Search for a vector-like quark T' decaying into top+Higgs in single production mode in full hadronic final state using CMS data collected at 8 TeV

Jose David Ruiz-Alvarez

To cite this version:

Jose David Ruiz-Alvarez. Search for a vector-like quark T' decaying into top+Higgs in single production mode in full hadronic final state using CMS data collected at 8 TeV . High Energy Physics - Experiment [hep-ex]. Université Claude Bernard Lyon 1, 2015. English. NNT : . tel-01328175

Numéro d'ordre 180-2015

Thèse
présentée devant
l'Université Claude Bernard – Lyon 1
École Doctorale de Physique et d'Astrophysique

pour l'obtention du

DIPLÔME DE DOCTORAT
Spécialité: Physique des Particules
(arrêté du 7 août 2006)

Search for a vector-like quark T' decaying into top+Higgs in single production mode in full hadronic final state using CMS data collected at 8 TeV

par
José David Ruiz-Álvarez

Sous la direction de:
Stéphanie Beauceron

Institut de Physique Nucléaire de Lyon (IPNL)

Soutenu publiquement le 21 octobre 2015
Jury

Patrizia Azzi	Rapportrice
Stéphanie Beauceron	Directrice de thèse
Sabine Crepé-Renaudin	Rapportrice
Aldo Deandrea	Examinateur
José Ocariz	Examinateur
Patrice Verdier	Président du jury

… when he came opposite Palodes, and there was neither wind nor wave, Thamus from the stern, looking toward the land, said the words as he had heard them: "Great Pan is dead." Even before he had finished there was a great cry of lamentation, not of one person, but of many, mingled with exclamations of amazement.

Plutarch, *De Defectu Oraculorum, 100-200*

After Buddha was dead, people showed his shadow for centuries afterwards in a cave,—an immense frightful shadow. God is dead: but as the human race is constituted, there will perhaps be caves for millenniums yet, in which people will show his shadow.—And we—we have still to overcome his shadow!

Friedrich Nietzsche, *Die fröhliche Wissenschaft, 1882*

Abstract

During 2012, the Large Hadron Collider (LHC) has delivered proton-proton collisions at 8 TeV center of mass energy to the ATLAS and CMS experiments. These two experiments have been designed to discover the Higgs boson and to search for new particles predicted by several theoretical models, as supersymmetry. The Higgs boson has been discovered by ATLAS and CMS experiments on July, 4th of 2012, starting a new era of discoveries in particle physics domain.

With the confirmation of the existence of the Higgs boson, searches for new physics involving this boson are of major interest. In particular, data can be used to look for new massive particles that decay into the Higgs boson accompanied with other particles of the standard model. One expected signature is a Higgs boson produced with a top quark, the two heaviest particles in the standard model. The standard model predicts a cross section of top-Higgs production, then any enhancement of their associated production will be a clear signature of physics beyond the standard model. In addition, the existence of physics beyond the standard model can also be reflected by resonances that decay into a top-quark and a Higgs boson.

In the first part of my work I describe the theoretical and experimental foundations of the standard model, as well as the experimental device. In the same theoretical chapter, I also discuss the formulation of an extension of the standard model. In addition, I describe a feasibility study of a search of one of the particles predicted by such model.

The second part contains the realization of the search for a top partner, T', within the CMS experiment. This top partner is a new particle very similar to the standard model top quark, but much heavier, that can decay into a top quark and a Higgs boson. The analysis looks for this particle in the full hadronic final state, where the Higgs boson decays into two b-quarks and the top quark decays into three standard model quarks, a b and two light quarks. In this channel, I reconstruct its mass from the identification of all its decay products. As a result of the analysis, I show the limits on the T' production cross section from the number of observed events in the specific signature.

Résumé

Le LHC (Large Hadron Collider) a produit en 2012 des collisions proton-proton à une énergie de 8 TeV dans le centre de masse pour les expériences ATLAS et CMS. Ces deux expériences ont été conçues pour découvrir le boson de Higgs et pour rechercher de nouvelles particules prédites par des modèles théoriques. Le boson de Higgs a été découvert le 4 juillet 2012 par les expériences ATLAS et CMS. Cette découverte marque le début d'une nouvelle période de recherche dans le domaine.

Avec la confirmation de l'existence du boson de Higgs, les recherches de nouvelle physique liées à ce boson sont devenues prioritaires. Par exemple, on peut chercher dans les données une nouvelle particule massive qui peut se désintégrer dans un boson de Higgs associé à d'autres particules du modèle standard. Une signature attendue est un boson de Higgs avec un quark top, les deux particules les plus lourdes du modèle standard. Le modèle standard prédit une section efficace pour la production du Higgs avec un quark top. Ainsi une mesure de cette section efficace montrant une valeur plus importante prouverait l'existence de physique au-delà du modèle standard. En outre, l'existence de physique au-delà le modèle standard pourrait montrer des résonances qui se désintègre dans un quark top et un boson de Higgs.

Dans la première partie de ce manuscrit, je présente les bases théoriques et expérimentales du modèle standard, ainsi que le dispositif expérimental. Dans le même chapitre théorique je discute une extension du modèle standard dans le cadre d'un modèle effectif englobant ce dernier. De plus, je détaille une étude de faisabilité d'une recherche d'une des nouvelles particules prédites par ce modèle, un quark vectoriel.

Dans la deuxième partie, la recherche dans CMS de ce quark vectoriel T', partenaire du quark top, est décrite. Ce partenaire du top est une nouvelle particule très similaire au quark top du modèle standard, mais beaucoup plus lourde. On considère le cas où ce nouveau quark se désintègre préférentiellement dans un quark top et un boson de Higgs. J'ai fait cette recherche dans le canal hadronique où le Higgs se désintègre en deux quarks b et le quark top se désintègre en trois quarks, un quark b et deux quarks légers. J'ai reconstruit la masse du T' à partir de l'identification de tous ses produits de désintégration. Le résultat obtenu est décrit sous forme des limites observées sur la section efficace de production du T' déduites à partir de cette analyse.

Remerciements

Si bien cette thèse se soit déroulée en France, je voudrais remercier touts ceux qui m'ont aidé ou accompagné pendant ma thèse dans leurs langues respectives afin que chacun d'entre eux puisse me comprendre.

En primer lugar quiero agradecerle profundamente a mi compañera en la vida, a mi esposa Adelaida Maria Rios Tabares. Gracias a ella, con su constante compañía y apoyo, llegamos a buen puerto. A mi familia le debo un gran agradecimiento. Aunque lejos, siempre estuvieron atentos para brindarme el apoyo que necesitaba en el momento preciso. Gracias a mi papa José Uriel Ruiz Marín, a mi mamá Gloria Amparo Álvarez Vargas y a mis hermanos : Verónica, Camilo y Vanessa. Finalmente quisiera agradecer a mis abuelos quienes siempre me apoyaron en mi carrera, Gildardo Álvarez Uribe y Marina Vargas Henao. A ellos dedico este trabajo, aunque no lo hayan logrado ver finalizado.

Je voudrais ensuite remercier l'IPNL et plus particulièrement le groupe CMS. Je remercie Stépahnie Beauceron, ma directrice de thèse, Bernard Ille, Patrice Verdier et Roberto Chierici qui m'ont guidé dans cette voie qui m'ont appris plein de choses. Je voudrais aussi remercier Viola Sordini, Stéphane Perriès, Jean Fay et Colin Bernet. Je remercie en particulier Camilo Carrillo, pour son amitié et de très intéressantes et nombreuses discussions. Je remercie aussi mon collègue de bureau Benoit, pour les nombreux "CC", les sorties et son impertinence. Je remercie aussi Martín, Elvire, David, Anne-Laure, Sébastien, Louis, Jean-Baptiste et Julien. Je voudrais remercier également le reste du groupe CMS. En dehors du groupe je voudrais grandement remercier Corinne Augier et Sandrine Hernandez qui m'ont aidé à traverser des moments difficiles. Je remercie Sylvie Florès pour toute son aide. Enfin, je voudrais remercier l'ensemble du laboratoire pour son accueil, en particulier le service financier qui a su gérer les difficultés relatives à mes missions et mes erreurs en français.

Je remercie très chaleureusement mon jury de thèse qui m'a fait des remarques très constructives et qui m'a donné une vision encore plus globale du domaine. Merci surtout à Patrizia Azzi et Sabine Crépé-Renaudin qui ont accepté de relire et commenter mon manuscrit.

Finalmente, quisiera agradecer a Jhovanny, Julián, Sergio y Alex, que desde la lejanía supieron darme una mano. Gracias para Ana Carolina y Maria del Carmen por los divertidos almuerzos. A mon groupe de son cubain “Clave 12” : Mario, Hernán, Olivier, Thierry, Franco, Rafael, Simon, Oscar.

Contents

Introduction

The seek to understand the nature of matter has been a long quest in human history. The ultimate objective to find the mechanisms that govern matter behavior and its composition has directed physics evolution and its discoveries. Concerning the mechanisms, four forces have been found in nature that dictate how different matter components interact between them. On the other hand, in the search for the fundamental components of matter the periodic table has been formulated. But even deeper in matter structure, a set of fundamental particles, without known inner structure up to present, have been also found.

One of the most astonishing findings has been the mechanics of the microscopic world. At the beginning of the XX century, physicists have described for the first time the mechanics of very small bodies. Such mechanics have been a challenge to human understanding of nature due to their radical difference to macroscopic mechanics. All these discoveries have been gathered and combined into a modern theory that describes the subatomic world: the standard model of particle physics.

This standard model has been confronted to nature during the last century with astonishing results. Up to now, almost all standard model predictions have been confirmed by several experiments. These confirmations have been a gigantic physics, mathematics and engineering task.

In particular, the LHC (Large Hadron Collider) has confirmed the last principal missing piece of the model. The LHC is the most powerful particle collider conceived and created by mankind up to our days. Even if it is a very big and complex machine, it operates with the same fundamental ideas of Rutherford of colliding particles ones against others to look at their inner structure. In 2012, with the finding of the Higgs boson, by ATLAS and CMS experiments, a long search of around half a century concluded.

Nonetheless, the standard model has limitations. Even though it constitutes the most complete picture of nature we have, it is not able to explain all phenomena in nature. LHC, as ATLAS and CMS experiments, have been also designed to look for extensions of the standard model designed to fill its lacunae.

This doctoral thesis is an effort to look for a prediction of an extension of the standard model using CMS data collected from collisions delivered by LHC during 2012. It has been structured in five chapters.

In the first chapter, I describe the standard model and its predictions. I proceed to describe the standard model limitations and an effective extension of this model and one of the new particles predicted by it. Afterward, I describe in detail the top quark and the Higgs boson, their properties, how they have been measured and their production according to the standard model.

In the second chapter, the experimental set up is detailed. In first instance, the Large Hadron Collider is described as well as the experiments located on it. Then, I make a detailed description of CMS experiment. This chapter is closed by the description of the techniques used in CMS to reconstruct particles from proton-proton collision events.

The third chapter is devoted to cast the phenomenological study I performed during my first year of thesis. I describe in detail the feasibility study done for a possible search in data of a Vector Like top partner. This chapter serves as a theoretical motivation to the data analysis described afterward in chapter 5.

During the first two years of my thesis I have performed different studies associated to the Monte-Carlo simulations of proton-proton collisions. This work is drawn up in the fourth chapter. I discuss there, what are Monte-Carlo simulations, how they are used in particle physics and the different packages in the market to produce them. I also show the specific task I have performed for CMS collaboration on the matter.

Finally, the fifth chapter presents the data analysis I performed using CMS data for proton-proton collisions provided by the LHC at 8 TeV. I describe in detail all the techniques used in the analysis and its results.

1 The Standard Model and beyond

From the Greeks, different theories about the composition and structure of the world have been formulated. In ancient Greece these theories were elaborated from a philosophical point of view. Nowadays, a very sophisticated set of tools and concepts has been developed, which allowed the construction of a general vision of nature, its components and structure. Moreover, on the subject of the constituents, or elemental constituents, a theory capable of describing the majority of known phenomena has been developed. This theory is the Standard Model (SM) of particle physics. It is a model constructed by the contribution of many people. One of its seminal papers is [1]. A modern formulation of this model is presented in [2].

The SM relies on two of the more elegant constructs of modern physics and mathematics. From the physics side, the quantum field theory and from the mathematics one group theory. Quantum field theory [3] is born from the understanding of processes that take place at very small spatial scales and in a regime where special relativity plays an important role. To describe this, a major part of the most brilliant minds of the 20th century dedicated their life: Paul Dirac [4], Richard Feynman [5], Enrico Fermi [6] among them. The theory of quantum fields has set in a common place two extraordinary achievements of physics: special relativity [7] and quantum mechanics. With it many phenomena have been described: β and α decay, solid state, among many other.

From the mathematics side, group theory has become one of the most powerful tools for particle physicists. A modern presentation of this theory, oriented to particle physics, is done in [8]. However, its development began quite early, with Galois around 1830 [9], and was used in other parts of physics. It is with Lie algebras [8] and the possibility of describing continuous symmetries that the most important steps were given. Also, this would have not been possible without the amazing connection found by Emmy Noether in 1918 [10]. She found that for every conserved quantity there is a preserved symmetry. Group theory can be seen, roughly speaking, as a way to mathematically describe symmetries and group theory became the tool to describe systems with conserved quantities.

In this chapter, the basics of the SM are presented. Its seminal ideas are presented, its structure and content and its ultimate consequences. Also its limitations are discussed and how a new theory addresses part of them. The chapter finishes with the description of the production of some of the SM particles in the Large Hadron Collider [11].

1.1 Fields, symmetries and interactions

From the very beginning of physics, one of the most fundamental questions has been how do bodies interact, and what is exactly an interaction. On the first type of interaction ever studied by physics, gravity, Newton proposed the concept of distant interaction [12]: the idea that bodies could interact without being in direct contact. But the question on how exactly that distant action was performed remained unanswered.

During the 19th and 20th century new phenomena were discovered pointing to brand new interactions, electricity, magnetism and radioactivity [13]. The very precise and complete description of electromagnetism developed by Gauss, Faraday, Ampère and finished by Maxwell succeeded in describing electricity and magnetism under the formalism of only one interaction within the mathematical formalism of classical fields. Further works addressed radioactivity, driving to a deeper understanding of nature and its composition.

A classical field is an assignment of a quantity to every point in space and time. For physics, the quantity that is attributed is a physical quantity such as mass, electrical charge or momentum. This quantity can be scalar or vector, giving rise to the notion of scalar or vector field, correspondingly. A comprehensive study of this concepts is done in [14]. The simplest example, is the temperature in a gas, that is a scalar quantity assigned to every point. Another example, a fluid can be described in terms of fields, being its velocity a vector field and its pressure a scalar field. Generic classical electromagnetic interactions can be described with the help of one vector field $\vec{A}(x)$, the vector potential, and one scalar field $\phi(x)$, the scalar potential. In Einstein's special relativity [7], dynamics of physical objects are described via the formalism of four vectors. In Newtonian dynamics [15] the position or speed of an object are described in three dimensional space by a three dimensional vector. In relativistic dynamics, one dimension is added to these vectors, for example the three dimensional position vector has an additional temporal dimension, or the momentum vector has an additional dimension with the energy of the object. In this formalism the scalar and vector potential can be organized in the four-potential $A_\mu = (-\phi, \vec{A})$. This can be used to define the strength field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, where $\partial_\mu = \left(-\frac{\partial}{\partial t}, \nabla\right)$ is the covariant derivative. From this tensor it is possible to obtain in a very generic and elegant way the equations of motion of the free field using the Lagrangian formalism, as in equation 1.1. With the Lagrangian density defined in equation 1.2. The Lagrangian formalism is a generalization of Newtonian dynamics that allows to obtain from a generic principle the equations of motion of a system. This formalism is described in [16].

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} \right) - \frac{\partial \mathcal{L}}{\partial A_\nu} = 0 \tag{1.1}$$

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

It is very important to notice that the equations of motion of the free field are invariant under the choice of the four-potential. More precisely, the covariant potential is not unique and the covariant derivative of a scalar field, $\partial_\mu \Lambda(x)$, can always been added,

$$A'_\mu = A_\mu + \partial_\mu \Lambda(x) \leftrightarrow \partial^\mu A_\mu = 0 \tag{1.3}$$

and describe the same physics. This non-uniqueness corresponds to the choice of a zero-point of the potential very well known in non-Lagrangian formalism of electrodynamics. When a specific value for this scalar field is chosen, $\Lambda(x)$, it is said that the gauge has been fixed. Further details are given in [17].

A four current vector can also be defined, $J_\mu = \left(\rho, \vec{J}\right)$ with ρ the electric charge density and $\vec{J}$ the current charge density. Then, plugging in this four current in the Lagrangian of the free field, defined in equation 1.2,

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} - A_\mu J^\mu \tag{1.4}$$

the complete set of equations of motion of the field with charges and currents are obtained.

The transformation stated from equation 1.3 can be understood as a transformation of the field. These type of transformations are mathematically understood under the group $U(1)$ [8], where the generic transformation operator can be written as $U = e^{i\theta(x)}$. It is said then that the electromagnetic vector potential is invariant under $U(1)$ transformations. This property identifies an essential characteristic of electromagnetism, its symmetric behavior under $U(1)$.

From this reasoning the most interesting results are drawn when the same symmetry is imposed to other fields. For example, the kinetic Lagrangian for a complex scalar field is $\mathcal{L} = (\partial^\mu \phi)^* \partial_\mu \phi$. To perform the transformation on the scalar field, it is sufficient to apply the operator as $\phi' = U\phi$ and $\phi'^* = \phi^* U^{-1}$. But it is evident that the Lagrangian is not the same after applying such transformation. Then, in order to preserve the Lagrangian under $U(1)$ it is necessary to change at the same time the derivative. Such transformation is given in equation 1.5, where g is a constant.

$$\mathcal{D}^\mu = \partial^\mu - igA^\mu \tag{1.5}$$

Then, the proposed Lagrangian can be rewritten, including the vector field, as

$$\mathcal{L} = (\mathcal{D}^{\mu}\phi)^{*}\mathcal{D}_{\mu}\phi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{1.6}$$

which is invariant under $U(1)$. An interaction term, of the form $igA^{\mu}\phi^{*}\partial_{\mu}$, between the scalar and the vector field, can be derived from the kinematic part of the Lagrangian $(\mathcal{D}^{\mu}\phi)^{*}\mathcal{D}_{\mu}\phi$. This shows that the requirement of the invariance under $U(1)$ of the scalar field leads to the introduction of an interaction with a vector field controlled by the constant g. The electromagnetic interaction is described precisely by a vector field and preserves $U(1)$ symmetry, which implies that this symmetry is the connection to electromagnetic interaction, identifying the interaction itself with the $U(1)$ symmetry. In addition, using Noether theorem [10] it can be shown that g is a conserved quantity, as the electric charge is.

But not only electromagnetism can be described via a continuous symmetry as $U(1)$. On 1896, radioactivity was discovered by the french physicist Henri Becquerel [18]. Three years after, Marie and Pierre Curie [19] studied in more detail the phenomenon and found Polonium and Radium elements. And later on, Ernst Rutherford [20] was able to describe radioactivity as coming in three types, alpha (α), beta (β) and gamma (γ). He also noticed that radioactivity was able to change matter, which allowed him to propose an atomic model, describing elements as basically and external core of negative charges and a nucleus positively charged. Consequently, this findings implied the existence of interactions different to electromagnetism, acting at the atomic scale.

The interaction that undergoes radioactivity, beta decay, is called the weak interaction. In 1934 Enrico Fermi made a first theoretical description of this interaction [6], but only in 1968 Sheldon Glashow, Abdus Salam and Steven Weinberg were able to describe the weak interaction with a symmetry group: $SU(2)$ [21, 22].

Finally, the interaction that keeps the nucleus components together, the strong interaction, was described with $SU(3)$ group mainly by Murray Gell-Mann in 1964 [23].

There have been many attempts to describe gravity with the same formalism, but up to present such attempts have been unsuccessful. This question remains one of the most important problems for modern particle physics.

1.2 Quantum fields and particles

Classical fields, introduced and described in last section 1.1, can be extended to a quantum theory. Such procedure is known as the quantization of fields and allows to unify special relativity and quantum mechanics in one theory, Quantum Field Theory (QFT). It describes the dynamics of systems in the regime where the speed is close to the speed of light in the atomic or smaller scales.

Quantum mechanics introduced two fundamental concepts: first, the description of the system by its states; and second, the identification of an observable with an operator. The state of a system is identified with a set of quantum numbers that characterizes the system state. For example, the energy is a quantum number of the hydrogen atom system, such that each state has a value for the energy related to the potential energy contained in the system. Quantum states are mathematically noted in Dirac notation as a ket,

$$|\alpha\rangle = |i, j, k, \ldots\rangle \tag{1.7}$$

with α the set of quantum numbers $i, j, k, \ldots$ This mathematical object lives in Hilbert space (a complex space $\mathbb{C}$ of functions), the conjugate of which, a bra, is noted $\langle\alpha|$, and their internal product $\langle\beta|\alpha\rangle$. The numerical value of $|\langle\beta|\alpha\rangle|^2$ gives the transition probability of the system from state β to state α, and $|\langle\alpha|\alpha\rangle|^2$ is the probability to find the system in the state α.

Physical observables like position, energy or momentum are described by complex operators such that $|\langle\alpha|\hat{O}|\alpha\rangle|^2$ has to be calculated in order to measure their value for a given state. The identification of observables and operators is called first quantization. In addition, the Schrödinger equation [24] describes the evolution of states,

$$\hat{\mathcal{H}}|\alpha\rangle = i\frac{d}{dt}|\alpha\rangle \tag{1.8}$$

with $\mathcal{H}$ the Hamiltonian of the system. The whole formalism is able to explain quantized systems, where the quantum numbers are discrete, such as the hydrogen atom or the black body.

Several functions or fields can be related to a system. These functions, wave functions, can be used to calculate the probability of a system to be in a given state or the transition probability between states. In second quantization, wave functions are upgraded into field operators. This procedure gives rise to the quantization of the state of the field, which is described by the quantum number n which is definite positive: $n = 0$ for the fundamental state and $n > 0$ for the excited states. Such excitations of the field are understood as physical particles, or quanta, that propagate in space-time, which means that $n = 0$ is the vacuum.

The first QFT ever created was born from the quantization of the electromagnetic field. Quantum Electro Dynamics (QED) is the quantized version of classical electrodynamics, that was developed by Tomonaga, Schwinger and Feynman around 1940 [25, 26, 27]. This theory describes electromagnetic interactions of a charged field and the electromagnetic vector field.

The charged field excitations correspond to electrons and the excitations of the vector fields are photons, responsible of light. Electrons are particles with negative electric charge that orbits around the nucleus in atoms. Discovered in 1897 by J. J. Thomson [28], it was fully described by P. A. Dirac in 1927 [4] with the Dirac equation that is the Schrödinger equation for a relativistic particle of spin 1/2. Spin, the intrinsic angular momentum carried by a particle, can be integer (0,1,2,...) or semi-integer ($\frac{1}{2}$,$\frac{3}{2}$,...). The particles with semi-integer spin, as electrons, are called fermions and particles with integer spin, as photons, are called bosons. Dirac equation predicted the existence of a particle identical to the electron but with positive charge, the positron. It was discovered in 1931 by Carl David Anderson [29].

The SM has three fundamental interactions: electromagnetic, weak and strong interactions. Each fundamental interaction is associated with a charge. Charged particles interact via the corresponding force. The electromagnetic interaction is associated with the electric charge. The weak charge is called the weak isospin. Finally, the strong force charge is called color.

Up to present days, 12 fundamental fermions and 5 fundamental bosons have been found. Fermions are organized in leptons, that do not interact strongly, and quarks, that do interact strongly. Leptons are organized in three families, the electron (e^-) and electron neutrino (ν_e), muon (μ^-) and muon neutrino (ν_μ) and tau (τ) and tau neutrino (ν_τ). Electron, muon and tau are electrically charged while neutrinos are neutral. Their respective anti-particles are equally organized, positron (e^+) with electron anti-neutrino ($\bar{\nu}_e$), anti-muon (μ^+) with muon anti-neutrino ($\bar{\nu}_\mu$) and anti-tau (τ^+) with tau anti-neutrino ($\bar{\nu}_\tau$). Quarks also come in three families, with the respective anti-quarks: up (u, $\bar{u}$) and down (d, $\bar{d}$), charm (c, $\bar{c}$) and strange (s, $\bar{s}$), top (t, $\bar{t}$) and bottom (b, $\bar{b}$). The fundamental bosons are the photon (γ), the $W^{+/-}$ (positively and negatively charged) and Z^0 that mediate the electroweak interaction, the gluons (g) mediating the strong interaction and the Higgs (H^0). The Higgs boson does not mediate an interaction but plays another role in the SM, that will be described in section 1.3. The strong force mediator, the gluon, was discovered by the PETRA [30] and PLUTO [31] experiments in 1979 at DESY. The weak bosons were discovered at CERN in 1983 by the UA1 and UA2 collaborations [32] at the SPS accelerator [33] (described on section 2.1.1), called at that time Sp$\bar{\text{p}}$S because it was a proton-antiproton collider. The Higgs boson has been discovered recently in 2012 by ATLAS [34] and CMS [35] experiments at the LHC [36, 37]. The 2013 physics Nobel prize [38] was awarded to Francois Englert and Peter W. Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider". The particle content of the SM is resumed in figure 1.1.

In the next section, one of the fundamental pieces of the SM is going to be

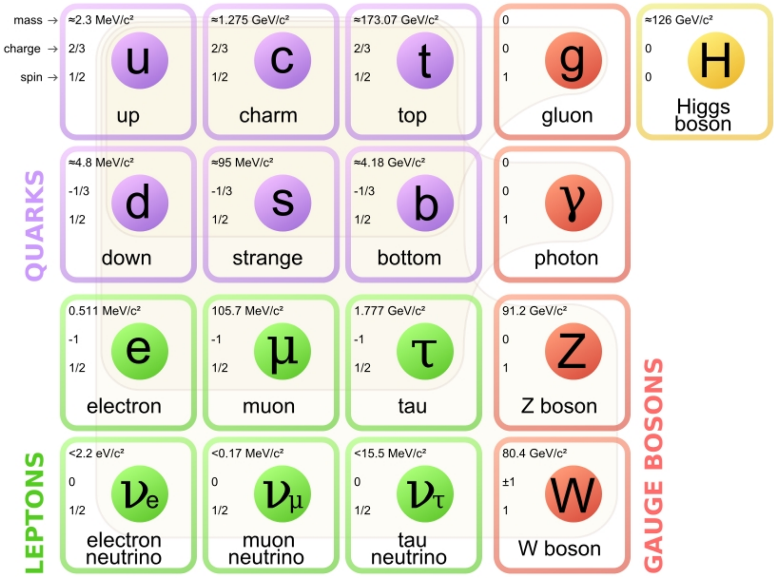


Figure 1.1: Particle content of the standard model.

described: how particles acquire mass. Whatsoever, even if the mechanism to give mass to particles is one of the most refined theoretical constructs, it leaves some unsolved puzzles in the heart of the SM from experimental observations. Later, the limitations of the SM and some theoretical insights to solve them are discussed.

1.3 The mass problem

Using the concepts developed on previous sections about QFT and symmetries it is possible to construct a whole theory giving rise to a precise description of particles and interactions between them. But such a theory does not allow to have massive bosons, whereas masses for fermions are allowed. A mass term for a fermion ψ is of the form

$$m_\psi \bar{\psi}\psi$$

where m_ψ is the mass of the field.

Under $U(1)$ transformations, $\psi' = U\psi$, the mass term remains the same, which means that it is invariant under $U(1)$ transformations. The same is not true for a boson. A mass term for a boson A, can be written as

$$m_A A^\mu A_\mu \tag{1.9}$$

where m_A is its mass.

The $U(1)$ transformation for the boson is $A'^{\mu} = A^{\mu} + \delta^{\mu}\theta(x)$, where $\delta^{\mu}\theta(x)$ represents an infinitesimal variation of the field by the coordinate depending parameter $\theta(x)$. Applying such a transformation on the mass term 1.9, the transformed term can be obtained

$$m_A\left(A^{\mu}A_{\mu} + A^{\mu}\delta_{\mu}\theta(x) + \delta^{\mu}\theta(x)A_{\mu} + \delta^{\mu}\theta(x)\delta_{\mu}\theta(x)\right) \tag{1.10}$$

where the last three terms make equation 1.10 not invariant under the $U(1)$ transformations. Consequently, a mass term for the boson destroys the invariance of the theory under $U(1)$ symmetry.

Nonetheless there is no need for a theory of a massive photon, however there is a need to have massive bosons for weak interactions. Indeed there is a relation between the mass of a boson and the range of the interaction mediated by it. Massless bosons transmit long range interactions, as electromagnetism, but short range interactions, as the weak interaction, must be mediated by massive bosons. More precisely the interaction range is inversely proportional to the mass of the boson, higher is the mass shorter is the range. Such relation can be seen from the structure of the propagator, which is a mathematical entity that describes the probability a particle has to travel a distance in a given time range. Such propagator, for a vector boson, has a generic form given in equation 1.11, where k_{μ} is the momentum carried by the boson, m its mass and ϵ the usual term to avoid the integration pole when $k^2 = m^2$. It is clear from this structure that a massive boson has less probability to travel a long distance than a massless boson in a given time.

$$\frac{1}{k^2 - m^2 + i\epsilon} \tag{1.11}$$

In the former discussion, the strong force is an exceptional case. Whereas the gluon is massless the interaction range of the strong force is limited by a phenomenon called confinement [2]. Due to the strength of the strong force, they can not be seen freely in nature but confined in hadrons. Therefore the interaction range is restricted to very short distances, typically nuclear distances less than one femtometer.

As massive bosons are required for weak interaction, somehow the $SU(2)$ symmetry has to be broken. There are basically two ways to break a symmetry:

- Explicit symmetry breaking: by the introduction of a symmetry breaking term in the Lagrangian, as a mass term for the bosons.

- Spontaneous symmetry breaking: when the ground state of one field fails to be invariant under the symmetry.

Explicit symmetry breaking is not an option, because the symmetry needs to be preserved in the lagrangian in order to introduce the interaction.

1.3.1 Spontaneous Symmetry Breaking

Several physical systems exhibit a spontaneous symmetry breaking. For example, a pencil balanced on its tip is a perfectly symmetric system around the vertical axis, however, because of the instability of the system the pencil will eventually fall over. The final state is stable but it is not symmetric under rotations around the z-axis. This change of state also decreased the potential energy of the system, driving the system to its ground state. This means that whereas the system had a symmetry the ground state does not show the symmetry. In general, symmetry breaking is linked to phase transitions, as liquid to gas transition or magnetization of a ferromagnet, covering a plethora of physical processes.

To achieve a spontaneous symmetry breaking of $SU(2)$ in QFT, a field should be chosen for which its ground state, vacuum, will fail the symmetry. This means, in practical terms, that such field will have a non-zero value in vacuum, i.e. there are particles from the field in vacuum. If a fermion field is chosen, the vacuum will show a preference on directionality depending on its spin orientation, which breaks Poincaré symmetry imposed by special relativity. The same is true if a spin-1 bosonic field is chosen for the task. In order to avoid this problem a spin-0 scalar field should be used to break the symmetry. In addition, this field should be electrically neutral to avoid having a vacuum with a non-zero electrical charge.

With all this properties in mind, a scalar doublet of $SU(2)$ can be defined as in equation 1.12, where ϕ^0 and ϕ^+ are complex fields. The most general potential can then be written from two auto-interaction terms as in equation 1.13.

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi^+ \\ \phi_{RE} - i\phi_{IM} \end{pmatrix} \tag{1.12}$$

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \tag{1.13}$$

Such a potential has a unique minimum for $\lambda > 0$ and $\mu^2 > 0$, but for $\lambda > 0$ and $\mu^2 < 0$, it has a set of minima with the shape of “Mexican hat”, shown in figure 1.2. Under $\lambda > 0$, $\mu^2 < 0$ configuration the field breaks spontaneously the symmetry reaching the ground state, acquiring an expectation value on vacuum v, different from zero. This procedure and the mechanism giving masses to particles in the SM was first proposed in [39, 40].

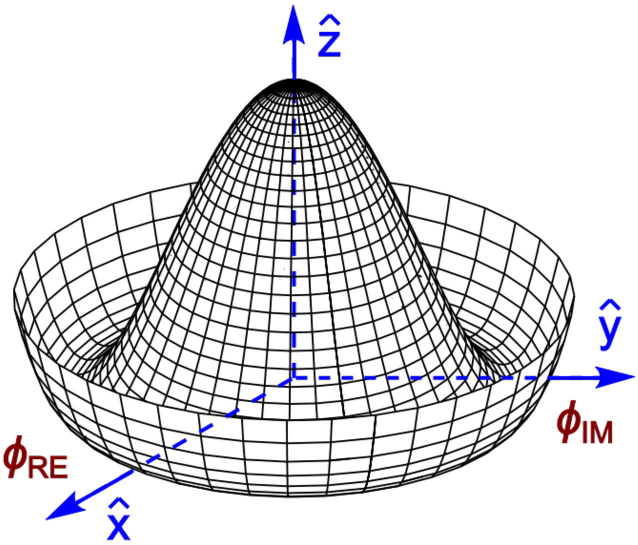


Figure 1.2: Complex scalar doublet potential.

1.3.2 Brout-Englert-Higgs mechanism

After the spontaneous symmetry breaking, the scalar doublet transforms into the form given in equation 1.14, where G^+ and G^0 are the Goldstone bosons from the breaking of the $SU(2)$ symmetry, and H^0 is the Brout-Englert-Higgs boson. From the Goldstone's theorem, when a symmetry is spontaneously broken, a massless boson appears for each broken generator. In our specific case, the three generators of $SU(2)$ are broken giving rise to three Goldstone bosons: G^+, G^- and G^0. This massless bosons are "eaten" by the W^+, W^- and Z^0 giving them an additional degree of freedom, the longitudinal polarization.

$$\Phi = \begin{pmatrix} G^+ \\ \frac{1}{\sqrt{(2)}}(H + v - iG^0) \end{pmatrix} \tag{1.14}$$

By this mechanism, the $W^{+/-}$ and Z^0 bosons acquire a mass, being its value set by the coupling constant of $SU(2)$ group and the vacuum expectation value of Brout-Englert-Higgs boson. In addition, the fermions in the theory also acquire a mass from the interactions with the scalar doublet. Such masses are in general of the form $m_f = \lambda_f v/\sqrt{(2)}$, where λ_f sets the interaction between the Brout-Englert-Higgs boson and the fermion. Finally, also the Brout-Englert-Higgs boson has a mass $m_H^2 = -2\mu^2$.

In summary, with this mechanism, the weak interaction bosons and fermions of the theory are given a mass at the price of introducing at least one additional scalar field to spontaneously break the $SU(2)$ symmetry.

1.4 Hierarchy problem and other SM limitations

The SM is certainly one of the most successful theories in the history of physics. With only 19 free parameters, it is able to make thousands of predictions that have been measured and tested over the last seventy to eighty years. However some aspects in the model are not completely understood. The most important one is the so-called hierarchy mass problem. At tree level, the Brout-Englert-Higgs boson has a mass $m_H^2 = -2\mu^2$, but the physical mass also contains the one-loop contributions from particles that interact with it, as the top quark. The Feynman diagram for such contribution can be seen in figure 1.3.

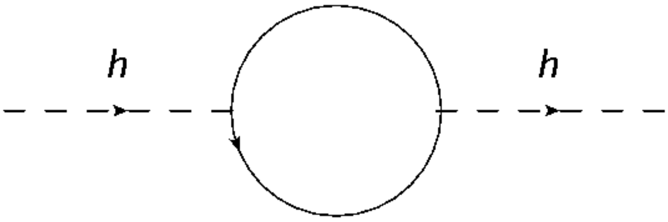


Figure 1.3: One loop diagram for contributions to the mass of the Brout-Englert-Higgs boson from interactions with fermions.

Such contributions add up giving a mass greater than the simple tree level mass. Each fermion contributes proportionally to its mass, which means that the top quark contributes the most. Moreover, if there are in nature heavier fermions that also interact with the Brout-Englert-Higgs boson they will also contribute to its mass. With such considerations, the Brout-Englert-Higgs boson mass is expected to be much greater than 125 GeV, and in principle not even of the order of 100 GeV but greater than 1 TeV.

The most famous proposed solution to this problem is supersymmetry (SUSY) [41, 42]. It proposes the existence of an additional symmetry between fermions and bosons. At a given point of the history of universe, nature did not distinguish between fermions and bosons. However, as it is known, this does not happen presently, that means this symmetry must be broken. Such symmetry implies the existence of a super-symmetric partner for each particle, a super-partner: a fermion for each boson and vice versa. This SUSY procedure doubles the particle content of the model. Before breaking SUSY, a particle and its partner have the same mass. This characteristic solves the hierarchy problem in SUSY. Figure 1.4

shows the one loop diagrams for the mass of the Brout-Englert-Higgs boson from the top and its super-partner the stop. Whereas, the top contribution is positive, the stop contribution is negative but equal in value, then cancelling each other.

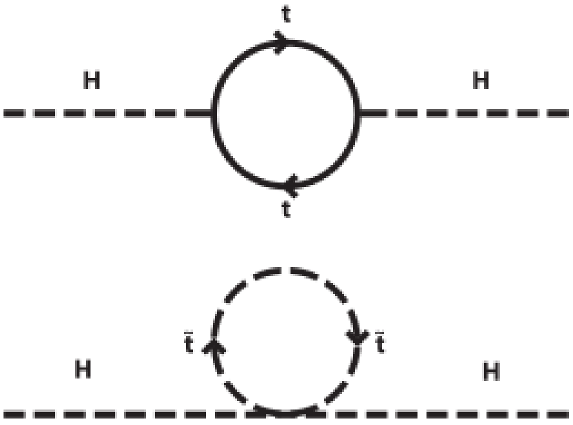


Figure 1.4: One loop diagrams for contributions to the mass of the Brout-Englert-Higgs boson from the top and the stop.

But this solution works exactly only if SUSY is not broken. As SUSY has to be broken, different ways have been proposed to break SUSY and still offer a solution to the hierarchy problem, leading essentially to solutions that need a fine adjustment of the parameters of the theory. This represents for some theoreticians a problem by itself, referred as fine-tuning or naturalness. The name fine-tuning refers to the need of choosing extremely tuned values for the parameters to adjust a model to the theoretical requirements Naturalness asks the justification of the specific values given to parameters in a fine-tuned model. Extensive searches for SUSY particles have been performed, according to different models: MSSM [43, 44], CMSSM [45], etc.

While hierarchy problem is an internal problem of the SM, there are however several questions that have not been solved. For example, how gravity is understood in the frame of QFT? Why are there only 3 generations of leptons and quarks? Why are there 4 fundamental forces? In addition, there have been experimental questionings to the SM. The most important one is the masses of neutrinos. In the SM, neutrinos are massless, but careful measurements [46, 47] have shown that neutrinos can oscillate between different flavors, phenomenon only possible if neutrinos have a mass. Measurements of solar and atmospheric neutrino oscillations have been the most important proofs of physics beyond the SM.

From cosmological measurements, the Wilkinson Microwave Anisotropy Probe (WMAP) has shown that the universe is not only made by visible matter, but suggests that around 24% of it is made of dark matter, a type of matter not visible by means of light. It has also shown that 71% of the universe is composed of dark energy, which makes the universe to be in an accelerated state of expansion. These results are shown in [48, 49]. Also the Planck probe has shown similar results, for example in [50]. The SM does not have any answer to these open problems so far.

Finally, it is known that the universe presents an asymmetry between matter and antimatter, the first being much more abundant than the second. Such asymmetry can be obtained by CP-violating processes (C for charge and P for parity). However the amount of CP violation in the SM is not compatible with the huge matter-antimatter asymmetry in nature. This problem, known as baryon asymmetry, represents an additional huge challenge for particle physics.

In conclusion, the SM has been a formidable model that has helped the understanding of a huge amount of physics. It has done thousands of predictions that have been verified experimentally in the last half-century. However, this is not the end of the story, perhaps only the beginning. There are theoretical and experimental motivations that support the idea that the SM is not the “final” theory that could explain all subatomic phenomena in nature. Currently, there are a lot of efforts, both theoretical and experimental, to understand and explain all the remaining pieces. The present work is one of them.

In the next section, an extension of the SM that looks for a solution to the discussed hierarchy problem is presented. Chapter 3 shows a strategy to look at LHC data for some of the predictions given by this SM extension.

1.5 Vector Like Quarks

From last sections, it has been shown how there are phenomena not described by the SM that need further understanding. From such problems some further models/theories have been developed. All this theories are commonly grouped under the term Beyond Standard Model or simply BSM. As already mentioned, one of the most famous BSM theory is supersymmetry (SUSY). This idea have given birth to a plethora of model realizations and physics predictions. So far, none of the new consequences of this theory have been confirmed but the experiments have an enormous investment on their search. But not only SUSY have seen the day light, there is on the market an astonishing amount of BSM theories addressing SM issues in different ways. Extra dimensions, fourth families, composite Higgs are a few of them.

In this section a set of models that introduce additional heavy quarks, heavier than the top, in order to solve the hierarchy problem, will be described.

1.5.1 Theoretical motivation

Even though the SM has 3 families, the number of families in nature is not restricted by any fundamental principle. However, searches for fourth families have not found any evidence of more than three families in nature [51]. But a fourth family formulation relies on the existence of a sequential additional family. For the quark sector, two additional quarks, heavier than the top with the same quantum number structure as the other 3 families. In particular, the 3 SM families are formed by doublets of the left components of the quark fields. A fourth family would introduce a fourth neutrino, this possibility that has been discarded by LEP [52] if the fourth neutrino couples to the Z^0 boson. Additionally, in [51], have been shown the consequences of last measurements of Higgs mass on a SM with a sequential fourth family. Such work shows that this fourth family model is excluded at 5.3 sigma.

Fermionic fields can be divided in terms of two chiral components. The chirality is defined with respect to chiral operators that decompose a field in its left and right components. The left chiral operator is $\frac{1-\gamma^5}{2}$ and right chiral operator is $\frac{1+\gamma^5}{2}$. Accordingly to the electroweak theory only the left fermions undergo weak interactions, reason why SM quark families are left. The chirality is related to the helicity, the later describes the relative orientation of a particle spin with respect to its momentum. A particle is said to have right helicity if its momentum and spin are aligned and left helicity if they are anti-aligned. The chirality and the helicity of a particle are the same for massless particles, as the photon.

Nothing prevents to formulate a model where new quarks right and left components are subjected to weak interaction indistinctly. A fermion having identical left and right chiralities is called vector-like. These type of models introduce only new quarks without the need of a new family.

These new states allow to have a suitable solution for the hierarchy problem. They arise typically in Kaluza-Klein extra dimension models [53, 54, 55, 56]. In more generic modeling of vector-like quarks (VLQ) [57, 58], these new states are added to the SM content as an electroweak representation (singlet or doublet) coupled to SM quarks.

1.5.2 Generic Formulation

The generic formulation that is going to be described is based on [58, 59].

In a very generic conception, vector-like quarks can be added to the SM content as being part of an electroweak representation. Introduced with the same color charge than SM quarks, they can come in 4 different forms depending of their electric charge:

- X with 5/3 electric charge

- T' with 2/3 electric charge, as the top quark.
- B' with -1/3 electric charge, as the bottom quark.
- Y with -4/3 electric charge.

The T' is also called in the literature a top partner. Given the VLQ electric charges, they can be arranged in different representations. Table 1.1 displays the plausible different electroweak representations of VLQ. The $\binom{X}{T'}$ and $\binom{T'}{B}$ doublet will be referred as the non-standard and standard doublet respectively.

	SM	Singlets		Doublets			Triplets	
	$\binom{u}{d}$ $\binom{c}{s}$ $\binom{b}{t}$	T'	B	$\binom{X}{T'}$	$\binom{T'}{B}$	$\binom{B}{Y}$	$\begin{pmatrix} X \\ T' \\ B \end{pmatrix}$	$\begin{pmatrix} T' \\ B \\ Y \end{pmatrix}$
$SU(2)_L$	2	1		2			3	
$U(1)_Y$	$q_L = 1/6$; $u_R = 2/3$; $d_R = -1/3$	2/3	-1/3	7/6	1/6	-5/6	2/3	-1/3

Table 1.1: Possible VLQ representations and corresponding $SU(2)_L \times U(1)$ charges and SM quarks.

The effective Lagrangian for a general formulation of VLQ, restricting it for T' and X type of vector-like quarks in the non-standard doublet is the following:

$$\begin{aligned} \mathcal{L} &= \kappa_T \left\{ \sqrt{\frac{\zeta_i \xi_Z^T}{\Gamma_Z^0}} \frac{g}{2c_W} \left[\bar{T'}_R Z_\mu \gamma^\mu u_R^i \right] \right\} \\ &- \kappa_T \left\{ \sqrt{\frac{\zeta_i \xi_H^T}{\Gamma_H^0}} \frac{M}{v} \left[\bar{T'}_L H u_R^i \right] + \sqrt{\frac{\zeta_3 \xi_H^T}{\Gamma_H^0}} \frac{m_t}{v} \left[\bar{T'}_R H t_L \right] \right\} \\ &+ \kappa_X \left\{ \sqrt{\frac{\zeta_i}{\Gamma_W^0}} \frac{g}{\sqrt{2}} \left[\bar{X}_R W_\mu^+ \gamma^\mu u_R^i \right] \right\} + h.c. \end{aligned} \tag{1.15}$$

neglecting terms proportional to the light quark masses. In addition, terms proportional to the top mass can also be neglected as they only introduce differences of only 10%-20% for T' masses below 600 GeV. The parameters κ_T, κ_X are the coupling strengths and determine the strength of the VLQ production. In equation 1.15, M represents the T' mass, m_t the top mass, ζ_i the coupling between VLQ and the ith SM quarks generation, ξ_Z^T and ξ_H^T the mixing of T' with the Z and Higgs boson respectively. Finally, v and g are the usual vacuum expectation value and electroweak coupling in the SM, and Γ_i^0 the width of the boson $i = W^{+/-}, Z^0, H^0$.

Production modes

VLQ's can be produced singly or pairwise, as any SM quark. Pair production is dominated by QCD processes via gluons, as shown in figure 1.5. Additional contributions are given via the t-channel exchange of a vector boson. For example, figure 1.5 shows the production process via the Z^0 boson that produce two same sign VLQ. Additionally, two different VLQ's can be associatively produced in proton-proton collisions, for example$\bar{q}q' \to \bar{T}'X, \bar{B}T', \bar{Y}B$ via the exchange of a $W^{+/-}$ boson.

Single production is mainly mediated by a vector boson via a t-channel in association with a SM quark. In figure 1.6 the Feynman diagram for T' single production with a SM quark is shown. In addition, the T' can be also produced with a Z^0 or $W^{+/-}$ boson or a Higgs from an initial gluon-quark fusion.

The VLQ production cross section in proton-proton collision strongly depends of their mixings with SM quarks. Historically, top partners have been modeled with exclusive mixings to third generation SM quarks. In the generic formulation being considered, the VLQ can mix to the three SM quark generations. In the present work, the case where the couplings of the T' to light generations are comparable to the third generation couplings has been considered. Figure 1.7, [59], shows the T' production cross sections in pairs and singly in proton-proton collisions. For T' masses higher than 500 GeV/c^2, the single production is dominant with regard to the pair production, at 8 TeV center of mass energy. Table 1.2 shows the cross sections for the production of the T' in single and pair modes at 8 TeV proton-proton collisions. For such cross sections the couplings of the T' to light generations have been maximized with respect to experimental bounds, as described in [59].

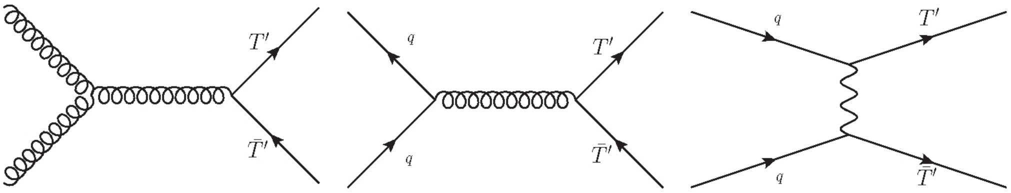


Figure 1.5: Feynman diagrams of T' production in pairs.

In the present work the single production of a T' in association with light quark is considered. However, another interesting possibility is the T' single production with an associated top-quark. Is also important to remark that up to now, only the case of single production with jets has been used in searches by ATLAS and CMS experiment. An interesting signatures for future searches could be also the single production in association with an electroweak boson.

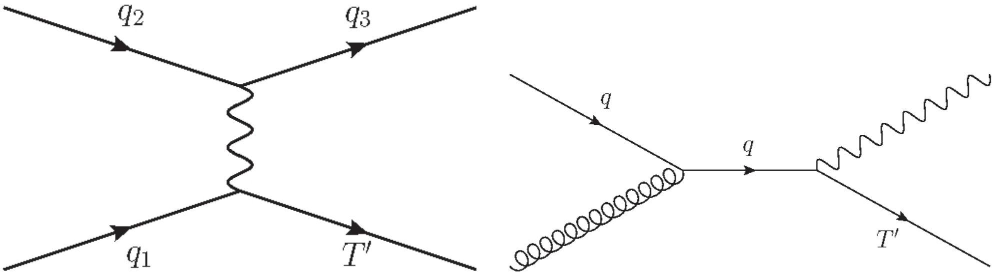


Figure 1.6: Single T' production Feynman diagrams.

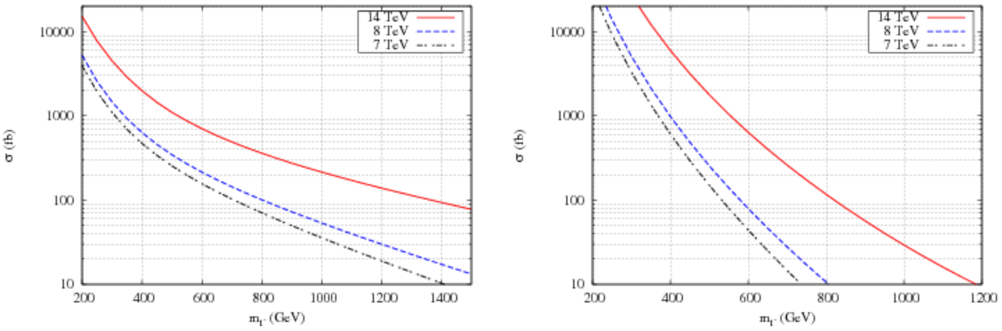


Figure 1.7: T' production cross section for single [left] and pair [right] channels as a function of the T' mass for different center of mass energy in proton-proton collisions. The cross sections have been calculated for the non-standard doublet with maximized couplings to light generations [59].

	Cross section [fb]	
T' Mass	Single production	Pair production
400	635.7	992.5
450	452.6	483.5
500	347.5	263.8
550	266.8	133.4
600	215.4	75.6
650	177.8	46.2
700	143.7	29.3
750	118.6	16.6
800	100.0	10.5

Table 1.2: T' production cross sections in single and pair production modes, for proton-proton collisions at 8 TeV center of mass energy. The cross sections have been calculated for the non-standard doublet with maximized couplings to light generations.

Decay modes

The parameters ζ_i and ξ_i, from the Lagrangian in equation 1.15, are directly linked to the branching ratios:

$$BR(T' \to Z^0 j) = \frac{\zeta_{jet}\xi_Z^T}{1+\zeta_3\xi_H\delta_H}\,, \qquad BR(T' \to Z^0 t) = \frac{(1-\zeta_{jet})\xi_Z^T}{1+\zeta_3\xi_H\delta_H}\,, \tag{1.16}$$

$$BR(T' \to H^0 j) = \frac{\zeta_{jet}(1-\xi_Z^T)}{1+\zeta_3\xi_H\delta_H}\,, \qquad BR(T' \to H^0 t) = \frac{(1-\zeta_{jet})(1-\xi_Z^T)(1+\delta_H)}{1+\zeta_3\xi_H\delta_H}\,,$$

$$BR(X \to W^+ j) = \zeta_{jet}\,, \qquad BR(X \to W^+ t) = (1-\zeta_{jet}) \tag{1.17}$$

where ζ_{jet} accounts for the contributions of the two light quark generations, understanding j for a jet coming from light quark (following the notation used in [58]).

The branching fractions of the top partner T' only depend on the two parameters ζ_{jet} and ξ_Z^T, as δ_H is a calculable function of the heavy mass scale for the vector-like quarks M:

$$\delta_H \sim 5\frac{m_t^2}{M^2}\,, \tag{1.18}$$

where only the leading order in $1/M^2$ has been kept. Note that in these formulae the decay rates into Z^0 and H^0 are kept arbitrary (parameterised by ξ_Z^T which is equal to 1/2 for the non-standard doublet case). These approximate but quite robust results allow to describe easily the phenomenology of the non-standard doublet. In figure 1.8 are presented the branching ratios of T' as a function of its mass [59]. It is important to notice that for masses higher than 300 GeV/c^2 the decay channel into top-Higgs becomes dominant.

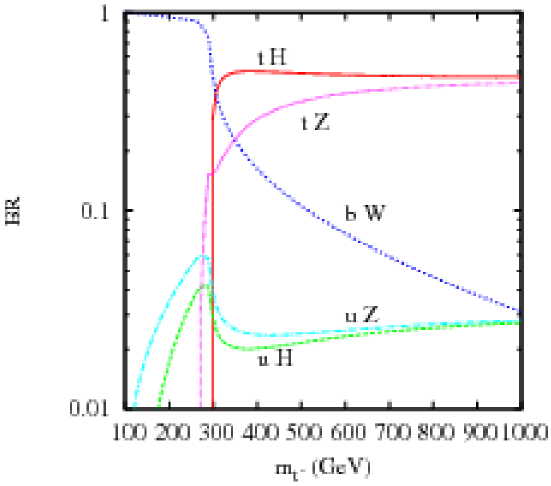


Figure 1.8: T' branching ratios as a function of its mass. The branching ratios correspond to the non-standard doublet with maximized couplings to light generations [59].

In the present work, a search for a T' will be described. This search looked for the T' when it decays into a top quark and a Higgs boson. Therefore, it is important to understand the properties of these particles and how they are produced in the LHC, as predicted by the SM. In the next sections the production of these SM particles at the LHC and their specific properties are going to be presented.

1.5.3 Other models

In the literature there are other approaches for the theoretical formulation of top-partners. They are motivated by a composite Higgs model, in which the Higgs boson is a bound state of the new strong dynamics from the top-partners. Two main examples of such models can be seen in [60, 55]. In this formulation they are naturally coupled with the third generation SM-quarks. However, the production and decay modes presented in last sections remain the same. But the production cross sections change. For the case of the vector-like T' the branching ratios to top-Higgs and top-Z channels is about 50%.

In figure 1.9 are presented the production cross sections for the pair and single production modes in proton-proton collisions at 8 TeV center of mass energy. Differently from the generic model, shown in last sections, the single production mode is dominant over the pair production for masses higher than 700 GeV/c^2. Also comparing the value of the cross sections, the generic model predicts a higher production cross section due to the enhanced couplings of the T' with light generation quarks, in both pair and single production modes.

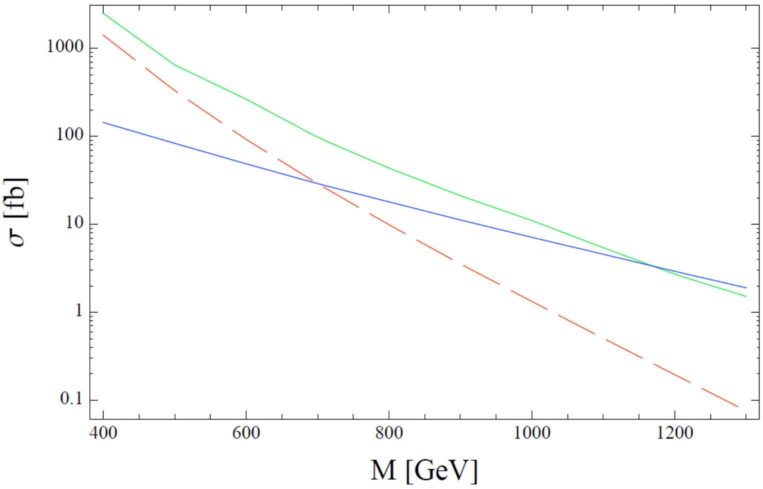


Figure 1.9: T' production cross section as a function of its mass for pair production (red dashed) and single production in association with a b-quark (blue). The green line is the cross section of the single production mode of X in association with a top-quark as a function of its mass [60].

1.5.4 Current experimental limits

VLQ models can be experimentally constrained from many sides. In [59] is shown how measurements involving the determination of the parameters of the CKM matrix constrain the introduction of new quarks that mix with SM quarks. Additionally, at the LHC during run 1, many analyses looking directly for top partners have been conducted by the ATLAS and CMS collaborations. This thesis is the product of one of this efforts.

Current searches have been focused in the pair production channel, existing only one search in the single production mode that was done using the ATLAS experiment. All this searches have set mass exclusion limits depending on the branching ratio of the T' to each decay channel: $bW^{+/-}$, tZ^0or tH^0. Figure 1.10

shows the summary of results from all analysis looking for a T' by ATLAS and CMS collaborations. From there, the highest limits achieved are 900 GeV/c^2 for $bW^{+/-}$ exclusive decay, 850 GeV/c^2 for tH^0 case and 800 GeV/c^2 for tZ^0 channel.

A more complete discussion of the ATLAS and CMS analyses is postponed to section 6.6 along with a comparison of the results from other public analyses and the search presented in this work.

1.6 Top at the LHC

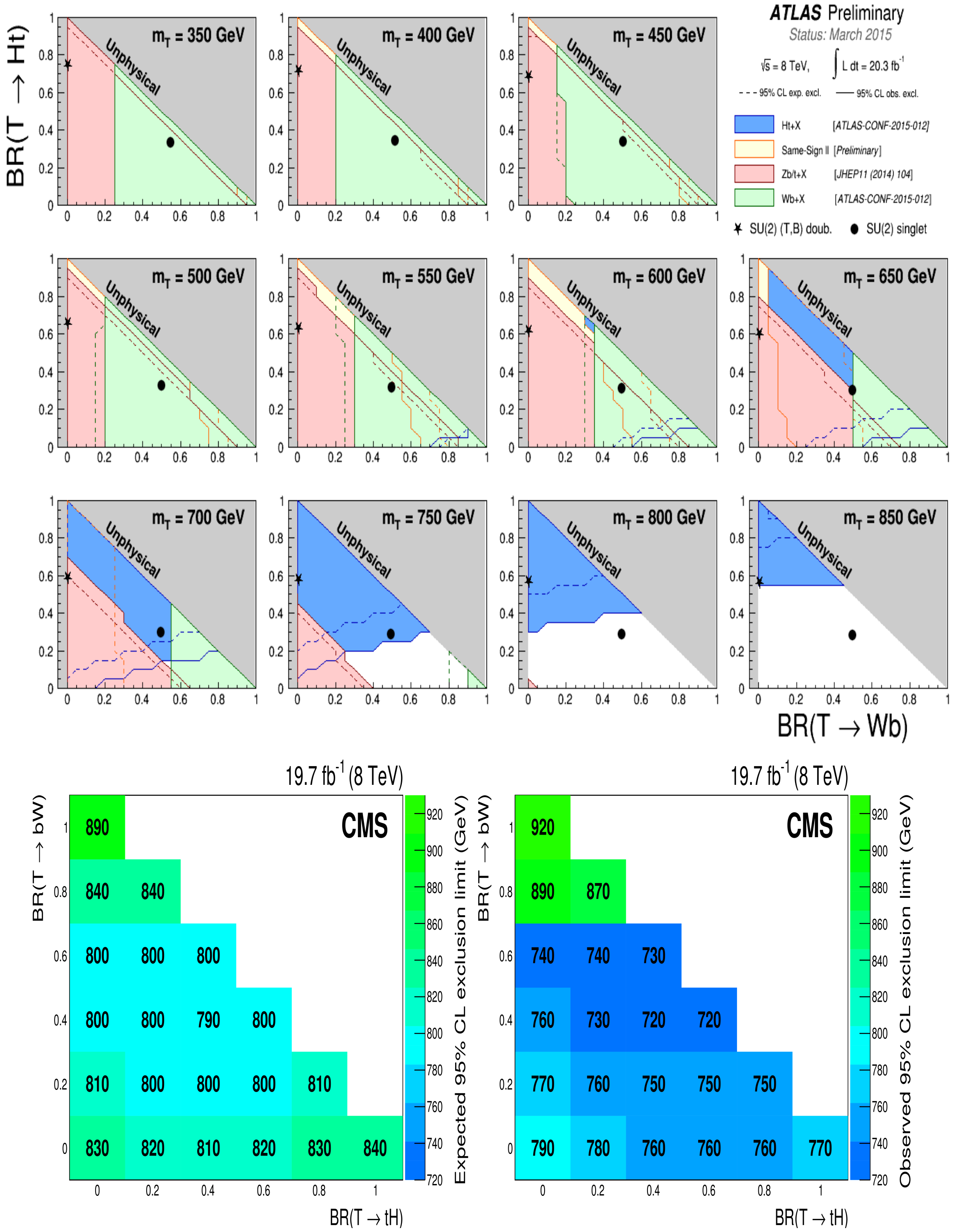
1.6.1 Top production

Discovered in 1995, [63, 64], by CDF [65] and DØ [66] collaborations at the Tevatron [67], it is the heaviest known fundamental particle. As the heaviest particle, many models beyond the SM predict a coupling of the top quarks with a heavier new physics sector. It forms a $SU(2)_L$ weak isospin doublet with the b-quark, discovered in 1977 [68]. Their mass difference, two orders of magnitude, is a fundamental question in the SM. Precision measurements of the top quark characteristics are fundamental inputs to test the SM and new physics searches.

The LHC can be seen as a top-factory, being the accelerator where the most of top-quarks can be produced. The top quark is produced in pairs or in single production mode. During run 1, taking into account 7 TeV and 8 TeV data, 5.6 millions of top pairs events and 2.7 millions of single top events were delivered by the LHC to ATLAS and CMS experiments. Taking into account the different cross sections of the production processes, that will be discussed in the following sections, and the instantaneous luminosity, cited in table 2.1, for 8 TeV center of mass energy at LHC, around 6 tops per second are produced, where 5 of them come from top-pair events and one from single-top events.

Pair production

Pair top-quark production constitutes the dominant production channel in proton-proton collisions at the LHC. Its high cross section makes this process one of the most important background for searches of new physics at the LHC, for example for the search shown in this work. Latest theoretical calculations at Next-to-Next-to-Leading-Order (NNLO) [69], see section 4.1.1, of top pair production from proton-proton collisions at $\sqrt{s} = 8$ TeV give a cross section $\sigma_{t\bar{t}} = 247.47$ pb with around 5% total uncertainty. The tree level diagrams that contribute to this production are summarized in figure 1.11, where gluon fusion dominates the pair production. At $\sqrt{s} = 13$ TeV, as for LHC Run 2, this cross section will scale about 3.3 times. Comparison between theoretical predictions and experimental measurements of top pair production is shown in figure 1.12 [70].



BR(T → bW) \ BR(T → tH)	0	0.2	0.4	0.6	0.8	1
1	890					
0.8	840	840				
0.6	800	800	800			
0.4	800	800	790	800		
0.2	810	800	800	800	810	
0	830	820	810	820	830	840

BR(T → bW) \ BR(T → tH)	0	0.2	0.4	0.6	0.8	1
1	920					
0.8	890	870				
0.6	740	740	730			
0.4	760	730	720	720		
0.2	770	760	750	750	750	
0	790	780	760	760	760	770

Figure 1.10: ATLAS expected and observed T' mass excluded at 95% CL as a function of the branching ratio [top] [61]. CMS expected [left-bottom] and observed [right-bottom] T' mass excluded at 95% CL as a function of the branching ratio [62].

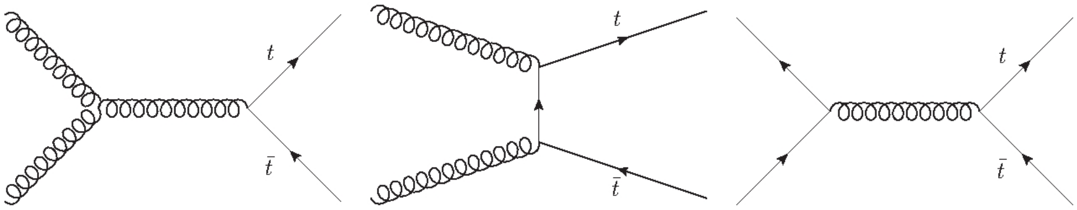


Figure 1.11: Feynman diagrams of top pair production processes for proton-proton collisions, via gluon fusion [left], gluon t-channel [middle] and quark-antiquark annihilation [right].

Single t production

A subdominant production process of top quark is the single production channel, where the top quark is produced in association with another quark, mediated by a W boson, or with a W boson. The Feynman diagrams for tree level single top production can be seen in figure 1.13. NNLO predictions have been also done for these three single production modes at $\sqrt{s} = 8$ TeV. The cross sections are respectively: $\sigma_{t,\,\text{s-channel}} = 5.56 \pm 0.22$ pb, $\sigma_{t,\,\text{t-channel}} = 84.34 \pm 1.69$ pb and $\sigma_{tW} = 22.2 \pm 0.67$ pb. Due to the high cross section of pair production compared to single production channels, the cross section measurement of each production mode has been complicated, due to the expected high number of background events and their small cross section values needing high statistics to be observable. The Tevatron was able to confirm the s and t-channels [73, 74], while at LHC the tW channel was confirmed today [75, 76]. The comparison of theoretical predicted cross section and experimental measurements, by ATLAS and CMS, at different center of mass energies is presented in figure 1.14 [70].

1.6.2 Top decay channels

The top quark decays extremely fast, a process taking only 3×10^{-25} second. Hadronization processes occur around 3×10^{-24} which means that the top quark decays even before hadronization begins. Therefore it is the only quark that can be studied before hadronization, in its free state.
According to CKM matrix (Cabibbo–Kobayashi–Maskawa) [77, 78], that describes how different quarks interact between them, the top quark decays preferentially to a b-quark and a W boson, almost 100% of the times [45].

Then $W^{+/-}$ boson can decay into a lepton and a neutrino, or into a pair of quarks. The branching ratio of leptonic decays of $W^{+/-}$ boson is around 33% while to quarks it is 67% [45]. The hadronic final state is thus the dominant one. In figure 1.15 the two decay modes of top quark with their respective branching

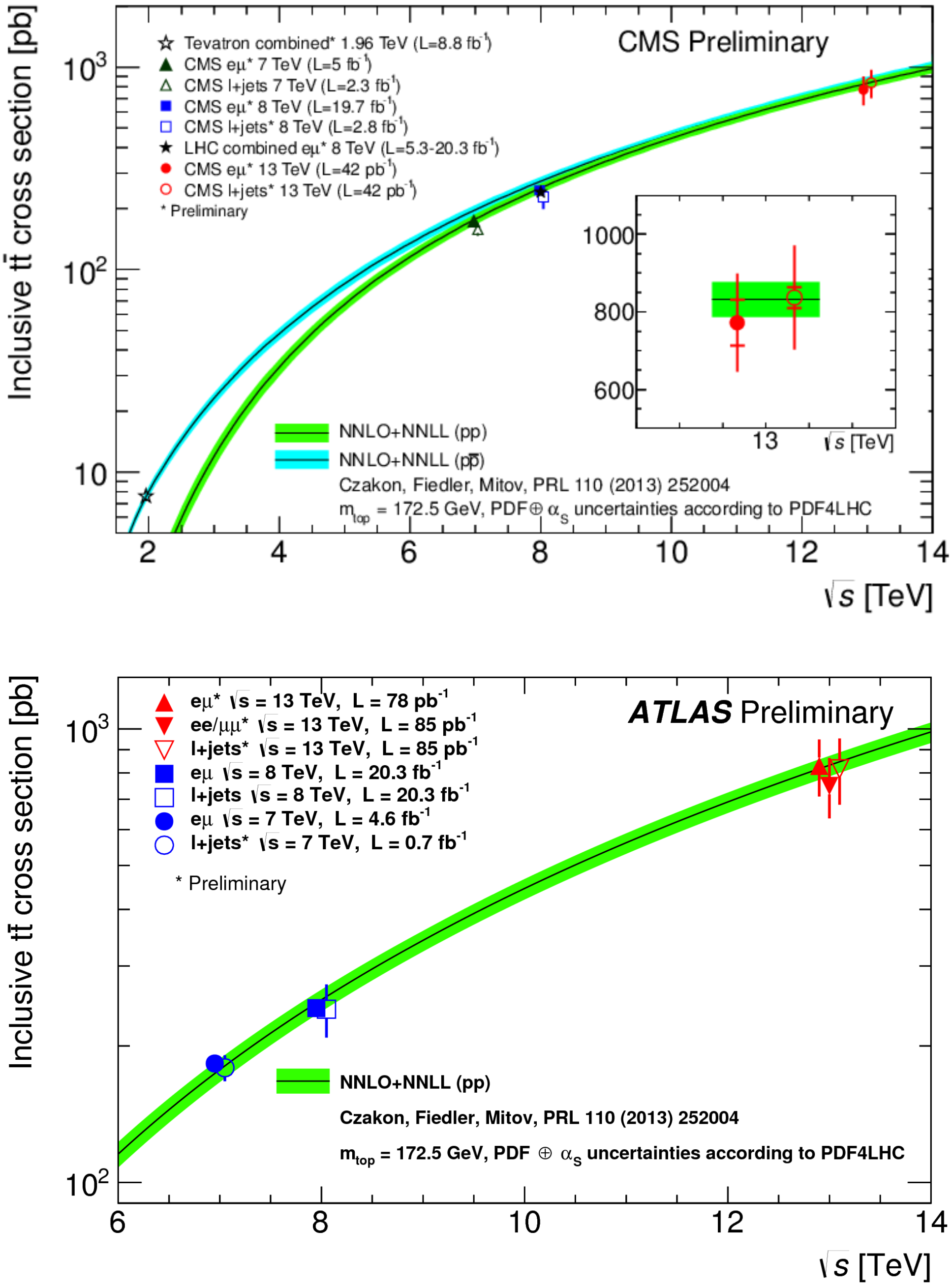


Figure 1.12: $t\bar{t}$ production cross section as a function of the center of mass energy in $p\bar{p}$ and pp collisions compared to theoretical predictions [70, 71, 72].

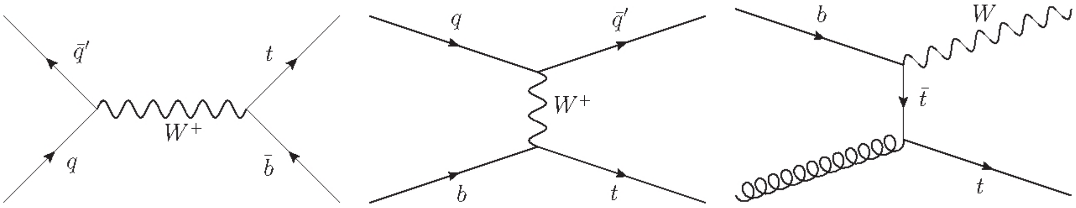


Figure 1.13: Feynman diagrams of single top production processes of proton-proton collisions, from left to right, s-channel, t-channel and associated $W^{+/-}$ production.

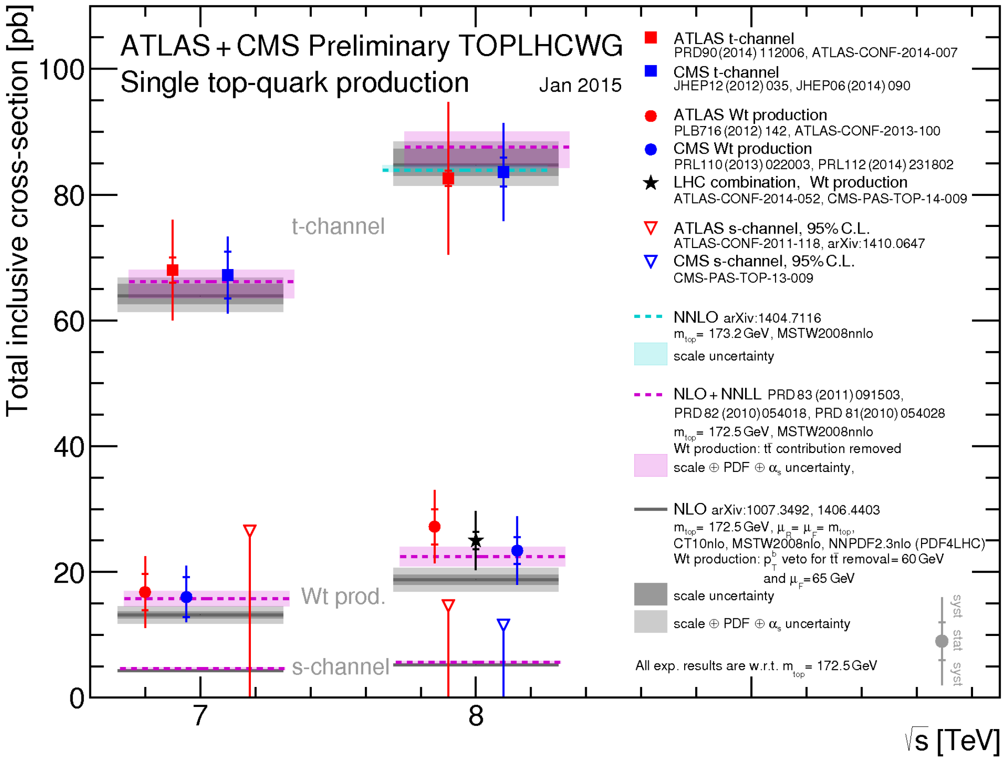


Figure 1.14: Single top production cross section as a function of the center of mass energy in proton-proton collisions compared to theoretical predictions for each production channel by ATLAS and CMS collaborations [70].

ratio are shown.

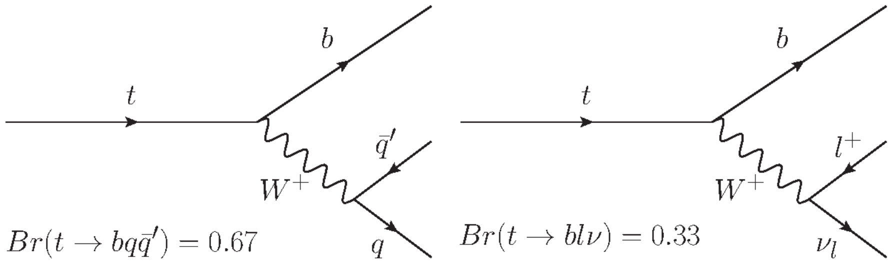


Figure 1.15: Feynman diagrams for top decay channels with respective branching ratios.

1.6.3 Mass and width of top quark

One of the most important quests in top physics is the measurement of top mass and width. They constitute two powerful tests of SM predictions. Whereas the top mass is a free parameter in the SM, it is constrained by other measurements, especially electroweak precision tests. Any deviation with respect to the expectation of the SM could point to the existence of new physics beyond the standard model. The top width is also predicted by the SM.

At Next-to-Leading-Order (NLO) [79], the SM expected value of top width is $\Gamma_t = 1.27$ GeV. It depends of the top mass, the strong force coupling α_S, the mass of the $W^{+/-}$ boson and the strength of the interaction between the top and b-quark. Present measurements have found $\Gamma_t = 2.00^{+0.47}_{-0.43}$ GeV [80], value in agreement with the SM predictions.

Top mass has been measured by ATLAS and CMS collaborations at LHC and by CDF and DØ at Tevatron for each decay channel of top both in single and pair production. Current world combination is $m_t = 173.34 \pm 0.76$. Results from current searches and their combinations are presented in figure 1.16 [70].

1.7 Higgs boson at the LHC

1.7.1 Higgs boson production

The SM Higgs boson is expected to be produced in proton-proton collisions at LHC by four different processes: gluon fusion via the fusion of two gluons via a top loop, Higgsstrahlung where a Higgs boson is radiated from a Z^0 or $W^{+/-}$ boson, vector boson fusion by the fusion of two vector bosons radiated from quarks, and quark

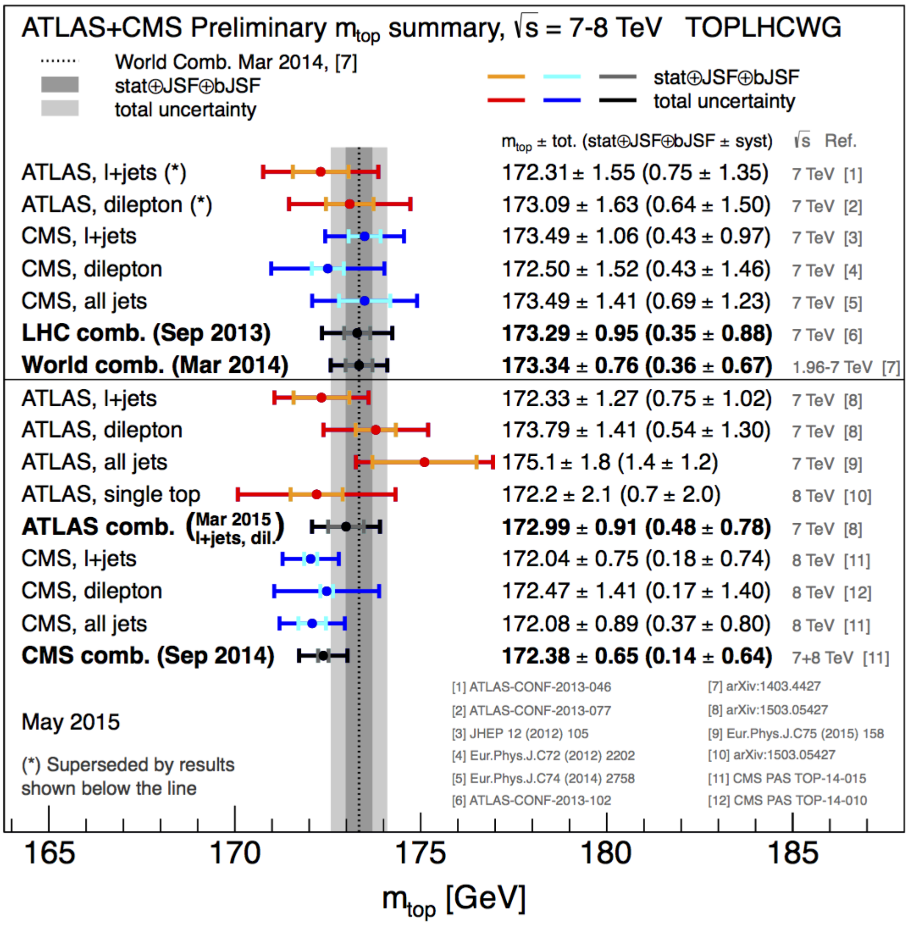


Figure 1.16: Top mass measurements from ATLAS and CMS collaborations and world combination including Tevatron results [70].

fusion from the fusion of quarks produced by a gluon splitting. Feynman diagrams for these processes are shown in figure 1.17.

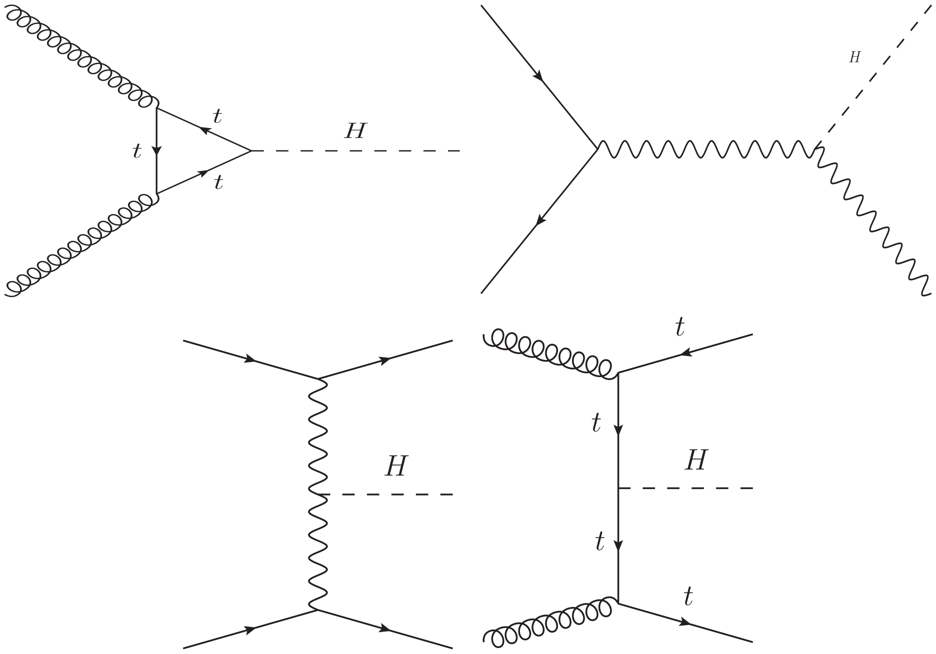


Figure 1.17: Feynman diagrams for Higgs boson production for proton-proton collisions: gluon fusion [left-up], Higgsstrahlung [right-up], vector boson fusion [left-down] and quark fusion [right-down].

All these processes have different contributions to the Higgs boson production cross section. In figure 1.18, the cross section of Higgs boson production for a Higgs mass of 125 GeV/c^2 for different center of mass energies at LHC is shown [81]. The dominant production is the gluon fusion, all the other processes are at least one order of magnitude smaller. The total Higgs boson production cross section at LHC is shown in figure 1.19 [82, 83, 84, 81], as a function of its mass and for different center of mass energies. At run 1 energies, 7-8 TeV, the total production cross section is about 20 pb, while at 13 TeV it will be 2.3 times higher, reaching $\sim$50 pb.

1.7.2 Higgs boson decay channels

As the Higgs boson couples to all massive particles in the SM it can decay in several ways. Figure 1.20, taken from [82, 83, 84, 81], shows the branching ratios of the decay channels of the Higgs boson as a function of its mass. The highest

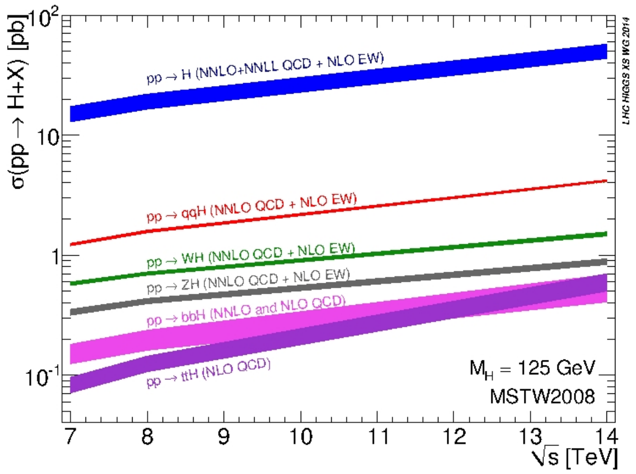


Figure 1.18: Higgs boson theoretical production cross section as a function of center of mass energy for a Higgs boson mass of 125 GeV/c^2 [81].

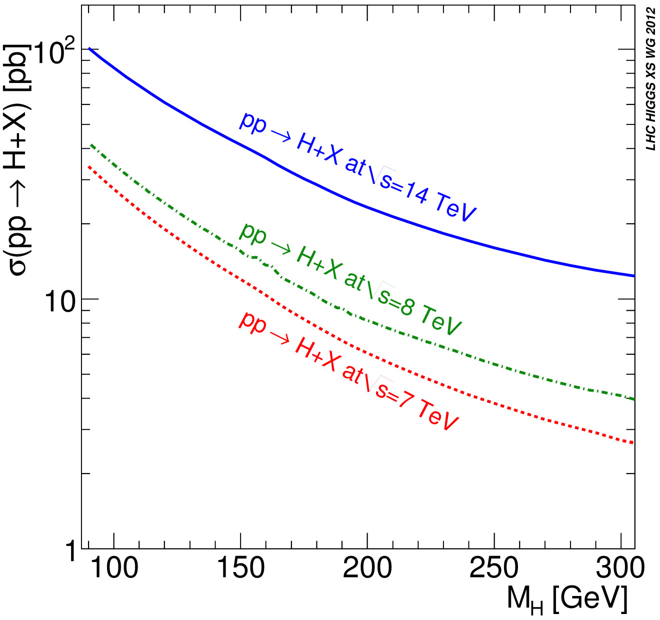


Figure 1.19: Total Higgs boson production cross section as a function of Higgs mass for different center of mass of proton-proton collisions [82, 83, 84, 81].

branching ratio corresponds to its decay to two b-quarks. This channels is difficult to observe due to the high QCD and $t\bar{t}$ backgrounds, however it gives the highest expected number of events as more than half of the Higgs bosons produced at LHC decay into $b\bar{b}$ final state. The Higgs boson decay to Z^0Z^0 is an important channel because it has a high branching ratio and one of its final states is 4 leptons, a very clean final state (low background rate) called the golden channel. Finally, even if the diphoton channel has a especially low branching ratio, it is very clean in terms of mass resolution. In the search that is going to be presented in chapter 5, the Higgs boson is going to be looked for in the $b\bar{b}$ decay. In figure 1.21 the Feynman diagrams of the discussed decay channels are presented.

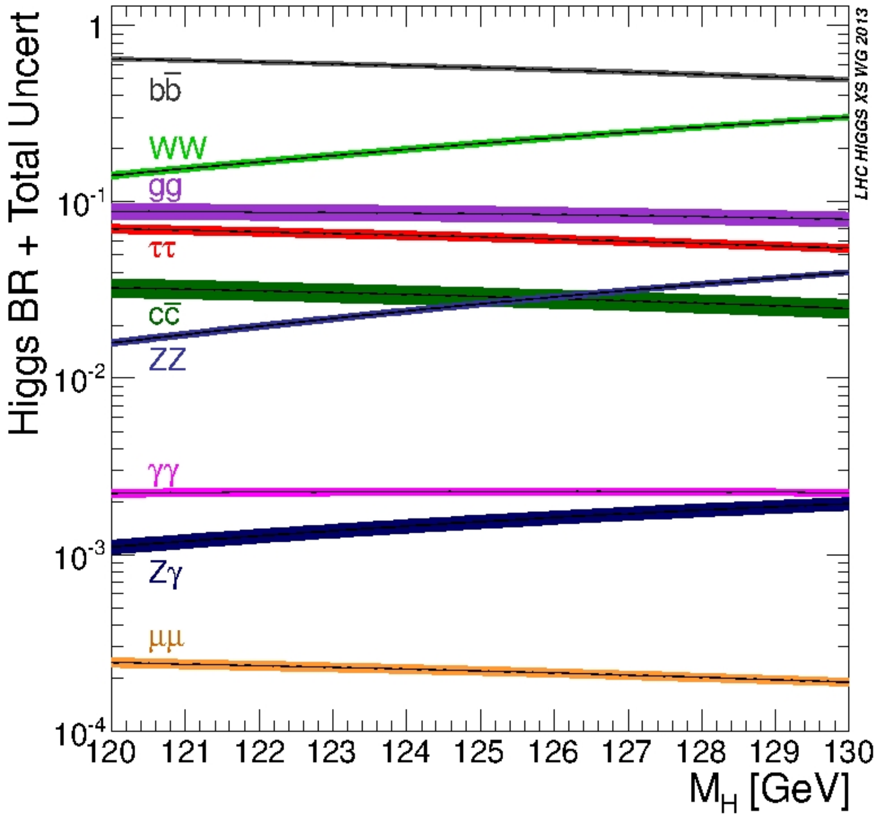


Figure 1.20: Higgs boson decay branching ratios as a function of its mass [82, 83, 84, 81].

The measurement of the branching ratios of the Higgs boson sets important constraint on new physics. For example, a sequential fourth family model has been excluded at a 5.3 sigma level using the signal strengths from searches of

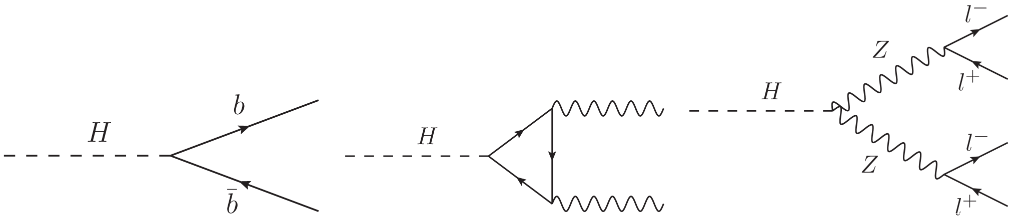


Figure 1.21: Feynman diagrams of Higgs boson decay: $b\bar{b}$ [left], diphoton [center] and golden channels [right].

the Higgs boson in different decay channels and electroweak precision observables in [51].

1.7.3 Mass and width of the Higgs boson

The 4th of July of 2012 was announced the discovery of a Higgs boson like particle at LHC by the ATLAS and CMS collaborations. It was discovered in several decay modes, but the most significant result was coming from the combination of results in all analyzed channels. Such searches have been refined during last years, consolidating the discovery.

To measure the mass of the Higgs boson, the diphoton and golden channel (Z^0Z^0) analyses have been combined between ATLAS and CMS [85]. The obtained value by the combination is $m_H = 125.09 \pm 0.24$. Combination and separated ATLAS and CMS results are shown in figure 1.22 [85, 86, 87]. In the same figure the results are shown, in terms of measured cross section, of all analyses done by CMS looking for a Higgs boson in different final states. The results are compatible with SM expectations.

Whereas nowadays there are plenty of experimental results for the measurement of Higgs boson mass, measuring the Higgs boson width constitutes an important challenge. The predicted width from the SM is ∼4 MeV for a 125 GeV Higgs. Figure 1.23 shows the width dependence on the Higgs boson mass, taken from [82, 83, 84, 81]. In the same figure a method proposed by CMS collaboration using the golden channel to constraint Higgs boson width is shown [88]. Such method set a limit of about 4.2 times the SM Higgs width (∼20 MeV).

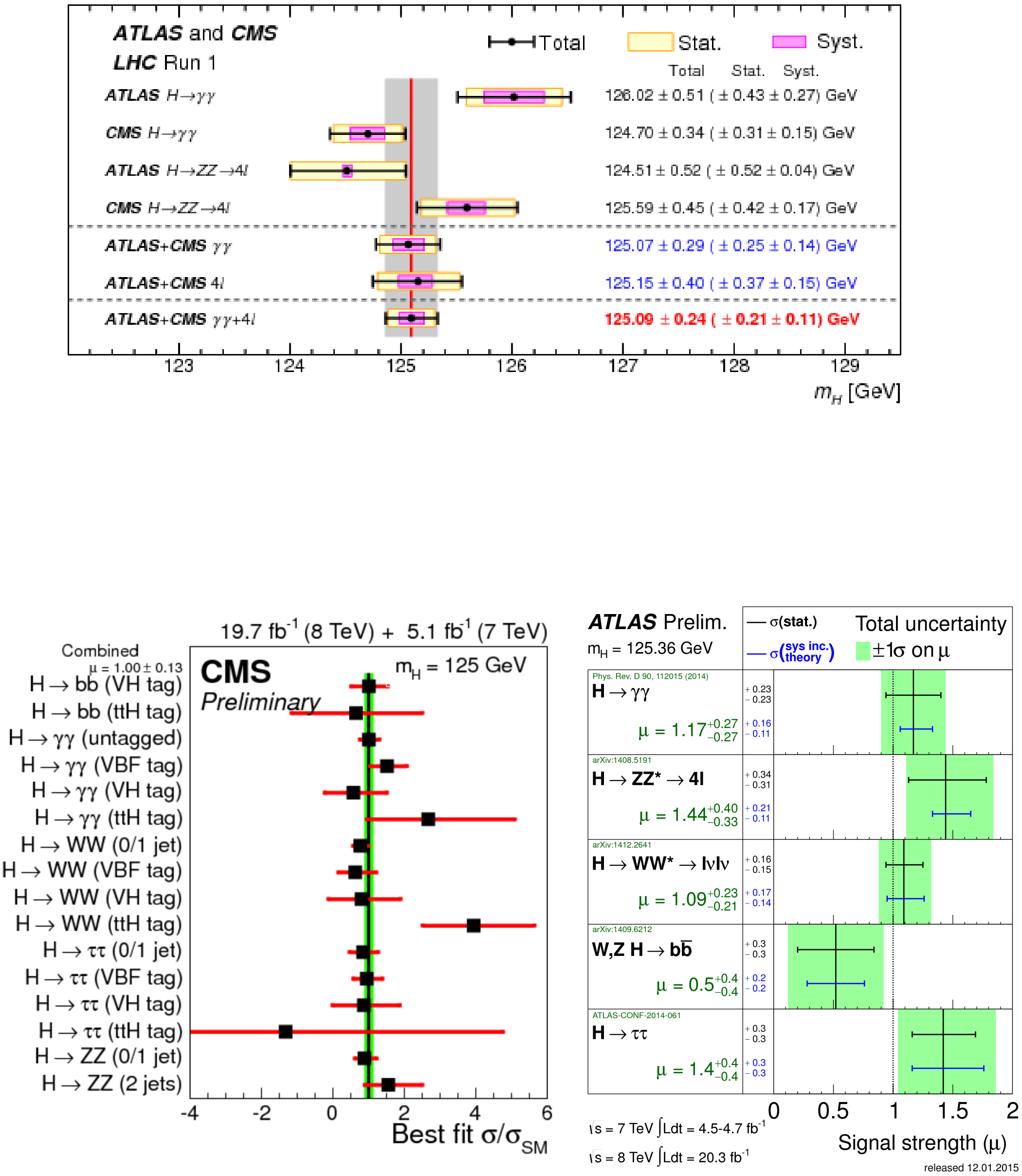


Figure 1.22: ATLAS and CMS combination of Higgs boson mass measurement [top] and σ/σ_{SM} (measured cross section over theoretical SM cross section) for searches performed by ATLAS and CMS in different Higgs boson decay channels [bottom] [85, 86, 87].

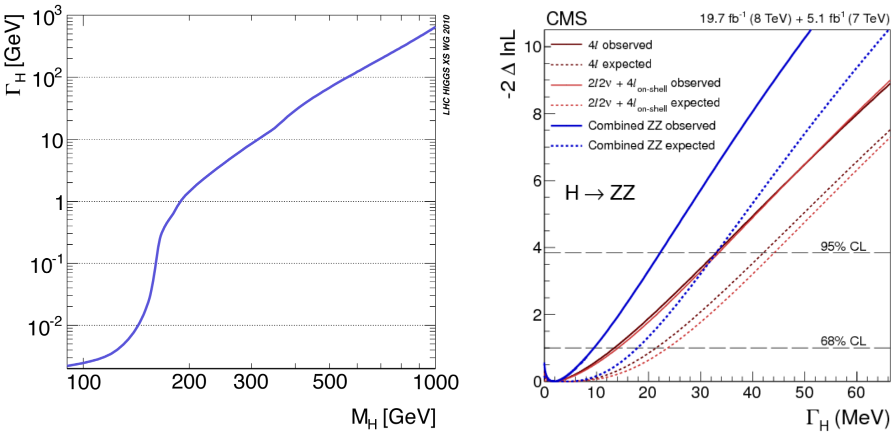


Figure 1.23: Higgs boson width as a function of its mass [82, 83, 84, 81] [left] and current limits from CMS measurement [88] [right].

2 The CMS experiment at LHC

The CMS experiment is one of the biggest particle physics experiments in the world. It is located on the ring of the LHC, which is the main accelerator managed by CERN, the European Organization for Nuclear Research or Conseil Européen pour la Recherche Nucléaire by its french name. This institution constitutes the biggest organization for research in particle physics all over the world. All along its 60 years of existence, since 1954, 21 member states have been part of it, but an overall of 113 countries participate in different ways to this organization.

In the present chapter different aspects of the LHC accelerator and the CMS experiment are discussed. In particular some emphasis is made in the CMS subdetectors related to jets, the objects that play the main role in the search that is the main subject of the present work. The present state of both CMS and LHC experiments, their achievements and the challenges that were overcome are also discussed. Finally, the expectations and goals for the upcoming run 2 are also mentioned.

2.1 The Large Hadron Collider

The Large Hadron Collider, or LHC [89], is a machine that accelerates and collides protons and lead ions. This machine is the biggest particle collider nowadays with a circumference of 27 km. It also achieves the highest collision energy by a collider up to present. The planned energy was 14 TeV at the center of mass of the collision. On the first run of the machine only 8 TeV were achieved, and the next run started with 13 TeV. The LHC is located on French-Swiss border near Geneva. The tunnel for the machine was carved around 100 m under the ground, 45 m under the Jura mountains and 170 m near the Léman lake with an inclination of around 1.4%, sloping down towards the lake. This machine has used as much as possible the old LEP buildings and sites. LEP was an electron-positron collider built between 1984 and 1989.

The protons and heavy ions accelerated by the machine collide in different points, where dedicated experiments are located to detect and to study the product from the collisions. The four main experiments located on the LHC ring are ALICE [90], ATLAS [91], CMS [92] and LHCb [93]. ALICE focuses on the study of the quark-gluon plasma produced in heavy ions collisions. ATLAS and CMS are experiments of generic purpose where searches for new physics and also precision measurements are performed. LHCb is dedicated to the physics of the b-quark. Even if one of the principal objectives of the construction of the LHC was the search for the Higgs boson, generic searches on new physics have been conducted from the very beginning of the first data taking in 2009.

The LHC is a complex machine composed of several parts. The two principal parts are the injector chain and the main ring. A diagram of the whole CERN accelerator complex is shown in figure 2.1. The injector chain has different stages that pre-accelerate protons and heavy ions to be injected into the main ring of LHC.

2.1.1 Injector chain

The injector chain begins with the proton source. Protons are extracted via the ionization of hydrogen gas in the Duoplasmatron Proton Ion Source. Such extraction is pulsed, which makes up the first bunch structure. The extracted protons are then accelerated up to 50 MeV in the linear accelerator, Linac2, that dates from 1978. After this first stage several steps are followed:

1. Linac2 injects proton bunches in the Proton Synchrotron Booster (PSB) where they are accelerated up to 1.4 GeV.

2. From PSB, the protons are delivered to the Proton Synchrotron (PS) where they reach an energy of 28 GeV. In the PS the bunches are also split from 6 initial bunches to 72, spaced by 25 ns.

3. Finally, the pre-acceleration chain is finished by the SPS, the Super Proton Synchrotron. There, the bunches are accelerated up to 450 GeV right before being inserted into the main LHC ring.

The whole pre-acceleration chain has been optimized to obtain the best possible performance on the final acceleration in the LHC main ring. All parameters are carefully controlled, for example the number of bunches, the separation between bunches, the separation between trains of bunches or the injection energy to each subsystem. The level of control achieved in the bunches manipulation is also remarkable, from old subsystems as the PS from 1959 or the newest, the SPS that dates from 1976.

2.1.2 LHC main ring

The main ring is composed of two rings that accelerate the proton bunches in opposite directions, clock-wise and counter clock-wise. A schematic view of the

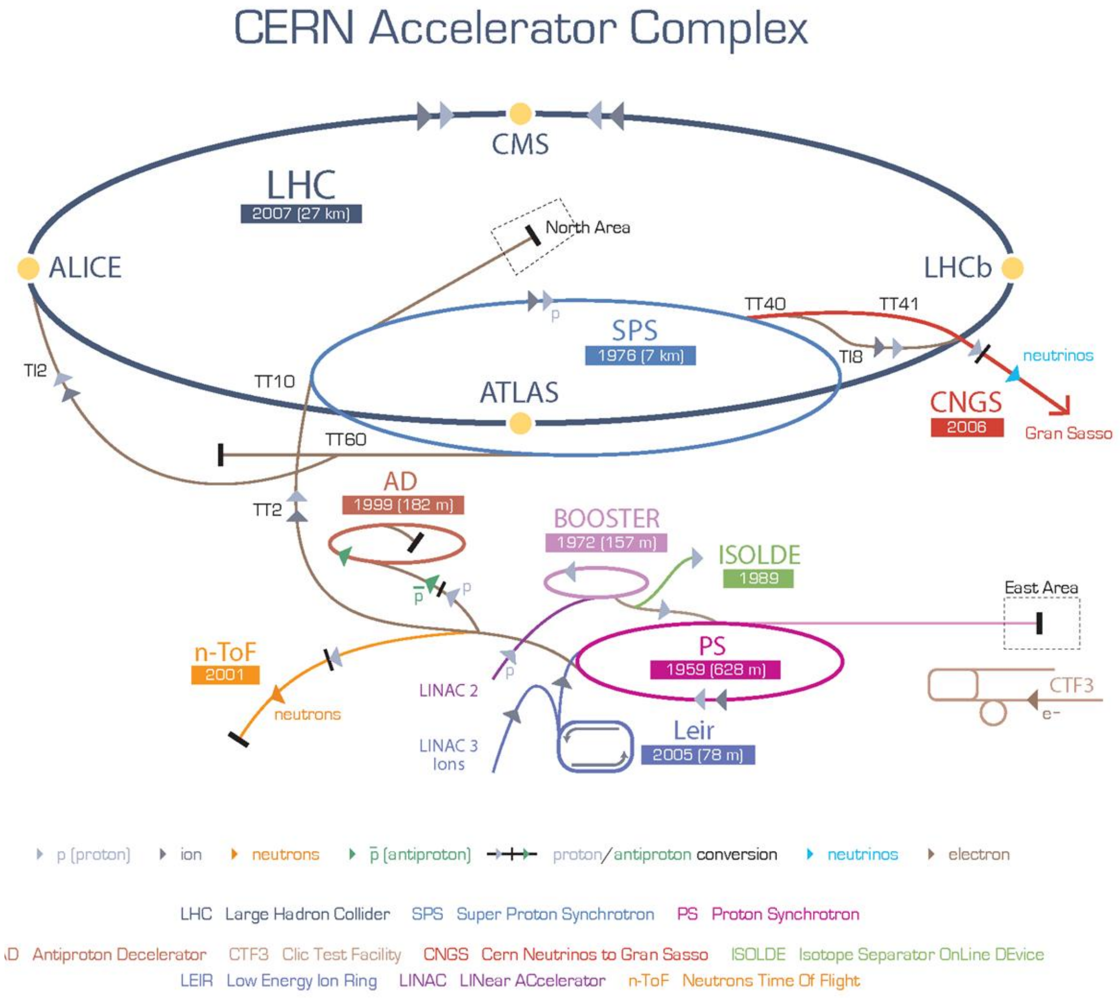


Figure 2.1: Organization of CERN accelerator complex.

design of the main ring is shown in figure 2.2. The rings cross in different points in order to collide the protons and they are divided in eight straight sections and eight arcs. In each octant bunches are controlled by dipole magnets. These complex magnets, shown in figure 2.3, need to produce a very strong magnetic field in order to be able to bend a 8 TeV beam of protons. This intense magnetic field of 8.33 T produced in opposite directions, is produced by electrical currents that are only achievable by means of superconductivity. All the 1232 dipoles operate at a temperature of 1.9 K, under cooling by liquid helium. They also operate under ultra-high-vacuum: the beam lines operate with a pressure less than 10^{-9} mbar and the whole dipole system with 10^{-6} mbar. This vacuum serves also as insulating system from the surroundings. In addition, the LHC main ring has other magnets that focus and correct different characteristics of the beam: 520 quadrupoles, 2464 sextupoles, 1232 octupoles.

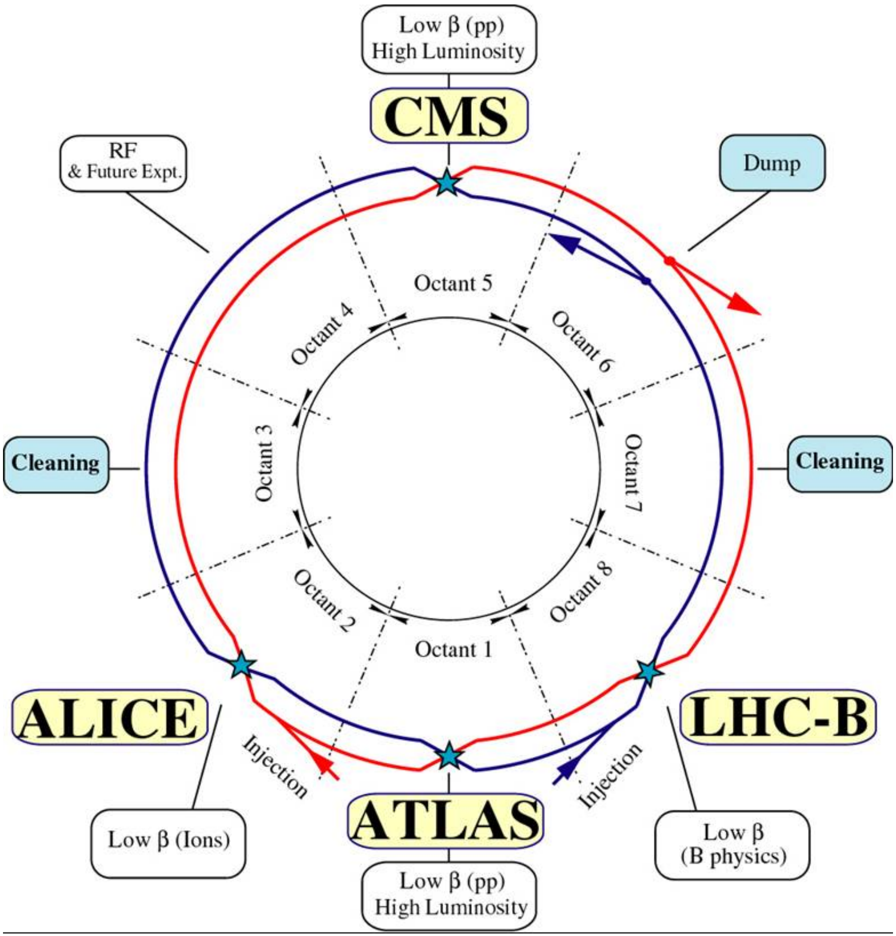


Figure 2.2: Schematic of the LHC main ring design.

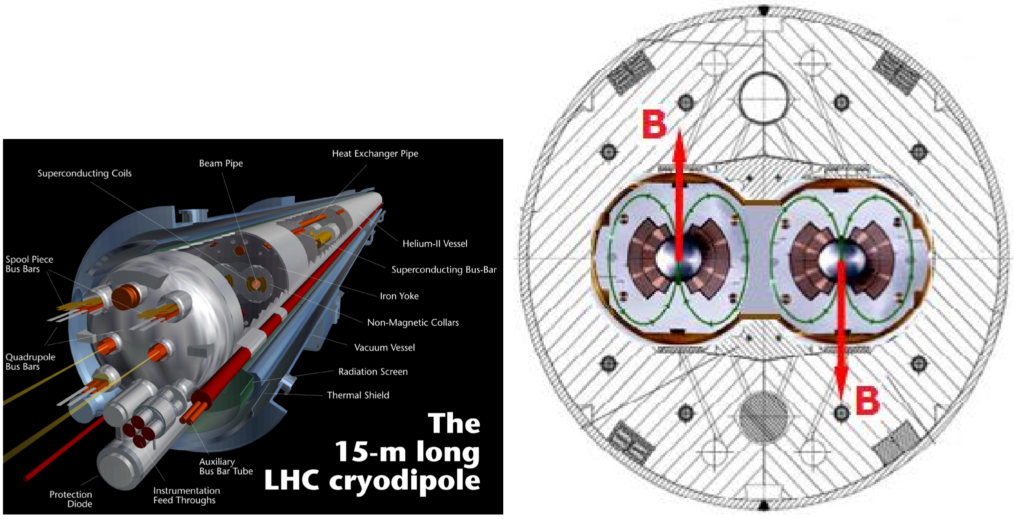


Figure 2.3: Design of LHC cryodipole and the magnetic field that bends the beam in the main ring.

Luminosity

In collider physics, the figure of merit is the luminosity, given in equation 2.1.

$$L = \frac{k_b N_b^2 f_{rev} \gamma}{4\pi \epsilon_n \beta^*} R = \frac{k_b N_b^2 f_{rev}}{4\pi \sigma_x^* \sigma_y^*} R \tag{2.1}$$

The number of events per second is proportional to the luminosity, hence is the quantity to be maximized by the design and operation of the accelerator. The collider characteristics depend on the number of bunches in the ring k_b, the number of protons per bunch N_b, the revolution frequency f_{rev}, the relativistic gamma factor γ, the normalized rms transverse beam emittance ϵ_n and the beta function at the interaction point β^*. The denominator on 2.1 can also be rewritten in terms of the horizontal and vertical width of the bunches at the crossing, σ_x^* and σ_y^*. In addition, there is the geometric reduction factor (R) that introduces a dependence with the crossing angle of the bunches at the interaction points, with a value around 0.8 for nominal LHC conditions. The crossing angle refers to the angle at which the bunches are crossed in the collision points of LHC. In table 2.1 the, LHC beam parameters at injection and collision are presented.

At the crossing points, the number of events coming from collisions and produced via a specific process, is directly proportional to the luminosity provided by the collider, as in equation 2.2. The constant of proportionality is the cross section of the process, noted as $\sigma_{process}$.

Table 2.1: LHC proton beam designed parameters.

Parameter/units	Injection	Collision
Energy [GeV]	450	7000
Luminosity [cm^{-2}s^{-1}]		10^{34}
k_b Number of bunches	2808	
Bunch spacing [ns]	24.95	
N_b intensity per bunch [protons/bunch]	1.15×10^{11}	
Beam current [A]	0.58	
ϵ_n normalized rms transverse beam emittance [μm]	3.5	3.75
f_{rev} revolution frequency [kHz]	11.25	

$$N_{events} = L\sigma_{process} \tag{2.2}$$

The total cross section of a proton-proton collision from the crossing of two bunches at 14 TeV is 100-110 mb [94], from three different scattering processes: elastic, diffractive and inelastic. In the elastic scattering the protons only exchange momenta but their structure remain unchanged, that is the case for the majority of the collisions. In diffractive scattering processes, not only momentum is exchanged but also new particles are produced in addition to the two final protons. Finally, in inelastic scattering, the constituents of the protons, the partons, interchange a big amount of momentum and produce a large quantity of particles. The inelastic processes contribute less than the diffraction to the total cross section. While inelastic collisions produce particles mainly in the central rapidity (defined in 2.2.1) region, diffractive and elastic final products have a large rapidity. Only in the hard interactions, inelastic scattering, the central rapidity region is filled up. In proton-proton collisions additionally color is exchanged.

From the crossing of two bunches not only one proton-proton interaction is expected. The number of expected interactions in each crossing depends on the number of protons per bunch and the geometric characteristics of the bunches at their crossing. From LHC design, a mean number of 20 pileup interactions are expected at 25 ns bunch space for 14 TeV. From them, only one is coming from an inelastic collision with a high exchange of momentum, which is the type of process of greater interest for detectors as ATLAS or CMS. This fact puts an additional difficulty to the detectors design in order to extract the hard interaction from all the elastic and diffractive collisions happening at the same time. Such phenomenon is known as Pile-Up, an illustration of a collision with high pile-up can be found on figure 2.4 as seen by the CMS detector.

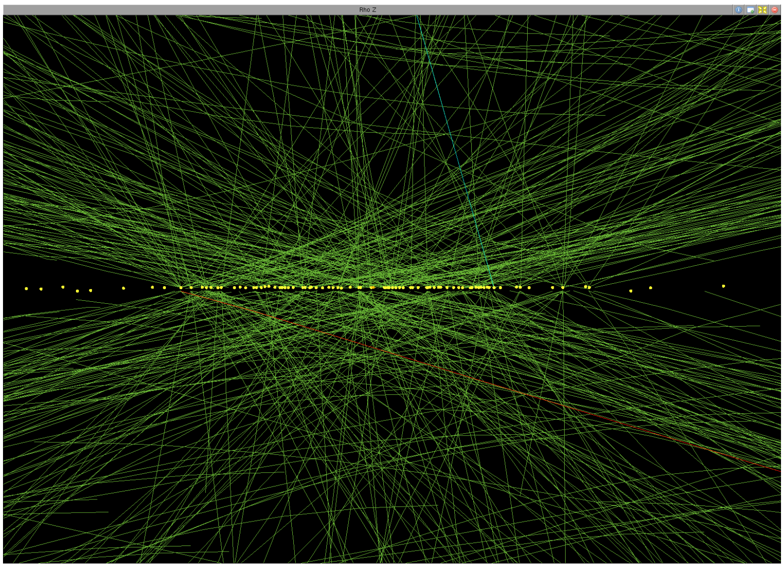


Figure 2.4: High pile-up event (78 interactions) seen by the CMS detector. Event 35655522 from run 198609 and lumi 56, recorded in 2012.

2.1.3 Run 1

On February 10th of 2013, the first stable run of the LHC reached an end. This run, now called Run 1, started on November 20th of 2009. LHC was originally planned to start in 2008, but an incident on one of the electric connections of one of the magnets forced to stop on the 19th of September of the same year. Since the restart in 2009, the energy was augmented from 450 GeV to 4 TeV per beam. The 23rd of September 2009, the first collisions were detected by the experiments. One week after, the achieved center of mass energy was $\sqrt{s} = 2.36$ TeV, already higher than the Tevatron (1.96 TeV).

In 2010, between 30th March to 6th December, 3.5 TeV per beam were reached delivering near 50 pb^{-1}. With the same energy, approximately 6 fb^{-1} were delivered in 2011.

In 2012, the center of mass energy reached one additional TeV, $\sqrt{s} = 8$ TeV, and around 20 fb^{-1} of integrated luminosity were delivered between April and December. The data produced in this period is going to be used in the analysis presented in chapter 5. Figure 2.5 shows the progress of the recorded luminosity by CMS for the 2010-2012 period. The first six weeks of 2013 were devoted to proton-lead collisions.

After this very successful run, the LHC was stopped for about 2 years for repair and maintenance of different systems in the experiments and in the LHC itself to achieve higher energies. After this period, known as the Long Shutdown 1 or LS1, the LHC restarted a new run in spring 2015.

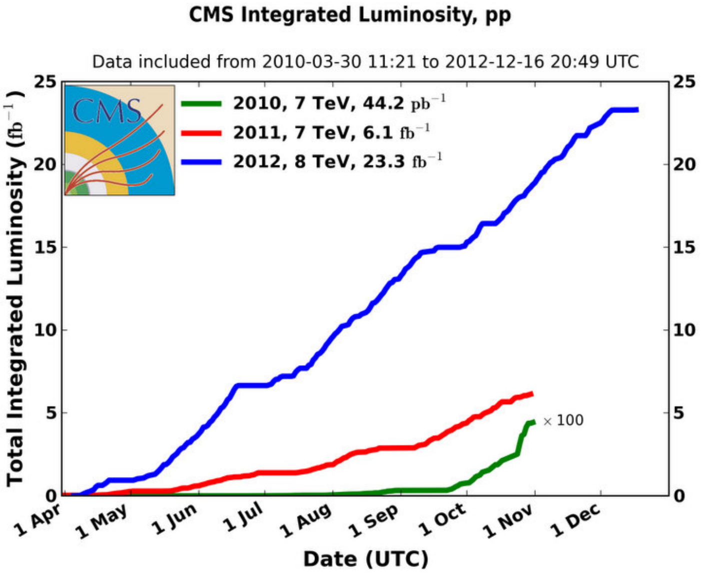


Figure 2.5: CMS integrated luminosity for proton-proton collisions delivered by LHC.

2.1.4 Other experiments at the LHC

The biggest LHC experiments are ATLAS and CMS, both of them generalist experiments designed to do precision measurements as well as new physics searches. Mainly recording proton-proton collisions, they have also recorded lead-lead and proton-lead collisions during the run 1. Both of them were designed for high instantaneous luminosity, $L = 10^{34}\mathrm{cm}^2\mathrm{s}^{-1}$.

In addition, there are two other experiments designed for specific purposes. ALICE built for the study of strongly interacting matter and LHCb that focuses on the study of the physics of the b-hadrons, specially related to the CP violation. The first of them record proton-proton collisions at an instantaneous luminosity of $10^{32}\mathrm{cm}^2\mathrm{s}^{-1}$ and the second record ion-ion collision with $L = 10^{27}\mathrm{cm}^2\mathrm{s}^{-1}$.

The CMS experiment is going to be described in detail in section 2.2. In the following sections the other three experiments mentioned above are going to be briefly presented.

ALICE

The ALICE experiment (A Large Ion Collider Experiment) is located at point 2 of the LHC main ring. It measures 16 m high, 16 m wide and 26 m long, and weights 10000 tons. Designed for heavy ion physics, it is able to detect an extremely high number of tracks per event. Its main subsystem is the Time Projection Chamber (TPC), a 90 m^3 gas chamber filled with a mixture of Ne, CO_2 and N_2 operated in a solenoid of 0.5 T. It allows to measure leptonic and hadronic charged particles in a momentum range from 0.5 to 10 GeV/c. The experiment structure can be seen on figure 2.6. ALICE collaboration counts around 1500 people, from 154 physics institutes in 37 countries.

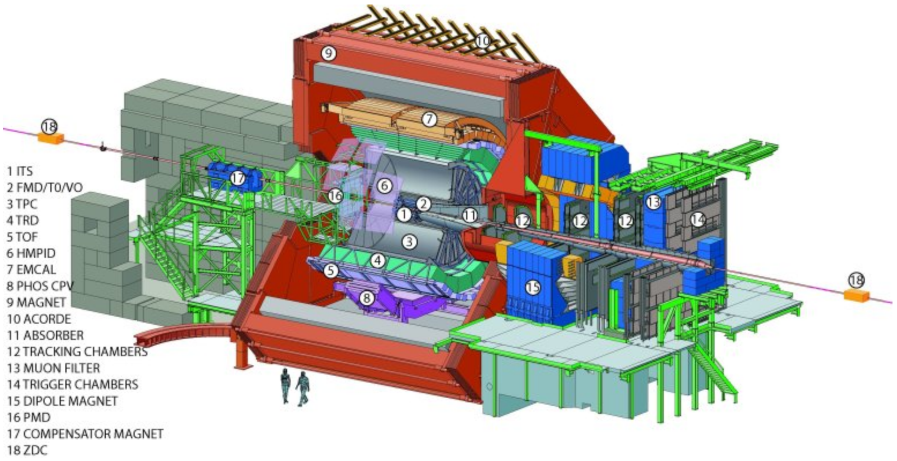


Figure 2.6: The ALICE detector.

ATLAS

The ATLAS experiment (A Toroidal LHC ApparatuS) is the biggest LHC experiment. It is located at point one, as displayed on figure 2.2, on the LHC main ring. It is a cylindrical detector similar to CMS, about 45 meter long, 25 meter high, and weights around 7000 tons. ATLAS main components are, from inside to outside, a tracking system, an electromagnetic calorimeter, a hadron calorimeter and muon chambers. In between these subsystems there is an internal solenoidal magnet and a set of external toroidal magnets. Its main technology for calorimetry is liquid Argon. The detector design is presented on figure 2.7.

ATLAS experiment configures a collaboration of around 3000 persons, coming from 117 universities around the world, from 38 countries.

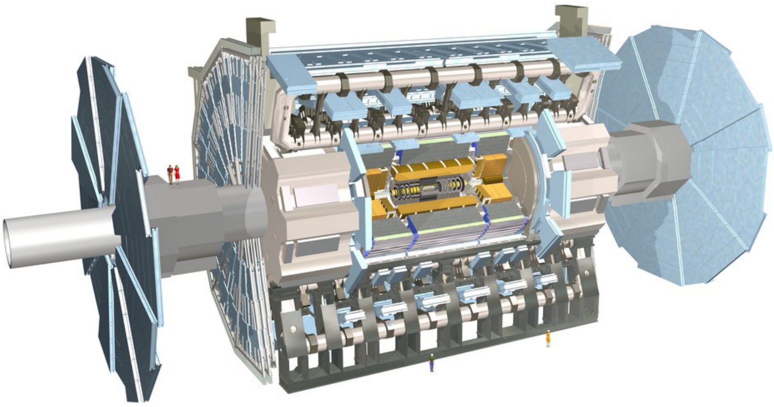


Figure 2.7: The ATLAS detector internal view.

LHCb

LHCb detector, hosted at point 8 of the LHC main ring, has a different design than ATLAS and CMS. Smaller than these experiments, it has been designed to be able to detect particles produced close to the beam direction. This is the reason why it is not cylindrically but conically shaped, as shown in figure 2.8. It also has the same main parts: a tracking system, electromagnetic and hadron calorimeters, muon chambers and magnets. Its major specificity is a system that allows to identify different hadrons, the RICH detectors, a crucial feature for the study of strong interacting matter. In addition, the tracker system counts with a very precise vertex locator system. The whole experiment measures 21 m long, 10 m high and 13 m wide, and weights 4500 tons. The LHCb collaboration groups around 700 persons from 69 different universities over 17 countries.

2.2 The Compact Muon Solenoid (CMS) experiment

The CMS detector, hosted at point 5 of the LHC main ring (see figure 2.2), is the second biggest LHC experiment. Cylindrically shaped, it measures 15 m of diameter and 28.7 m long, and it weights 14000 tons, making it the heaviest LHC experiment. Its subsystems are concentrically organized around the beam line. It is called compact because the whole calorimetry is inside the solenoid magnet, and muon solenoid because it has a very precise muon detection. Its main characteristic is the strong 3.8 T superconductor solenoid magnet. A representation of the detector can be found in figure 2.9. The CMS collaboration is formed by around

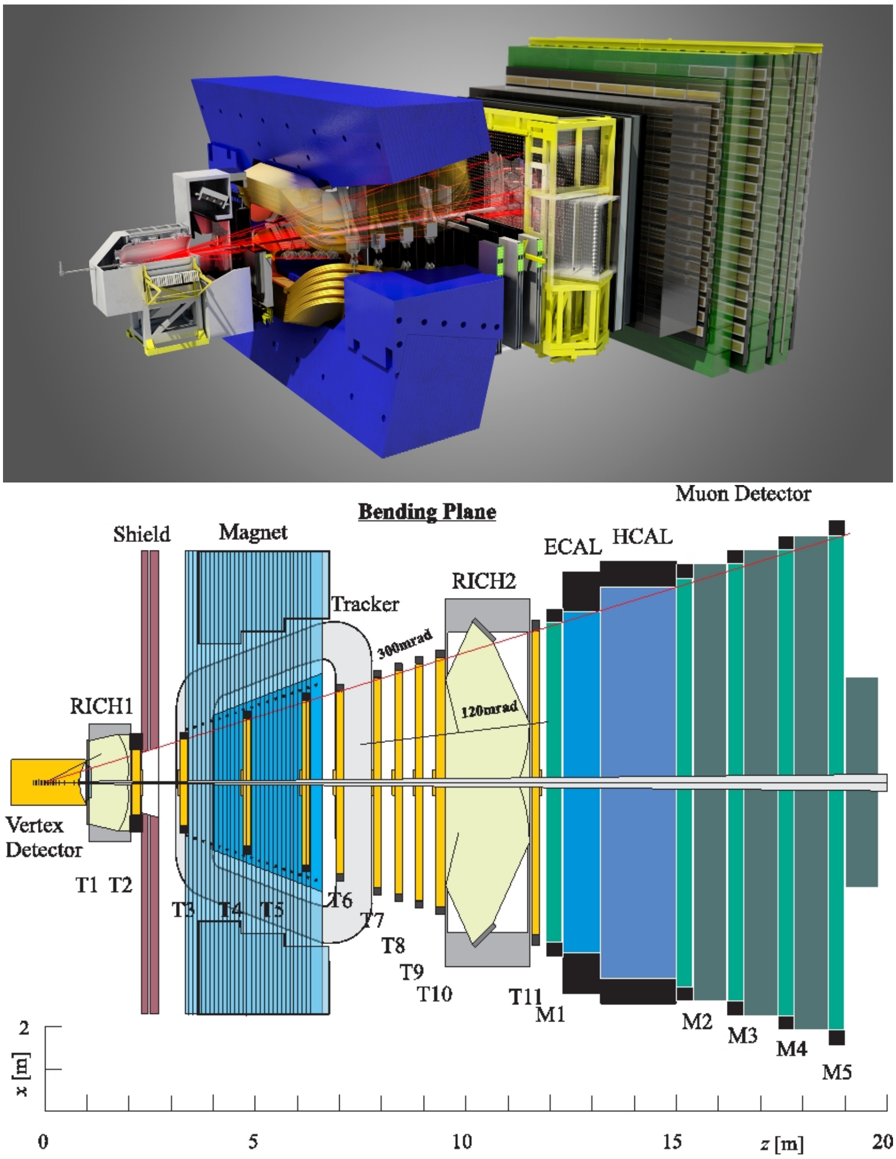


Figure 2.8: LHCb detector 3D view [top] and view from the top [bottom].

3500 scientists from 181 institutes over 41 countries.

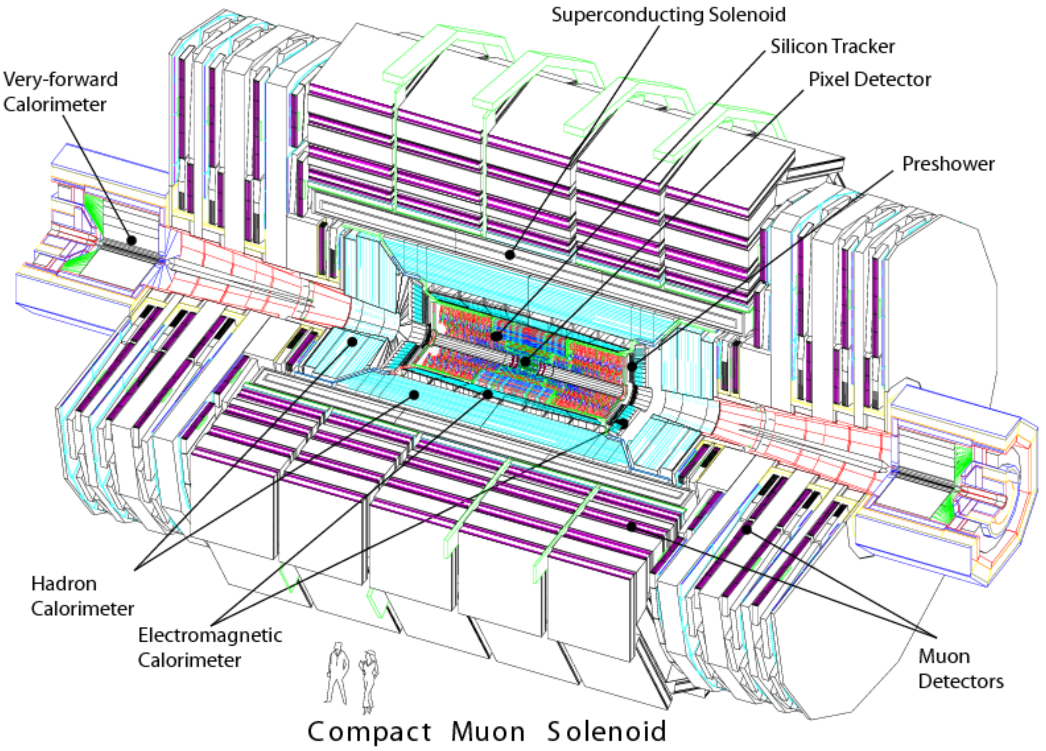


Figure 2.9: The CMS detector.

CMS has been designed to be able to do very precise identification of particles and measurements of their properties. For example, the energy of jets can be determined with an uncertainty smaller than 4% for jets with a $p_T > 20$ GeV/c. Details on other objects will be given later in this section. For the measurement of the momentum of the charged particles, CMS has a very powerful magnet that allows to bend very energetic particles. In addition, the calorimeters allow to measure accurately the energy from hadrons, electrons and photons. At the most external layer, the muons chambers measures muons properties, and in the innermost the tracking system reconstructs the collision points and the charged particles tracks. In figure 2.10 a representation of the different subsystems of the CMS and how particles are reconstructed from them is displayed.

In the following subsections the different subsystems of the CMS experiment are described. In first place, the definition of the coordinate system followed by the description of CMS magnet are given. As last point, a description of the trigger system is presented.

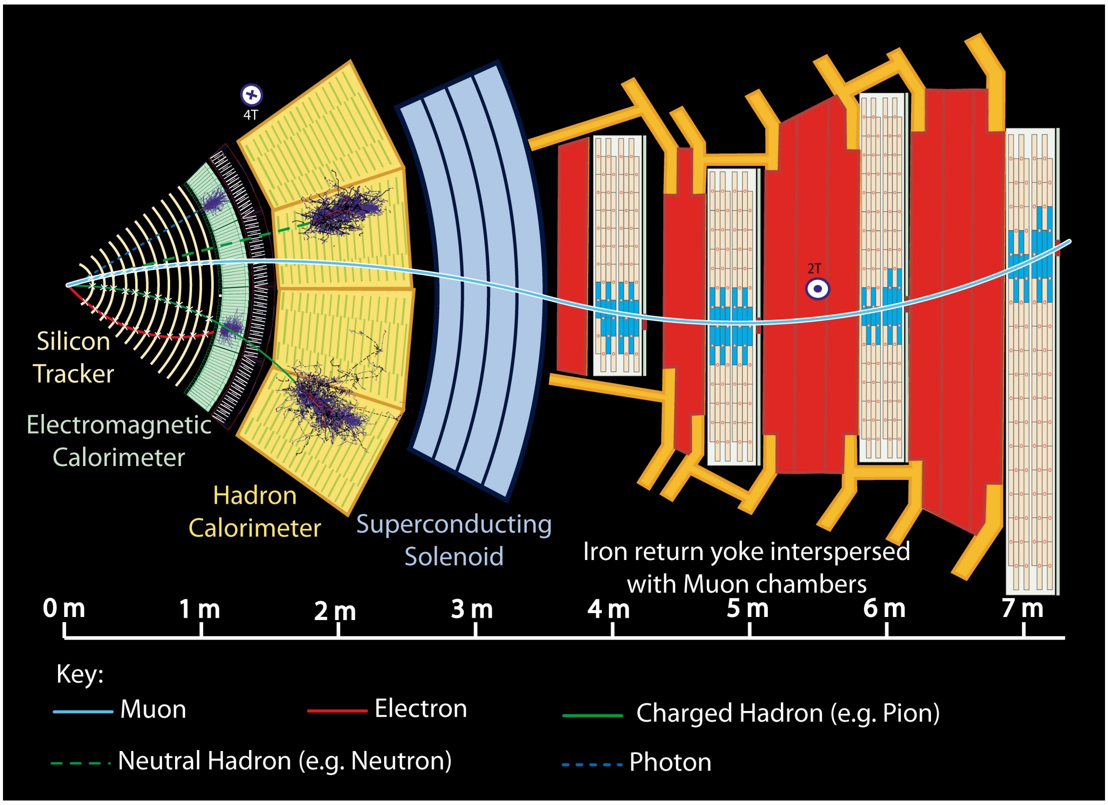


Figure 2.10: CMS sub-detectors and particle identification.

2.2.1 Coordinate system

The origin of coordinates for the CMS detector is located at its center. This center is the nominal collision point, the “interaction point”. From there, the z-axis is defined along the beam pipe line pointing towards the Jura mountains. The positive/negative z-axis directions define the positive/negative sides of the detector. The y-axis is defined towards the zenith and the x-axis towards the center of the LHC ring. Due to the inclination of the LHC plane, this coordinate system is slightly tilted with respect to the true vertical. A representation of the coordinate system definition can be found in figure 2.11.

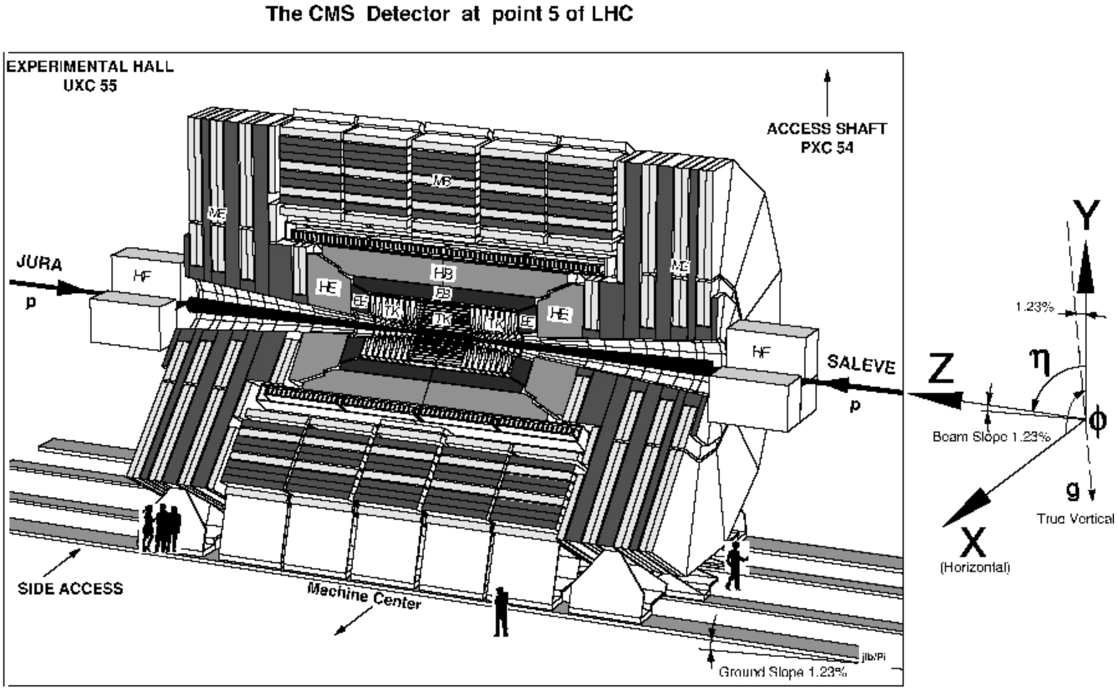


Figure 2.11: The CMS coordinate system.

Two angles are defined: the ϕ angle in the x-y plane from the x-axis towards the positive y-axis, and the θ angle in the z-y plane from z-axis towards the positive y-axis. In experimental particle physics, it is preferred to work with relativistic invariant quantities, reason why instead of working with θ, the pseudorapidity η is defined (equation 2.3).

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{2.3}$$

Another relativistic invariant quantity can be defined: the rapidity y as from equation 2.4. With $\vec{p}$ being the momentum vector and E the energy of a given particle, p_L denotes its longitudinal component, that in our case is the same as the z-component.

$$y = \frac{1}{2}\ln\left(\frac{E + p_L}{E - p_L}\right) \tag{2.4}$$

On the limit that the mass of the particle is very small compared to its momentum, the particle energy can be approximated by the momentum magnitude, giving rise to the definition of the pseudorapidity in terms of the momentum of the particle $\eta = \frac{1}{2}\ln\left(\frac{\boldsymbol{p}+p_L}{\boldsymbol{p}-p_L}\right)$

The radial coordinate is defined over the x-y plane, plane that is called the transverse plane being orthogonal to the longitudinal direction, the z-axis. In such a plane are also defined the transverse quantities of particles, as the transverse momentum p_T. Finally, for any two objects, an angular distance can be defined in the $\eta - \phi$ plane, as in equation 2.5.

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.5}$$

2.2.2 Magnet

In order to measure the momentum of the charged particles going inside the detector, it is crucial to apply a magnetic field. The magnetic field should be able to bend particles sufficiently for the detector to measure the bending. Therefore, a very strong magnetic field is needed for very energetic particles. The momentum of a charged particle inside an uniform magnetic field can be written as

$$p = \gamma m v = qBr \tag{2.6}$$

where B is the magnitude of the magnetic field, γ the usual relativistic factor, m the mass of the particle, v its rapidity, q its charge and r the bending radius. The sagitta of the arc is

$$s = \frac{L^2}{8r} = \frac{qBL^2}{8p}, \tag{2.7}$$

with L the trajectory length that the particle moved inside the magnetic field. Inside a solenoid L is equal to its radius.

From relation 2.7 it is possible to deduce that the resolution on the momentum of the particle has an inverse dependence with the magnetic field and the radius of the solenoid, as shown in equation 2.8.

$$\frac{dp}{p} \propto \frac{p}{BL^2} \tag{2.8}$$

The design of the CMS magnet targets both features, a large solenoid and a strong magnetic field, utilizing a solenoid of 6 m of diameter and 13 m long. It is made of 4 layers of windings of NbTi cable that is cooled to 4.45 K in order to achieve the superconducting state. This magnet is able to produce an internal uniform magnetic field of 3.8 T. From this values, and using equation 2.8, the expected resolution for a charged particle with a $p_T = 10$ GeV/c is around the 0.02%. Outside the magnet, 5 wheels and 3 disks of iron are placed in order to return the magnetic field flux, inducing an external 2 T radial magnetic field. This iron yoke is the main contribution to the detector weight, 12000 tons. The muon chambers are located in between the iron yoke, that bends the muons in opposite way for better resolution. The tracker system and the calorimeters are located inside the magnet.

2.2.3 Tracker system

The tracker has been built with two different technologies: Pixels and Silicon Strips. They are arranged concentrically in cylindrical volumes in the barrel, being the pixel detector the innermost. In the endcap the two technologies are arranged in parallel disks. The CMS tracker extends to a radius of 1.1 m and around 2.7 m on each z direction, reaching a coverage in $|\eta|$ between 0 and 2.5. The pixel system is in the region with a radius below ≈ 20 cm and the silicon detector surrounding the pixel system.

The pixel system is formed by three barrel layers ($r = 4.3$ cm, $r = 7.3$ cm and $r = 10.4$ cm) and four endcap disks ($z = \pm 35.5$ cm and $z = \pm 46.5$ cm). It has an approximate active surface of one square meter with approximately 70×10^6 pixels with a cell size of 100 μm by 150 μm. This pixel system allows to obtain three highly precise points that are mainly used for reconstructing vertexes. Its resolution in $r - \phi$ is shown in figure 2.12 [95]. The pixel system achieves a resolution in $r - \phi$ around 9μm. In the same figure the efficiency to find hits per each barrel layer and endcap disk is shown. This efficiency is higher than 98% for all layers and disks. This efficiency degrades with integrated luminosity due to radiation damage from collisions. In figure 2.13, the disposition of all the tracker subsystems is shown. The most of the layers of the tracker are contained in the $|\eta| < 1.6$ region, as shown in figure 2.14.

The Silicon Strip system is formed by an inner (20 cm $< r <$ 55 cm) and outer tracker (55 cm $< r <$ 116 cm). It has an approximate active surface of 300 m^2. The inner tracker is composed of a 4-layer barrel (TIB for Tracker Inner Barrel) and 3 disks (TID for Tracker Inner Disks) in each endcap. The outer tracker systems is composed of 6 layers in the barrel (TOB for Tracker Outer Barrel) and 9 disks in each endcap (TEC for Tracker EndCap).

The tracker system has been designed to specifically address the reconstruction

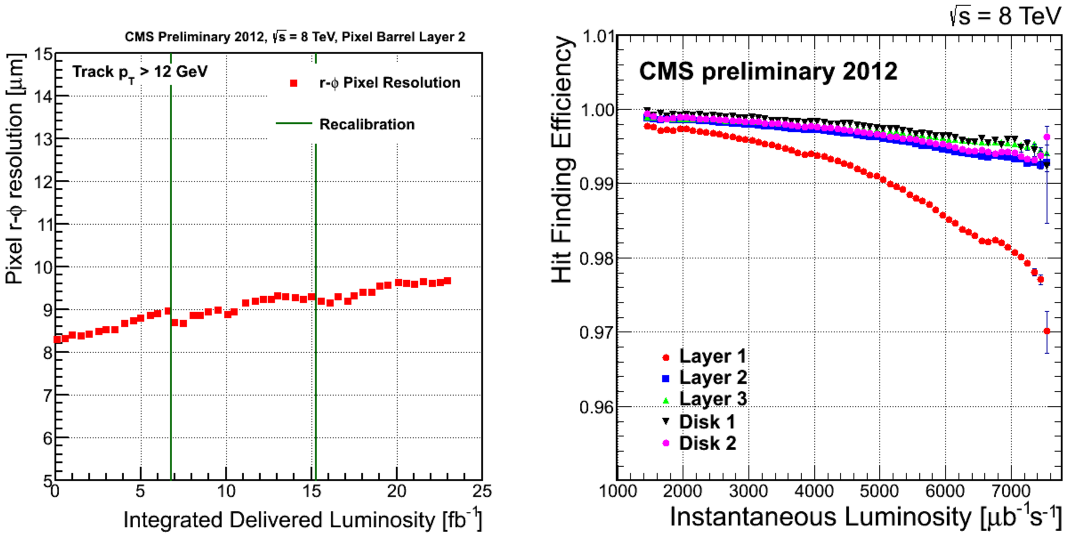


Figure 2.12: Resolution of the pixel detector in $r-\phi$ as a function of the luminosity delivered to CMS detector [left] and efficiency of finding hits by the same detector [right]. The hit finding efficiency is higher than 98% for the entire tracker. A degradation of pixel resolution is clearly visible as a function of the recorded luminosity. A high hit finding efficiency is important for a precise p_T measurement [95].

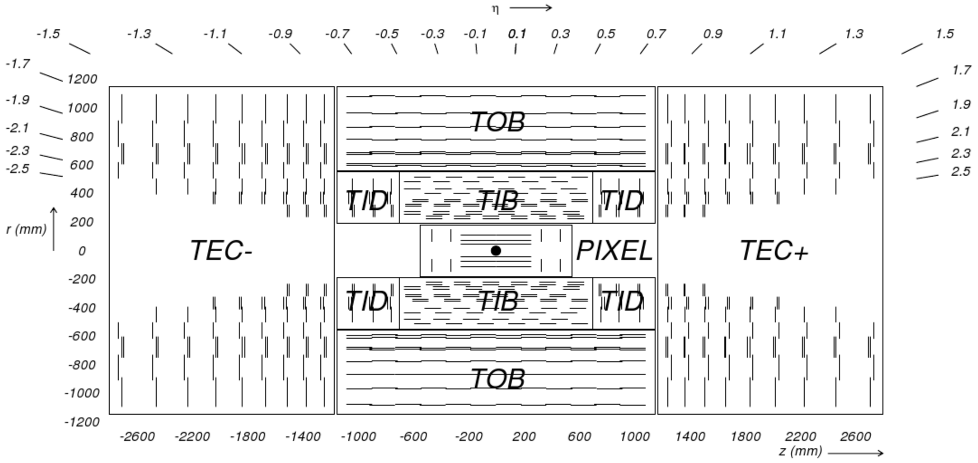


Figure 2.13: CMS tracker system configuration: Pixel, Tracker Inner Barrel (TIB), Tracker Inner Disks (TID), Tracker Outer Barrel (TOB) and Tracker EndCap (TEC).

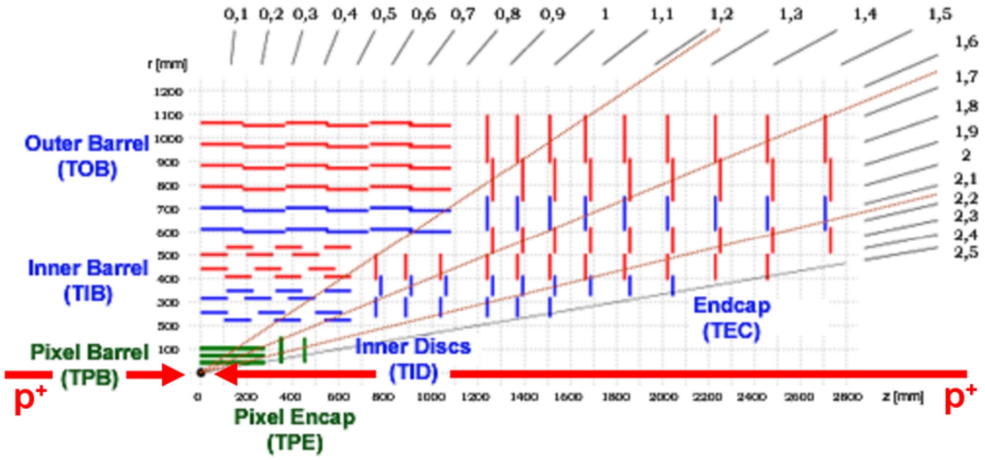


Figure 2.14: Tracker acceptance in η.

of high p_T leptons, with particular interest in the isolation of electrons and, as a consequence, to isolate photons. This system measures the energy of all electrically charged particles. The tracker plays a central role in jet reconstruction because charged hadrons are their most important constituents. Also the tracker fulfills granularity requirements to reconstruct secondary vertexes to tag and reconstruct B-hadrons. The tracker system is able to reconstruct tracks of particles with at least 0.1 GeV of p_T with $|\eta| < 2.5$. Charged hadrons are reconstructed with an efficiency of at least 85% for $p_T = 1$ GeV and up to 95% for p_T above 10 GeV. The full tracker system was built to be highly resistant to radiation damage, as it is the closest subsystem to the interaction point, and is expected to last for around 10 years.

2.2.4 Electromagnetic calorimeter

The CMS ECAL (Electromagnetic CALorimeter) is the detector subsystem designed to stop photons and electrons in order to measure their energy. It is an hermetic cylindrical calorimeter made of 61200 crystals in the barrel ($|\eta| < 1.479$) and 7324 in each endcap ($1.48 < |\eta| < 3$). The crystals material is lead-tungstate that is transparent, very dense (8.28 g/cm^3), has a small Moliere radius (2.2 cm) and a short radiation depth ($X_0 = 0.89$ cm). This material has been chosen for the characteristics already described, but also because of its fast emitting scintillation light (about 25 ns) and its good energy resolution, typically with $\sigma(E)/E \sim 0.5\%$ for an energy of 50 GeV and $\sigma(E)/E \sim 0.4\%$ for an energy of 100 GeV . The main disadvantage of this material is its high dependent response to the temperature (about 2% /°C), making crucial to maintain a stable temperature in the ECAL system. The crystals are distributed in 36 super-modules, 1700 crystals each, in the barrel (EB for ECAL Barrel) and in four 'Dee's, of 3662 crystals each, in the endcaps (EE for ECAL endcap). Each endcap is divided in two equal semicircles, called 'Dee's. In the EB, the scintillation light is collected by Avalanche Photo-Diodes, or APD, and by Vacuum Photo-Triodes, or VPT, in the EE. A preshower system is installed in front of each endcap to allow a better discrimination between photons and π^0's. A representation of the CMS ECAL is presented in figure 2.15.

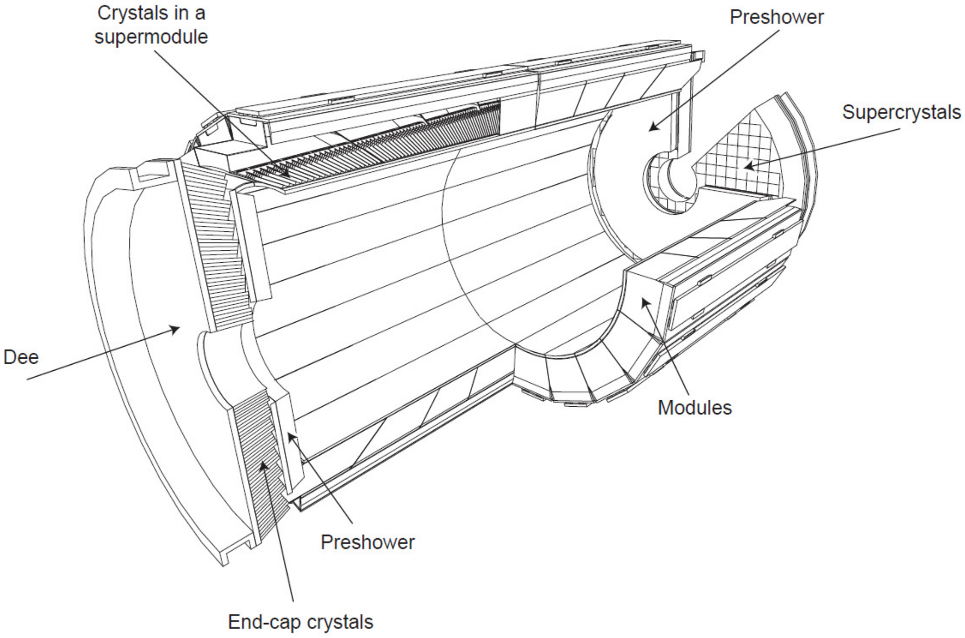


Figure 2.15: CMS ECAL representation showing its different components: barrel modules, the preshower system and endcap 'Dee's.

The ECAL system requires constant correction to compensate the loss of transparency due to aging by radiation. Figure 2.16 shows the relative response of the crystals during LHC run 1 and the di-electron mass from electrons measured in the ECAL barrel [96]. The aging of the crystals is measured by a laser monitoring system that measures their transparency. To validate the corrections derived from the laser monitoring, a comparison is performed between the reconstruction of electrons by the tracker and the ECAL. Such comparison in function of time is shown in figure 2.17 [96].

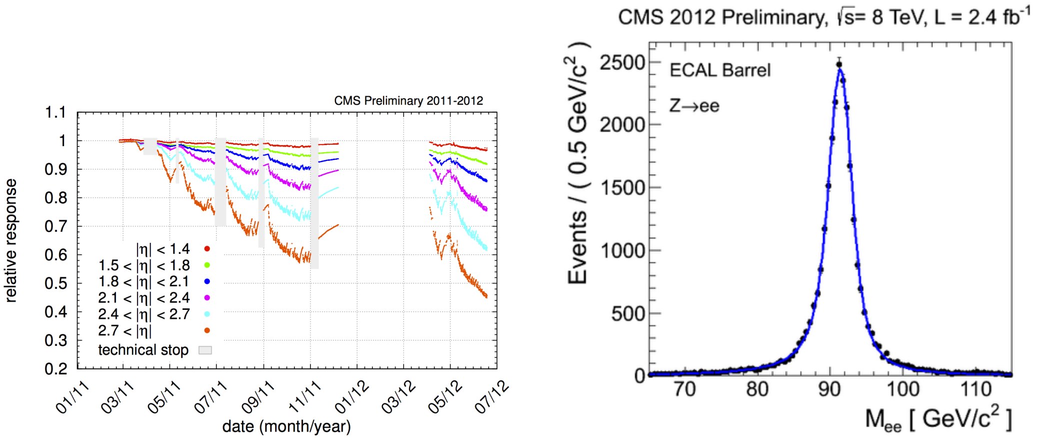


Figure 2.16: Relative response of the ECAL crystals as a function of time, for different η regions [left]. Di-electron invariant mass from electrons identified in the ECAL barrel [right] [96].

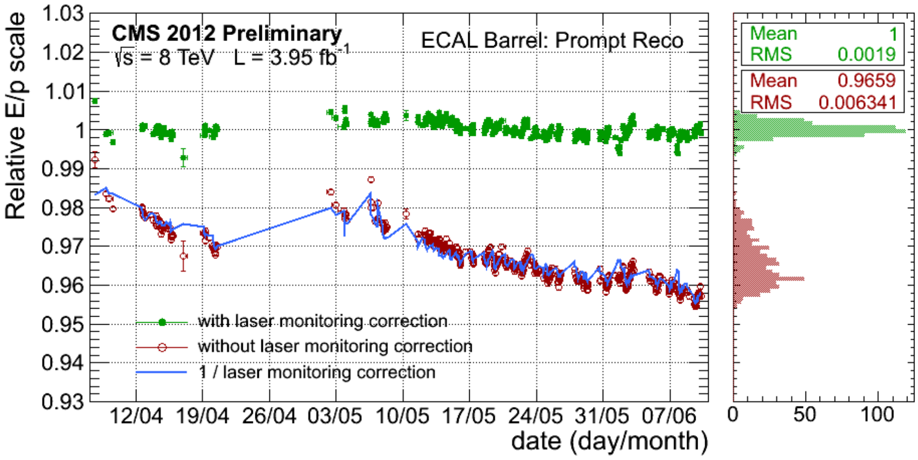


Figure 2.17: Relative E/p for electrons with and without laser monitoring as a function of time. The electron energy E is measured by the ECAL while its momentum p is measured using the tracker information. Laser monitoring is able to effectively correct ECAL lost of transparency [96].

2.2.5 Hadronic Calorimeter

The CMS HCAL, for Hadronic CALorimeter, is the subdetector designed to measure the energy of hadrons produced in the collisions. It is an hermetic set of subsystems covering up to $|\eta| < 5.2$:

- Hadron Barrel Calorimeter (HB): Covering $|\eta| < 1.4$ is located between the ECAL barrel and the magnet.
- Hadron Endcap Calorimeter (HE): Extends the coverage of the barrel in the region $1.4 < |\eta| < 3$.
- Hadron Outer Calorimeter (HO): Located outside the magnet, uses it as an additional absorber.
- Hadron Forward Calorimeter (HF): Completes the coverage of the system from $|\eta| = 3$ up to $|\eta| = 5.2$.

The CMS HCAL layout is shown in figure 2.18. The system is made of two parts, an absorber to develop the hadronic showers and a scintillator to measure the particles energy. The length scale of hadronic calorimetry is defined as the interaction length corresponding to the mean free path of an hadron in a material. The typical length of an hadronic shower is around 300 mm. The HB absorber is made of 40 mm thick steel plate, eight layers of brass plates of 50.5 mm thick, six brass plates of 56.5 mm thick and a steel plate of 75 mm thick. The HE uses the same absorber but with thicker plates of 79 mm. Between the absorber plates a plastic scintillator, Kuraray SCSN81, 3.7 mm thick, is placed. In the region with $|\eta| < 1.6$ the achieved granularity is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ in the region with $|\eta| > 1.6$. This design gives a total of 70000 tiles. The produced light in the HB is collected by optical fibers and transferred to the Hybrid Photo Diodes (HPDs). These diodes were chosen because of their small sensitivity to the magnetic field, an important feature because HCAL is inside the solenoid magnet.

The HF design is very different from the rest of the HCAL subsystems. The most important challenge for the HF is the high resistance to radiation from collisions: while in the central rapidity region, 100 GeV are deposited in average, in the forward region, this average is 760 GeV. For this reason a Cherenkov detector made of quartz fibers with a steel absorber was chosen. The light produced in the HF is collected by photo multipliers.

HCAL response is defined as the energy measured by it with respect to the momentum of a track, being measured by the tracker system, for a single particle. In the most central region, $|\eta| < 0.52$, the HCAL response is ~ 0.6 for

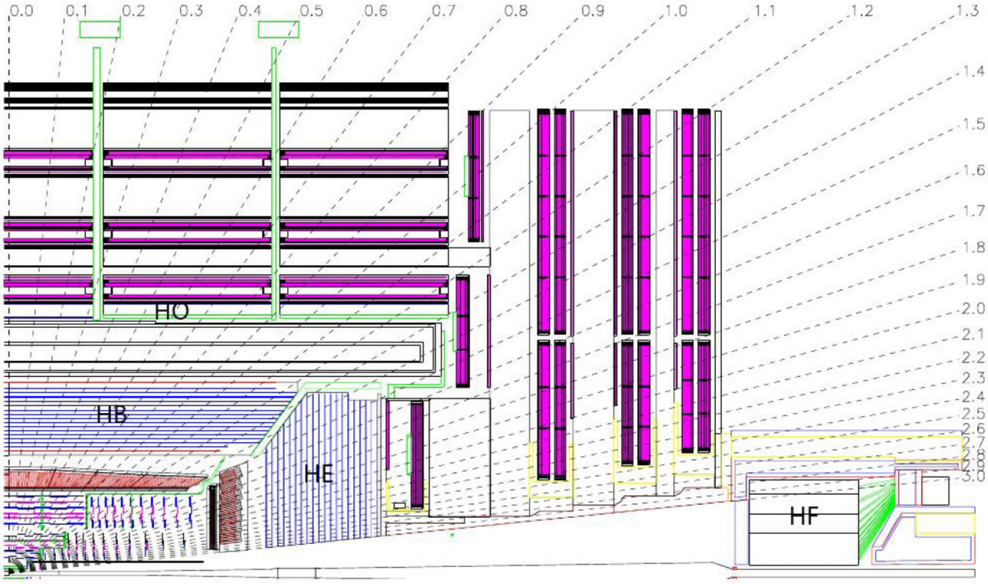


Figure 2.18: CMS HCAL representation.

$p_{Track} > 8$ GeV/c. In the region $1.39 < |\eta| < 2.01$ the response is ~ 0.4 for $p_{Track} > 8$ GeV/c.

Calorimeters, ECAL and HCAL, and tracker are the fundamental pieces to identify the hadronic products of collision events. Whereas HCAL is designed to measure the energy of hadrons (charged and neutral) produced from collisions, its information alone its not sufficient to correctly reconstruct them. The method used to combine the calorimeters information to reconstruct physics objects called jets will be presented in section 2.3.3.

2.2.6 Muon chambers

The muon system of CMS is located at the most exterior layer of the detector due to the penetration power of this particle. Muons are not stopped by the calorimeters and, with neutrinos, they are able to escape the detector. The muon chambers are placed in a cylinder around the HO and in disks in the endcaps. Three main characteristics have been fulfilled from the design: efficient muon identification, precision measurement of muon charge and momentum, and fast measurement to provide trigger capabilities. These chambers are made of three different subsystems:

- Drift Tubes Chambers (DT): located in the region with $|\eta| < 1.2$ and disposed in four layers. They consist of individual drift tubes of 50 μm of diameter anode wire with two electrode plates creating a drift electric field. The wall of the cell act as cathode. The cells are filled with a gas mixture, 85% Argon and 15% CO_2. The tubes are organized in plaques that are also organized in SuperLayers (SL) each one made of 4 plaques. The barrel is made of 250 DT's disposed in four cylinders separated by iron yokes.

- Cathode Strip Chambers (CSC): installed in the endcaps, provide a coverage up to $|\eta| = 2.4$ from $|\eta| = 0.9$. These chambers are multi-wire proportional chambers made of six planes of anode wires with 7 cathode planes. Four CSC stations are placed in each endcap. The wires are oriented in azimuthal direction while the cathode planes are radially oriented, allowing a complete measurement of the particle position. This system is able to measure with a precision between the 75 μm and 150 μm.

- Resistive Plate Chambers (RPC): This subsystem is made of gaseous parallel plate detectors. This detector is specially useful for triggering as it is very fast and have a good position resolution. There are 480 RPC distributed in 6 layers in the barrel with the DT and in 3 layers in the endcaps with the CSC, and covers the region with $|\eta| < 1.6$.

Figure 2.19 shows a representation of the muon system with the different components. The DT and CSC system cover $|\eta| < 2.4$ without any gap.

Information from DT, CSC and RPC subsystems are utilized to reconstruct muons in CMS detector. In figure 2.20 the muon reconstruction efficiency for the three subsystems [97] is shown. The muon system measures muons p_T with a resolution $\sigma(p_T)/p_T \sim 0.2$ for $|\eta| < 1.5$ and $\sigma(p_T)/p_T \sim 0.4$ for $1.5 < |\eta| < 2.1$.

2.2.7 Trigger

LHC has been designed to collect data from proton-proton collisions every 25 ns, meaning a frequency of 40 MHz. Each recorded event by CMS has a nominal size between 0.5 and 1 MB, corresponding to a data flow of around 10^9 MB/s = 1PB/s that is extremely big for transfer and for storing. Therefore, an on-line selection of events has to be done. The trigger system of CMS does this task in two fold, a level 1 (L1) and a high level trigger (HLT). The L1 is hardware based and the HLT is software based.

The interesting events produced on proton-proton collisions for new physics searches are very rare. The enormous majority of events coming from proton-proton collisions correspond to well understood phenomena, while new physics events are 'exotic' with regards to the most common type of events. Then it is interesting to keep only a part of the events, what actually makes easier the analyses afterward.

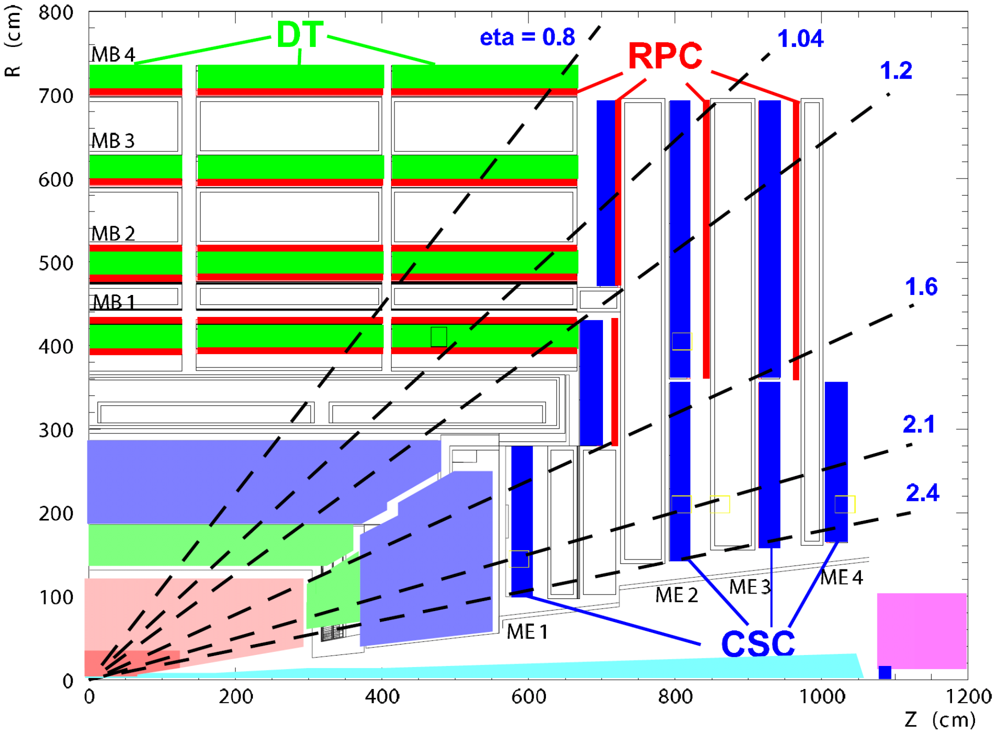


Figure 2.19: CMS muon chambers representation.

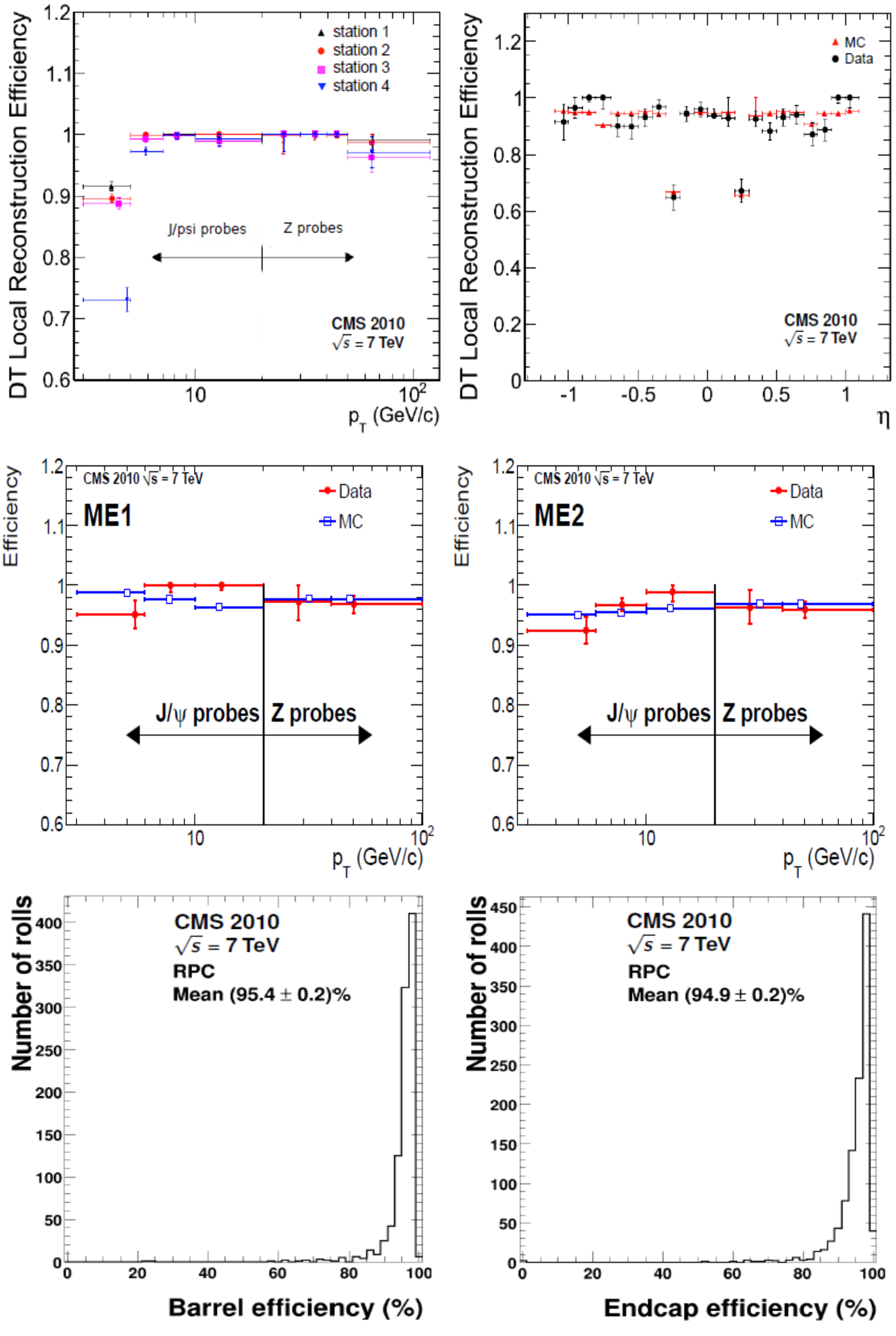


Figure 2.20: Muon reconstruction efficiency in the 4 barrel DT stations as a function of p_T and η [up], muon reconstruction efficiency for data and Monte-Carlo simulation as a function of p_T for the two CSC end-cap stations [middle], muon reconstruction efficiency for RPC barrel and end-cap [down] [97].

The CMS trigger system is designed to keep only 100 kHz by the L1 and 1 kHz by the HLT. L1 is reducing the data flux by 2 orders of magnitude and the HLT by another 3 orders of magnitude.

Level 1 trigger

The L1 is designed to trigger over coarse data coming from the calorimeters and muon chambers, holding data in pipe-lined memories in the front-end electronics. Therefore, it relies on very fast reconstruction of objects coming from this subsystems: muons, electrons, photons, jets, total energy and total hadronic energy (with their corresponding total missing energy and hadronic missing energy). The total energy is calculated as the sum of all the objects in the event, while for the hadronic energy only calojets are considered in the sum. This reconstruction differs from the final reconstruction of the objects, that has a smaller granularity. For example a jet for the L1 consists on successive energy deposits in the ECAL and HCAL, while the off-line reconstruction takes into account also the tracker information.

The L1 starts from data coming from the subsystems which is afterward combined in order to build ranked trigger objects in localized regions of the detector. Global Muon and Calorimeter triggers sort the objects and send the best ranked to the Global Trigger (GT). Before the GT, no events are rejected. It is only with the GT that the selection is applied. The GT combines the information and can apply topological requirements and takes a decision on keeping or disregarding the event. On figure 2.21 the work-flow of the L1 is shown [98].

The L1 cards are distributed between the detector and a nearby cavern at 7 m distance from the detector. The latency time that the L1 disposes between the collision and the taking of the decision is about 3.2 μs. Therefore, the front-end memory in the cards should be able to keep in memory up to 128 bunch crossings.

High Level Trigger

Events passing the L1 pass to the High Level Trigger (HLT), a set of filtering algorithms that run in CPU farms of about 10^4 cores. The reconstruction of objects done by the HLT differs slightly from the final off-line reconstruction. Final off-line reconstruction takes into account the whole information of the detector following the procedure described in section 2.3, while reconstruction for the trigger does a simpler procedure due to computing time limitations. During 2012, the decision taking process was 110 ms, around 10^5 times more than for the L1.

HLT is divided in several paths (128), where each path corresponds to a specific filtering algorithm. Each path contributes with a rate to the HLT, such that the sum of the rates of all path sum up the total HLT rate. The events selected

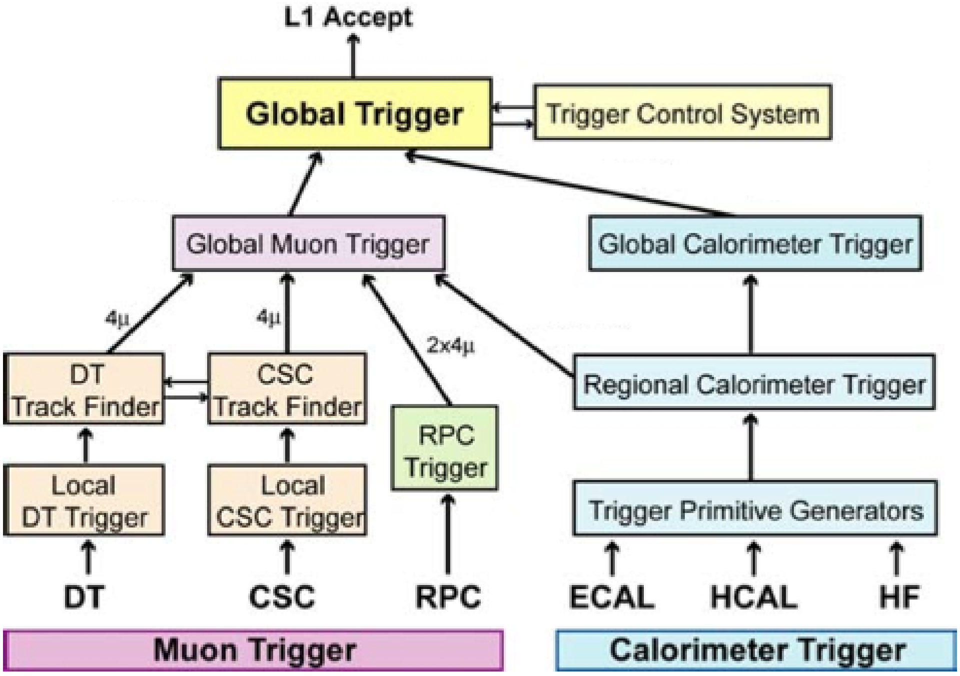


Figure 2.21: The L1 architecture [98].

by HLT are finally stored on disks under several paths depending on the selection performed. From the HLT, events are written to disks with a speed of 100 MBytes/s. There is a constant development of HLT paths focusing on different analysis requirements in order to obtain the best possible selection efficiency for specific signal types.

2.3 Object reconstruction

The different subsystems of CMS play specific roles in the identification of particles from collisions and in the measurement of their properties. In order to achieve such measurements, the information collected by the different parts of a subsystem should be combined. Also, for some particles, the information between different subsystems might be used to achieve a successful identification.

2.3.1 Particle Flow (PF) algorithm

CMS uses a sophisticated algorithm called Particle flow [99, 100, 101, 102] to reconstruct the particles detected in the subsystems: charged and neutral hadrons, electrons, muons and photons. The algorithm uses the information from all the subsystems to ameliorate the reconstructed objects resolution.The quality of the objects reconstructed by this algorithm greatly affects the analyses, as they rely on the information of these objects. The Particle flow algorithm has been optimized to give the best performance in terms of minimal fake rate, correct energy and momenta reconstruction. It is based in an iterative process that searches for the optimal combination of the information of an event in order to reconstruct different objects. Precisions will be given later, when the objects will be described.

2.3.2 Track and vertex reconstruction

A vertex is defined as the point from where several tracks in the detector are originated. Reconstruction of vertices and tracks is one of the most important parts of the reconstruction of events.

Track reconstruction proceeds from the clustering of pixel and silicon strips signals into hits. The found hits in the different layers of the tracker are used to reconstruct the trajectory of the charged particle that left the hits. In order to reconstruct trajectories, at least three hits or two hits and a vertex constraint are needed. The reconstruction proceeds from the propagation of hits in different layers. Such propagation is done through an iterative procedure where all the different combinations between tracks are considered. This process could lead to multiple possible tracks. In order to discard fake tracks and to choose the most likely tracks, the vertex compatibility and the number of hits per tracks are used. Most compatible tracks and the ones with the highest number of hits are preferred. Finally, a χ^2 fitting is used in order to discriminate between multiple tracks.

A way to estimate the goodness of the reconstruction is to measure the reconstruction efficiency of charged pions and muons, such efficiencies are shown in figure 2.22 [103]. This efficiency is obtained using Monte-Carlo simulations and it is defined as the fraction of charged particles that can be associated with their corresponding reconstructed tracks. The reconstruction is near 100% efficient for muons and pions with $p_T > 1$ GeV.

Vertex reconstruction

Vertex reconstruction proceeds from clustering of tracks. The problem of vertex finding relies on the association of a set of tracks up to the beam line. A figure of merit is used in order to measure the compatibility of a set of tracks pointing to a given vertex. For such figure of merit, the z position of the tracks, from trajectory extrapolation, and the vertex position are used. The final assignment of vertices responds to the minimization of the figure of merit. Finally, the primary vertex associated to a set of tracks is defined as the one that minimizes the sum of the p_T squared, $\sum_{tracks} p_T^2$, of the associated tracks. The resolution, in x and z, and efficiency of primary vertices is presented in figure 2.23 [103]. The identification of the primary vertex of an event is crucial to separate objects produced in hard interactions from other interactions, as pile up. The primary vertex reconstruction is performed by clustering tracks together and fitting them to determine how likely they came from a common vertex. The reconstructed vertex with the largest total p_T is considered as the primary vertex.

Calorimeter clustering

Identification of calorimeter clusters is an important step in the reconstruction of particles in CMS. They are defined as a group of deposits in an specific part of a calorimeter. Their reconstruction proceeds from the cell energy maxima, then adjacent deposits are associated to it up to an energy threshold determined to avoid adding noise. Thresholds are 80 MeV for ECAL barrel and 300 MeV for ECAL endcaps, it is 800 MeV for the HCAL. The clusters are used in the particle flow algorithm to reconstruct objects, as jets. The procedure is done both in the ECAL and HCAL. Calorimeter clusters are used to identify, for example, neutral hadrons and reconstruct the energy of charged hadrons outside the tracker acceptance.

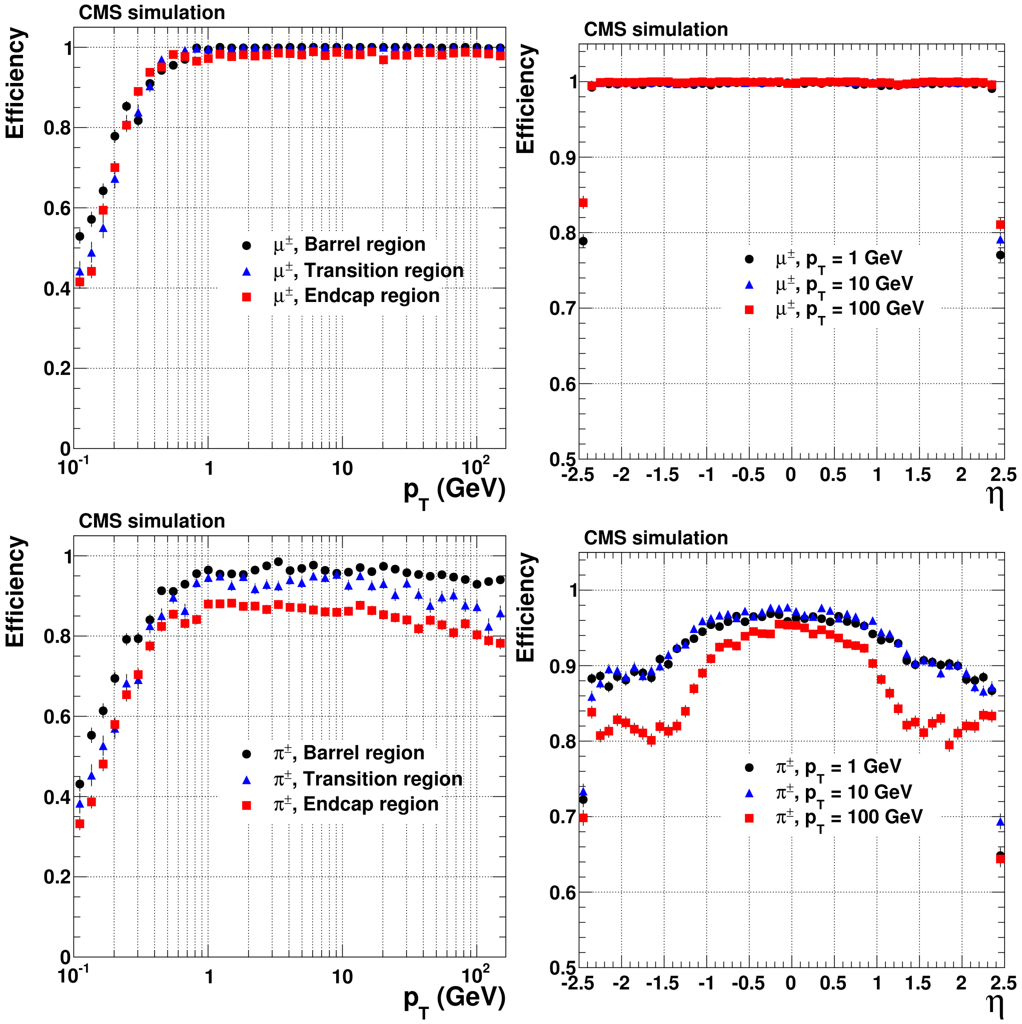


Figure 2.22: Reconstruction efficiency of charged pions and muons in the tracker system as a function of p_T and η. Pions and muons are reconstructed with an efficiency close to one for a p_T >1 GeV [103].

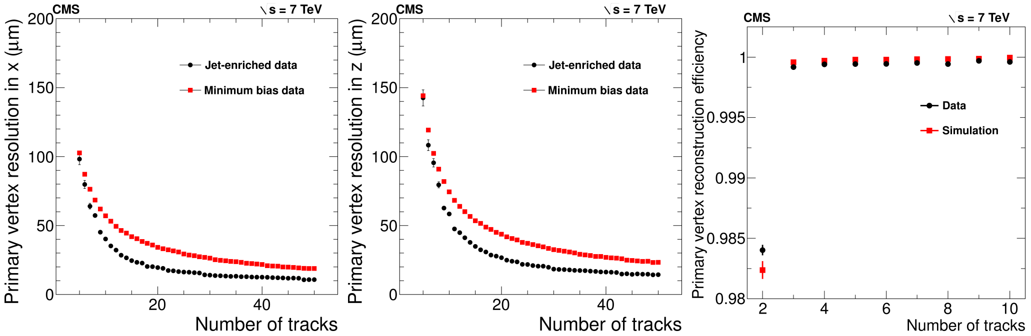


Figure 2.23: Resolution in x [left] and z [center] and efficiency [right] of reconstruction of primary vertices. Small differences are seen in the primary vertex reconstruction efficiency between data and Monte-Carlo simulation for events with only 2 tracks [103].

Link between subdetectors

A single particle traversing CMS may interact with several subdetectors. For example, a charged pion would leave a track, a cluster in the ECAL and a cluster in the HCAL. Then, taking into account the information of different subdetectors and connecting them help to correctly identify each type of particle. Thanks to the excellent granularity of CMS, which allows to have a highly performing linking between different subdetectors, efficient reconstruction algorithms are available. For example, using Monte-Carlo simulations the difference of the reconstructed jet p_T with the generated p_T pondered by the generated p_T for jets with a $p_T = 40 - 60$ GeV/c and $|\eta| < 2.5$ is around 0.2. The reconstructed jet p_T greatly depend of the capacity of the reconstruction algorithm to link traces with cluster in the ECAL and HCAL. Indeed jets are composed of charged hadrons that leave tracks and cluster and neutral hadrons that only leave clusters.

The links are done from an extrapolation of tracks towards the calorimeters. A track and a cluster are linked if the extrapolation point is within the cluster boundaries. The extrapolation to the ECAL is performed up to a cluster on it. On the other hand, the extrapolation to the HCAL is performed up to a depth of a nuclear interaction, characterizing an hadronic shower. Figure 2.24 shows the links from two tracks linking ECAL and HCAL clusters in $\eta - \phi$ plane, taken from [95].

2.3.3 Jet reconstruction

Quarks and gluons produced from a collision cannot be observed directly due to the confinement of strong interactions. After their production, they follow an hadronization process (see section 4.1.2), making principally hadrons that are de-

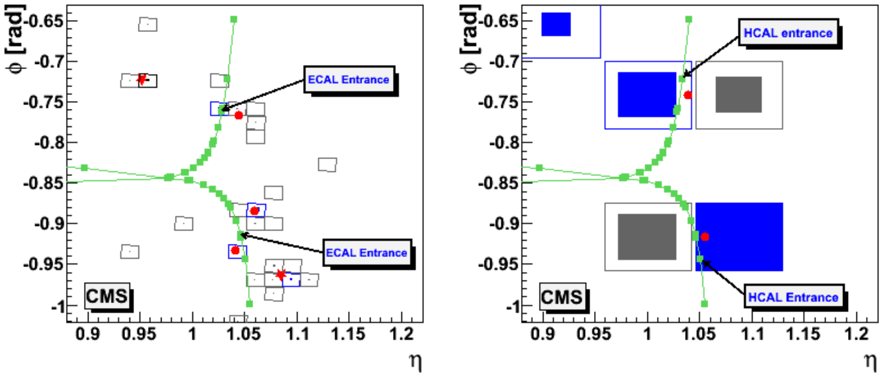


Figure 2.24: Connection between tracks reconstructed in the tracker (in green) and energy clusters in the ECAL [left] and HCAL clusters [right]. Each box represents a calorimeter cell. The red circles represents clusters that were linked to a track, while red stars mark clusters not linked to tracks [95].

tected in CMS. The set of hadrons produced in the process is called an hadronization cascade. All the particles produced from an initial quark or gluon define a jet. Ideally a jet should contain the energy of the initial parton that undergoes the hadronization processes. Requiring this, jets could be connected to the partons produced in the collision. The analysis presented in chapter 5 relies on these objects.

Jet clustering algorithms

Particle flow algorithm reconstructs individually each particle produced from a collision event. In order to associate all the different particles identified with particle flow reconstruction procedure, additional algorithms are implemented: clustering algorithms. The CMS procedure for jet reconstruction only looks at hadronic activity. Therefore, all isolated leptons in an event are removed from the particle content delivered to the clustering process. Within particle physics experiments, different algorithms have been used. ATLAS and CMS mainly use the anti-kt but CMS also uses Cambridge-Aachen algorithm for many analysis. In the following, these algorithms are going to be described with their specific characteristics. Two types of algorithms exist: cone type clustering (SISCone) and sequential clustering (kT, Cambridge-Aachen, anti-kT). The first type uses a fixed radius circle to try to identify clusters of particles. The second type clusters particles by a distance until a fixed maximal radius.

SISCone SISCone algorithm [104] uses a circle to explore randomly the $\eta - \phi$ plane until a particle comes into contact with it. Then the circular enclosure is pivoted until a second particle is in contact with the edge. All the matches define the stable cones. Particles in the found stable cones are removed and the procedure is repeated until no more cones are found. From all found stable cones, an algorithm is used to split or merge the cones in final jets based on an overlap parameter. In figure 2.25 [105], an example of the jets reconstructed by SISCone is shown for a circular enclosure of radius 1.

kT kT-algorithm [106] begins with the calculation of the kT distance for all pair of particles (i,j), defined as

$$d_{ij} = min(p_{Ti}^2, p_{Tj}^2)\Delta R_{ij}^2/R^2 \tag{2.9}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + \Delta(\phi_i, \phi_j)^2$ and R is the radius parameter. The distance of a particle to the beam is $d_{iB} = p_{Ti}^2$.

Then the minimal distance between all d_{ij} and d_{iB} is found, if it is a d_{ij} the two particles are merged, if it is a d_{iB} the particle is declared to be a final jet. The process is repeated until no particles are left. A result of this reconstruction can be found in figure 2.25.

Cambridge-Aachen Cambridge-Aachen (CA) jet algorithm is similar to kT algorithm but changes the distance definition to $d_{ij} = \Delta R_{ij}^2/R^2$ and $d_{iB} = 1$. In figure 2.25 the jets produced by this algorithm are shown.

Anti-kT Anti-kT (AK) algorithm proceeds equivalently to kT and CA algorithms changing the definition of distance to

$$d_{ij} = min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right)\Delta R_{ij}^2/R^2;\ d_{iB} = \frac{1}{p_{Ti}^2} \tag{2.10}$$

This algorithm produces circular jets in the $\eta - \phi$ plane. An example of the reconstruction performed by this algorithm can be seen in figure 2.25.

Infrared and collinear safety Two extremely important characteristics that should be achieved correctly by jet algorithms are to be collinear and infrared safe. In QCD processes, a parton can split in two collinear components or can emit soft radiation. Such processes will give rise to extra hadrons that could lead to ambiguities in the jet reconstruction, giving different jet compositions of an event. Collinear and infrared safety correspond to the ability of an algorithm to give the same results independently if an initial parton radiated or not. In other

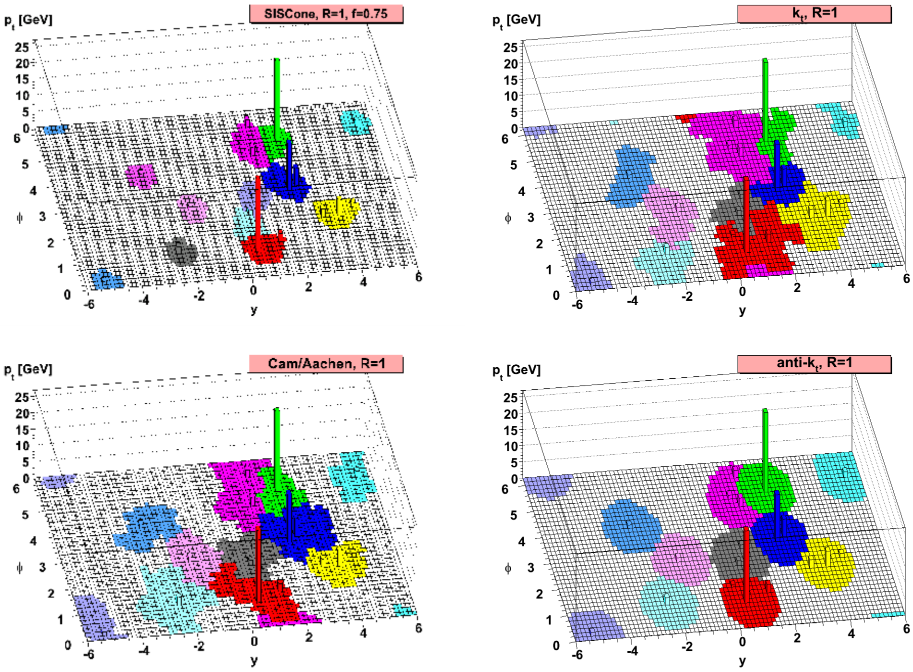


Figure 2.25: Jets in a Monte-Carlo event reconstructed with four different clustering algorithms: SISCone [left-up], kT [right-up], Cambridge-Aachen [left-down] and anti-kT [right-down]. All algorithms used same radius $R = 1$. Jets are displayed in the $y - \phi$ plane and the reconstructed p_T in the z-axis [105].

words, the jet reconstruction should include this radiation correctly to the same jet and do not give rise to an additional jet coming only from one of these radiations. The jet algorithms described in previous sections are collinear and infrared safe.

Jet area Jet area, discussed in [107], is a concept used to estimate the sensitivity of jet algorithms to radiation or pileup or underlying event. Naively, the area of the jet is expected to correspond to πR^2, where R is the radius parameter used by the algorithm. Obviously, this parameter is also related to the final shape of jets given by the algorithm. Figure 2.25 shows the area by SISCone, kT, CA and AK algorithms in the $y - \phi$ plane. A close correspondence of the area to the expected value shows a low sensitivity of the clustering algorithm to radiation, making it more stable to real experimental conditions, where pileup and underlying event contributions could be high. A more precise way to estimate the stability of an algorithm, in terms of area, is to measure the jet area compared to πR^2 as a function of the jet p_T. Figure 2.26 presents such study [105], showing that the AK algorithm is the most stable one. This behavior constitutes a powerful reason to prefer AK over other algorithms.

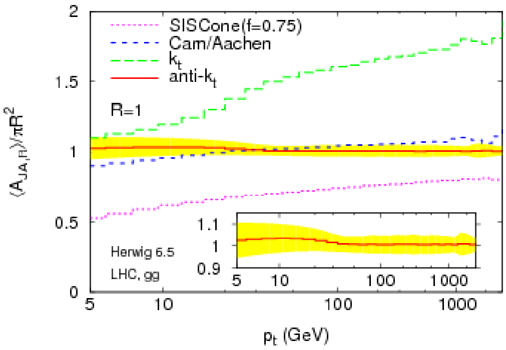


Figure 2.26: Jet areas over expected geometrical jet area for different jet algorithms as a function of the reconstructed p_T for dijets events [105].

Reconstruction at CMS CMS experiment uses AK algorithm with a radius of 0.5 to reconstruct jets. Alternatively CA with radius of 0.8, 1.0 or 1.5 are also used to identify jets substructure, jets inside jets. AK reconstruction relies on particles reconstructed by the particle flow algorithm, being then called particle flow jets (PFJets). Figure 2.27 shows the composition of PFJets.

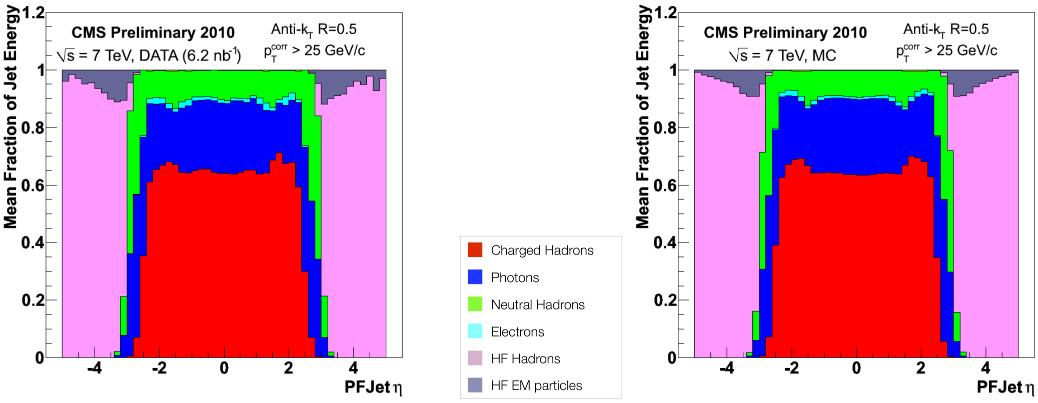


Figure 2.27: Jet composition fraction of particle flow jets as a function of η for data [left] and Monte-Carlo simulation [right]. Central jets $|\eta| < 2.5$ are composed $\sim 60\%$ by charged hadrons and the rest by neutral particles (hadrons and photons). In the legend, HF stands for the Hadron Forward Calorimeter [108].

Jet energy corrections

Whereas jet reconstruction algorithms have been tuned to give the best performance possible, in real experiments the reconstruction is not perfect. Some times all the information in the detector, tracks or calorimeter clusters, are not available or just partially. This impacts jet properties, as for example the p_T. Data-driven methods, as in [109], have been implemented in CMS to correct such imperfections and to measure the resolution of the reconstructed jets in p_T. Several correction levels have been considered in order to take into account subdetectors response to pileup, particle content in different eta regions, among others. In general, such corrections are called Jet Energy Corrections (JEC).

The most important information to determine from jet reconstruction are the energy scale and resolution, correspondingly called Jet Energy Scale (JES) and Jet Energy Resolution (JER). Uncertainties in these quantities depend on the reconstructed p_T of the jets and on η, they are presented in figure 2.28 [95]. JEC uncertainties are of the order of 1% for central jets with $p_T > 100$ GeV/c.

b-jets identification

B-quarks produced in a collision hadronize forming B-hadrons. These hadrons differentiate from the others due to their stability, taking more time to decay in average than unstable hadrons with only u or d quarks. This additional time of

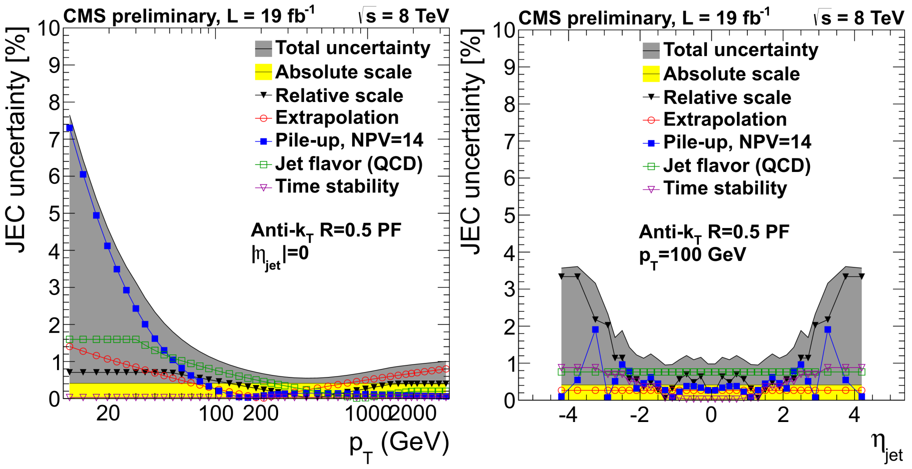


Figure 2.28: JEC uncertainties as a function of p_T [left] and η [right]. JEC have an uncertainty smaller than 4% for jets with $p_T > 20$ GeV, what shows the reliability of jet reconstruction [95].

flight is seen in the detector as a displaced vertex. This characteristic is one of the principal handles to identify jets coming from an initial b-quark. Complex algorithms have been developed in CMS to study and combine jet information to attribute a probability of coming from a b-quark to each jet.

CMS collaboration uses for b jet identification, called b-tagging, the Constrained Secondary Vertex (CSV) algorithm [110, 111]. The main variable used by the algorithm is the existence of a displaced vertex. Specific parameters of the displaced vertex are used, as the time of flight, its mass, the number of tracks associated to it. The impact parameter, defined as the distance between the secondary and primary vertex, is taken into account, because for jets coming from B-hadrons this variable has usually a higher value than for the rest of jets. CSV algorithm takes all these variables and use a MVA (Multivariate Variable Analysis) to build up a discriminant. On the value of the discriminant three cuts are defined. They define the working points that allow to select either with a high efficiency (loose working point: CSVL) or b-tags with a highly pure sample (tight working point: CSVT). The middle working point, CSVM, allows a high efficiency while preserving high purity. In figure 2.29 the number of secondary vertices and the final discriminator for CSV algorithm are shown [111].

CSVM corresponds to a cut on the discriminator greater than 0.679. This working point has an average efficiency of correctly tagged b-jets of ϵ_b^{CSVM} ∼70%, also the efficiency to tag as b-jet a c-jet is around ϵ_c^{CSVM}20% and any light jet ϵ_l^{CSVM} ∼1%, these two last efficiencies are known as the fake rates from c and light quarks respectively. Two additional working points are defined, a tight working

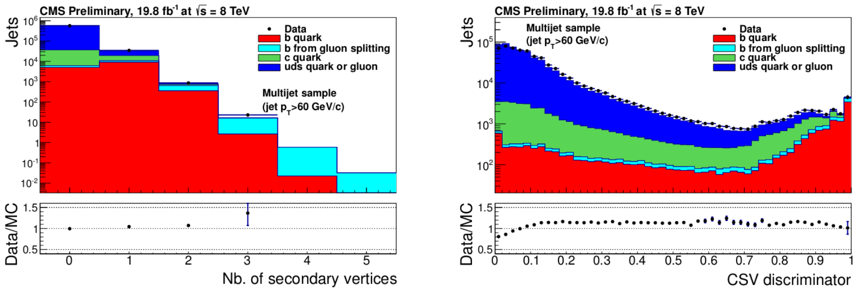


Figure 2.29: Number of secondary vertices [left] and CSV final discriminator [right] for each jet flavor, from a QCD Monte-Carlo sample [111].

point with the lowest fake rate of b-jets ($\epsilon_b^{CSVT} = 50\%$, $\epsilon_c^{CSVT} = 7\%$, $\epsilon_l^{CSVT} = 0.2\%$) and a loose working points with a high b-jet b-tagging efficiency ($\epsilon_b^{CSVL} = 85\%$, $\epsilon_c^{CSVL} = 45\%$, $\epsilon_l^{CSVL} = 10\%$). When the algorithm is applied to Monte-Carlo, a correction has to be performed in order to match the same behavior than in data. Scale factors are derived from the comparison of data and Monte-Carlo. For CSVM a correction of around 5% to correctly tagged b-jets is needed and 17% and to incorrectly b-tagged jets. Figure 2.30 shows the b-tagging efficiencies for CSVM working point as a function of jets p_T and η, from reference [111].

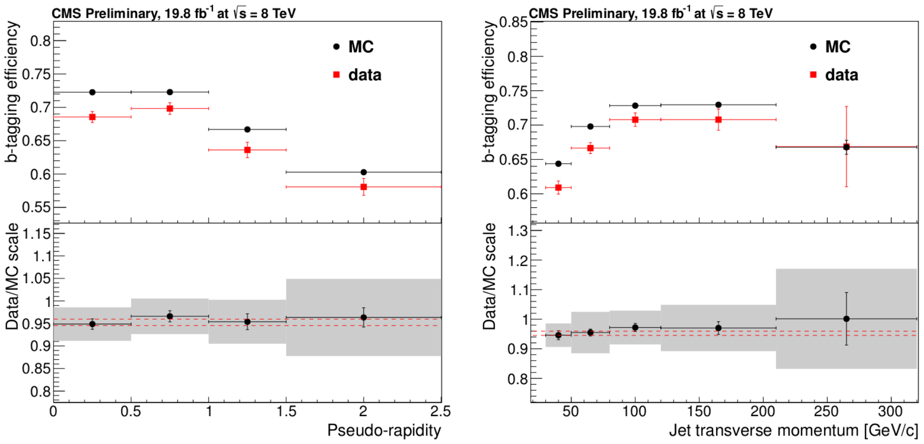


Figure 2.30: Efficiency of CSV b-tagging algorithm in the medium working point as a function of jet p_T and η [111].

3 Feasibility study for a search of a T' at LHC at 8 TeV

In this chapter a strategy is developed to look for a T' in the full hadronic final state. This channel is very challenging. Mainly because the number of expected events for backgrounds in the full hadronic channel is several orders of magnitude higher than the number of expected events for signal in proton-proton collisions at 8 TeV center of mass energy. Additionally, it is a channel poorly explored by current analysis due to its intrinsic difficulties, what makes this an original proposal. In section 1.5 a generic model of VLQ has been introduced. In addition some of their properties predictions have been described. As discussed, a T' can be produced in proton-proton collisions at LHC in several ways. For T' masses higher than 500 GeV/c^2 single production mode gives a higher cross section with regard to pairwise production. Consequently, if there is a T' in nature, it has a higher observability in data considering single production process.

For example, with 20 fb^{-1} at 8 TeV, around 4000 events are expected for a single produced T' with a mass of 700 GeV/c^2, whereas only 500 are expected in the pair production process. But, all these events split in all the possible different final states. In the case of a 700 GeV/c^2 mass T', the dominant decay channel is to top-Higgs, that is around 50% from figure 1.8. Following the calculation, this means about 2000 events are expected where the T' decays into top and Higgs. In this study the choice of the mass value for the T' is based on the argument that it could be accessible to run 1 data collected by CMS at 8 TeV center of mass energy. The principal objective of the study is to motivate a full data search for VLQ. However, it is clear that for higher masses, specially higher than 1 TeV, special techniques, different from the ones used in this study, will be necessary. For such energies, it will be perhaps needed to use for example boosted techniques [112, 113, 114]. These techniques are useful when several decay products of a massive particle are merged in a single jet, because of its high p_T. They allow to recover underlying substructure of jets.

Finally, from Higgs branching ratios (figure 1.20), the Higgs decays the most of the times, 57%, to a pair of b-quarks. Additionally, the top quark decays in a b-quark and a $W^{+/-}$ boson with a branching ratio about 100%. And the $W^{+/-}$ boson decays into a pair of quarks 66% of times. Using these three branching ratios, the

$Br(T' \to 3b, 2j) = 37\%$. Thus the expected number of events of T' into 5 quarks (or jets) is of around 700 events. For collected events by the CMS experiment in LHC run 1, the full hadronic final state constitutes the channel with the highest number of expected events. In the following sections a tentative strategy to extract these events from backgrounds is described. A majority of the results exposed in this chapter are part of the reference [115]. A data analysis inspired in this phenomenological study will be presented in chapter 5.

3.1 Samples used in the study

This feasibility study relied on Monte-Carlo simulations for the signal and backgrounds. MadGraph 5 [116, 117], version 1.5.11, has been used for the generation of parton level samples. Signal model has been implemented with the FeynRules toolkit [118]. For the simulation of hadronization processes of parton samples, Pythia 6 [119] has been used. These tools are well known to describe correctly high jet multiplicity final states, specifically an unstable particle (as the $W^{+/-}$ boson, Z^0 boson or t-quark) with up to four additional jets. For a detailed description of Monte-Carlo techniques and tools see chapter 4.

The cross sections and expected number of events for signal and each background used in the study, at 8 TeV for 20 fb^{-1}, are listed in table 3.1. All SM processes giving a final state with at least 5 jets in the final state have been considered. As it will be described in the next section, the strategy of the study is based in the fully hadronic final state where the signal produce 5 jets coming from T' and one additional jet in the forward region.

At production level some loose cuts were required to facilitate the sample generation. QCD was produced requiring all jets with a $p_T > 30$ GeV/c and within a pseudo-rapidity of $|\eta| < 5$. All the other background samples were produced with jets having $p_T > 10$ GeV/c and no cut on the pseudo-rapidity. For samples with at least one Z^0 boson, di–boson processes Z^0Z^0, $W^{+/-}Z^0$and Z^0+jets, the mass of the di–lepton pair was required to satisfy $M_{ll} > 50$ GeV/c^2 in order to avoid integration troubles (when the di–lepton invariant mass approaches zero the cross-section diverges). After hadronization, the jets were built up with Anti-Kt algorithm using $R = 0.5$ with the standard implementation provided by the FastJet package [120].

The signal sample was produced with jets with $p_T > 10$ GeV/c with the same packages. The vector-like mass was set around 700 GeV/c^2, and the mixing to their maximal allowed values when both mixing to third and first generation are allowed (however the couplings to light generations were maximized to the experimental constraints, the coupling to the second generation is negligible). This choice corresponds to the following set of parameters used in the simulation for the signal:

$\xi_Z^T = 0.5145$, $V_R^{41} = 0.078$, $V_R^{42} = 0.0041$, $\kappa_T = 0.087$ and $BR(T' \to H^0 t) = 0.472$. With this choice, the physical mass of the T' is $M_{T'} = 734$ GeV/c^2. This also set the cross section to 200 fb. These parameters correspond to a benchmark point similar to the one defined in [59], it was set to obtain a VLQ mixed with light and heavy SM quark generations and a T' mass close to 700 GeV/c^2.

The samples used in this study were analyzed using MadAnalysis package [121, 122].

3.2 Strategy for the full hadronic final state

The final state of interest for this study, the full hadronic final state, contains 3 b-quarks and 2 additional quarks as decay products of the T'. Two b's coming from the Higgs boson, a third b from the top-quark decay, and 2 light jets from the $W^{+/-}$ boson decay. In addition, the T' is produced in association with a light quark. Figure 3.1 shows the η distribution of the jet produced in association with the T'. Consequently, the signal events are expected to have at least 6 jets, 5 with an $|\eta| < 2.5$ and one with an $|\eta| < 5$. Unlike the leptonic channels all the decay products of the T' are seen by a detector (mainly produced with an $|\eta| < 2.5$), as CMS, circumstance that allows to do a full mas reconstruction of it.

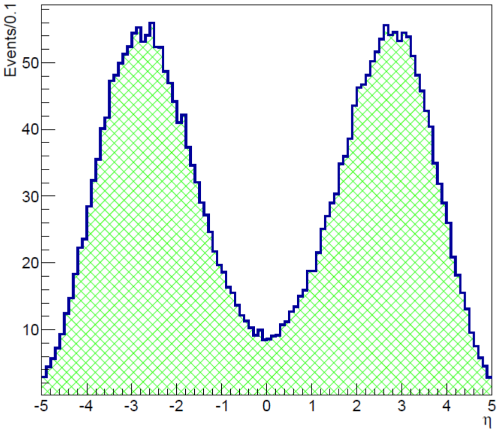


Figure 3.1: η distribution of the forward jet produced in association with the T'. Signal sample is normalized to theoretical cross section and to 20 fb^{-1}.

The main difficulty in full hadronic final states is the large QCD background. In this study all the possible SM backgrounds giving a full hadronic final state were included, in decreasing order of contribution: QCD production, vector plus jets (V+jets, where V is a $W^{+/-}$ or a Z^0 boson), $t\bar{t}$, single top, and di–boson ($W^{+/-}W^{+/-}$, $W^{+/-}Z^0$, Z^0Z^0) have been considered, as shown in table 3.1.

Process	$\sigma_{8\text{TeV}}$ (pb)	Expected Events
Signal ($T'j$)	0.2	700
QCD (bbjjj)	500	10,000,000
$W^{+/-}$+jets	37,509	750,180,000
Z^0+jets	3,503.71	70,074,200
$t\bar{t}$	234	4,680,000
single-t	114.85	2,297,000
Di-boson	96.82	1,936,400

Table 3.1: Cross sections and expected number of events for background processes and signal for a luminosity of 20 fb^{-1}. The single top and di-boson samples were produced with up to 3 additional jets.

The analysis strategy relies on the correct identification of the 5 jets coming from the T'. For this purpose a jet association method has been designed to select the 5 jets from the T' based on the characteristics of the signal. The method used to identify the jets coming from T' is the following:

- Tag b–jets and keep events with at least two b-jets ($n_b \geq 2$).
- To reconstruct the Higgs, only pairs of b-jets with a $\Delta R_{bb} < 2.5$ are considered. If more than one pair is found, the pair with the closest mass to the Higgs boson mass (125 GeV/c^2) is chosen. The two jets chosen for the Higgs reconstruction are not longer considered in the jet collection for the reconstruction procedure of the $W^{+/-}$ boson and top quark.
- From the remaining jets, the pair of jets with mass closest to the W-mass (80 GeV/c^2) is identified as the $W^{+/-}$ boson.
- Finally, from the remaining jets, a third jet is chosen and coupled to the previously identified W-jets. The jet giving the closest mass to the top mass (172 GeV/c^2), combined with the W-jets, is selected as the top b-jet.

From the objects reconstructed with the jet association method and other signal characteristics, a selection to discriminate signal events from backgrounds has been designed. As in the SM backgrounds there is no real Higgs, the selection strongly relies on the presence of a real Higgs in signal events. Also the presence of a real top in signal events is used.

In the following section, the selection applied is described. The different criteria have been chosen in order to increase the discrimination of signal from background. The objective of the study is not to obtain the best possible discrimination but moreover to illustrate a possible selection to extract the signal and to give a qualitative understanding of signal characteristics.

3.3 Event selection

All cuts were applied one after the other in the order given in the following list:

- *Cut 0*: In first instance only the events with at least 6 jets with $p_T > 30$ GeV/c are kept. From them, at least five jets within $|\eta| < 2.5$ and at least one jet within $2 < |\eta| < 5$. The T' decays into five central jets, but the associated jet produced with it tends to be in the forward direction, as shown in figure 3.1. This cut tries to mimic the detector acceptance.
- *Cut 1*: The first kinematic cut requires $p_T > 150$ GeV/c for the leading jet, $p_T > 80$ GeV/c for the sub-leading jet and $p_T > 60$ GeV/c for the 3^{rd} and 4^{th} leading jets in each event. The p_T distribution of the six leading jets is shown in figure 3.2.
- *Cut 2*: The following criteria uses the total hadronic energy ($H_T = \sum |p_T(j)|$, with the sum over all jets in the event), which is plotted in figure 3.3 for each backgrounds and the signal. Signal events have higher H_T than background events, reflecting the presence of the very massive T'. Events with $H_T >$ 630 GeV/c were selected.
- *Cut 3*: At least two b-jets were required in order to perform the jet association procedure. The identification of jets coming from b-quarks has been described in section 2.3.3. However, as this study has been done up to the hadronization level, the same techniques were not used. As a substitute, the following method was used to emulate the performance of b-jet identification algorithms:
    1. Select a working point for a b-tagging algorithm. This sets the efficiency of the algorithm to tag jets coming from a b-quark (ϵ_b^{b-tag}), from a c-quark (ϵ_c^{b-tag}) and from a light-quark (ϵ_l^{b-tag}). As reference, the loose working point of the CSV algorithm has been chosen (discussed in section 2.3.3). The CMS results [111] were used: $\epsilon_b^{b-tag} = 0.9$, $\epsilon_c^{b-tag} = 0.6$ and $\epsilon_l^{b-tag} = 0.1$.
    2. Throw a random number r between 0 and 1 for each event.
    3. Loop over all the jets from an event and, depending on their flavor and the random number from last step, declare each jet to be or not to be b-tagged. A jet is b-tagged if: it is coming from a b-quark and $r \leq \epsilon_b^{b-tag}$,

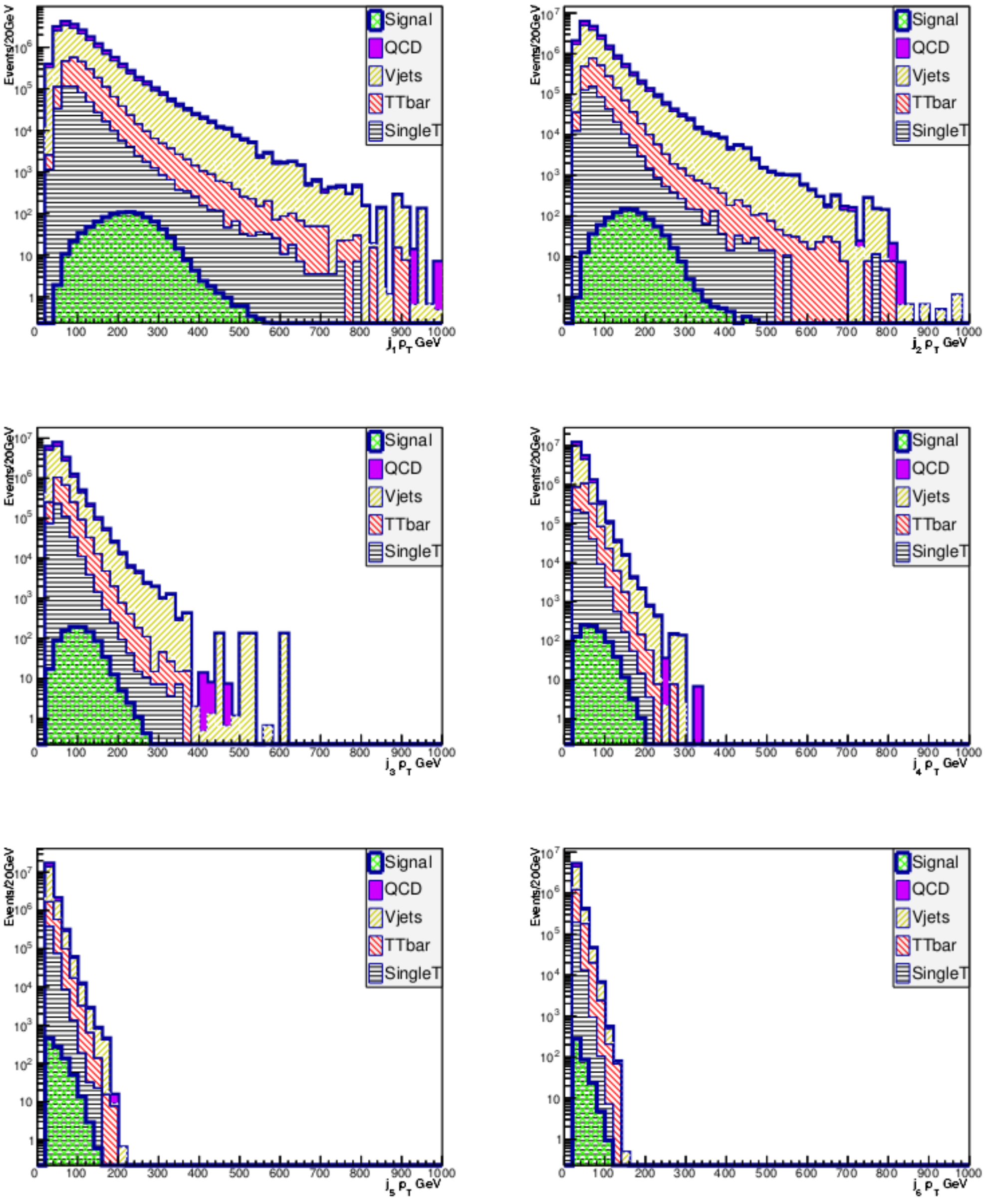


Figure 3.2: p_T of the six leading jets for backgrounds (stacked) and signal (over–imposed) normalized to 20 fb^{-1} luminosity. QCD background is on top of the stack of backgrounds.

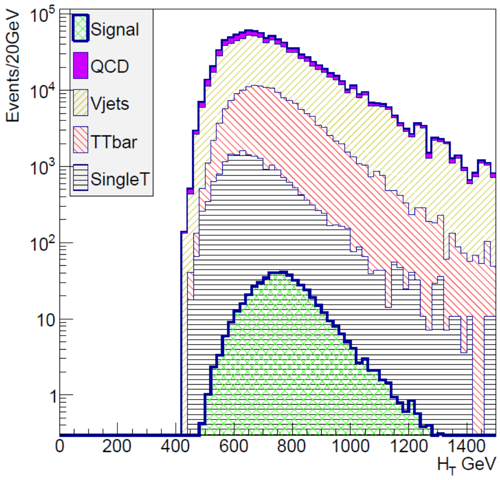


Figure 3.3: Total hadronic energy for backgrounds (stacked) and signal (over–imposed) normalized to 20 fb^{-1} luminosity. H_T is higher in signal than in background events.

or it is coming from a c-quark and $r \leq \epsilon_c^{b-tag}$, or it is coming from a light-quark and $r \leq \epsilon_l^{b-tag}$.

The number of b-tagged jets using the method described above is displayed in figure 3.4.

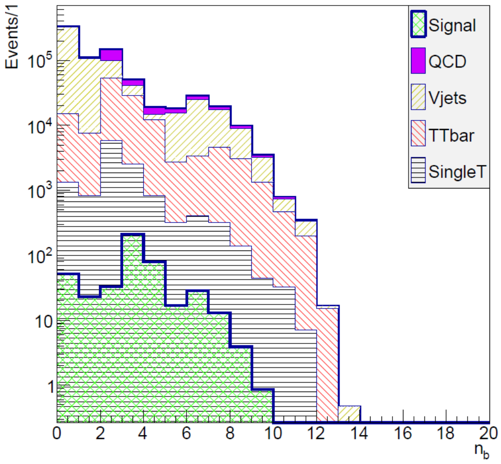


Figure 3.4: B-tagged jet multiplicity for backgrounds (stacked) and signal (over–imposed) normalized to 20 fb^{-1} luminosity. The signal has as mean value 3 b-tagged jets.

- *Cut 4*: Events are kept if the two jets assigned to the Higgs have a $\Delta R_{bb} < 1.8$. As the Higgs boson comes from the decay of the massive T', it is produced with a momentum different from zero. As a result the two b's from the Higgs boson are produced close by in $\eta - \phi$ space.

- *Cut 5*: As the Higgs boson and top quark produced by the signal come from a very heavy particle, they are expected to have a greater p_T than fakes reconstructed in backgrounds. Accordingly, only events which have a Higgs candidate with $p_T > 200$ GeV/c and a top candidate with $p_T > 300$ GeV/c were selected. The p_T of the reconstructed Higgs candidate and top quark candidate for signal and backgrounds are shown in figure 3.5, in a 2D histogram.

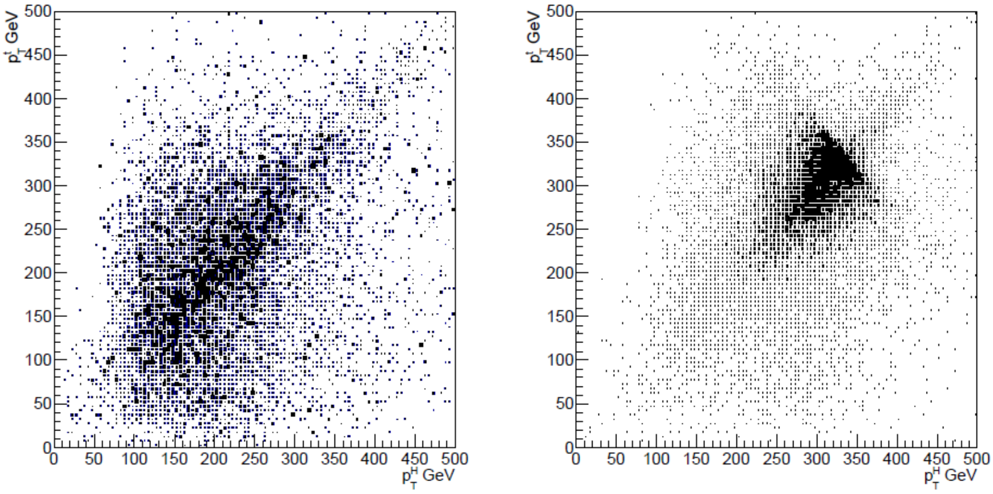


Figure 3.5: Reconstructed Higgs candidate p_T in the x axis and reconstructed top candidate p_T in the y axis for backgrounds (left) and signal (right). Signal events have reconstructed Higgs and top with higher p_T than backgrounds.

- *Cut 6*: The distance in $\eta - \phi$ plane between the reconstructed $W^{+/-}$ and Higgs candidates is preferentially around 3 for signal, while for backgrounds the distribution is much more spread as there is no real Higgs boson. ΔR_{HW} is plotted in figure 3.6. Selecting only the events within $2.2 < \Delta R_{HW} < 3.5$ helps to reduce QCD and $W^{+/-}$+jets background events.

- *Cut 7*: The $\Delta\phi_{bb}$ of the b jets identified as coming from the Higgs boson candidate and the $\Delta\phi_{bW}$ between the reconstructed W candidate and the jet which formed the top are expected to be mainly close in ϕ for the signal

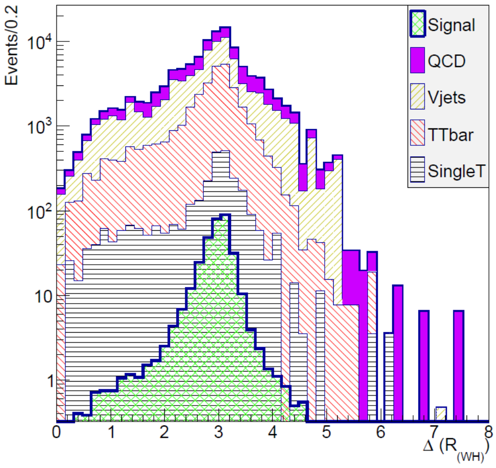


Figure 3.6: ΔR between the reconstructed Higgs and $W^{+/-}$ candidates for backgrounds (stacked) and signal (over–imposed) normalized to 20 fb^{-1} luminosity. Although signal and backgrounds have a mean value of $\Delta R_{HW} = 3$, backgrounds tend to have lower values than signal, as well as larger tails.

while more evenly distributed for backgrounds. Only events with $\Delta\phi_{bb} < 2.0$ and $\Delta\phi_{bW} < 3.3$ were kept. This cut is specially useful for reducing QCD and $W^{+/-}$+jets background events.

- *Cut 8*: As in cut 7, the $W^{+/-}$ boson produced from the T' is expected to have a non-zero momentum. Thus, the $\Delta\phi_{jj}$ between the jets of the $W^{+/-}$ boson are expected to be more centered around zero for the signal with respect to the backgrounds. Events were required to have $\Delta\phi_{jj} < 2.3$ to reduce single–top background.

- *Cut 9*: Events with a reconstructed Higgs mass close to the Higgs mass were kept. Events with a Higgs candidate with a mass between 100 GeV/c^2 and 135 GeV/c^2 were selected for the analysis. The distribution of the reconstructed Higgs mass is shown in figure 3.7.

- *Cut 10*: For the final cut the relative total hadronic energy is defined as the ratio between the p_T of the decay products identified as the Higgs candidate and top quark candidate and the total hadronic energy of the event:

$$\frac{p_T^H + p_T^t}{H_T}\,.$$

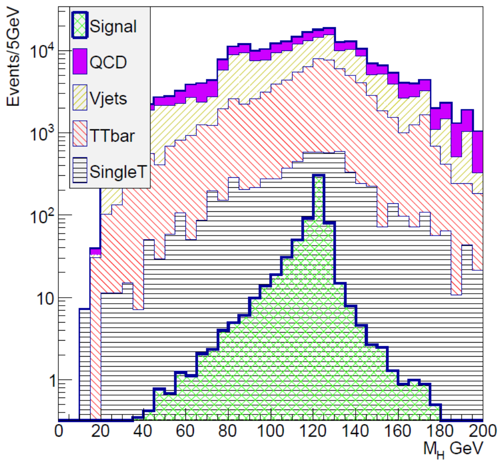


Figure 3.7: Mass of the reconstructed Higgs candidate for backgrounds (stacked) and signal (over–imposed). Backgrounds have larger tails than signal for the reconstructed Higgs mass.

This variable is specially useful to discriminate signal from $t\bar{t}$ events. Events were required to have a relative total hadronic energy larger than 0.65. The relative total hadronic energy is shown in figure 3.8.

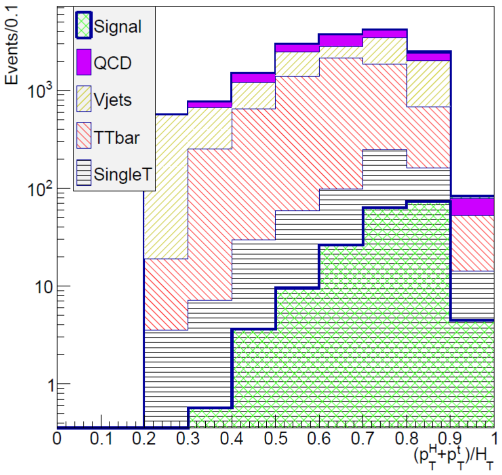


Figure 3.8: Relative total hadronic energy for backgrounds (stacked) and signal (over–imposed) normalized to 20 fb^{-1} luminosity.

After cut 3, but before cut 4, the $W^{+/-}$, top and Higgs candidates are formed with the jet association method. The mass distribution of $W^{+/-}$ and t candidates is plotted in figure 3.9.

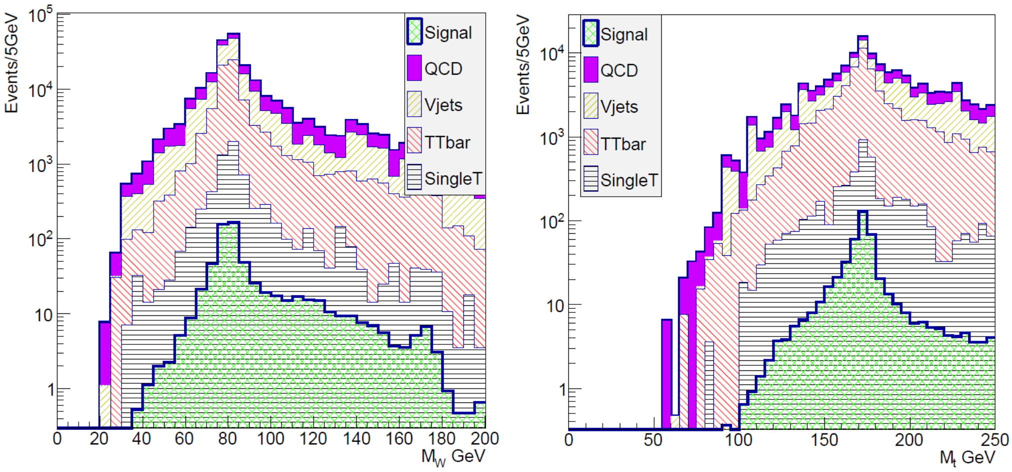


Figure 3.9: Reconstructed $W^{+/-}$ [left] and top [right] mass for backgrounds (stacked) and signal (over–imposed) normalized to 20 fb^{-1} luminosity.

In the selection, the characteristics of the signal have been exploited in order to differentiate it from the backgrounds. Table 3.2 shows the efficiency of each cut in the procedure described above: the first line contains the number of events after the initial cut 0 normalized to an integrated luminosity of 20 fb^{-1}, while in the following lines the efficiencies of the cuts for signal and background samples are shown after each of the 10 cuts. The bottom line of the table contains the overall selection efficiency.

Due to the low statistics of our samples, no numbers are quoted for the Z^0+jets background (producing a Z^0 boson with a high jet multiplicity is computationally very expensive). Furthermore, the inclusive Z^0+jets cross section is one order of magnitude smaller than the $W^{+/-}$+jets one and, due to the similar branching ratios and kinematics, similar efficiencies are expected as for the $W^{+/-}$+jets background. This background is therefore ignored, and a contribution of at most with an additional 10% of the $W^{+/-}$+jets is assumed. In addition, from the study it was checked that the di–boson contribution is negligible. No di–boson background is quoted either.

The selection relied on angular variables that are not greatly changed by the introduction of detector effects. Correspondingly, nonetheless this study has been performed up to hadronization level in the samples production, it should not greatly change when considering detector simulation.

	Signal	QCD	$W^{+/-}$+jets	$t\bar{t}$	t+ jet	tW
Cut 0	554 ± 3	$203,930 \pm 1,150$	$1,015,294 \pm 11,567$	$337,024 \pm 1,608$	$25,349 \pm 300$	$19,416 \pm 469$
Cut 1	0.91 ± 0.01	0.571 ± 0.007	0.67 ± 0.02	0.439 ± 0.005	0.45 ± 0.01	0.42 ± 0.03
Cut 2	0.92 ± 0.01	0.68 ± 0.01	0.74 ± 0.02	0.81 ± 0.01	0.61 ± 0.02	0.70 ± 0.06
Cut 3	0.84 ± 0.01	0.86 ± 0.02	0.22 ± 0.01	0.83 ± 0.01	0.82 ± 0.04	0.85 ± 0.08
Cut 4	0.93 ± 0.01	0.68 ± 0.01	0.74 ± 0.06	0.56 ± 0.01	0.49 ± 0.03	0.45 ± 0.05
Cut 5	0.92 ± 0.01	0.60 ± 0.02	0.56 ± 0.05	0.53 ± 0.01	0.61 ± 0.05	0.56 ± 0.09
Cut 6	0.92 ± 0.01	0.61 ± 0.02	0.56 ± 0.07	0.74 ± 0.03	0.66 ± 0.07	0.72 ± 0.15
Cut 7	0.75 ± 0.01	0.67 ± 0.03	0.67 ± 0.11	0.71 ± 0.03	0.77 ± 0.09	0.75 ± 0.18
Cut 8	0.87 ± 0.02	0.76 ± 0.04	0.82 ± 0.15	0.84 ± 0.04	0.77 ± 0.11	0.90 ± 0.24
Cut 9	0.91 ± 0.02	0.33 ± 0.02	0.41 ± 0.10	0.51 ± 0.03	0.52 ± 0.09	0.48 ± 0.16
Cut 10	0.87 ± 0.02	0.54 ± 0.06	0.55 ± 0.19	0.49 ± 0.04	0.79 ± 0.17	0.72 ± 0.31
combined	0.284 ± 0.005	$(7.5 \pm 0.5) \times 10^{-3}$	$(3.1 \pm 0.7) \times 10^{-3}$	$(9.3 \pm 0.5) \times 10^{-3}$	$(11 \pm 1) \times 10^{-3}$	$(10.5 \pm 3) \times 10^{-3}$

Table 3.2: Number of events for signal and backgrounds after the first selection cut (cut 0), and efficiencies of each stage of the cutting procedure. The errors indicated are statistical only, based on the number of events.

3.4 Results

The T' peak reconstruction after full selection is shown in figure 3.10. An excess of signal events over backgrounds is found around 730 GeV/c^2 for the signal over the ensemble of backgrounds. The lack of smoothness of the distribution is due to the lack of statistics in the Monte-Carlo samples for the backgrounds, especially for $W^{+/-}$+jets. These fluctuations due to the poor statistics can change the final estimate of the number of background events entering the peak of the signal.

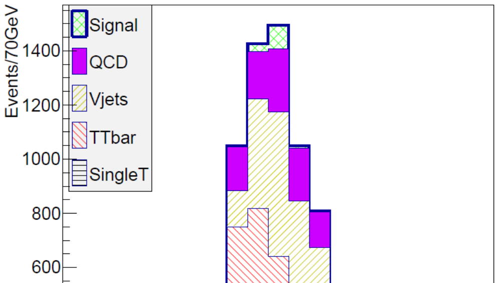


Figure 3.10: Reconstructed T' mass after all cuts for backgrounds and signal (stacked) normalized to 20 fb^{-1} luminosity. Signal peak is visible on top of the sum of backgrounds.

	unweighted events		weight	weighted events
	after Cut 10	in mass window		
Signal	8601	3780	0.018	69 ± 1
$t\bar{t}$	409	57	7.7	437 ± 58
W+jets	24	3	132	395 ± 228
QCD	235	34	6.48	220 ± 38
tW	18	3	11.3	34 ± 20
t+ jet	75	7	3.55	25 ± 9
total background				1112 ± 352

Table 3.3: Number of signal and background events from utilized MC samples: in the first column the simulated events that pass all kinematic cuts, in the second column the events that fall in the mass window $710 < M_{jjjjj} < 750$ GeV/c^2, finally in the fourth column the number of weighted events in the mass window normalized to the physical cross section (the applied weight is listed in the third column). All the errors are statistical only. For the total background, the linear sum of errors was considered.

From the peak reconstruction, the number of events falling into a window of 20 GeV around the T' mass was selected, i.e. within $710 < M_{jjjjj} < 750$ GeV. The number of events contained in this window, for each process, is listed in table 3.3. An enhanced signal over background ratio is obtained, with:

$$\frac{S}{\sqrt{S+B}} = 2.0 \pm 0.3\,, \qquad \text{and} \quad \frac{S}{B} = 0.06 \pm 0.02\,. \tag{3.1}$$

To the quoted uncertainties, uncertainties linked to the cross section calculation for the signal should be added, PDF's and possible loop contributions. From similar studies and analyses done by ATLAS and CMS collaborations the uncertainties linked to these sources are not bigger than 10% to 15% (see for instance [123]), therefore their inclusion should not change significantly the conclusions of this study.

As final remark, the signal sample used for the study was generated with a T' width of 1 GeV/c^2. However, the theoretical model predicts a width of 11 GeV/c^2 for the benchmark point that has been used. For this larger width, an opening up to 30 GeV/c^{2}on the integration window is needed to include 2-σ of the signal. Nonetheless, with these changes the estimator $S/\sqrt{S+B}$ does not change significantly within the statistical error. Then, the conclusion of the study does not change taking into account the larger T' width.

In conclusion, this study shows a plausible selection in order to perform a data analysis to test a hypothetical T' that decays into a top quark and a Higgs boson in the full hadronic channel. Moreover this study shows, how mixed couplings of VLQ to light and heavy quark generations enhance the VLQ production cross section; and also give specific signatures to search for these VLQ beyond SM particles.

It is important to stress that the present study, as the CMS analysis that will be presented afterward, rely in a VLQ model that has been poorly explored in experiments. This model allows the VLQ to be mixed to the three generations of SM quarks, enhancing the production cross section in proton-proton collisions. The most common VLQ models assume that the top-partner mixes preferentially to third SM-quark generation. The model used in this work represents then a generalization with respect to the most widely used models in the literature.

In chapter 5, a data analysis using data collected by CMS experiment during run 1 is presented. The present work concludes with the experimental limits set on this model.

4 Monte-Carlo event simulation

Although nowadays a very elegant and complete theoretical description of particle physics exists, it is not always evident how to translate this theory in actual predictions to compare with measurements. Moreover, on the case of hadronic colliders, as the LHC, it is even more difficult due to the characteristics of the strong interaction. On this subject, a set of tools and approaches have been developed in order to be able to make accurate predictions from theory that could be directly tested for experimentally, for example by ATLAS or CMS experiments. In the present chapter, such tools and formalisms and a set of studies comparing their predictions to data are described.

4.1 Monte-Carlo simulations

The Monte-Carlo (MC) simulations use random numbers and large samplings to calculate mathematical quantities in complex configurations, as integrals or probabilities. The typical example is on how to calculate the integral of a one-dimensional function. Several random points can be thrown in the Cartesian plane and count how many of them are under the function. Then the integral of the function will be proportional to the fraction of points under the curve to the total used points. Larger the number of points, closer the estimation to the real value. An illustration of the procedure is shown in figure 4.1.

A similar method is utilized to simulate proton-proton collisions. This simulation is used to generate “random” events and to calculate quantities, as the cross section, for a given physical process. Each event represents the final state of a collision, i.e. the set of particles produced from the collision and detected by the apparatus. Such simulations comprehend different steps: first, the partonic processes making reference to the interaction between the partons inside the proton; second, the hadronization of the particles produced from parton interactions; and third, the simulation of the interaction between the hadrons (from second step) and the detector materials. Such events are used to evaluate predictions from theory in the frame of a specific experiment. Whereas the hadronization and detector simulation are well-known physical processes, new theoretical predictions rely basically on the partonic level, where the fundamental interaction processes take part.

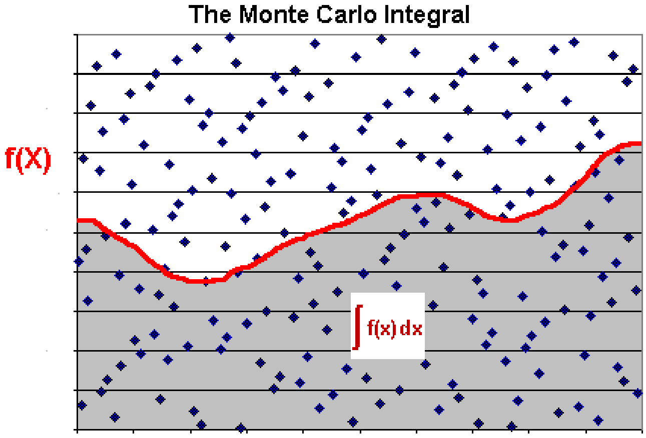


Figure 4.1: Integration using Monte-Carlo methods. The points in the plane have been chosen randomly. The integral of the plotted function is proportional to the ratio of points under the curve to the total number of used points.

4.1.1 Parton simulation

The parton model was initially proposed by Richard Feynman in 1969 [124], as a method to understand collisions of non-fundamental particles. The model considers a composed particle, as a proton or a neutron, formed by a given number of point-like fundamental particles. When a collision occurs the point-like particles inside have a major probability to scatter. For example, when an electron collides with a proton, most of the interactions will be between the electron and the fundamental charged components of the proton, u and d quarks. This "hard" components are called valence quarks. Surrounding them, there is the sea of quarks and gluons.

However, as the energy of the collision increases the probability to scatter a sea component, quark or gluon, increases. In addition, even if the valence quarks of a proton are the u and d quarks, heavier quarks can appear in the sea, as the b, c or s quarks. The probability to interact with a component, a valence one or from the sea, is described by parton distribution function, commonly called PDF. A PDF function $f \equiv f(x, Q^2)$ represents the number density of a given quark or gluon as a function of the resolution scale Q^2 and the fraction of momentum carried by the parton x. The determination of a PDF is done via a fit of large data samples from experiments specifically designed to test the inner structure of nucleons. The SLAC center (Stanford Linear Accelerator Center) [125], in California, United States, first probed the existence of partonic structure inside nucleons using leptons as probes scattered against nucleons, with experiences known as deep inelastic scattering (DIS) experiments [126]. Another important experiment was the HERA [127] accelerator at DESY in Hamburg, Germany, which used electrons to study the inner structure of protons [128].

In figure 4.2 the Martin-Stirling-Thorne-Watt [129] (MSTW) PDF for two resolution scales is shown, taken from [129]. The MSTW PDF is one of the experimental fits combining data from DIS and HERA. From this PDF, it is shown that u and d quarks carry the most of the momentum of the proton. The rest of the momentum is spread mainly over a huge amount of gluons and some, less probable, sea quarks as $\bar{u}, \bar{d}$ or c and s. One important feature is that the composition of the proton changes depending on the resolution scale. At $Q^2 = 10\,\mathrm{GeV}^2$ there is no probability to find a b-quark in the proton while at $Q^2 = 10^4\,\mathrm{GeV}^2$ there is a non-negligible probability to find it.

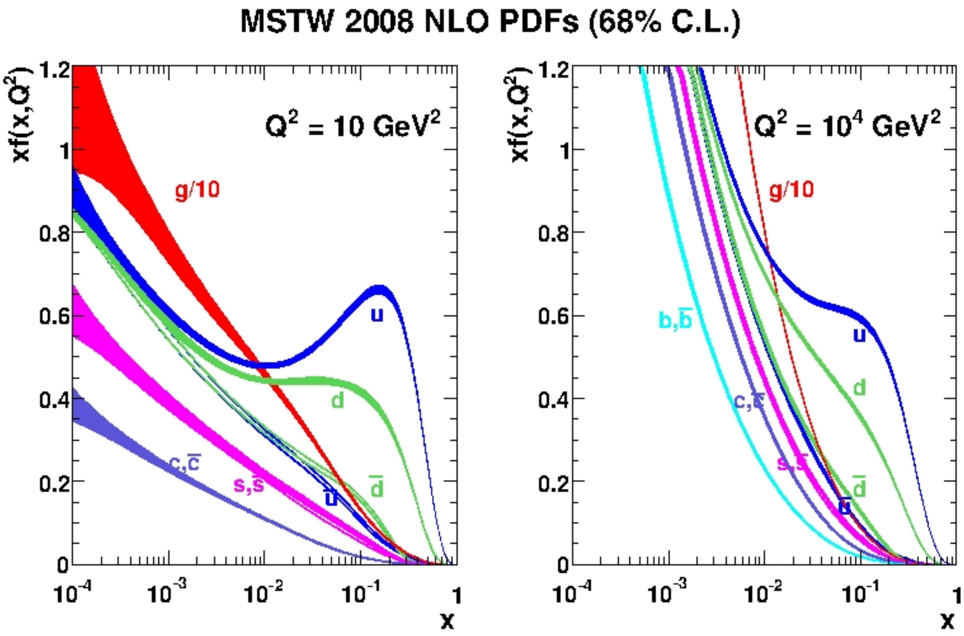


Figure 4.2: Martin-Stirling-Thorne-Watt proton PDF for $Q^2 = 10\,\mathrm{GeV}^2$ [left] and $Q^2 = 10^4\,\mathrm{GeV}^2$ [right]. x is the fraction of momentum carried by the parton and $f(x, Q^2)$ the PDF function [129].

Two other important PDF fits are CTEQ [130] and NNPDF [131]. Together with MSTW, they are the most used PDF sets in the CMS experiment for MC production.

The knowledge of the proton content allows to compute the cross section of processes proton-proton collisions. This cross section depends of the momentum carried by each parton inside the parton, and are modulated by the probability of the interaction between different quarks during the collision.

For a scattering process, the hard part of the differential cross section can be written as,

$$\begin{aligned} d\sigma_{ij\to lm} &= \left(\int_0^1\int_0^1 f_i(x_i,Q^2)f_j(x_j,Q^2)dx_i dx_j\right) \\ &\times \frac{d^3p_l}{(2\pi)^2 2E_l}\frac{d^3p_m}{(2\pi)^2 2E_m}\delta^4\left(p_i+p_j-p_l-p_m\right) \\ &\times |\mathcal{M}_{ij\to lm}|^2 \end{aligned} \tag{4.1}$$

where $f_{i,j}$ corresponds to the PDF's of the initial partons. $\mathcal{M}_{ij\to lm}$ is the matrix element of the process which is the part of the S-matrix that contains the amplitude of the process, and governs the transition from the initial to the final state [132]. The matrix element could account effectively for all processes mediating the transition from the initial to a given final state, but in practice it is calculated only with a given number of processes. The calculation can achieve different levels, usually tree level or Leading Order (LO), but modern calculation could arrive, depending on the process, to one loop or Next-to-Leading-Order (NLO) or even two loops the Next-to-Next-to-Leading-Order (NNLO). This limit depends exclusively on the feasibility of the theoretical calculations. Figure 4.3 shows an example of a leading order and its corresponding NLO diagrams for a fermion scattering.

Parton simulations try to simulate the interactions between partons in hadronic collisions. They are the basic level for many particle physics simulation of events, where the basic pieces, as the matrix element of the quarks interaction, are calculated.

4.1.2 Hadron simulation

Quarks produced in a collision are not seen freely due to the strong interaction. They will take quarks from vacuum to form hadrons. What reaches the detector as final state are the stable hadrons resulting from the hadronization process. From a single parton several hadrons can result, producing a shower of particles, as seen in figure 4.4. Two main processes are simulated, the showering that corresponds to radiations emitted by quarks, from initial or final state, and hadronization the process of forming hadrons from quarks and gluons.

As QCD strong interaction imposes several theoretical restrictions to have a first principle understanding of these phenomena, the description of hadronization relies on the construction of effective models. There are basically two main models to

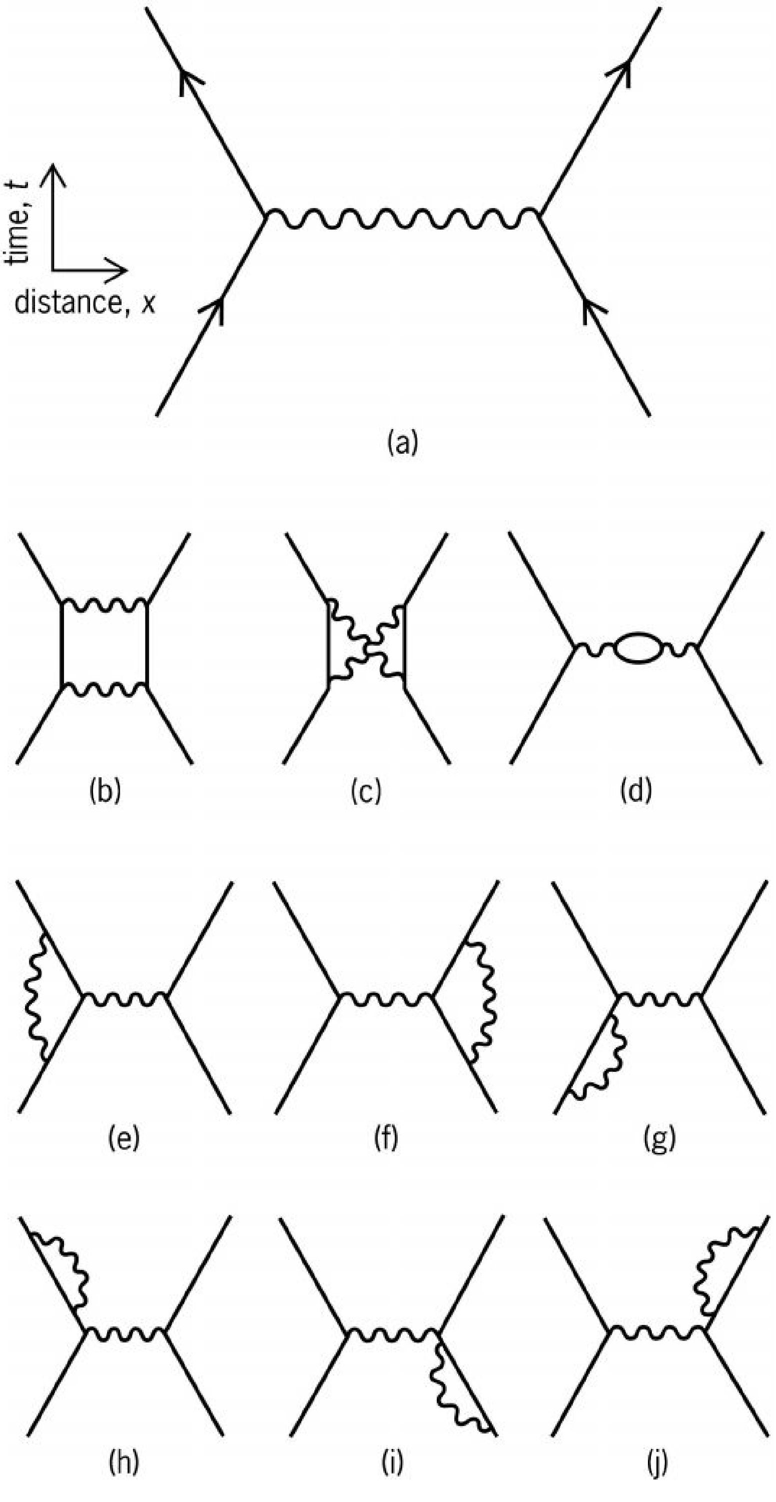


Figure 4.3: LO (a) and NLO (b)-(j) processes contributing to fermions scattering. Several processes have to be considered in order to have accurate predictions for experimental measurements.

simulate hadronization and showering of quarks and gluons, both giving comparable predictions. It is important to remark that this is a crucial step in simulations as from an accurate hadron production simulation depends the correctness of MC description of jets.

4.1.3 Detector simulation

After final particles from a collision are simulated, the next step is to simulate the interaction of these particles with the detector. The principle is that the detector response should be simulated as close as possible to the real detector. In that sense, the objective is to have a detector response using MC simulations as with real data. However, not everything can be simulated. For example, during data taking different problems might arise (subdetector misbehaving, dead cells, etc.) that can not be adjusted in MC simulation, and needs further correction and tuning. Moreover, in practice, MC simulations for ATLAS and CMS experiments are prepared before data taking. Then, if one wants to reproduce precisely the details of what happened with the detector during the data taking, MC simulation should be redone to reproduce the data taking particularities.

CMS has used GEANT 4 [133] software to simulate the detector. Precise implementation of detector geometry was done in order to correctly simulate each subdetector, their details as size, number of channels, cells, electronic cards, were taken into account. This simulation allows to get, as for data, the same output from the detector to be used for the reconstruction and identification of the objects.

4.2 Tools

There are several tools on the market to perform the different steps of MC simulations of proton-proton collisions. I'll briefly describe some of them, the most used in CMS.

4.2.1 Matrix-element generators

Regarding parton simulation, MadGraph [116, 117] package is one of the most widely used both by experimental and theoretical particle physics communities. It calculates matrix-elements, LO cross sections and particle widths. With such information, it generates MC events for parton collisions.

The latest series of releases also include an additional package to perform NLO QCD corrections to SM processes. The framework also includes the possibility to work with BSM models, making it a powerful tool for evaluating predictions from them. Another parton generator is POWHEG, in its most recent version

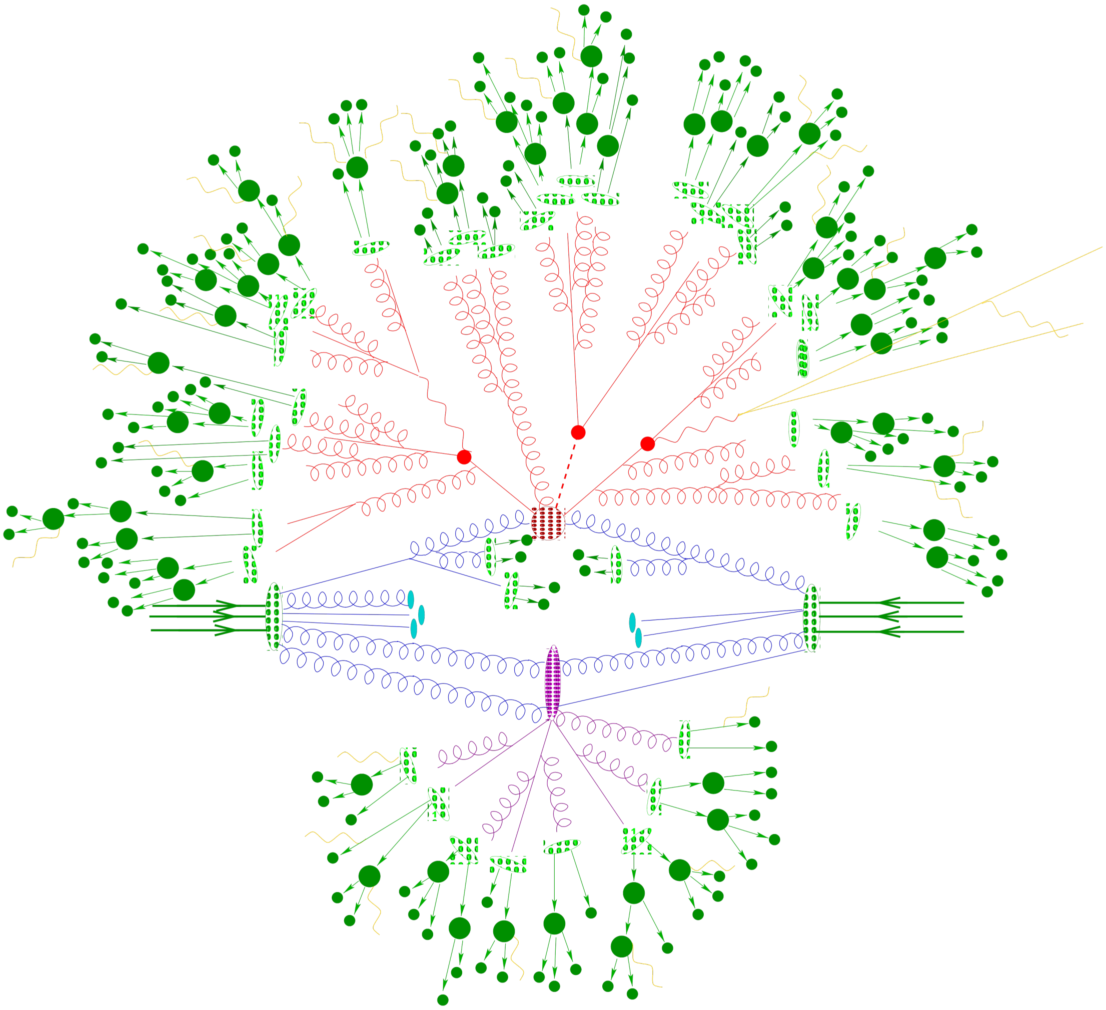


Figure 4.4: Graphical representation of the hadronization process of partons resulting from a proton-proton collision. The biggest green ellipses, preceded by three green arrows, represent the initial protons. The small green circles inside them represent the partons inside the protons. The big red circle, with small red circles inside, represent the hard interaction from partons (in blue lines). The hard interaction produces three particles, red full small circles, that decay afterward. The red lines represent the decay products from the particles produced in the hard interaction, and the additional particles produced from showering. Final hadrons are drawn in the most exterior part of the graph. These final hadrons are represented by the small full green circles.

POWHEG BOX [134, 135, 136], that is specially accurate to simulate events with top quarks.

In CMS, all physics groups analyzing data collected by the experiment, have to do simulations of signals and backgrounds. For this purpose, MC simulation tools as MadGraph, need to be interfaced with the central CMS software CMSSW (CMS SoftWare). The generators group do this work preparing sets of validated code to be used later by CMS users. This group is also in charge of central MC production, the technical process of MC production of samples doing all the needed steps to obtain validated samples to compare with CMS data: parton generation, hadronization and detector simulation.

Central production of samples using MadGraph is done via the generation of stand-alone packs of code used to launch parallelized production. This utility is known as gridpack. Gridpack generation must include all possible configurations needed for the CMS central production: settings to run within CMSSW in batch systems and additional physics requirements depending on the process to be generated. A special script has been put in place to allow users to do their own gridpack generation with small effort in terms of automatization of common tasks.

In addition, the generator group also provide constant support for users to test the validity of their production, their MadGraph configuration, the usage of the interface software in CMSSW and to test implementations of BSM models.

Preparation of MadGraph production

Any production of partonic MC samples with MadGraph relies on 4 key elements, a model and three production cards. In order to launch any event generation in MadGraph, a set of three cards has to be prepared. These cards are text files where the user can specify several parameters used by MadGraph to generate a sample. In more detail, the 4 elements are:

- Model: The implementation of the theoretical model according to which the production will be done. The implementation of a particle physics model, for computational calculations, is not a trivial task. Tools are available to perform this. MadGraph has incorporated as default the model implementation developed by FeynRules [118], the Universal FeynRules Output (UFO) [137]. A complete set of tools to facilitate the implementation of theoretical models for computational work with different packages was developed in FeynRules. MadGraph includes by default the implementation of the SM and the Minimal SuperSymmetric Model (MSSM), but in order to generate events for private models, its implementation has to be provided. The facility to generate events for any model has been one of the most appealing features of MadGraph for the particle physics community.

- Process card: In this card, the user defines the model to be used and the process to be generated. In general, the process definition could use the definition of multi-particle labels that MadGraph utilizes to loop over all the particles contained in the label. Multi-particle labels are variables defined by the user to design sets of particles from the model. Figure 4.5 shows the lines contained in the process card to generate $t\bar{t}$ with 0 and 1 additional jet. In this example, the proton is defined as a multi-particle label containing all the quarks (except the top) and the gluon. Afterward, positively and negatively charged leptons are grouped under multi-particle labels "l+" and "l-". Similar labels are declared for neutrinos and anti-neutrinos. After the declaration of labels, the process for the production is declared, in this case $t\bar{t}$ production with 0 and 1 additional jet.

```
import model sm
# Define multiparticle labels
define p = u c s d b u~ c~ s~ d~ b~ g
define j = p
define l+ = e+ mu+ ta+
define l- = e- mu- ta-
define vl = ve vm vt
define vl~ = ve~ vm~ vt~
# Specify process(es) to run
generate p p > t t~  @0
add process p p > t t~ j @1
# Output processes to MadEvent directory
output -f
```

Figure 4.5: MadGraph process card example.

- Parameters card: This card is used to set the parameters of the model invoked for the production. Following the example of top pair production, it might be interesting to produce different samples varying the top mass. In figure 4.6 an extract of the card that controls the masses of the particles in the model is shown. The first column refers to the PDG ID of the particles, 6 for the top quark, and the second column is the value of the mass. In this case, the top mass has been set to 172.5 GeV/c^2.

- Run card: In the run card several settings can be fixed related to the generation of the events. For example, the center of mass energy or the minimal and maximal ΔR between jets. For generation of events requiring more than one

```
###################################
## INFORMATION FOR MASS
###################################
Block MASS
    4 1.420000e+00 # MC
    5 4.800000e+00 # MB
    6 1.725000e+02 # MT
   11 5.110000e-04 # Me
   13 1.056600e-01 # MM
   15 1.777000e+00 # MTA
   23 9.118800e+01 # MZ
   25 1.250000e+02 # MH
```

Figure 4.6: Extract of MadGraph parameters card. The first column refers to the PDG ID of the particles, 6 for the top quark, and the second column is the value of the mass. In this case, the top mass has been set to 172.5 GeV/c^2.

additional extra jet, a parameter called xqcut can also be set in this card. This parameter is relevant for the merging procedure, important to avoid double counting, a topic that will be discussed in section 4.2.2. In figure 4.7 an extract of a run card is shown, where the parameters corresponding to the center of mass energy of the collision, the number of events to be generated, the gridpack generation, the choice of PDF set and the merging procedure are displayed.

MadGraph release validation

MadGraph is a very dynamic tool, where the input from the users is being taken into account permanently: users give reports of possible bugs or recommendations in the way tool works. New releases of MadGraph are constantly delivered to solve found issues or to improve weaknesses. For CMS production, not every new release is automatically incorporated. In order to begin using a new release a validation procedure is done, which is time and resources demanding. Basic processes of the SM are considered using the last release used by CMS and then compared to the new release. Such comparison is performed to control that predictions from both releases are identical, for processes where no changes have been performed between both releases.

For example, one of the control processes for this validation is the SM production of a Z^0 boson that decays into neutrinos produced with extra jets. MadGraph is

```
#*********************************************************************
# Run to generate the grid pack                                      *
#*********************************************************************
  .true.     = gridpack  !True = setting up the grid pack
#*********************************************************************
# Number of events and rnd seed                                      *
#*********************************************************************
  1000       = nevents ! Number of unweighted events requested
      0       = iseed   ! rnd seed (0=assigned automatically=default))
#*********************************************************************
# Collider type and energy                                           *
#*********************************************************************
        1     = lpp1  ! beam 1 type (0=NO PDF)
        1     = lpp2  ! beam 2 type (0=NO PDF)
     6500     = ebeam1  ! beam 1 energy in GeV
     6500     = ebeam2  ! beam 2 energy in GeV
#*********************************************************************
# PDF CHOICE: this automatically fixes also alpha_s and its evol.    *
#*********************************************************************
 'cteq6l1'    = pdlabel     ! PDF set
#*********************************************************************
# Matching - Warning! ickkw > 0 is still beta</div>
#*********************************************************************
 1        = ickkw            ! 0 no matching, 1 MLM, 2 CKKW matching
 1        = highestmult      ! for ickkw=2, highest mult group
 1        = ktscheme         ! for ickkw=1, 1 Durham kT, 2 Pythia pTE
 1        = alpsfact         ! scale factor for QCD emission vx
 F        = chcluster        ! cluster only according to channel diag
 F        = pdfwgt           ! for ickkw=1, perform pdf reweighting
#*********************************************************************
# Jet measure cuts                                                   *
#*********************************************************************
 20   = xqcut   ! minimum kt jet measure between partons
#*********************************************************************
```

Figure 4.7: Extract of MadGraph run card.

able to produce extra jets up to a multiplicity of 4. In figure 4.8 the comparison between samples produced with MadGraph 1.4.8 and 1.5.11 is displayed. The results of both releases are equivalent for the specific considered process.

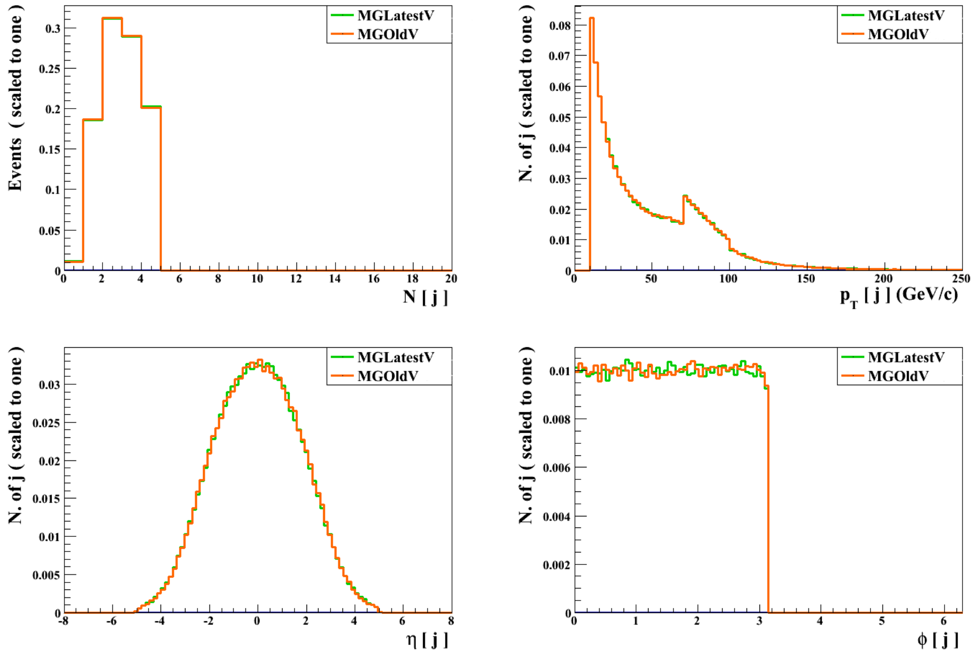


Figure 4.8: Comparison between 1.4.8 (MGOldV) and 1.5.11 (MGLatestV) MadGraph releases for SM Z^0+jets production with Z^0 decaying into neutrinos. Number of jets [left-up], p_T of jets [right-up], η of jets [left-down] and ϕ of jets [right-down]. As expected, no deviation is observed.

In addition, figure 4.9 shows two global variables, the missing energy and the total hadronic transverse energy (defined in equation 4.2), for the same process where the validation was performed.

$$\not{E}_T = |\sum_{visible\ particles} \vec{p}_T|, \quad H_T = \sum_{hadronic\ particles} |\vec{p}_T| \tag{4.2}$$

In the study shown in the mentioned figures, there is no difference between the two compared releases for the processes used.

Incorporating new releases improvements

New MadGraph releases are interesting not only because of bugs being fixed but also because they include new features and improvements. For example, latest set

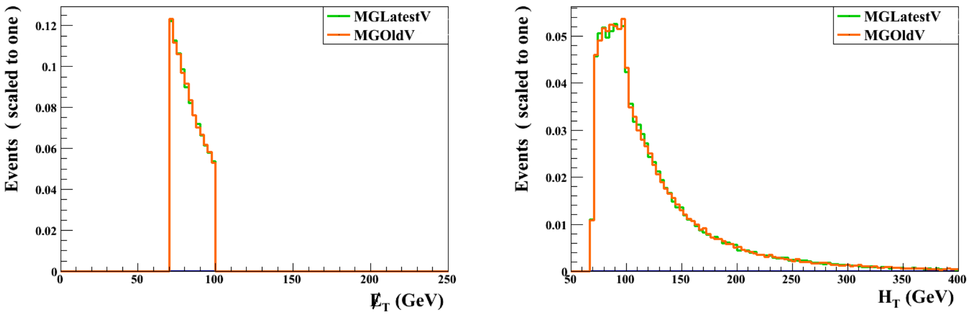


Figure 4.9: Comparison between 1.4.8 (MGOldV) and 1.5.11 (MGLatestV) MadGraph releases for SM Z^0+jets production with Z^0 decaying into neutrinos. Missing energy [left] and total hadronic transverse energy [right]. No difference is seen between both versions.

of releases include NLO corrections for SM processes, improvement not present before. NLO corrections are crucial to do precise measurements, that is one of the major interests in various SM sectors. Top physics has become a very important sector to perform precise measurements. Any small deviation of SM predictions related to the top quark could be related to the presence of BSM physics. To this extent, there is a growing need to have better MC simulations of top quark processes.

From a technical point of view, a complete simulation of top decay chain was not feasible at MadGraph level. It was extremely time and CPU consuming for old MadGraph releases. Modern releases (from MadGraph 5) have improved the performance of these calculations but also have included new tools that allow to simulate the entire top decay at parton level chain with high precision and small resources. One of these tools is MadSpin [138, 139], which was developed to preserve spin correlations between decay products. However MadGraph alone is able to simulate the top decay, it is still extremely CPU and time consuming, and then, not suitable for big central production at CMS. On the other hand, decaying the top quark with Pythia erases all spin correlations. In CMS, MadSpin has been included in the gridpack generation production mainly to generate top pair production samples. MadSpin is the modern version of the old DECAY tool, that did not preserve spin correlations.

However to consider using MadSpin in CMS, an improvement with respect to former samples should be shown. MadSpin is expected to have the same results as MadGraph, but with an improvement in terms of computational resources consumption. Figure 4.10 shows the top quark mass reconstructed from its hadronic decay (b, u and $\bar{d}$ quarks). This figure demonstrate that MadSpin and MadGraph produce comparable results with a non-zero top width, while DECAY gives a zero width. The ΔR between the positively and negatively charged decay products of the $W^{+/-}$ bosons produced in top pair events is displayed in figure 4.11. This observable is different in the case where there are or no spin correlations.

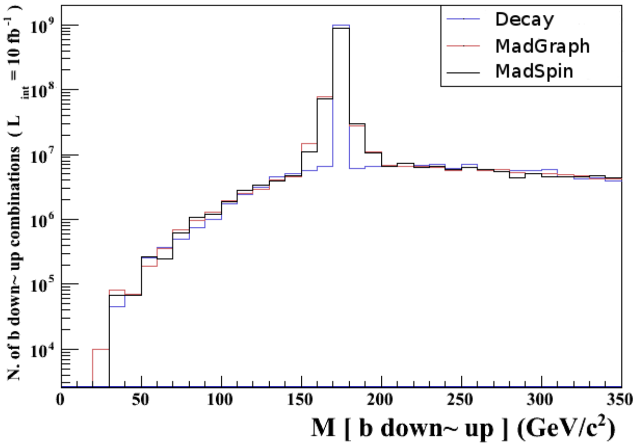


Figure 4.10: Top mass reconstructed from its hadronic decay using different tools to simulate the top decay for $t\bar{t}$ events at $\sqrt{s} = 8$ TeV. MadSpin and MadGraph correctly reproduce the top width.

4.2.2 Hadron generators

Taking as input partonic events, hadron generators perform their corresponding showering and hadronization. The most used hadronizers in CMS are Pythia [119], in releases 6 and 8, and Herwig ++ [140]. They are used to simulate a wide range of SM processes simulation. They can be used as MC generators, without the parton step, which is specially used to simulate QCD processes.

Hadronization processes are QCD mediated. As the strong force is a non-perturbative interaction, different hadrons simulations implement different effective models. Reason why when comparing simulations to data, different hadron generators are used in order to understand the set of data independently from the hadronization model.

Interface between partonic and hadronic generation of events

Matrix element (ME) generation is responsible for the simulation of the hard processes. This corresponds to generating all the final quarks and gluons coming

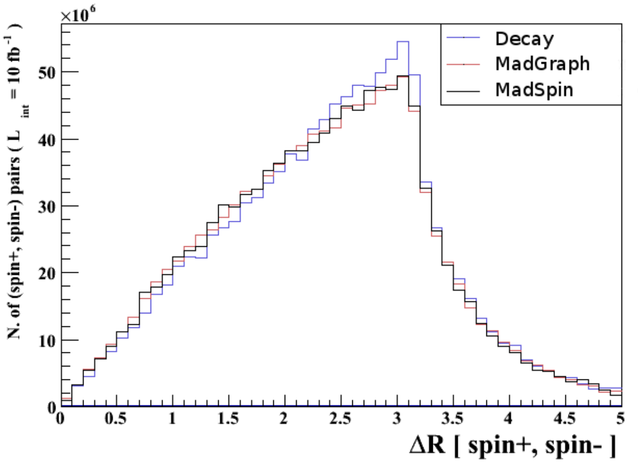


Figure 4.11: ΔR from positively and negatively charged decay products using different tools to simulate the top decay for $t\bar{t}$ events at $\sqrt{s} = 8$ TeV. MadSpin and MadGraph produce different results than the DECAY tool, as these first two preserve spin correlations.

from the collision. Parton showering (PS) is responsible for the simulation of "soft" processes coming from the strong interaction. For example, an initial or final quark can radiate a gluon that will lead to additional jets in the event. The generation of extra jets (coming from extra quarks or gluons) can suffer from an overlap between the ME and PS. In figure 4.12, a graphical description of the procedure followed by ME and PS to add extra jets to a process is shown.

This overlap can lead to double counting events, which has as consequence an overestimation of the cross section of the process. To solve this issue a procedure has been designed to study the overlap between MadGraph and Pythia. This procedure is known as merging [141, 142, 143]. It is based on the application of one or two cuts to separate hard and soft phase space, hard region being generated by ME and soft region by PS. MadGraph uses MLM merging formalism to do this task [144, 145].

In general, the merging procedure follows a set of common steps:

- A jet algorithm is chosen and all relevant cross sections are calculated accordingly to the process under consideration. The calculation is performed for each final jet multiplicity. For example, when generating top pair production up to 4 additional jets, the cross section is calculated for $t\bar{t}$ with 0, 1, 2, 3, and 4 jets separately.

- ME samples are produced with a probability corresponding to their cross

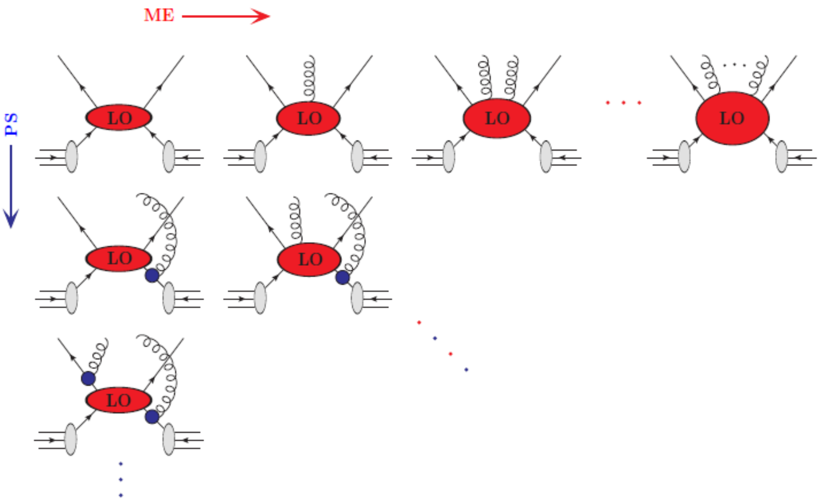


Figure 4.12: ME and PS contributions to events with several partons in the final state in a leading order MC simulation.

section.

- Events or partons are rejected based on specific procedure requirement. For MLM formalism, events are rejected if not all partons are matched to jets.
- Parton shower is invoked, constrained to not produce any extra jet.0

MadGraph utilizes a MLM merging procedure with the k_T jet algorithm. The MLM prescription proceeds as follows:

1. Generation of parton level events constrained to the requirements defined by the user:

   $$p_T^{parton} > p_T^{min}, \ |\eta^{parton}| < \eta^{max}, \ \Delta R_{partons} > R_{min}$$

   where *min* and *max* stand for the minimal and maximal values required by the user for each variable.

2. The k_T jet algorithm is used to find the jets of the event from partonic events. This first reconstruction uses a cut on the minimal k_T distance as defined in equation 2.9. This first minimal cut is set, by the user, in MadGraph with the variable xqcut (see 4.7). Q^X_{cut} will be used to refer to it. This cut is applied at generator level, what means that MadGraph will only deliver events where every pair of partons satisfy this requirement.

3. Running of the parton shower and hadronization over the partonic events, and rerunning of the jet algorithm with a different minimal k_T distance. The second minimum, the Q_{cut}, is set in Pythia and is known as the merging scale. The two minima are chosen to be $Q_{cut} > Q^X_{cut}$. Q_{cut} is also a generator cut at Pythia level. Pythia will give then only the events where every pair of hadrons pass this criterion.

4. Comparison of the set of jets obtained from step 2 and 3. If a jet from the parton event falls in a radius smaller than $1.5 \times R$ of a jet from the showered event, it is said to be matched to it (R being the radius parameter of the jet algorithm). If a match is found for all partons the event is kept, otherwise is rejected.

5. Finally, if an event after hadronization has more jets than the generated partons, is discarded.

The goodness of the procedure depends of the choice of Q_{cut} and Q^X_{cut}. The optimal values are highly dependent of several parameters, for example the p_T of the partons or the center of mass energy of the collision.

In general, for each process there are an optimal Q_{cut} and Q^X_{cut}. But in principle there is no way to deduce them without producing the samples and performing the merging procedure. If the procedure was successful, it can be known only from the simulation results. Reason why, multiple generation of samples scanning both Q_{cut} and Q^X_{cut}, are needed to find an optimal solution. As the main interest is to cover correctly the entire phase space with ME+PS, Q^X_{cut} can be fixed and then Q_{cut} scanned over it, with $Q_{cut} > Q^X_{cut}$. If no Q_{cut} value optimize the merging procedure then another Q^X_{cut} is used.

The effectiveness of the merging procedure is checked through the differential jet rate (DJR) diagrams. Changing the merging scale Q_{cut} might change the jet multiplicity of an event during the parton showering and hadronization by Pythia. In the DJR diagrams the transition between n and $n-1$ jet multiplicities is examined as a function of the merging scale for different values of n. The Q_{cut} controls the number of jets found in an event from parton showering. Its optimal value leads to a continuum in the transition phase between ME and PS. The choice of the correct merging scale is reflected in the smoothness of the transition between n and $n-1$ jet multiplicities. A non-optimal value leads to an under or overpopulation of the phase space. In both cases the sample does not represent a meaningful physics sample.

The optimal value of the merging scale, Q_{cut}, depends on several details of the generated events. For example, the minimal p_T cuts set in the generation. But it strongly depends of the process and the center of mass energy of the collisions.

For MC samples generation for the run 2 of the LHC, optimal values for Q_{cut} and Q^X_{cut} should be found according to the increase of the center of mass energy from 8 TeV to 13 TeV. Figure 4.13 shows the DJR diagrams for $t\bar{t}$ production with up to 3 additional jets with MadGraph and Pythia 8. For this study $Q^X_{cut} = 20$ GeV and the optimal Q_{cut} was found to be 100 GeV (top figure). The result using $Q_{cut} = 60$ GeV is also shown (bottom figure) where a discontinuity in the transition point between different jet multiplicities. In the figure at the bottom, the transition from 3 to 2 jets ($DJR(3 \to 2)$) shows a discontinuity in the meeting point of blue-dashed and green-dashed curves, respectively 3 and 2 partons curves. In figure 4.14 the same study is shown for Drell-Yan process with up to 3 additional jets at 13 TeV with MadGraph and Pythia 6.

4.3 Validation on data

MC simulations are theory based, in the sense that generated events reflect predictions of a given model. For example, the simulation of top quark decay depends on the theory predictions of the branching ratios of all its possible final states. Such branching ratios are utilized by generators as probabilities, to evaluate with random numbers if a top quark in a single event decay into a specific channel.

A considerable number of SM processes have been measured by experiments very accurately. Such processes have also served to test theoretical predictions. Such well understood processes can be then used as reference to test the accuracy of MC simulations. All MC generators are tested against known experimental and theoretical results to prove their validity. For example, MadGraph versions used in CMS have been carefully validated internally by the collaboration before being used for central production.

4.3.1 RIVET

There are different tools designed to validate MC generators. The Rivet project (Robust Independent Validation of Experiment and Theory) [146] is a toolkit for MC generators validation, providing a large set of experimental analyses. This tool has been extensively used by experiments and MC generators developers for MC development, tuning and validation.

Rivet not only provides the analyses, it also provides the experimental data resulting from them. These data have been processed to unfold detector effects from them. This procedure tries to inverse detector simulation to obtain event information of particles before interacting with the detector. This constitutes a suitable method to compare data with MC simulations up to hadronization.

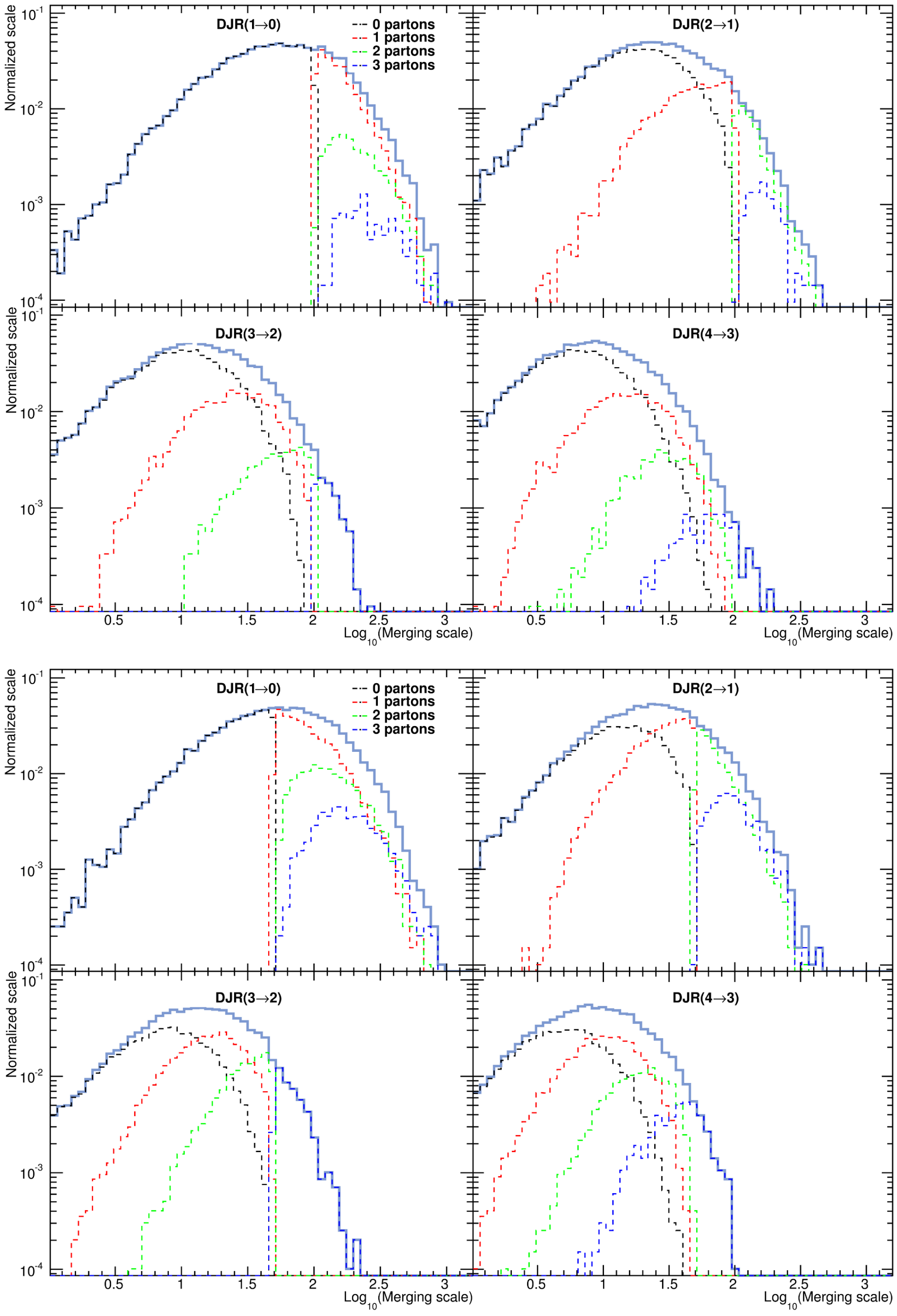


Figure 4.13: DJR diagrams for 13 TeV top pair production with $Q_{cut} = 100$ GeV and $Q^X_{cut} = 20$ GeV [optimal case] (top) and $Q_{cut} = 60$ GeV and $Q^X_{cut} = 20$ [non-optimal] (bottom). Bottom figure shows a discontinuity in the transition from 3 to 2 partons at the point where blue-dashed and green-dashed curves meet.

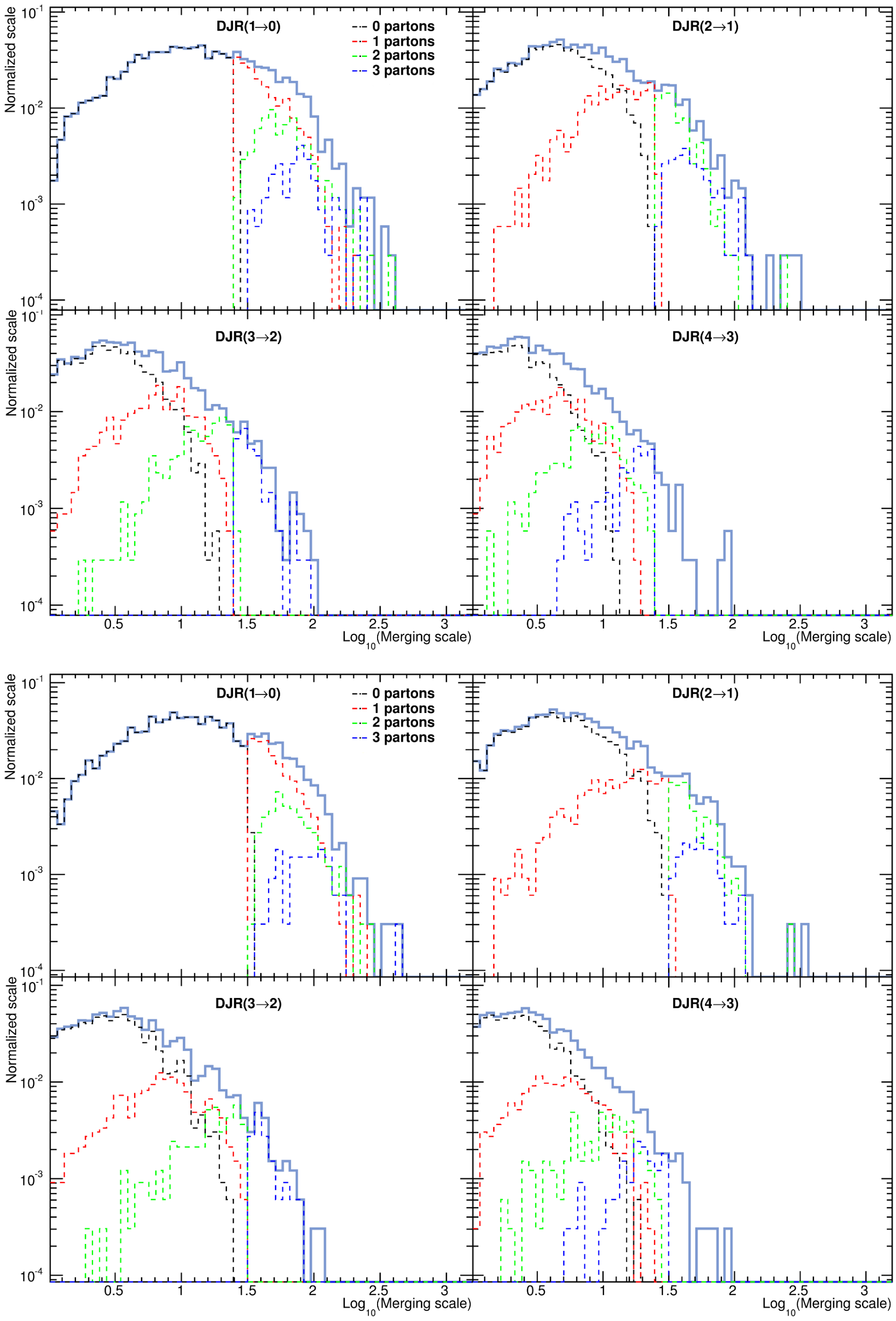


Figure 4.14: DJR diagrams for 13 TeV Drell-Yan production with $Q_{cut} = 125$ GeV and $Q^X_{cut} = 10$ GeV [optimal case] (top) and $Q_{cut} = 32$ GeV and $Q^X_{cut} = 10$ [non-optimal] (bottom). Bottom figure shows a discontinuity in the transition from 1 to 0 partons at the point where black-dashed and red-dashed curves meet.

ATLAS and CMS experiments contribute constantly by providing unfolded results to Rivet toolkit database.

Figures 4.15 and 4.16 display examples of MC validation using Rivet tool. For the first figure, a comparison between several MC simulations have been compared to ATLAS experiment data in the measurement of $W^{+/-}$ boson p_T. In figure 4.16, the same MC generators have been compared to Z^0 boson p_T measurements performed by ATLAS and CMS. In the three validations, MadGraph interfaced with Pythia 6 describe better the real data than the other generators used.

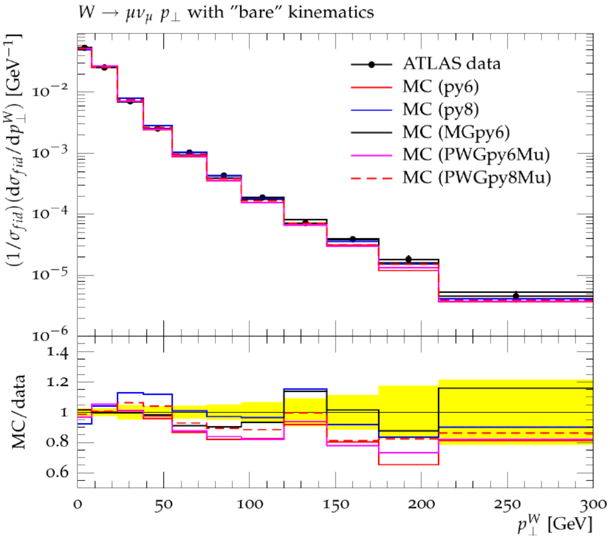


Figure 4.15: $W^{+/-}$ boson p_T measured by ATLAS experiment in muon final states compared to different MC simulations. py6 stands for Pythia 6, py8 for Pythia 8, MGpy6 for MadGraph interfaced with Pythia 6 and PWG for Powheg. MadGraph with Pythia 6 gives the best description of experimental data for the shown observable.

In general, present MC generators describe correctly SM physics. However, there are known observables that are not correctly described by MC simulations. The most important issue has been seen when measuring the p_T of the top quark in top-pair production. Figure 4.17 shows the ratio between data and MC, produced with MadGraph and Pythia 6, for the differential cross section of top pair production [147]. This figure has been produced from the combination of different CMS analyses [148, 149, 150]. A clear trend can be seen, showing that top p_T in MC tend to be higher than in data. Due to this issue, MadGraph with Pythia 6 simulations of top pair production should be corrected. These corrections are considerable for high p_T tops. They have been studied for tops with a p_T smaller

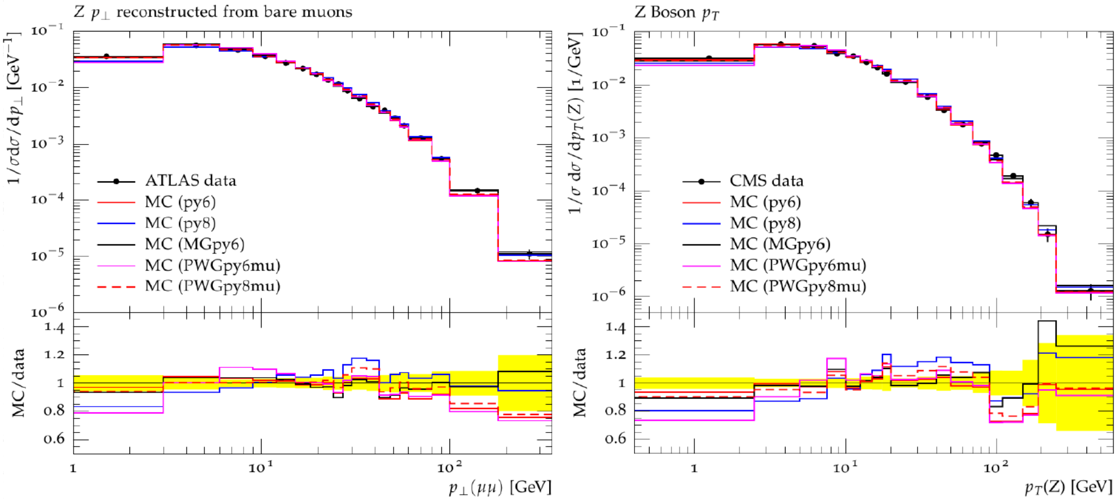


Figure 4.16: Z^0 boson p_T measured by ATLAS [left] and CMS [right] experiments in di-muon channel. py6 stands for Pythia 6, py8 for Pythia 8, MGpy6 for MadGraph interfaced with Pythia 6 and PWG for Powheg. MadGraph with Pythia 6 gives the best description of experimental data in both cases.

than 400 GeV/c.

In conclusion, though MC simulations are useful to understand particle physics processes and to test theoretical predictions, they are not always fully valid. For example, high p_T spectra of particles or high jet multiplicity are not always well reproduced. For this reason it is very important for analyses looking for new physics that have to deal normally with non-explored or poorly explored physics, to develop strategies to estimate backgrounds from data. MC simulations are however useful to compare theoretical predictions with experimental measurements, in particular in this work they were used to model the signal characteristics in a search for a top-partner, described in chapter 5.

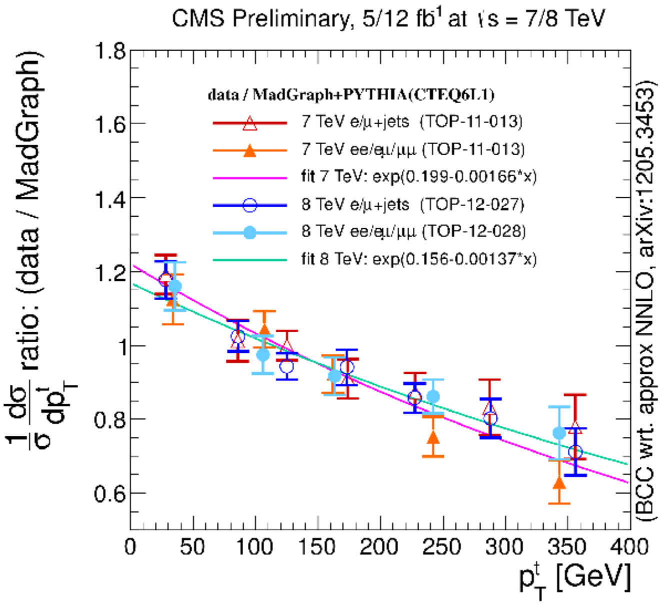


Figure 4.17: Data to MC ratio of normalized differential top pair production cross section as s function of top p_T. The p_T of the top quark measured in MC tends to be higher than in data for a p_T greater than 100 GeV/c [147].

5 Search for a single produced T' decaying into top and Higgs in the full hadronic final state

In the present chapter the search performed using 2012 data collected by CMS for a T' in the full hadronic final state is described. The theoretical formalism for such object has been described on section 1.5. In addition, a feasibility study for this search has been presented in chapter 3. As discussed in this latter, a resonance in the five jets invariant mass is looked for.

In this chapter the following terms will be used:

- **Signal MC sample**: This term refers to the MC samples used to simulate signal.
- **Background samples**: It refers to MC samples used to simulate backgrounds.
- **Signal sample**: The sample or set of data events passing the analysis selection.
- **Control sample**: The sample or set of data event passing the selection designed for data-driven background estimation.

5.1 Analysis Strategy

The strategy of the analysis is based on a high signal efficiency while keeping under control the background. The main background for the hadronic final states is the multijet production. This background should not present any resonance in the 5-jets invariant mass variable but should be a continuum.

Several variables to distinguish between SM backgrounds and the signal have been identified, the most important one being the number of b-tagged jets. In the signal, at least three jets coming from b-quarks are expected, and consequently, at least three b-tagged jets are expected. After requiring such conditions, $t\bar{t}$ become an important background for the analysis.

5.2 Datasets

The analysis uses the MultiJet primary dataset processed with the reconstruction of January 22th 2013. This reconstruction procedure is the most expensive (in terms of CPU resources) processing of data, where all the final objects used in analyses are reconstructed. All detector information is used in the vertexing and tracking algorithms to reconstruct global objects as muons or to reconstruct jets by different algorithms with anti-kt or Cambridge-Aachen algorithms (see section 2.3.3). A primary dataset is a set of events passing trigger selections based on a common object without any further selection. They are named accordingly to their high level trigger path, in our case trigger paths requiring at least 2 jets. The processed datasets are listed in table 5.1.

Dataset name	Int. Luminosity (pb^{-1})
/MultiJet/Run2012A-22Jan2013-v1/AOD	889.4
/MultiJet1Parked/Run2012B-05Nov2012-v2/AOD	4429.0
/MultiJet1Parked/Run2012C-part1-05Nov2012-v2/AOD	494.6
/MultiJet1Parked/Run2012C-part2-05Nov2012-v2/AOD	6654.0
/MultiJet1Parked/Run2012D-part1-10Dec2012-v1/AOD	5955.1
/MultiJet1Parked/Run2012D-part2-17Jan2013-v1/AOD	734.0
/MultiJet1Parked/Run2012D-part2-PixelRecover-17Jan2013-v1	538.4
Total	19694.5

Table 5.1: List of Multijet Primary Dataset used in the analysis and the corresponding integrated luminosity calculated using the golden JSON (Java Script Object Notation) file. The golden JSON file contains the information about the luminosity sections considered as good for all runs. A good luminosity section is defined as a luminosity section where the detector was fully functioning, that is all subsystems were taking data without any problems.

In addition, the MC samples processed to study the different backgrounds entering the analysis are presented in table 5.2. QCD, in p_T hat (for initial partons p_T ranges), and diboson samples have been generated using Pythia 6 [119], while the ones in H_T bins were generated with MadGraph 5 [116, 117] interfaced with Pythia 6. The same generators were used to generate $t\bar{t}$ MC samples with the additional usage of MadSpin [138, 139]. The single top samples were generated using Powheg [134, 135, 136] generator.

Finally, for the simulation of signal, several samples for different T' masses have been processed. Nine different mass points between 600 GeV/c^2 and 1 TeV/c^2 in steps of 50 GeV/c^2 have been utilized. The processed signal samples are shown in table 5.3. These samples were generated with MadGraph 5 and Pythia 6.

Samples	Cross-Section (pb)	Number of events
QCD_Pt-120to170_TuneZ2star_8TeV_pythia6	16×10^4	5.9M
QCD_Pt-170to300_TuneZ2star_8TeV_pythia6	34×10^3	5.8M
QCD_Pt-300to470_TuneZ2star_8TeV_pythia6	18×10^2	5.9M
QCD_Pt-470to600_TuneZ2star_8TeV_pythia6	114	3.9M
QCD_Pt-600to800_TuneZ2star_8TeV_pythia6	27	3.9M
QCD_Pt-800to1000_TuneZ2star_8TeV_pythia6	3.5	3.9M
QCD_HT-500To1000_TuneZ2star_8TeV-madgraph-pythia6	84×10^2	30M
QCD_HT-1000ToInf_TuneZ2star_8TeV-madgraph-pythia6	2×10^2	14M
DYToCC_M_50_TuneZ2star_8TeV_pythia6	31×10^2	2M
DYToBB_M_50_TuneZ2star_8TeV_pythia6	38×10^2	2M
TTJets_MSDecays_central_TuneZ2star_8TeV-madgraph-tauola	247.7 [NNLO]	62M
T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11.1 [NNLO]	497k
T_s-channel_TuneZ2star_8TeV-powheg-tauola	3.79 [NNLO]	260k
T_t-channel_TuneZ2star_8TeV-powheg-tauola	54.9 [NNLO]	3.7M
Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11.1 [NNLO]	493k
Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola	1.76 [NNLO]	140k
Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola	29.7 [NNLO]	1.9M
WZ_TuneZ2star_8TeV_pythia6_tauola	33.6 [NLO]	10M
ZZ_TuneZ2star_8TeV_pythia6_tauola	7.6 [NLO]	9.8M
WW_TuneZ2star_8TeV_pythia6_tauola	56 [NLO]	10M
TTH_Inclusive_M-125_8TeV_pythia6	0.13 [NLO]	100K

Table 5.2: List of the Monte-Carlo background samples used in the analysis, their corresponding cross-section and their number of events.

The signal samples have been produced with only two decay channels for the Higgs boson: $b\bar{b}$ and $\tau^-\tau^+$. Thus, in order to obtain the correct branching ratio of the Higgs to $b\bar{b}$, a rescaling factor must be applied. The branching ratio of this channel for SM Higgs of 125 GeV/c^2 is 0.57 [84]. However, in signal samples the effective branching ratio is 0.94. The mass point signal samples have 94% of times the Higgs boson decaying into $b\bar{b}$ and 6% to $\tau^-\tau^+$. In consequence, a weight of 0.61 has been applied to these samples to obtain the correct branching ratio.

5.3 Event selection

In the following sections, the event processing and selection for the analysis are presented. A pre-processing of events is performed, which includes several stages to filter them. For example, only the events in luminosity sections where the detector was working correctly, all subdetectors were on and running without problematic behavior, were selected and processed. From these pre-processed events the analysis selection is performed.

Sample	T' Mass (GeV/c^2)	Cross-Section (fb)	Number of events
TprimeJetToTH_M-600_TuneZ2star_8TeV-madgraph_tauola	600	215.4	95K
TprimeJetToTH_M-650_TuneZ2star_8TeV-madgraph_tauola	650	177.8	99K
TprimeJetToTH_M-700_TuneZ2star_8TeV-madgraph_tauola	700	143.7	99K
TprimeJetToTH_M-750_TuneZ2star_8TeV-madgraph_tauola	750	118.6	99K
TprimeJetToTH_M-800_TuneZ2star_8TeV-madgraph_tauola	800	100	96K
TprimeJetToTH_M-850_TuneZ2star_8TeV-madgraph_tauola	850	84.3	99K
TprimeJetToTH_M-900_TuneZ2star_8TeV-madgraph_tauola	900	72.6	99K
TprimeJetToTH_M-950_TuneZ2star_8TeV-madgraph_tauola	950	62.6	96K
TprimeJetToTH_M-1000_TuneZ2star_8TeV-madgraph_tauola	1000	53.9	99K

Table 5.3: List of Monte-Carlo signal samples used in the analysis, their corresponding cross-section and T' mass.

5.3.1 Event processing

As first step, data and MC events were processed with PAT (Physics Analysis Toolkit) [151], to produce PAT-tuples. PAT simplifies access to reconstructed objects and tools developed for analyses. It constitutes a common work-ground where analysts can find a simplified access to event content and to the full set of CMS tools. This step makes a first selection of objects with very basic cuts as a minimal jet p_T. Only jets within $|\eta| < 5$ (to cover the majority of the HCAL acceptance and include forward jets from signal) and with at least a p_T of 20 GeV/c are considered. In addition, several noise filters are applied. For example, only events with at least one good primary vertex (n.d.o.f. ≥ 4, $|z| <$ 24 cm, $|\rho| <$ 2 cm) are kept: where the reconstruction of primary vertices, the number of degrees of freedom (n.d.o.f) is related to the number of tracks associated to the vertex, z is the vertex position in the beam direction, and ρ is its "mass", a weight assigned by the reconstruction algorithm associated to its compatibility to form a track cluster. In addition, jets by using particles reconstructed via the Particle Flow algorithm [99, 100, 101, 102] and Charge Hadron Subtracted (CHS) technique [152] are corrected with the jet energy corrections (JEC, described in section 2.3.3).

An important setup for the processing of samples in CMS is the selection of the global tag. For data, it has calibration and alignment information of the detector; and for MC, it is used to mimic real detector conditions, in order to get closer MC simulations to data.

In MC samples, the pileup has to be corrected to observed pileup in data. For these samples a model to simulate expected pileup in data, called scenario, is used. The pileup scenario used for MC simulation does not describe correctly the data, therefore the MC samples have been weighted in correspondence. Figure 5.1 shows the pileup variable for data and MC (for the scenario 10 - S10), and the weight from the ratio for each bin between data and MC. The obtained weights were applied to MC events as a function of their true number of interactions. For data, the true number of interactions represents the expected number of interactions per crossing for a given lumi section from the average bunch instantaneous luminosity with respect to the total inelastic cross section.

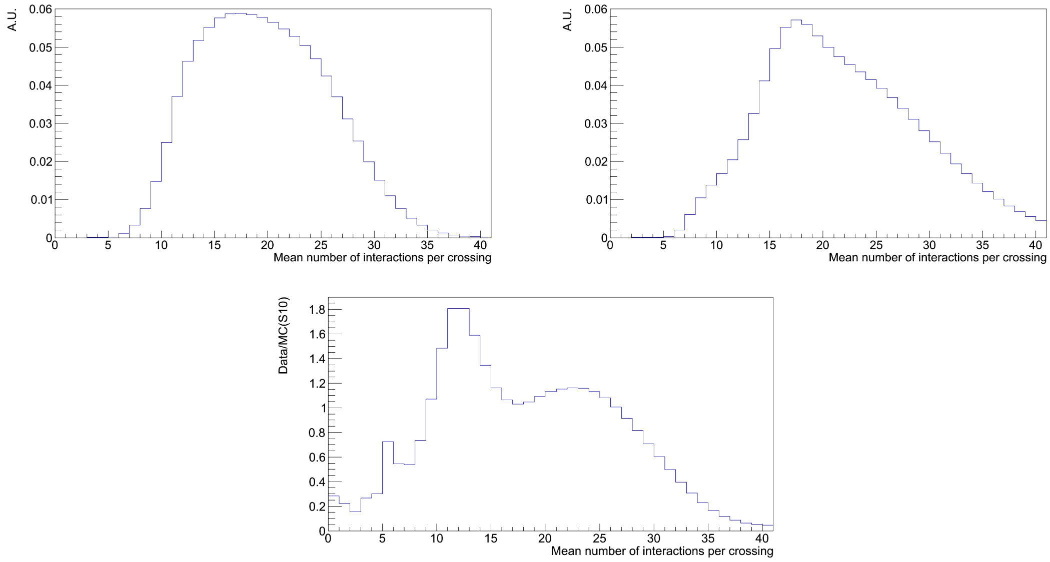


Figure 5.1: Pileup for data [up-left], MC S10 [up-right] and ratio between them [bottom].

To check the correctness of the procedure, the number of vertices in data and MC were compared. This comparison can be found in figure 5.2, and it is performed before cutting on the number of b-tagged jets. After this reweighting procedure, the sum of the MC samples describes correctly the number of vertices in data.

5.3.2 Basic selection

The selection starts with the trigger requirements. The level 1 trigger selects multijet events with at least 4 central jets ($|\eta| < 3$) with a $p_T > 32$ GeV/c or $p_T > 36$ GeV/c or $p_T > 40$ GeV/c, or events with at least two central jets with $p_T >$ 52 GeV/c or $p_T > 56$ GeV/c or $p_T > 64$ GeV/c, or events with $H_T > 125$ GeV/c or $H_T > 150$ GeV/c or $H_T > 175$ GeV/c with at least 4 central jets. To obtain a fully unprescaled trigger selection, on top of the loosest trigger requirement that is prescaled, tighter unprescaled trigger requirements were added. The HLT path chosen require at least 6 jets with a $p_T > 20$ GeV/c, 4 with a $p_T > 60$ GeV/c and

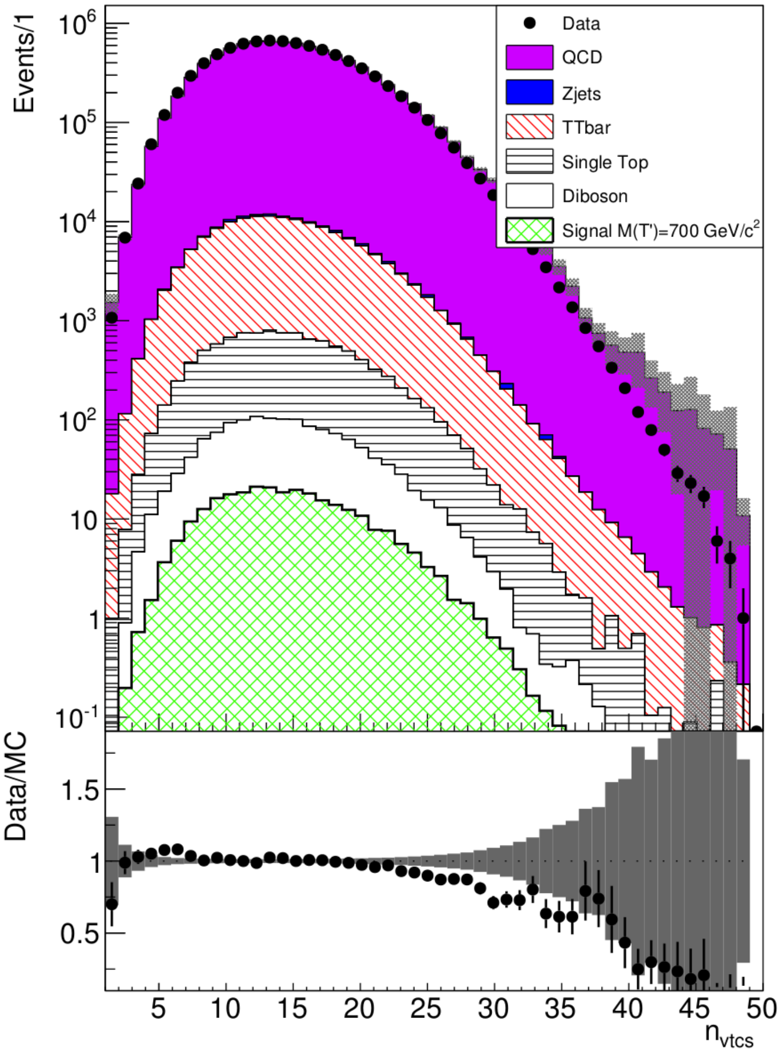


Figure 5.2: Number of vertices distribution for data and MC samples. The comparison has been performed after basic selection except number of b-tagged jets (the basic selection is described in the next section, 5.3.2). The gray band correspond to the statistical error of MC samples sum. Normalization of MC samples was done to the 19.7 fb^{-1}.

two with a $p_T > 80$ GeV/c. These jets are required to have $|\eta| < 3$. This HLT path is named *DiJet_80_DiJet_60_DiJet_20*. The HLT works with CaloJets, jets reconstructed with calorimeter information, slightly different from the PF jets from the final event reconstruction. Right after the requirement from the trigger, a first cut is applied, to close up the selection to the parameter space of the trigger selection. Only events with two jets with $p_T > 90$ GeV/c, two jets with $p_T > 70$ GeV/c and two jets with $p_T > 30$ GeV/c were kept. This selection is applied over PF jets.

As a second selection step, at least 5 jets with $p_T > 30$ GeV/c and $|\eta| < 2.5$ and at least one additional jet with $p_T > 30$ GeV/c and $|\eta| < 5$ were required for all events. This cut is driven by signal properties. The T' is produced in association with a light jet, that is produced in the forward η region. In addition, the T' decay products (5 jets in the full hadronic final state) are essentially produced in the tracker acceptance ($|\eta| < 2.5$). Figure 5.3 shows the η distribution of the accompanying jet from signal MC sample with M=700 GeV/c^2 using MC truth information. At this stage of the selection, all events have at least 6 jets with $p_T > 30$ GeV/c.

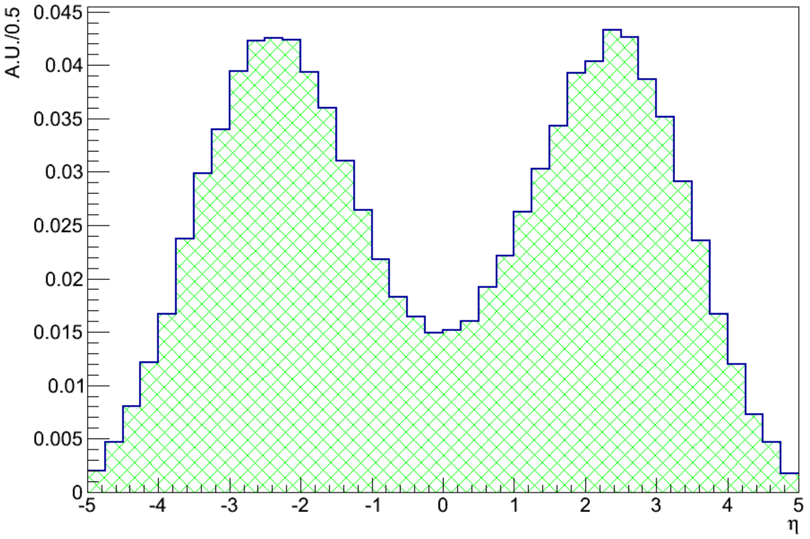


Figure 5.3: Distribution of pseudorapidity of the accompanying jet produced with the T'. The distribution is taken from the signal MC sample with M=700 GeV/c^2 and it is normalized to unity.

Moreover the leading jet was required to have a $p_T > 150$ GeV/c. With this cut $t\bar{t}$ is reduced by 32% and QCD by 24% (estimation performed from MC samples shown in table 5.2), while signal (M=700 GeV/c^2) is only reduced by 9%. As the leading jet in signal events is coming from a heavy T', it is expected to have higher p_T than in SM background events. In figure 5.4 and 5.5 the p_T and η of the 6 leading jets for data and MC are shown. The comparison is performed before selecting the number of b-tagged jets. Some disagreements between MC samples and the data are expected at very high jet p_T, because QCD MC samples were processed for a p_T up to 1000 GeV/c (see table 5.2). The MC samples are used in this analysis only for illustration purposes, as final background estimation has been derived directly from data. Additionally, in the η distributions of jets a disagreement between data and MC is visible for $|\eta| > 3$. This disagreement is caused by the simulation of the trigger on MC, that do not correspond exactly with the behavior of real trigger in data. This mismatch of the trigger in MC is a motivations to estimate backgrounds from data instead of using MC.

After the cut on the leading jet p_T, $H_T > 550$ GeV/c was required. Multijet background events have smaller H_T than signal events. The H_T distribution for data and MC samples is shown in figure 5.6, this comparison is performed before b-tagged jets multiplicity criterion. The QCD MC samples were combined to increase the statistics on this process. The QCD H_T binned and p_T hat samples were normalized to half the luminosity observed in the data. The combination of these samples should then correctly represent the expected multijet background for an $H_T > 550$ GeV/c. However, as p_T hat samples were processed only up to 1000 GeV/c some disagreement is expected at high H_T.

The final cut of the basic selection concerns the request of a minimum number of b-tagged jets. The CSV algorithm (described in section 2.3.3) was used to identify jets coming from a b quark. In the following, b-tagged jets are defined as jets that were b-tagged by the CSV algorithm and b-jets as jets matched at generator level as coming from a b-quark (from matrix element). The medium working point was chosen as it allows to have a high efficiency on b-jet identification (70%) with a low rate from c-quark (20%) and light quarks (1%) [110, 111]. In the full hadronic final state, the T' decays in three b-quarks, a $b\bar{b}$ system coming from the Higgs boson and an additional b-quark from the t-quark, and two light jets from the $W^{+/-}$ boson produced by the top. Accordingly, signal events are expected to have at least 3 b-tagged jets, while $t\bar{t}$ and QCD events should have mainly 2. QCD events can have also 4 b-tagged jets but in smaller proportion. In addition, the b-jets from backgrounds have different properties than the b-jets from signal, as ΔR or p_T.

In principle several working points can be used to establish the b-tagged jets content of the events. For the CSV algorithm three working points have been defined as a function of the cut made on the discriminator value, 0.244 for the loose working point (CSVL), 0.679 for the medium (CSVM) and 0.898 for the tight

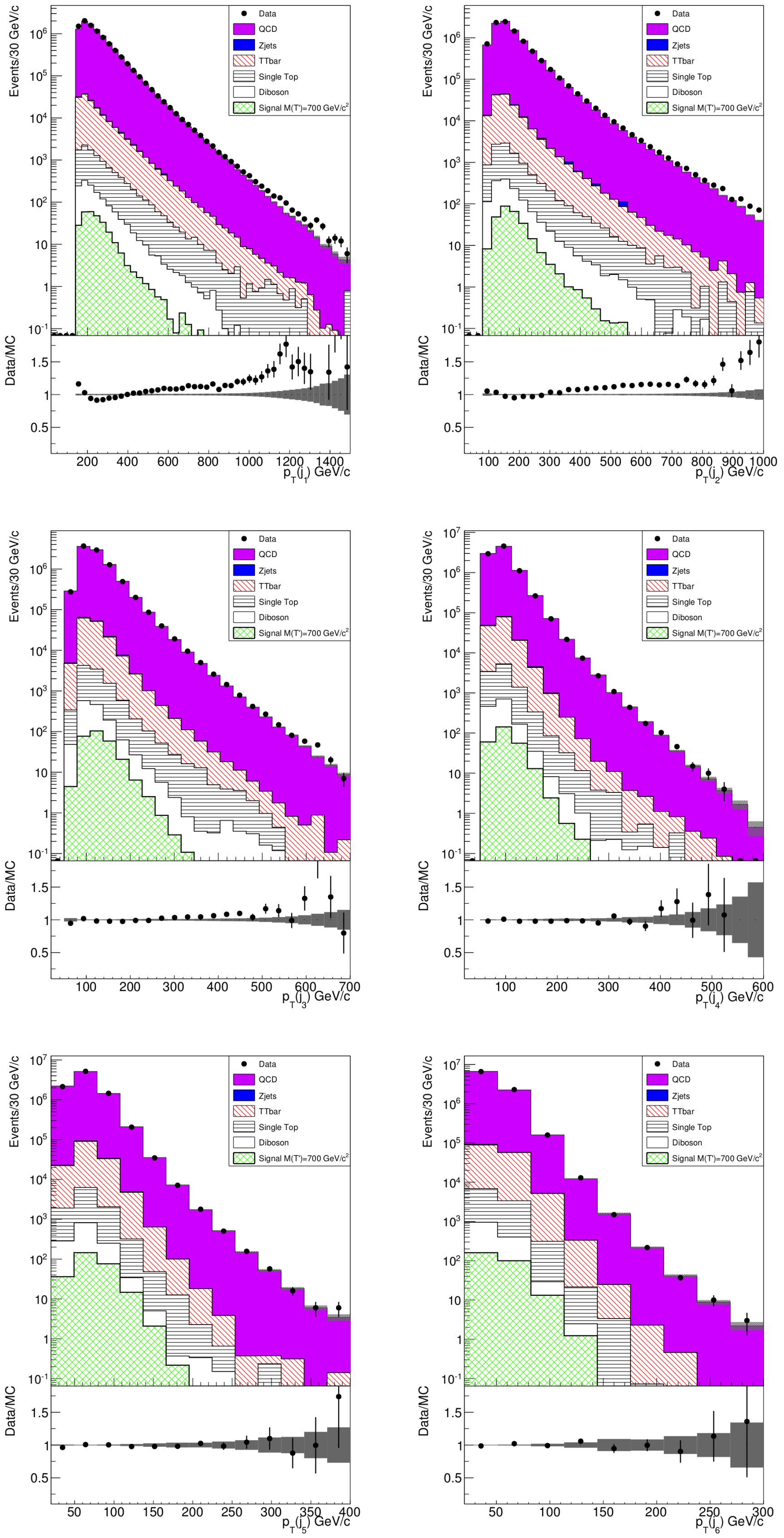


Figure 5.4: Distribution of the transverse momentum of the 6 leading jets. The gray band represents the statistical uncertainties from the sum of the MC background. Reasonable agreement is observed, with the multijet process as the dominant process at this stage. Normalization of samples was done to 19.7 fb^{-1}.

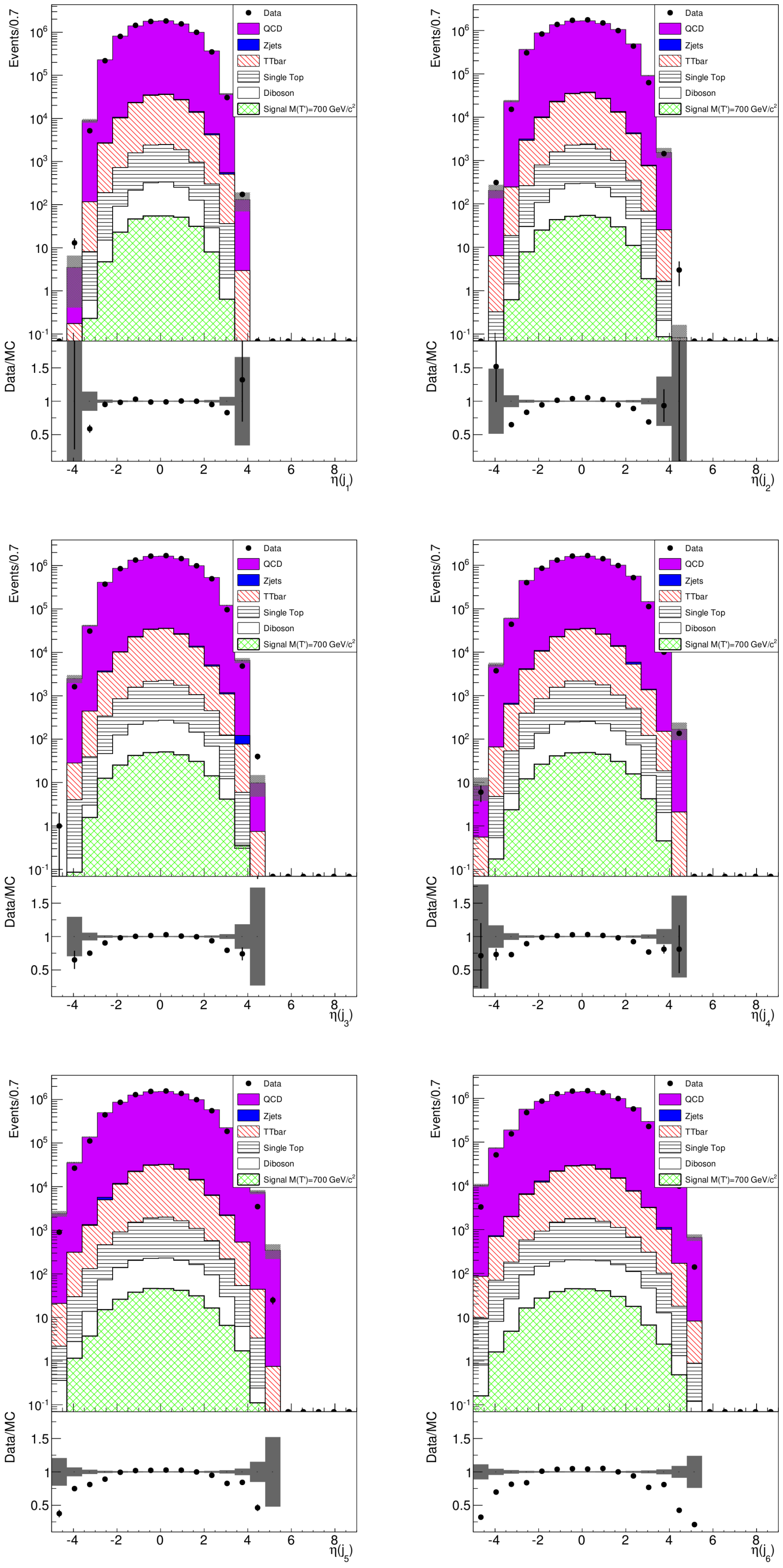


Figure 5.5: Distribution of η of the 6 leading jets. The gray band represents the statistical uncertainties from the sum of the MC background. Reasonable agreement is observed, with the multijet process as the dominant process at this stage. Normalization of samples was done to luminosity.

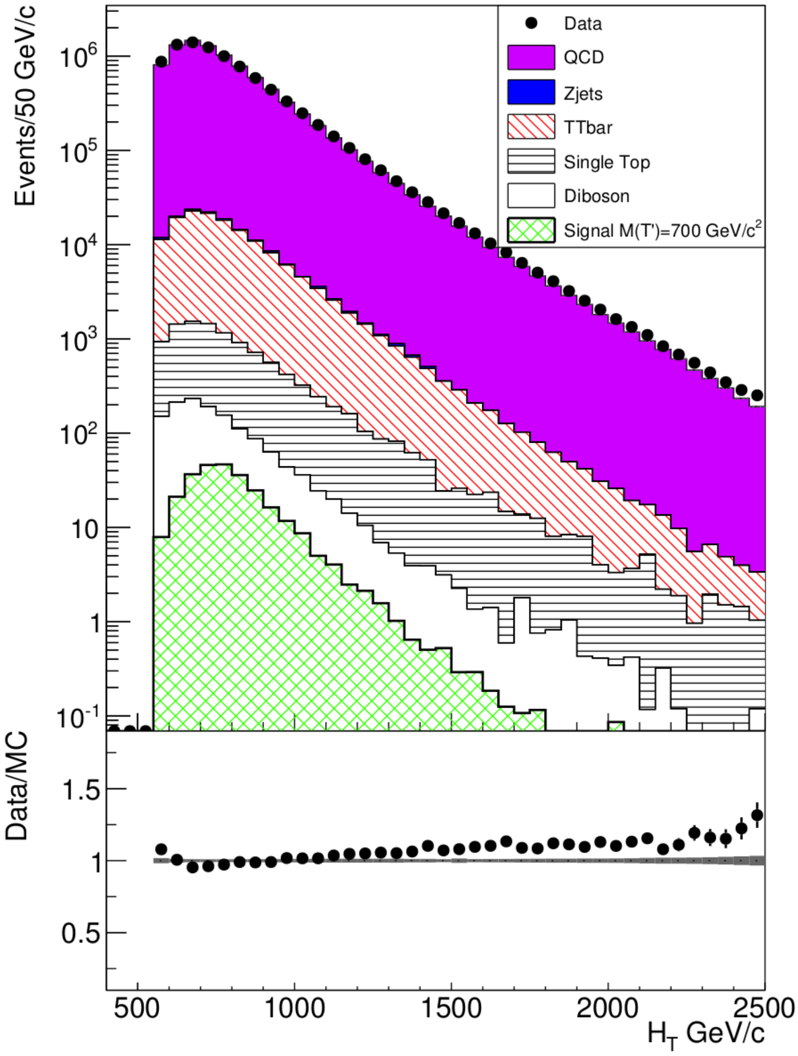


Figure 5.6: Distribution of the H_T variable for data and the sum of the MC samples normalized to luminosity. The signal sample (M=700 GeV/c^2) is over-imposed on top of the stack of the MC samples. The gray band represents the statistical uncertainties from the sum of the MC background. Reasonable agreement is observed, with the multijet process as the dominant process at this stage. Normalization of samples was done to luminosity.

working point (CSVT). For example, events can be required to have at least 3 b-tagged jets one with tight working point and two with medium working point. All possible combinations have been studied from the three available working points (loose (CSVL), medium (CSVM) and tight (CSVT)) to require at least 3 CSV b-tagged jets in order to establish which combination gave the best S/B while keeping the most of signal events. For the study, the $M = 700$ GeV/c^2 signal sample has been used as signal and as backgrounds, the $t\bar{t}$, QCD_HT-500To1000 and QCD_HT-1000ToInf MC samples have been utilized. The study was performed applying the basic selection, after H_T cut. The results of the study are contained in table 5.4. This table shows that the most efficient cuts to discriminate signal from backgrounds and to keep signal are to require at least 3 CSVM b-tagged jets or to require at least 2 CSVM and 1 CSVT b-tagged jets. Selecting at least 3 CSVM b-tagged jets has the same S/B and $S/\sqrt{S+B}$ than requiring at least 2 CSVM and 1 CSVT b-tagged jets. However, the 3 CSVM criterion has a higher efficiency on signal. Thus, the 3 CSVM requirement is preferred over the other combinations.

At least	$\epsilon(S)$ [%]	$\epsilon(t\bar{t})$ [%]	ϵ(QCD_HT-500To1000) [%]	ϵ(QCD_HT-1000ToInf) [%]	$\frac{S}{B} \times 10^3$	$\frac{S}{\sqrt{S+B}} \times 10^2$
3 CSVL	65 ± 0.4	38 ± 0.04	6 ± 0.02	7 ± 0.02	0.4 ± 0.005	24.8 ± 0.3
3 CSVM	31 ± 0.4	8 ± 0.02	1 ± 0.01	0.6 ± 0.01	1.8 ± 0.05	38.2 ± 0.8
1 CSVL and 2 CSVM	55 ± 0.4	27 ± 0.03	2 ± 0.01	2 ± 0.01	0.8 ± 0.01	33.7 ± 0.5
2 CSVL and 1 CSVM	64 ± 0.4	37 ± 0.04	5 ± 0.02	5 ± 0.02	0.5 ± 0.007	28.3 ± 0.3
2 CSVM and 1 CSVT	29 ± 0.4	8 ± 0.02	0.5 ± 0.006	0.5 ± 0.006	1.9 ± 0.05	38.4 ± 0.8
1 CSVM and 2 CSVT	22 ± 0.4	5 ± 0.02	0.3 ± 0.005	0.3 ± 0.003	2.2 ± 0.07	35.8 ± 0.9
3 CSVT	9 ± 0.2	1 ± 0.01	0.1 ± 0.003	0.09 ± 0.002	3.1 ± 0.2	27.3 ± 1.1
1 CSVL and 2 CSVT	33 ± 0.4	13 ± 0.03	0.9 ± 0.01	0.8 ± 0.007	1.1 ± 0.03	31.5 ± 0.6
2 CSVL and 1 CSVT	57 ± 0.4	30 ± 0.03	3 ± 0.02	3 ± 0.01	0.7 ± 0.01	31.5 ± 0.4
1 CSVL and 1 CSVM and 1 CSVT	51 ± 0.4	24 ± 0.03	2 ± 0.01	2 ± 0.01	0.9 ± 0.02	30.8 ± 0.4

Table 5.4: Comparative study of different possible combinations to require at least 3 b-tagged jets with CSVL, CSVM and CSVT working points. Efficiencies of cuts over signal and principal MC background samples are presented, as well as $\frac{S}{B}$ and $\frac{S}{S+B}$. High values of $\frac{S}{S+B}$ point to a good discrimination while keeping the signal efficiency high.

Additionally, the 2D plots linking the number of b-tagged jets in the three working points for signal after requiring $j_1 > 150$ GeV/c and $H_T > 550$ GeV/c can be found in figure 5.7.

As described in section 2.3.3, the b-tagging is performed using a procedure where several jet variables are taken into account. It strongly depends on the ability to find displaced vertices, reason why it has been restricted to jets in $|\eta| <= 2.4$. Even if the MC simulations have been tuned to be as close as possible to the real detector, differences with data remain. Therefore the b-tagging performance is different in data and in MC. Thus, a correction must be applied to MC to mimic b-tagging response on data. In general, the CSV algorithm is slightly more efficient

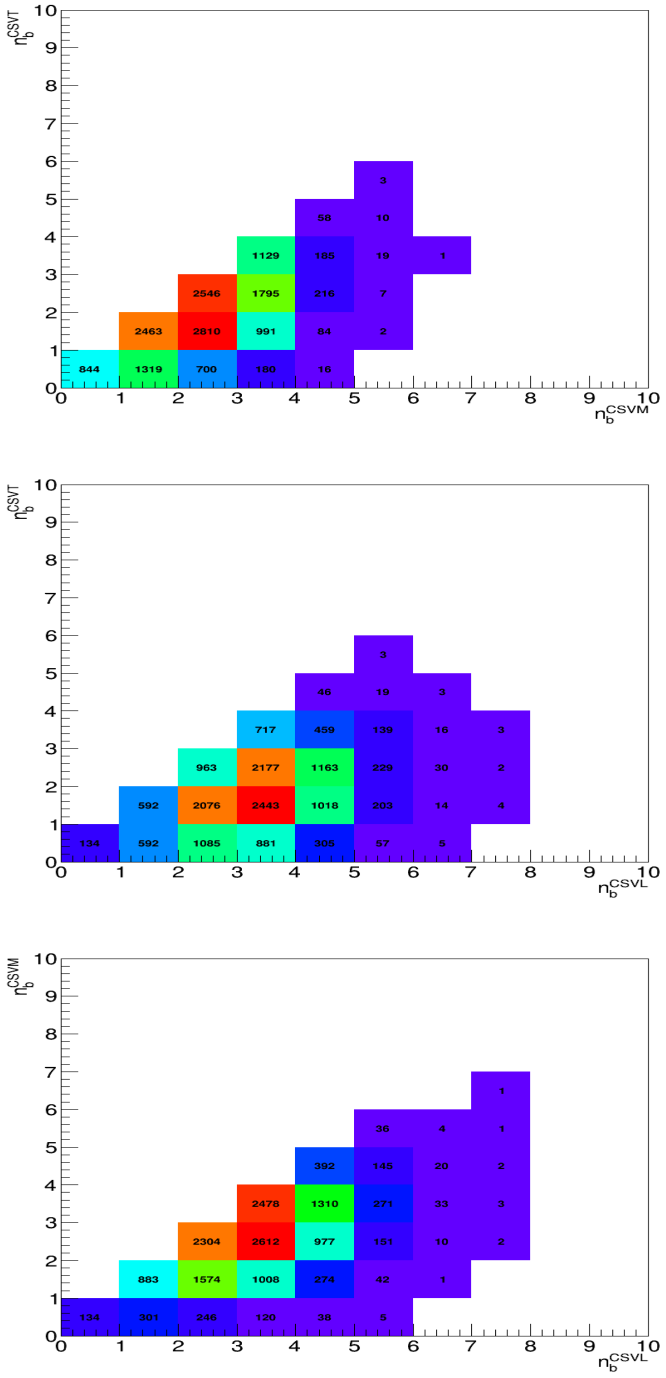


Figure 5.7: 2D plots showing the number of b-tagged jets for between different b-tagging working points for signal MC sample (M=700 GeV/c^2) with selection up to the cut $H_T > 550$ GeV/c. CSVM with CSVT [top], CSVT with CSVL [medium] and CSVM with CSVL [bottom]. Numbers in each bin correspond to the number of entries in the bin. The plots show also the link between the number of b-tagged jets in the different working points.

for b-tagging b-jets in MC than in data, and more light jets are tagged as b-jets in data than in MC. To match MC to data a scale factor has been derived by the collaboration. It is defined for each jet depending on its flavor (b and c or light), p_T and η. In equation 5.1 [111], the parametrization of the scale factors for CSVM working point can be found. The functions are defined as $SF_{\eta}^{flavor}(p_T)$.

$$\begin{aligned}
SF_{|\eta|\leq 2.4}^{b\ or\ c}(x) &= 0.938887 + 0.00017124x - 2.76366\times 10^{-7}x^2 \\
SF_{|\eta|\leq 0.8}^{light}(x) &= 1.07541 + 0.00231827x - 4.74249\times 10^{-6}x^2 \\
&\quad +2.70862\times 10^{-9}x^3 \\
SF_{0.8<|\eta|\leq 1.6}^{light}(x) &= 1.05613 + 0.00114031x - 2.56066\times 10^{-6}x^2 \\
&\quad +1.67792\times 10^{-9}x^3 \\
SF_{1.6<|\eta|\leq 2.4}^{light}(x) &= 1.05625 + 0.000487231x - 2.22792\times 10^{-6}x^2 \\
&\quad +1.70262\times 10^{-9}x^3 \qquad (5.1)
\end{aligned}$$

In order to apply the b-tagging scale factors to MC samples, a method approved by the collaboration was used. The chosen method allows to calculate a weight per event in terms of its jet flavor content. The weight definition can be found in equation 5.2,

$$w = \frac{P(\mathrm{DATA})}{P(\mathrm{MC})} \qquad (5.2)$$

where

$$P(\mathrm{MC}) = \prod_{i=\mathrm{tagged}} \varepsilon_i \prod_{j=\mathrm{not\ tagged}} (1-\varepsilon_j) \qquad (5.3)$$

$$P(\mathrm{DATA}) = \prod_{i=\mathrm{tagged}} \mathrm{SF}_i\varepsilon_i \prod_{j=\mathrm{not\ tagged}} (1-\mathrm{SF}_j\varepsilon_j) \qquad (5.4)$$

the products are defined over all jets in the event. ε represents the b-tagging efficiency. Efficiencies were calculated for each MC sample as a function of flavor, p_T and η. In figure 5.8 the CSVM b-tagging efficiencies for b, c and light jet flavors for $t\bar{t}$ MC sample are displayed. Also, the calculated weights for each MC sample are shown in figure 5.9.

Efficiencies were calculated using the following formula:

$$\varepsilon_f(i,j) = \frac{N_f^{\text{b-tagged}}(i,j)}{N_f^{\text{Total}}(i,j)} \qquad (5.5)$$

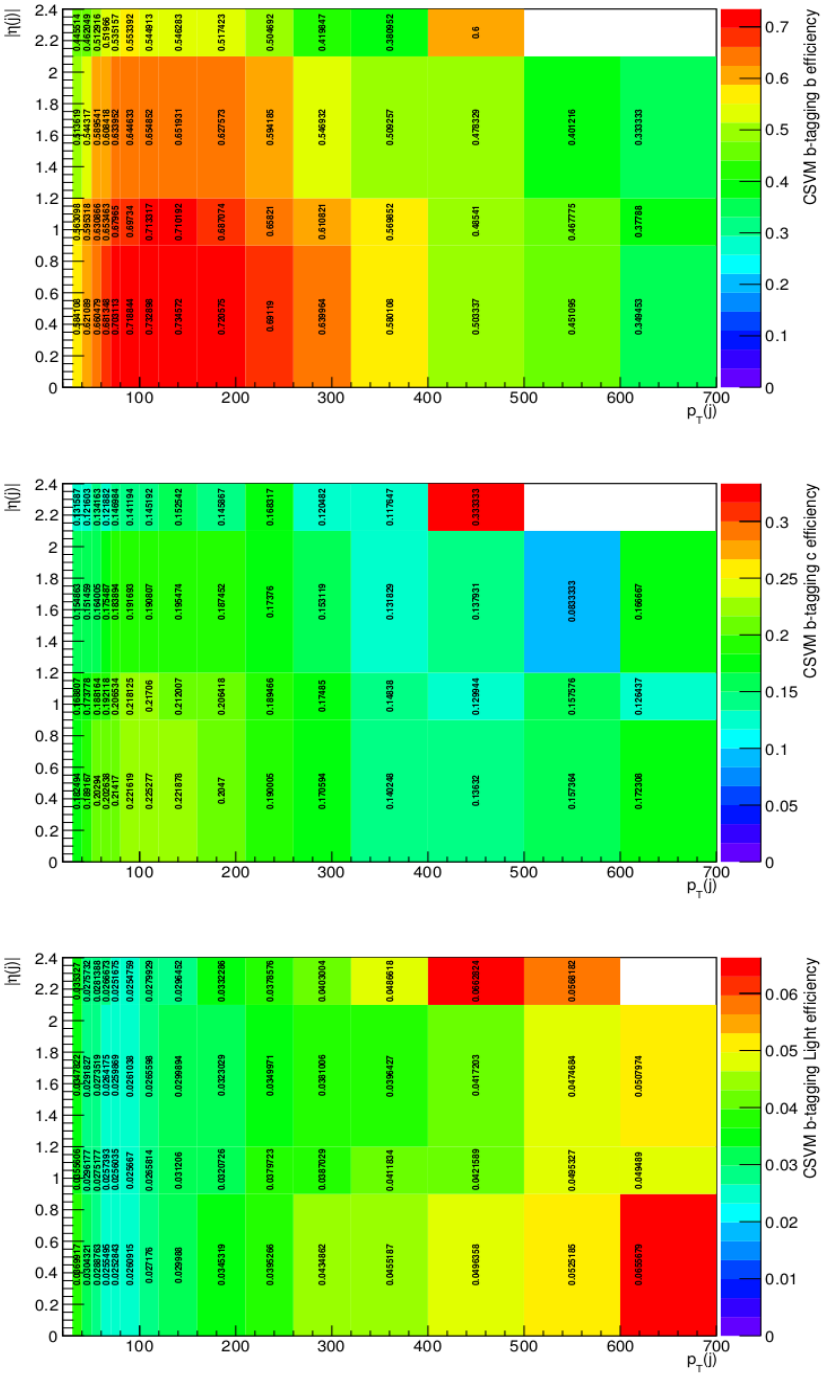


Figure 5.8: CSVM b-tagging efficiency for b-jets [left], c-jets [center] and light jets [right] as a function of p_T and η for $t\bar{t}$.

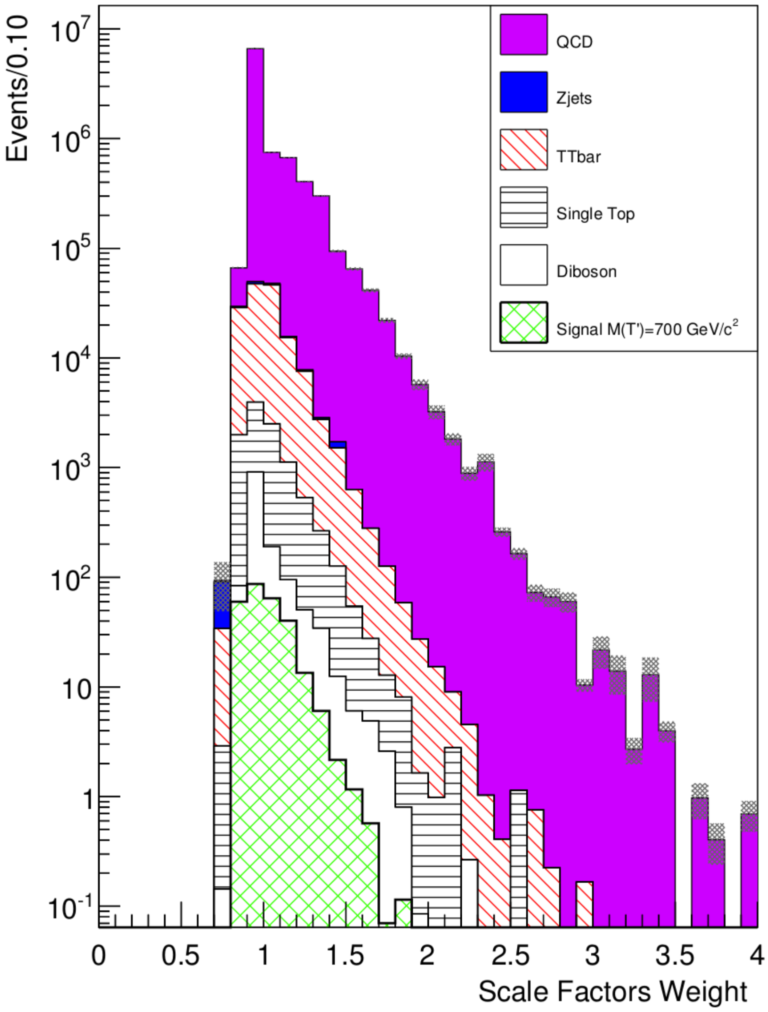


Figure 5.9: Distribution of the weights from b-tagging scale factors for all MC samples.

where $N_f^{\text{Total}}(i,j)$ and $N_f^{\text{b-tagged}}(i,j)$ are the total number of jets and the number of b-tagged jets, respectively, of flavor f in the (p_{T},η) bin (i,j). The determination of the jet flavor is done via the matching of jets to quarks from parton level simulation. The matching is not always successful and some jets can have no parton matched, i.e. no associated flavor. These jets, as well as jets coming from gluons, were included in the light flavor category. The b-tagging efficiencies are not calculated in an event per event basis but from the whole MC sample being considered.

Finally, the number of CSVM b-tagged jets in data was compared to the MC predictions in figure 5.10. The different data/MC comparisons presented are shown for illustration and to check that pileup and b-tagging corrections in MC are correctly applied. For the final analysis results, only signal MC samples are used, as backgrounds were estimated directly from the data.

5.3.3 T' reconstruction with a χ^2 sorting algorithm

A crucial step to reconstruct the T' mass is to correctly identify the five jets coming from its decay. Events passing the basic selection have at least 6 jets, but jet multiplicity might be higher. The jet multiplicity distribution is shown in figure 5.11 after requiring at least 3 CSVM b-tagged jets for all MC samples. Additionally, there are several possible combinations between jets to reconstruct the Higgs boson and top quark from the T' decay. In consequence, the process of identification of jets coming from the T' is not a trivial task.

A χ^2 sorting algorithm has been used to identify the T' decay products and to reconstruct the Higgs and top candidates. This technique relies on the definition of a χ^2 variable for each jets combination in an event. The combination that minimizes this variable gives the best fit of the objects under reconstruction. An example of this method can be found in [153] or [95].

The χ^2 variable is defined in equation 5.6. The values used as inputs were: $M_H = 125$ GeV/c^2, $M_W = 84.06$ GeV/c^2, $M_t = 175.16$ GeV/c^2, $\sigma_H = 12.4$ GeV/c^2, and $\sigma_W = 10.12$ GeV/c^2 and $\sigma_t = 17.35$ GeV/c^2. These values were taken from [95, 154], where similar studies of a χ^2 reconstruction of these particles in full hadronic final state were performed. The Higgs boson width and mass were extracted from an analysis looking for the associated production of a Z^0 and Higgs boson, with the Z^0 decaying into leptons and the Higgs boson in $b\bar{b}$. The top quark and $W^{+/-}$ boson width and mass were taken from an analysis looking for $t\bar{t}$ resonances with one top quark decaying in the leptonic channel and the second top quark going into the hadronic channel. For the Higgs reconstruction only CSVM b-tagged jets are considered. For the $W^{+/-}$ reconstruction all jets with a $p_T > 30$ GeV/c were considered, while for the top reconstruction one b-tagged jet and the pair of jets used for the $W^{+/-}$ were utilized.

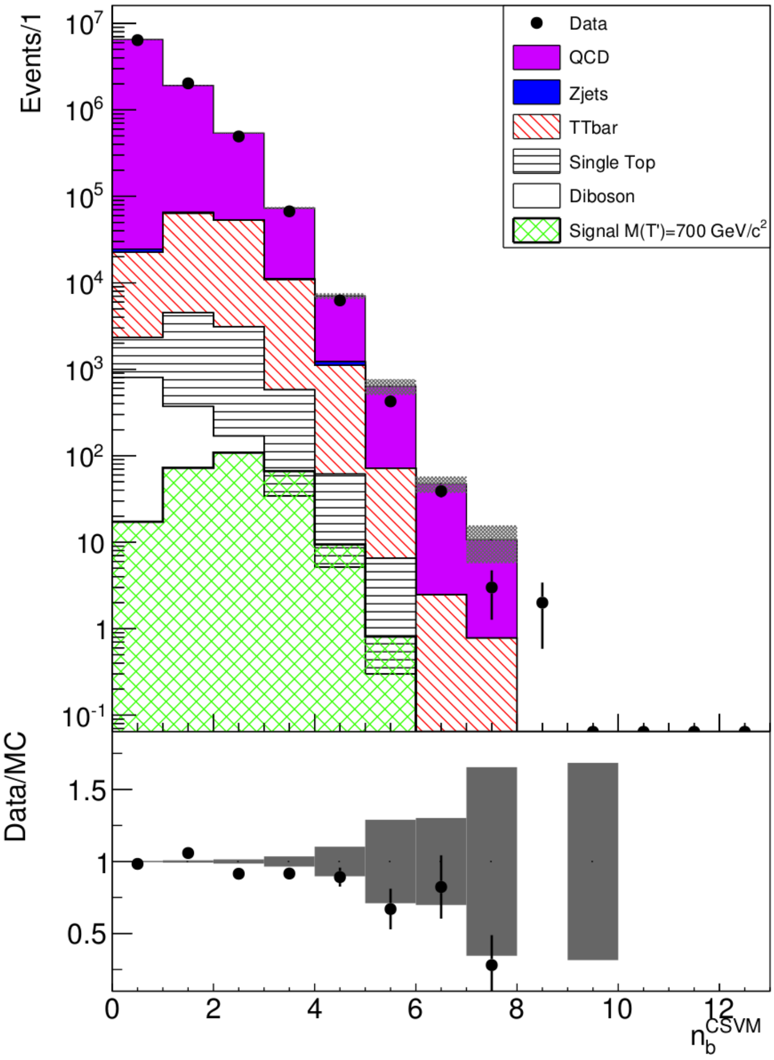


Figure 5.10: B-tagged CSVM jet multiplicity for data and MC samples before requiring at least 3 CSVM b-tagged jets. The sum of MC samples is normalized to the integrated luminosity.

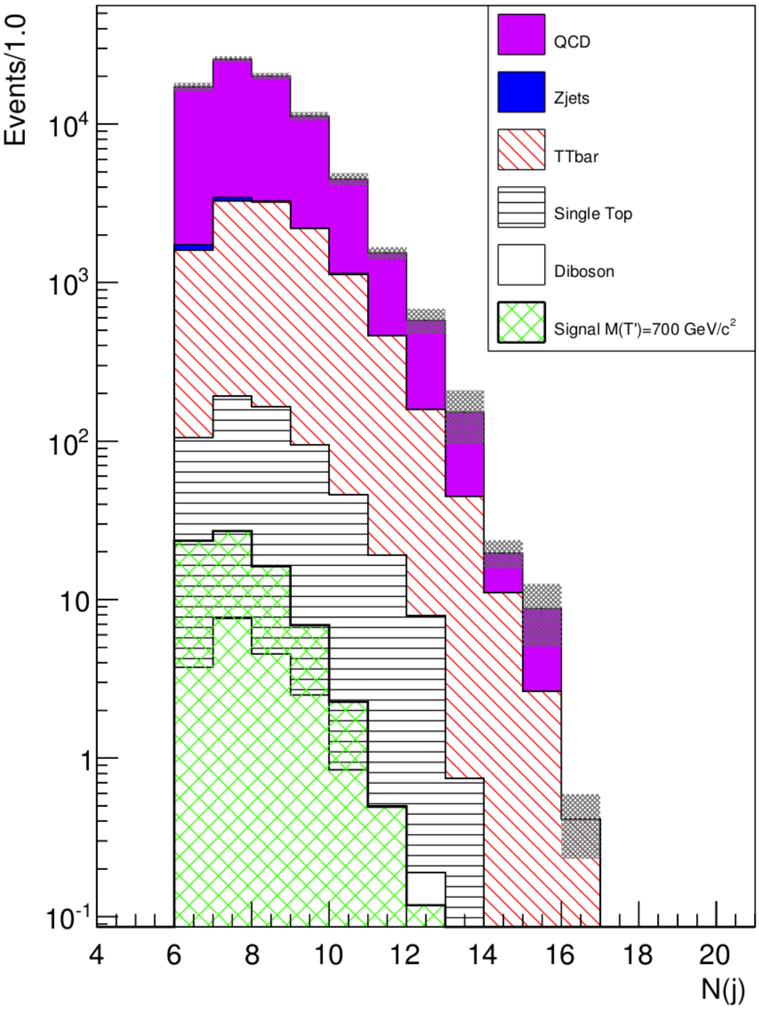


Figure 5.11: Jet multiplicity for MC samples after requiring at least 3 CSVM b-tagged jets. The sum of MC samples is normalized to the integrated luminosity. Signal is overlaid.

$$\chi^2 = \frac{(M_H - M_{bb})^2}{\sigma_H^2} + \frac{(M_W - M_{jj})^2}{\sigma_W^2} + \frac{(M_t - M_{bjj})^2}{\sigma_t^2} \tag{5.6}$$

The efficiency of the reconstruction of each particle (Higgs boson, $W^{+/-}$ boson, top quark and T') can be evaluated using signal MC samples. Two efficiencies are defined: the exclusive efficiency as the ratio between the number of events where the particle was correctly reconstructed by the χ^2 algorithm and the number of events where all jets were correctly matched to a parton, and the inclusive efficiency where the ratio is done with respect to the total number of events. Both efficiencies were evaluated after basic selection. The Higgs candidate is considered to be correctly reconstructed if the two jets chosen by the χ^2 algorithm correspond to the two b-jets coming from the Higgs boson. The same definition applies for the $W^{+/-}$ boson, while for the top candidate it applies only if the third b-tagged jet was correctly chosen to the MC truth matched jet from the top decay. The T' is considered to be correctly reconstructed if the five jets chosen to reconstruct the Higgs, $W^{+/-}$, and top candidates correspond to their jets from MC truth, independently if they were or not correctly reconstructed. The efficiency values are shown in figure 5.12. The exclusive efficiency for the T' is of around 70% independently of its mass.

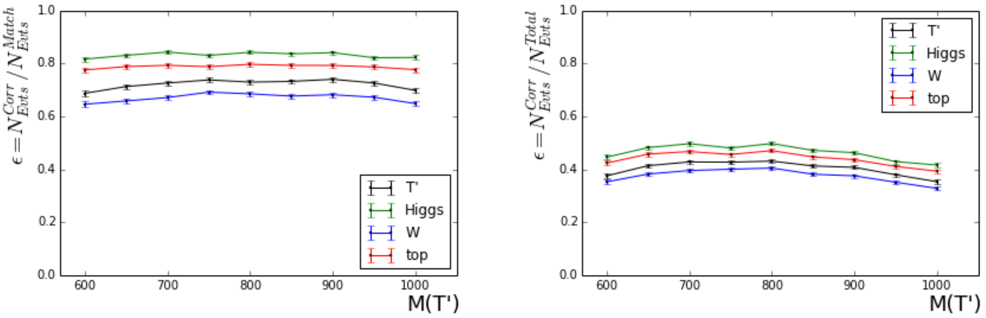


Figure 5.12: Reconstruction efficiency by the χ^2 algorithm of the Higgs boson, $W^{+/-}$ boson, top quark and T', as the ratio of the number of events where the particle was correctly reconstructed to the number of events where jets could be matched to partons [left] and to the total number of events [right].

In figure 5.13 the $W^{+/-}$, Higgs and top candidate masses reconstructed by the χ^2 sorting algorithm are shown. From the gaussian fit on each mass, the reconstructed masses and widths are close to the values used for the χ^2 variable definition. This correspondence shows a reliable reconstruction procedure of all three resonances. Additionally, in figure 5.14 the reconstructed T' mass for all used signal MC samples is shown.

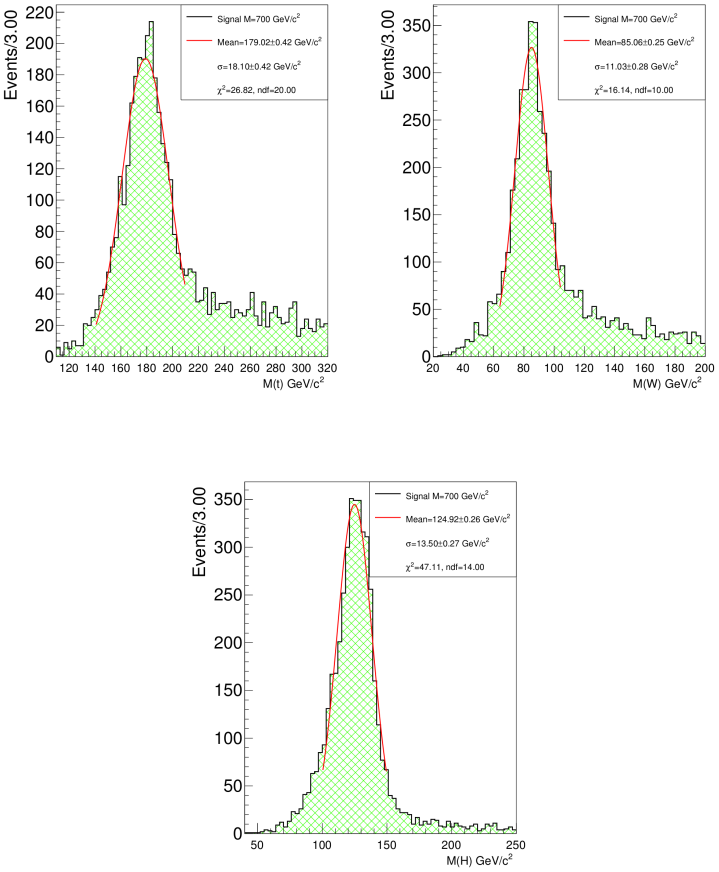


Figure 5.13: Reconstructed top, $W^{+/-}$ and H^0 masses for the T' mass point of 700 GeV/c^2. The reconstructed masses and widths of the three resonances, $M_H^{reco} = 124.92 \pm 0.26$ GeV/c^2, $M_W^{reco} = 85.06 \pm 0.26$ GeV/c^2, $M_t^{reco} = 179.02 \pm 0.42$ GeV/c^2, $\sigma_H^{reco} = 13.50 \pm 0.27$ GeV/c^2, and $\sigma_W^{reco} = 11.03 \pm 0.28$ GeV/c^2 and $\sigma_t^{reco} = 18.10 \pm 0.42$ GeV/c^2. The corresponding values used for the reconstruction procedure are: $M_H = 125$ GeV/c^2, $M_W = 84.06$ GeV/c^2, $M_t = 175.16$ GeV/c^2, $\sigma_H = 12.4$ GeV/c^2, and $\sigma_W = 10.12$ GeV/c^2 and $\sigma_t = 17.35$ GeV/c^2.

The χ^2 distribution is plotted in figure 5.15. Signal events have preferentially a $\chi^2 < 50$, while multijet and $t\bar{t}$ backgrounds have much longer tails at higher χ^2 values.

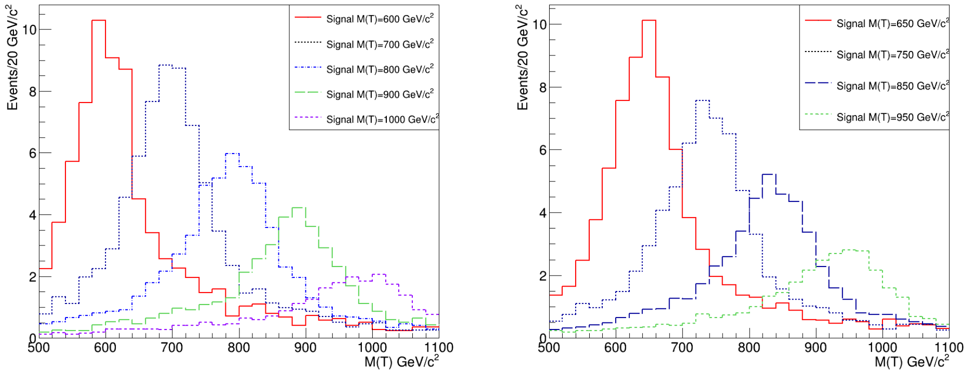


Figure 5.14: Reconstructed T' mass for all mass points from the χ^2 sorting algorithm after basic selection. Each mass point is normalized to luminosity and its corresponding cross section. A gaussian fit of these distributions will be presented afterward in section 5.3.4, accompanied with a discussion about the resolution on the reconstruction of the T'.

5.3.4 Selection based on reconstructed objects

After the reconstruction procedure, a final step in the selection is performed based on the properties of the reconstructed resonances. This last set of cuts has been optimized via a multidimensional scan to obtain the highest discrimination between signal and backgrounds. The discrimination of the selection has been evaluated from the estimator S/B, where the signal MC sample with $M(T) = 700$ GeV/c^2, and the $t\bar{t}$ and QCD_HT-500To1000 samples for the backgrounds have been used. These two background samples have the highest contribution to the invariant mass of five jets between 600 and 1000 GeV/c^2. For optimization purposes, in order not to have a high statistical error, the selection has been adjusted to keep at least 10 signal events, for the 700 GeV/c^2 mass point, after the full selection. In the following discussion, efficiency curves will be presented to show how different cut values, on each variable, select backgrounds and signal.

The main limitation for the optimization of the selection is the lack of statistics in the multijet MC sample. It is to be kept in mind that the plots with data MC comparison are shown for illustration only, since the limit setting and final results were produced with an estimation of backgrounds derived from the data.

The first variable considered for the selection is the χ^2 distribution. As seen in figure 5.15, the signal tends to have smaller values of χ^2 than the backgrounds. In figure 5.16, a scan of the efficiency of cutting on the χ^2 for the values dis-

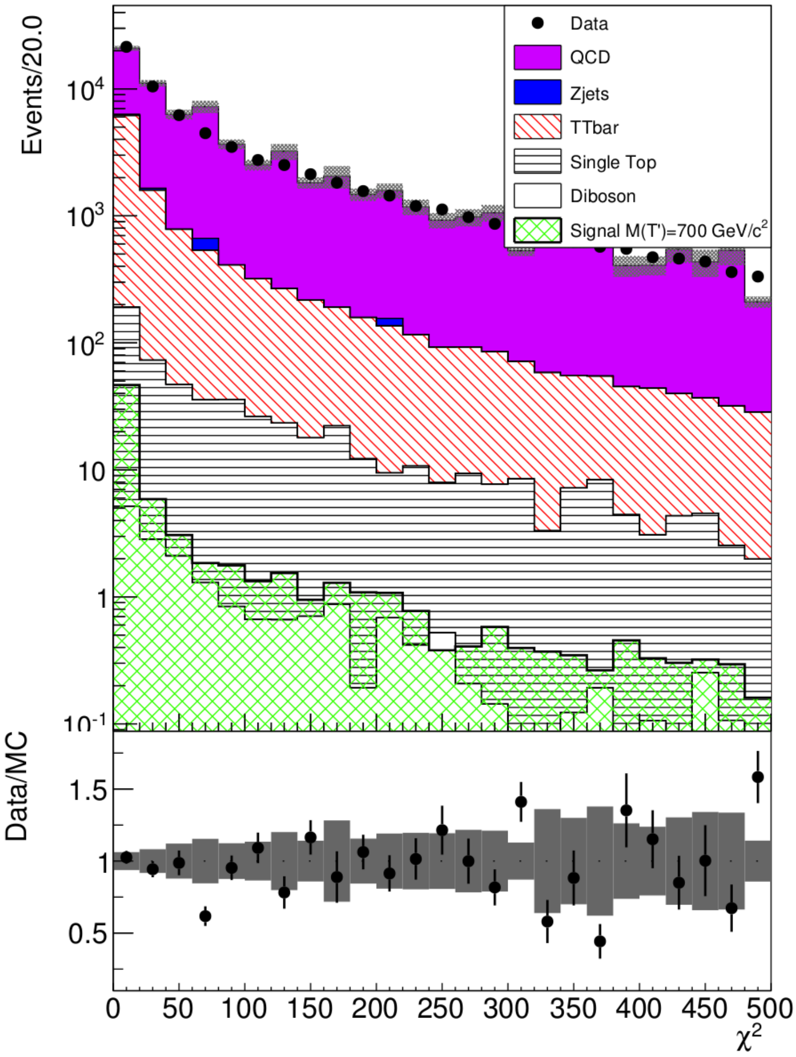


Figure 5.15: Distribution of the χ^2 variable for data and MC samples. The signal sample used has a T' mass of 700 GeV/c^2. Backgrounds present high tails. The sum of MC is normalized to the integrated luminosity. Signal has been weighted to the luminosity and its theoretical cross section.

played in the x-axis is shown. These efficiencies were evaluated for the signal ($M(T') = 700$ GeV/c^2), $t\bar{t}$ and QCD_HT-500To1000 MC samples. In addition, the ratio of the efficiency for signal over the efficiency for $t\bar{t}$ and QCD samples are shown. These ratios are directly proportional to S/B. In agreement with the increase of S/B as measure of the cut value, a cut on $\chi^2 < 8$ has been chosen. One additional feature of this cut, that will be seen in section 5.4, is that it ensures the agreement of the control sample with the signal sample for the background estimation (see section 5.4.1 and figure 5.32).

The cut value on the χ^2 variable has been chosen for various reasons. First, it was optimized to obtain a high S/B while not cutting too much on the signal, with the constraint of 10 signal events after full selection for the 700 GeV/c^2 mass point. Second, the scan performed for the optimization of the selection showed that it was preferable not to take an extremely tight cut on the χ^2 variable. A relatively loose cut on the χ^2 combined with additional criteria gave a higher S/B than a tight χ^2 selection and kept the desired 10 signal ($M = 700$GeV/c^2) events after full selection. Thus, the cut value was not chosen to be extremely tight in terms of signal efficiency (around 50%, as it will be shown in table 5.7).

After cutting on the χ^2, events were required to have the two b-quarks used to reconstruct the Higgs boson at a distance $\Delta R(bb) < 1.2$. This cut value has the characteristic of keeping the vast majority of the signal (90%), while reducing backgrounds by a half. As in signal events, the Higgs boson is produced by the massive T', it is expected to be produced with a high p_T. Therefore, the decay products from the Higgs, the $b\bar{b}$ system, are expected to be close in the $\eta - \phi$ plane. In figure 5.17, the $\Delta R(bb)$ distribution in data is compared to MC before cutting on this variable. In the same figure the efficiency of cutting on different values for each MC sample used in the optimization is shown, altogether with the efficiency ratios.

One important difference between signal and backgrounds is that in signal events there is a $b\bar{b}$ system coming from the decay of a Higgs boson. Thus, a resonance peak is expected to appear for the $b\bar{b}$ system for signal events, while it is not the case for backgrounds. In consequence, events were required to have a Higgs candidate mass between 105 and 145 GeV/c^2. This variable is controlled by the χ^2 cut, in consequence the tails of the Higgs candidate mass are greatly reduced by the criterion $\chi^2 < 8$. The efficiency plot of cutting on events outside a window around 125 GeV/c^2 in the Higgs candidate mass is shown in figure 5.18. In the same figure the Higgs candidate mass is shown, before cutting on it, for the data and MC samples. The cut value chosen on the Higgs candidate mass allows to select 90% of the signal while reducing by 50% the multijet background.

Due to the presence of a real top in $t\bar{t}$ events, this background is able to mimic the signal properties more likely than QCD. To understand how the $t\bar{t}$ system is

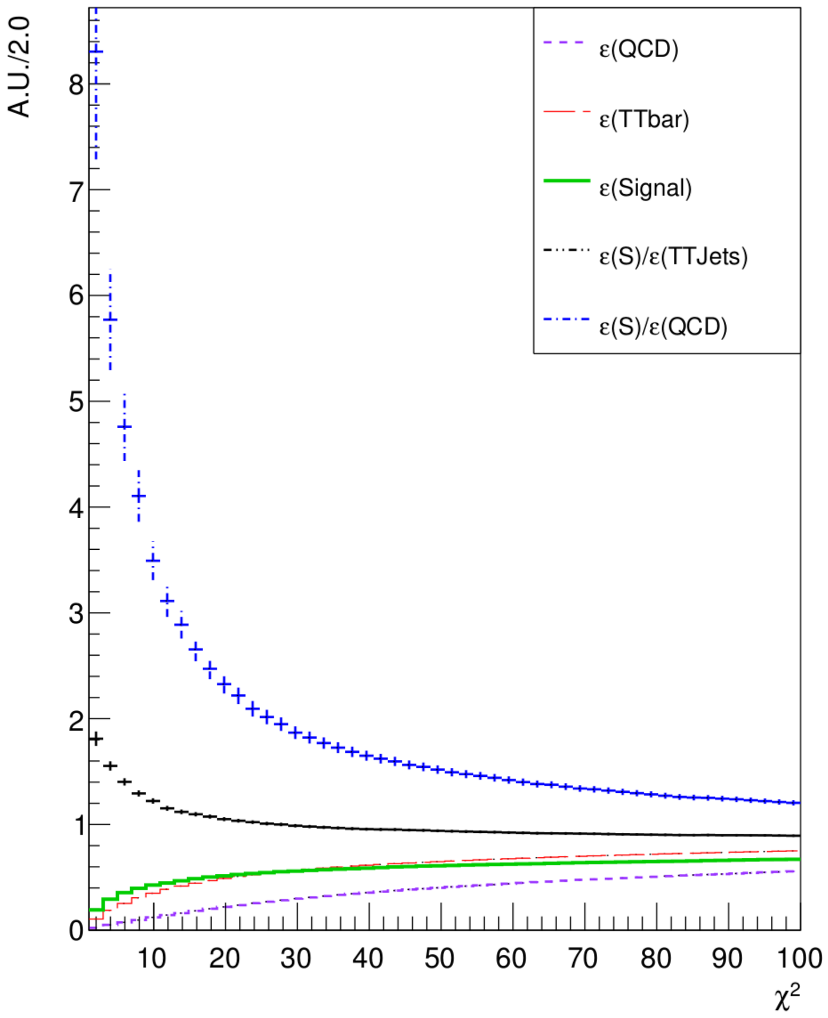


Figure 5.16: Efficiency of the selection using criterion $\chi^2 < x$, with x the cut value, as a function of the cut value for T' with $M = 700$ GeV/c^2, $t\bar{t}$ and QCD_HT-500To1000 MC samples. Ratios between efficiency for signal and each background are also displayed.

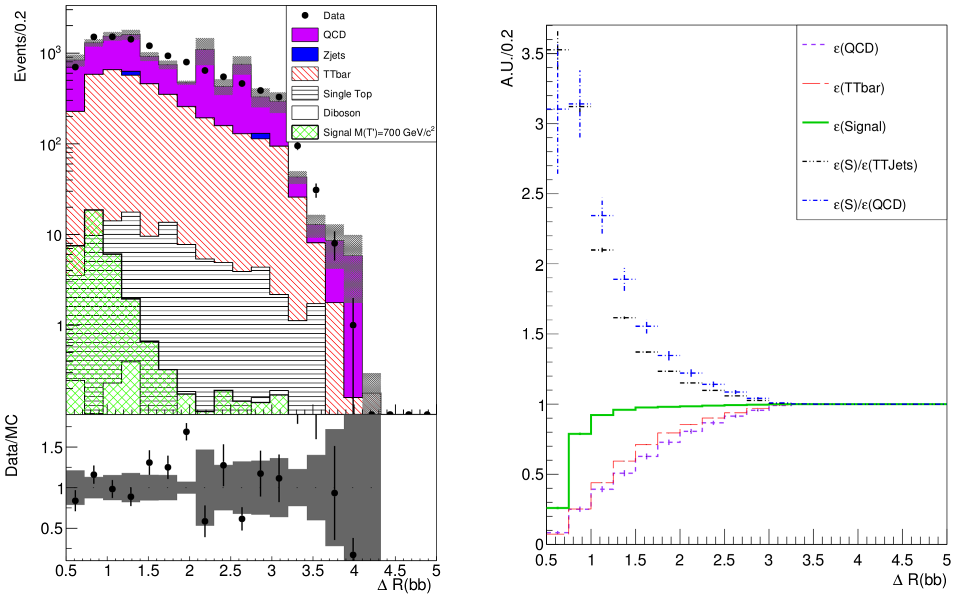


Figure 5.17: ΔR of the 2 b-tagged jets used to reconstruct the Higgs candidate after χ^2 cut. The signal which is simply overlaid prefers low ΔR while backgrounds have larger distribution at higher value. The gray band represents the statistical uncertainties from the sum of the MC background. Normalization was done to luminosity [left]. Efficiency of the selection with the criterion $\Delta R(bb) < x$, with x the cut value, as a function of the cut value for $M = 700$ GeV/c^2 signal sample, $t\bar{t}$ and QCD_HT-500To1000 MC samples and ratios between efficiency for signal and each background [right].

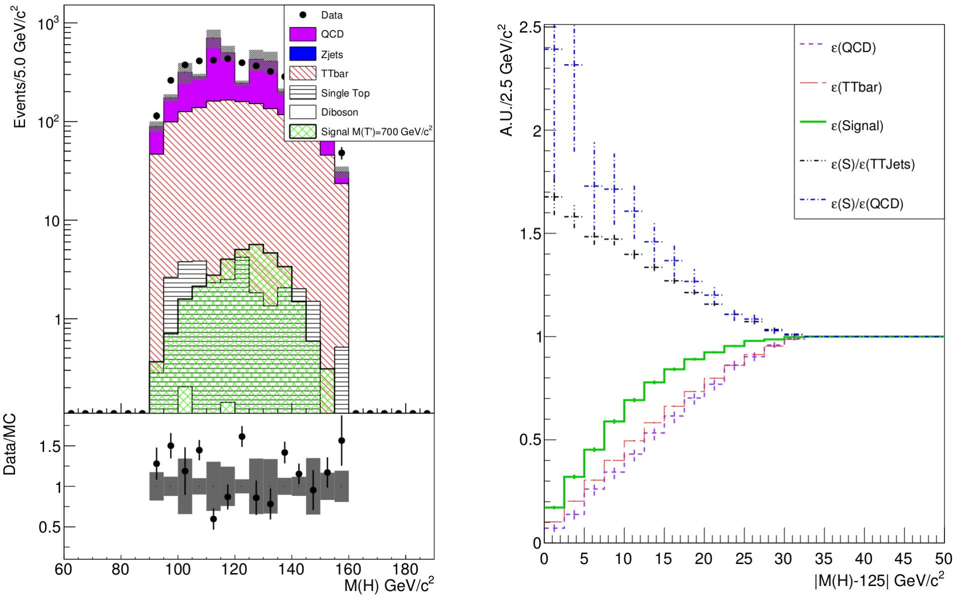


Figure 5.18: Distribution for $M(H_{cand})$ for data and the sum of Monte Carlo samples [left]. Efficiency of the selection with the criterion $|M(H_{cand}) - 125| < x$, with x the cut value, as a function of the cut value for the $M = 700$ GeV/c^2 signal sample, $t\bar{t}$ and QCD_HT-500To1000 MC samples and ratios between efficiency for signal and each background [right]. All others selection criteria are applied up to this one.

mimicking a Higgs boson, the MC truth information of the jets chosen by the χ^2 sorting algorithm as coming from the Higgs have been studied. In figure 5.19 the Higgs reconstruction cases are shown.

This figure shows that the Higgs candidate is preferentially reconstructed in $t\bar{t}$ events from a b-quark coming from a top quark and one quark coming from a $W^{+/-}$ boson. As one top quark is also reconstructed, this means that the Higgs is being reconstructed from the second top. Thus, to identify $t\bar{t}$ events, the b-tagged jets used for the Higgs and one additional jet not utilized by the χ^2 algorithm can be taken to reconstruct the second top quark. As this extra jet is coming from a top quark, it is expected to have a significant amount of p_T, in principle with higher p_T than extra jets coming from radiation processes, then the second top quark is reconstructed from the two Higgs jets and the leading jet not used by the χ^2 algorithm. A second $W^{+/-}$ boson is also defined from the sub-leading Higgs jet and leading jet not used by the χ^2 algorithm. From the top quark decay, the b-quark coming from it, is expected to have higher p_T than the two jets from the $W^{+/-}$ boson. Then the second $W^{+/-}$ boson definition follows the expectation that the two jets coming from the $W^{+/-}$ boson decay have lower p_T than the b-quark coming directly from top quark decay. The distribution for the second top quark mass is shown in figure 5.20 for data and MC samples. The top quark mass peak is clearly visible in $t\bar{t}$ events.

A variable with the second top quark and second $W^{+/-}$ boson to discriminate signal from background (mainly $t\bar{t}$) has been built. Taking the sum of the second top quark candidate and second $W^{+/-}$ boson candidate mass, divided by the Higgs boson candidate mass, a better discrimination is achieved between signal and backgrounds than with the second top quark candidate mass alone. The distribution for this variables is shown in figure 5.21 for data and the sum of the MC samples. Events with $\left[M(top^{2nd}) + M(W^{2nd})\right]/M(H) > 6.8$ were selected. The cut value was chosen in function of the optimization of the S/B and with the constraint to keep at least 10 signal ($M = 700$ GeV/c^2) events after full selection, with a signal efficiency of 60%, rejecting 70% of $t\bar{t}$ and 40% of QCD_HT-500To1000 background events.

The selection continues taking advantage of the angular relation between the T' and the quark produced in association with it for signal events. They are produced preferentially with high ΔR while for backgrounds the two objects have lower ΔR. The leading jet not used by the reconstruction procedure is defined as the jet produced in association to the T'. This jet will be referred as the 6th jet, j^6. The distribution of $\Delta R(T', j^6)$ can be found in figure 5.22 for data and MC samples along with the efficiency plot of cutting on $\Delta R(T', j^6)$ for $M = 700$ GeV/c^2 signal sample, $t\bar{t}$ and QCD_HT-500To1000 MC samples. At this stage the statistics in QCD MC samples begin to be very low. Events that have $\Delta R(T', j^6) > 4.8$ were

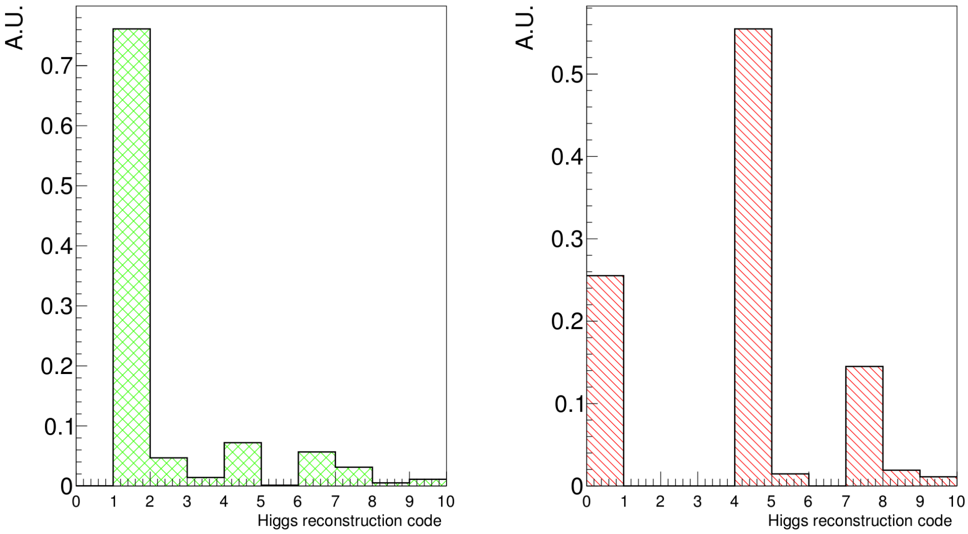


Figure 5.19: Higgs candidate reconstruction MC truth for signal (M=700 GeV/c^2) [left] and $t\bar{t}$ [right] MC samples. In signal the Higgs is reconstructed preferentially from two jets coming from the Higgs (bin 1) while for $t\bar{t}$ the Higgs candidate is reconstructed preferentially from a b-quark from a top and a quark from a $W^{+/-}$ boson (bin 4). The study is performed after χ^2 reconstruction, before cut over the χ^2 variable. The bins are defined as follows:

Bin 0: Higgs candidate reconstructed from two b-quarks coming each one from a top quark.

Bin 1: Higgs candidate reconstructed from Higgs boson $b\bar{b}$ system.

Bin 2: Higgs candidate reconstructed from one quark from the Higgs boson and one b-quark from a top quark.

Bin 3: Higgs candidate reconstructed from one quark from the Higgs boson and one quark from a $W^{+/-}$ boson.

Bin 4: Higgs candidate reconstructed from one b-quark from the top quark and one quark from a $W^{+/-}$ boson.

Bin 5: Higgs candidate reconstructed from two quarks coming from $W^{+/-}$ bosons.

Bin 6: Higgs candidate reconstructed from one jet from the Higgs boson and one additional quark (not from Higgs boson, $W^{+/-}$ boson nor top quark).

Bin 7: Higgs candidate reconstructed from one b-quark from a top quark and one additional quark (not from Higgs boson, $W^{+/-}$ boson nor top quark).

Bin 8: Higgs candidate reconstructed from one jet from a $W^{+/-}$ boson and one additional quark (not from Higgs boson, $W^{+/-}$ boson nor top quark).

Bin 9: Higgs candidate reconstructed from two additional quarks (not from Higgs boson, $W^{+/-}$ boson nor top quark).

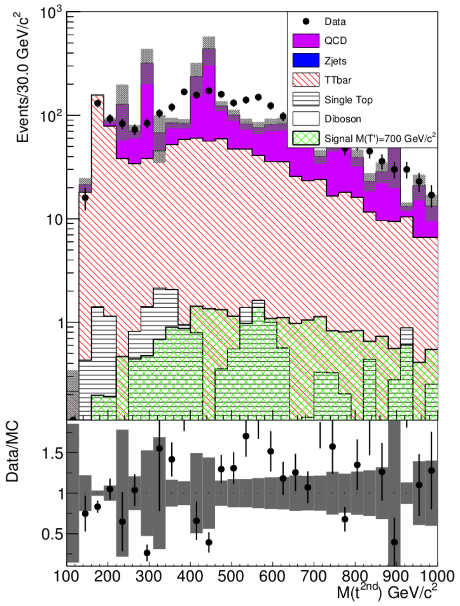


Figure 5.20: Second top mass distribution for data and the sum of the Monte Carlo samples. Selection criteria are applied up to the Higgs mass cut. The second top is reconstructed from the two Higgs jets and the leading jet not used by the χ^2 algorithm. The normalization was done with respect to luminosity.

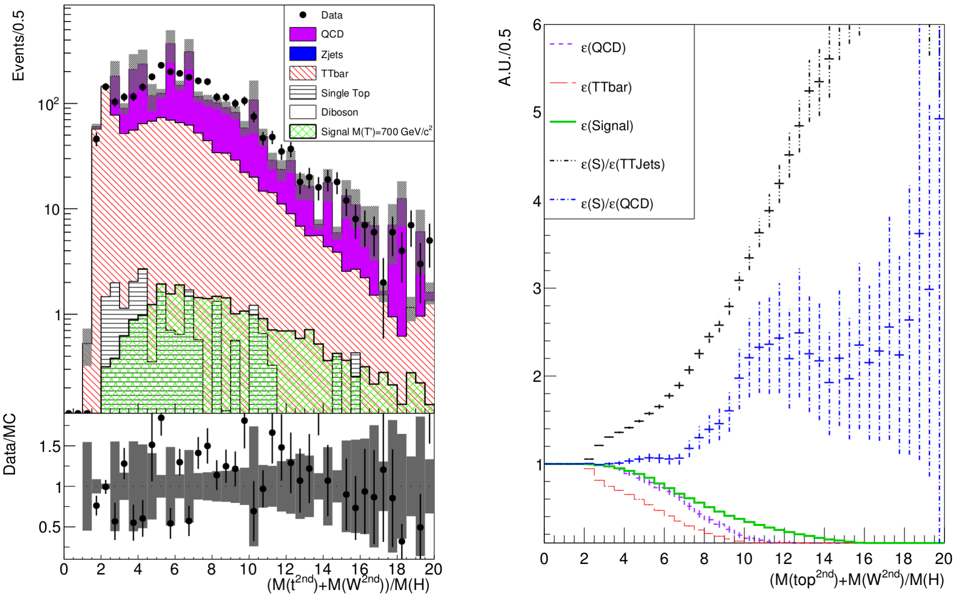


Figure 5.21: Distribution of $\left[M(top^{2nd}) + M(W^{2nd})\right]/M(H)$ for data and the sum of the Monte Carlo samples. Selection criteria are applied up to the Higgs mass cut. The low statistics in the multijet (QCD) MC sample is visible at this stage. The gray band represents the statistical uncertainties from the sum of the MC background and it is dominated by QCD samples [left]. Efficiency of selection criterion $\left[M(top^{2nd}) + M(W^{2nd})\right]/M(H) > x$, with x the cut value, as a function of cut value for $M = 700$ GeV/c^2 signal sample, $t\bar{t}$ and QCD_HT-500To1000 MC samples and ratios between efficiency for signal and each background [right].

selected, giving 60% efficiency on signal and rejecting 90% of $t\bar{t}$ and QCD_HT-500To1000 backgrounds.

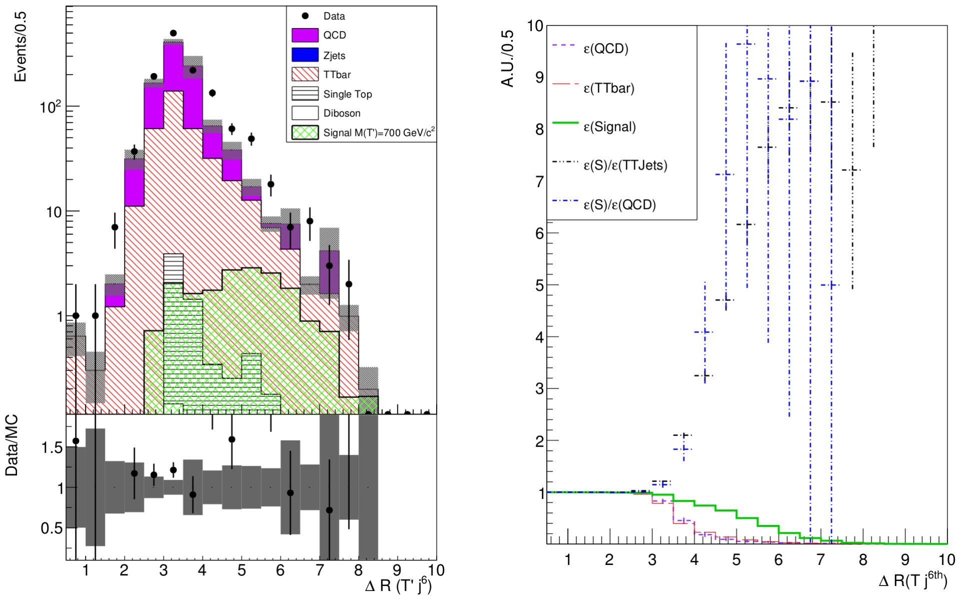


Figure 5.22: Distributions for $\Delta R(T', j^6)$ for data and the sum of Monte Carlo samples. All others criteria are applied up to this one. The low statistics in the multijet (QCD) MC sample is visible at this stage. The gray band represents the statistical uncertainties from the sum of the MC background and it is dominated by QCD samples [left]. Efficiency of selection criterion $\Delta R(T', j^6) > x$, with x the cut value, as a function of cut value for $M = 700$ GeV/c^2 signal sample, $t\bar{t}$ and QCD_HT-500To1000 MC samples and ratios between efficiency for signal and each background [right].

For the last criterion the Relative H_T variable is defined as the sum of the p_T of the Higgs candidate plus the p_T of the top candidate divided by H_T of the event. As for signal events, the top quark and Higgs boson candidates are coming from a heavy object, signal events have a Relative H_T closer to 1 than backgrounds events. The relative H_T distribution is shown in figure 5.23 for data and the sum of MC samples before cutting on it, with the rest of the selection applied. At this stage the QCD MC samples have very low statistics. In the same figure the efficiency plot of the cut can be found. Events that have a Relative $H_T > 0.67$ have been kept. For this last cut, the scan performed for the optimization of the selection took specially into account the S/B with respect to $t\bar{t}$ background, because the QCD sample used for the optimization had very low statistics. Thus, the cut value over the relative H_T was optimized to obtain the highest S/B with $t\bar{t}$ as background, keeping at least 10 signal events for the 700 GeV/c^2 mass point. This cut kept 90% of signal and rejected 50% of $t\bar{t}$ events.

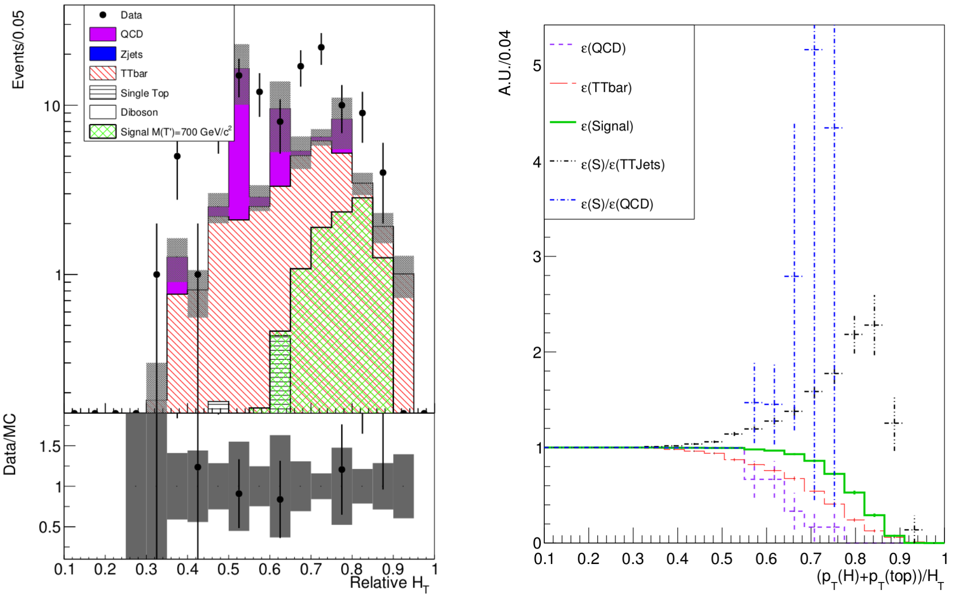


Figure 5.23: Distribution of Relative H_T variable for data and the sum of the Monte Carlo samples. All others criteria are applied up to this one. At this stage multijet (QCD) MC sample have very low statistics. The gray band represents the statistical uncertainties from the sum of the MC background and it is dominated by QCD samples [left]. Efficiency of the selection criterion Relative $H_T > x$, with x the cut value, as a function of the cut value, for $M = 700$ GeV/c^2 signal sample, $t\bar{t}$ and QCD_HT-500To1000 MC samples and ratios between efficiency for signal and each background [right].

To check the evolution of the selection, in table 5.5 the S/B after each cut is shown after the reconstruction of resonances with the χ^2 algorithm. In the table an increase of the estimator is visible at each step. As a reminder, $M = 700$ GeV/c^2 signal, $t\bar{t}$ and QCD_HT-500To1000 MC samples were used for this calculation.

Finally, in table 5.6, the results are presented after full selection with all mass points fitted with a gaussian. The found widths for the T' mass vary between 30 to 50 GeV/c^2, which represents about 6% of T' mass. This fit is used to calculate

Cut	S/B	$S/\sqrt{S+B}$
$\chi^2 < 8$	$3.4 \times 10^{-2} \pm 2.85 \times 10^{-3}$	0.96 ± 0.05
$\Delta R(bb) < 1.2$	$4.76 \times 10^{-2} \pm 4.52 \times 10^{-3}$	1.10 ± 0.07
105 GeV/c^2 $< M(H_{cand}) <$ 145 GeV/c^2	$6.37 \times 10^{-2} \pm 6.74 \times 10^{-3}$	1.22 ± 0.08
$(M(top^{2nd}) + M(W^{2nd}))/M(H_{cand}) > 6.8$	0.15 ± 0.03	1.45 ± 0.16
$\Delta R(Tj^6) > 4.8$	0.42 ± 0.19	1.67 ± 0.32
Relative $H_T > 0.67$	1.16 ± 0.17	2.13 ± 0.17

Table 5.5: S/B and $S/\sqrt{S+B}$ using MC samples for each step of the selection after reconstruction of the resonances with the χ^2 sorting algorithm. Only $M = 700$ GeV/c^2 signal, $t\bar{t}$ and QCD_HT-500To1000 MC samples were used.

the expected yields for the signal in order to calculate the limits (see sections 6.2 and 6.3).

Generated		Reconstructed		
Mass (GeV/c^2)	Width (GeV/c^2)	Mass (GeV/c^2)	Width (GeV/c^2)	χ^2/ndf
600	0.62	604.60 ± 14.18	32.44 ± 10.37	4.99/4
650	0.80	644.56 ± 12.64	35.53 ± 9.54	8.03/4
700	1.02	691.79 ± 13.65	41.16 ± 9.75	10.80/7
750	1.27	736.26 ± 15.53	45.38 ± 11.46	10.24/7
800	1.56	782.77 ± 18.17	49.52 ± 13.54	24.10/7
850	1.89	832.86 ± 18.09	47.89 ± 13.44	16.06/7
900	2.26	881.53 ± 19.12	45.69 ± 14.23	11.50/7
950	2.67	929.02 ± 24.97	51.48 ± 18.91	14.11/7
1000	3.13	970.48 ± 32.15	53.45 ± 25.13	10.42/7

Table 5.6: Reconstructed mass and width for T' candidates after full analysis selection using a gaussian fit for each signal mass generated.

5.3.5 Efficiencies

Trigger

The trigger efficiency has been computed using as reference the prescaled trigger requiring $H_T > 400$ GeV/c (HLT_HT400), where the H_T is defined as the scalar sum of the p_T of the calojets. From other analyses using H_T triggers, as in [155], the trigger plateau is reached around 100 GeV after the threshold. Assuming it is the case also for HLT_HT400, as the selection is requiring $H_T > 550$ GeV/c, it selects events in the efficiency plateau. As the procedure is intended to measure the turn on curve of HLT_Dijet80_Dijet60_Dijet20, it is important that the used events are in the plateau of HLT_HT400 in order not to introduce effects from the turn on curve from this later trigger. The trigger efficiency is normally evaluated after full selection. The trigger efficiency has been monitored at an earlier stage and after full selection. At the earlier stage of the selection used, the major contribution is coming from backgrounds, thus not only the signal MC samples have been used but also the $t\bar{t}$ MC sample. The differences between this last sample and data were used to compute a systematic uncertainty on signal that enter the calculation of limits.

The p_T of the 6th leading jet has been used to parametrize the trigger efficiency. This efficiency is defined as the ratio between the number of events passing both triggers (HLT_HT400 and HLT_Dijet80_Dijet60_Dijet20) and the number of events passing only HLT_HT400. HLT_HT400 is used as an unbiased sample with respect to HLT_Dijet80_Dijet60_Dijet20. In figure 5.24, the trigger efficiency is presented as a function of the 6th leading jet p_T, and after the χ^2 cut for all signal MC samples.

Additionally, trigger efficiencies were computed after full selection. They are shown in figure 5.25. An underestimation of signal events with respect to data by trigger selection is observed for a 6th jet p_T between 30-50 GeV/c and 70-90 GeV/c. However, $t\bar{t}$ has a slightly higher efficiency than data. As one of the main components of data is $t\bar{t}$, and there are not enough statistics in QCD MC samples to measure their trigger efficiency, $t\bar{t}$ trigger efficiency has been used to estimate a systematic uncertainty on the trigger selection for signal MC samples as a function of 6th leading jet p_T. The entire procedure of the estimation of the systematic uncertainty from trigger on signal yields is described in section 6.2.

Selection

In figure 5.26 the total efficiency of the selection for each signal mass point is shown. The lowest efficiency is for 600 GeV/c^2 mass point, around 1%, while the highest is for 850 GeV/c^2 mass point, around 2%. This efficiency has been calculated with respect to events passing the trigger selection.

Additionally, in table 5.7 the efficiencies of each step of the selection for signal sample ($M = 700$ GeV/c^2), multijet background, top background ($t\bar{t}$ + single top), and diboson background MC samples are quoted. These efficiencies have been calculated with respect to the number of events passing the trigger. Table 5.8 presents the cut flow of the entire selection showing the number of events expected in MC samples and observed in data. Finally, for completeness, tables 5.9, 5.10

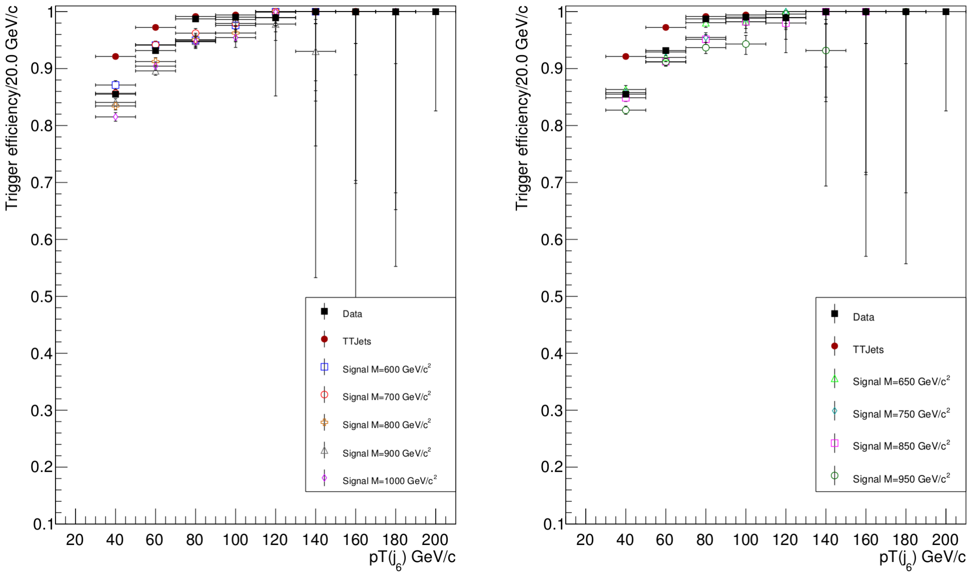


Figure 5.24: Efficiency for data and the MC signal samples for events passing trigger HLT_Dijet80_Dijet60_Dijet20 with respect to trigger bit HLT_HT400 after standard selection up to the χ^2 cut (included). At this early stage of the selection, discrepancies between 10% and 6% at higher p_T are observed between data and signal MC samples. Differences between $t\bar{t}$ and data are at a maximum of 7%. This efficiency is parametrized as a function of the 6^{th} jet p_T. Efficiencies for signal MC samples with T' masses equal to 600, 700, 800, 900 and 1000 GeV/c^2 are shown as well as for $t\bar{t}$ and data [left]. Efficiencies for signal MC samples with T' masses equal to 650, 750, 850 and 950 GeV/c^2 are shown with $t\bar{t}$ and data [right].

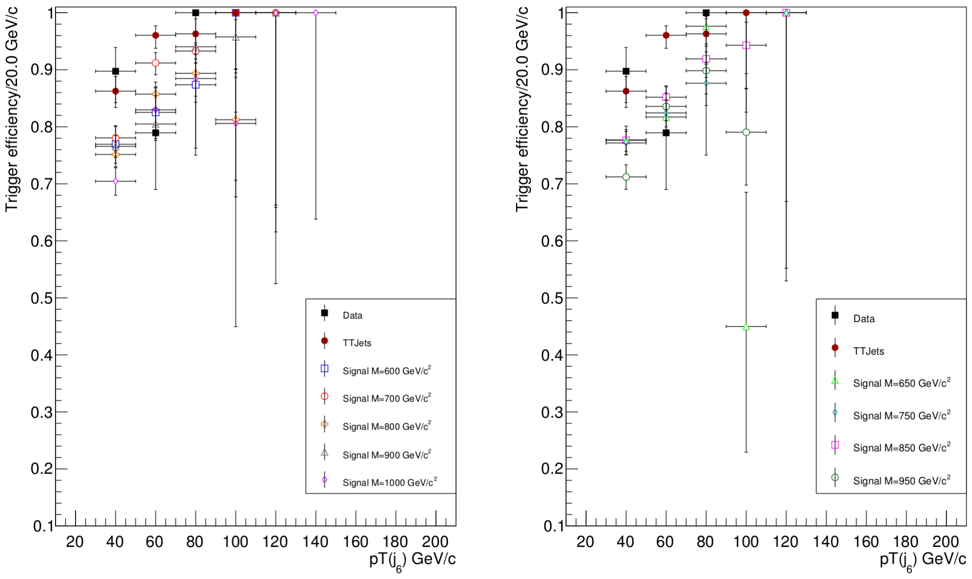


Figure 5.25: Efficiency for data and the MC signal samples for events passing trigger bit HLT_Dijet80_Dijet60_Dijet20 with respect to trigger bit HLT_HT400 after full selection. This efficiency is parametrized as a function of the 6^{th} jet p_T. The dispersion observed is about 10% between data and signal MC samples, while only about 4% for $t\bar{t}$. Efficiencies for signal MC samples with T' masses equal to 600, 700, 800, 900 and 1000 GeV/c^2 are shown with $t\bar{t}$ and data [left]. Efficiencies for signal MC samples with T' masses equal to 650, 750, 850 and 950 GeV/c^2 are shown with $t\bar{t}$ and data [right].

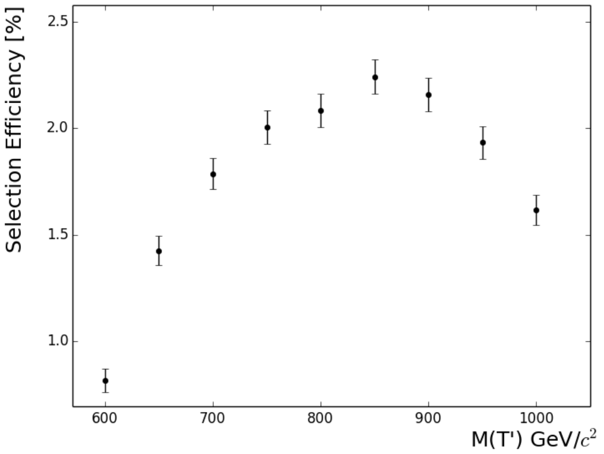


Figure 5.26: Selection efficiency for each signal mass point. Highest efficiency is achieved for 850 GeV/c^2 mass point. Error bars represent the binomial error of the efficiency for each mass point. These efficiencies have been calculated with respect to the number of events that passed the trigger selection.

and 5.11 show the number of unweighted events, with respect to lumi, for all the MC samples used in the analysis at each step of the selection. It also presents the weights used for each MC sample. The table for QCD MC samples, table 5.9, shows how quickly the statistics for multijet background diminish and eventually go to zero, which leads to an underestimation of the error associated to these samples.

5.4 Background estimation from data

As seen the in last sections, there are not enough statistics in the background MC samples in order to properly estimate the expected SM contamination after full selection. Moreover, MC predictions for the QCD component are affected by large systematic uncertainties that would diminish their reliability. This is why it is preferable to estimate the backgrounds making use of data.

The main difficulty to estimate the backgrounds comes from the correlation between the variables used in the selection. This made difficult to use established data driven background estimation methods as the ABCD method (See for example [155]). Another difficulty comes from the need of deriving the two dominant backgrounds after full selection: multijets and $t\bar{t}$. Despite of these difficulties, a

Selection	Cut	Signal (M=700 GeV/c^2)	Multijet	$t\bar{t}$ + single top	Diboson
Basic	Trigger cut and p_T,η selection	52.68 ± 2.11	$18.53 \pm 5 \times 10^{-3}$	36.55 ± 0.06	16.14 ± 0.34
	$j_1 > 150$ GeV/c	47.94 ± 2.11	$14.04 \pm 4 \times 10^{-3}$	24.73 ± 0.06	11.58 ± 0.30
	$H_T > 550$ GeV/c	47.65 ± 2.11	$13.53 \pm 4 \times 10^{-3}$	24.22 ± 0.06	11.16 ± 0.29
	$n_b^{CSVM} >= 3$	14.57 ± 1.49	$0.06 \pm 3 \times 10^{-4}$	1.98 ± 0.02	0.15 ± 0.04
Analysis	$\chi^2 < 8$	7.09 ± 1.09	$5 \times 10^{-3} \pm 8 \times 10^{-5}$	0.64 ± 0.01	0.01 ± 0.01
	$\Delta R(bb) < 1.2$	6.47 ± 1.04	$2 \times 10^{-3} \pm 5 \times 10^{-5}$	$0.26 \pm 7 \times 10^{-3}$	$7 \times 10^{-3} \pm 7 \times 10^{-3}$
	105 GeV/c^2 < $M(H_{cand})$ < 145 GeV/c^2	5.76 ± 0.99	$1 \times 10^{-3} \pm 4 \times 10^{-5}$	$0.19 \pm 6 \times 10^{-3}$	$3 \times 10^{-3} \pm 5 \times 10^{-3}$
	$\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$	3.63 ± 0.79	$5 \times 10^{-4} \pm 3 \times 10^{-5}$	$0.06 \pm 3 \times 10^{-3}$	$2 \times 10^{-3} \pm 4 \times 10^{-3}$
	$\Delta R(T', j^6) > 4.8$	2.02 ± 0.60	$2 \times 10^{-5} \pm 6 \times 10^{-6}$	$6 \times 10^{-3} \pm 1 \times 10^{-3}$	—
	$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	1.80 ± 0.56	$4 \times 10^{-6} \pm 2 \times 10^{-6}$	$3 \times 10^{-3} \pm 8 \times 10^{-4}$	—

Table 5.7: Cumulative efficiencies, in %, for signal and main background as a function of the applied cuts. After the $\Delta R(T', j^6) > 4.8$ cut, there are no left events in the diboson MC samples. These efficiencies have been calculated with respect to the number of events that passed the trigger selection.

Cut	Signal (M=700 GeV/c^2)	Multijet	$t\bar{t}$ + single top	Diboson	Data
Trigger selection	560.31±3.13	74803879.03±190145.80	601988.30±512.93	11718.97±47.35	451250111±21242.65
p_T and η selection on jets	295.15±2.27	13863750.21±73389.97	219998.68±275.69	1891.41±19.02	12865712±3586.88
$j_1 > 150$ GeV/c	268.60±2.17	10501566.69±58350.08	148893.06±232.38	1357.09±16.11	9303286±3050.13
$H_T > 550$ GeV/c	267.01±2.16	10123680.68±56326.87	145792.65±229.61	1307.81±15.82	9001871±3000.31
$n_b^{CSVM} >= 3$	81.65±1.19	48381.01±3554.13	11920.15±61.87	17.09±1.74	73879±271.81
$\chi^2 < 8$	39.49±0.83	4284.06±947.99	3858.73±29.84	1.65±0.51	10581±102.86
$\Delta R(bb) < 1.2$	36.00±0.79	1343.66±249.14	1552.02±18.01	0.78±0.29	3874±62.24
105 GeV/c^2 < $M(H_{cand})$ < 145 GeV/c^2	32.05±0.75	1023.03±220.82	1138.84±15.37	0.40±0.21	2820±53.10
$\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$	20.19±0.59	400.00±102.61	359.50±8.78	0.19±0.14	1242±35.24
$\Delta R(T', j^6) > 4.8$	11.27±0.44	17.90±7.34	36.00±2.16	0±0	—
$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	10.01±0.42	3.16±3.07	20.95±1.28	0±0	—

Table 5.8: Cut flow of expected events from MC samples and observed events in data as a function of the applied cuts. After the $\Delta R(T', j^6) > 4.8$ cut, there are no longer events in the diboson MC samples.

method to derive the shape of backgrounds contribution to $M(5j)$ from the data, the invariant mass of the 5-jets chosen to reconstruct the T', after full selection and one additional method to estimate its normalization have been derived.

5.4.1 Method for background shape estimation

For the shape estimation of the backgrounds a control sample enriched by backgrounds while depleted in signal events has been built. From the selection, the most stringent criterion against the background is $n_b^{CSVM} >= 3$ (see table 5.7). This criterion is selecting only 0.4% of multijet events and 8% of $t\bar{t}$ events. Reason why the number of b-tagged jets was taken as variable suited to define the control sample.

The control sample is defined as the region in which the selection is the same but $n_b^{CSVL} >= 3$ and vetoing events with $n_b^{CSVM} >= 3$. With the veto on events

Cut	QCD_HT-500To1000	QCD_HT-1000ToInf	QCD_Pt-120to170	QCD_Pt-170to300	QCD_Pt-300to470	QCD_Pt-470to600	QCD_Pt-600to800	QCD_Pt-800to1000
Trigger selection	4341732 ± 2083.68	3626698 ± 1904.39	84995 ± 291.54	273227 ± 522.71	604608 ± 777.57	515380 ± 717.90	547153 ± 739.70	544121 ± 737.65
p_T and η selection on jets	1239652±1113.40	1590777±1261.26	10695±103.42	66092±257.08	214821±463.49	206651±454.59	226714±476.14	227544±477.02
$j_1 > 150$ GeV/c	1111168±1054.12	1590771±1261.26	5873±76.64	57578±239.95	213922±462.52	206606±454.54	226707±476.14	227542±477.01
$H_T > 550$ GeV/c	1097972±1047.84	1590771±1261.26	5329±73.00	56575±237.85	213880±462.47	206605±454.54	226707±476.14	227542±477.01
$n_b^{CSVM} >= 3$	5522±74.31	9114±95.47	18±4.24	303±17.41	1325±36.40	1196±34.58	1314±36.25	1283±35.82
$\chi^2 < 8$	526±22.93	548±23.41	1±1	32±5.66	102±10.10	70±8.37	57±7.55	40±6.32
$\Delta R(bb) < 1.2$	195±13.96	286±16.91	0±0	10±3.16	51±7.14	42±6.48	23±4.80	20±4.47
105 GeV/$c^2 < M(H_{cand}) <$ 145 GeV/c^2	137±11.70	222±14.90	0±0	8±2.83	42±6.48	34±5.83	16±4	14±3.74
$\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$	77±8.77	201±14.18	0±0	1±1	28±5.29	31±5.57	14±3.74	13±3.61
$\Delta R(T', j^6) > 4.8$	6±2.45	5±2.24	0±0	0±0	0±0	0±0	0±0	0±0
$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	1±1	2±1.41	0±0	0±0	0±0	0±0	0±0	0±0
Weight	28.59×10^{-1}	14.95×10^{-2}	518.02	58.06	29.24×10^{-1}	28.45×10^{-2}	6.76×10^{-2}	0.89×10^{-2}

Table 5.9: Cut flow of unweighted events from QCD MC samples as a function of cuts applied. Errors are calculated as $\sqrt{N}$ of the central value. In the last line, the weights corresponding to each sample from normalization to luminosity are presented.

Cut	TTJets	T_tW-channel	T_t-channel	T_s-channel	Tbar_tW-channel	Tbar_t-channel	Tbar_s-channel
Trigger selection	6970016±2640.08	33485±182.99	49803±223.17	5304±72.83	33001±181.66	23459±153.16	2416±49.15
p_T η selection on jets	2633335±1622.76	9080±95.29	9863±99.31	1172±34.23	8827±93.95	4477±66.91	494±22.23
$j_1 > 150$ GeV/c	1772135±1331.22	6696±81.83	7384±85.93	954±30.89	6467±80.42	3257±57.07	390±19.75
$H_T > 550$ GeV/c	1735831±1317.51	6549±80.93	7155±84.59	931±30.51	6321±79.50	3140±56.04	377±19.42
$n_b^{CSVM} >= 3$	143984±379.45	403±20.07	439±20.95	67±8.19	413±20.32	216±14.70	27±5.20
$\chi^2 < 8$	47840±218.72	66±8.12	47±6.86	10±3.16	81±9	23±4.80	8±2.83
$\Delta R(bb) < 1.2$	19350±139.10	21±4.58	18±4.24	4±2	21±4.58	8±2.83	2±1.41
105 GeV/$c^2 < M(H_{cand}) <$ 145 GeV/c^2	14201±119.17	14±3.74	14±3.74	3±1.73	16±4	6±2.45	1±1
$\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$	4467±66.84	3±1.73	9±3	1±1	5±2.24	4±2	0±0
$\Delta R(T', j^6) > 4.8$	446±21.12	0±0	3±1.73	0±0	0±0	0±0	0±0
$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	266±16.31	0±0	0±0	0±0	0±0	0±0	0±0
Weight	78.77×10^{-3}	44.01×10^{-2}	28.80×10^{-2}	28.76×10^{-2}	44.38×10^{-2}	30.30×10^{-2}	24.79×10^{-2}

Table 5.10: Cut flow of unweighted events from $t\bar{t}$ and single top MC samples as a function of cuts applied. Errors are calculated as $\sqrt{N}$ of the central value. In the last line, the weights corresponding to each sample from normalization to luminosity are presented.

with at least 3 CSVM b-tagged jets a high contamination from signal events in the control sample is prevented. However, some contamination remains (7% maximum, see table 5.12). By opposition, the set of events passing the standard selection will be referred as signal sample. This contamination will be studied later in this section. The control sample then is formed by background events, mainly QCD and $t\bar{t}$. In figure 5.27 a schematic diagram for the construction of the control and signal samples is presented.

From this control sample the background shape in the signal sample in the five jets invariant mass is estimated. Then the hypothesis is that the $M(5j)$ is equivalent in both control and signal sample, apart from a possible signal contribution in the signal sample. In the next section the procedure to test the validity of this hypothesis will be described.

Cut	WZ	ZZ	WW	Signal (M=700 GeV/c^2)
Trigger selection	66144±257.18	65230±255.40	51613±227.18	32081±179.11
p_T and η selection on jets	10744±103.65	9934±99.67	8353±91.39	16899±130.00
$j_1 > 150$ GeV/c	7679±87.63	7197±84.84	6027±77.63	15379±124.01
$H_T > 550$ GeV/c	7404±86.05	6965±83.46	5795±76.12	15288±123.46
$n_b^{CSVM} >= 3$	97±9.85	253±15.91	38±6.16	4675±68.37
$\chi^2 < 8$	9±3	34±5.83	2±1.41	2261±47.55
$\Delta R(bb) < 1.2$	4±2	22±4.69	0±0	2061±45.40
105 GeV/$c^2 < M(H_{cand}) <$ 145 GeV/c^2	2±1.41	12±3.46	0±0	1835±42.84
$\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$	1±1	5±2.24	0±0	1156±34.00
$\Delta R(T', j^6) > 4.8$	0±0	0±0	0±0	645±25.40
$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	0±0	0±0	0±0	573±23.94
Weight	11.04×10^{-2}	15.29×10^{-3}	66.24×10^{-3}	17.47×10^{-3}

Table 5.11: Cut flow of unweighted events from diboson and signal (M=700 GeV/c^2) MC samples as a function of cuts applied. Errors are calculated as $\sqrt{N}$ of the central value. In the last line, the weights corresponding to each sample from normalization to luminosity are presented.

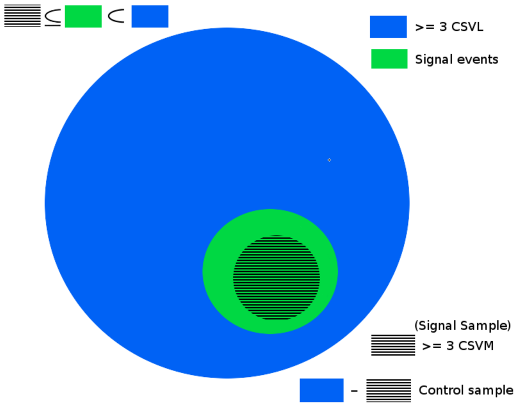


Figure 5.27: Schematic representation of signal and control sample. The big blue circle represents the ensemble of events with at least 3 CSVL b-tagged jets. The green circle represents the signal MC events while the circle filled with horizontal lines represents the signal sample (events with at least 3 CSVM b-tagged jets). The control sample then is the blue circle minus the horizontal line filled circle.

Validation

In order to validate that the $M(5j)$ shape in the control sample can be used to estimate the background shape in the signal sample, a shape comparison between signal and control samples at different stages of the selection was performed. An agreement of shapes is then shown at several stages between both samples. From this validation a source of systematic uncertainty has been estimated. The comparison is only performed in the $M(5j)$ range between 550 and 1100 GeV/c^2, the range of explored signal masses.

The validation is performed at 4 different stages of the selection. The signal sample $M(5j)$ shape, after the selection with at least 3 CSVM b-tagged jets, is compared to the control sample, selection with at least 3 CSVL b-tagged jets and veto on events with at least 3CSVM b-tagged jets, at each stage. The 4 selected stages for the comparison are:

- Stage A: Selection up to $\chi^2 < 8$, included.
- Stage B: Selection up to $\Delta R(bb) < 1.2$, included.
- Stage C: Selection up to $\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$, included.
- Stage D: Full selection.

In addition, in order to check that the $M(5j)$ shape in the control sample is independent of the b-tagging working point, the validation procedure is redone for intermediate working points between CSVL and CSVM. As seen in section 2.3.3, the CSVM working point corresponds to a cut in the b-tagging discriminator greater than 0.679, while CSVL corresponds to a cut at 0.244. Two additional working points were chosen to perform the validation, corresponding to cuts at 0.389 and 0.534. These two points are equally separated between CSVL and CSVM.

As it is unknown if there is a signal in data, the data-data (control sample and signal sample in data) comparison is performed only for stages A, B and C where background dominates as it is shown in figure 5.28. In addition QCD-QCD MC samples (control sample and signal sample in QCD MC) comparison is only performed for A, B and C stages due to low statistics at the end of the selection (shown in figure 5.29). The comparison for $t\bar{t}$ (control sample and signal sample in $t\bar{t}$ MC sample) and the sum of MC samples is shown in figures 5.30 and 5.31, respectively. After full selection it is shown that both shapes are in agreement for $t\bar{t}$ and the sum of MC samples, thus it can be deduced that the estimation is also valid in data. Another reason to drive this conclusion, is that data-data validation shows an agreement for stages A, B and C. Additionally, a general agreement is shown at all stages for all working points.

The cut on the χ^2 is not only useful for signal discrimination but also drives the agreement in the $M(5j)$ between control and signal samples. This cut ensures that the events in the control sample have very similar characteristics to signal sample events at an early stage of the selection. More precisely, this cut grants similar

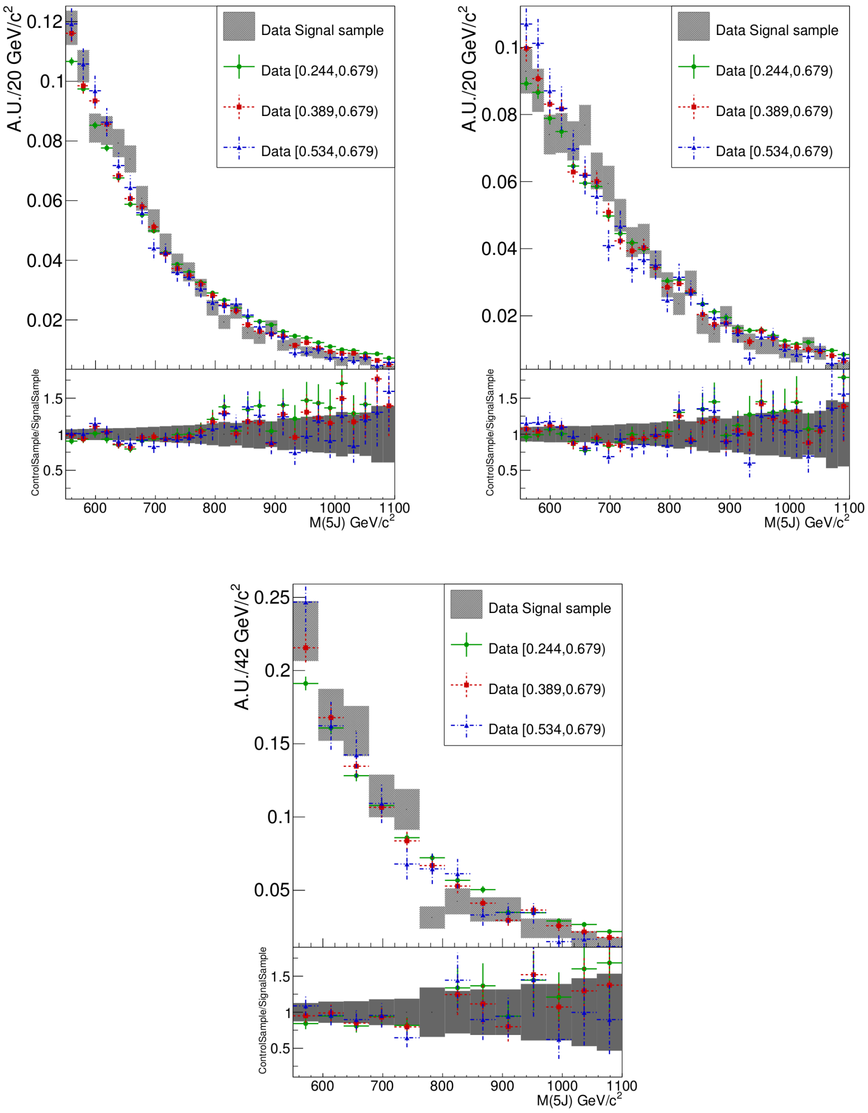


Figure 5.28: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done for data within 3 stages of selection: A [top left], B [top right] and C [bottom]. The 3 working points are given in different colors. Within statistical error, the 3 shapes for the control sample are in agreement at all stages with the signal sample, in gray. All histograms have been normalized to unity.

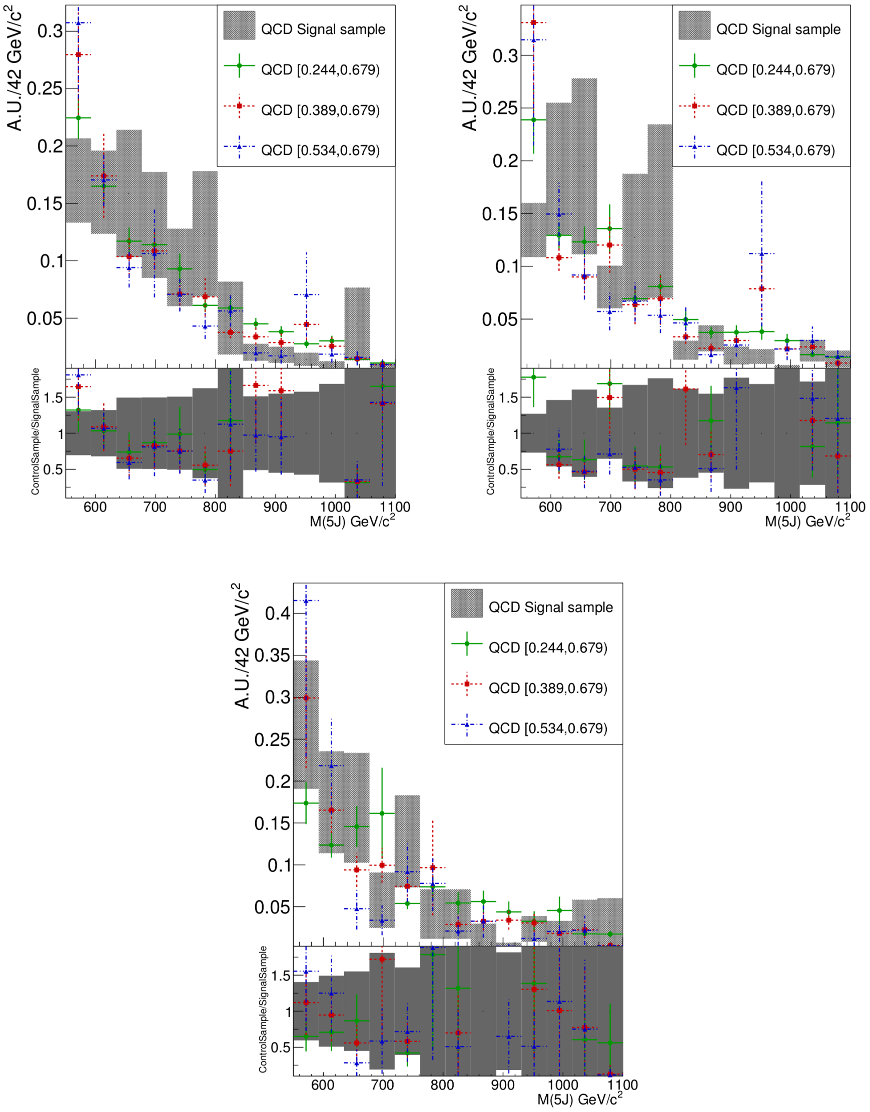


Figure 5.29: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done for QCD Monte Carlo samples within 3 stages of selection: A [top left], B [top right] and C [bottom]. The 3 working points are given in different colors. The signal sample is displayed in gray. QCD MC as all the other Monte-Carlo samples are purely used for illustration. All histograms have been normalized to unity. The errors of the QCD samples are underestimated due to their low statistics. To correctly estimate the associated errors, for each MC sample an error of 1.8 events (times the corresponding weight) should be added to each bin with zero entries.

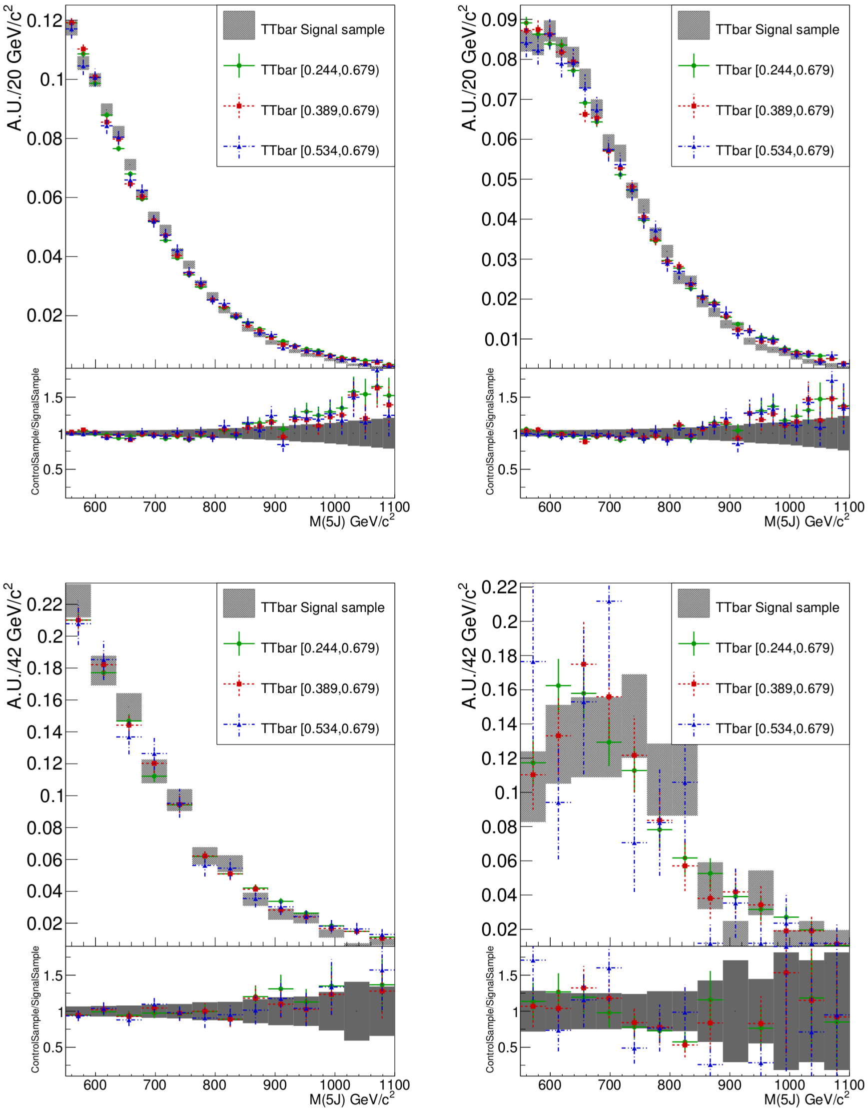


Figure 5.30: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done for $t\bar{t}$ Monte Carlo samples within 4 stages of selection: A [top left], B [top right], C [bottom left] and D [bottom right]. The 3 working points are given in different colors. Within statistical error, the 3 shapes for the control sample are in agreement at all stages with the signal sample, in gray. $t\bar{t}$ MC as all the other Monte-Carlo samples are purely used for illustration. All histograms have been normalized to unity.

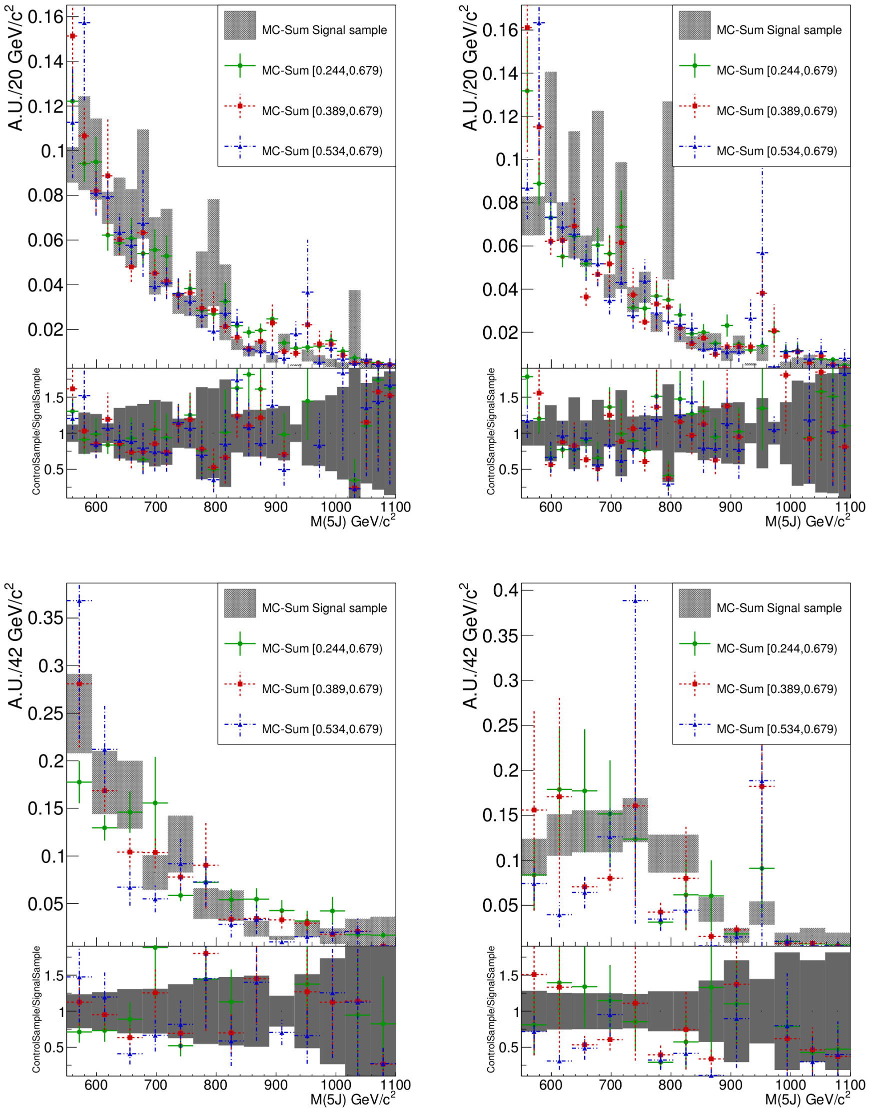


Figure 5.31: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done for the weighted sum of background Monte Carlo samples within 4 stages of selection: A [top left], B [top right], C [bottom left] and D [bottom right]. The 3 working points are given in different colors. Within statistical error, the 3 shapes for the control sample are in agreement at all stages with the signal sample, in gray. Monte-Carlo samples are purely used for illustration. All histograms have been normalized to unity. The errors of the sum of MC samples are underestimated due to the low statistics of the QCD samples.

reconstructed objects in both samples. In order to assess this statement, a χ^2-test (see [156]) has been used to evaluate the agreement of $M(5j)$ between control and signal sample scanning different values of the χ^2 cut, after applying some of the later selection criteria. At early stage of the selection, the $M(5j)$ shape from the control sample is not expected to agree to the signal sample as the background composition is different. With the χ^2-test, it is shown how the χ^2 cut closes up the phase space of both samples. This test evaluates the similarity of two histograms in terms of shape, delivering the χ^2/ndf of the comparison. If $\chi^2/ndf \sim 1$ the two histograms shape are compatible. Figure 5.32 shows that at early stages of the selection, the agreement is achieved for cut values much smaller than 100. The rest of the selection criteria also contribute to have a closer phase space between the signal and control sample, but their contribution is less important than the one performed by the cut over the χ^2.

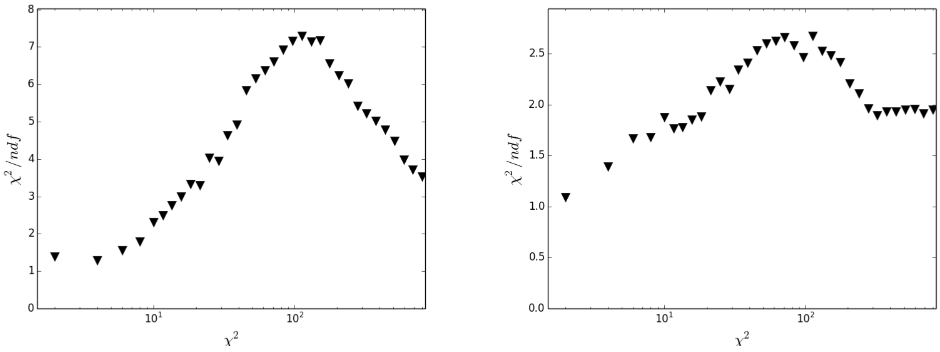


Figure 5.32: Distribution of the agreement between control sample and signal sample in data at early stage of the selection as a function of the χ^2 cut value. The y-axis represents the χ^2/ndf from a shape comparison made in data between control sample and analysis sample. The study is performed after requiring $\Delta R(bb) < 1.2$ [left] and up to $\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$ criterion [right] on top of basic selection and reconstruction. The control sample $M(5j)$ shape tend to agree with signal sample for lower values of χ^2 cut.

It is very important to have a control sample strongly dominated by backgrounds. A high contamination of signal in the control sample would mean considering signal events in the signal sample as background. As the control sample definition takes out the events with signal characteristics, the signal contamination is expected to be not high with respect to the background content of the control sample. To estimate this contamination the number of events for each signal mass point MC sample in the control sample were taken around the T' mass value in a 1-sigma window, and compared to the number of events from data in the control sample in the same windows after full selection. The procedure to determine these windows, used afterward to calculate the signal and background yields for the limit calculation, is explained in section 6.2. Table 5.12 presents the total number of events for each MC mass point in the control sample after full selection and for data. In figure 5.33 the M=700 GeV/c^2 signal MC sample and data after full selection in the control sample is shown. The highest expected contamination is from the M=750 GeV/c^2 and M=800 GeV/c^2 mass points, around 7%.

Events in control sample at mass point	Signal	Data	Contamination [%]
Signal M=600 GeV/c^2	3.37 ± 0.30	112 ± 10.58	3 ± 0.6
Signal M=650 GeV/c^2	4.50 ± 0.31	116 ± 10.77	4 ± 0.6
Signal M=700 GeV/c^2	5.85 ± 0.32	106 ± 10.30	6 ± 0.8
Signal M=750 GeV/c^2	5.40 ± 0.28	82 ± 9.06	7 ± 1.1
Signal M=800 GeV/c^2	5.30 ± 0.26	81 ± 9.00	7 ± 1.0
Signal M=850 GeV/c^2	4.22 ± 0.21	81 ± 9.00	5 ± 0.8
Signal M=900 GeV/c^2	3.89 ± 0.19	64 ± 8.00	6 ± 1.1
Signal M=950 GeV/c^2	3.56 ± 0.17	73 ± 8.54	5 ± 0.8
Signal M=1000 GeV/c^2	2.19 ± 0.12	46 ± 6.78	5 ± 1.0

Table 5.12: Number of events in the control sample for signal MC samples and data after full selection in a window corresponding to one σ for each mass point, as it will be explained in section 6.2. The contamination is evaluated as the ratio of the number of events in the control sample for each mass point and data.

Finally, an additional check on the working point dependence of the control sample was performed. The first validation procedure compared 3 different working points vetoing CSVM. This is, jets were considered to be b-tagged in the control sample for a discriminator between $[0.244, 0.679)$, $[0.389, 0.679)$ and $[0.534, 0.679)$, respectively. The second and third ranges are contained in the first one. The first validation procedure is then inclusive in terms of working points definition. However, this could lead to bias in the validation procedure as the success of the validation in the $[0.534, 0.679)$ range could drive the conclusion in the other cases. Then it is preferable to repeat the same validation for the exclusive cases: $[0.244, 0.389)$, $[0.389, 0.534)$ and $[0.534, 0.679)$. The exclusive validation can be found in figures 5.34 for $t\bar{t}$, in 5.35 for QCD, in 5.36 for MC background samples sum and in 5.37 for data.

This validation procedure, shows that the background shape of $M(5j)$ from the control sample can be used to estimate the background shape in the signal sample.

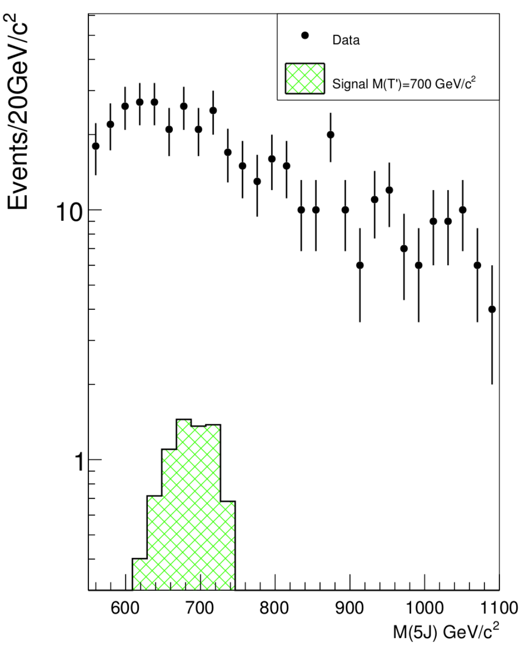


Figure 5.33: Signal contamination in the control sample, comparing 5-jets invariant mass between data and MC signal (M=700 GeV/c^2).

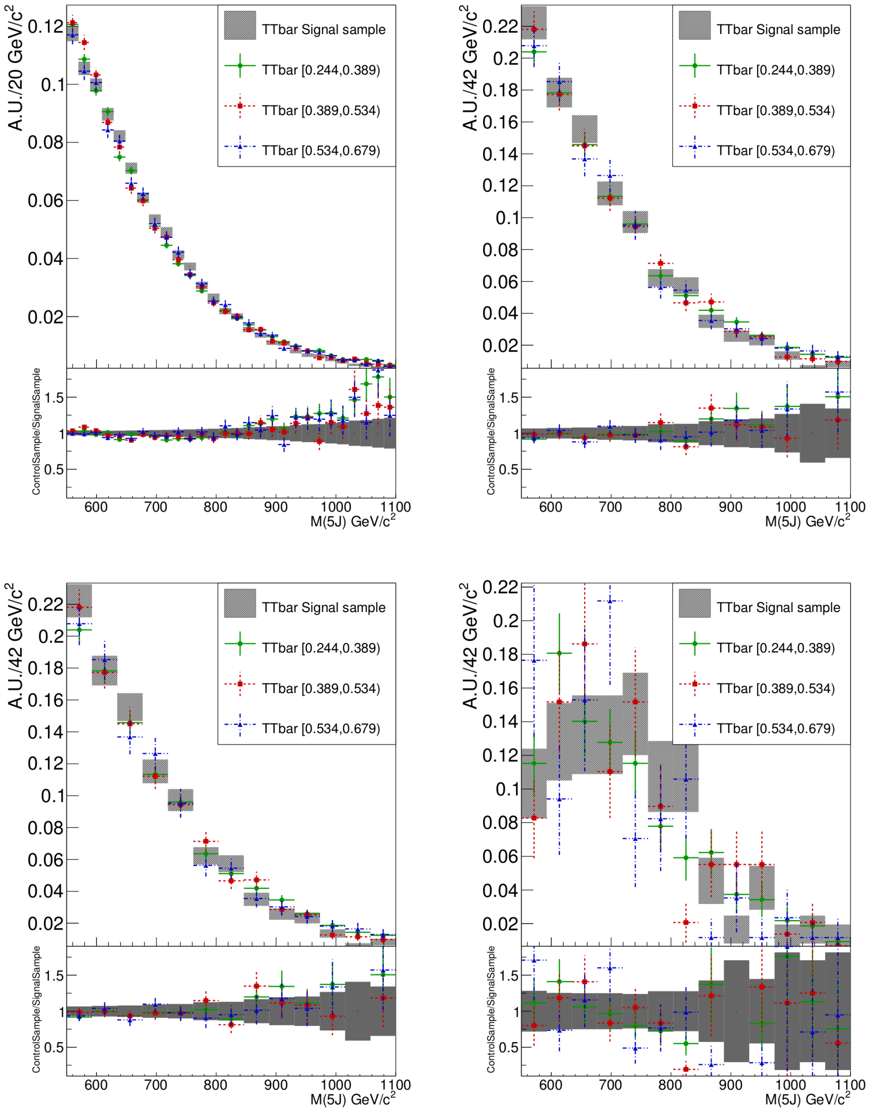


Figure 5.34: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done, in exclusive regions, for $t\bar{t}$ Monte Carlo samples within 4 stages of selection: A [top left], B [top right], C [bottom left] and D [bottom right]. The 3 working points are given in different colors. Within statistical error, the 3 shapes for the control sample are in agreement at all stages with the signal sample, in gray. $t\bar{t}$ MC as all the other Monte-Carlo samples are purely used for illustration. All histograms have been normalized to unity.

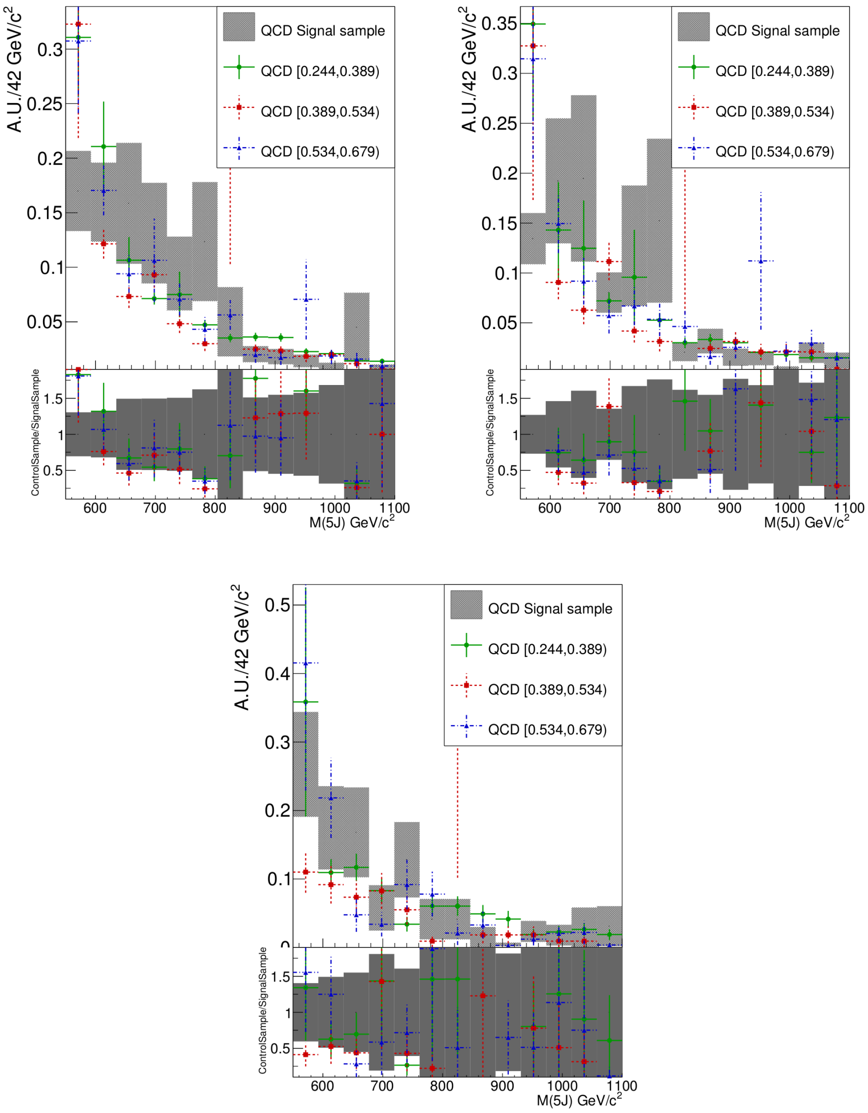


Figure 5.35: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done, in exclusive regions, for QCD Monte Carlo samples within 3 stages of selection: A [top left], B [top right] and C [bottom]. The 3 working points are given in different colors. The signal sample is displayed in gray. A lack of statistics is visible. All histograms have been normalized to unity. The errors of the QCD samples are underestimated due to their low statistics.

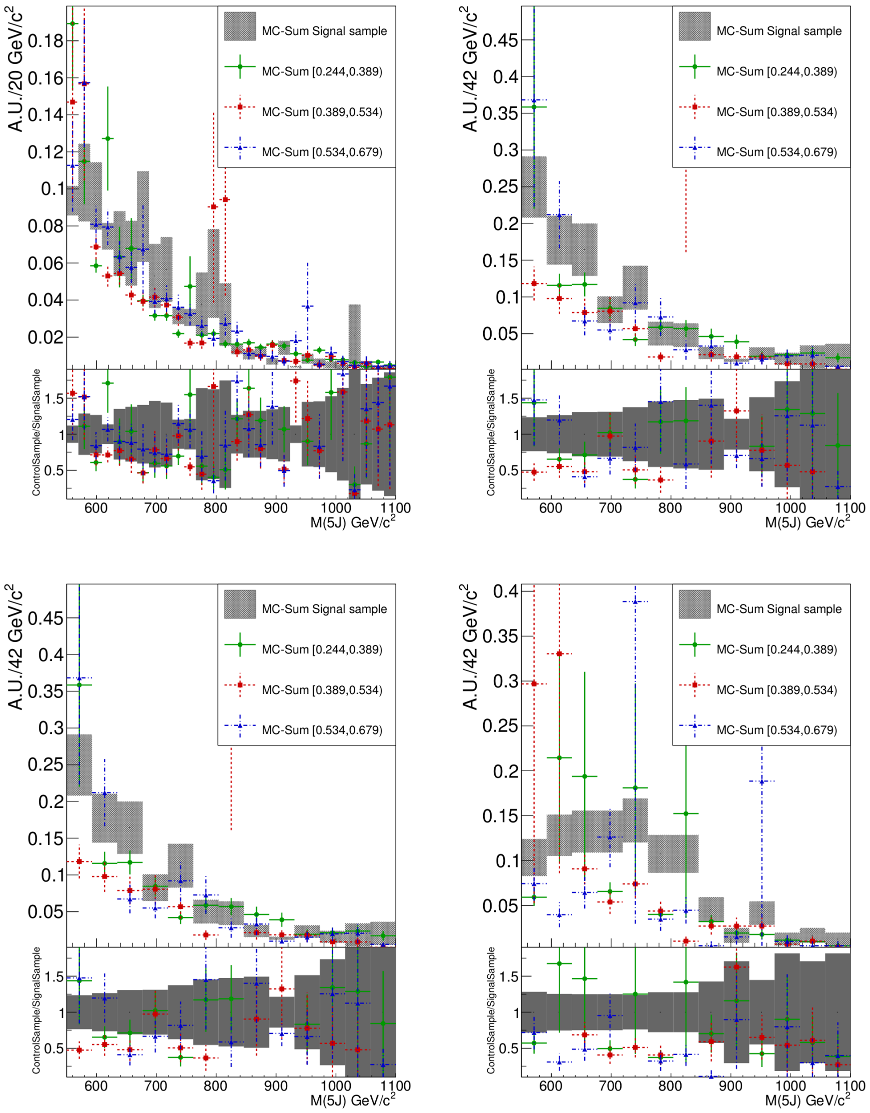


Figure 5.36: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done, in exclusive regions, for the weighted sum of background Monte Carlo samples within 4 stages of selection: A [top left], B [top right], C [bottom left] and D [bottom right]. The 3 working points are given in different colors. The signal sample is displayed in gray. A lack of statistics is visible. All histograms have been normalized to unity. The errors of the sum of MC samples are underestimated due to the low statistics of the QCD samples.

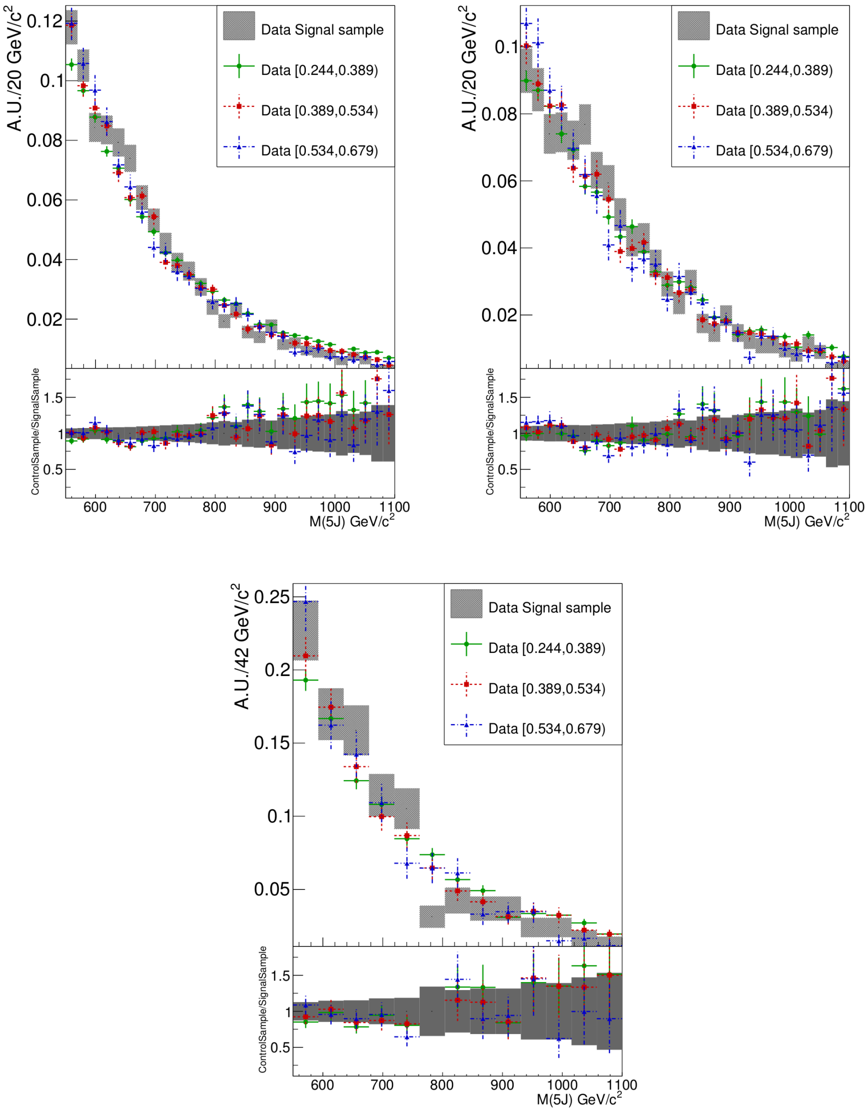


Figure 5.37: Comparison of the 5-jets invariant mass in the signal and control samples. In the control sample, different b-tagging working points are studied. This comparison is done, in exclusive regions, for data within 3 stages of selection: A [top left], B [top right] and C [bottom]. The 3 working points are given in different colors. Within statistical error, the 3 shapes for the control sample are in agreement at all stages with the signal sample, in gray. All histograms have been normalized to unity.

In this estimation all backgrounds are estimated together, even though they are composed mainly by two sources, $t\bar{t}$ and QCD. However, the normalization that need to be applied to this shape is unknown, and will be derived by other means.

5.4.2 Method for the estimation of the background normalization

To estimate the normalization of backgrounds in the signal sample, i.e. the number of background events in the signal sample, a sideband method has been used. The Higgs boson candidate mass was used for this purpose. The method consists in taking the ratio between the number of events inside N_{in}^{CS} and outside N_{out}^{CS} the cut window in the control sample. The cut window is the same used in the selection, 105 GeV/c^2 $< M(H_{cand}) <$ 145 GeV/c^2 (see figure 5.18). The ratio between them is defined as $R^{CS} = N_{in}^{CS}/N_{out}^{CS}$. This ratio describes the proportion of events inside the window with respect to the number of events outside in the control sample. The same ratio can be also defined in the signal sample, $R^{SS} = N_{in}^{SS}/N_{out}^{SS}$. However, in the signal sample it might be signal events inside the window. The hypothesis is that if both ratios are compatible in the signal and control sample, at selection stages where backgrounds are dominant, the ratio from the control sample R^{CS} can be used to predict the number of background events inside the Higgs candidate mass in the signal sample. Thus, the estimated number of background events in the signal sample is:

$$N_{in}^{SS_{BKG}} = R^{CS} N_{out}^{SS} \tag{5.7}$$

Validation

To test the validity of the method, equation 5.7 has been used at each stage of the selection to compare the prediction with the real content inside the cut window in the signal sample. This test can not be done after full selection, when a significant signal component in the cut window could be present. However, as the method is used to predict the background content, it can be validated at earlier stages of the selection, where backgrounds are dominant and signal is negligible. This is, to validate the method it has to be proven that $R^{CS} = R^{SS}$ and $N_{in}^{SS_{BKG}} = N_{in}^{SS}$ at earlier stages of the selection.

For this validation procedure, the Higgs candidate mass cut has been omitted from the selection and the other cuts have been applied in the same order to check that $N_{in}^{SS_{BKG}} = N_{in}^{SS}$ and $R^{CS} = R^{SS}$. In table 5.13 the results of the procedure are shown. At each step of the selection $N_{in}^{SS_{BKG}} = N_{in}^{SS}$ and $R^{CS} = R^{SS}$, within errors. The last line of the table corresponds to full selection, stage at which the validation is not performed. However, from this line the normalization of backgrounds is obtained from the method: $N_{in}^{SS_{BKG}} = 52.96 \pm 17.95$. The validation is not performed also after $\Delta R(T', j^6) > 4.8$ because the signal contamination is around 10% inside the Higgs candidate mass window in the signal sample.

Cut	N_{in}^{CS}	N_{out}^{CS}	N_{in}^{SS}	N_{out}^{SS}	χ^2/ndf
$\chi^2 < 8$	163589 ± 404.46	49112 ± 221.61	8071 ± 89.84	2510 ± 50.10	0.99
$\Delta R(bb) < 1.2$	35266 ± 187.79	13247 ± 115.10	2820 ± 53.10	1054 ± 32.47	2.20
$\frac{M(top_{cand}^{2nd})+M(W_{cand}^{2nd})}{M(H_{cand})} > 6.8$	19269 ± 138.81	8001 ± 89.45	1242 ± 35.24	528 ± 22.98	1.36
$\Delta R(T', j^6) > 4.8$	1566 ± 39.75	636 ± 25.22	—	40 ± 6.32	—
$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	519 ± 22.78	196 ± 14.00	—	20 ± 4.47	—

Cut	R^{CS}	R^{SS}	$N_{in}^{SS_{BKG}}$
$\chi^2 < 8$	3.33 ± 0.02	3.22 ± 0.10	8360.65 ± 225.28
$\Delta R(bb) < 1.2$	2.66 ± 0.04	2.68 ± 0.13	2805.95 ± 125.75
$\frac{M(top_{cand}^{2nd})+M(W_{cand}^{2nd})}{M(H_{cand})} > 6.8$	2.41 ± 0.04	2.35 ± 0.17	1271.60 ± 78.72
$\Delta R(T', j^6) > 4.8$	2.46 ± 0.16	—	98.49 ± 21.97
$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	2.65 ± 0.31	—	52.96 ± 17.95

Table 5.13: Results of the validation procedure of the normalization estimation. The validation is started after the χ^2 cut. All other cuts are applied progressively. The R^{SS} and N_{in}^{SS} have been blinded in the last two lines for the validation procedure, because the amount of signal events in the signal sample is not negligible with respect to background events. In addition, the Higgs candidate mass distribution in the control sample is compared to its distribution in the signal sample via a χ^2 test, the results of this test are shown in the table. They show the shape agreement between both samples.

As an additional check on the validity of the method, table 5.14 shows the number of MC signal events in the control and signal samples at each stage of the selection. It is important to control that the signal contamination is not high. In this case, the contamination, for the normalization estimation, is negligible. After the full selection, the maximal signal contamination in the control sample is around 1%. Finally, figure 5.38 shows the comparison of the Higgs candidate mass from the data in the control and signal samples after the $\chi^2 < 8$ and the $\frac{M(top_{cand}^{2nd})+M(W_{cand}^{2nd})}{M(H_{cand})} > 6.8$ criteria, separately. In both figures, the shapes of the distributions are in close agreement as it was checked by the χ^2-test (in table 5.13).

Cut	Signal (M=700GeV/c^2) N_{in}^{CS}	Signal (M=700GeV/c^2) N_{out}^{CS}	Signal (M=700GeV/c^2) N_{in}^{SS}	Signal (M=700GeV/c^2) N_{out}^{SS}
$\chi^2 < 8$	34.91±0.78	5.83±0.32	34.90±0.78	4.59±0.28
$\Delta R(bb) < 1.2$	28.35±0.70	4.12±0.27	32.05±0.75	3.95±0.26
$\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$	17.85±0.56	2.79±0.22	20.19±0.59	2.53±0.21
$\Delta R(T', j^6) > 4.8$	9.27±0.40	1.15±0.14	11.27±0.44	1.36±0.15
$\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$	7.77±0.37	1.01±0.13	10.01±0.42	1.24±0.15

Table 5.14: Signal MC events for the 700 GeV/c^2 mass point in the control and signal sample at each stage of the selection used for the estimation of the background normalization method. The maximal contamination in the control sample, after the $\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T} > 0.67$ criterion, is around 1%. The maximal contamination in the signal sample, after the $\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$ cut, is around 1% as well. The signal contamination in the control sample has been evaluated as $N_{in}^{CS}(Signal)/N_{in}^{CS}(Data)$, and in the signal sample as $N_{in}^{SS}(Signal)/N_{in}^{SS}(Data)$.

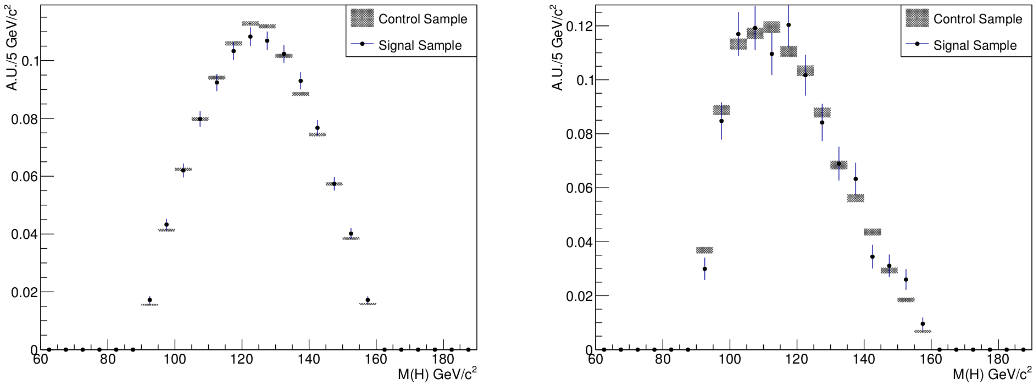


Figure 5.38: Higgs candidate mass distribution for the data in the signal and control sample after the $\chi^2 < 8$ cut [left] and after the $\frac{M(top^{2nd}_{cand})+M(W^{2nd}_{cand})}{M(H_{cand})} > 6.8$ criterion [right].

6 Signal extraction

In this chapter are going to be described the techniques to extract the number of expected events for signal and background after applying the selection. Also the calculation of systematic uncertainties is performed. The final statistical interpretation of results is based on a cut-and-count method. The signal and background events after full selection will be extracted for each mass point analyzed. Nine T' mass points were used between 600 GeV/c^2 and 1000 GeV/c^{2}in 50 GeV/c^2 steps.

The following concepts will be used in this chapter:

- **Signal yield**: Number of events from the integration of 5-jets invariant mass in a window around the mean value from each MC sample mass point (Y).
- **Background yield**: Number of events from the integration of 5-jets invariant mass in a window around the mean value defined by each MC sample mass point from the data-driven background estimation.
- **Variation of a yield from a systematic uncertainty**: When applying the 1-sigma variations of a systematic uncertainty a new signal or background yield (Y') is obtained. The variation obtained on the yield, presented in %, is calculated as $|Y' - Y|/Y$. In the summary tables the absolute value of the variation is presented, while in the figures also the sign is included.

6.1 Definition of integration mean and window values

For the calculation of the signal yields the integral over 1-sigma around the mean value of each mass point has been used. In figure 6.1 the results from the gaussian fit applied to each mass point for the mean value and standard deviation are shown. To be less sensitive to statistical fluctuations, linear fits of these distributions have been performed, and the parametric dependence of mean and width as a function of the mass as reference has been used. The values are presented in table 6.1 and they were used for the computation of yields and variation on yields from systematic uncertainties. In the same table are shown the statistical uncertainties of signal yields in percentage. In the calculation of systematics for signal a conservative approach will be followed. If a source of systematic uncertainty is found to be smaller than the statistical uncertainties of the yield, the latter will be taken as the value of the systematic uncertainty.

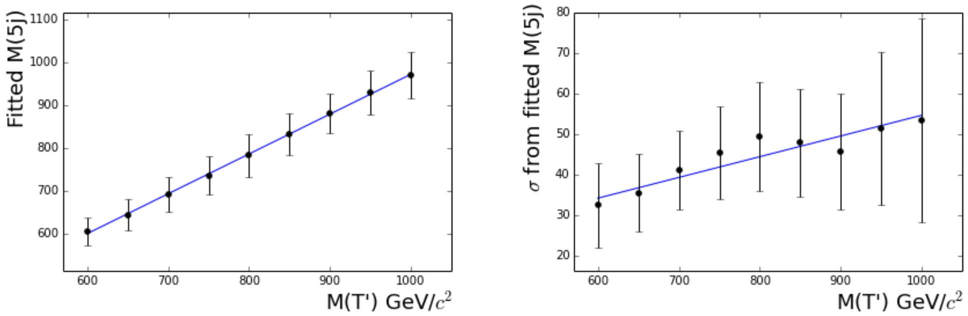


Figure 6.1: Gaussian fit results for each MC mass point, in terms of mean value and standard deviation. The errors shown for the mean values are the corresponding standard deviations from the gaussian fit. The error of the standard deviation itself is coming from the gaussian fit, related to the statistics of the samples. A linear fit of these results has been applied, represented by the blue line in each plot.

6.2 Systematics

In order to compute exclusion limits, the systematic uncertainties of the analysis should be evaluated. Two sets of sources have been identified, related either to signal or backgrounds. As our signal is modeled with the help of MC samples, the following sources of uncertainties for MC simulation have been considered: Jet Energy Scale, b-tagging, PDF+α_S, pileup reweighting, trigger and luminosity. As backgrounds were derived from data, no systematic uncertainties from MC corrections are considered. For background estimation only systematic uncertainties related to the data-driven methods (shape and normalization) were considered. Finally, the luminosity uncertainty considered is 2.6% for signal and background yields.

The first source of systematic uncertainty considered is related to the scale factors determination for b-tagging. For each scale factor, there is an associated error, depending on the jet flavor. The scale factors for b and c jets were treated fully correlated while they are considered uncorrelated with the light jets ones. In addition, the error for c-jets scale factors is considered to be the double of the one associated to b-jets. The scale factors for b and c-jets were varied simultaneously within their errors and signal yields were re-computed. The procedure was

Results from linear fit				
Sample Name	Mass (GeV/c^2)	Width (GeV/c^2)	Yield	Error on yield (%)
$Tj \rightarrow tHj$ 600 GeV/c^2	600.70	34.21	4.39 ± 0.35	7.89%
$Tj \rightarrow tHj$ 650 GeV/c^2	647.07	36.77	6.64 ± 0.38	5.72%
$Tj \rightarrow tHj$ 700 GeV/c^2	693.44	39.33	7.84 ± 0.37	4.72%
$Tj \rightarrow tHj$ 750 GeV/c^2	739.81	41.89	7.24 ± 0.32	4.46%
$Tj \rightarrow tHj$ 800 GeV/c^2	786.18	44.46	6.06 ± 0.28	4.56%
$Tj \rightarrow tHj$ 850 GeV/c^2	832.55	47.02	5.65 ± 0.24	4.26%
$Tj \rightarrow tHj$ 900 GeV/c^2	878.92	49.58	4.78 ± 0.21	4.30%
$Tj \rightarrow tHj$ 950 GeV/c^2	925.29	52.14	3.80 ± 0.17	4.56%
$Tj \rightarrow tHj$ 1000 GeV/c^2	971.67	54.70	2.28 ± 0.12	5.37%

Table 6.1: Mean value and standard deviation from the linear fit, done using gaussian fit of MC mass points after the full selection. In the last column the yields for each mass point are presented from the integration around one sigma of the mean value from the linear fit. The error of the yields are presented as a percentage of the nominal value.

repeated for light jets independently. In table 6.2, the obtained uncertainty of the yields from the up ($+1\sigma$) and down (-1σ) with respect to the nominal value is displayed in %. Variations for b and c-jets give an uncertainty of around 6% while for light jets only 1%. The results for b and c-jets and an overall systematic of 1% for light flavor were considered to derive a total b-tagging systematic uncertainty. In addition figure 6.2 shows the full uncertainties from b-tagging. The signal yields after applying 1-sigma variations on the uncertainties of b-tagging scale factors are compatible, within errors, with nominal signal yields. Accordingly to the conservative approach proposed, the highest error between the b-tagging systematic uncertainty and the statistical error (quoted in table 6.1) is taken as the b-tagging uncertainty for limits calculation.

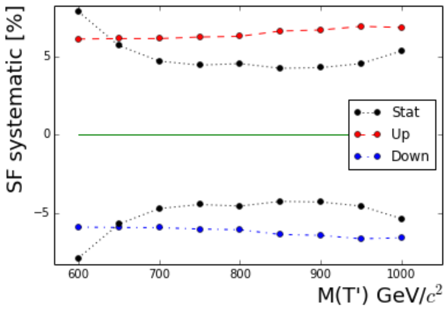


Figure 6.2: Total b-tagging uncertainties for signal yields.

Sample Name	b or c quark		Light flavours	
	up [%]	down [%]	up [%]	down [%]
$Tj \to tHj$ 600 GeV/c^2	6.05	5.82	0.89	0.89
$Tj \to tHj$ 650 GeV/c^2	6.08	5.85	0.51	0.51
$Tj \to tHj$ 700 GeV/c^2	6.08	5.86	0.48	0.49
$Tj \to tHj$ 750 GeV/c^2	6.18	5.95	0.66	0.66
$Tj \to tHj$ 800 GeV/c^2	6.23	5.99	0.35	0.38
$Tj \to tHj$ 850 GeV/c^2	6.57	6.30	0.88	0.88
$Tj \to tHj$ 900 GeV/c^2	6.63	6.36	0.52	0.51
$Tj \to tHj$ 950 GeV/c^2	6.87	6.58	0.88	0.89
$Tj \to tHj$ 1000 GeV/c^2	6.80	6.51	0.55	0.54

Table 6.2: B-tagging uncertainties for signal yields.

The second source of systematic uncertainty considered is the jet energy scale determination. In MC, this scale is corrected via the Jet Energy Corrections (JEC) to match the data. These corrections were varied by $\pm 1\sigma$, with sigma the error of the corrections, and the signal yields were recomputed. JEC are composed of Jet Energy Scale and Jet Energy Resolution corrections. Each of them were treated as uncorrelated, and thus varied independently. The systematic values are presented in table 6.3. In addition, figure 6.3 shows the full uncertainties from JEC. The mean value of JES systematic is 6%, while for JER is 1.6%. For some mass points the JEC systematic uncertainty is fully contained in the statistical error of yields (from table 6.1), thus the highest uncertainty between the systematic uncertainty and the statistical error is taken as the JEC uncertainty for limits calculation.

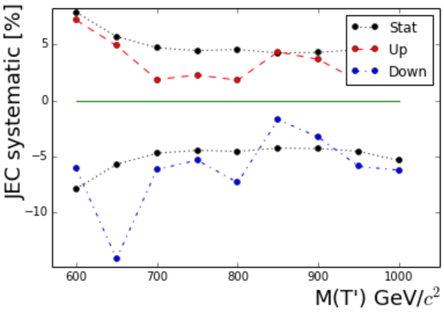


Figure 6.3: Total JEC uncertainties for signal yields.

Sample Name	JER		JES	
	up [%]	down [%]	up [%]	down [%]
$Tj \rightarrow tHj$ 600 GeV/c^2	2.87	1.92	7.02	5.90
$Tj \rightarrow tHj$ 650 GeV/c^2	0.43	1.29	4.66	14.09
$Tj \rightarrow tHj$ 700 GeV/c^2	2.61	1.26	1.05	5.98
$Tj \rightarrow tHj$ 750 GeV/c^2	1.94	1.58	1.67	5.12
$Tj \rightarrow tHj$ 800 GeV/c^2	1.15	1.73	0.94	7.18
$Tj \rightarrow tHj$ 850 GeV/c^2	0.40	2.27	4.06	0.74
$Tj \rightarrow tHj$ 900 GeV/c^2	1.53	1.72	3.33	2.89
$Tj \rightarrow tHj$ 950 GeV/c^2	0.74	0.32	0.77	5.69
$Tj \rightarrow tHj$ 1000 GeV/c^2	2.31	1.99	0.67	6.04

Table 6.3: JEC uncertainties for signal yields.

Another source of uncertainties for MC signal samples is related to the PDF and α_S used in the simulation. MC signal samples were produced with the PDF set CTEQ6.6 with the best fit values among PDF members. The possible variations from all 44 members of CTEQ6.6 to obtain an overall systematic effect on signal yields have been considered. In addition two additional PDF sets have been considered: MSTW2008 and NNPDF2.0. For MSTW2008 40 members and for NNPDF2.0 50 members were added in quadrature. For the α_s, 1-σ variations were considered around the central value for each PDF set: $\alpha_s(M_Z) = 0.118$ for CTEQ6.6, $\alpha_s(M_Z) = 0.12018$ for MSTW2008 and $\alpha_s(M_Z) = 0.119$ for NNPDF2.0.

In principle, to calculate the systematic variations when changing the PDF set used in the MC simulation, it should be required to regenerate the MC simulation with each PDF member. Such a procedure is inefficient and requires a high availability of computing resources and time. Therefore, to calculate the observable with the varied members of the PDF a reweighting method was used. A weight was calculated for each member j, in equation 6.1 where $x_{1,2}$ is the fraction of the momentum carried by the two initial partons entering the hard interaction, $f_{1,2}$ is their corresponding pdg code, and Q is the scale. The weights used to calculate the yields and then the systematics for the cases of CTEQ6.6 and MSTW2008 were given by equation 6.2 while for NNPDF2.0 were given by equation 6.3. In these equations, $\mathcal{O}^j$ is the observable evaluated for the member j, being $j = 0$ for the nominal value, and $C = 1.64485$ for CTEQ6.6 to correctly obtain 1-σ variation, and $C = 1$ for MSTW2008.

$$w^j = \frac{PDF^j(x_1, f_1, Q)PDF^j(x_2, f_2, Q)}{PDF^0(x_1, f_1, Q)PDF^0(x_2, f_2, Q)} \tag{6.1}$$

$$\begin{aligned}\sigma_{up} &= \frac{1}{C}\sqrt{\sum_{i=1}^{N_{vectors}/2} (max\{\mathcal{O}^{2i-1} - \mathcal{O}^0, \mathcal{O}^{2i} - \mathcal{O}^0, 0\})^2} \\ \sigma_{down} &= \frac{1}{C}\sqrt{\sum_{i=1}^{N_{vectors}/2} (max\{\mathcal{O}^0 - \mathcal{O}^{2i-1}, \mathcal{O}^0 - \mathcal{O}^{2i}, 0\})^2}\end{aligned} \tag{6.2}$$

$$\sigma = \sqrt{\frac{1}{N_{replicas} - 1} \sum_{i=1}^{N_{replicas}} (\mathcal{O}^i - \mathcal{O}^0)^2} \tag{6.3}$$

The results of the calculation of the systematic uncertainties for the PDF's can be found in table 6.4. The full systematic for PDF+α_s were considered as the highest variation between the three considered. This overall systematic can be found in figure 6.4. The overall systematic effect of PDF variations is around 3%. The PDF systematics are fully contained in the statistical error of yields. Consequently, the PDF uncertainty is taken to be the same value as the statistical error for each mass point.

Sample Name	CTEQ6.6		MSTW2008		NNPDF2.0
	up [%]	down [%]	up [%]	down [%]	up,down [%]
$Tj \to tHj$ 600 GeV/c^2	2.54	1.76	3.09	2.15	2.37
$Tj \to tHj$ 650 GeV/c^2	2.51	1.75	3.23	2.21	2.25
$Tj \to tHj$ 700 GeV/c^2	2.69	1.84	3.25	2.23	2.82
$Tj \to tHj$ 750 GeV/c^2	2.77	1.88	3.10	2.15	2.77
$Tj \to tHj$ 800 GeV/c^2	2.83	1.97	3.16	2.20	2.58
$Tj \to tHj$ 850 GeV/c^2	2.91	1.94	3.16	2.19	2.64
$Tj \to tHj$ 900 GeV/c^2	3.02	2.00	3.27	2.24	2.25
$Tj \to tHj$ 950 GeV/c^2	3.38	2.18	3.38	2.38	2.74
$Tj \to tHj$ 1000 GeV/c^2	3.19	2.08	3.30	2.29	3.07

Table 6.4: PDF+α_s uncertainties for signal yields.

In addition, the pileup reweighting MC correction procedure was considered as a possible source of systematic uncertainty. To estimate the size of the uncertainty the minimum bias inelastic cross section, 69.4 mb for 2012 , was varied by 5% (up and down). For the up and down variations, the weights to correct pileup

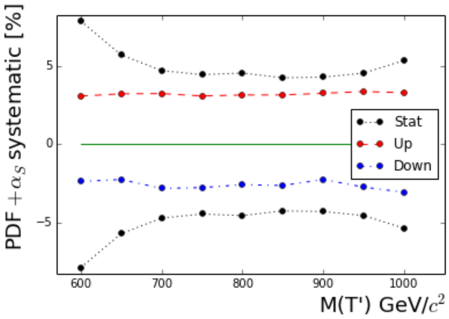


Figure 6.4: Overall PDF+α_s uncertainties for signal yields.

in MC simulation were recalculated, and with the varied weights the yields were re-computed. The effect of these variations is shown in table 6.5 and in figure 6.5. As the systematic uncertainties found are fully contained in the statistical errors of the yields (from table 6.1), this statistical error is considered as the maximal uncertainty for pileup and is used for limits calculation.

Sample Name	Pileup	
	up [%]	down [%]
$Tj \to tHj$ 600 GeV/c^2	1.36	1.01
$Tj \to tHj$ 650 GeV/c^2	1.54	1.78
$Tj \to tHj$ 700 GeV/c^2	3.05	3.37
$Tj \to tHj$ 750 GeV/c^2	2.67	2.76
$Tj \to tHj$ 800 GeV/c^2	1.18	1.26
$Tj \to tHj$ 850 GeV/c^2	2.25	2.18
$Tj \to tHj$ 900 GeV/c^2	1.60	1.66
$Tj \to tHj$ 950 GeV/c^2	2.23	2.33
$Tj \to tHj$ 1000 GeV/c^2	3.08	3.76

Table 6.5: Pileup uncertainties for signal yields.

Finally, the observed differences between $t\bar{t}$ MC samples and the data in the trigger efficiency in section 5.3.5 are considered as a possible source of systematic uncertainty. The ratio between $t\bar{t}$ and data has been used to estimate such uncertainty on the signal yields as a function of the 6th jet p_T. After applying this ratio to the signal MC samples, the yields have been recomputed and compared to the nominal yields. The difference between both is quoted in table 6.6 as the systematic uncertainty associated to the trigger. As for pileup uncertainties, the trigger systematics are fully contained in the statistical error and then the statistical error

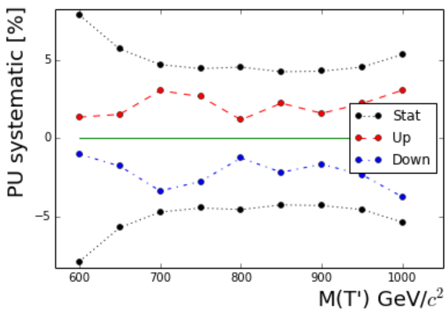


Figure 6.5: Total pileup uncertainties for signal yields.

will be used as the trigger systematic for the limits calculation.

Sample Name	Trigger
	up/down [%]
$Tj \rightarrow tHj$ 600 GeV/c^2	2.51
$Tj \rightarrow tHj$ 650 GeV/c^2	3.31
$Tj \rightarrow tHj$ 700 GeV/c^2	4.02
$Tj \rightarrow tHj$ 750 GeV/c^2	3.72
$Tj \rightarrow tHj$ 800 GeV/c^2	3.79
$Tj \rightarrow tHj$ 850 GeV/c^2	3.46
$Tj \rightarrow tHj$ 900 GeV/c^2	3.13
$Tj \rightarrow tHj$ 950 GeV/c^2	3.40
$Tj \rightarrow tHj$ 1000 GeV/c^2	3.60

Table 6.6: Trigger uncertainties for signal yields.

As a summary, table 6.7 shows the systematic uncertainties for each MC mass point. The dominant systematics are coming from JEC and b-tagging sources. The different sources of systematic errors are considered to be uncorrelated, and consequently summed in quadrature.

For background, the systematics are associated to the estimation methods of the shape and normalization. In the case of the shape, the dispersion of the ratio between signal and control sample on data are used for the determination of the uncertainty. After the full selection, there is no sufficient statistics to determine a systematic so an early stage in the selection is chosen. The dispersion measurement is done after requiring $\Delta R(bb) < 1.2$ in the control sample (see top right plot of figure 5.28). In figure 6.6 the dispersion of the ratio plot is shown. This dispersion is fitted with a gaussian function. The total systematic uncertainty is then taken

Sample Name	b-tagging		JEC		PDF+α_S	Pileup	Trigger
	up [%]	down [%]	up [%]	down [%]	up/down [%]	up/down [%]	up/down [%]
$Tj \to tHj$ 600 GeV/c^2	7.89	7.89	7.89	7.89	7.89	7.89	7.89
$Tj \to tHj$ 650 GeV/c^2	6.16	5.93	5.72	14.17	5.72	5.72	5.72
$Tj \to tHj$ 700 GeV/c^2	6.16	5.94	4.72	6.18	4.72	4.72	4.72
$Tj \to tHj$ 750 GeV/c^2	6.26	6.03	4.46	5.35	4.46	4.46	4.46
$Tj \to tHj$ 800 GeV/c^2	6.31	6.08	4.56	7.34	4.56	4.56	4.56
$Tj \to tHj$ 850 GeV/c^2	6.64	6.38	4.35	4.26	4.26	4.26	4.26
$Tj \to tHj$ 900 GeV/c^2	6.71	6.44	4.30	4.30	4.30	4.30	4.30
$Tj \to tHj$ 950 GeV/c^2	6.95	6.65	4.56	5.90	4.56	4.56	4.56
$Tj \to tHj$ 1000 GeV/c^2	6.87	6.59	5.37	6.24	5.37	5.37	5.37

Table 6.7: Summary of the uncertainties for signal yields. In addition, all mass points have 2.6% uncertainty from luminosity measurement.

from the standard deviation of the fit. The dispersion corresponds to 20% around the nominal value. In addition, a linear fit of the ratio between the control and signal samples was performed at the same selection stage to check if there was a possible trend. The fit found a positive slope of $(1.27 \pm 1.18) \times 10^{-3}$, which is very close to zero. Therefore, no further corrections due to a trend were considered.

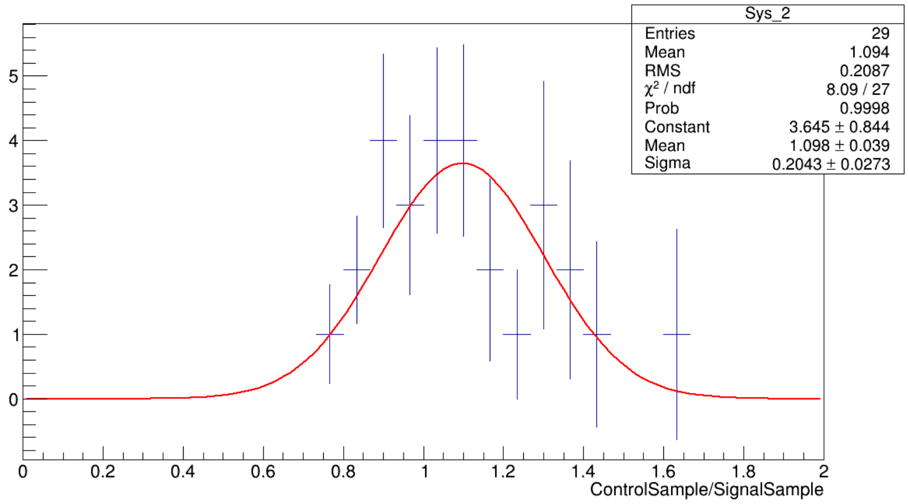


Figure 6.6: Dispersion of the ratio between the control and signal samples for shape data-driven estimation of background shape. A dispersion of 20% is observed after the $\Delta R(bb)$ cut. This figure corresponds to the projection on the y-axis of the ratio between signal and control samples shown in top right plot of figure 5.28.

In addition, for the background normalization estimation, from table 5.13 after the full selection, the normalization is estimated to be 52.96 ± 17.95. The error represents 34% of the central value. This value is taken as systematic uncertainty for the background normalization method. This systematic is dominant with respect to the rest of the systematics for signal and background. Summing in quadrature the different sources of systematic uncertainties for the background from the estimation methods, the total uncertainty on background is 39%. Also, it was shown in the last section that the highest signal contamination expected is of 7% for 750 GeV/c^2 and 800 GeV/c^2mass points. Consequently, any further effect from signal contamination is neglected.

Finally, Table 6.8 shows the systematics for backgrounds and signal MC sample with M=700 GeV/c^2. The dominant systematic source is coming from the normalization estimation method for backgrounds.

Systematics Name	Signal	Background
PDF	4.72%	–
Luminosity	2.6%	–
Trigger	4.72%	–
B-tag	+6.16% / -5.94%	–
JEC	+4.72% / - 6.18%	–
Pileup	4.72%	–
Background Shape determination	–	20%
Background Normalization	–	34%

Table 6.8: Summary of the uncertainties in the case of the signal yield for the mass point at 700 GeV/c^2 and for background yields.

6.3 Results

Using the definitions introduced in section 6.1, the background yields and the number of observed events in data are calculated and quoted in table 6.9.

Figure 6.7 shows two signal MC mass points compared to observed data with subtraction of the estimated background. Additionally, figure 6.8 presents separately the $M(5j)$ distribution after full selection observed in data, the estimated background, and the same signal MC samples as in previous figure.

As there is no evidence of an excess, the yields and systematics given in last section were used to compute expected and observed exclusion limits. The limits were calculated using Theta package [157] using asymptotic CL_S limits. A simple counting experiment was considered. For the Theta package a model is build for

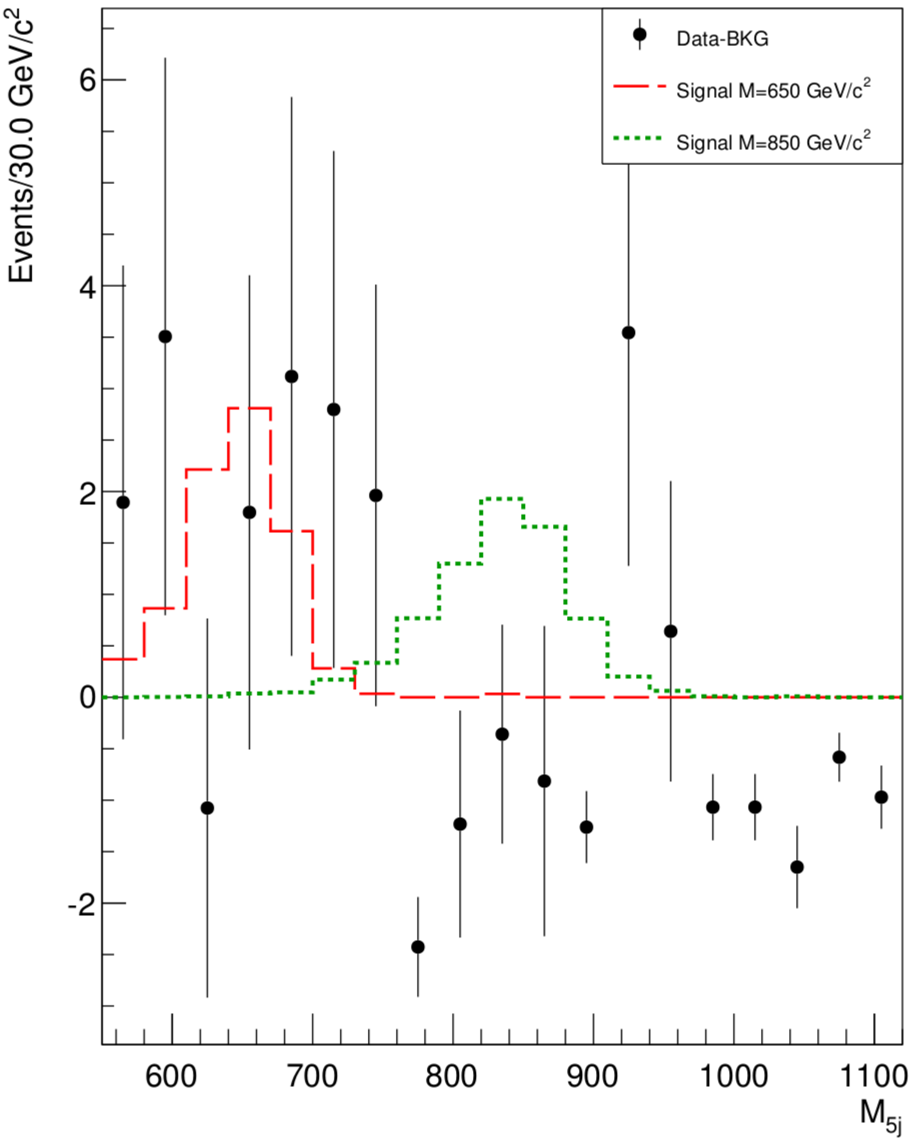


Figure 6.7: $M(5j)$ after the full selection for data minus estimated background and signal mass points at 650 GeV/c^2 and 850 GeV/c^2.

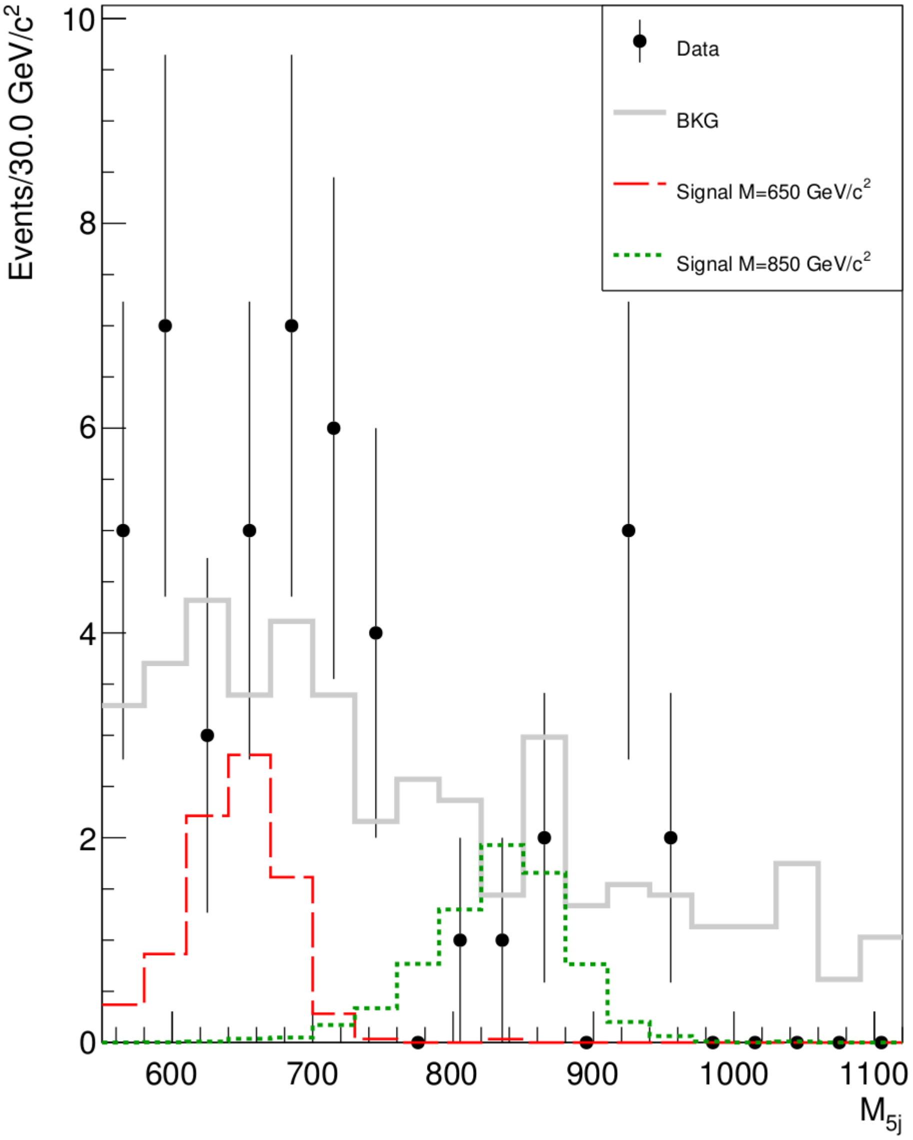


Figure 6.8: $M(5j)$ after the full selection for data and the estimated background and signal mass points at 650 GeV/c^2 and 850 GeV/c^2.

T' Mass GeV/c^2	Signal	Background	Observed Data
600	4.39±0.35	11.31	15
650	6.64±0.38	11.83	15
700	7.84±0.37	13.06	22
750	7.24±0.32	12.24	17
800	6.06±0.28	8.54	6
850	5.65±0.24	9.36	4
900	4.78±0.21	7.30	8
950	3.80±0.17	8.43	9
1000	2.28±0.12	5.24	7

Table 6.9: Expected number of events for the signal, estimated background and observed data after the full selection in 1-σ integration window. The errors for signal yields represent only the statistical uncertainty.

each mass point given by the relation $\lambda(\beta, \theta) = \beta S(\theta) + B(\theta)$, where β designs the signal strength (called also μ in the literature), S the signal events expected from the mass point and B the background events from the estimation performed in the analysis. In this model both signal and background can have associated nuisance parameters θ linked to the different sources of systematic uncertainties. Corresponding to a counting experiment a likelihood is built $L(\beta, \theta) = \frac{\lambda(\beta,\theta)^n}{n!} e^{-\lambda(\beta,\theta)}$, where n is the number of observed events. Two specific likelihoods are then defined, $L_{s+b} = L(1, \theta)$ corresponding to the positive signal hypothesis and $L_b = L(0, \theta)$ corresponding to the null signal hypothesis. To determine upper limits, a test statistics is defined $q = -2ln(L_{s+b}/L_b)$. Then, the probability distribution function can be constructed from simulation for each hypothesis as a function of the test statistics q. These probability distribution functions are denoted as $f(q|s+b)$ for the positive signal hypothesis, and $f(q|b)$ for the null signal hypothesis. An illustration of the procedure that has been described is shown in figure 6.9, described in [158].

The p-value of a given q is defined as the area to the right of the probability distribution function from the specified q. Thus, the upper limit on β at a confidence level $CL = 1 - \alpha$ is the value of β for which the p-value is α. Accordingly, the CL_{s+b} and CL_b are defined with respect to the corresponding hypothesis. In addition, a confidence level for signal only can be defined as $CL_s = CL_{s+b}/CL_b$. With respect to this later definition, a potential signal is expected to be excluded at 95% confidence level when the associated signal strength β gives $CL_s = 0.05$. There are also 1-σ and 2-σ error bands associated to $CL_s = 0.05$ limit. Such limit and bands constitute the expected bands of exclusion for the signal strength, than

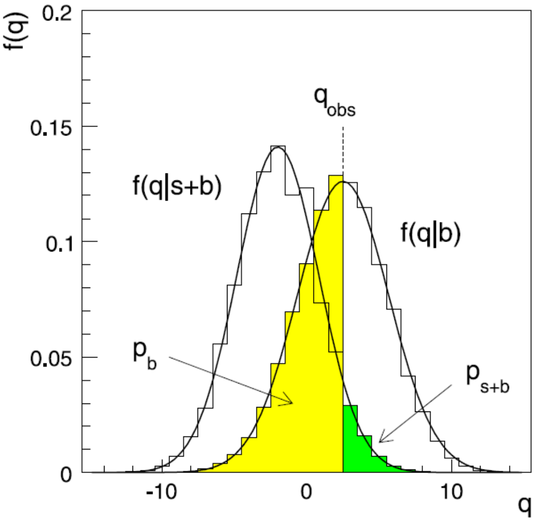


Figure 6.9: The distribution of the test statistics $q = -2ln(L_{s+b}/L_b)$ under the hypothesis of $\beta = 0$ (null signal hypothesis) and $\beta = 1$ (positive signal hypothesis) [158]. q_{obs} is the observed value of q in data, p_{s+b} and p_b the probabilities the observation come from the positive signal and null signal hypotheses, correspondingly.

can be also interpreted as the excluded signal cross section $\sigma_{excluded} \geq \beta\sigma_s$ (with σ_s the signal cross section).

Finally, using the number of observed events in data to obtain a q_{obs}, an observed signal strength can be calculated for which $CL_s = 0.05$. If the observed signal strength is higher than 2-σ on top of the expected excluded signal strength, it is said that a signal could not be excluded in data, what requires other studies to determine the significance of the potential signal. On the contrary, if the observed signal strength is contained in the exclusion bands, the associated cross section is said to be excluded.

In figure 6.10 the expected and observed exclusion limits in terms of the cross section as a function of $M(5j)$ are shown. The theoretical prediction is represented by the red line. The theoretical cross sections have been calculated with MadGraph 5 [116, 117] utilizing the FeynRules [118] model implementation developed in [58, 59]. Comparing this theoretical prediction to expected limits, there is no expectation of excluding any masses in the studied range. However, from the observed limits the 850 GeV/c^2 mass point has been excluded to 95% CL (confidence level). In addition, following a linear interpolation of the observed limits for the mass points around 850 GeV/c^2, masses between 813 and 862 GeV/c^2 have been excluded to 95% CL (confidence level). These observed limits are due to a downward statistical fluctuation of data around 850 GeV/c^2.

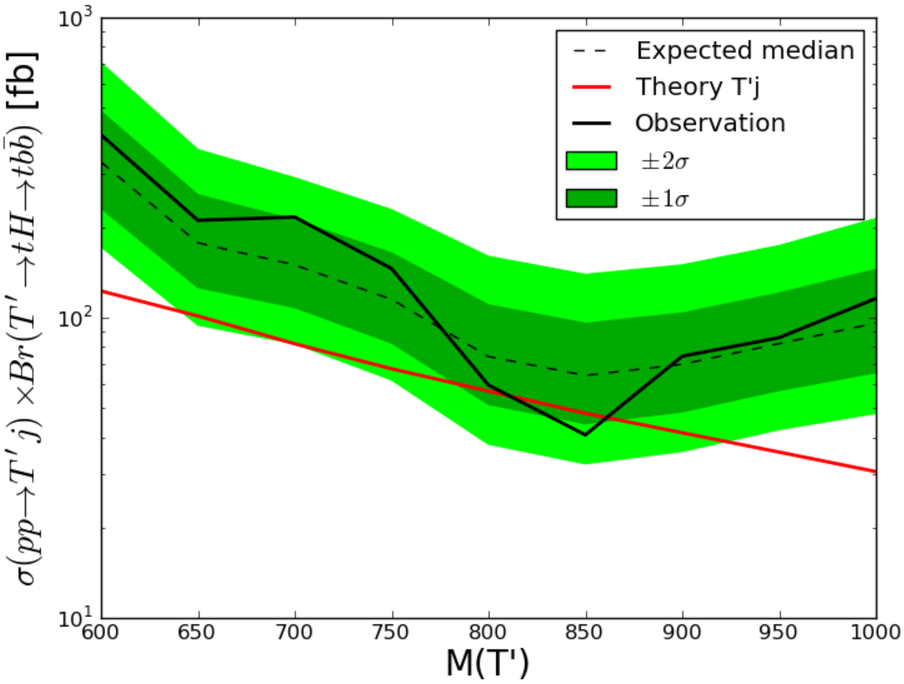


Figure 6.10: Expected and observed limits in terms of T' production cross section as a function of $M(5j)$. The red line represents the theoretical prediction of the cross section [58, 59]. Following a linear interpolation of the observed limits for the mass points around 850 GeV/c^2, masses between 813 and 862 GeV/c^2 are excluded at 95% CL.

6.4 Perspectives

The run 2 of the LHC opens different possibilities for the type of search presented in this chapter. First of all, as seen in figure 1.7, the production cross section of the T' in the single production mode will be multiplied by about 3.5 times for 14 TeV center of mass energy, with respect to the cross section at 8 TeV. However, the gain from this increase should be attenuated by the increase of the $t\bar{t}$ production cross section (of the same order), that is one of the main backgrounds in the full hadronic channel.

On the other hand, even if this analysis could have some sensitivity at 13 TeV, it should be adapted to the new conditions of the run 2. The increase in the center of mass energy will allow searches for higher masses than the ones explored in the analysis described in this work. For such objects, the p_T and H_T cuts will need to be re-tuned.

Additionally, to explore T' masses beyond the TeV, it will be necessary to adapt the analysis. For a T' with a mass of 1 TeV/c^2 produced at rest, the top quark will be produced with a p_T of 325 GeV/c and the Higgs boson with a p_T around 375 GeV/c. The decay products of a particle are reconstructed in one jet, if the $p_T \gtrsim M/R$, with R the radius of the jet and M the mass of the decaying particle. For run 2, CMS collaboration has decided to work by default with AK4 jets. Therefore, CMS will see boosted Higgs bosons if they have been produced with a $p_T > 310$ GeV/c, and top quarks if their $p_T > 440$ GeV/c. This means that a T' with a mass of 1 TeV/c^2 will produce a non-negligible proportion of boosted Higgs, and then a search for a T' with the Higgs decaying in the hadronic channel will need to implement techniques to tag a boosted Higgs. Techniques of this kind have been used, for example in [155]. These techniques have put in place the possibility of detecting several b-subjets in one jet. A crucial point, because the high b-jet multiplicity is a very important tool to discriminate signal from QCD background.

6.5 Possible improvements

Some improvements to the actual setting of the analysis, presented in this chapter, can be considered. In the first place, the dominant uncertainty of the limit calculation is the one related to the normalization estimation method. This uncertainty is diminishing the precision of the measurement in terms of excluded cross section. Additionally, with this uncertainty decreased, the next step could be to gain also in the uncertainty related to the background shape estimation, that is the sub-dominant systematic uncertainty.

In the second place, to gain some sensitivity to high mass T', it could be envisaged to add an additional category to the selection to include the cases where the $W^{+/-}$ boson is produced boosted. This is, for a $W^{+/-}$ boson produced with a $p_T > 160$ GeV/c. This could help to accept more signal events for T' masses higher than 900 GeV/c^2.

Additionally, it can be considered also to multiply the selection categories based on the different b-tagging working points and in the b-tagged jets multiplicity. For example, to add a selection category with exactly two CSVM b-tagged jets could contribute to add the events where one of the three b-quarks in the signal fails to be b-tagged. These possible categories would need a reformulation of the background estimation methods, that are fully based on the b-tagging criterion.

Finally, a detailed study of all the possible models for a T', up to detector simulation, would ease the inclusion of selection criteria to have a selection more optimized to all the possible signal cases. In particular, the study of the subleading single production mode of the T' with two additional jets, instead of just one, would allow the analysis to look for a set of processes, T'+1 jet plus T'+2 jets, with a significant higher cross section.

6.6 Other T' searches

During the run 1 of LHC several analyses have been done by the ATLAS and CMS collaborations looking for a T' in several final states and production modes. In the following these analyses are going to be briefly presented grouped by the production mode of the searched T'. Summary results of the discussed analyses will be presented at the end of this section and compared to the results from section 6.3. Whereas, there are several searches for vector like quarks very similar to the T' used in this work, as a $B^{'}$ or a T' with an exotic electric charge of 5/3, this section will present only searches for a T' with an electric charge of 2/3.

The searches that are going to be presented are based on a different model than the one used in this work. They consider that the T' mixes exclusively with the third generation SM quarks. The model used in this thesis explores a more generic model that assumes that the T' can be mixed with the three generations. In particular, the model used for the generation of the signal MC samples have been constructed with a T' that mixes with all generations of SM quarks.

6.6.1 Searches for pair produced T'

As discussed in section 1.5, the T' can be produced in pairs in proton-proton collisions, via the diagrams from figure 1.5. In this case, two heavy T' are produced and each one can follow one of the three possible decay channels: b$W^{+/-}$, tZ^0 and tH^0. Analyses performed during run 1 looked normally for one of the T''s going into a specific decay, and then applied a strategy depending on the secondT' channel(s) considered.

The simplest case is to have both T''s going into the same decay channel, for example where $T', T' \rightarrow bW^{+/-}, bW^{+/-}$. CMS collaboration has performed an analysis looking for this specific signature [159], where one of the $W^{+/-}$ bosons decays into two quarks and the other one decay into a lepton (muon or electron) and a neutrino. This search was able to exclude T' masses below 912 GeV/c^2.

Two additional analyses have been done in CMS where both T''s are considered to decay via the same channel: $T', T' \rightarrow tH^0, tH^0$ [155, 160]. One searches for the full hadronic final state using jet substructure techniques, where the top quark and Higgs boson decay products are fully contained in one jet. Exclusion limits for T' masses below 747 GeV/c^2 was obtained. The second analysis requires one Higgs boson going into two photons, thus two photons are required with at least two jets. It considers two channels where the $W^{+/-}$ boson is decaying leptonically or hadronically. For the lepton channel, at least one lepton is required and for the hadronic one at least one b-tagged jet. This last analysis exclude T' masses up to 540 GeV/c^2.

ATLAS collaboration has two searches [61] for the cases where $T^{'}, T^{'} \to tH^0, X$ and $T', T' \to bW^{+/-}, X$, where the X means any of the three channels. The second case was optimized to $T^{'}, T^{'} \to bW^{+/-}, bW^{+/-}$, being then comparable to the CMS analysis previously presented. Both analyses rely on a pre-selection requiring one electron or muon, missing energy, at least 4 jets and at least one b-tagged jet. From this pre-selection, several categories relying basically in the b-tagged jet and jet multiplicities are designed to increase sensitivity to different channels. For the $bW^{+/-}, X$ case, it proceeds as CMS analysis in the sense that the selection was optimized to select events where the hadronic $W^{+/-}$ boson was produced in boosted or in resolved topology. The two analyses have been combined to present exclusion limits not only in terms of the T' mass but also as a function of the T' branching ratio to each channel. Assuming 100% branching ratio of the T' to top-Higgs, these analysis excludes T' masses below 900 GeV/c^2. For the case of a T' exclusively decaying to $bW^{+/-}$, masses below 750 GeV/c^2 were excluded at 95% CL.

In addition both ATLAS and CMS have analysis not focused in specific decays but have implemented strategies to be sensitive to the three decay channels [161, 162]. Both searches are in leptonic final states, with muons or electrons. ATLAS analysis require at least two same charged leptons with at least 2 jets, at least one b-tagged jet and $\not{E}_T > 40$ GeV/c, while CMS divide the analysis in two channels, one with just one lepton and the other for multi-lepton final states. In the single lepton channel, the lepton is required to be in the detector acceptance. In addition, at least 3 jets are required where one of them should be W-tagged, as described in the $T', T' \to bW^{+/-}, bW^{+/-}$ analysis. For the multi-lepton channel, missing energy is required, at least one b-tagged jet, and 4 categories are constructed, to increase sensitivity to each decay channel, from the jet multiplicity requirement. Results from ATLAS and CMS analyses are shown in figure 6.11.

Finally, a search for $T', T' \to tZ^0, X$ has been performed by the ATLAS collaboration [163]. In this analysis both single and pair production are considered. It selects events with at least 2 leptons, where one pair of leptons is a Z^0-candidate. Two channels are then build, one for di-leptonic events (exactly 2 leptons) and one for tri-leptonic events (at least 3 leptons). In the di-lepton channel at least 2 b-tagged jets are required, while in the tri-lepton channel at least one b-tagged jet is required. Combining both channels, under the pair production hypothesis,

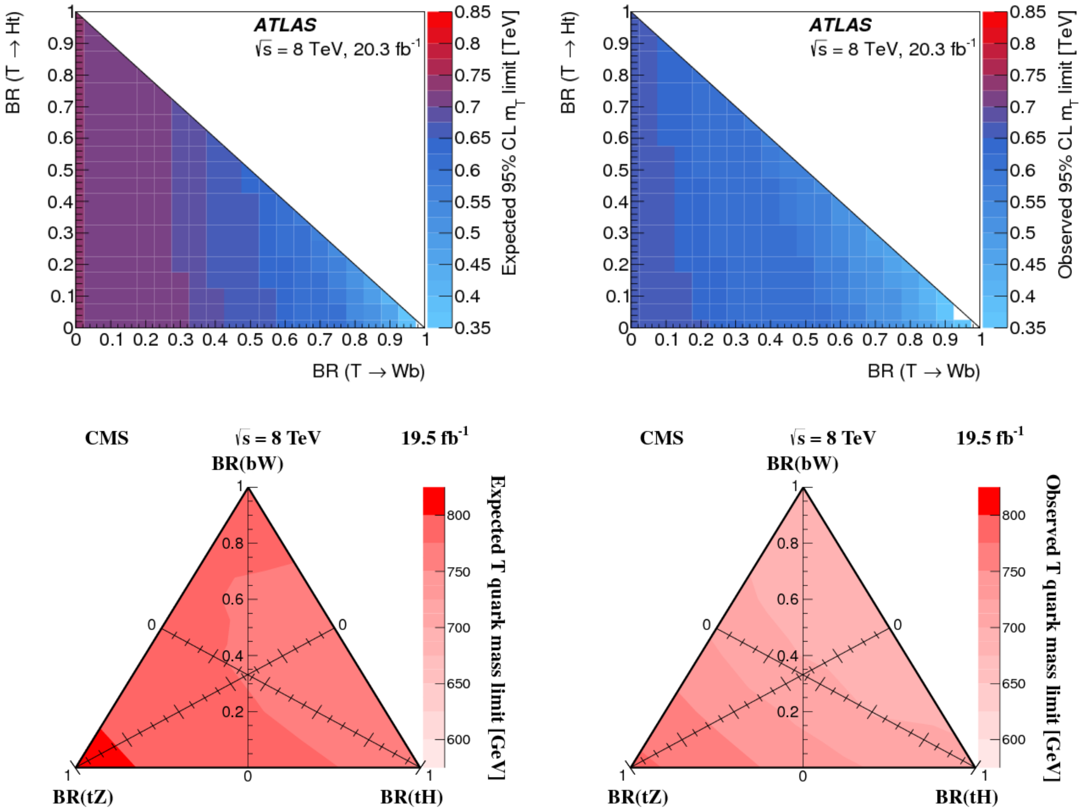


Figure 6.11: ATLAS expected [left-top] and observed [right-top] T' mass excluded at 95% CL as a function of the branching ratio [161]. CMS expected [left-bottom] and observed [right-bottom] T' mass excluded at 95% CL as a function of the branching ratio [162]. These triangular plots correspond to a scan over all the possible branching fractions of the T' into its three possible decay channels: $bW^{+/-}$, tH^0 and tZ^0. The scan is constrained to have one as the sum of the three individual branching ratios. In ATLAS plots, the (0,0) corner corresponds to $Br(T' \to tZ^0) = 1$, the up corner to $Br(T' \to tH^0) = 1$ and the right corner to $Br(T' \to bW^{+/-}) = 1$. In CMS plots, these exclusive branching ratios are left, up and right corners respectively.

T' masses below 800 GeV/c^2 were excluded.

6.6.2 Searches for single produced T'

The only analysis in the literature exploring the T' single production mode is part of the last discussed analysis [163] in the previous subsection. For the single production mode at least one forward jet is required instead of requiring a minimum H_T. This forward jet is used as the single production mode signature. This analysis did not put exclusion limits on T' mass, but did established an upper limit in the single mode production cross section. Figure 6.12 shows these results [163]. It is important to remark that the production mode studied in this search is different from the one used in this work. ATLAS analysis assumes that in the single production mode the T' is produced with two additional quarks. The model used in this thesis assumes that the T' is produced with only one additional quark.

ATLAS analysis in the single production mode puts an upper limit in the production cross section of the T' multiplied by the $Br(T' \to tZ^0)$ of around 3×10^2 fb as mean value. The CMS analysis, presented in this work, puts an upper limit on the production cross section of the single production mode of the T' multiplied by the $Br(H^0 \to b\bar{b})$ of around 1×10^2 fb as mean value. With respect to pair production, CMS analyses have put upper limits between 10 to 100 fb in the production cross section. These limits are more stringent than the ones achieved in the single production mode. ATLAS have put similar limits in the T' pair production mode.

6.7 Summary and conclusions

Figure 1.10, presented in section 1.5.4, shows the summary of results from all analysis looking for a T' by ATLAS and CMS collaborations. From there, the highest limits achieved are 900 GeV/c^2 for $bW^{+/-}$ exclusive decay, 850 GeV/c^2 for tH^0 case and 800 GeV/c^2 for tZ^0 channel. Taking into account these results, the search presented in this thesis adds important information in two ways. First, it puts limits in a theoretical model not explored by other searches. Second, it achieves upper limits in the production cross section of the single production mode, comparable but lower than in other analyses.

Two main conclusions can be derived from the analyses presented. The single production mode has been poorly explored, only one analysis, and the TeV territory (masses higher than 1 TeV/c^2) has not been excluded up to now. The single production mode represents a challenging mode for searches, but it can contribute to overall understanding of VLQ's, as it was shown in this thesis. For LHC run 2, the searches should be able to begin the exploration of the TeV territory with a greater collision energy.

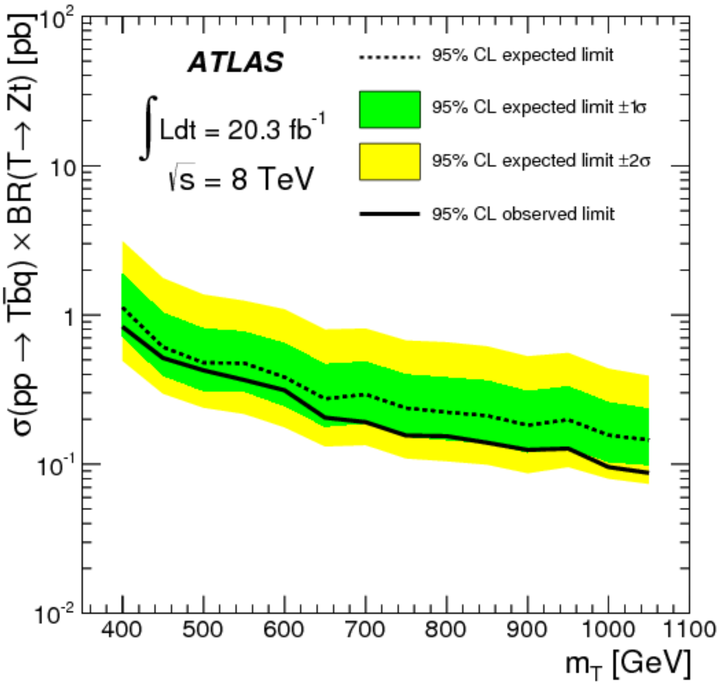


Figure 6.12: Upper limit (at 95% CL) on the T' single production mode cross section times the branching ratio as a function of its mass.

Conclusions

This thesis studies followed two axis: Monte-Carlo simulations in the context of the CMS collaboration and a search for the vector like quark T' in proton-proton collisions from LHC run 1.

The Monte-Carlo studies performed are grouped in two main activities. First of all, the maintenance and improvement of CMS production of simulations using the MadGraph package. This task was also accompanied by constant support to CMS users of MadGraph. As a result, several new pieces of code were introduced to the central CMS software, CMSSW, and other codes were optimized and updated to new versions of MadGraph or CMSSW. Additionally, valuable support was given to several users for the production and validation of Monte-Carlo samples of their interest. The second main activity was the physics validation of recent versions of MadGraph and new tools for Monte-Carlo simulations. This validation process probed the reliability of this package, an important step for later CMS usage. From this validation activities, various releases of MadGraph were validated. These new releases were included in CMS simulation production process, replacing old ones. Additionally, the recent developed tool MadSpin was added to the production of $t\bar{t}$ samples with MadGraph, increasing the capability of CMS to generate this process, requiring less time and cpu resources, in addition of having samples with a better description of the $t\bar{t}$ production.

The second axis studies followed in two directions. In first instance, a phenomenological feasibility study for a search in LHC data collected in run 1 was performed. Such study showed the observability of a single produced T' in the full hadronic channel, where the T' decays into a top quark and a Higgs boson, and both decaying in the hadronic modes. It performed a mass reconstruction in the full hadronic channel using two main tools: a procedure to optimize the selection of the 5 jets coming from the T', and jet b-tagging as a powerful tool to discriminate signal events from backgrounds. On top of this, several variables were identified to further discriminate signal events. As a result, a selection was designed that gave a satisfactory significance with S/B and $S/\sqrt{S+B}$ used as discriminators.

Based on the phenomenological study a search for a single produced T' was performed in CMS run 1 data. An exclusive efficiency of 70% was found for the identification process of the decay products from the T'. This identification procedure was based in the requirement of at least 3 b-tagged jets, allowing a high discrimination of signal events with respect to QCD background events. The identification process also reconstructed the top quark and Higgs boson coming from the T'. Using these objects, a final selection was performed. Finally, a novel data-based background estimation procedure was put in place. This procedure developed a method to estimate the shape of the invariant mass of 5 jets. A separate method was also developed to estimate the normalization of the backgrounds, this is the number of expected events from backgrounds after the full selection. With the usage of these methods, sensible exclusion limits were achieved. With respect to theoretical predictions of the production cross section of the T' as a function of its mass, no expected exclusion limits are achieved but an upper limit on the cross section is set. However, a T' mass of 850 GeV/c^2 was excluded at 95% CL in observed limits.

Finally, the results from the analysis were put in perspective with respect to similar analyses done by ATLAS and CMS collaborations. It was shown that the analysis presented in this thesis adds an important contribution to the current knowledge of vector like quarks, in terms of exclusion limits and techniques for searching them in the single production mode. Also it is important to stress that the analysis explores the full hadronic final state, a difficult channel not widely explored as leptonic ones. Currently, the LHC has restarted proton-proton collisions at 13 TeV center of mass energy, exploring territories never reached before. From these collisions, new searches for T' vector like quarks are expected, with an increased sensitivity to higher masses. It is expected then that for the first time T' masses higher then 1 TeV/c^2 will be tested.

A Limits with different integration windows

In the present appendix a study of the dependence of limits with respect to the chosen integration window is presented. In chapter 5, the limits, yields and systematics were computed from the integration over one sigma of the mean values and widths presented in table 6.1 (see section 6.2). However the integration over one sigma of signal widths for the calculation of signal yields could seem restricted, because it is only taking into account 68% of signal events, it is not evident that opening up the integration window will led to better limits due to the increase of background events entering in the window. Reason why it has been chosen to repeat the exercise of limit calculation with 1.5 and 2 sigma.

For 1.5 sigma the entire set of systematic uncertainties have been recalculated as well as the yields. The same procedure described in section 6.2 has been followed. Table A.1 shows the set of systematics for each mass point with 1.5 integration window. Also in table A.2 the yields with 1.5 sigma window are presented. The limits with 1.5 sigma are presented in figure A.1. Background systematics are independent of the integration window, then the same values quoted in table 6.8 have been used for limits calculation.

Sample Name	b-tagging		JEC		PDF+α_S	Pileup	Trigger
	up [%]	down [%]	up [%]	down [%]	up/down [%]	up/down [%]	up/down [%]
$Tj \rightarrow tHj$ 600 GeV/c^2	7.28	7.28	10.79	11.18	7.28	7.28	7.28
$Tj \rightarrow tHj$ 650 GeV/c^2	6.15	5.92	8.25	10.52	5.29	5.29	5.29
$Tj \rightarrow tHj$ 700 GeV/c^2	6.12	5.91	4.51	5.16	4.51	4.51	4.51
$Tj \rightarrow tHj$ 750 GeV/c^2	6.29	6.05	4.68	8.03	4.46	4.27	4.27
$Tj \rightarrow tHj$ 800 GeV/c^2	6.33	6.10	5.47	7.07	4.22	4.22	4.22
$Tj \rightarrow tHj$ 850 GeV/c^2	6.70	6.43	4.00	4.36	4.00	4.00	4.00
$Tj \rightarrow tHj$ 900 GeV/c^2	6.74	6.46	3.94	3.94	3.94	3.94	3.94
$Tj \rightarrow tHj$ 950 GeV/c^2	6.95	6.66	4.29	6.15	4.29	4.29	4.29
$Tj \rightarrow tHj$ 1000 GeV/c^2	6.86	6.58	5.04	5.59	4.81	4.81	4.81

Table A.1: Summary of uncertainties for signal samples with 1.5σ integration window. In addition, all mass points have 2.6% uncertainty from luminosity measurement.

Finally, the same procedure is repeated for 2 sigma integration window. For this

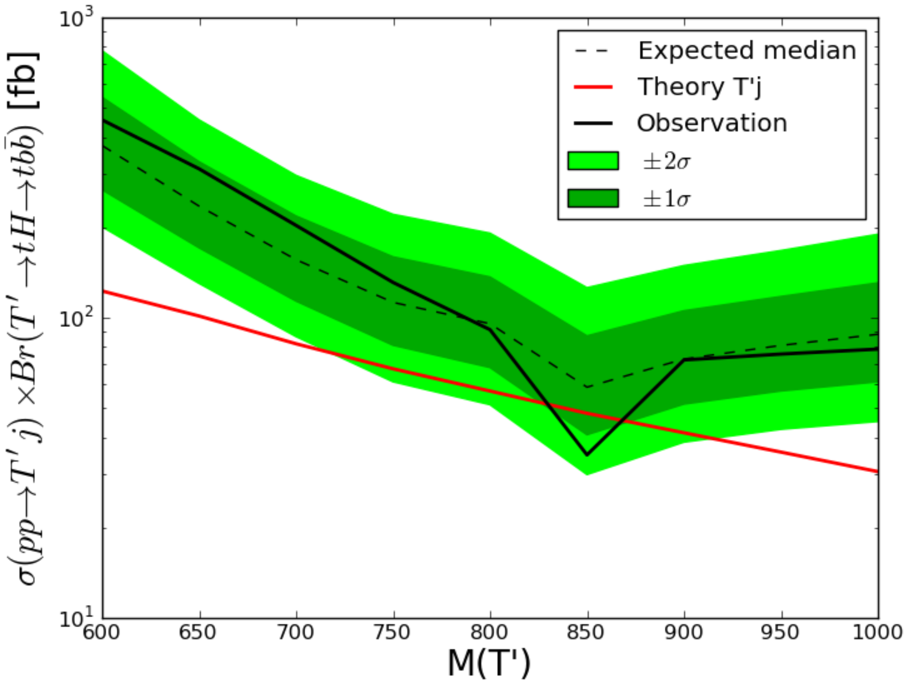


Figure A.1: Expected and observed limits in terms of T' production cross section as function of $M(5j)$ with 1.5 sigma integration window. The red line represents the theoretical prediction of the cross section [58, 59]. The 850 GeV/c^2 mass point is excluded at 95% CL. With a linear approximation it is possible to exclude masses between 830 and 870 GeV/c^2 at 95% CL.

T' Mass GeV/c^2	Signal	Background	Observed Data
600	5.16±0.38	16.97	22
650	7.78±0.41	18.92	28
700	8.59±0.39	17.38	25
750	7.93±0.34	14.60	18
800	7.08±0.30	14.91	14
850	6.42±0.26	10.69	4
900	5.69±0.22	11.11	11
950	4.28±0.18	11.00	10
1000	2.84±0.14	8.33	7

Table A.2: Expected number of events for the signal, estimated background and observed data after full selection with 1.5σ integration window.

window, signal systematics can be seen in table A.3 as well as yields in table A.4. Limits for 2 sigma integration window are shown in figure A.2.

Sample Name	b-tagging		JEC		PDF+α_S	Pileup	Trigger
	up [%]	down [%]	up [%]	down [%]	up/down [%]	up/down [%]	up/down [%]
$Tj \to tHj$ 600 GeV/c^2	7.28	7.28	10.80	11.18	7.28	7.28	7.28
$Tj \to tHj$ 650 GeV/c^2	6.16	5.94	6.71	10.51	5.17	5.17	5.17
$Tj \to tHj$ 700 GeV/c^2	6.15	5.93	4.44	6.24	4.44	4.44	4.44
$Tj \to tHj$ 750 GeV/c^2	6.25	6.02	4.65	7.32	4.12	4.12	4.12
$Tj \to tHj$ 800 GeV/c^2	6.33	6.10	4.13	5.34	4.13	4.13	4.13
$Tj \to tHj$ 850 GeV/c^2	6.67	6.40	4.25	5.43	3.84	3.84	3.84
$Tj \to tHj$ 900 GeV/c^2	6.72	6.44	3.87	3.87	3.87	3.87	3.87
$Tj \to tHj$ 950 GeV/c^2	6.95	6.66	4.29	6.15	4.29	4.29	4.29
$Tj \to tHj$ 1000 GeV/c^2	6.85	6.58	8.22	4.65	4.65	4.65	4.65

Table A.3: Summary of uncertainties for signal samples with 2σ integration window. In addition, all mass points have 2.6% uncertainty from luminosity measurement.

The best limits achieved from the analysis are obtained using one sigma integration window. This opening limits the number of background events entering the considered T' masses.

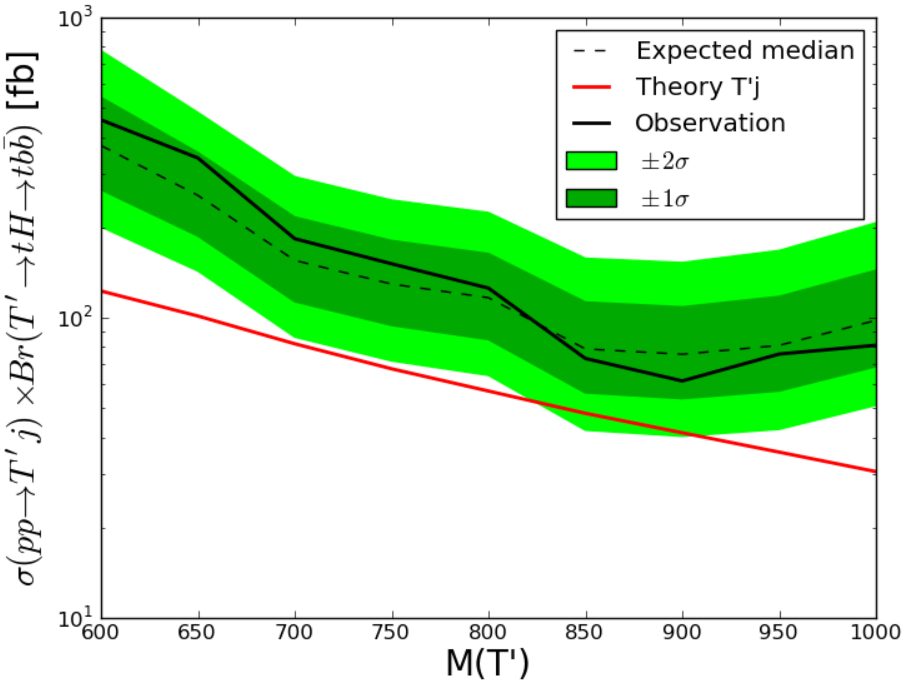


Figure A.2: Expected and observed limits in terms of T' production cross section as function of $M(5j)$ with 2 sigma integration window. The red line represents the theoretical prediction of the cross section [58, 59]. No observed exclusion limits are reached.

T' Mass GeV/c^2	Signal	Background	Observed Data
600	5.16±0.38	16.97	22
650	8.15±0.42	22.21	33
700	8.87±0.39	19.95	25
750	8.52±0.35	19.44	24
800	7.38±0.31	19.02	21
850	6.95±0.27	14.40	13
900	5.91±0.23	14.81	11
950	4.28±0.18	11.00	10
1000	3.04±0.14	11.93	9

Table A.4: Expected number of events for the signal, estimated background and observed data after full selection with 2σ integration window.

Bibliography

[1] S. Weinberg. *A Model of Leptons*. *Phys. Rev. Lett.* 19 (1967), 1264–1266.

[2] D. Griffiths. *Introduction to elementary particles* (2008).

[3] W. Pauli. *Relativistic Field Theories of Elementary Particles*. *Rev. Mod. Phys.* 13 (1941), 203–232.

[4] P. A. M. Dirac. *Quantum theory of emission and absorption of radiation*. *Proc. Roy. Soc. Lond.* A114 (1927), 243.

[5] R. P. Feynman. *Mathematical formulation of the quantum theory of electromagnetic interaction*. *Phys. Rev.* 80 (1950), 440–457.

[6] E. Fermi. *An attempt of a theory of beta radiation. 1.*. *Z. Phys.* 88 (1934), 161–177.

[7] A. Einstein. *On the electrodynamics of moving bodies*. *Annalen Phys.* 17 (1905), 891–921. [Annalen Phys.14,194(2005)].

[8] H. Georgi. *Lie algebras in particle physics*. Frontiers in Physics. Perseus Books, Reading, MA (1999).

[9] P. M. Neumann. *The matehmatical writings of Évariste Galois*. Heritage of European Mathematics. European Mathematical Society, University of Oxford, UK (2013).

[10] E. Noether. *Invariant variation problems*. *Transport Theory and Statistical Physics* 1 (1971), 186–207. `arXiv:physics/0503066`.

[11] L. Evans et al. *LHC Machine*. *Journal of Instrumentation* 3 (2008)(08), S08001.

[12] I. Newton. *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press. University of California Press, California (1999).

[13] W. Pauli. *Dear radioactive ladies and gentlemen*. *Phys. Today* 31N9 (1978), 27.

[14] R. Feynman. *The Feynman Lectures on Physics*. Basic Books. Basic Books, A Member of the Perseus Books Group, Philadelphia, PA (2013).

[15] S. Goldstein et al. *Classical Mechanics*. Addison-wesley (2001).

[16] M. Maggiore. *A Modern introduction to quantum field theory* (2005).

[17] J. D. Jackson. *Classical Electrodynamics*. Wiley (1998).

[18] H. Becquerel. *On the rays emitted by phosphorescence*. *Compt. Rend. Hebd. Seances Acad. Sci.* 122 (1896)(8), 420–421.

[19] M. G. B. P.Curie, M. Curie. *Sur une nouvelle substance fortement radioactive, contenue dans la pechblende*. *Compt. Rend. Hebd. Seances Acad. Sci.* 127 (1898)(7), 1215–1217.

[20] E. Rutherford. *The scattering of alpha and beta particles by matter and the structure of the atom*. *Phil. Mag.* 21 (1911), 669–688.

[21] S. L. Glashow et al. *Breaking chiral symmetry*. *Phys. Rev. Lett.* 20 (1968), 224–227.

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

[23] M. Gell-Mann. *A Schematic Model of Baryons and Mesons*. *Phys. Lett.* 8 (1964), 214–215.

[24] E. Schrödinger. *Quantisierung als Eigenwertproblem*. *Annalen der Physik* 384 (1926)(4), 361–376.

[25] S.-I. Tomonaga et al. *On Infinite Field Reactions in Quantum Field Theory*. *Phys. Rev.* 74 (1948), 224–225.

[26] J. S. Schwinger. *Quantum electrodynamics. I A covariant formulation*. *Phys. Rev.* 74 (1948), 1439.

[27] R. P. Feynman. *Space-time approach to nonrelativistic quantum mechanics*. *Rev. Mod. Phys.* 20 (1948), 367–387.

[28] J. J. Thomson. *Cathode rays*. *Phil. Mag.* 44 (1897), 293–316.

[29] C. D. Anderson. *The Positive Electron*. *Phys. Rev.* 43 (1933), 491–494.

[30] R. Brandelik et al. (TASSO Collaboration). *Evidence for Planar Events in e+ e- Annihilation at High-Energies*. *Phys. Lett.* B86 (1979), 243.

[31] C. Berger et al. (PLUTO Collaboration). *Evidence for Gluon Bremsstrahlung in e+ e- Annihilations at High-Energies*. *Phys. Lett.* B86 (1979), 418.

[32] P. Hansen. *Results from the UA1 and UA2 Experiments*. In *International School of Physics Enrico Fermi: Elementary Particles Varenna, Italy, June 26-July 6, 1984* (1984) .

[33] *Report on the design study of a 300 GeV proton synchrotron*. CERN, Geneva (1964). French version available.

[34] T. A. Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. *Journal of Instrumentation* 3 (2008)(08), S08003.

[35] T. C. Collaboration. *The CMS experiment at the CERN LHC*. *Journal of Instrumentation* 3 (2008)(08), S08004.

[36] G. Aad et al. (ATLAS Collaboration). *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. *Phys. Lett.* B716 (2012), 1–29. `arXiv:1207.7214`.

[37] S. Chatrchyan et al. (CMS Collaboration). *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. *Phys. Lett.* B716 (2012), 30–61. `arXiv:1207.7235`.

[38] N. N. M. A. 2014. *The Nobel Prize in Physics 2013* (2013).

[39] F. Englert et al. *Broken Symmetry and the Mass of Gauge Vector Mesons*. *Phys. Rev. Lett.* 13 (1964), 321–323.

[40] P. W. Higgs. *Broken Symmetries and the Masses of Gauge Bosons*. *Phys. Rev. Lett.* 13 (1964), 508–509.

[41] P. Fayet. *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*. *Phys. Lett.* B69 (1977), 489.

[42] S. P. Martin. *A Supersymmetry primer*. *Adv.Ser.Direct.High Energy Phys.* 21 (2010), 1–153. `arXiv:hep-ph/9709356`.

[43] V. Khachatryan et al. (CMS Collaboration). *Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions*. *JHEP* 1410 (2014), 160. `arXiv:1408.3316`.

[44] G. Aad et al. (ATLAS Collaboration). *Search for neutral Higgs bosons of the minimal supersymmetric standard model in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*. *JHEP* 1411 (2014), 056. `arXiv:1409.6064`.

[45] K. Olive et al. (Particle Data Group). *Review of Particle Physics*. *Chin.Phys.* C38 (2014), 090001.

[46] Y. Ashie et al. (Super-Kamiokande). *Evidence for an oscillatory signature in atmospheric neutrino oscillation*. *Phys.Rev.Lett.* 93 (2004), 101801. `arXiv:hep-ex/0404034`.

[47] C. Weinheimer et al. *Neutrino Masses*. *Annalen Phys.* 525 (2013)(8-9), 565–575. `arXiv:1307.3518`.

[48] C. L. Bennett et al. *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*. *The Astrophysical Journal, Supplement* 208 (2013), 20. `arXiv:1212.5225`.

[49] G. Hinshaw et al. *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*. *The Astrophysical Journal, Supplement* 208 (2013), 19. `arXiv:1212.5226`.

[50] P. Ade et al. (Planck Collaboration). *Planck 2015 results. XIII. Cosmological parameters* (2015). `arXiv:1502.01589`.

[51] O. Eberhardt et al. *Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations*. *Phys.Rev.Lett.* 109 (2012), 241802. `arXiv:1209.1101`.

[52] S. Schael et al. (SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3). *Precision electroweak measurements on the Z resonance*. *Phys. Rept.* 427 (2006), 257–454. `arXiv:hep-ex/0509008`.

[53] T. Kaluza. *On the Problem of Unity in Physics*. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* 1921 (1921), 966–972.

[54] R. Contino et al. *Light custodians in natural composite Higgs models*. *Phys.Rev.* D75 (2007), 055014. `arXiv:hep-ph/0612048`.

[55] O. Matsedonskyi et al. *Light Top Partners for a Light Composite Higgs*. *JHEP* 1301 (2013), 164. `arXiv:1204.6333`.

[56] G. Dissertori et al. *Discovery potential of top-partners in a realistic composite Higgs model with early LHC data*. *JHEP* 1009 (2010), 019. `arXiv:1005.4414`.

[57] J. A. Aguilar-Saavedra et al. *Handbook of vectorlike quarks: Mixing and single production*. *Phys. Rev.* D88 (2013)(9), 094010. `arXiv:1306.0572`.

[58] M. Buchkremer et al. *Model Independent Framework for Searches of Top Partners*. *Nucl.Phys.* B876 (2013), 376–417. `arXiv:1305.4172`.

[59] G. Cacciapaglia et al. *Heavy Vector-like Top Partners at the LHC and flavour constraints*. *JHEP* 1203 (2012), 070. `arXiv:1108.6329`.

[60] A. De Simone et al. *A First Top Partner Hunter's Guide*. *JHEP* 04 (2013), 004. `arXiv:1211.5663`.

[61] G. Aad et al. (ATLAS Collaboration). *Search for production of vector-like quark pairs and of four top quarks in the lepton plus jets final state in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*. Tech. Rep. ATLAS-CONF-2015-012, CERN, Geneva (Mar 2015).

[62] V. Khachatryan et al. (CMS Collaboration). *Search for vector-like charge 2/3 T quarks in proton-proton collisions at* $\sqrt{s} = 8$ *TeV* (2015). `arXiv:1509.04177`.

[63] F. Abe et al. (CDF Collaboration). *Observation of top quark production in* $\bar{p}p$ *collisions*. *Phys. Rev. Lett.* 74 (1995), 2626–2631. `arXiv:hep-ex/9503002`.

[64] S. Abachi et al. (D0 Collaboration). *Observation of the top quark*. *Phys. Rev. Lett.* 74 (1995), 2632–2637. `arXiv:hep-ex/9503003`.

[65] R. Blair et al. (CDF Collaboration). *The CDF-II detector: Technical design report* (1996).

[66] V. M. Abazov et al. (D0 Collaboration). *D0 Run IIB upgrade technical design report* (2002).

[67] T. Group. *Design Report Tevatron 1 project* (1984).

[68] S. W. Herb et al. *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*. *Phys. Rev. Lett.* 39 (1977), 252–255.

[69] M. Czakon et al. *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through* $O(\alpha_S^4)$. *Phys.Rev.Lett.* 110 (2013), 252004. `arXiv:1303.6254`.

[70] L. P. C. C. Top LHC working Group. *Top LHC Working Group Summary Plots* (2015).

[71] *Measurement of the inclusive and differential tt production cross sections in lepton + jets final states at 13 TeV*. Tech. Rep. CMS-PAS-TOP-15-005, CERN, Geneva (2015).

[72] *Measurements of the $t\bar{t}$ production cross-section in the dilepton and lepton-plus-jets channels and of the ratio of the $t\bar{t}$ and Z boson cross-sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. Rep. ATLAS-CONF-2015-049, CERN, Geneva (Sep 2015).

[73] V. M. Abazov et al. (D0 Collaboration). *Observation of Single Top Quark Production*. *Phys. Rev. Lett.* 103 (2009), 092001. `arXiv:0903.0850`.

[74] T. Aaltonen et al. (CDF Collaboration). *First Observation of Electroweak Single Top Quark Production*. *Phys.Rev.Lett.* 103 (2009), 092002.

[75] G. Aad et al. (ATLAS Collaboration). *Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7$ TeV*. *Phys. Lett.* B716 (2012), 142–159. `arXiv:1205.5764`.

[76] S. e. a. Chatrchyan (CMS Collaboration). *Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s} = 8$ TeV*. Tech. Rep. arXiv:1401.2942. CMS-TOP-12-040. CERN-PH-EP-2013-237, CERN, Geneva (Jan 2014). Comments: Submitted to Phys. Rev. Lett.

[77] N. Cabibbo. *Unitary Symmetry and Leptonic Decays*. *Phys. Rev. Lett.* 10 (1963), 531–533.

[78] M. Kobayashi et al. *CP-Violation in the Renormalizable Theory of Weak Interaction*. *Progress of Theoretical Physics* 49 (1973)(2), 652–657.

[79] M. Jeżabek et al. *QCD corrections to semileptonic decays of heavy quarks*. *Nuclear Physics B* 314 (1989)(1), 1–6.

[80] V. M. Abazov et al. (D0 Collaboration). *An Improved determination of the width of the top quark*. *Phys.Rev.* D85 (2012), 091104.

[81] LHCHXSWG. *LHC Higgs Cross Section Working Group Picture Gallery* (2015).

[82] LHC Higgs Cross Section Working Group et al. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. *CERN-2011-002* (CERN, Geneva, 2011). `arXiv:1101.0593`.

[83] LHC Higgs Cross Section Working Group et al. *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*. *CERN-2012-002* (CERN, Geneva, 2012). `arXiv:1201.3084`.

[84] LHC Higgs Cross Section Working Group et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. CERN-2013-004* (CERN, Geneva, 2013). `arXiv:1307.1347`.

[85] G. Aad et al. (ATLAS, CMS Collaborations). *Combined Measurement of the Higgs Boson Mass in pp Collisions at* $\sqrt{s} = 7$ *and 8 TeV with the ATLAS and CMS Experiments*. *Phys.Rev.Lett.* 114 (2015), 191803. `arXiv:1503.07589`.

[86] S. Chatrchyan et al. (CMS Collaboration). *Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model* (2014).

[87] G. Aad et al. (ATLAS Collaboration). *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at* $\sqrt{s}$ *= 7 and 8 TeV in the ATLAS experiment*. Tech. Rep. ATLAS-CONF-2015-007, CERN, Geneva (Mar 2015).

[88] V. Khachatryan et al. (CMS Collaboration). *Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs*. *Phys.Lett.* B736 (2014), 64. `arXiv:1405.3455`.

[89] O. S. Brüning et al. *LHC Design Report*. CERN, Geneva (2004).

[90] P. Cortese et al. (ALICE Collaboration). *ALICE physics performance: Technical Design Report*. Technical Design Report ALICE. CERN, Geneva (2005). Revised version submitted on 2006-05-29 15:15:40.

[91] G. Aad et al. (ATLAS Collaboration). *ATLAS detector and physics performance: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva (1999). Electronic version not available.

[92] G. L. Bayatian et al. (CMS Collaboration). *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva (2006).

[93] J. Alves, A. Augusto et al. (LHCb Collaboration). *The LHCb Detector at the LHC*. *JINST* 3 (2008), S08005.

[94] C. Augier et al. (UA4/2 Collaboration). *Predictions on the total cross-section and real part at LHC and SSC*. *Phys.Lett.* B315 (1993), 503–506.

[95] S. Brochet et al. *Recherche de physique au-delà du modèle standard dans le secteur du quark top et calibration de l'échelle en énergie des jets avec l'expérience CMS du LHC*. Ph.D. thesis, Lyon, IPN (2014).

[96] S. Chatrchyan et al. (CMS Collaboration). *ECAL Performance Plots, 2012 Data* (2012).

[97] S. Chatrchyan et al. (CMS Collaboration). *The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC*. *JINST* 8 (2013), P11002. `arXiv:1306.6905`.

[98] T. Lenzi. *Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors*. Ph.D. thesis (2013). `arXiv:1306.0858`.

[99] V. Khachatryan et al. (CMS Collaboration). *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. Rep. CMS-PAS-PFT-09-001 (2009).

[100] V. Khachatryan et al. (CMS Collaboration). *Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV*. Tech. Rep. CMS-PAS-PFT-10-002 (2010).

[101] V. Khachatryan et al. (CMS Collaboration). *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*. Tech. Rep. CMS-PAS-PFT-10-001 (2010).

[102] V. Khachatryan et al. (CMS Collaboration). *Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV*. Tech. Rep. CMS-PAS-PFT-10-003 (2010).

[103] S. Chatrchyan et al. (CMS Collaboration). *Description and performance of track and primary-vertex reconstruction with the CMS tracker*. *JINST* 9 (2014)(10), P10009. `arXiv:1405.6569`.

[104] G. P. Salam et al. *A Practical Seedless Infrared-Safe Cone jet algorithm*. *JHEP* 0705 (2007), 086. `arXiv:0704.0292`.

[105] M. Cacciari et al. *The Anti-k(t) jet clustering algorithm*. *JHEP* 0804 (2008), 063. `arXiv:0802.1189`.

[106] S. D. Ellis et al. *Successive combination jet algorithm for hadron collisions*. *Phys.Rev.* D48 (1993), 3160–3166. `arXiv:hep-ph/9305266`.

[107] M. Cacciari et al. *The Catchment Area of Jets*. *JHEP* 0804 (2008), 005. `arXiv:0802.1188`.

[108] F. Beaudette (CMS Collaboration). *The CMS Particle Flow Algorithm* (2014), 295–304. `arXiv:1401.8155`.

[109] Chatrchyan, Serguei and others. *Determination of jet energy calibration and transverse momentum resolution in CMS*. *Journal of Instrumentation* 6 (2011), 11002. `arXiv:1107.4277`.

[110] S. Chatrchyan et al. (CMS Collaboration). *Identification of b-quark jets with the CMS experiment*. *JINST* 8 (2013), P04013. `arXiv:1211.4462`.

[111] S. Chatrchyan et al. (CMS Collaboration). *Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events*. Tech. Rep. CMS-PAS-BTV-13-001, CERN, Geneva (2013).

[112] S. Chatrchyan et al. (CMS Collaboration). *Search for Anomalous Top Quark Pair Production in the Boosted All-Hadronic Final State using pp Collisions at sqrt(s) = 8 TeV*. Tech. Rep. CMS-PAS-B2G-12-005 (2013).

[113] G. Aad et al. (ATLAS Collaboration). *Performance of boosted top quark identification in 2012 ATLAS data*. Tech. Rep. ATLAS-CONF-2013-084, CERN, Geneva (Aug 2013).

[114] E. Usai (ATLAS, CMS Collaborations). *Boosted top: experimental tools overview* (2015). `arXiv:1501.00900`.

[115] S. Beauceron et al. *Fully hadronic decays of a singly produced vectorlike top partner at the LHC*. *Phys. Rev.* D90 (2014)(11), 115008. `arXiv:1401.5979`.

[116] J. Alwall et al. *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*. *JHEP* 1407 (2014), 079. `arXiv:1405.0301`.

[117] J. Alwall et al. *MadGraph 5 : Going Beyond*. *JHEP* 1106 (2011), 128. `arXiv:1106.0522`.

[118] A. Alloul et al. *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*. *Comput.Phys.Commun.* 185 (2014), 2250–2300. `arXiv:1310.1921`.

[119] T. Sjostrand et al. *PYTHIA 6.4 Physics and Manual*. *JHEP* 0605 (2006), 026. `arXiv:hep-ph/0603175`.

[120] M. Cacciari et al. *FastJet User Manual*. *Eur.Phys.J.* C72 (2012), 1896. `arXiv:1111.6097`.

[121] E. Conte et al. *MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology*. *Comput.Phys.Commun.* 184 (2013), 222–256. `arXiv:1206.1599`.

[122] E. Conte et al. *Designing and recasting LHC analyses with MadAnalysis 5*. *Eur.Phys.J.* C74 (2014)(10), 3103. `arXiv:1405.3982`.

[123] G. Aad et al. (ATLAS Collaboration). *Search for heavy vector-like quarks coupling to light quarks in proton-proton collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*. *Phys.Lett.* B712 (2012), 22–39. `arXiv:1112.5755`.

[124] R. P. Feynman. *Very high-energy collisions of hadrons*. *Phys. Rev. Lett.* 23 (1969), 1415–1417.

[125] *SLAC linear collider conceptual design report*. SLAC, Stanford, CA (1980).

[126] L. W. Whitlow et al. *Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross sections*. *Phys. Lett. B* 282 (1991)(SLAC-PUB-5442), 475–482. 20 p.

[127] *HERA - A Proposal for a Large Electron Proton Colliding Beam Facility at DESY*. Tech. Rep. DESY-HERA-81-10 (1981).

[128] S. Chekanov et al. (ZEUS Collaboration). *Measurement of the Longitudinal Proton Structure Function at HERA*. *Phys. Lett.* B682 (2009), 8–22. `arXiv:0904.1092`.

[129] A. Martin et al. *Parton distributions for the LHC*. *Eur.Phys.J.* C63 (2009), 189–285. `arXiv:0901.0002`.

[130] P. M. Nadolsky et al. *Implications of CTEQ global analysis for collider observables*. *Phys.Rev.* D78 (2008), 013004. `arXiv:0802.0007`.

[131] R. D. Ball et al. *A first unbiased global NLO determination of parton distributions and their uncertainties*. *Nucl.Phys.* B838 (2010), 136–206. `arXiv:1002.4407`.

[132] M. E. Peskin et al. *An introduction to quantum field theory*. Advanced book program. Westview Press Reading (Mass.), Boulder (Colo.) (1995). Autre tirage : 1997.

[133] S. Agostinelli et al. (GEANT4). *GEANT4: A Simulation toolkit*. *Nucl.Instrum.Meth.* A506 (2003), 250–303.

[134] P. Nason. *A New method for combining NLO QCD with shower Monte Carlo algorithms*. *JHEP* 0411 (2004), 040. `arXiv:hep-ph/0409146`.

[135] S. Frixione et al. *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*. *JHEP* 0711 (2007), 070. `arXiv:0709.2092`.

[136] S. Alioli et al. *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*. *JHEP* 1006 (2010), 043. `arXiv:1002.2581`.

[137] C. Degrande et al. *UFO - The Universal FeynRules Output*. *Comput.Phys.Commun.* 183 (2012), 1201–1214. `arXiv:1108.2040`.

[138] P. Artoisenet et al. *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*. *JHEP* 1303 (2013), 015. `arXiv:1212.3460`.

[139] S. Frixione et al. *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*. *JHEP* 0704 (2007), 081. `arXiv:hep-ph/0702198`.

[140] M. Bahr et al. *Herwig++ Physics and Manual*. *Eur.Phys.J.* C58 (2008), 639–707. `arXiv:0803.0883`.

[141] J. Alwall et al. *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*. *Eur.Phys.J.* C53 (2008), 473–500. `arXiv:0706.2569`.

[142] J. Alwall et al. *QCD radiation in the production of heavy colored particles at the LHC*. *JHEP* 0902 (2009), 017. `arXiv:0810.5350`.

[143] N. Lavesson et al. *Merging parton showers and matrix elements: Back to basics*. *JHEP* 0804 (2008), 085. `arXiv:0712.2966`.

[144] M. L. Mangano et al. *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b} + n$ jets as a case study*. *Nucl.Phys.* B632 (2002), 343–362. `arXiv:hep-ph/0108069`.

[145] M. L. Mangano et al. *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*. *JHEP* 0701 (2007), 013. `arXiv:hep-ph/0611129`.

[146] A. Buckley et al. *Rivet user manual*. *Comput.Phys.Commun.* 184 (2013), 2803–2819. `arXiv:1003.0694`.

[147] M. Goerner et al. *Differential Cross Sections for Top-Quark-Pair Production in the e/μ+Jets Final State at $\sqrt{s} = 8$ TeV in CMS*. Ph.D. thesis, Hamburg U. (2014).

[148] S. Chatrchyan et al. (CMS Collaboration). *Measurement of differential top-quark pair production cross sections in pp colisions at* $\sqrt{s} = 7$ *TeV*. *Eur.Phys.J.* C73 (2013)(3), 2339. `arXiv:1211.2220`.

[149] S. Chatrchyan et al. (CMS Collaboration). *Measurement of masses in the* $\mathrm{t\bar{t}}$ *system by kinematic endpoints in pp collisions at* $\sqrt{s} = 7$ *TeV*. *Eur.Phys.J.* C73 (2013), 2494. `arXiv:1304.5783`.

[150] V. Khachatryan et al. (CMS Collaboration). *Measurement of the differential cross section for top quark pair production in pp collisions at sqrt(s) = 8 TeV*. Tech. Rep. CMS-TOP-12-028 (2015). `arXiv:1505.04480`.

[151] W. Adam et al. *PAT: The CMS physics analysis toolkit*. *J. Phys. Conf. Ser.* 219 (2010), 032017.

[152] H. Kirschenmann (CMS Collaboration). *Jet performance in CMS*. Tech. Rep. CMS-CR-2013-325, CERN, Geneva (Oct 2013).

[153] S. Chatrchyan et al. (CMS Collaboration). *Measurement of jet multiplicity distributions in* $\mathrm{t\bar{t}}$ *production in pp collisions at* $\sqrt{s} = 7$ *TeV*. *Eur. Phys. J.* C74 (2015), 3014. [Erratum: Eur. Phys. J.C75,no.5,216(2015)], `arXiv:1404.3171`.

[154] S. Chatrchyan et al. (CMS Collaboration). *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*. *Phys.Rev.* D89 (2014)(1), 012003. `arXiv:1310.3687`.

[155] V. Khachatryan et al. (CMS Collaboration). *Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure*. *JHEP* 06 (2015), 080. `arXiv:1503.01952`.

[156] N. D. Gagunashvili. *Comparison of weighted and unweighted histograms*. *ArXiv Physics e-prints* (2006). `arXiv:physics/0605123`.

[157] *Theta framework web page*. `http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto/`.

[158] G. Cowan et al. *Asymptotic formulae for likelihood-based tests of new physics*. *The European Physical Journal C* 71 (2011)(2), 1554.

[159] S. Chatrchyan et al. (CMS Collaboration). *Search for vector-like quarks in final states with a single lepton and jets in pp collisions at* $\sqrt{s} = 8$ *TeV*. Tech. Rep. CMS-PAS-B2G-12-017, CERN, Geneva (2014).

[160] S. Chatrchyan et al. (CMS Collaboration). *Search for vector-like top quark partners produced in association with Higgs bosons in the diphoton final state.* Tech. Rep. CMS-PAS-B2G-14-003, CERN, Geneva (2014).

[161] G. Aad et al. (ATLAS Collaboration). *Analysis of events with b-jets and a pair of leptons of the same charge in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector* (2015). `arXiv:1504.04605`.

[162] S. Chatrchyan et al. (CMS Collaboration). *Inclusive search for a vector-like T quark with charge* $\frac{2}{3}$ *in pp collisions at* $\sqrt{s} = 8$ *TeV*. *Phys. Lett.* B729 (2014), 149–171. `arXiv:1311.7667`.

[163] G. Aad et al. (ATLAS Collaboration). *Search for pair and single production of new heavy quarks that decay to a Z boson and a third-generation quark in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*. *JHEP* 11 (2014), 104. `arXiv:1409.5500`.