$ were considered.

The previous analyses and Figures 2.8, 2.9 and 2.10 were taken from [57].

2.3.5 Missing Energy

Many detectors in particle physics experiments are blind to neutral weakly interactive particles. It is the case for CMS experiment and neutrinos. The neutrinos are neutral leptons with a

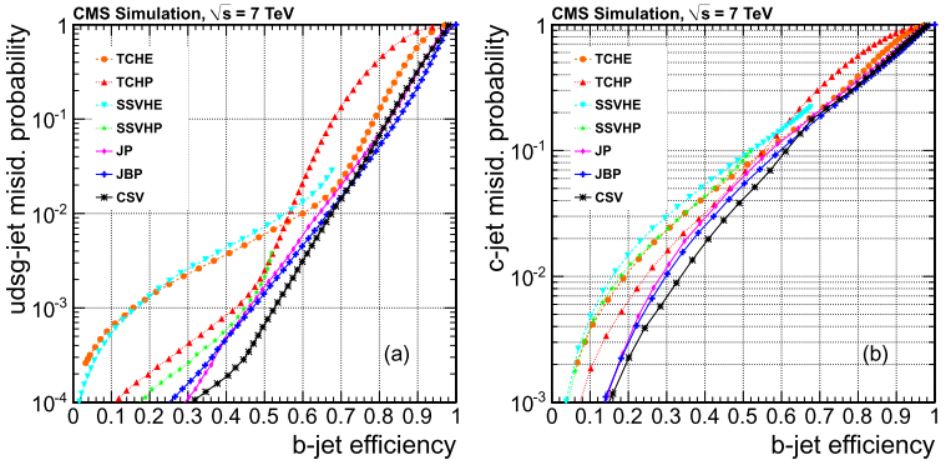


Figure 2.8: *Performance curves obtained from simulation for the different algorithms. (a) light-parton, and (b) c−jet misidentification probabilities as a function of the b−jet efficiency*

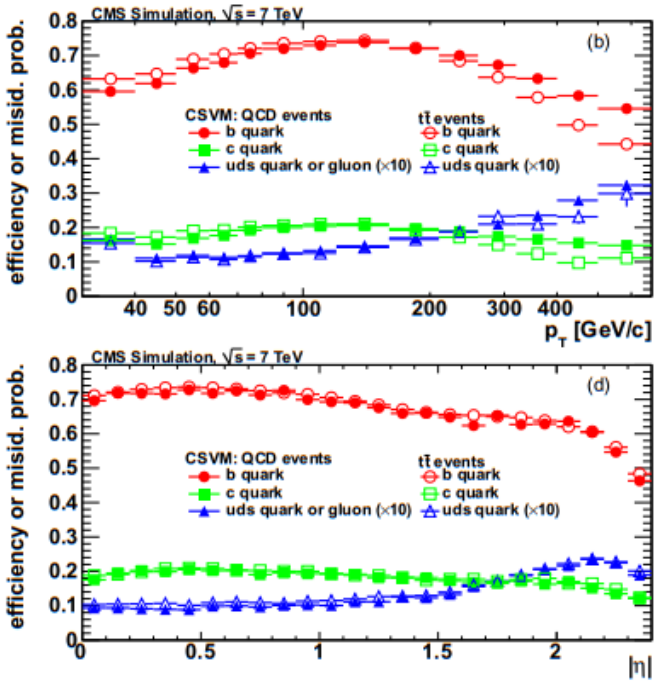


Figure 2.9: *Efficiency for b−jets and misidentification probabilities for c and light-parton jets of the CSVM taggers as a function of jet p_T and jet pseudorapidity in QCD multijet events (filled makers) and $t\bar{t}$ events (empty markers).*

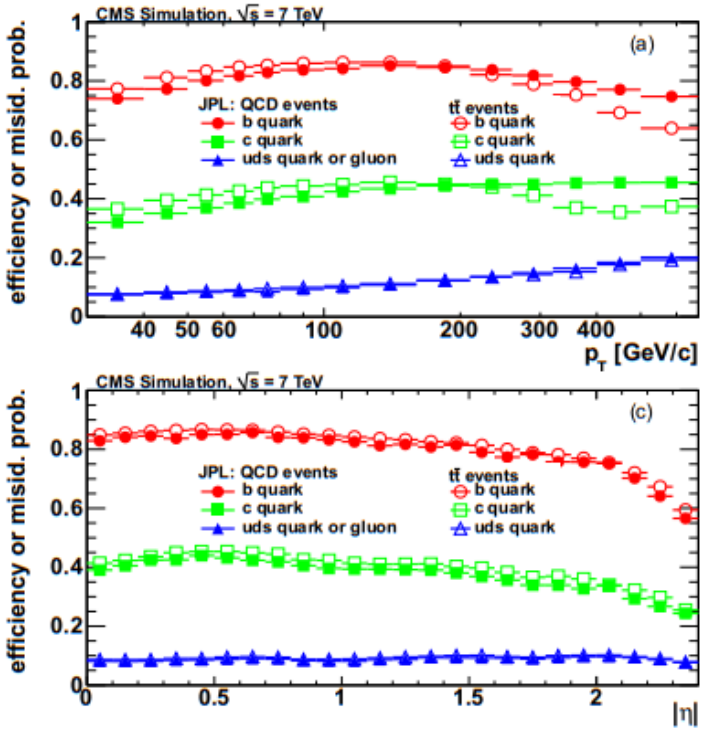


Figure 2.10: *Efficiency for b−jets and misidentification probabilities for c and light-parton jets of the JPL taggers as a function of jet p_T and jet pseudorapidity in QCD multijet events (filled markers) and $t\bar{t}$ events (empty markers).*

small mass compared with any other particle in the SM. These particles can travel through every single part in the detector without leaving any clue, more than unbalanced energy collisions. As the LHC collides protons at a scale of energy of 13 TeV in the center of mass, with the whole momenta being contained outside the transverse plane, the final state particles should have the same amount of energy in order to conserve energy-momentum. In that way in a collision where all the particles are detected and reconstructed but a neutrino, the event will have missing energy. Particularly, as the momentum in the transverse plane is zero before the collision, any other value for the momentum after the collision is considered as an unbalanced event and the missing energy tagged as an object.

The p_T^{miss} reconstruction algorithms.

The p_T^{miss} for the PF algorithm reefers to the negative value of the sum of p_T or the momentum in transverse plane of the objects involved for the PF per event. It is the most used algorithm since it provides a simple, robust but yet performant estimate of the p_T^{miss} reconstruction. However there is a second algorithm, and it relies on the "pileup per particle identification" (PUPPI) method [58].

The PUPPI method uses a local shape value called α that is sensitive between the collinear configuration of particles produced via hadronization of quarks and gluon produced in QCD mechanisms. This variable is calculated from each neutral particle, using the surrounding charged particles compatibles with the primary vertex.

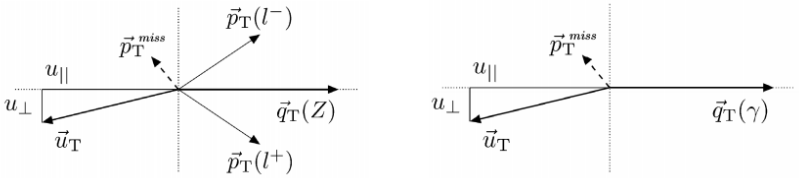


Figure 2.11: *Illustration of the Z boson(left) and photon(right) event kinematics in the transverse plane. u_T is the resultant of the sum over all the reconstructed particles in the event, except for the two leptons from the Z decay (left) or the photon (right).*

PUPPY uses PF candidates. For α for an i candidate is given by:

$$\alpha_i = \log \sum_{j \neq i, \Delta R_{ij} < 0.4} \left(\frac{p_{Tj}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5,\ j \text{ are charged PF candidates from primary vertex} \\ \text{for } |\eta_i| > 2.5,\ j \text{ are all kinds of reconstructed PF candidates} \end{cases} \tag{2.2}$$

where j refers to neighboring charged PF candidates originating from the PV within a cone of radius R in the $\eta - \phi$ plane around i, and ΔR_{ij} is the distance in the same plane between the i and j PF candidates.

An approximation of χ^2 is made

$$\chi^2 = \frac{(\alpha_i - \bar{\alpha}^2_{\mathrm{PU}})}{\mathrm{RMS}^2_{\mathrm{PU}}} \tag{2.3}$$

to determinate the likelihood that a PF candidate came from the pileup. Here the $\bar{\alpha}^2_{\mathrm{PU}}$ is the mean value of the α_i distribution for the pileup particles in a single event. And $\mathrm{RMS}_{\mathrm{PU}}$ is the corresponding root-mean-square (RMS) of the α_i distribution.

Performance of the PF algorithm

To test and understand the performance of the p_T^{miss} algorithms, the response and resolution of this object is studied in samples with an identified Z boson decaying into a pair of electrons or muons, or with an isolated photon. Such events have a small or no genuine p_T^{miss}, and the performance is measured by comparing the momenta of the vector boson to that of the hadronic recoil system. The hadronic recoil system is defined as the vector p_T^{miss} sum of all PF candidates except for the vector boson. In the Figure 2.11 the kinematic representation of the transverse momenta of the vector boson and the hadronic recoil, $\vec{q}_T$ and $\vec{u}_T$ are shown. In order to conserve the momentum in the transverse plane $\vec{q}_T + \vec{u}_T + \vec{p}_T^{\mathrm{miss}} = 0$.

The p_T^{miss} distribution in dilepton and photon samples are shown in Figure 2.12.

The resolution for the p_T^{miss} in these events is determined mainly by the hadronic activity. The uncertainty shown in the Figures 2.13 includes uncertainties in the Jet Energy Scale (JES) and the Jet Energy Resolution (JER)

Finally Figure 2.13. shows the response as a function of q_T indata and simulation in $Z \to \mu^+\mu^-$, $Z \to e^+e^-$ and photons events. The response reaches unity for boson p_T. Deviations from unity indicate imperfect calibration of the hadronic energy scale.

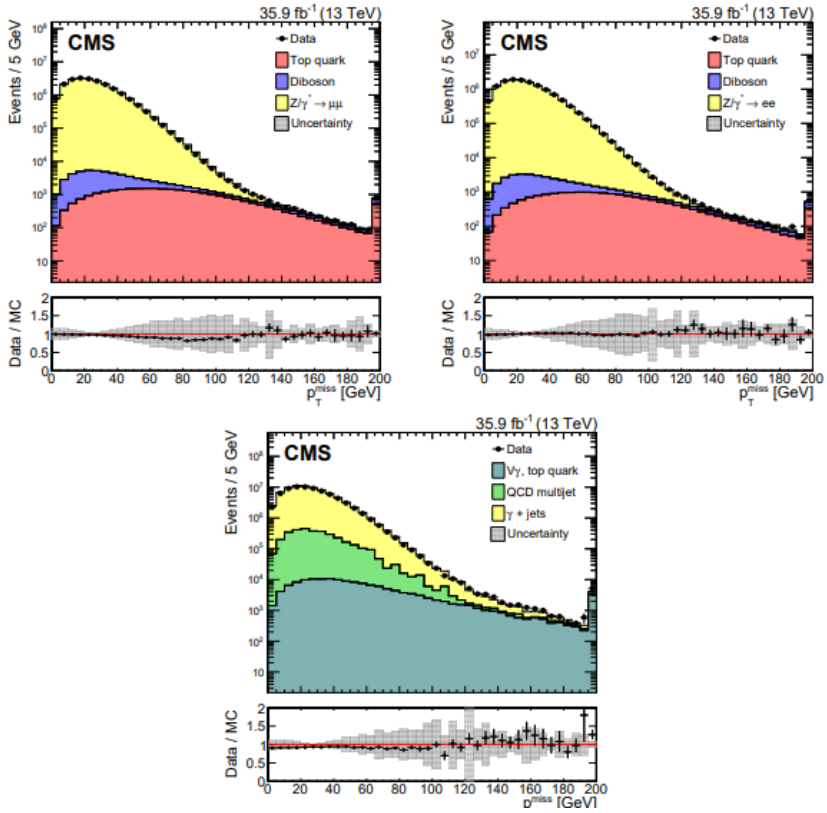


Figure 2.12: *Upper panel: Distribution of p_T^{miss} in $Z \to \mu^+\mu^-$ (left) $Z \to e^+e^-$ (right) and γ + jets (bottom) in data and simulation. The last bin contains information of all events with $p_T^{miss} > 195$ GeV. The lower panel: Data to simulation ratio, plot taken from [59].*

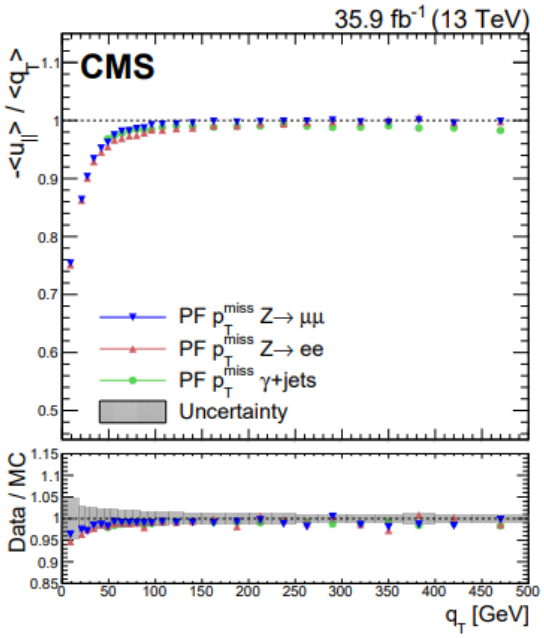


Figure 2.13: *Upper panel: Response of p_T^{miss} defined as $-\langle u_{\parallel}\rangle/\langle q_T\rangle$, in data in $Z \to \mu^+\mu^-$ (blue) and $Z \to \mu^+\mu^-$ (red) and γ + jets (green) events. Lower panel: Ratio of the p_T^{miss} response in data and simulation. The band corresponds to the systematic uncertainties due to the JES, the JER, plot taken from [59].*

The previous information and plots are detailed in [59].

Chapter 3

The problem

In the first chapter it was said that the SM was not a definitive theory. In that sense, it can be said that it does not explain every single observation one can make in the nature, for example it doesn't say anything about masses of neutrinos, or dark matter, or dark energy, or $g-2$, etc. Between all these discrepancies is the $R_{D^{(*)}}$ anomaly, a stain in the LFU, that says that the SM charged leptons should coupled the same way to the gauge-bosons, no matter it's flavor. To be more specific, the equation 1.4 details what the R_D and R_{D^*} quantifies, i.e. the ratio of the branching fractions of B mesons decaying in D mesons, τ and ν_τ, to the B mesons decaying in D mesons, l and ν_l. It can be seen that the SM predicts the second value of the equation 3.1, different from the measured value that is the first-one in the same equation. One could think, as the SM predicts the same decay rates for the charged leptons, then the value for $R_{D^{(*)}}$ should be $= 1$, Nevertheless, this is not a correct way to think about this quantity, because the $R_{D^{(*)}}$ measure the decay ratio of D or D^* to third generation leptons relative to the other possible modes of leptonic decay. These deviation from the theoretical model, actually differs from the observation at a global significance between 3 experiments of 3.4σ [60]. In this chapter are explored certain topologies related with the R_{D^*} anomaly and with physics beyond the SM that could be studied in the CMS experiment.

$$R\left(D^*\right) = \begin{cases} 0.304 \pm 0.013 \pm 0.007, & \text{Exp} \\ 0.257 \pm 0.005, & \text{SM} \end{cases} \tag{3.1}$$

3.1 The topology

Our final state of interest is a B hadron, a τ_h and p_T^{miss}. At a partonic level, it could be seen as a $\mathrm{b} + \tau + \nu$, as it is shown in the Figure 3.1, our proposal is to hunt an unknown particle that leads to that final state. In that sense, this mediator should couple to the three particles. The question would be how? We shall begin discussing the parton distribution function for this purpose.

3.1.1 Parton Distribution Function

Nowadays physicist know that hadrons are a composition of quarks and gluons, there are some "special" quarks that determine the quantum numbers of the hadrons, particularly three quarks for baryons and two for mesons. For example, protons are composed by two quarks up and one

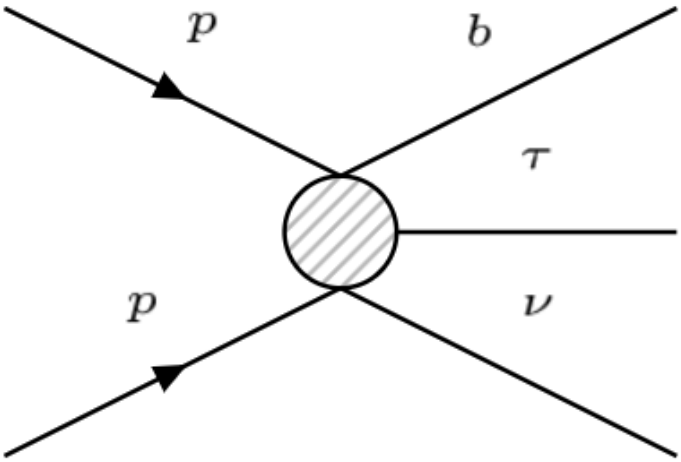


Figure 3.1: *Topology for the search. The most probable initial state includes a gluon and an up type quark. The circle in the midele of the diagram represents the mediation that is model dependent.*

quark down, the mass of those quarks round about 1 MeV, while the proton's mass rounds about 1 TeV, the explanation for the rest of the mass is energy in interaction and it's due to the gluons, permanent creating quarks and antiquarks, something that it's been called the sea quarks [61].

Nevertheless this knowledge is kind of recent. The first one to use the expression parton, was Richard Feynman, and he referred to them as a generic name for any constituent particle within the proton, neutron or any other hadron.

The determination of the PDF starts from a parametrization of the nonperturbative PDFs at a low scale, fits to various sets of experimental data (mainly to deep inelastic scattering data) that are performed within the DGLAP evolution scheme (An equation for QCD evolution, it's name reefers to its authors Dokshitzer-Gribov-Lipatov-Altarelli-Parisi). The resulting PDFs depend on the choice of the input data, the order in which the perturbative QCD calculation is performed, the assumptions about the PDFs, the treatment of heavy quarks, the correlation between α_s and the PDFs and the treatment of the uncertainties [62].

In the Figure 3.2 we observe the PDF at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2, the second one is a typical energy of a process in the LHC. In the vertical axis we have the probability to expect a parton, normalized to the expected number of a given parton in the proton, in the horizontal axis we have x, the fraction of the momentum carried by the parton. In the figure on the left we can observe different distributions for different partons, the most remarkably are the u, the d and the g (note that g is divided by ten). We can interpret that it's much more probable to find an up-quark carrying 1/10 of the proton's momentum, than finding an anti-up carrying that fraction of the momentum.

In a high energy process (the right-one) the landscape has changed a bit, now we have a new distribution. First, we can see that there is a new distribution, the $b, \bar{b}$. Also the distributions seem to be more stacked, but not just that, the distributions now show that there is more likely to find any parton carrying for example 1/1000 of the protons momentum in comparison with the process where $Q^2 = 10$ GeV^2. However, there is still a common characteristic between the

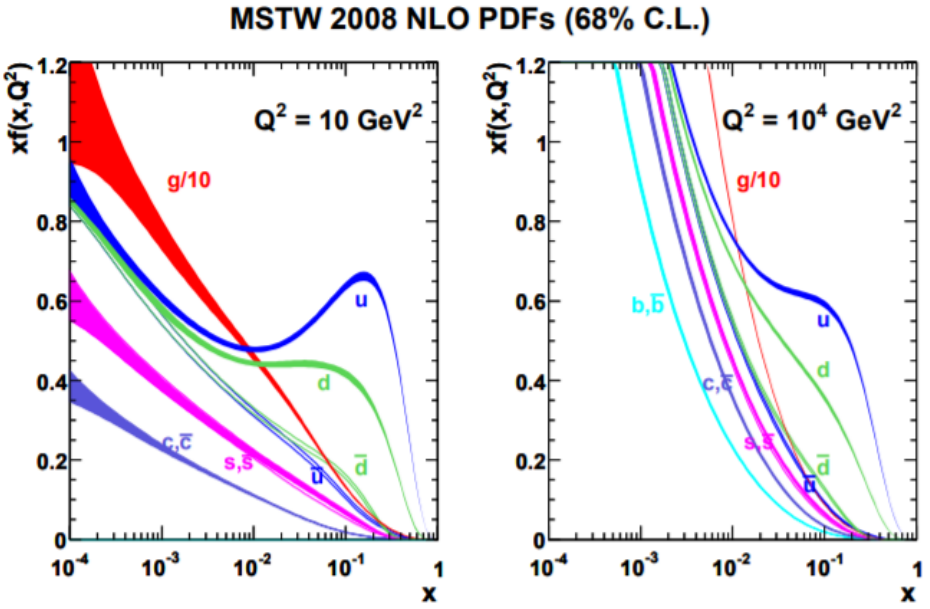


Figure 3.2: *MSTW 2008 NLO PDFs at* $Q^2 = 10\ GeV^2$ *and* $Q^2 = 10^4\ GeV^2$

two plots. There is much more likely to find an u, d or a g carrying a big amount of the proton's momentum than any other partons.

Taking all this into account and considering that we will need enough energy to create a ≈ 1TeV or 2 TeV W' which means that it will be needed a high amount of momenta in the initial state to produce an on-shell mediator. In that sense, the most probable initial state at tree level will be given by a an up-quark and a gluon.

3.2 The spectra

By this point an important question must be made. Detectors as CMS can detect certain signals, and consider them as candidates for jets, muons, electrons and photons, as we discussed in the past chapter. But mediators as gluons or $W^{\pm}$, Z^0 or $W'^{\pm}$ cannot be observed directly by the experiment. The only chance to hunt this mediators is to explore the kinematic of the objects that the detector can measure, and look for signals of those mediators. There is were we have the mass spectrum.

In this thesis as we are looking for study the phenomenology of 3 different models, we will consider 3 different mass spectres.

3.2.1 Invariant Mass

The invariant mass is a spectrum of mass consider in two-body decays. The idea of these mass is to consider the conservation of the 4^{th}- momenta in a two body decay

$$M_i^2 = E_f^2 - \mathbf{p}_f^2. \tag{3.2}$$

As we are considering a two boy decay, both the energy and the 3-momentum will have two contributions

$$\begin{aligned} M_i^2 &= (E_1 + E_2)^2 - \|\mathbf{p}_1^2 + \mathbf{p}_2^2\|^2 \\ M_i^2 &= E_1^2 + 2E_1E_2 + E_2^2 - \|\mathbf{p}_1 + \mathbf{p}_2\|^2. \end{aligned} \tag{3.3}$$

As the energy has the contribution of the rest mass, and the momenta per particle

$$M_i^2 = m_1^2 + \mathbf{p}_1^2 + 2E_1E_2 + m_2^2 + \mathbf{p}_2^2 - \|\mathbf{p}_1 + \mathbf{p}_2\|^2. \tag{3.4}$$

Now, if we consider for the final state particles that $\mathbf{p}^2 \gg m^2 \Rightarrow E^2 = \mathbf{p}^2$

$$\begin{aligned} M_i^2 &= \mathbf{p}_1^2 + 2\|\mathbf{p}_1\|\|\mathbf{p}_2\| + \mathbf{p}_2^2 - \|\mathbf{p}_1 + \mathbf{p}_2\|^2 \\ M_i^2 &= \mathbf{p}_1^2 + 2\|\mathbf{p}_1\|\|\mathbf{p}_2\| + \mathbf{p}_2^2 - \mathbf{p}_1^2 - 2\mathbf{p}_1 \cdot \mathbf{p}_2 - \mathbf{p}_2^2 \\ M_i^2 &= 2\|\mathbf{p}_1\|\|\mathbf{p}_2\| - 2\mathbf{p}_1 \cdot \mathbf{p}_2. \end{aligned} \tag{3.5}$$

By taking into account the Figure 2.4 we can write the Cartesian contributions of the 3-momenta the experimental framework

$$\begin{aligned} p_{\mathrm{x}} &= p_{\mathrm{T}} \cos\phi, \\ p_{\mathrm{y}} &= p_{\mathrm{T}} \sin\phi, \\ p_{\mathrm{z}} &= p_{\mathrm{T}} \sinh\eta, \end{aligned} \tag{3.6}$$

and by making some trigonometric algebra we can obtain

$$M_i^2 = 2p_{T_1}p_{T_2}\left[\cosh\left(\eta_1 - \eta_2\right) - \cos\left(\phi_1 - \phi_2\right)\right]. \tag{3.7}$$

This spectrum of mass will be consider for LQ models, where the b quark couples to the τ in the final state.

3.2.2 Transverse Mass

The transverse mass is a mass spectrum that assumes that the whole physics happens in the transverse plane. That means that $\theta_1 = \theta_2 = \pi/2$ and $\eta_1 = \eta_2 = 0$ which also means that Equation 3.7 can be written as

$$m_T^2 = 2p_{T_1}p_{T_2}\left[1 - \cos\left(\phi_1 - \phi_2\right)\right]. \tag{3.8}$$

This mass spectra is specially useful when p_T^{miss} is taken into account, because these object, it is measured only in the transverse plane, so there is no information of η associated. This spectra is specially determinant in the SSM where the τ couples to the ν.

3.2.3 Total Mass

The total mass is a 3-body-decay spectra. It can be said that is a generalization of the transverse mass, and that is because it also assumes that the whole physics happens in the

transverse plane. The procedure is analogous of the determination from the Equation 3.2 until the 3.5, but considering 3 bodies at the final state [63].

$$m_{\mathrm{T}}^{\mathrm{tot}} \equiv \sqrt{\left(p_{\mathrm{T}}^{\tau_1} + p_{\mathrm{T}}^{\tau_2} + E_{\mathrm{T}}^{\mathrm{miss}}\right)^2 - \left(\mathbf{p}_{\mathrm{T}}^{\tau_1} + \mathbf{p}_{\mathrm{T}}^{\tau_2} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}\right)^2} \tag{3.9}$$

This mass spectra is specially useful in the EFT models where the 3 particles in the final state are coupled in the same vertex.

Chapter 4

Analysis and Results

So far, it's been discussed the different characteristics of the detector, the models, and the objects related to the thesis. However to make an analysis, information is needed, and that information must be analyzed. In this chapter we present the proposal for a search for the models mentioned before with the objects related with this thesis. Starting by the some tools specialized in simulating and emulating decays, collisions and scatters in high energy process useful for this same search.

The second part of the chapter will be a discussion about how we work with these files results of the simulation. This section will be also separated in two parts: one, where the resulting files from the simulation are transformed, and another where the two steps of the analysis are discussed.

In the final part the results from the analysis will be shown in plots and tables.

4.1 Simulation

The data that come from collider experiments is the result of the whole possible combination of processes that could happen in nature at the energy scale of the collider. In that sense, in the data there are processes that can be explained with the SM, but also, there could be a bunch of events that cannot be explained by the SM. That means that there is another way needed to determine if there are events carrying a sign of new physics.

Simulating process is the answer to this need. With simulations, one can explore the phenomenology and behaviour of the particles and the detectors by only considering a certain type of process involved. The simulation chain in this work consist of three parts.

4.1.1 MadGraph5

MadGraph5 is a framework that provides all the elements needed for SM and BSM, that calculates numerically cross sections at LO, and in some times, at NLO if the model allows it, integrates with different software, and is provided with tools for event manipulation and analysis [64]. It needs a model, specified in a `UFO` (Universal FeynRules Output) which allows to translate all the information about a given particle physics model into a Python module that can easily be linked to existing matrix-element generators [65].

MadGraph is the first step in the simulation chain. It simulates a certain process at a partonic level. The main result is a file named `unweighted_events.lhe.gz` which contains the whole

information of the particles per event of the simulation.

4.1.2 Pythia8

Pythia8 is a software for high energy physics event generation. It can be implemented in the same frame-work of madgraph. It can estimate total and partial cross-sections, and contain soft interactions, PDF's, initial and final state showering, matching and merging matrix elements, multi-parton interaction, hadronization and fragmentation, etc [66].

Pythia8 is a crucial part in this study, because MadGraph5 only compute the process at a partonic label, nevertheless, in nature, quarks and gluons cannot be seen bare. As they live in nature, they suffer a process called hadronization, as it was discussed in 3.2 as in our final state, two of the particles decay hadronically, so it's very important to have a very precise simulation for the hadronic and showering processes.

4.1.3 Delphes

As some greeks would say that Delphes is the ancient state-city where the oracle used to predict people's future, particle physicist would say that is a `C++` framework where physics make predictions of particle physics. Specifically, Delphes allows to "emulate" the response from the detector front certain particles [67].

As it's been said CMS only can reconstruct certain objects as jets, muons, electrons, photons and p_T^{miss}. It's mandatory to simulate the response and the efficiency of the detector in order to understand the type of signal physicist can expect from the search.

The simulation chain is then made by 3 steps, partonic simulation, performed by MadGraph5, hadronization, and decays simulation, made by Pythia8 and finally the emulation of the detector response carried out by Delphes. At the end of the simulation chain the result is a `.root` file. A tree-type data structure, with the whole information per-event, per-object. In the next section it'll be discussed how this structure is converted into a table-type structure and how it was analyzed.

4.2 Monte Carlo

It's needed different type of information in order to have an idea of what one can expect from the SM and what one can expect from BSM physics. From now and on, we are calling signal to those montecarlo events that have the final state that we are interested in, and also are mediated with BSM physics. Monte carlo events that lead to the interesting final state, mediated by interaction explained by the SM are called backgrounds.

4.2.1 Signals

As it's seen already, our signals will be mediated by processes ruled by SSM, EFT, and LQ. we made 50 k of Monte Carlo (MC) events for each sample, with 3 different samples for each model. The parameters are specified in the Table 4.1. The rest of the parameters are set by default by models built for [20].

Model	Parameters	Cross_Section(pb)
SSM	$m_{W'} = 600$ GeV	5.25
	$m_{W'} = 1$ TeV	0.45
	$m_{W'} = 1.6$ TeV	0.03
EFT	$\epsilon_l^{cb} = 1.0$	0.13
	$\epsilon_{sL}^{cb} = 1.0$	0.08
	$\epsilon_t^{cb} = 1.0$	0.71
LQ_U(1)	$mLQ = 1$ TeV	0.02
	$mLQ = 2$ TeV	3×10^{-4}
	$mLQ = 3$ TeV	9.82×10^{-6}

Table 4.1: *Signal parametrization and cross sections*

Process	Parameters	Cross_Section(pb)
$t\bar{t}$	–	5.04×10^2
$W + jets$	`ptl` $= 190$ GeV, `misset` $= 160$ GeV	0.64
$DY + jets$	`ptl` $= 190$ GeV	0.25

Table 4.2: *Backgrounds parametrization and their respective cross section*

4.2.2 Backgrounds

The idea with the backgrounds is to consider the processes which can lead to the same final state as the signal, but can be explained by the SM. In the context of the LHC, the most probable processes explained by the SM which can lead into $B + \tau_h + \mathbf{p}_T^{\text{miss}}$ are $t\bar{t}$ →semileptonic, $W + jets$ and $Z + jets$. The information is in Table 4.2.

4.3 Data Processing

By its self, Monte Carlo events would not say too much. The aim of these section is to clarify and specify the technical process that was made in order to obtain the results. The it's described the usage of the packages Awkward and Uproot and the design of the data structure used.

4.3.1 Uproot & Awkward

Root is an open-source framework for data analysis specialized in high energy physics. It can be manipulated as a `C++` interpreter. In that sense, `.root` files can be manipulated via Root, with a `C++` syntax. Nevertheless, in the recent years many physicist have migrate to high level programming languages as Python, so other frameworks have been developed in order to manipulate the `.root` files with a Python interpreter. Uproot is a library that allows to read root files with a python interpreter, and Awkward helps to work with "jagged" arrays, or n-dimensional arrays with no need that every column has the same number of instances [68]. This is needed because not every event has the same number of objects as jets or electrons or others.

As it's been already said this `.root` files are tree-like data structures. One of our first main tasks was to develop a framework where these structures are transformed into table-type structures. This supposes a first difficulty, because there could be events with more than one jet, and

Event	Jets	p_T	η	ϕ
1	1	132	2.1	0.8
	2	58	1.4	-1.3
2	1	43	-1.7	0.5
3	1	81	-0.5	1.4
	2	46	0.1	-0.9
	3	46	-0.1	0.9
4	1	87	2.3	3.14
	2	31	1.1	-0.6
6	1	18	0.2	-2.5

Table 4.3: *Example of an initial data-structure built by uproot. The total jets per-event are considered, in the case there are no jets in a event, events with no jets are not considered.*

other with no jets. If we try to put this raw data with uproot, in the case of jets we wold obtain something like the table 4.3. After a transformation the data structure would look like Table 4.4.

Event	jet1_p_T	jet1_η	jet1_ϕ	jet2_p_T	jet2_η	jet2_ϕ
1	132	2.1	0.8	58	1.4	-1.3
2	43	-1.7	0.5	NaN	NaN	NaN
3	81	-0.5	1.4	46	0.1	-0.9
4	87	2.3	3.14	31	1.1	-0.6
5	NaN	NaN	NaN	NaN	NaN	NaN
6	18	0.2	-2.5	NaN	NaN	NaN

Table 4.4: *Example of processed jet kinematic data from the Table 4.3, in this case only two jets per-event are considered.*

In our case, apart from the kinematic information from the first 4 jets per-event ordered descending in p_T, the *jet-tau-tag* and *jet-b-tag* information, that are binary variables were used, considered 1 in the case the jet is tagged as a τ or a b and 0 if they are not. Also the $\mathbf{p}_T^{\text{miss}}$ is included.

4.3.2 Python Analysis

Once the table-type-structure is ready, an event selection criteria is implemented. The idea is to economize computational costs, in that sense we begin with very loose cuts. Only events with one τ - jet and one b -jet are needed, so only were considered events with more than one jet. After that, a filter of at least a jet tagged as a τ and other different from the first one, to be tagged as a b. Those jets are required to have a $p_T \geq 25$ GeV and 30 GeV respectively as it's explained in [55, 57]. If there are more than one jet tagged as τ (or a b) the one with a greater p_T is prioritized. There also could happend to be two jets tagged as both τ and b in that case, the τ is considered as the one with the higher p_T.

Once this is considered, the result will be a Dataframe with $p_T(\tau)$, $\eta(\tau)$, $\phi(\tau)$, $p_T(b)$, $\eta(b)$, $\phi(b)$, p_T^{miss}, $\phi(p_T^{\text{miss}})$. With this information the kinematic can be made, and the constructions of the mass spectra described in 4.2.

The main task of this work is to determine the difference between the SM and possible theories

of BSM. In that sense, in this chapter are shown the results of SSM, EFT and LQ separately in comparison with the SM.

Results are motivated by the $R_{D^{(*)}}$ anomaly. Having this into account, and reminding that the information processed comes from MC events, the comparison leads to significance curves and mass spectra histograms that could motivate future searches for hunting or discarding parameter spaces for the BSM models in pp collider experiments.

4.4 Unfiltered Data

Once one can process the tree-like data structure, it's wanted to know what was the behavior of the different objects in the different kinematic variables. So there're produced different histograms for the different variables comparing signals and backgrounds in order to discover the variables one must use to produce linear cuts to maximize the statistical significance [69]

$$Z = \frac{N_S}{\sqrt{N_S + N_B}}. \tag{4.1}$$

In the first place one can study the problem at a CMS-object level. In the case of this study, it was wanted to see if there was any specific variable for the p_T^{miss} or the jets which has a different distribution for the signals and for the backgrounds. The reason for this is that it is desired to differentiate kinematically between signal, remembering that these are montecarlo events from BSM and background or SM montecarlo events. The results for the jet's kinematics are in Figures 4.1, 4.2, and 4.3. The results for the missing energy are shown in the Figures 4.4 for the SSM. The same results for EFT are shown in Figures 4.5, 4.6, 4.7 and 4.8. The LQ analogous distributions are in the Figures 4.9, 4.10, 4.11 and 4.12.

From now and on, the 3 backgrounds $t\bar{t}$, W + jets and $Z + jets$ will be denoted by the colors cyan, purple and lemon-green for the solid histograms. The signals for each model will be transparent histograms with blue, red and green colors.

There are quiet remarkably characteristics across the first jet results for all the models. In the Figures 4.1, 4.5 and 4.9, the plot in the upper left panel, that represents the p_T from the first jet per event, shows a clear differentiation respect the backgrounds. If we think about the diagram in the Figure 4.1 one can imagine that the first jet must be associated either with the b or the τ, and the second jet would be associated with the other particle. In that sense, a big amount of energy that comes from the collision is carried by the first two jets. Another two remarkable characteristics come from the upper right and lower left panels from the Figures 4.1, 4.2, 4.3, 4.5, 4.6, 4.7, 4.9, 4.10, 4.11. In those distributions is easily to see that the jets had no preferred direction in the azimuthal angle ϕ. This distribution serves as a control variable to check the quality from the simulations, besides that there also can be seen that the favourite value for the pseudorapidity η is zero, and seems to be a very symmetric distribution.

The last remarkable characteristic comes from the lower right panel from Figures 4.2, 4.3 4.6, 4.7, 4.10 and 4.11. These histograms represent the second jet with the highest p_T in the event. In an ideal scenario, the whole physics after the collision happens in the transverse plane i.e. $\eta = 0$. In that way, the partons from the two protons collide, produces a, fro the SSM, a W' and a b quark, then, this W' decays into a τ and ν. In this case, is likely that the first jet comes from

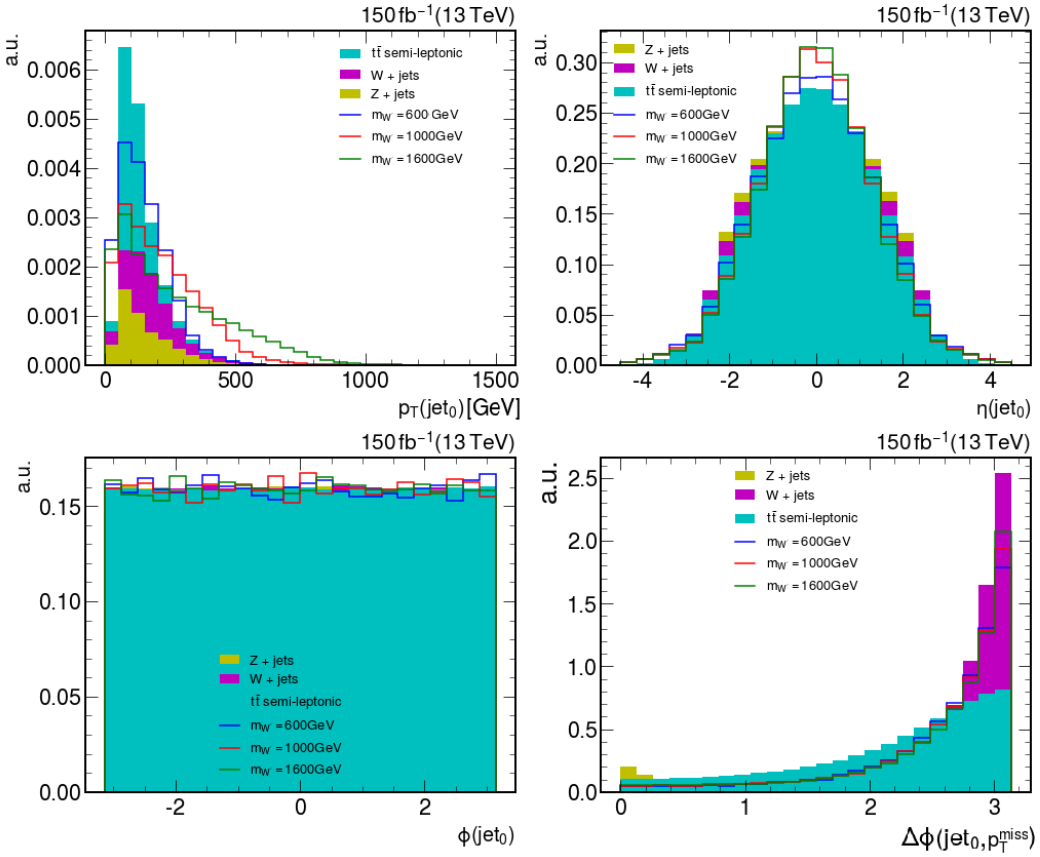


Figure 4.1: *First jet kinematic behavior per event for SSM signals and background. In the upper-left panel is plotted the $p_T(jet_0)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_0, p_T^{miss})$.*

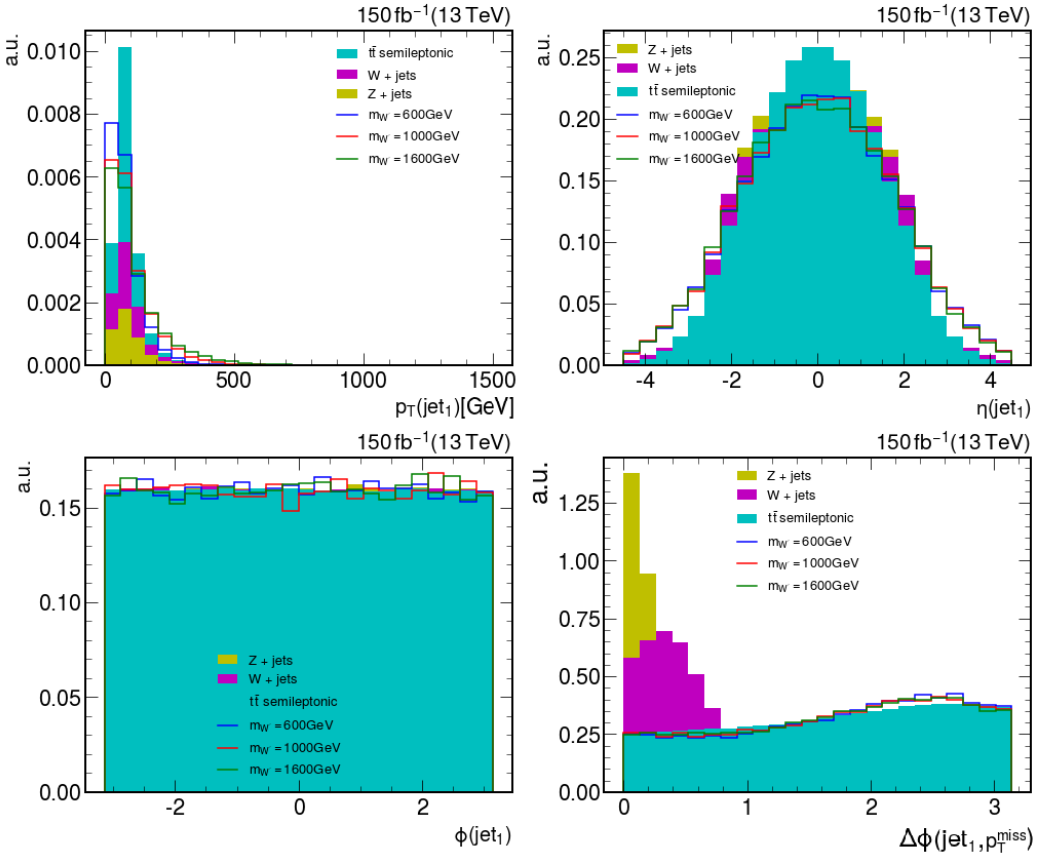


Figure 4.2: *Second jet kinematic behavior per event for SSM signals and background. In the upper-left panel is plotted the $p_T(jet_1)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_1, p_T^{miss})$.*

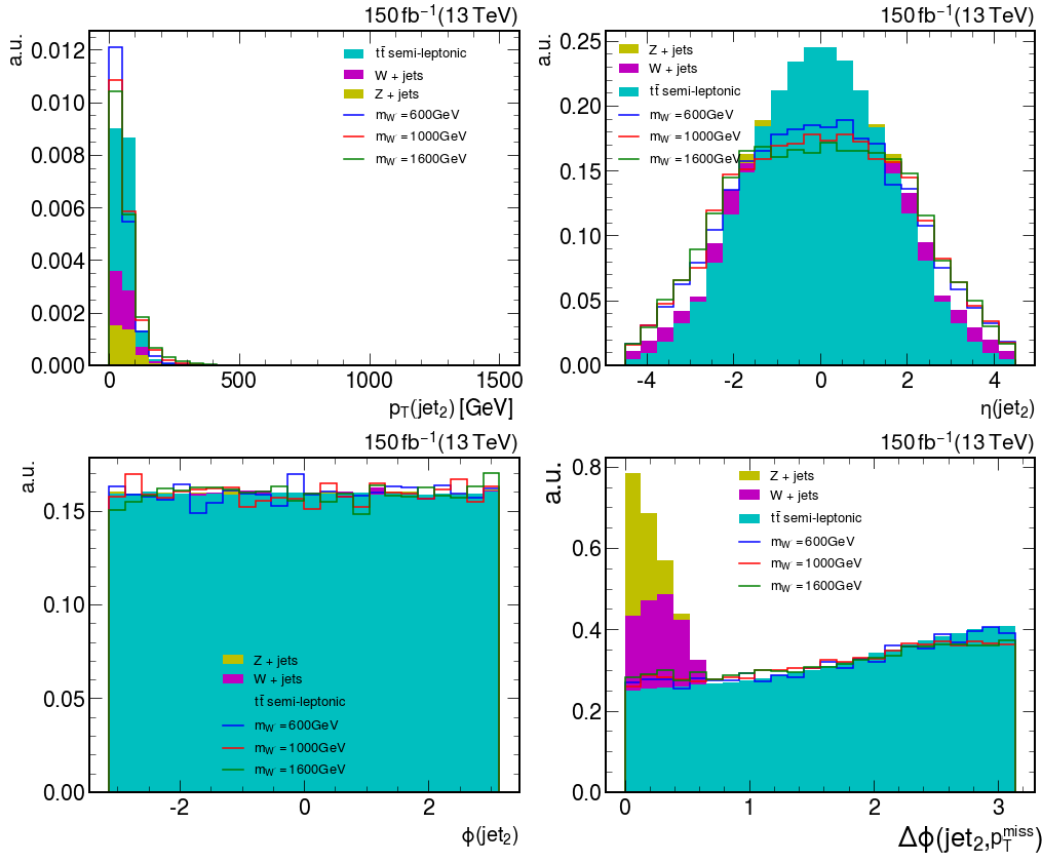


Figure 4.3: *Third jet kinematic behavior per event for SSM signals and background. In the upper-left panel is plotted the $p_T(jet_2)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_2, p_T^{miss})$.*

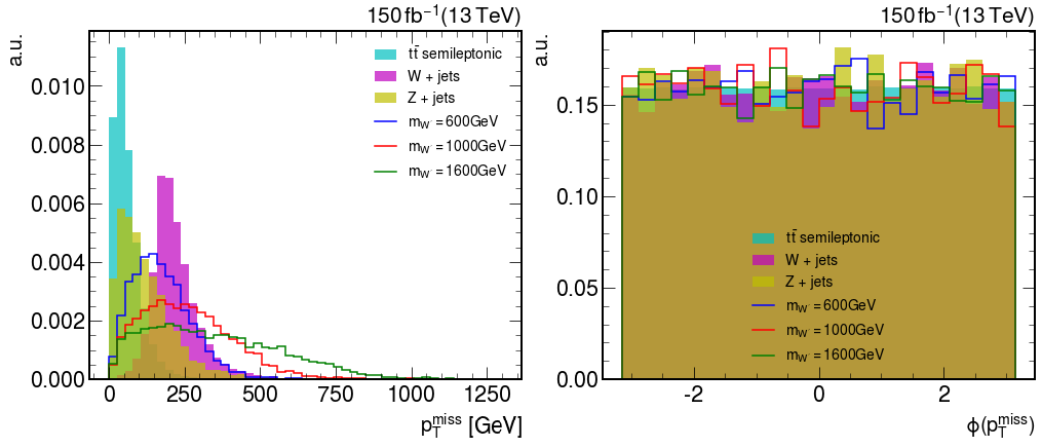


Figure 4.4: $\boldsymbol{p}_T^{miss}$ *information for the SSM. The left panel shows the distribution for the magnitude of the missing transverse momentum, the right panel shows the distribution of $\phi(p_T^{miss})$. As this histogram shows the distribution of an object in the final state, the backgrounds are not stacked and present a transparency in order to understand the difference between each backgrounds and the signals.*

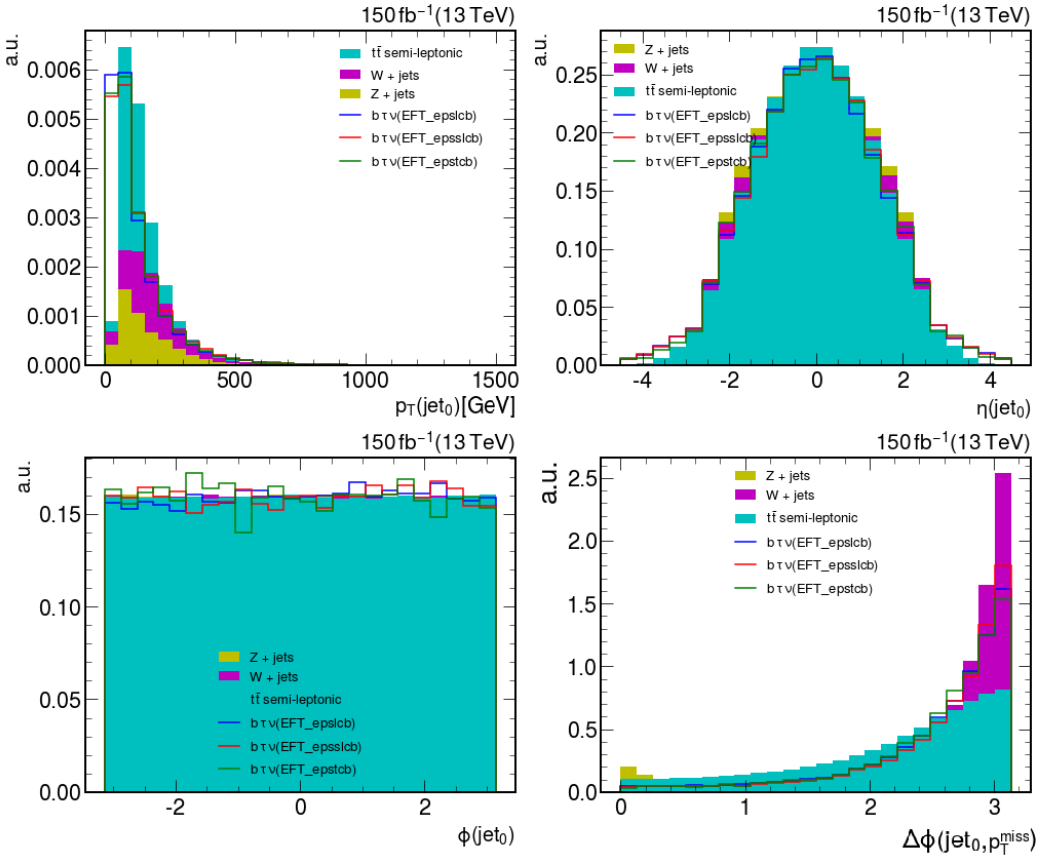


Figure 4.5: *First jet kinematic behavior per event for EFT signals and background. In the upper-left panel is plotted the $p_T(jet_0)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_0, p_T^{miss})$.*

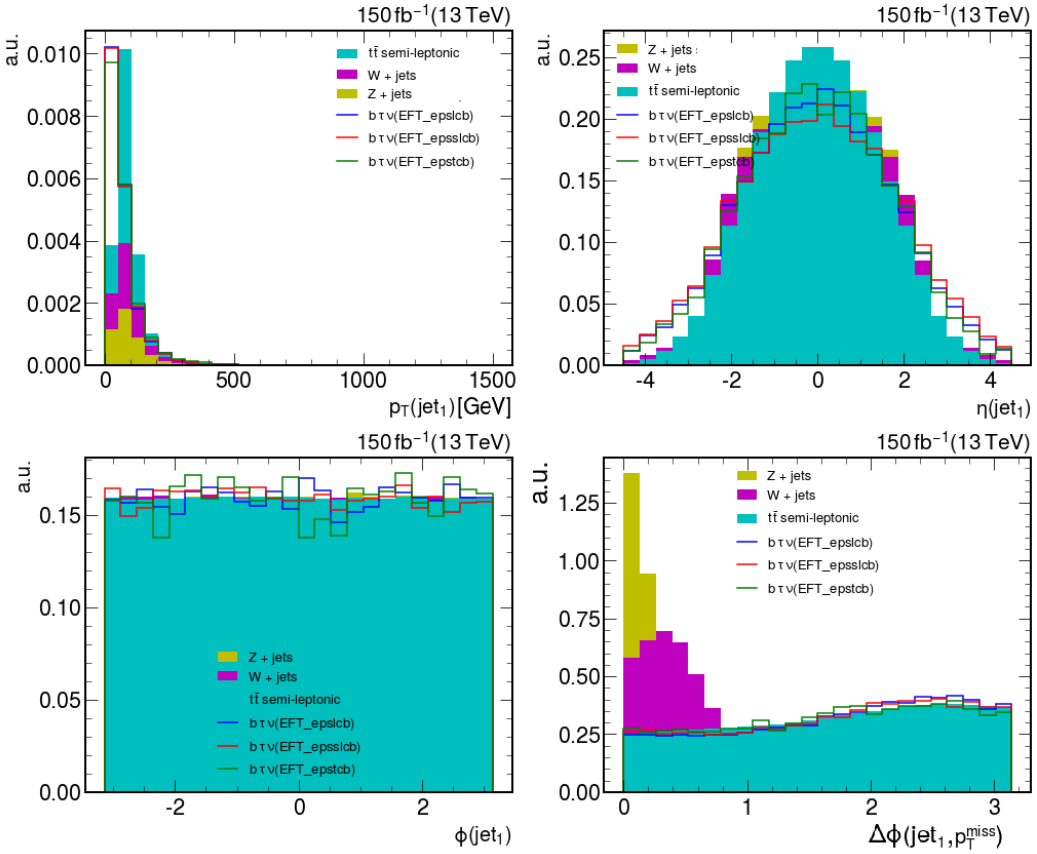


Figure 4.6: *Second jet kinematic behavior per event for EFT signals and background. In the upper-left panel is plotted the $p_T(jet_1)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_1, p_T^{miss})$.*

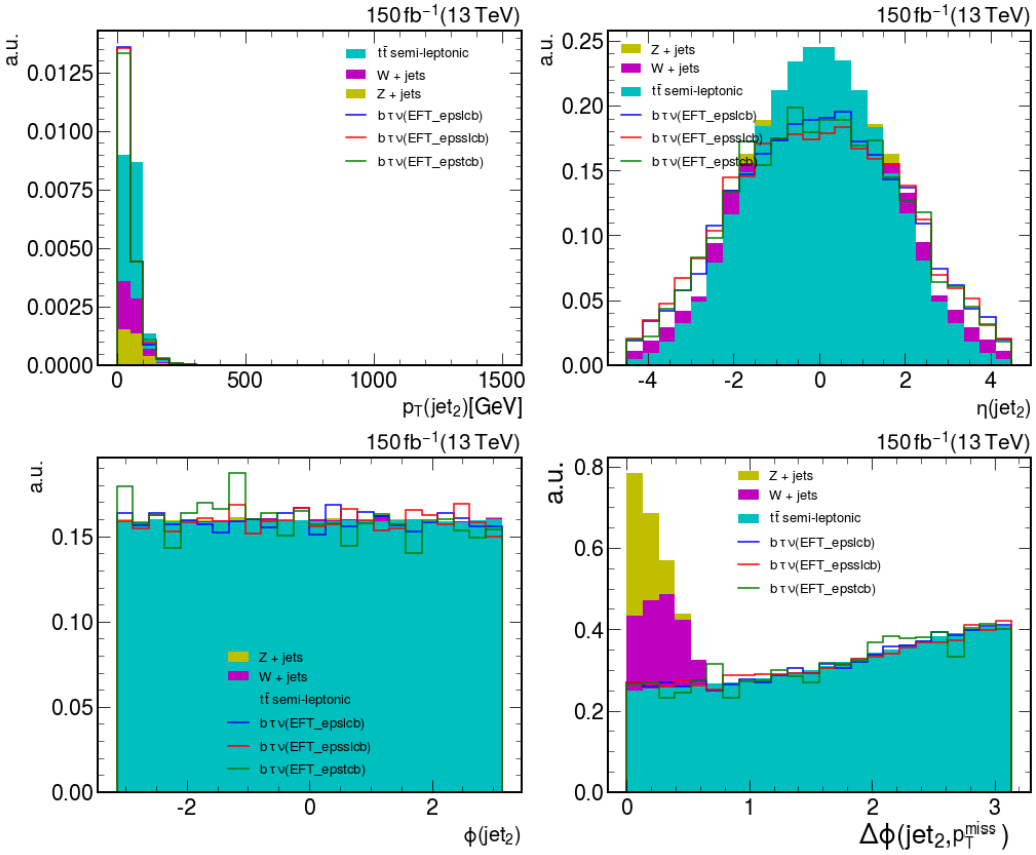


Figure 4.7: *Third jet kinematic behavior per event for EFT signals and background. In the upper-left panel is plotted the $p_T(jet_2)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_2, p_T^{miss})$.*

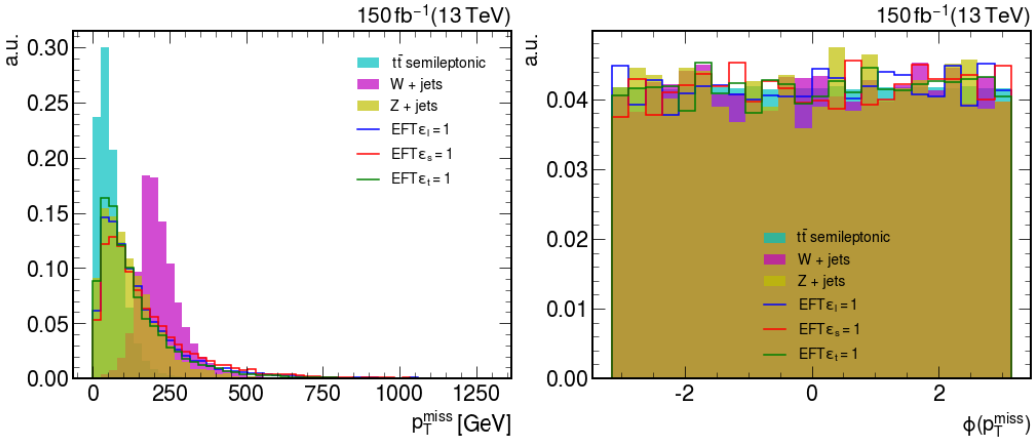


Figure 4.8: $\boldsymbol{p}_T^{miss}$ *information for the EFT. The left panel shows the distribution for the magnitude of the missing transverse momentum, the right panel shows the distribution of $\phi(p_T^{miss})$. As this histogram shows the distribution of an object in the final state, the backgrounds are not stacked and present a transparency in order to understand the difference between each backgrounds and the signals.*

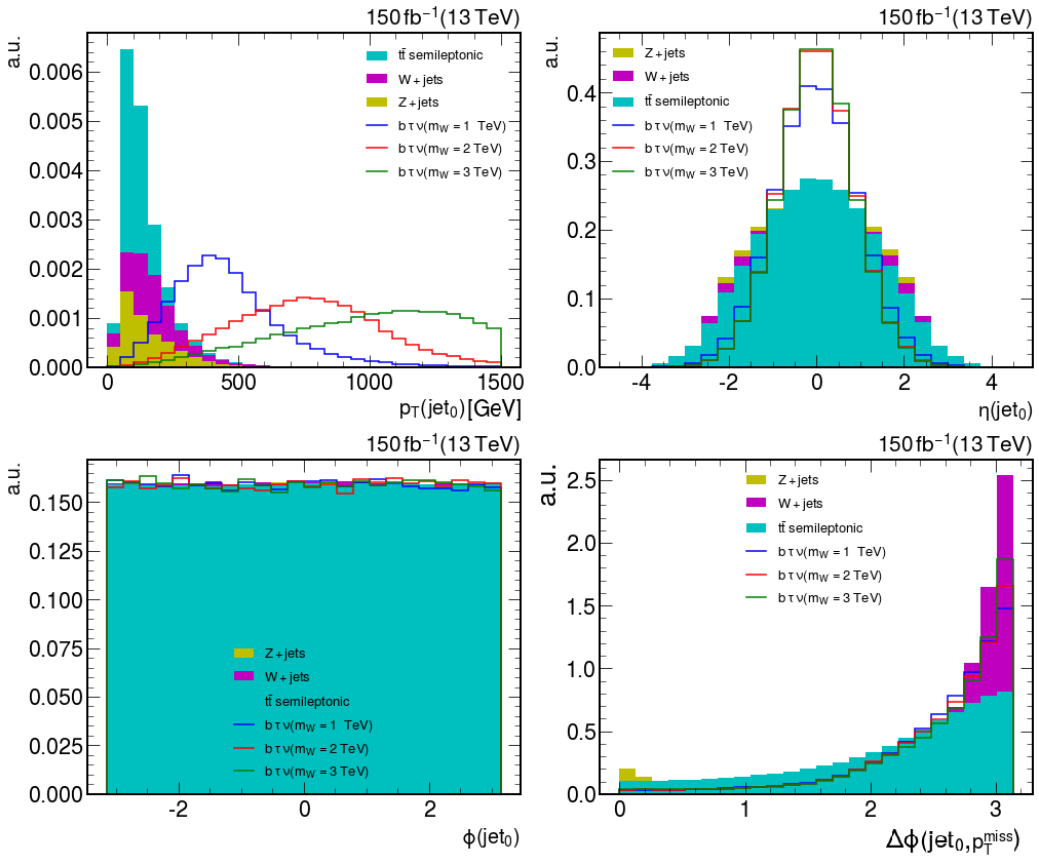


Figure 4.9: *First jet kinematic behavior per event for LQ signals and background. In the upper-left panel is plotted the $p_T(jet_0)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_0, p_T^{miss})$.*

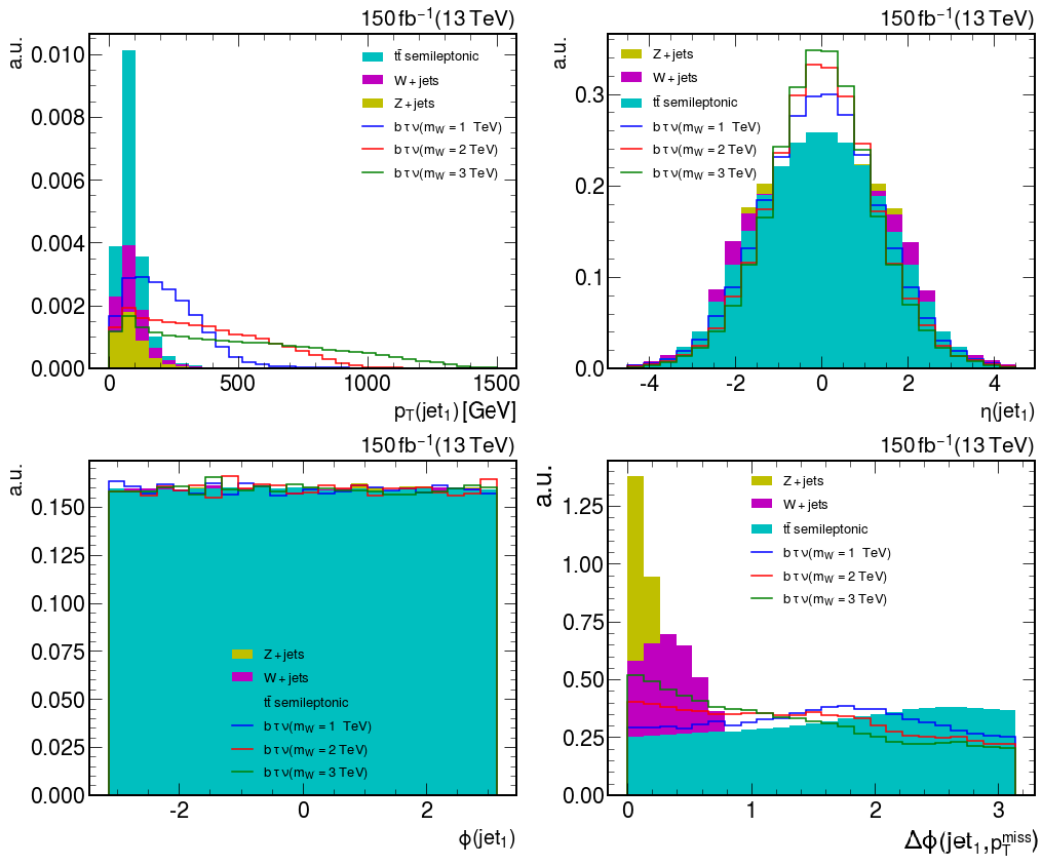


Figure 4.10: *Second jet kinematic behavior per event for LQ signals and background. In the upper-left panel is plotted the $p_T(jet_1)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_1, p_T^{miss})$.*

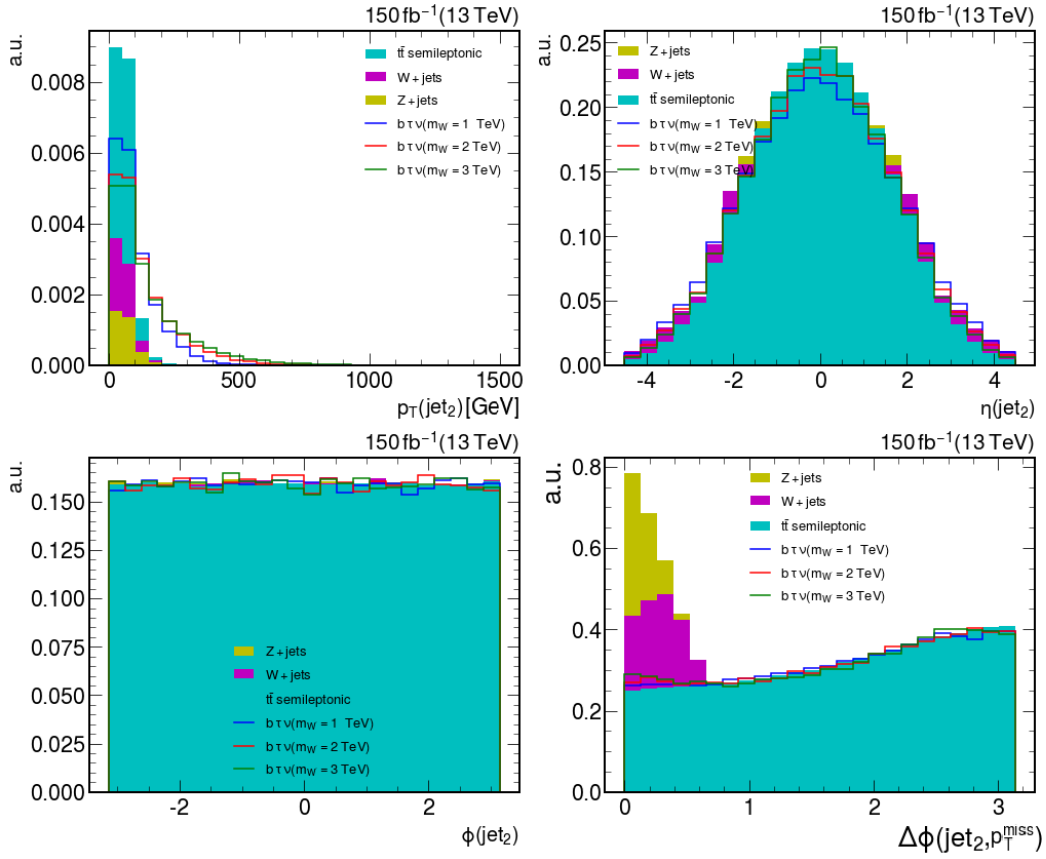


Figure 4.11: *Third jet kinematic behavior per event for LQ signals and background. In the upper-left panel is plotted the $p_T(jet_2)$, the upper-right one shows the pseudrapidity η, the bottom at left shows the azimuthal angle ϕ, and the bottom-right panel shows the $\Delta\phi(jet_2, p_T^{miss})$.*

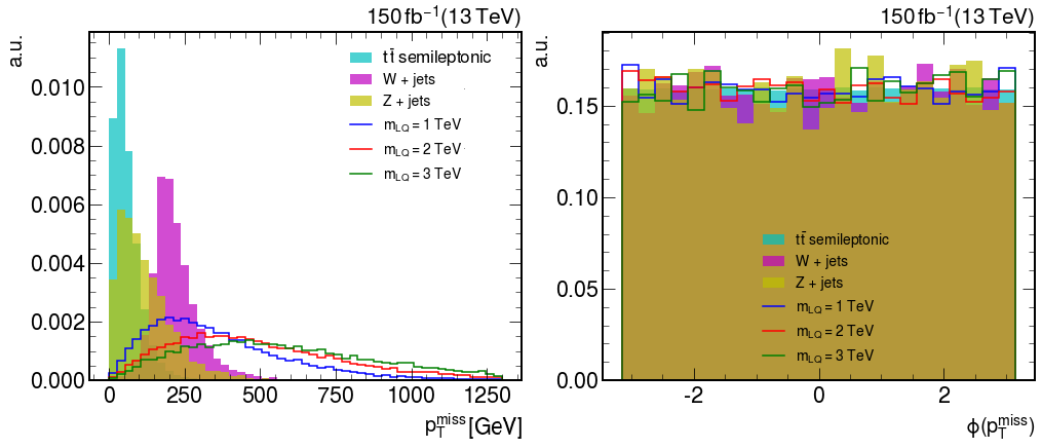


Figure 4.12: $\boldsymbol{p}_T^{miss}$ *information for the LQ. The left panel shows the distribution for the magnitude of the missing transverse momentum, the right panel shows the distribution of $\phi(p_T^{miss})$. As this histogram shows the distribution of an object in the final state, the backgrounds are not stacked and present a transparency in order to understand the difference between each backgrounds and the signals. As this histogram shows the distribution of an object in the final state, the backgrounds are not stacked and present a transparency in order to understand the difference between each backgrounds and the signals.*

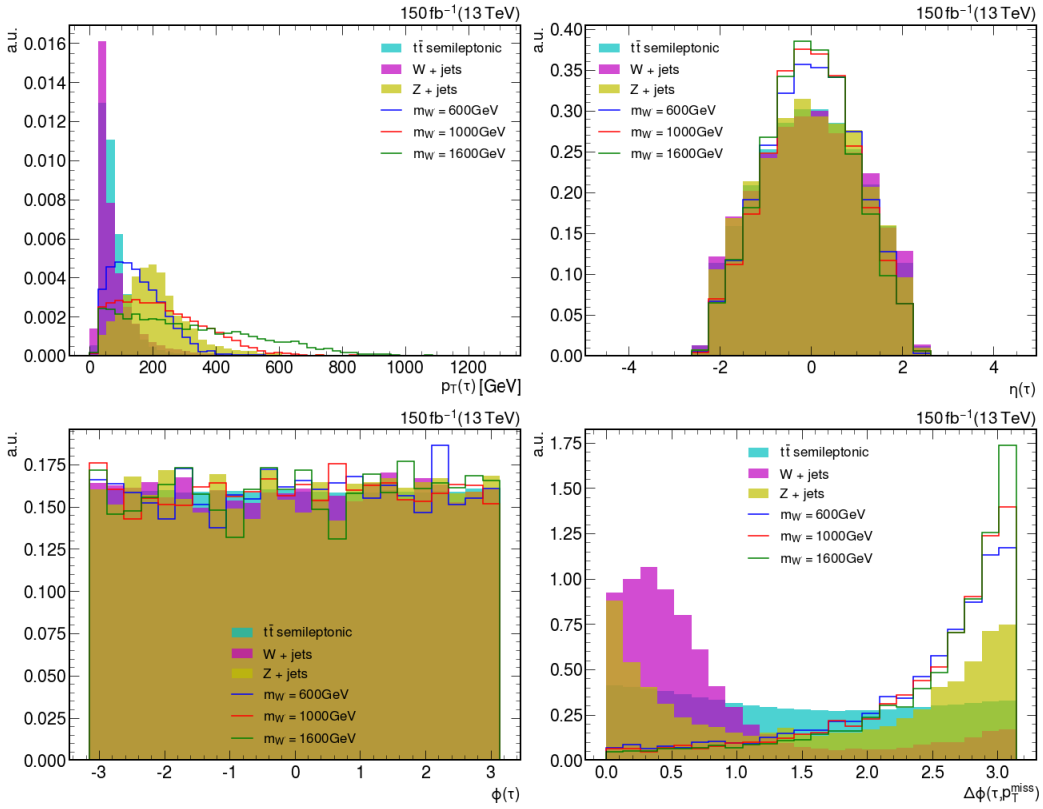


Figure 4.13: *τ kinematic information for the SSM. The upper panels are the p_T and the η from left to right, the lower are the azimuthal angle ϕ and $|\Delta\phi(\tau, \boldsymbol{p}_T^{miss})|$ in the same order. As this histogram shows the distribution of an object in the final state, the backgrounds are not stacked and present a transparency in order to understand the difference between each backgrounds and the signals.*

the b quark, and the second one comes from the τ. Since the decay from the W' produces both τ and p_T^{miss} and the τ decays also into p_T^{miss} and hadrons, some of the information from the τ gets lost in this process in a p_T^{miss} way. Also, if the W' is produced at rest, the τ and the p_T^{miss} will be produced back to back. So it's not weird that the maximum value for the histogram of $\Delta\phi$ between the second or the third jet and the p_T^{miss} for the signals of SSM to be $\approx \pi$

In the case of the EFT. The 3 particles comes from the same vertex. Nevertheless the same behavior is shown for the second and third jet, and similar output is observed in the LQ results.

This generalized behavior motivates to search the best candidates to τ and b in order to find the cuts that maximize the statistical significance. That means that only events with two jets must be considered. Also two of these jets per event must be tagged one as τ and other different from the one tagged as τ, must be tagged as b. These candidates for τ and b also had to pass kinematics considerations due to the detector's capacity to observe them in agreement with the sections 2.3.3 and 2.3.4. These requirements suppose the baseline requirements for the interesting events.

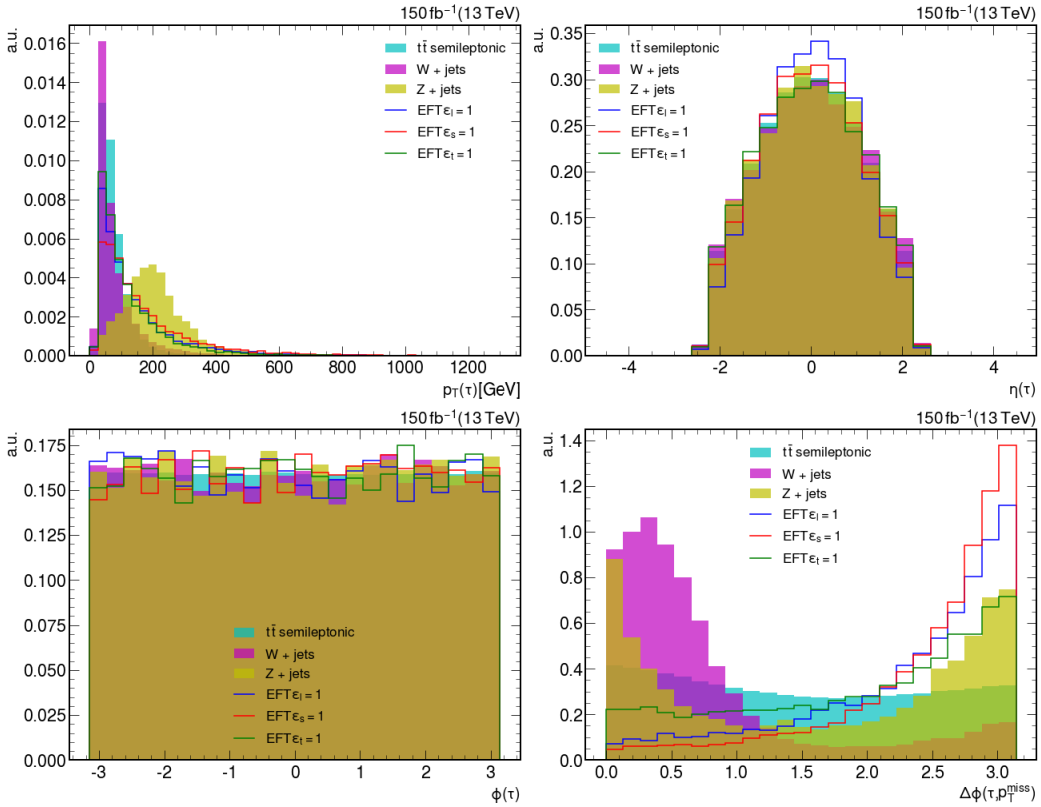


Figure 4.14: *τ kinematic information for the EFT. The upper panels are the p_T and the η from left to right, the lower are the azimuthal angle ϕ and $|\Delta\phi(\tau, \boldsymbol{p}_T^{miss})|$ in the same order. As this histogram shows the distribution of an object in the final state, the backgrounds are not stacked and present a transparency in order to understand the difference between each backgrounds and the signals.*

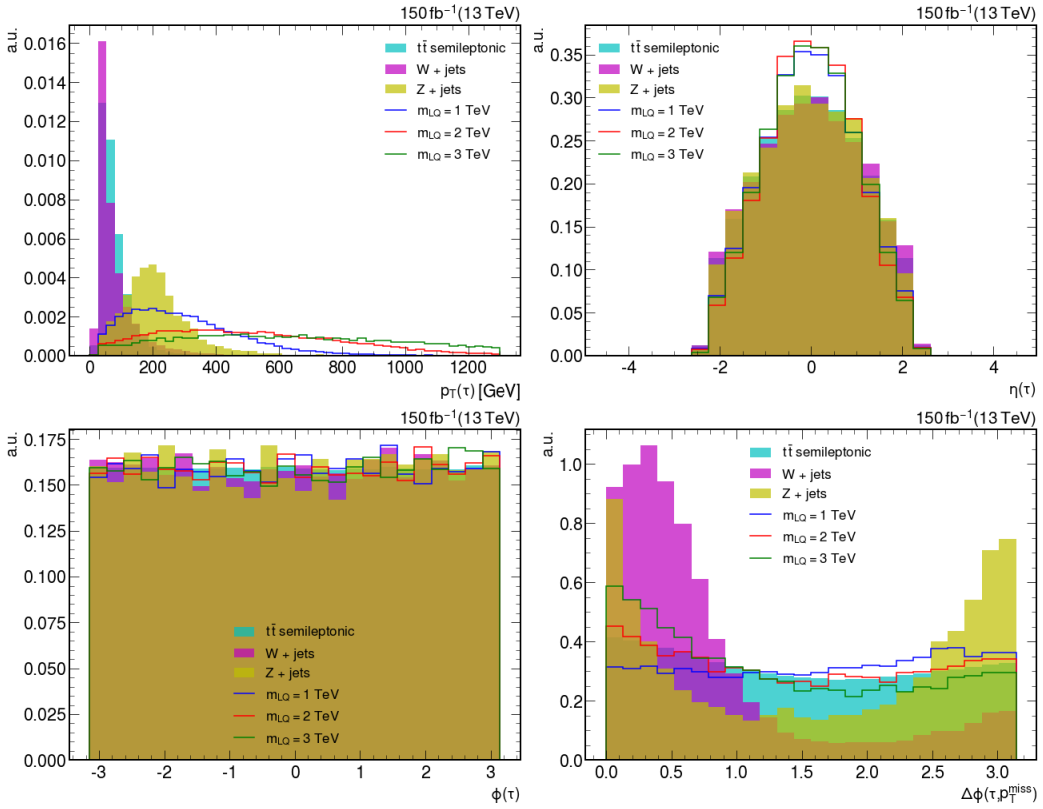


Figure 4.15: *τ kinematic information for the LQ. The upper panels are the p_T and the η from left to right, the lower are the azimuthal angle ϕ and $|\Delta\phi(\tau, \boldsymbol{p}_T^{miss})|$ in the same order.*

4.5 Selection Criteria

First, a baseline selection is required. As previous stated, a τ and a $b-jet$ is required in the final state, but more restrictively, the number of $b-jets$ are restricted to 1, this with the purpose of minimize the contribution of the $t\bar{t}$ background. Also a veto over a second τ is required in this baseline selection. This restrictions were imposed as in [70]. The specific information can be found in the Table 4.5.

We studied the behavior of the τ and the $\mathbf{p}_T^{\text{miss}}$ that are two out of three objects that for the three models carry information from the interaction related with the new physics. The behavior of the kinematic variables of the τ are shown in the Figures 4.13, 4.14 and 4.15. We can say that for SSM, LQ and EFT the $p_T(\tau)$ and $\Delta\phi(\tau, p_T^{miss})$ marks the difference between

Criterion	Selection
N_τ	≥ 1
$\lvert\eta_\tau\rvert$	≤ 2.3
$2^{\text{nd}} - \tau$	$p_T > 50\,\text{GeV}$ & $\lvert\eta\rvert < 2.3$
$N_{e/\mu}$ ó $p_T(e/\mu) > 15\,\text{GeV}$	$= 0$
$N_{\text{b-jets}}$	$= 1$
$p_T(b)$	$> 20\,\text{GeV}$
$\lvert\eta_{\text{b-jets}}\rvert$	< 2.5

Table 4.5: *Baseline selection, developed to minimize the $t\bar{t}$ contribution.*

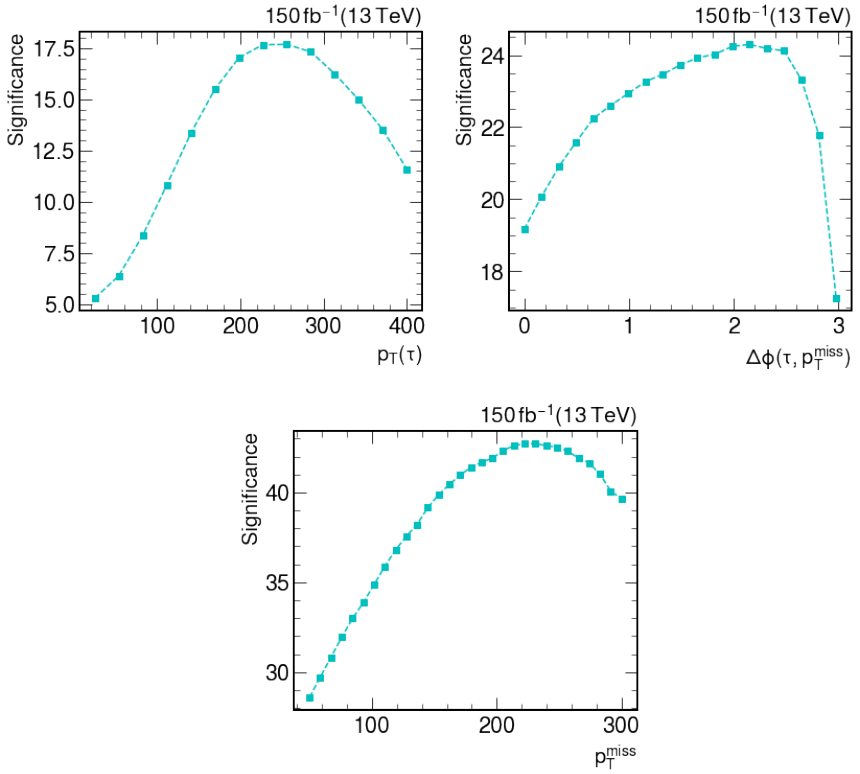


Figure 4.16: *Significance versus the 3 variables. The upper left panel is Z vs $p_T(\tau)$, the upper right-one is Z vs $|\Delta\phi(\tau, \boldsymbol{p}_T^{miss})|$. The lower panel shows Z vs p_T^{miss}. The 3 plots for the SSM where the signal is the one with $m_{W'} = 1$ TeV*

signals and backgrounds. So if we want to maximize Z, then we can calculate the number of events for signal and backgrounds after cutting in a certain value of these two variables and then increasing the value of the variable in a small number and repeating the process. The $\mathbf{p}_T^{\text{miss}}$ was also considered. We wanted to implement the same cuts for each model, independently of the value of the mass or the type of coupling, so we chose as signals for the Z maximization the SSM with $m_{W'} = 1\,\text{TeV}$, the EFT with $\epsilon_T^{cb} = 1$ and LQ with $m_{LQ} = 1\,\text{TeV}$. The maximum value for the statistical significance is searched the same way for the 3 models. First, varying the $p_T(\tau)$. Once a maximum is reached, only events with transverse momentum above the found value are considered, and then the process is repeated for the $|\Delta\phi(\tau, \mathbf{p}_T^{\text{miss}})|$ and once again for the p_T^{miss}.

The values that maximize statistical significance Z, are presented in the table 4.6.

Parameter	SSM	EFT	U1_LQ
$p_{\text{T}}(\tau) >$	250 GeV	200 GeV	300 GeV
$\lvert\Delta\phi(\tau, p_{\text{T}}^{\text{miss}})\rvert >$	1.5	2.0	1.0
$p_{\text{T}}^{\text{miss}} >$	200 GeV	300 GeV	400 GeV

Table 4.6: *Selection criteria for each type model for the three different models,* W′, LQ *and EFT, considered in the analysis. Thresholds have been selected based on best signal significance.*

This selection criteria are applied for signals and backgrounds, separating by model, i.e. all the SSM have the same criteria, different from the selection criteria applied to the EFT, and on with the LQ. The resultant events after the baseline selections and after applying the selection

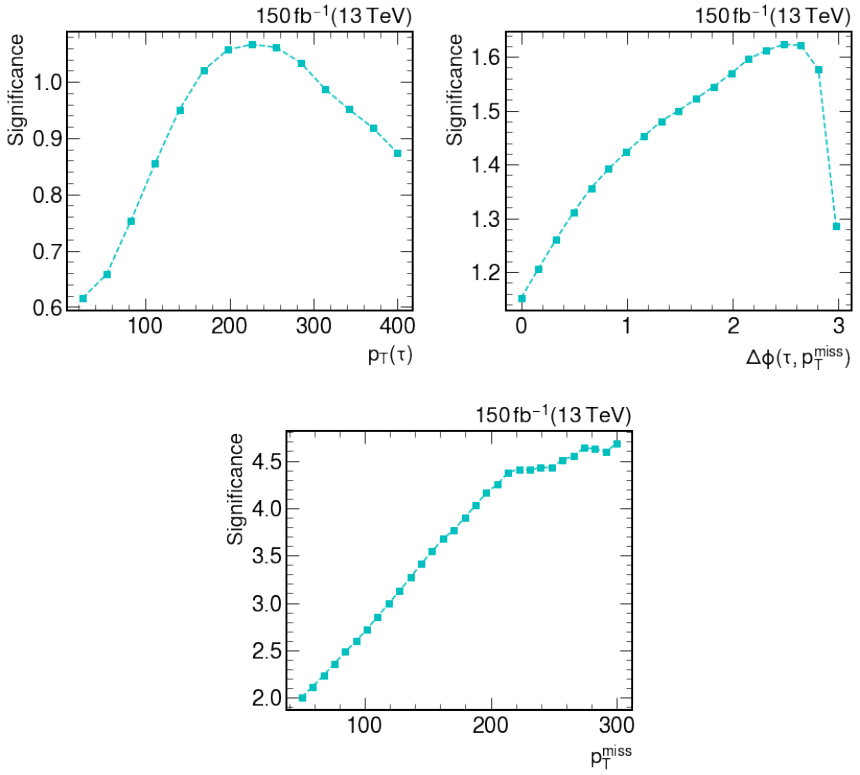


Figure 4.17: *Significance versus the 3 variables. The upper left panel is Z vs $p_T(\tau)$, the upper right-one is Z vs $|\Delta\phi(\tau, \boldsymbol{p}_T^{miss})|$. The lower panel shows Z vs p_T^{miss}. The 3 plots for the EFT where the signal is the one with $\epsilon_S^{cb} = 1$*

criteria are presented in the Table 4.7. Each type of selection present 2 types of cuts, the baseline selection and the ones from the selection criteria.

After applying these selection criteria we explored the 3 different mass spectra discussed in the section 4.2.

As the objects have been already selected, and the events that satisfy the different selection criteria (applied to the different models), now the histograms for the different signals and backgrounds can be produced. For this section it will be shown the 3 types of mass spectra for each model for the baseline selection and after applying the selection criteria.

4.6 Sequential Standard Model

4.6.1 Baseline Selection

The baseline selection in all the cases were composed by the consideration exclusive of the events with at least one jet to be candidate to be a τ, and only one different jet to tagged as b. Those jets must satisfy that $p_T(\tau) > 25\,\text{GeV}$, $\eta(\tau) \leq 2.4$, $p_T(b) > 30\,\text{GeV}$, $\eta(b) \leq 2.5$. The results after this selection are the Figures 4.19, 4.20 and 4.21.

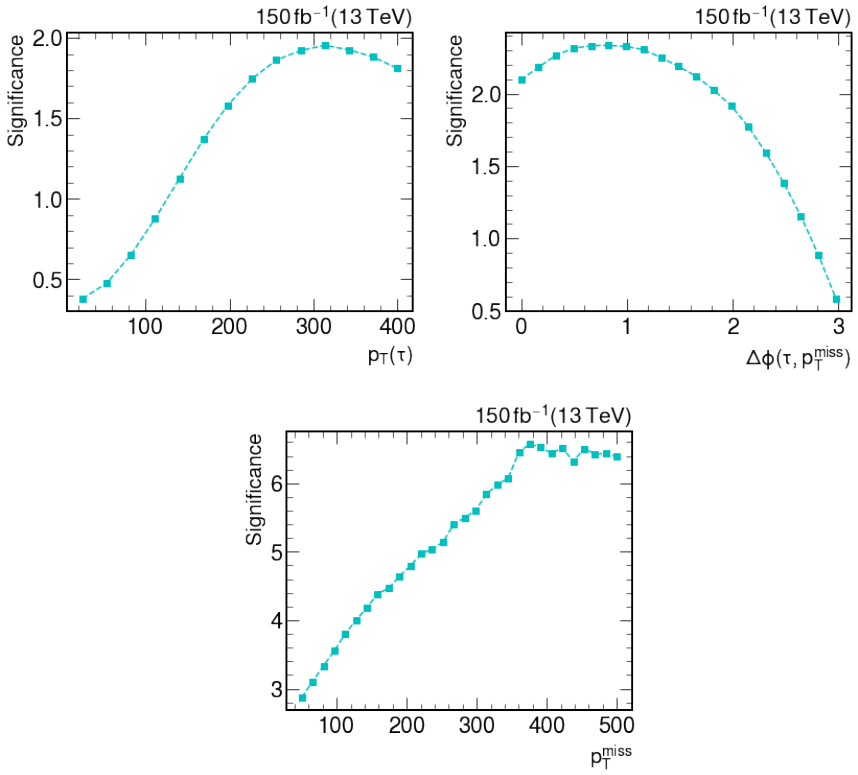


Figure 4.18: *Significance versus the 3 variables. The upper left panel is $Z\,vs\,p_T(\tau)$, the upper right-one is $Z\,vs\,p_T^{miss}$. The lower panel shows $Z\,vs\,|\Delta\phi(\tau, \boldsymbol{p}_T^{miss})|$. The 3 plots for the LQ where the signal is the one with $m_{LQ} = 1$ TeV*

	baseline	Selection SSM	Selection EFT	Selection LQ
$m_{W'} = 0.6$ TeV	8.55×10^4	6.57×10^3		
$m_{W'} = 1.0$ TeV	7.78×10^3	2.461×10^3		
$m_{W'} = 1.6$ TeV	5.09×10^2	2.48×10^2		
EFT $\epsilon_l = 1$	1.52×10^3		1.94×10^2	
EFT $\epsilon_s = 1$	9.09×10^2		1.81×10^2	
EFT $\epsilon_t = 1$	1.09×10^4		8.41×10^2	
$m_{LQ} = 1$ TeV	5.56×10^2			1.24×10^2
$m_{LQ} = 2$ TeV	6.71			2.48
$m_{LQ} = 3$ TeV	1.82×10^{-1}			6.86×10^{-2}
$t\bar{t}$	1.80×10^6	1.11×10^3	1.45×10^3	3.11×10^2
$W +$ jets	3.48×10^3	2.86×10^1	4.26×10^1	8.80
$Z +$ jets	7.91×10^1	2.79	3.69	7.16×10^{-1}

Table 4.7: *Table with the expected number of events per sample with the baseline selection from Table 4.5 and the selection criteria from Table 4.6.*

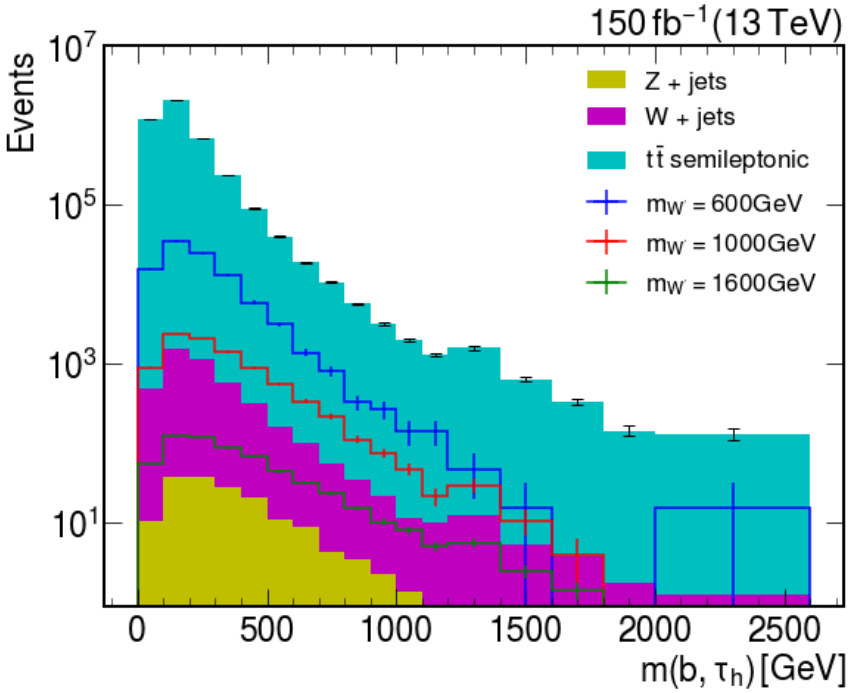


Figure 4.19: *Invariant mass between the b and the τ_h for the SSM after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

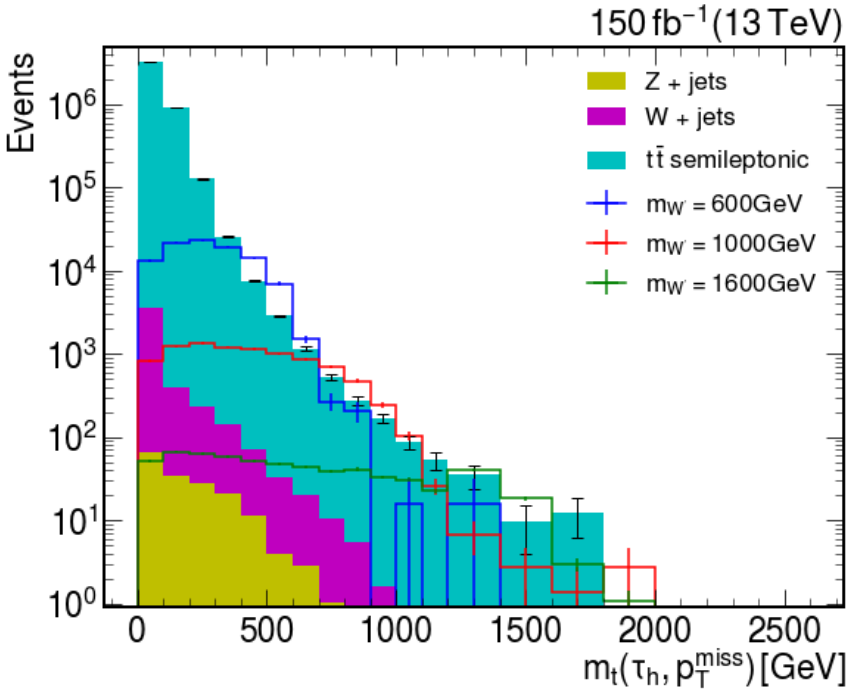


Figure 4.20: *Transverse mass of the τ_h for the SSM after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

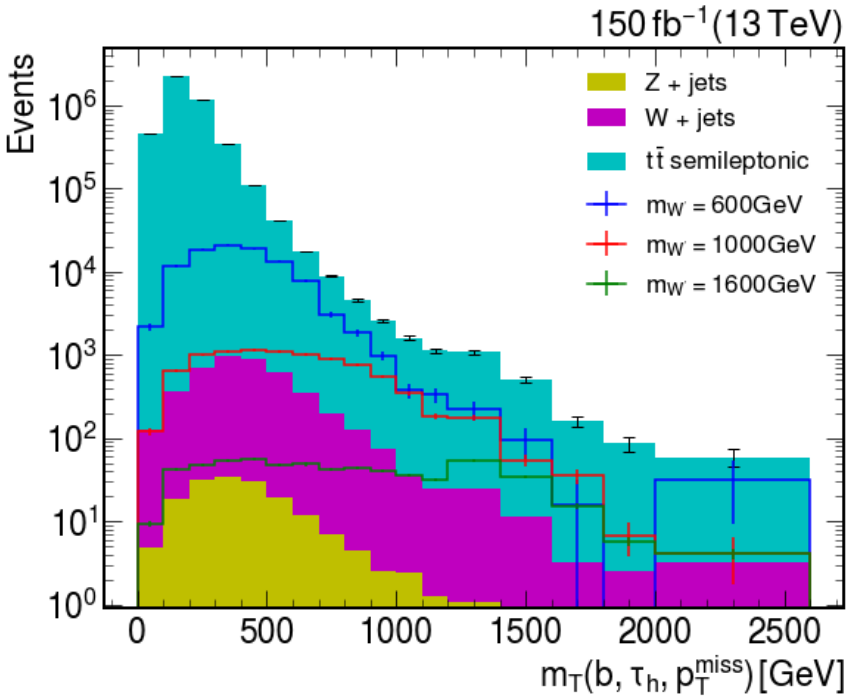


Figure 4.21: *Total mass between the b and the τ_h and $\boldsymbol{p}_T^{miss}$ for the SSM after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

4.6.2 Significance Selection Criteria

The statistical significance was maximized for the sample of $m_{W'} = 1\,\text{TeV}$ the objective was to observe the change in these histograms based on the selection criteria specified. These changes are presented in the Figures 4.22, 4.23 and 4.24

4.7 Effective Field Theory

4.7.1 Baseline Selection

The baseline selection is once again the same that was specified in the table 4.5, the kinematic selection criteria is stored in the table 4.6 and the number of events after the whole selection criteria are in the table 4.7. Its important by this point to remember that the 3 different signals were made with 3 different couplings, a scalar-one, a vector-one and a tensor-one. The mass spectra in this time are in the Figures 4.25, 4.26 and 4.27.

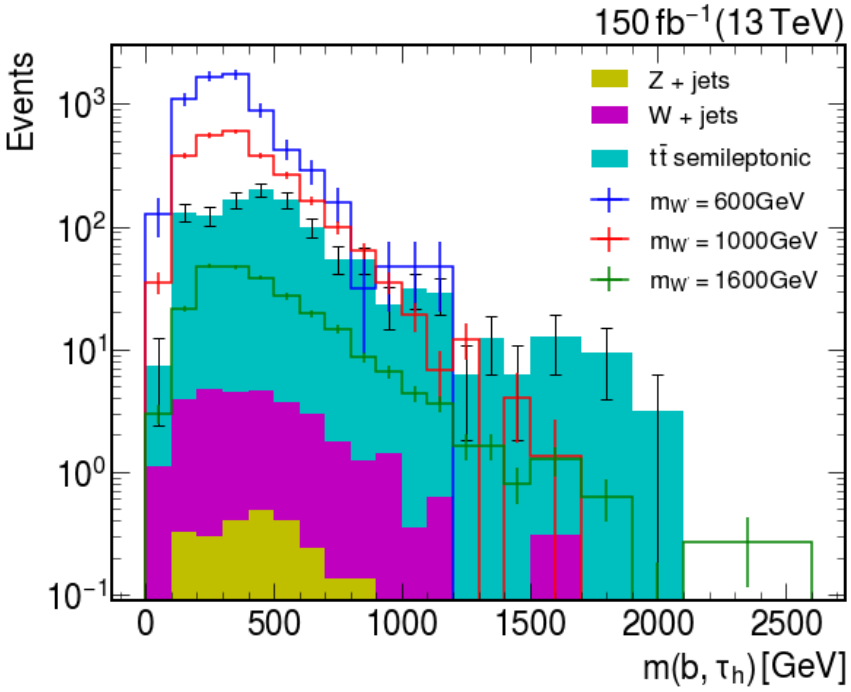


Figure 4.22: *Invariant mass between the b and the τ_h for the SSM after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

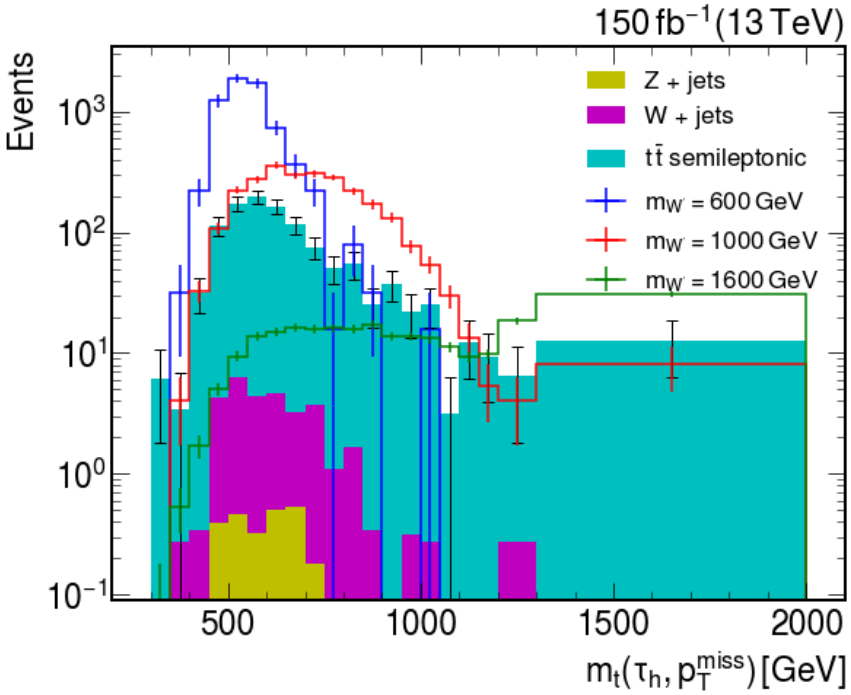


Figure 4.23: *Transverse mass of the τ_h for the SSM after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

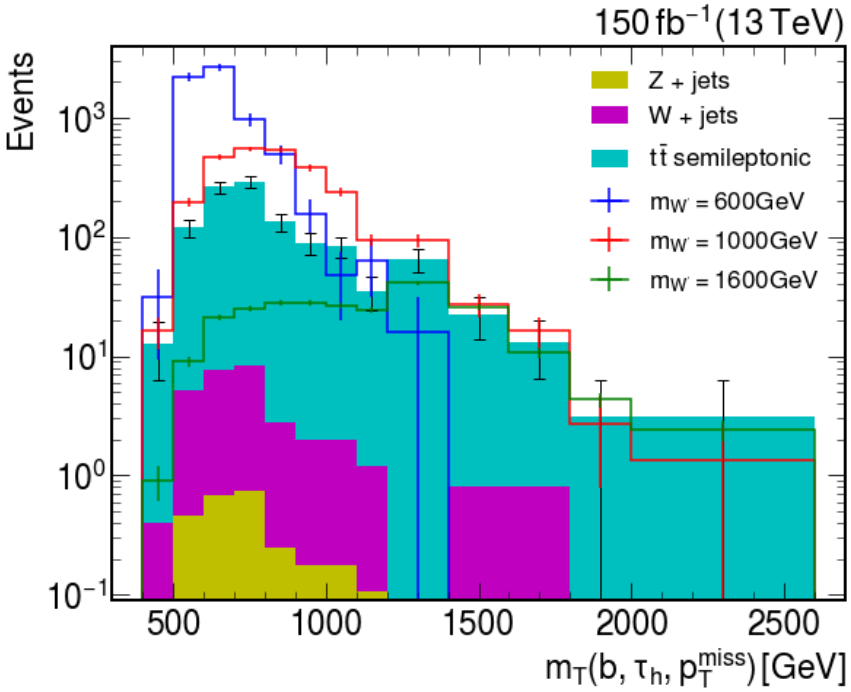


Figure 4.24: *Total mass between the b and the τ_h and $\boldsymbol{p}_T^{miss}$ for the SSM after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

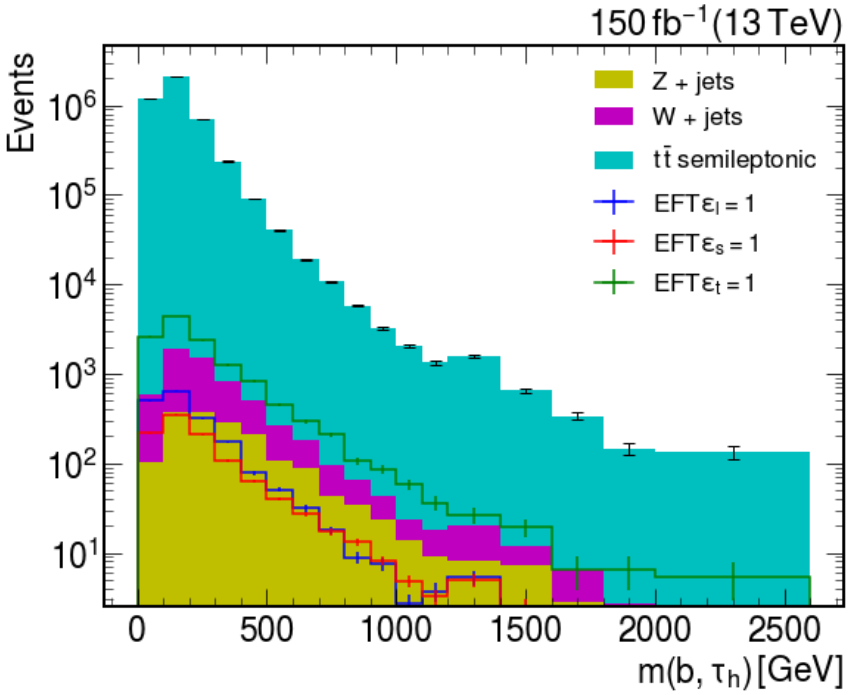


Figure 4.25: *Invariant mass between the b and the τ_h for the EFT after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

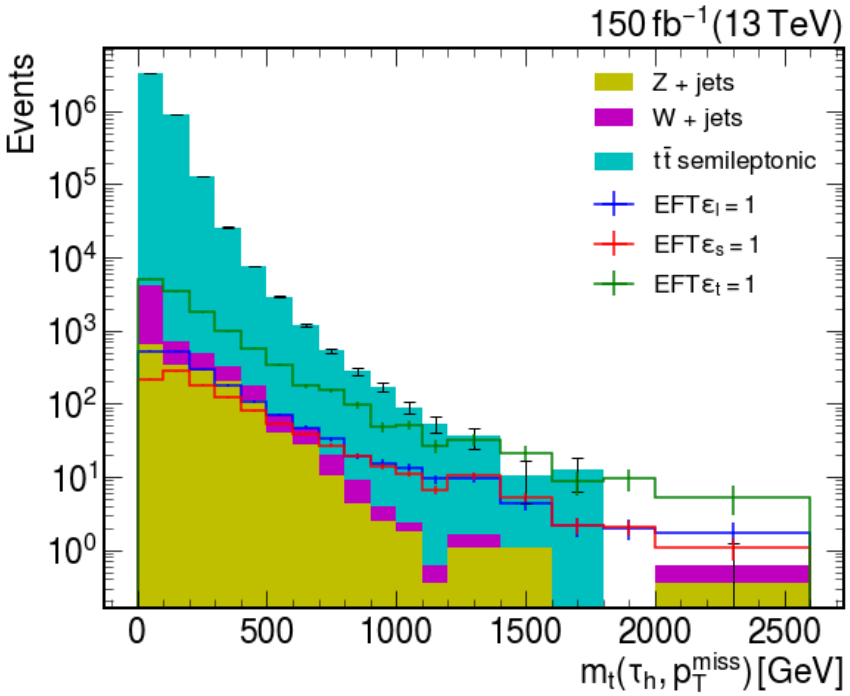


Figure 4.26: *Transverse mass of the τ_h for the EFT after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

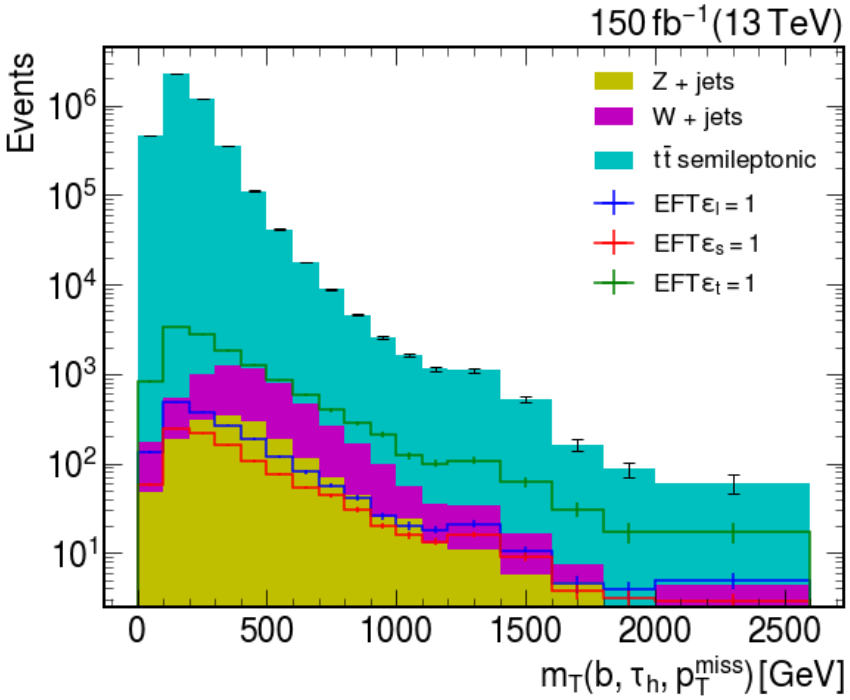


Figure 4.27: *Total mass between the b and the τ_h and $\boldsymbol{p}_T^{miss}$ for the EFT after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

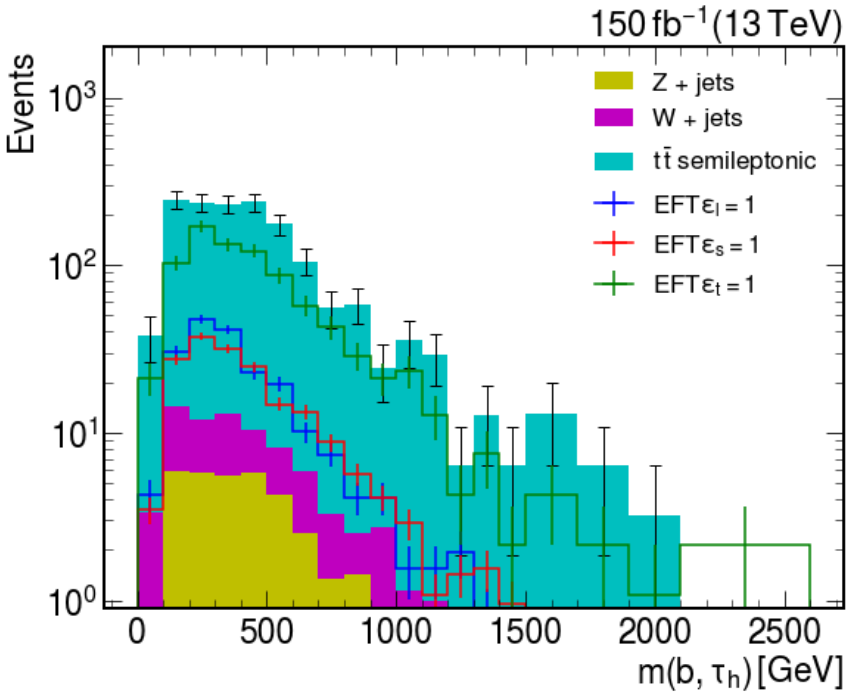


Figure 4.28: *Invariant mass between the b and the τ_h for the EFT after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

4.7.2 Significance Selection Criteria

Once the statistical significance was maximized for the sample of $\epsilon_S^{cb} = 1$. The objective was to observe was to observe the change in these histograms. These changes are presented in the Figures 4.28, 4.29 and 4.30

4.8 Leptoquark

4.8.1 Baseline Selection

The baseline selection is once again the same that was specified in the table 4.5, the selection criteria in table 4.6 and the number of events after the whole selection in table 4.7. Its important by this point to remember that the 3 different signals were made with 3 different couplings, a scalar-one, a left-one and a tensor-one. The mass spectra in this time are in the Figures 4.31, 4.32 and 4.33.

4.8.2 Significance Selection Criteria

Once the statistical significance was maximized for the sample of $m_{LQ} = 1\,\text{TeV}$. The objective was to observe was to observe the change in these histograms. These changes are presented in the Figures 4.34, 4.35 and 4.36

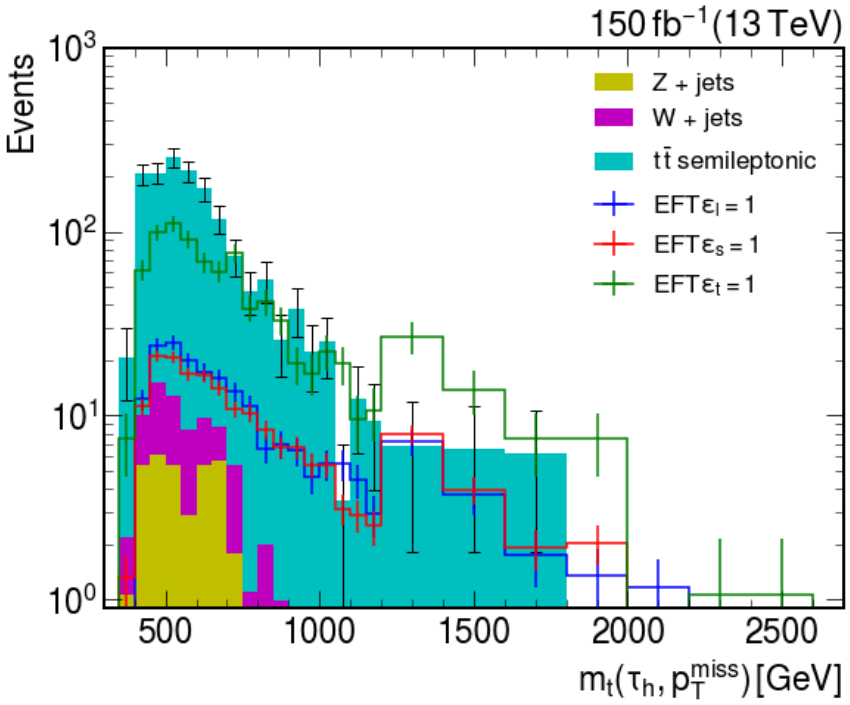


Figure 4.29: *Transverse mass of the τ_h for the EFT after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

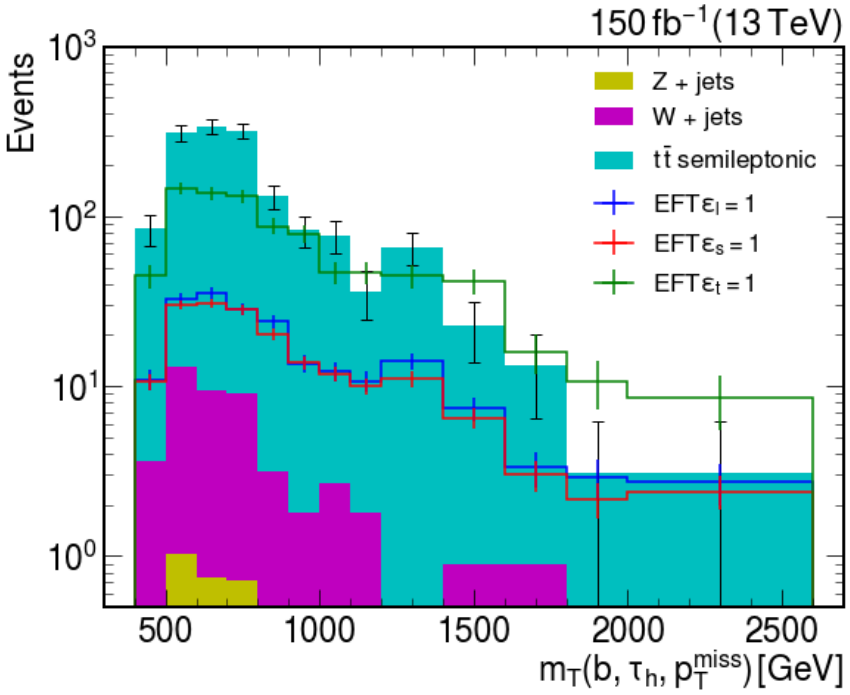
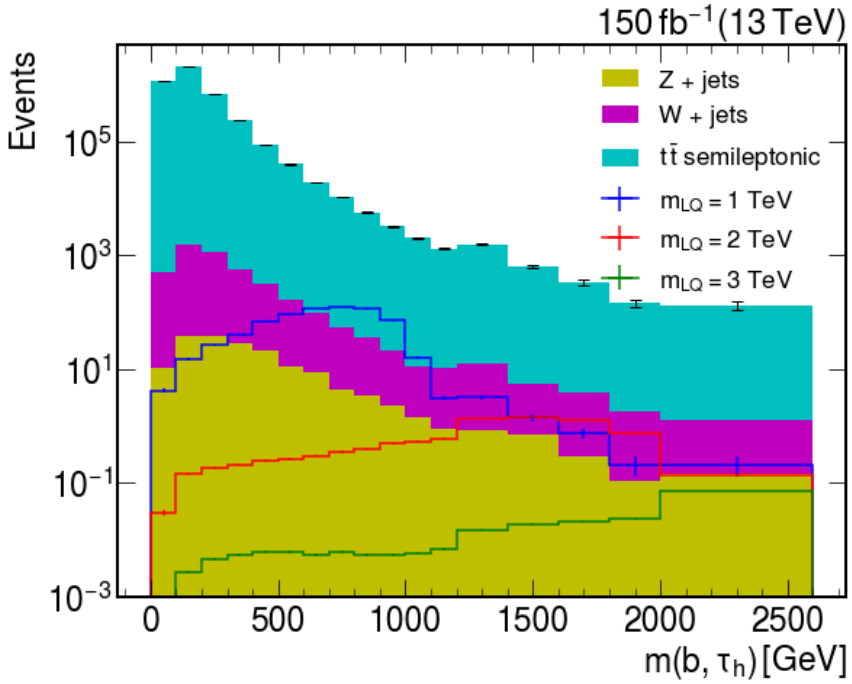


Figure 4.30: *Total mass between the b and the τ_h and $\boldsymbol{p}_T^{miss}$ for the EFT after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*



|1

Figure 4.31: *Invariant mass between the b and the τ_h for the LQ models after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

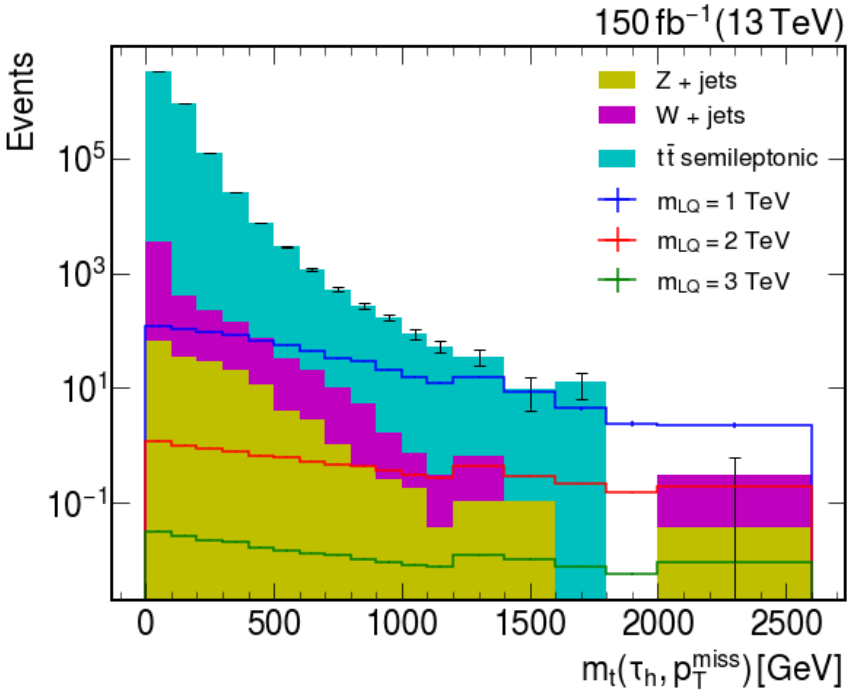


Figure 4.32: *Transverse mass of the τ_h for the LQ models after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

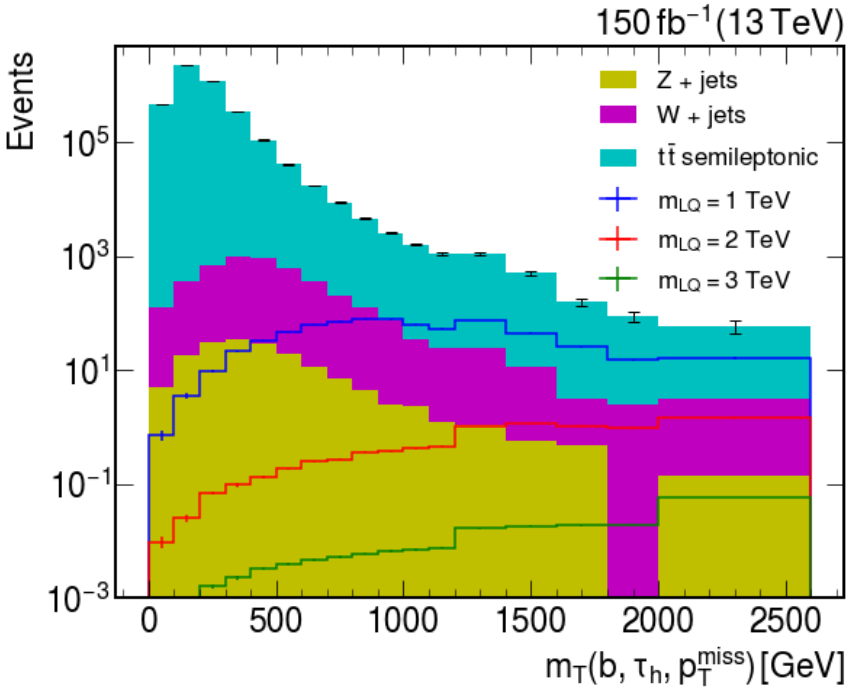


Figure 4.33: *Total mass between the b and the τ_h and $\boldsymbol{p}_T^{miss}$ for the LQ model after the baseline selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

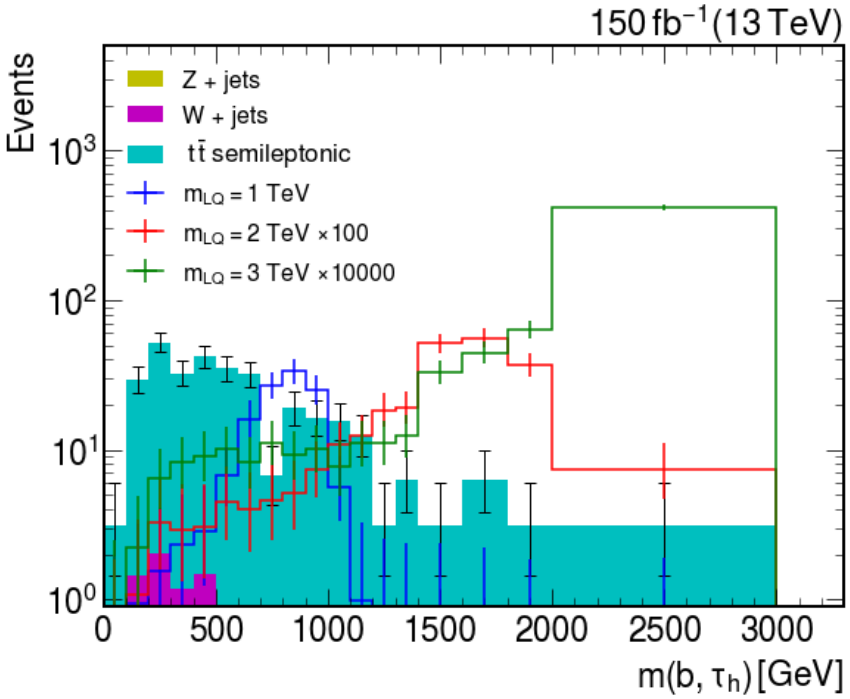


Figure 4.34: *Invariant mass between the b and the τ_h for the LQ model after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

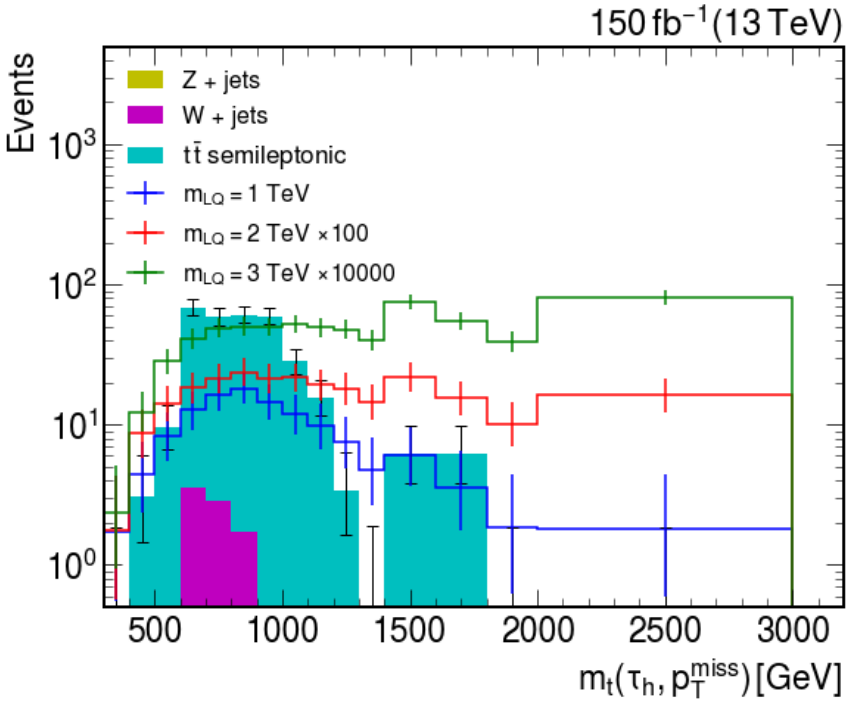


Figure 4.35: *Transverse mass of the τ_h for the LQ model after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

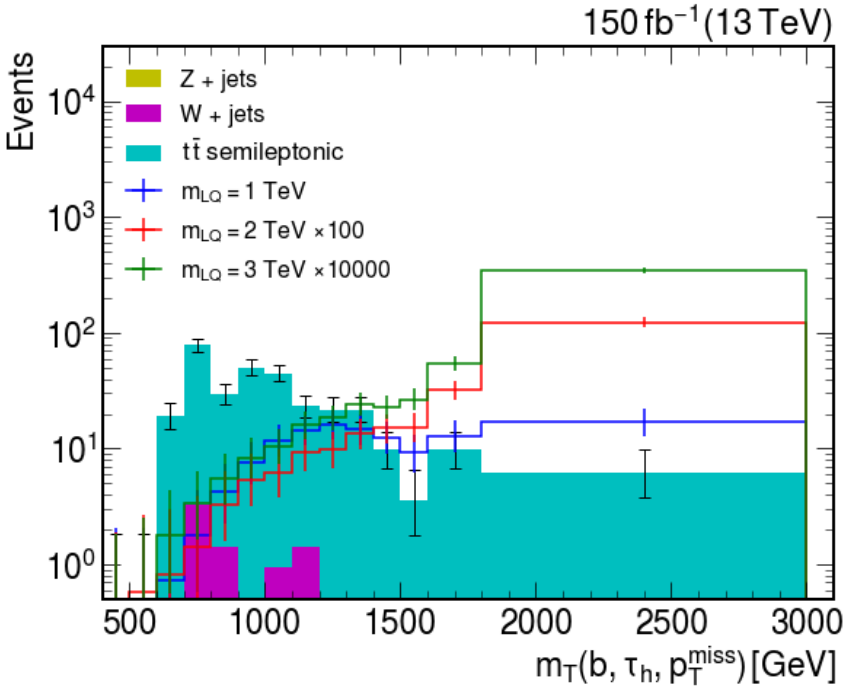


Figure 4.36: *Total mass between the b and the τ_h and $\boldsymbol{p}_T^{miss}$ for the LQ model after the statistical significance selection criteria. Each signal presents its own error-bars, while the background presents the error-bars for the contribution of the 3 backgrounds per bin.*

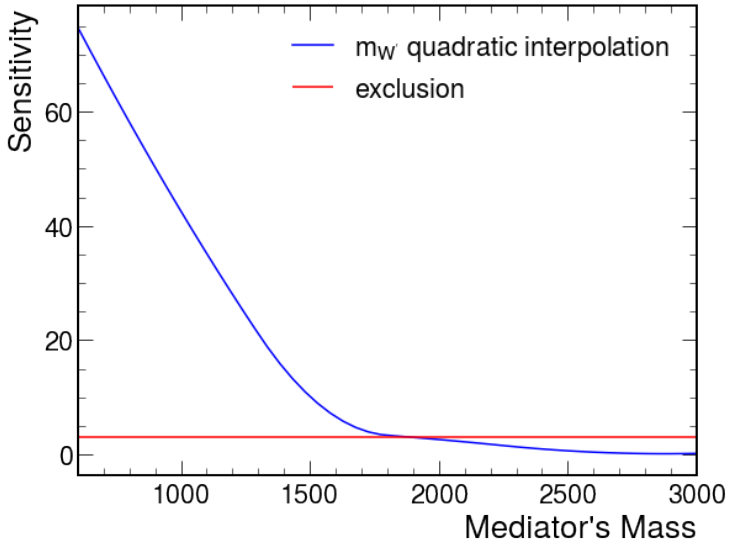


Figure 4.37: *Significance curve for the Sequential Standard Model. Masses simulated* $m_{W'} = [0.6, 1, 1.6, 2, 2.53]$ *TeV. The interpolation was made with a quadratic function, and the exclusion line is set in* 3σ.

4.9 Exclusion

So far, the mass histograms were presented the results for each model after a baseline selection criteria, and a kinematical selection criteria that also included that first baseline. These results were obtained by montecarlo simulations of pp collisions at a $\sqrt{s} = 13\,\text{TeV}$ at a $L = 150\,\text{fb}^{-1}$. In other words, in a LHC context, taking into account the CMS objects description.

In this section these results will be discussed.

4.9.1 SSM

In the SSM histograms showed in the section 4.6 correspond to the 3 mass spectra discussed in the section 3.2 and were conceived for 3 different masses for the W' boson, 600 GeV , 1 TeV and 1.6 TeV. In order to obtain the exclusion limits, masses for 2 TeV , 2.5 TeV and 3 TeV were also simulated and processed with the same baseline and the same selection criteria discussed for the SSM signals in the chapter 4. With the significance Z achieved for these W' an interpolation was made in order to obtain the significance curve between the lowest and the highest mass simulated.

The results shown in the Figure 4.37 show a exclusion for the W' in the final state $b, \tau_H, p_T^{\text{miss}}$ for masses below 1.9 TeV with a statistical significance of 3σ, while masses up to 1800 GeV could potentially be discovered with 5σ signal significance above the background expectation. Masses up to 4 TeV were excluded in [59], but this analysis was made specifically for the final state τ_H, p_T^{miss}. So this results represents a motivation for searches in collider experiments in order to exclude this masses in this specific final state.

4.9.2 EFT

The effective field theory represents a way to study the type of the mediator without compromising with a specific scale of mass, specially if that scale of mass has a low probability to

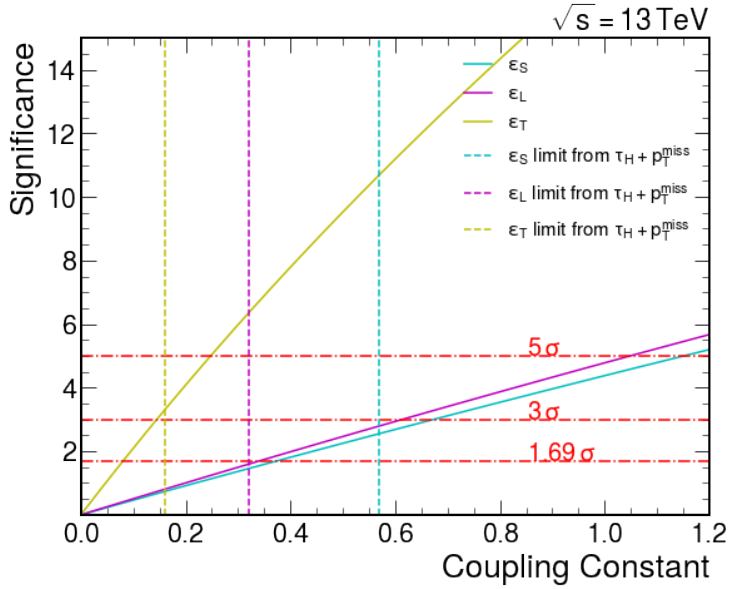


Figure 4.38: *Significance curve for the Effective Field Theory. Types of couples simulated were* $\epsilon_S, \epsilon_L, \epsilon_T$*. The vertical lines represent the exclusion limits for [20]. The horizontal lines are the exclusion values of* $1.69\sigma, 3\sigma$ *and* 5σ*.*

be produced in the LHC. Previous studies were made considering B decays and τ's in the final state as [20]. Particularly this study predicted certain exclusion limits for the EFT. Our and their results are presented in the Figure 4.38, with the vertical lines being their exclusion limits and our curves being the statistical significance for each type of couple, for each value of that couple. As it can be seen in Figure 4.38, this work is more sensitive in the tensor coupling than the previous studies. For this coupling, this search sets the limit for the tensor coupling in 0.13 for exclusion at 3σ and 1.69σ the limit is set on 0.07. Considering only τ_H, p_T^{miss} the limit for exclusion were set on 2σ at a value for the tensor coupling on 0.10 which shows the sensitivity of this work.

4.9.3 Leptoquark

The Leptoquark as it was mentioned in the chapter 1, is a theoretical proposal that couples the leptons to the quarks directly. In this work was considered a Leptoquark that couples the *up*-type quarks with the neutral leptons, and the *down*-type quarks with the charged leptons. Six different samples of this model were simulated, $m_{LQ} = [0.5, 1, 1.25, 1.5, 2, 3]$ TeV. And the statistical significance reached for those masses involved in the process we are interested are plotted in the Figure 4.39

From the Figures 4.37, 4.38 and 4.39 can be seen that for the SSM W' masses below the 1.8 TeV could be excluded at a statistical significance of 3σ. In the case of the SSM W', previous restrictions were imposed in [59] for masses below 4 TeV. Nevertheless that research was made for a final state of τ, ν. The closest final state considered in the literature for LQ searches, study final states with two τ_H and one b-jet candidates. The exclusion limit achieved by that search sets a limit on the LQ mass, which should be greater than 1.1 TeV [71]. Therefore, our proposal reaches higher and more stringent limit for the LQ mass and also explores a new channel relevant

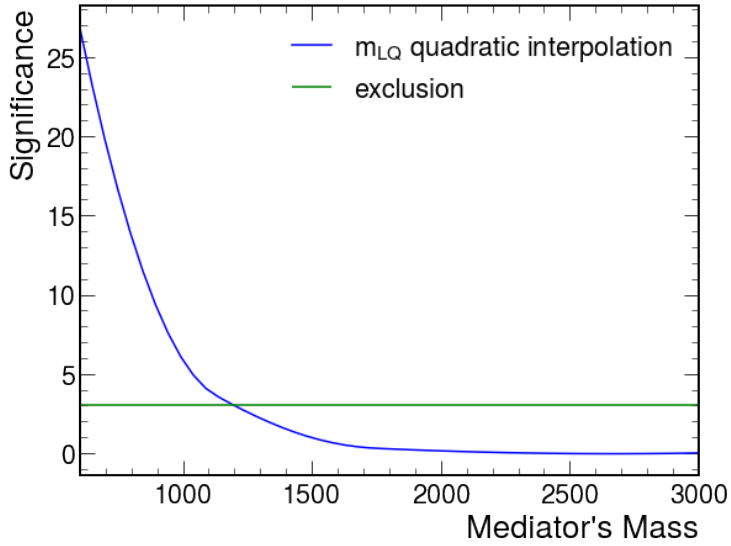


Figure 4.39: *Significance curve for the Leptoquark Model. Masses simulated* $m_{LQ} = [0.5, 1, 1.25, 1.5, 2, 3]$ *TeV. The interpolation was made with a quadratic function, and the exclusion line is set in* 3σ.

to study B meson anomalies. The gain in experimental sensitivity is associated with more optimal selection criteria, specially when considering only one τ_h candidate instead of two.

Chapter 5

Conclusions

So far, motivated by the $R_{D^{(*)}}$ anomaly, a final state of $b-\tau-\nu$ in pp collisions was considered. MC samples for signals of new physics and background processes were produced and analyzed, taking into account previous studies. The analysis led in a first place to a selection criteria that includes different cuts for the MC events, and this selection led to results in the form of mass spectra. And finally these results allow us to improve the sensitivity to relevant new physics parameter space for 3 models considered.

Run 3 in the LHC opens different possibilities for searches of new physics. The integrated luminosity is expected to increase up to $300\,\text{fb}^{-1}$ and $\sqrt{s} = 13.6\,\text{TeV}$. The results from chapter 4 could be reproduced for those conditions. Also this research could be extended to an experimental context for experiments like CMS or ATLAS, in order to probe these masses and couplings in this final state.

Emulation is a process different to a simulation in the sense that the emulation process copies the response from the detector to the different objects that travel through it. In the case of this study, the response of the detector was emulated by Delphes. This same response could be simulated by Geant4 to reproduce the response of the different parts of and experiment as the CMS experiment while the particles travel and interact in it.

The process of selecting the best values for cuts in the $p_T(\tau)$, $\Delta\phi(\tau, p_T^{\text{miss}})$ and p_T^{miss} was done by first comparing the behavior of the kinematic variables and then selecting the variables that seem to differ from signals and backgrounds. This process can be done with recent methods adapted from machine learning (ML) and deep learning (DL) in order to see if a certain ML or DL model could select the same variables, optimize the statistical significance and to probe if it reaches the same values for the cuts as it was done in this work. Also the process to obtain the kinematic values can also be implemented via a python module called 'Relaxed'. Basically, it implements a generalized process that applies all the cuts in all possible combinations of order in the selected variables [72].

Acknowledgements

I would like to thanks all the people that in some way have helped me with the elaboration of this thesis.

First, I would like to thanks to my mother Gladys, the only one who had always've been there for me, and pushed me to start this master degree. And in general for supporting me in every single dream I've ever had. To my brother and my father, both of them for being those type of people who cares about the wellness and happiness of their family. To my girlfriend Natalia, who gave me a north, and taught me that the eyes must stay in the prey not on the horizon and for accompanying me in this last year in which this thesis was not only written, but also in the production of the results. To my friends, Esteban, Mayi, Maya, and Florez, who were always there to bring me out a smile whenever I needed. To Daniel, Manuel and Alexis, my partners on board in this Graduated trip, who helped me in different instances, not just academically talking, but also when I needed to talk about frustration or personal difficulties. To my advisor José David, who had the patience to teach me and start working with me, even when I was not registered in the master's program, and for taking me into account for the different events and jobs. To the University of Antioquia, for being my home for these 10 years.

Medellín, 2022

Tomas Atehortua Garcés

Bibliography

[1] J. Beringer *et al.*, "Particle data group," *Phys. Rev. D*, vol. 86, no. 010001, 2012.

[2] E. R. Scerri, "The periodic table," in *Philosophy of chemistry*. Elsevier, 2012, pp. 329–338.

[3] J. J. Thomson, *The structure of the atom*. Academie Royale de Belgique, 1913.

[4] E. Rutherford, "Lxxix. the scattering of α and β particles by matter and the structure of the atom," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 21, no. 125, pp. 669–688, 1911.

[5] J. Chadwick, "Possible existence of a neutron," *Nature*, vol. 129, no. 3252, pp. 312–312, 1932.

[6] B. Thaller, *The dirac equation*. Springer Science & Business Media, 2013.

[7] R. P. Feynman, "Space-time approach to quantum electrodynamics," *Physical Review*, vol. 76, no. 6, p. 769, 1949.

[8] J. Schwinger, "Quantum electrodynamics. i. a covariant formulation," *Physical Review*, vol. 74, no. 10, p. 1439, 1948.

[9] C.-N. Yang and R. L. Mills, "Conservation of isotopic spin and isotopic gauge invariance," *Physical review*, vol. 96, no. 1, p. 191, 1954.

[10] W. Heisenberg, "Über den bau der atomkerne. ii," in *Original Scientific Papers/Wissenschaftliche Originalarbeiten*. Springer, 1989, pp. 208–216.

[11] M. Gell-Mann, "The eightfold way: A theory of strong interaction symmetry," California Inst. of Tech., Pasadena. Synchrotron Lab., Tech. Rep., 1961.

[12] M. Gell-Mann, "A schematic model of baryons and mesons," *Physics Letters*, vol. 8, no. 3, 1964.

[13] D. Griffiths, *Introduction to elementary particles*. John Wiley & Sons, 2020.

[14] L. Di Lella and C. Rubbia, "The discovery of the W and Z particles," in *60 Years of CERN Experiments and Discoveries*. World Scientific, 2015, pp. 137–163.

[15] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Physical Review Letters*, vol. 13, no. 16, p. 508, 1964.

[16] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Physical Review Letters*, vol. 13, no. 9, p. 321, 1964.

[17] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, and et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, p. 3061, Sep 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2012.08.021

[18] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al., "Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, p. 129, Sep 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2012.08.020

[19] M. Peskin, *An introduction to quantum field theory*. CRC press, 2018.

[20] A. Greljo, J. M. Camalich, and J. D. Ruiz-Álvarez, "Mono-τ signatures at the LHC constrain explanations of B-decay anomalies," *Physical review letters*, vol. 122, no. 13, p. 131803, 2019.

[21] D. Bardhan, P. Byakti, and D. Ghosh, "A closer look at the R_D and R_{D^*} anomalies," *Journal of High Energy Physics*, vol. 2017, no. 1, pp. 1–37, 2017.

[22] F. Capozzi, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, "Neutrino masses and mixings: Status of known and unknown 3ν parameters," *Nuclear Physics B*, vol. 908, pp. 218–234, 2016.

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

[24] K. Garrett and G. Duda, "Dark matter: A primer," *Advances in Astronomy*, vol. 2011, 2011.

[25] V. Barger, W.-Y. Keung, and E. Ma, "Sequential W and Z bosons," *Physics Letters B*, vol. 94, no. 3, pp. 377–380, 1980.

[26] J. Y. Araz, M. Frank, B. Fuks, S. Moretti, and Özer Özdal, "Cross-fertilising extra gauge boson searches at the LHC," 2021.

[27] A. Grozin, "Effective field theories," 2020.

[28] R. Alonso, B. Grinstein, and J. M. Camalich, "Lepton universality violation with lepton flavor conservation in b-meson decays," *Journal of High Energy Physics*, vol. 2015, no. 10, pp. 1–30, 2015.

[29] M. Freytsis, Z. Ligeti, and J. T. Ruderman, "Flavor models for $B \to D^{(*)}\tau\nu$," *Physical Review D*, vol. 92, no. 5, p. 054018, 2015.

[30] D. Robinson, B. Shakya, and J. Zupan, "Right-handed neutrinos and R_{D^*}," *Journal of High Energy Physics*, vol. 2019, no. 2, pp. 1–32, 2019.

[31] Y. Sakaki, R. Watanabe, M. Tanaka, and A. Tayduganov, "Testing leptoquark models in $B \to D^{(*)}\tau\nu$," *Physical Review D*, vol. 88, no. 9, p. 094012, 2013.

[32] CERN, "Cern facts and faqs," 2022 [Online]. [Online]. Available: https://home.cern/resources/faqs

[33] R. Bruce, B. Salvant, T. Pieloni, K. S. B. Li, E. Metral, J. Jowett, R. Tomas Garcia, D. Mirarchi, G. Rumolo, S. Redaelli *et al.*, "LHC Run 2: Results and challenges," Tech. Rep., 2016.

[34] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*, ser. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. [Online]. Available: https://cds.cern.ch/record/782076

[35] F. Carminati, P. Foka, P. Giubellino, A. Morsch, G. Paic, J. Revol, K. Safarik, Y. Schutz, U. A. Wiedemann, A. Collaboration *et al.*, "ALICE: Physics performance report, volume i," *Journal of Physics G: Nuclear and Particle Physics*, vol. 30, no. 11, p. 1517, 2004.

[36] A. A. Alves Jr, L. Andrade Filho, A. Barbosa, I. Bediaga, G. Cernicchiaro, G. Guerrer, H. Lima Jr, A. Machado, J. Magnin, F. Marujo *et al.*, "The LHCb detector at the lhc," *Journal of instrumentation*, vol. 3, no. 08, p. S08005, 2008.

[37] ATLAS Collaboration, "ATLAS detector and physics performance technical design report," *http://aslasinfo.cern. ch/Atlas/GROUPS/PHYSICS/TDR/access.html*, 1999.

[38] "CMS experiment web page," https://cmsexperiment.web.cern.ch/index.php/detector, accessed: 2021-09-30.

[39] D. Campi, B. Curé, A. Gaddi, H. Gerwig, A. Hervé, V. Klyukhin, G. Maire, G. Perinic, P. Brédy, P. Fazilleau *et al.*, "Commissioning of the CMS magnet," *IEEE transactions on applied superconductivity*, vol. 17, no. 2, pp. 1185–1190, 2007.

[40] W. Erdmann, "The CMS pixel detector," *International Journal of Modern Physics A*, vol. 25, no. 07, pp. 1315–1337, 2010.

[41] CMS Collaboration, "Performance and operation of the CMS electromagnetic calorimeter," *Journal of Instrumentation*, vol. 5, no. 03, p. T03010, 2010.

[42] C. collaboration *et al.*, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data," *Journal of Instrumentation*, vol. 5, no. 03, p. T03012, 2010.

[43] ——, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC," *Journal of Instrumentation*, vol. 8, no. 11, p. P11002, 2013.

[44] G. Bayatian, S. Chatrchyan, G. Hmayakyan, A. Sirunyan, W. Adam, T. Bergauer, M. Dragicevic, J. Eroe, M. Friedl, R. Fruehwirth *et al.*, "CMS physics: Technical design report," 2006.

[45] M. Della Negra, P. Jenni, and T. S. Virdee, "The construction of ATLAS and CMS," *Annual Review of Nuclear and Particle Science*, vol. 68, pp. 183–209, 2018.

[46] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," CERN, Geneva, Tech. Rep. CMS-PAS-PFT-09-001, Apr 2009. [Online]. Available: https://cds.cern.ch/record/1194487

[47] Y. Allkofer, C. Amsler, D. Bortoletto, V. Chiochia, L. Cremaldi, S. Cucciarelli, A. Dorokhov, C. Hörmann, R. Horisberger, D. Kim *et al.*, "Design and performance of the silicon sensors for the CMS barrel pixel detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 584, no. 1, pp. 25–41, 2008.

[48] L. Borrello, E. Focardi, A. Macchiolo, and A. Messineo, "Sensor design for the CMS silicon strip tracker," CERN-CMS-NOTE-2003-020, Tech. Rep., 2003.

[49] CMS collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *Journal of Instrumentation*, vol. 9, no. 10, p. P10009, 2014.

[50] J. Greensite, *An introduction to the confinement problem.* Springer, 2011, vol. 821.

[51] A. Heister, V. Konoplyanikov, C. Tully, S. Petrushanko, J. Rohlf, O. Kodolova, and A. Ulyanov, "Measurement of jets with the CMS detector at the LHC," CERN-CMS-NOTE-2006-036, Tech. Rep., 2006.

[52] G. Bagliesi, "Tau tagging at ATLAS and CMS," *arXiv preprint arXiv:0707.0928*, 2007.

[53] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Ero, C. Fabjan, M. Friedl *et al.*, "Performance of tau-lepton reconstruction and identification in CMS," *Journal of Instrumentation*, vol. 7, 2012.

[54] "Level 1 Tau trigger performance in 2016 data and VBF seeds at Level 1 trigger," Jul 2017. [Online]. Available: https://cds.cern.ch/record/2273268

[55] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. E. Del Valle *et al.*, "Performance of reconstruction and identification of tau leptons decaying to hadrons and nu (tau) in pp collisions at $\sqrt{s} = 13\,\text{TeV}$," 2018.

[56] CMS collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ tev," *arXiv preprint arXiv:1809.02816*, 2018.

[57] C. collaboration *et al.*, "Identification of b-quark jets with the cms experiment," *Journal of Instrumentation*, vol. 8, no. 04, p. P04013, 2013.

[58] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification," *Journal of High Energy Physics*, vol. 2014, no. 10, Oct 2014. [Online]. Available: http://dx.doi.org/10.1007/JHEP10(2014)059

[59] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. E. D. Valle, and et al., "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS

detector," *Journal of Instrumentation*, vol. 14, no. 07, p. P07004P07004, Jul 2019. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/14/07/P07004

[60] Y. Li and C.-D. Lu, "Recent anomalies in b physics," *arXiv preprint arXiv:1808.02990*, 2018.

[61] D. F. Geesaman and P. E. Reimer, "The sea of quarks and antiquarks in the nucleon," *Reports on Progress in Physics*, vol. 82, no. 4, p. 046301, Mar 2019. [Online]. Available: http://dx.doi.org/10.1088/1361-6633/ab05a7

[62] R. Placakyte, "Parton distribution functions," 2011.

[63] M. Aaboud, G. Aad, B. Abbott, B. Abeloos, S. Abidi, O. AbouZeid, N. Abraham, H. Abramowicz, H. Abreu, R. Abreu *et al.*, "Search for additional heavy neutral higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2018, no. 1, pp. 1–54, 2018.

[64] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "Madgraph 5: going beyond," *Journal of High Energy Physics*, vol. 2011, no. 6, pp. 1–40, 2011.

[65] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, "Ufo the universal feynrules output," *Computer Physics Communications*, vol. 183, no. 6, p. 12011214, Jun 2012. [Online]. Available: http://dx.doi.org/10.1016/j.cpc.2012.01.022

[66] R. Ciesielski and K. Goulianos, "Mbr monte carlo simulation in pythia8," *arXiv preprint arXiv:1205.1446*, 2012.

[67] S. Ovyn, X. Rouby, and V. Lemaitre, "Delphes, a framework for fast simulation of a generic collider experiment," 2010.

[68] J. Pivarski. Pyhep 2021: Uproot and awkward array tutorial. Youtube. [Online]. Available: https://www.youtube.com/watch?v=s47Nz0h0vcg

[69] G. Cowan, *Statistical data analysis*. Oxford university press, 1998.

[70] M. Abdullah, J. Calle, B. Dutta, A. Flórez, and D. Restrepo, "Probing a simplified W model of $R_{D^{(*)}}$ anomalies using b tags, τ leptons, and missing energy," *Physical Review D*, vol. 98, no. 5, p. 055016, 2018.

[71] "The search for a third-generation leptoquark coupling to a τ lepton and a b quark through single, pair and nonresonant production at $\sqrt{s} = 13$ TeV," CERN, Geneva, Tech. Rep., 2022. [Online]. Available: https://cds.cern.ch/record/2815309

[72] N. D. Simpson, "Analysis optimisation with differentiable programmingű," 2022. [Online]. Available: https://github.com/gradhep/relaxed

Abbreviations

ATLAS	A Toroidal LHC Apparatus
BSM	Beyond Standard Model
CERN	Conseil Européen pour la Recherche Nucléaire
CMS	Compact Muon Solenoid
EFT	Effective Field Theory
EW	Electroweak
LO	Leading Order
LHC	Large hadron collider
LFU	Lepton Flavour Universality
LEP	Large Electron-Positron Collider
LQ	Leptoquark
UFO	Universal FeynRules Output
MET	Missing Energy Transverse
NLO	Next to leading-order
PF	Particle Flow
PDF	Parton Distribution Function
SSM	Sequential Standard Model
SM	Standard model