

Searches for heavy resonances in all-jet final states with top quarks using jet substructure techniques with the CMS Experiment

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

EMANUELE USAI

aus Ivrea, Italien

Hamburg
2017

Gutachter der Dissertation:

Prof. Dr. Johannes Haller
Dr. Alexander Schmidt

Zusammensetzung der Prüfungskommission:

Prof. Dr. Elisabetta Gallo-Voss
Prof. Dr. Caren Ines Hagner
Prof. Dr. Johannes Haller
Prof. Dr. Gudrid Moortgat-Pick
Dr. Alexander Schmidt

Vorsitzende der Prüfungskommission:

Prof. Dr. Caren Ines Hagner

Datum der Einreichung:

27. Juli 2017

Vorsitzender Fach-Promotionsausschusses PHYSIK:

Prof. Dr. Wolfgang Hansen

Leiter des Fachbereichs PHYSIK:

Prof. Dr. Michael Potthoff

Dekan der Fakultät MIN:

Prof. Dr. Heinrich Graener

Abstract

While the Standard Model is very successful in describing subnuclear phenomena, it is not a complete theory of particle physics. Several new theories have been developed to address its issues. Many extensions of the Standard Model predict the existence of high-mass resonances. In some cases these resonances have an enhanced coupling to third generation quarks or to a hypothetical new generation of heavy non-chiral quarks. This thesis describes two searches for new phenomena compatible with these theories, in particular a search is presented for resonant top-antitop production, and the first search for heavy resonances decaying to a top quark and a heavy top quark partner T is shown.

The searches target the all-jets decay channels and use data collected by the CMS Experiment at the CERN LHC between 2012 and 2015 at a center-of-mass energy of 8 TeV and 13 TeV. Due to the high mass of the resonances considered, the final state particles have a high Lorentz-boost. To reconstruct the hadronic decay of the top quarks and W bosons, jet substructure techniques such as top quark and W boson tagging algorithms, and boosted b jet identification are employed. These algorithms are studied with a particular focus on their validation and performance assessment.

No signs of physics beyond the Standard Model are observed, but stringent limits are placed on the production of heavy resonances decaying to top-antitop quark pairs or a top quark and a heavy T quark.

Zusammenfassung

Zwar beschreibt das Standardmodell sehr erfolgreich subatomare Phänomene, jedoch ist es keine vollständige Theorie der Teilchenphysik. Viele neue Theorien wurden entwickelt mit dem Ziel, Lösungsansätze für dessen Mängel zu bieten. Viele Erweiterungen des Standardmodells sagen die Existenz von Resonanzen bei hohen Massen voraus. In einigen Fällen haben diese Resonanzen eine starke Kopplung an Quarks der dritten Generation, oder an eine hypothetische, neue Generation von schweren, nicht-chiralen Quarks. Diese Arbeit beschreibt zwei Suchen nach neuen Phänomenen, die mit diesen Theorien kompatibel sind. Insbesondere werden eine Suche nach resonanter Top-Antitop-Produktion präsentiert, sowie die erste Suche nach schweren Resonanzen, die zu einem Top-Quark und einem schweren Top-Quark Partner T zerfallen.

Die Suchen basieren auf den vollhadronischen Zerfallskanälen dieser Prozesse und verwenden die Datensätze des CMS-Experiments am LHC, die zwischen 2012 und 2015 bei Schwerpunktsenergien von 8 TeV und 13 TeV gesammelt wurden. Aufgrund der hohen Masse der hypothetischen Resonanzen haben die Teilchen im Endzustand einen hohen Lorentz-Boost. Daher werden Jet-Substruktur Methoden zur Identifikation von hadronischen Top-Quark oder W-Boson Zerfällen genutzt sowie Algorithmen zur Identifikation von geboosteten b-Jets. Es werden Studien präsentiert zur Validierung dieser Algorithmen und zur Beurteilung ihrer Leistungsfähigkeit.

Es werden keine Anzeichen von Physik jenseits des Standardmodells beobachtet, jedoch können hohe Ausschlussgrenzen auf die Produktion von schweren Resonanzen, die zu einem Top-Antitop-Paar oder zu einem Top-Quark und einem schweren T Quark zerfallen, gesetzt werden.

Sommario

Sebbene il Modello Standard descriva con precisione molti fenomeni a livello subnucleare, non è una teoria completa per la fisica delle particelle. Molte nuove teorie sono state proposte in risposta alle sue lacune.

Diverse estensioni del Modello Standard ipotizzano l'esistenza di risonanze ad alta massa. In alcuni casi, queste risonanze hanno degli accoppiamenti grandi con i quark di terza generazione o con una ipotetica nuova generazione di quark pesanti non chirali. Questa tesi presenta due ricerche di segnali compatibili con queste teorie, in particolare vengono presentate una ricerca di produzione risonante di top-antitop e la prima ricerca di risonanze ad alta massa che decadono in un quark top e in un partner pesante del quark top, T .

Le ricerche sono concentrate sugli stati finali contenenti esclusivamente getti adronici e usano dati raccolti dall'Esperimento CMS al LHC del CERN tra il 2012 e il 2015 ad una energia del centro di massa di 8 e 13 TeV. A causa dell'alta massa delle risonanze considerate, le particelle nello stato finale hanno un Lorentz boost molto elevato. Per ricostruire il decadimento adronico del quark top e del bosone W , tecniche di sottostruttura dei getti sono impiegate, come algoritmi di tagging del quark top e del bosone W , e identificazione dei b getti ad alto impulso trasverso. Questi algoritmi sono studiati con particolare attenzione alla loro validazione e alla valutazione delle prestazioni.

Nessun segno di fisica oltre il Modello Standard è osservato, ma limiti stringenti sono derivati sulla produzione di risonanze pesanti che decadono in coppie top-antitop o in un quark top e un quark pesante T .

Ai miei genitori.

Contents

Introduction	1
1 Theoretical introduction	3
1.1 A summary of the Standard Model of particles and interactions	3
1.1.1 Gauge symmetry and the Higgs mechanism	5
1.2 Issues with the Standard Model	6
1.3 Vector-like quarks	9
1.4 Composite Higgs theories	11
1.4.1 Little Higgs theories	12
1.5 Theories with an extra dimension	13
1.5.1 An effective theory for warped extra dimensions	16
2 The Large Hadron Collider and the CMS Experiment	21
2.1 The Large Hadron Collider	21
2.2 The Compact Muon Solenoid Experiment	25
2.2.1 The solenoidal magnet	26
2.2.2 The tracking system	29
2.2.3 Calorimetry	31
2.2.4 The muon detection system	34
2.2.5 The data acquisition and trigger system	35
2.2.6 The computing infrastructure	36
2.2.7 Upgrades	39
3 Event reconstruction	41
3.1 Luminosity calculation	41
3.2 The particle flow paradigm and object reconstruction	42
3.3 Hadronic jets	45
3.3.1 Jet clustering	45
3.3.2 Jet energy corrections	46
3.4 B-jet identification	47
3.5 Event simulation	48
4 Jet substructure tools	51
4.1 Jet grooming	51
4.2 The N-subjettiness algorithm	53
4.3 The HEPTOP Tagger algorithm	54

4.4	The CMS Top Tagger algorithm	59
4.4.1	Version used with 8 TeV data	59
4.4.2	Version used with 13 TeV data	60
4.4.3	Tagging W bosons	61
4.5	Triggers using jet substructure techniques	62
4.5.1	Introduction and trigger description	62
4.5.2	Performance assessment	64
5	Subject b tagging commissioning on 8 TeV data	71
5.1	Introduction	71
5.2	Data and simulation samples	71
5.3	Event reconstruction and selection	72
5.4	Commissioning studies	73
5.5	Conclusions	76
6	Commissioning of the HEPTopTagger with 8 TeV data	81
6.1	Algorithm performance in simulated events	81
6.1.1	Discriminating variables	81
6.1.2	Algorithm performance in simulation	82
6.2	HEPTopTAGGER performance in data	87
6.2.1	Introduction	87
6.2.2	Samples	88
6.2.3	Selection	88
6.2.4	Top-jet clustering at low p_T	93
6.2.5	Scale factor measurement	96
6.3	Conclusions	97
7	Search for heavy resonances decaying to $t\bar{t}$ with 8 TeV data	99
7.1	Introduction	99
7.2	Samples	100
7.2.1	Simulated event samples	100
7.2.2	Data samples	100
7.2.3	Trigger selection	102
7.3	Selection of events and optimization	104
7.3.1	Normalization of simulated samples	108
7.4	Background estimation	108
7.4.1	Kinematic distributions and event yields	112
7.5	Systematic Uncertainties	119
7.6	Summary of the very-boosted decay category	121
7.7	Statistical interpretation and results	122
7.7.1	Combination of HEPTopTAGGER and CMS Top Tagger analyses	126
7.8	Combination with leptonic final states	126
7.9	Conclusions	134
8	Search for heavy resonances decaying to a top quark and a vector-like T quark with 13 TeV data	135
8.1	Introduction	135
8.2	Samples	137
8.2.1	Modeling of the spin correlations in signal samples	140
8.3	Selection	144

8.3.1	Trigger	145
8.3.2	Final selection	147
8.4	Background estimate	152
8.4.1	Top-quark background estimate	152
8.4.2	Non-top background estimate	152
8.5	Systematic uncertainties	164
8.6	Results	168
8.7	Conclusions	170
Conclusions		177
A Tables of data and simulated samples		179
B Systematic uncertainties templates for the search for resonant $t\bar{t}$ production at 8 TeV		185
C Systematic uncertainties templates for the search for resonant top plus vector-like T quark production at 13 TeV		191
D Observed limits arranged on a bidimensional grid of masses		197

Introduction

The Standard Model (SM) is an extremely successful theory: its predictions span an energy range of several orders of magnitude, a wide range of phenomena, and have been successfully tested in many experiments, from the early days of accelerator physics to the LHC. In July 2012, the CMS and ATLAS collaborations at the CERN Large Hadron Collider (LHC), announced the discovery of a particle compatible with the SM Higgs boson [1,2]. This epochal discovery fulfills one of the design goals of the LHC and fills in the last, undiscovered, missing piece of the Standard Model. However, other than the fact that the SM does not describe the gravitational interaction, several issues plague the theory: cosmological observations suggest that the SM might be missing a sector of particles or interactions, while technical issues and naturalness arguments point towards the idea that the SM is a low-energy approximation of a more complete theory. Many theories have been proposed to solve these issues. Direct experimental evidence of phenomena unexplained by the SM is needed to eventually confirm any of the extensions of the SM. These theories include supersymmetry, composite Higgs models, and theories with extra dimensions. Therefore, the full exploitation of the discovery potential of the LHC is a top priority in the worldwide strategy of research in high energy physics. The experimental results obtained by the LHC will shape the future of theoretical research and drive the construction of new experimental facilities to continue the quest for knowledge of the fundamental components of matter.

This thesis presents two searches for beyond-the-standard-model (BSM) effects in the collision data collected by the CMS experiments at the LHC. The first is a search for top-antitop resonant production. Such resonances are a solid prediction of a wide range of BSM theories and make the analysis described a model-independent search for new physics. The search is conducted in the particularly challenging all-jet final state. To increase the sensitivity of the search a number of advanced reconstruction techniques have been used. These are collectively referred to as “jet substructure algorithms”: they are a class of algorithms that focuses on the reconstruction of decays of particles with high Lorentz-boost. Jet substructure tools have become increasingly important to analyze LHC data; this work will discuss in detail some of these techniques and their performance in data. The second analysis is a search for heavy resonances decaying to a top quark and a heavy top quark partner. This novel idea explores the hypothesis that a heavy resonance might predominantly decay to heavy partners of the SM quarks or a heavy partner and a SM quark. This can happen in theories with a composite Higgs boson or extra dimensions. In such a case the detection of the resonance might escape traditional searches. This search is also performed in the all-jet final state and makes use of jet substructure algorithms.

The organization of the thesis is detailed in the following. The first part of the thesis includes an introduction to the topic of searches for physics beyond the standard model, with

theoretical motivation and description of the experimental setup. Chapter 1 briefly summarizes the Standard Model and its issues, then composite Higgs boson and extra dimension theories are introduced as well as the models used for the interpretation of the experimental data; Section 2.1 describes the Large Hadron Collider, while Section 2.2 describes the CMS Experiment; details about the standard event reconstruction techniques are given in Chapter 3. The second part of the thesis describes more in detail the advanced reconstruction tools employed: Chapter 4 describes the jet substructure techniques used in the thesis and presents the implementation of jet substructure in the software trigger, Chapter 5 describes in detail b jet identification in boosted top decays and its validation in data, while Chapter 6 presents the implementation, optimization, and validation of top jet identification in CMS. The third and last part of the thesis presents two analyses of CMS data in search for physics beyond the standard model, using the jet substructure techniques described in the previous chapters. In particular, Chapter 7 describes the search for resonant top-antitop production in the all-jet final state. A short summary of the combination of the analysis with other decay channel is discussed. Finally, Chapter 8 presents the first search for heavy resonances decaying to a top quark and a vector-like quark.

Theoretical introduction

1.1 A summary of the Standard Model of particles and interactions

The Standard Model (SM) is a quantum field theory describing three of the four known interactions between elementary particles, namely the electromagnetic, the weak, and the strong forces. The gravitational force, relevant for the understanding of the large scale structure of the universe, is not included in the SM, and the development of a quantum field theory that can include gravity is still one of the main challenges of modern physics. The elementary particles included in the Standard Model are twelve fermions: three leptons, three neutrinos, and six quarks; and six bosons: the Higgs boson, and five gauge bosons mediators of the three fundamental forces. Electromagnetic interactions affect quarks and leptons except neutrinos; only quarks can interact by the strong force, while neutrinos can interact only by the weak force. A summary of the particles in the Standard Model and their properties is shown in Figure 1.1.

A remarkable property of the Standard Model is that electromagnetic interactions, responsible for the atom and molecular structure, and weak interactions, responsible for nuclear decays, are interpreted as two different low-energy manifestations of a single unified interaction, the electroweak force. The Higgs mechanism explains why the two forces appear to be so different in our universe, in particular why only the electromagnetic force is a long range interaction, while weak forces are relevant only at the nuclear scale. The Higgs field plays a crucial role for the electroweak symmetry breaking. The recent discovery of the Higgs boson at LHC has been a milestone that completed the experimental evidence of all particles predicted by the Standard Model [1, 2].

Quantum Chromodynamics (QCD) is the sector of the SM that describes the strong interactions between quarks and gluons, responsible for their binding in bound states like hadrons and nuclei. Due to asymptotic freedom, the strength of the strong interaction becomes weaker and weaker at high energy, so that the elementary constituents of protons and neutrons can be probed in deep inelastic scattering. Confinement occurs at low energy and, as a consequence, single isolated quarks cannot be found in nature. The mechanism of confinement are still not fully understood due to the failure of perturbation theory at low energy.

The photon, the gluon, and the neutrinos are massless in the SM. Quarks and charged leptons have masses ranging from hundreds of KeV to hundreds of GeV. A particular role in the Standard Model is played by the top quark, the most massive elementary particle currently known, with a mass of 173 GeV [4]. The top quark is crucial for the analyses presented in this thesis and it is used as a tool to search for physics beyond the Standard Model. Given its large mass, the top quark has a very short lifetime of about 10^{-25} s. The lifetime is shorter than the

Standard Model of Elementary Particles

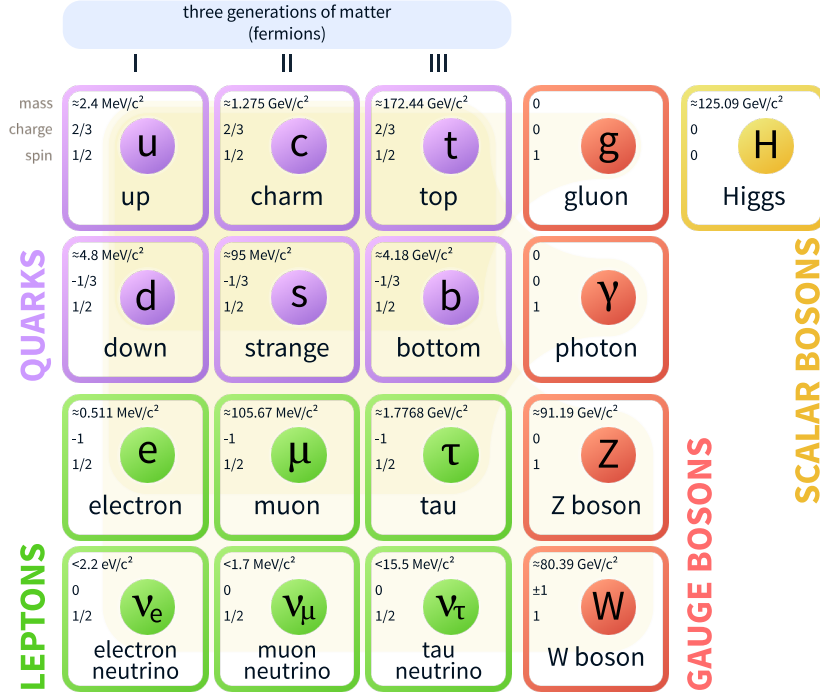


Figure 1.1: Summary of the particle content of the Standard Model [3].

average hadronization time, therefore the top quark decays before forming bound states (such as mesons and baryons). This particular feature allows for very precise measurements of its properties. The top quark was discovered by the D0 and CDF Collaborations in 1995 at the Fermilab Tevatron [5,6]. At the LHC the top quark is produced in pairs through gluon fusion and $q\bar{q} \rightarrow g$, or as a single particle through the weak interaction. The top quark pair cross section at the LHC is 832 pb (245 pb) at 13 TeV (8 TeV) [7], therefore a large number of $t\bar{t}$ events are produced, making the LHC the ideal machine to study the properties of the top quark. Given that the CKM matrix element $V_{tb} \approx 1$, the top quark decays almost exclusively to a W boson and a b quark. Due to its large mass and coupling to the Higgs boson, the top quark is an important direct and indirect probe for physics beyond the Standard Model. Figure 1.2 shows two representative Feynman diagrams for $t\bar{t}$ and single-top production at the LHC.

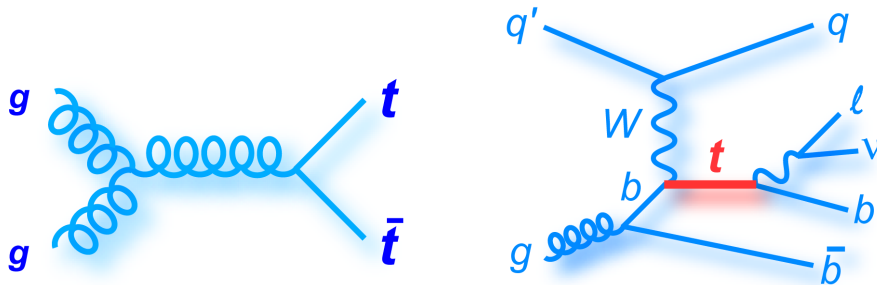


Figure 1.2: Representative Feynman diagrams for $t\bar{t}$ (left) and single-top t-channel (right) production at the LHC [8].

1.1.1 Gauge symmetry and the Higgs mechanism

The Standard Model is based on the principle of local invariance (gauge symmetry) [9]. The path integral is invariant under local transformations of fermion and boson matter fields of the form

$$\psi^a(x) \rightarrow \Omega(x)_{ab} \psi^b(x),$$

where $\Omega(x)$ is an element in a given representation of the Lie group $SU(N)$. For instance in QCD, the internal degree of freedom a of a quark field is commonly called “color”, and the gauge matrix $\Omega(x)$ is an element of $SU(3)$. Additional fields are required to fulfill the local invariance of the above transformation. The action of the Standard Model is invariant if the simple derivatives ∂_μ are replaced by covariant derivatives D_μ

$$D_\mu = \partial_\mu + ig A_\mu(x).$$

The gauge field $A_\mu(x) = A_\mu(x)^a T_a$ is a field of variables in the Lie algebra of the gauge group and T_a are the generators of the groups. The field $A_\mu(x)$ transforms as

$$A_\mu^a(x) \rightarrow \Omega^\dagger(x) A_\mu(x) \Omega(x) + \frac{i}{g} \Omega^\dagger(x) \partial_\mu \Omega(x),$$

where g is the gauge coupling, which represents the strength of the interactions between matter and gauge fields.

The gauge group of the Standard Model is $U(1) \times SU(2) \times SU(3)$. The symbol $\times$ denotes the tensor product of two groups, meaning that a given matter field can have in general a set of many different internal degrees of freedom (like quarks). Electroweak interactions are a unified theory modeled through the subgroup $U(1) \times SU(2)$. The quantum numbers associated are the isospin and the weak hypercharge. The Higgs mechanism breaks the symmetry of the vacuum down to the $U(1)$ symmetry of electromagnetism, already manifest in the Maxwell equations. A scalar field ϕ is introduced with a potential

$$V(\phi) = \frac{\mu}{2} \phi^2 + \frac{\lambda}{4!} \phi^4.$$

The field acquires a non-vanishing expectation value $\langle \phi \rangle \neq 0$ driven by the negative mass term μ . As a consequence, the constant field configuration $\phi_M = 0$ is a maximum and not a minimum of the potential. The true minimum of the potential is at $\phi_m = 6\mu/\lambda$, and perturbation theory has a defined behavior to handle quantum fluctuations around ϕ_m .

Another important consequence driven by the negative mass term, is that the kinetic term of the field ϕ :

$$D_\mu \phi D_\mu \phi,$$

that includes the covariant derivative and the interactions with the gauge fields, generates a mass term in the Lagrangian for the gauge bosons equal to

$$\langle \phi \rangle^2 A_\mu A_\mu.$$

Many other interaction terms are included, that will preserve the full gauge symmetry of the theory, while still allowing for vacuum solutions where the W and Z boson are effectively massive. In other words, the full $U(1) \times SU(2)$ gauge symmetry is only hidden by the Higgs mechanism, and not explicitly manifest in nature at low energy.

The remaining degree of freedom manifests itself in the form of a scalar neutral particle: the Higgs boson. The Higgs mechanism provides a mass term for the charged fermions as well, via an interaction between the Higgs field and the fermion, in the form of

$$\mathcal{L}_{\text{Yukawa}} = -G[\bar{\psi}_R \phi^\dagger \psi_L + \bar{\psi}_L \phi \psi_R] + h.c. = -m_f \bar{\Psi} \Psi - \frac{m_f}{v} H \bar{\Psi} \Psi,$$

where the mass of the fermion is $m_f = \frac{vG}{\sqrt{2}}$. This is called the Yukawa coupling and defines the coupling of the Higgs boson with fermions, which is proportional to the fermion mass. Thus the top quark has the largest coupling to the Higgs boson among all the fermions. In the expression above, the fermion fields are expressed in terms of a left-handed doublet, composed for instance by a lepton and the corresponding neutrino. Additionally, a right-handed singlet is present. The Higgs doublet ϕ ensures the correct gauge invariance of the Yukawa interaction term. This particular combination of left- and right-handed particles is required to correctly take into account the parity violations of weak interactions. In the Lagrangian of the Standard Model, neutrinos appear only with their left-handed components.

A boson with properties compatible with that of the SM Higgs boson has been discovered in 2012 at the LHC [1,2]. Over the following five years the larger amount of data collected allowed to perform precise measurement of the properties of the particle such as the cross section, branching fractions, mass, spin, and width. These measurements continue to be compatible with the hypothesis of a SM Higgs boson, but do not completely exclude the possibility of deviations with respect to the SM hypothesis [10]. Figure 1.3 shows a comparison of the measured couplings of the Higgs boson for different particles in function of their mass, compared to the SM expectation and the best fit of the measured values.

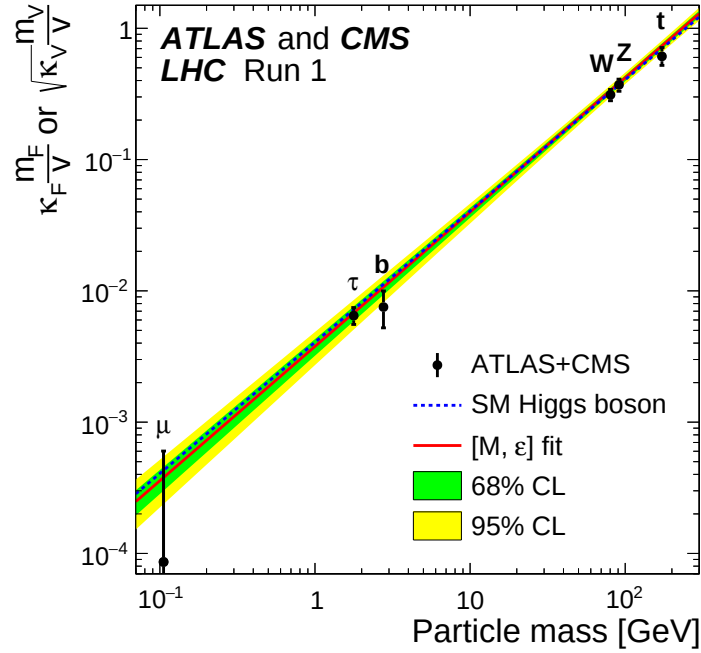


Figure 1.3: Comparison of the measured couplings of the Higgs boson for different particles in function of their mass, compared to the SM expectation and the best fit of the measured values [10].

1.2 Issues with the Standard Model

Despite the recent experimental success of the SM, there are many evidences that it cannot be the complete theory of all fundamental forces of the universe. The most striking shortcoming of the SM is that it does not include the gravitational interaction. Gravity is non renormalizable in quantum field theory, at least in a perturbative framework. The development of

a theory of quantum gravity is one of the main open problems in fundamental theoretical physics. Radical approaches such as String Theory and Loop Quantum Gravity are being studied in order to provide a suitable candidate for a theory describing all the fundamental forces [11]. The effects of quantum gravity are expected to be relevant however only near the Planck scale, for black holes and for the description of the very early instants of the Big Bang.

Measurements of cosmological quantities such as the rotational velocities of the galaxies shows that an unknown form of matter, called “dark matter” (DM) [12], composes a large amount of the mass present in the universe. DM must interact very weakly with ordinary matter and no particle in the Standard Model has the properties required to be a candidate for dark matter. Therefore an additional sector of the Standard Model is needed to explain the large-scale dynamics of the universe. Many experiments in the field of astroparticle and collider physics are trying to shed light on the properties of dark matter. Another issue of cosmological nature is the large discrepancy between the measurement of the cosmological constant (leading to the prediction that most of the energy in the universe is composed by dark energy) and the prediction of the vacuum energy in quantum field theory.

Apart from the elusive quantum properties of gravity and dark matter, the SM does not explain why the part of universe that can be probed is composed almost exclusively by particles, and not antiparticles. The SM does provide a source for matter-antimatter asymmetry (often referred to as baryon asymmetry) in the form of a CP-violating complex phase in the CKM matrix. However, the amount of CP violation in the SM is insufficient to explain the observed baryon asymmetry, therefore a source of baryon asymmetry beyond the Standard Model is assumed to exist.

Neutrinos are considered as massless particles in the SM, therefore the discovery of neutrino oscillations requires the inclusion of mass for these particles [13,14]. Moreover several experiment are being conducted to determine whether the neutrinos are Dirac or Majorana particles and their mass hierarchy.

Assuming that the Standard Model is the correct theory up to the Planck scale, and assuming the currently measured values for the top quark and the Higgs boson, the theory could possibly allow the universe to be in a metastable phase of the Higgs field, that is, a non-global minimum of the Higgs potential. This could lead to the possibility of a quantum phase transition of the universe to the true vacuum [15,16]. The predicted mean-life of the metastable vacuum should not be smaller than the current age of the universe.

The Standard Model does not explain the values of the masses of its particles, their large difference and the reason for the number of lepton and quark generations. The very large difference between the strength of the different forces is also unexplained. Therefore the SM has a conspicuous number of free parameter in the model, that leads to believe that the SM might be the just an “effective” low-energy approximation of a more complete theory.

Finally, since the Higgs boson is a scalar particle, the radiative corrections to its mass are quadratically divergent in function the energy cut-off Λ . Chiral symmetry prevents large quantum corrections to the mass of fermions, but there are no symmetries that protect the mass of a scalar. In quantum field theory language, the mass of a fermion renormalizes multiplicatively, while the mass of a scalar has also an additive renormalization constant. The energy cut-off marks the scale at which the SM loses prediction power in favour of a more complete theory. The radiative corrections to the Higgs boson mass can be written as:

$$m_H^2 = m_{H,\text{bare}}^2 - \frac{3}{8\pi^2} y_t^2 \Lambda^2 + \frac{1}{16\pi^2} g^2 \Lambda^2 + \frac{1}{16\pi^2} \lambda^2 \Lambda^2,$$

where the first correction term is relative to the top loop correction, the second to the W/Z bosons correction, and the third to the Higgs self-coupling loop. For cutoff energies around

10 TeV the corrections become very large and a fine tuning of the parameters undergoing renormalization at a per-cent level is required to recover the small mass of the Higgs boson. Pushing the validity of the SM beyond 10 TeV quickly leads to levels of fine tuning that appear very unnatural [17]. This is referred to as the little hierarchy problem.

A similar problem of fine tuning happens when considering the radiative corrections to the Higgs boson mass due to GUT-scale particles (i.e. the ones appearing at the scale of the unification of the EW and strong forces). The GUT radiative corrections are large because of the large mass of the participating fields, but the fine tuning has to happen at the EW scale which is approximately 28 order of magnitude smaller. This is referred to as the big hierarchy problem.

Several theories have been proposed to address the little hierarchy problem: Supersymmetry (SUSY) is a class of theories introducing a symmetry between bosons and fermions [18]. The corresponding conserved charge, given by a fermion operator Q , transforms a boson in the corresponding supersymmetric fermionic partner, and vice versa:

$$Q|\text{fermion}\rangle = |\text{boson}\rangle, \quad Q|\text{boson}\rangle = |\text{fermion}\rangle.$$

In fully supersymmetric theories, every particle in the SM has a supersymmetric partner with the opposite spin statistics. The radiative correction to the Higgs boson mass from SM particles is cancelled by the contribution from the corresponding supersymmetric partner. Figure 1.4 shows two diagrams contributing to the Higgs boson mass radiative corrections that cancel each other's divergent contribution in SUSY.

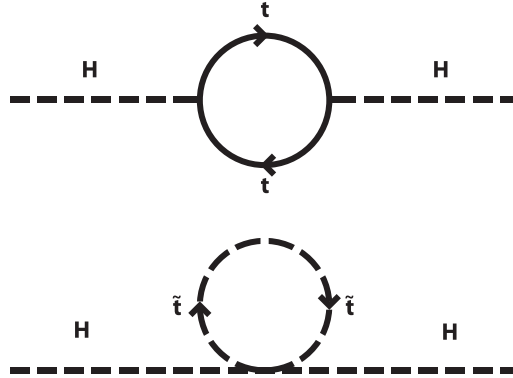


Figure 1.4: Two diagrams of Higgs boson mass radiative corrections that cancel each other's divergent contribution in SUSY (top quark and stop squark) [19].

SUSY theories also provide candidates for dark matter, however supersymmetric particles have not been observed in nature at masses identical to their SM partners, therefore if SUSY exists it must be some sort of broken symmetry, where the breaking could be either spontaneous, explicit or driven by quantum anomalies. Searches for supersymmetric partners at higher masses have been conducted at LEP, Tevatron, and the LHC, and are still ongoing. No sign of supersymmetry has been found so far and the experimental results point at masses of the supersymmetric partners in the TeV range, at least.

Two other approaches are commonly pursued to address the little hierarchy issue: composite Higgs theories and theories with extra dimensions. These theories often feature heavy neutral resonances that are the main target of the searches presented in this thesis, and will be described more in detail in Sections 1.4 and 1.5.

1.3 Vector-like quarks

A possible extension of the SM quark sector is to assume the existence of a fourth generation of quarks. It is natural to assume that the new quarks couple to the electro-weak force in the same way as the other generations, that is in a chiral way that breaks the symmetry between the left-handed and right-handed components of the spinor. In such a case, the fourth generation quarks receive mass through the Higgs mechanism, leading to a Yukawa coupling proportional to the mass of the quark itself. Given that a heavy fourth generation is expected to be significantly heavier than the top quark, these quarks have a large coupling to the Higgs boson, which would modify several parameters of the Higgs boson. For example, the branching fraction of the Higgs boson to two photons, which depends on a fermion loop, would change significantly with the presence of a heavy fermion with a large Higgs coupling running in the loop. The precise measurement of the properties of the Higgs boson exclude the presence of a chiral fourth generation of quarks. However, it is possible to assume that the quarks behave in a non-chiral, or vector-like, way. In this case the quarks do not have the classic $V - A$ structure of the coupling with the electroweak force, but only the vector component V , that does not contain the γ^5 term, and therefore treats the left-handed and the right-handed components of the spinor in the same way. These are called vector-like quarks (VLQs) [20, 21]. The consequence of this fact is that the Higgs mechanism is not necessary anymore to provide a mass term to the vector-like quarks, but a simple mass term $-m\bar{\psi}\psi$ is used instead. The absence of a Yukawa coupling for the vector-like quarks greatly reduces their impact on the Higgs boson parameters such as the branching fraction to photons. Therefore VLQs are compatible with the measured properties of the Higgs boson. Generally speaking, however, the presence of VLQ and their mixing with the other three generations does have an impact on other parameters of the SM, such as the Z boson branching fractions and its forward-backward asymmetry. Some of these parameters are very precisely measured and constrain the coupling parameters of the VLQs. Additionally, the VLQs introduce a source of flavour-changing neutral currents (FCNCs) at tree level, which poses tight constraints on models containing a VLQ sector as well.

There are several possible representation of VLQs that leads to well-behaved models (renormalizable lagrangian, and well-defined quantum numbers for the quarks) [20, 21]. For example singlet representations

$$T_{L,R} \quad B_{L,R} ,$$

doublets

$$(X, T)_{L,R} \quad (T, B)_{L,R} \quad (B, Y)_{L,R} ,$$

and triplets

$$(X, T, B)_{L,R} \quad (T, B, Y)_{L,R} .$$

Where the T quark has a charge of $2/3$ (sometimes referred to as T' quark), the B of $-1/3$ (or B'), the X of $5/3$ (or $T_{5/3}$), and the Y of $-4/3$. Other more complicated representations are possible by assuming the existence of additional VLQs. Generally speaking, the mixing with the other generation of quarks leads to the following mass scale:

$$m_T \geq m_X, \quad m_B \geq m_Y ,$$

while the mass relationship between the T and the B is model dependent.

The vector-like quarks mix with the SM quarks and with other VLQs, however due to the constraints imposed by the SM measurements and the mass scale of the VLQs, the coupling to third generation SM quarks is favoured. Therefore in the literature describing the

phenomenology of the VLQ, often only the decay modes to third generation are considered [20,21]. The main decay modes for the T quark are bW , tH , and tZ , while for the B quark the decay modes are bH , bZ , and tW . The decay modes for the X and Y quarks are bW and tW , respectively. Figure 1.5 shows a graphical representation of the decay modes of the vector-like quarks. Vector-like quarks are produced at the LHC in pairs mediated by a gluon,

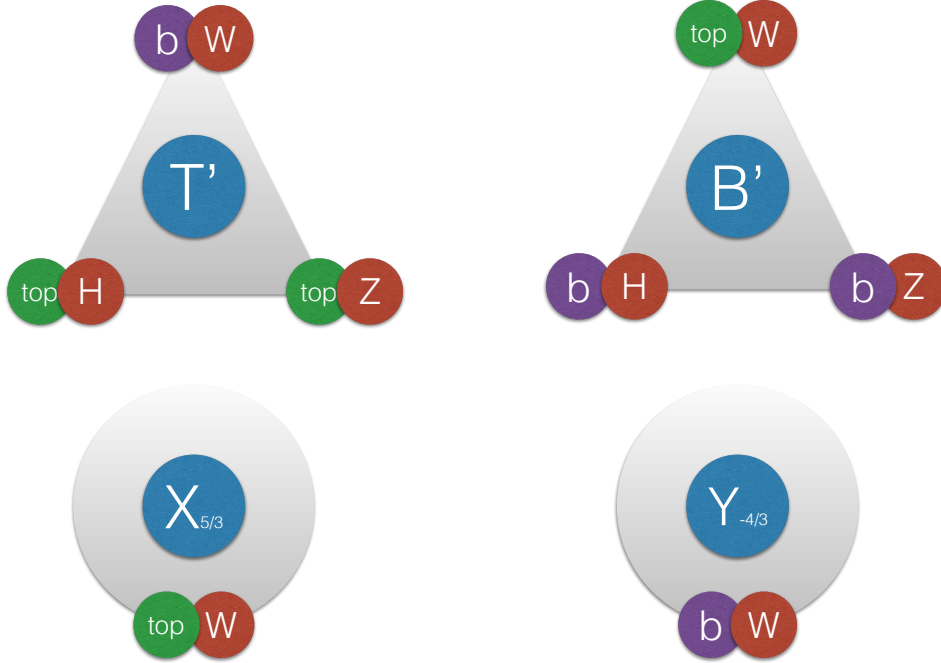


Figure 1.5: Graphical representation of the decay modes of the vector-like quarks.

similarly to $t\bar{t}$, in a pure QCD process. Additionally VLQs can be produced singly through the coupling with vector bosons (thus an EW process), similarly to single-top production in the t-channel. Figure 1.6 shows two representative Feynman diagrams for pair and single VLQ production at the LHC.

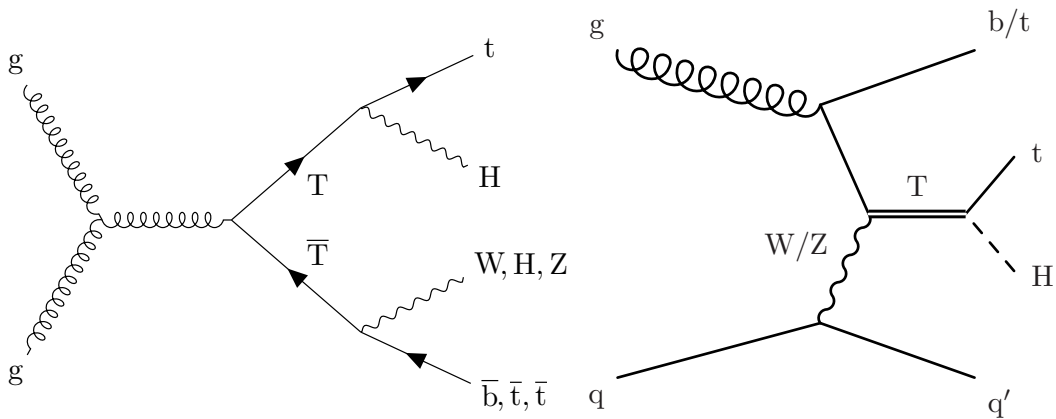


Figure 1.6: Representative Feynman diagrams for pair (left) and single (right) VLQ production at the LHC [22,23].

Single-VLQ production depends on the coupling strength of the VLQ to the EW sector. Figure 1.7 shows the cross section of different VLQs for some production modes and for

different multiplets, at the LHC. Single production is the preferred production mode at the LHC at 13 TeV, for high masses of the VLQs.

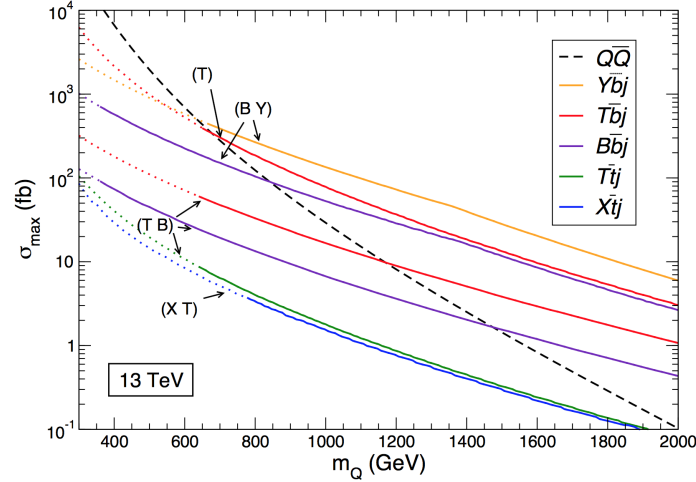


Figure 1.7: Maximum single and pair heavy quark production cross sections at the LHC for different multiplets, in function of the mass of the particle [20].

Direct searches for VLQ pair production at the LHC exclude masses between 800 and 1300 GeV, depending on the quark and final state considered [22–27]. Vector-like quarks appear often in theories beyond the Standard Model, such as the ones described in the following sections.

1.4 Composite Higgs theories

An elegant solution to the problem of fine tuning of the Higgs boson mass can be found if the Higgs boson itself is not considered as a fundamental scalar particle but rather as a composite bound state made of some other elementary constituents. From an historical perspective, a similar situation occurred when the pion was discovered in the previous century. The pion was initially thought as the mediator of the nuclear interactions between protons and neutrons, but later it was discovered that the pion is just only one of the many other existing mesons, and that both mesons and baryons are bound states of quarks and gluons. The pion plays an important role only in the low energy effective lagrangian of QCD, in the framework of chiral perturbation theory, where a triplet of light pseudoscalar particles appears as would-be Goldstone bosons of the spontaneous chiral symmetry breaking.

The key observation is that electroweak symmetry would be broken anyhow even in the case where the Higgs mechanism would not be there. Chiral condensation in QCD, given by a non-vanishing expectation value of the light quark bilinear

$$\langle \bar{u}u + \bar{d}d \rangle \neq 0,$$

would break the electroweak symmetry and would provide a mass to the vector bosons proportional to

$$M^2 \propto g^2 F_\pi^2 / 4,$$

being F_π the pion decay constant. Pions would become effectively the longitudinal mode of the vector bosons, without the need of the extra degrees of freedom of the Higgs field. The

mass of the vector boson would be however only of a few hundreds MeV, but it could be possible to imagine the existence of an another QCD-like theory whose typical scale is of the order of ≈ 1 TeV, instead of ≈ 1 GeV as in QCD. The theory corresponding to the scenario outlined here is called technicolor [28–30]. Despite its appealing structure, many problems arise when realistic mechanism to provide the mass to the Standard Model fermions are considered. The role of the Higgs boson is in fact twofold: on one side, it breaks electroweak symmetry, on the other side, it provides a mass to leptons and quarks by Yukawa couplings. Precision experimental measurements on the FCNCs have strongly constrained the basic technicolor theory, that has been definitely ruled out after the discovery of the Higgs boson, since a QCD-like theory has no possible candidate for a light scalar resonance. A possible solution to the experimental constraints of technicolors which has been intensively investigated is the so-called “extended technicolor”.

1.4.1 Little Higgs theories

Another alternative to classic technicolor models is the “little composite Higgs model” [31–34], similarly to previous models the Higgs is a Goldstone boson arising from an additional strong force sector, however the new force is not introduced explicitly, but through a new global symmetry breaking. Moreover this class of theories predicts a low mass Higgs (hence the term “little”) compatible with the boson discovered at the LHC. The global symmetry group G should be large enough to include at least the EW symmetry group. The Goldstone boson of the theory would be the Higgs boson, which is massless at tree level. The Goldstone Higgs can however acquire mass if the new global symmetry is explicitly broken. Generally speaking, a nontrivial group structure is needed to recover the SM in the low-energy limit, cancel the quadratic divergences of the Higgs boson mass corrections at tree level, and assign the correct masses to the gauge bosons and the Higgs. Realistic models often implement the “collective symmetry breakdown” [35, 36], which consists in assuming that the global symmetry G contains two copies of $SU(2) \times U(1)$:

$$G \supset H_1 \times H_2 = [SU(2)_1 \times U(1)_1] \times [SU(2)_2 \times U(1)_2] .$$

The group is arranged in a way that when part of G is broken, the Higgs boson mass corrections are still protected by the global symmetry. Two sets of Goldstone scalar fields are introduced, one for every copy of H_i , some of the degree of freedoms are absorbed by the gauge bosons to generate mass, while the remaining ones form the Higgs bosons. The gauge boson sector of the theory will contain SM gauge bosons and a number of additional bosons with a mass at the scale of the vacuum expectation value of the new symmetry breaking. These resonances are often called Z' and W' in the literature. The Yukawa sector of the theory will contain an additional quark sector in order to cancel the top loop corrections to the Higgs boson propagator. The Z' and W' bosons arising from these theories tend to behave differently with respect to the ones in weakly-coupled extensions of the SM: they tend to be broader, since they interact strongly with SM bosons. Some models implement “partial compositeness” [35–37], that is every SM fermion mixes with a heavy partner arising from the composite dynamics. This leads to SM fermions coupling to the new sector proportionally to their mass, therefore the new heavy resonances preferably decay to the heaviest fermions available, that is the top quark and the new heavy fermionic sector. To summarize, two important phenomenological features of modern composite Higgs models are the presence of heavy resonances that have an enhanced coupling to top quarks and heavy fermions, and vector-like spin-1/2 heavy fermions. In particular, bosonic resonances commonly referred to as Z' and W' arise when the group G of the grand unification theory (GUT) contain a subgroup other than that of the Standard Model ($SU(3) \times SU(2) \times U(1)$).

A minimal composite effective theory of the Higgs boson, described in Ref. [38], is used for the interpretation of the analysis in Chapter 8. The model is based on the global symmetry $G = SO(5) \times U(1)_X$ as an approximate group of the UV theory. The global symmetry spontaneously breaks to $H = SO(4) \times U(1)_X \sim SU(2)_L \times SU(2)_R \times U(1)_X$. The model features heavy resonances $\rho_{L/R}^\pm$ and $\rho_{L/R}^0$ arising from the strong dynamics (reducing to $\rho_{L/R}^X$ for a singlet representation). Additionally a VLQ sector is present and is organized as:

$$\Psi = \frac{1}{\sqrt{2}} \begin{bmatrix} iB - iX_{5/3} \\ B + X_{5/3} \\ iT + iX_{2/3} \\ -T + X_{2/3} \end{bmatrix}.$$

The quadruplet Ψ decomposes in two doublets under $SU(2) \times U(1)$: (T, B) and $(X_{5/3}, X_{2/3})$, where $X_{2/3}$ is a second top-like VLQ.

The model focuses on the description of the phenomenology of heavy vector resonances, particularly the interactions with top quark partners. The results of the analysis in Chapter 8 are compared with the production cross section of the neutral spin-1 resonance ρ_L^0 decaying to a top quark and the top-like VLQ T . This model assumes T branching fractions to tH/tZ channels of 50/50%. Benchmark values of the model parameters are used, based on constraints from experimental observation and the choice to expose decays of the heavy resonances to VLQs: $y_L = c_3 = c_2 = 1$, and $g_{\rho_L} = 3$. Where g_{ρ_L} regulates the interaction strength between the vector resonances and the other particles in the composite sector, c_3 regulates the interaction strength between the heavy resonance and the heavy top partners, c_2 regulates the interaction between the top quark and the heavy top partners, and y_L regulates the Yukawa couplings. Figure 1.8 reports the decay branching fractions of the ρ_L^0 resonance in function of its mass for two different benchmarks of the model parameters.

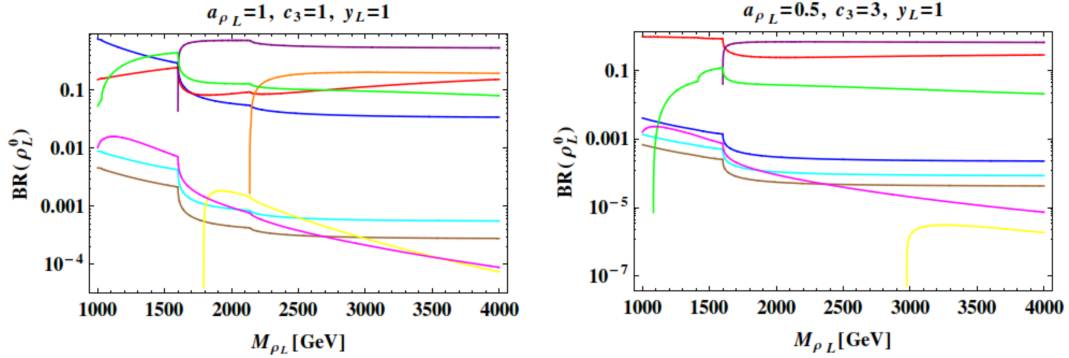


Figure 1.8: Decay branching fractions of the ρ_L^0 as a function of the resonance mass for $g_{\rho_L} = 3$, $m_\Psi = 800 \text{ GeV}$ and two sets of the model parameters. The different decay modes are: $WW + Zh$ (blue), $t\bar{t} + b\bar{b}$ (red), l^+l^- (brown), $u\bar{u} + d\bar{d}$ (cyan), $X_{5/3}X_{5/3}^- + X_{2/3}X_{2/3}^-$ (purple), $T\bar{T} + B\bar{B}$ (orange), $X_{2/3}\bar{T}$ (yellow), $X_{2/3}\bar{t}$ (magenta), and $T\bar{t} + B\bar{b}$ (green) [38].

1.5 Theories with an extra dimension

The idea of Extra Dimensions (ED) dates back to shortly after the proposal of Einstein's theory of general relativity, by Kaluza and Klein [39,40]. One of the appealing features of such theories is that it can explain the gauge hierarchy problem, that is the large difference in strength of the gravitational force and the other forces. One can start assuming that there

are $D = 4 + \delta$ dimensions and that gravity acts in $D - 1$ spatial dimensions, therefore the gravitational interaction must decrease $\sim 1/r^{2+\delta}$. This way one can rewrite Newton's laws as:

$$F(r) = \begin{cases} -G_N \frac{m_1 m_2}{r^2} = -\frac{1}{M_{Pl}^2} \frac{m_1 m_2}{r^2}, & D = 4 \\ -G_D \frac{m_1 m_2}{r^{2+\delta}} = -\frac{1}{M_D^{2+\delta}} \frac{m_1 m_2}{r^{2+\delta}}, & D = 4 + \delta \end{cases}$$

Where M_{Pl} and G_N are the Planck mass and the Newton's constant, respectively; M_D is the energy scale of the gravitational force in the D -dimensional theory. To solve the hierarchy issue, M_D should not be too different from the EW scale, e.g. around 1 TeV, therefore it can assumed that the gravitational interaction in the ED saturates at a scale $R \sim 1/M_D$, and for $r \gg R$ Newton's law is recovered. However, for $r \ll R$ the theory violates Newton's law. Taking the large r limit, assuming the saturation of ED gravity, and comparing to the 4D limit of the theory:

$$\frac{1}{M_D^2} \frac{m_1 m_2}{r^2} \frac{1}{(M_D R)^\delta} = \frac{1}{M_{Pl}^2} \frac{m_1 m_2}{r^2}.$$

From this expression one can extract the saturation distance scale R in function of the number of extra dimension. $\delta < 3$ is excluded by small-scale measurements of gravity, but under these assumptions a larger amount of extra dimensions is not excluded. In this toy model, SM particles should not propagate to the extra dimensions otherwise its effect would appear well below the TeV scale, therefore only gravity and the graviton live in the ED. Generally speaking, when probing the phenomenology of ED models it is customary to limit the study to the effects of one extra dimension. In the ED language, the whole spacetime is referred to as the "bulk", while a limited subspace of the multidimensional space where some of the dynamics may be confined is commonly called the "brane". In more complicated models, the particles might reside either on the bulk (universal ED models) or on a brane. The original Kaluza-Klein model assumed the extra dimension to be size of the Planck scale, however models trying to solve the hierarchy problem often feature large or warped extra dimensions [41, 42].

Extra dimensions have been extensively considered in the recent years not only in string theory, but also as a possibility to explain the properties of electroweak symmetry breaking and of the Higgs boson dynamics, as mentioned above. Assuming only one ED, the momentum of a particle living in the extra dimension appears as a mass in the 4D space:

$$E = \sqrt{\vec{p}^2 + p_5^2 + m^2} = \sqrt{\vec{p}^2 + m_{KK}^2}$$

Where m_{KK} is the Kaluza-Klein mass. The fifth component of the momentum p_5 is a continuous real number if the ED is infinite, however if the ED is a "compactified" closed circle, the momentum can assume an infinite number of discrete values. This translates to a spectrum of particles with different masses, referred to as the Kaluza-Klein tower.

Let τ denotes the coordinate of the fifth dimension and x the coordinate in the usual four-dimensional Minkowski space. Assume now that the fifth dimension is compactified with periodic boundary conditions and has length R . In this case the spacetime coordinates can be written as $X^M = (x^\mu, \tau)$, with $0 \leq \tau \leq 2\pi R$. The vacuum metric can be written as:

$$ds^2 = \eta_{MN} dX^M dX^N = \eta_{\mu\nu} dx^\mu dx^\nu - d\tau^2.$$

Considering just a simple massless scalar theory whose action is

$$S = \int d^4x d\tau \partial_\mu \phi^\dagger(x, \tau) \partial_\mu \phi(x, \tau) + \partial_\tau \phi^\dagger(x, \tau) \partial_\tau \phi(x, \tau),$$

the field $\phi(x, \tau)$ can be expanded in Fourier modes as

$$\phi(x, \tau) = \sum_n \phi_n(x) e^{(-i2\pi n\tau/R)}.$$

The last term with the partial derivatives ∂_τ of the action given above now reads as

$$- \sum_n \frac{4\pi^2 n^2}{R^2} \phi_n^\dagger(x) \phi_n(x),$$

the integral over the τ variable is trivial, and the action reduces to

$$S = \int d^4x \sum_n \partial_\mu \phi_n^\dagger(x) \partial_\mu \phi_n(x) - \frac{4\pi^2 n^2}{R^2} \phi_n^\dagger(x) \phi_n(x),$$

that is the action for an infinite number of scalar fields whose mass is $2\pi n/R$. The spectrum of the effective four dimensional theory will therefore contain a massless particle together with a tower of infinitely many scalar particles whose mass is linearly increasing and inversely proportional to R . If R is sufficiently small, all the massive scalars could be heavy enough to not be detected. The result can be generalized to a number of additional dimensions, where a field ϕ with a 4D-mass m_0 generates a KK tower of particles with masses

$$m^2 = m_0^2 + \frac{n_5^2}{R_5^2} + \frac{n_6^2}{R_6^2} + \dots$$

where R_i is the size of the i -th dimension. Realistic models for electroweak symmetry are based on a Yang-Mills theory in five dimensions. As before, fermion and gluon fields are Fourier expanded.

In the previous calculation, the ED is assumed to be compactified, as no evidence of extra dimensions is currently found in nature. However, theories referred to as Randall-Sundrum models do not assume the ED to be compactified, instead a warped space is introduced (hence the name warped extra dimension). The starting point is a 5D space where the ED contains only vacuum energy:

$$ds^2 = \eta_{MN} dX^M dX^N = e^{-2\sigma(\tau)} g_{\mu\nu} dx^\mu dx^\nu - d\tau^2,$$

where $e^{-2\sigma(\tau)}$ is the “warp factor”. The equation has a form similar to the Friedmann-Robertson-Walker (FRW) metric for a vacuum-energy dominated universe, transposed to 5D, where matter does not permeate the fifth dimension and the exponent of the warping factor plays the role comparable to that of a 5D cosmological constant. Assuming matter is localized in two branes at $\tau_1 = 0$ and $\tau_2 = \pi R$, it can be shown that the metric is a solution of Einstein’s equations for $\sigma(\tau) = k|\tau|$. Therefore the vacuum metric solution of the model can be written as:

$$ds_5^2 = e^{-2k\tau} \eta_{\mu\nu} dx^\mu dx^\nu - d\tau^2, \quad 0 \leq \tau \leq \pi R,$$

leading to, in the two branes:

$$(ds_4^{(1)})^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \quad (ds_4^{(2)})^2 = e^{-2k\pi R} \eta_{\mu\nu} dx^\mu dx^\nu.$$

These are Minkowski metrics with a null 4D cosmological constant. The first brane is called the Planck brane, while the second is called the TeV brane, describing the physics at the TeV scale. The gravitational interaction is strong in the Planck brane but suppressed by the

warping in the TeV brane. Assuming a 4D Higgs boson is located in the TeV brane, the action of the Higgs field can be written as:

$$S_{\text{Higgs}} = \int d^4x \sqrt{-g^{(2)}} \left[g^{(2)\mu\nu} \partial_\mu H \partial_\nu H - \lambda (|H|^2 - v_0^2)^2 \right] .$$

Substituting the expression for the interaction with gravity in the TeV brane:

$$g_{\mu\nu}^{(2)} = e^{-2k\pi R} g_{\mu\nu}^{(0)} ,$$

and renormalizing the Higgs field $H' = H e^{-k\pi R}$, the physical renormalized VEV becomes:

$$v = e^{-k\pi R} v_0 .$$

The physical Higgs VEV is scaled down by a factor $e^{-k\pi R}$ with respect to the original VEV. Assuming the original Higgs VEV is at the Planck scale $v_0 \sim M_{Pl}$, the physical VEV is naturally brought to the weak scale by a $kR \sim O(10)$ scaling, hence removing the hierarchy problem. Any mass parameter from the bulk theory will be scaled down by the same factor in the TeV brane. The quartic coupling of the Higgs boson λ is left untouched by the renormalization of the field. This is known as the RS1 scenario.

Generally speaking, the phenomenology of warped dimension theories will include gravitons and radions (scalar field emerging as quantum excitation of the extra-dimensional part of the metric). However, if one allows the other SM particles to propagate in the bulk, KK modes of these particles will be present. In this scenario, heavy fermions are often present as KK excitations of SM quarks. In models where the Higgs boson is a Goldstone boson of a larger symmetry group, as explained in Section 1.4, the KK top quark is essential to cancel the quadratic divergences of the SM top quark in the Higgs boson mass corrections, hence addressing the little hierarchy issue. A particularly robust prediction of such models is the presence of KK excitations of the gluon with masses in the TeV range [36]. Many theoretically well-motivated models feature mass-dependent couplings to the KK gluon, where the KK gluon has the largest coupling to the top quark and its heavy partners.

The results of the search for resonances presented in Chapter 7 are interpreted in the framework of RS1 models providing stringent limits on the production of the the first KK excitation of the gluon [43].

1.5.1 An effective theory for warped extra dimensions

The search for resonances presented in Chapter 8 is also interpreted in the framework of KK theories with a warped dimension [44]. The model proposed in Ref. [44] is an effective theory with one warped extra dimension that considers only the lowest-energy spin-1 and spin-1/2 resonances. The model describes, in particular, the decays of the lightest Kaluza–Klein excitation of the gluon, G^* , to one SM particle and one heavy fermion. In some configurations the KK gluon has a large decay width and its detection would be experimentally challenging. The model, while being specifically tailored for warped ED theories, has a broader application as a benchmark for BSM theories containing a heavy spin-1 color octet coupling to heavy partners of third generation SM quarks. The effective model is build using the partial compositeness paradigm explained in Section 1.4, where the heavy particles arise from a new strong sector and couple to SM particles through linear mixing terms. The Higgs boson is also a composite state. In this scenario the coupling of the G^* to heavy fermions is typically large, thus if kinematically allowed the heavy gluon will predominantly decay to heavy fermions or a combination of SM fermions and heavy fermions. The effective theory is based on a

global symmetry $SU(3)_c \times O(4) \times U(1)_X$, with $O(4) \supset SO(4) \sim SU(2)_L \times SU(2)_R$. The non-composite sector of the model contains the SM without the Higgs. The composite sector contains the heavy gluon G^* transforming as $(\mathbf{8}, \mathbf{1}, \mathbf{1})_0$ under $SU(3)_c \times SU(2)_L \times SU(2)_R \times U(1)_X$, the composite Higgs boson

$$\mathcal{H} = \begin{bmatrix} \phi_0^\dagger & \phi^+ \\ -\phi^- & \phi_0 \end{bmatrix}$$

transforming as $(\mathbf{1}, \mathbf{2}, \mathbf{2})_0$, and a set of vector-like quarks:

$$\begin{aligned} \mathcal{Q} &= (\mathbf{3}, \mathbf{2}, \mathbf{2})_{2/3} = \begin{bmatrix} T & T_{5/3} \\ B & T_{2/3} \end{bmatrix}, \quad \tilde{T} = (\mathbf{3}, \mathbf{1}, \mathbf{1})_{2/3}, \\ \mathcal{Q}' &= (\mathbf{3}, \mathbf{2}, \mathbf{2})_{-1/3} = \begin{bmatrix} B_{-1/3} & T' \\ B_{-4/3} & B' \end{bmatrix}, \quad \tilde{B} = (\mathbf{3}, \mathbf{1}, \mathbf{1})_{-1/3}. \end{aligned}$$

In complete theories, other composite states will be present, but these will not be considered here for simplicity. The lagrangian of the theory can be written as the sum of three different terms $\mathcal{L} = \mathcal{L}_{\text{elementary}} + \mathcal{L}_{\text{composite}} + \mathcal{L}_{\text{mixing}}$. Omitting the terms involving the EW interaction, since they are not relevant in the phenomenology of the model, the different terms of the lagrangian are:

$$\begin{aligned} \mathcal{L}_{\text{elementary}} &= -\frac{1}{4g_{el3}^2} G_{\mu\nu} G^{\mu\nu} + \bar{q}_L^i i \not{D} q_L^i + \bar{u}_R^i i \not{D} u_R^i + \bar{d}_R^i i \not{D} d_R^i \\ \mathcal{L}_{\text{composite}} &= -\frac{1}{4g_{*3}^2} G_{\mu\nu}^* G^{*\mu\nu} + \frac{1}{2} \text{Tr} \left\{ \partial_\mu \mathcal{H}^\dagger \partial^\mu \mathcal{H} \right\} - V(\mathcal{H}^\dagger \mathcal{H}) \\ &\quad + \text{Tr} \left\{ \bar{\mathcal{Q}} (i \not{D} - \mathcal{G}^* - \bar{m}_Q) \mathcal{Q} \right\} + \bar{\tilde{T}} (i \not{D} - \mathcal{G}^* - \bar{m}_{\tilde{T}}) \tilde{T} \\ &\quad + \text{similar terms for down-type multiplets} \\ &\quad - Y_{*U} \text{Tr} \left\{ \bar{\mathcal{Q}} \mathcal{H} \right\} \tilde{T} - Y_{*D} \text{Tr} \left\{ \bar{\mathcal{Q}}' \mathcal{H} \right\} \tilde{B} + h.c. , \\ \mathcal{L}_{\text{mixing}} &= \frac{1}{2} \frac{\bar{M}_{G_*}^2}{g_{*3}^2} \left(G_\mu - G_\mu^* \right)^2 - \Delta_{L1} \bar{q}_L^3 (T, B) - \Delta_{L2} \bar{q}_L^3 (T', B') - \Delta_{tR} \bar{t}_R \tilde{T} - \Delta_{bR} \bar{b}_R \tilde{B} + h.c. , \end{aligned}$$

where V is the Higgs potential and D is the $SU(3)$ covariant derivative. Other than the usual degrees of freedom of the SM, the model introduces four composite VLQ masses, and the G^* mass. Additionally, two Yukawa couplings $Y_{*U,D}$ and two new gauge couplings $g_{el3,*3}$ are added, one for the elementary and one for the composite particles. Finally, there are four elementary-composite mixing terms $\Delta_{L1,L2,tR,bR}$. $\mathcal{L}_{\text{mixing}}$ describes the interaction between the composite sector and the elementary sector; the mixing with the first and second generation of SM quarks is assumed to be small and is not considered in the model. A change of base from the elementary-composite basis to the mass eigenstate basis is needed to diagonalize the lagrangian, this naturally introduces mixing angles that are a function of the free parameters of the non-diagonal lagrangian:

$$\tan \phi_{tR,bR} = \frac{\Delta_{tR,bR}}{\bar{m}_{\tilde{T},\tilde{B}}}, \quad \tan \phi_L = \frac{\Delta_{L1}}{\bar{m}_Q}.$$

They can be interpreted as the degree of compositeness of the mass eigenstate quarks. In the gauge sector the two couplings become the rotation angle:

$$\tan \theta_3 = \frac{g_{el3}}{g_{*3}}.$$

Given these definitions, the physical masses, that is the mass eigenstates, are given by:

$$m_{\tilde{T},\tilde{B}} = \frac{\tilde{m}_{\tilde{T},\tilde{B}}}{\cos \phi_{tR,bR}}, \quad m_{T',B',B_{-4/3},B_{-1/3}} \simeq \tilde{m}_{Q'} ,$$

for the Q multiplet:

$$m_{T,B} = \frac{\tilde{m}_Q}{\cos \phi_L}, \quad m_{T_{5/3},T_{2/3}} = m_T \cos \phi_L ,$$

finally, for the G^* resonance:

$$m_{G^*} = \frac{\tilde{m}_{G^*}}{\cos \theta_3} .$$

The masses of the third generation SM quarks and the elementary gauge coupling are a function of the Yukawa couplings and the mixing angles. Therefore imposing the measured SM values, three equations can be obtained constraining the parameters. Additionally, the following simplifications are considered to reduce the amount of free parameters: the Yukawa coupling is the same for all the composites $Y_{*U} = Y_{*D}$, $\sin \phi_{bR} = \sin \phi_L$, and the mass of top-like and b-like VLQs is assumed to be the same $m_{T'} = m_T = m_{\tilde{B}} = m_{\tilde{T}}$. This leaves the model with only five free parameters: the mass of the T vector-like quark m_T and of the heavy resonance m_{G^*} , the Yukawa coupling for the composites Y_* , the mixing angle in the gauge sector $\tan \theta_3$, and the mixing angle between the top and the T composite VLQ $\sin \phi_{tR}$. Given the relationships between the different mixing angles and coupling parameters this is not the only choice of independent parameters.

For the interpretation of the results of the search in Chapter 8, the specific case is considered where the G^* gluon decays to a top quark and the heavy fermion T. The model assumes branching fractions to be 50/25/25% for the T quark decay to the bW/tH/tZ channels, as commonly considered in the literature [20]. Benchmark values of $\tan \theta_3 = 0.44$, $\sin \phi_{tR} = 0.6$, and $Y_* = 3$ are used for the model parameters. This choice of values enhance the decays of the KK gluon to a SM quark and a heavy quark. As shown above, this model predicts the existence of other vector-like quarks such as the $T_{5/3}$ quark, with a charge of five-thirds and a mass lower than the mass of the T quark: $T_{5/3} = m_T \cos \phi_L$, with $\cos \phi_L = 0.83$, given the benchmark parameter chosen. These other heavy quarks have a moderate impact on the branching fraction of the G^* resonance to Tt and the total decay width of the resonance. However, the contribution of the other heavy quarks is properly taken into account when computing the theoretical prediction. In particular, the $T_{5/3}$ quark becomes relevant when $m_{T_{5/3}} \gtrsim 2m_{G^*}$. Generally speaking, top quark flavour-changing neutral currents are generated in the interaction between the top quark, the G^* gluon, and T quark. However this effect is estimated to be below the reach of current measurements [45] because of the large suppression factor due to off-shell effects of the heavy resonance and the T quark.

Figure 1.9 shows the production cross section of the heavy resonance in function of its mass, Fig. 1.10(left) and 1.11 show the branching fractions of the G^* boson decay modes in function of the resonance mass, while Fig. 1.10(right) shows the width of the heavy resonance in function of its mass.

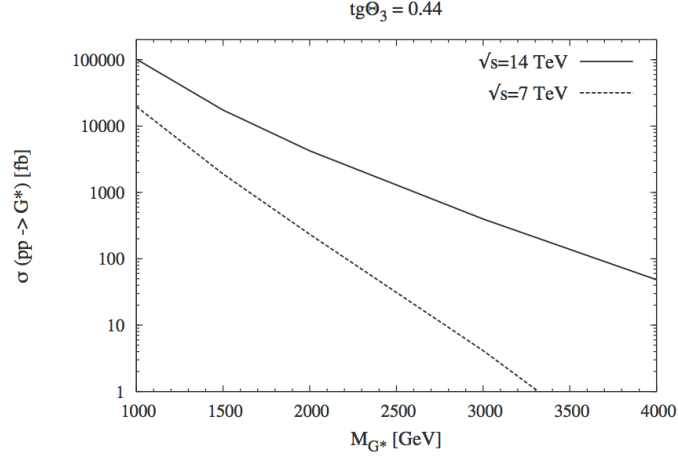


Figure 1.9: Production cross section of the G^* resonance at the LHC (Drell Yan production mode), for $\tan \theta_3 = 0.44$. The cross section scales quadratically in function of the $\tan \theta_3$ parameter [44].

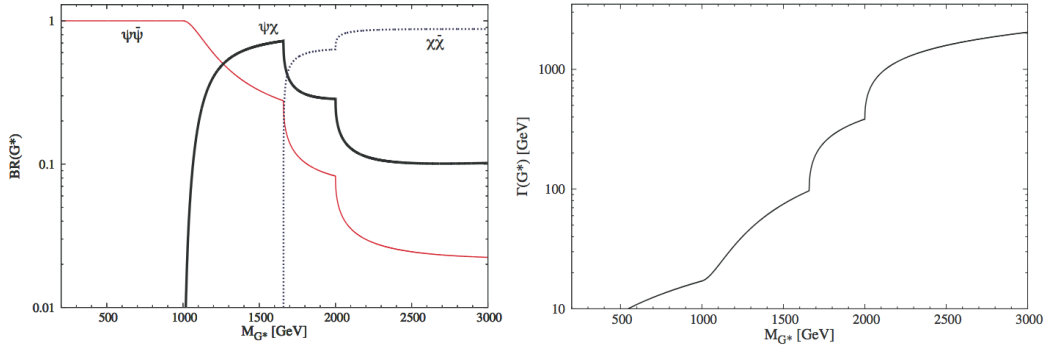


Figure 1.10: Left panel: decay branching fractions of the G^* to two SM quarks $\psi\bar{\psi}$, two vector-like quarks $\chi\bar{\chi}$, and one SM quark plus one heavy quark, $\chi\bar{\psi} + \psi\bar{\chi}$, as a function of the mass of the heavy resonance. Right panel: total decay width of the heavy resonance as a function of its mass. The plots use the benchmark model parameters: $m_T = 1$ TeV, $\tan \theta_3 = 0.44$, $\sin \phi_{tR} = 0.6$, and $Y_* = 3$ [44].

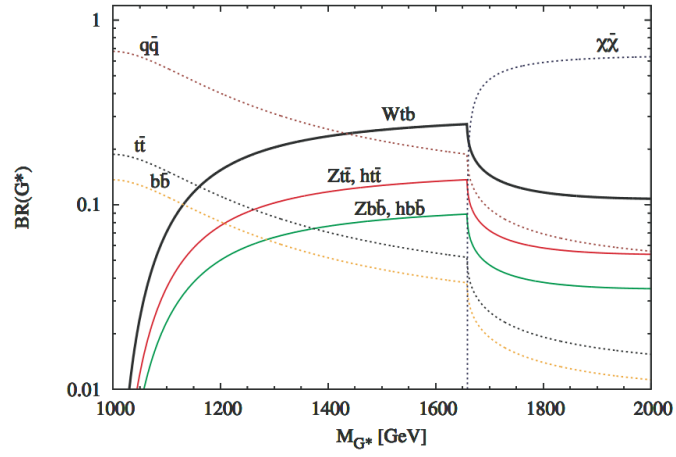


Figure 1.11: Decay branching fractions of the G^* resonance as functions of its mass for benchmark values of the model parameters: $m_T = 1$ TeV, $\tan \theta_3 = 0.44$, $\sin \phi_{tR} = 0.6$, and $Y_* = 3$. The decay channels $Ht\bar{t}$ and $Hb\bar{b}$ have the same branching ratio of $Zt\bar{t}$ and $Zb\bar{b}$, respectively [44].

The Large Hadron Collider and the CMS Experiment

2.1 The Large Hadron Collider

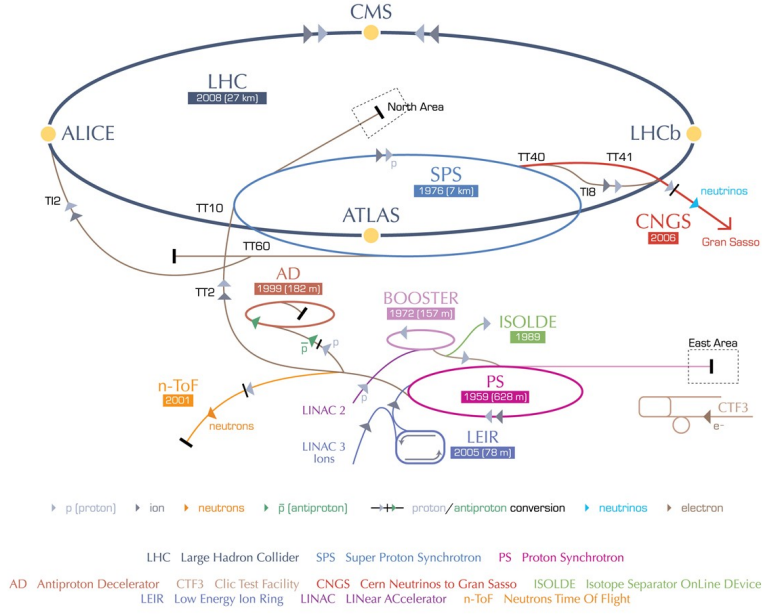
The Large Hadron Collider (LHC) is a proton-proton collider built at CERN on the French-Swiss border near the city of Geneva. As of 2017 the Large Hadron Collider is the highest-energy hadron collider ever built. The LHC is a discovery machine: it was built to explore the mechanism for the origin of mass in the SM, and possible extensions of the SM, including supersymmetry, dark matter, and extra dimensions. The LHC is also able to collide lead ions with other lead ions or lead ions with protons, allowing to study the behavior of matter at high density and high temperature, such as the quark gluon plasma. Generally speaking the goal is to probe particle physics at the TeV energy scale, continuing the efforts started at the Tevatron collider at the Fermi National Accelerator Laboratories, the first collider to reach a center-of-mass energy at the TeV scale.

Figure 2.1 shows a schematic representation of the accelerator complex at CERN. Several of the accelerators in the complex are necessary to the functionality of the LHC and serve to prepare and pre-accelerate the beams of protons or lead ions before they are injected in the LHC.

The study of physics processes at the TeV scale is a difficult technological challenge. The machine must be able to accelerate particles at high energy and, at the same time, provide a very high rate of interaction (luminosity). These requirements drive the main features of the accelerator. In a hadron collider, the fundamental constituents participating in the hard scattering are the partons, carrying a fraction x of the four-momentum of the particles in the beam. Therefore, the energy of the hard scattering process can span several orders of magnitude. For example, assuming a center-of-mass energy of $\sqrt{s} = 13$ TeV, and typical parton momentum fractions $x_1, x_2 \simeq 0.15 - 0.20$ of the incoming protons momenta, the center-of-mass energy of the hard-scatter process is $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s} \simeq 1 - 2$ TeV. Despite the loss in accumulated energy, protons are easier to accelerate with respect to electrons, since the energy lost for synchrotron radiation, proportional to γ^4 (where $\gamma = E/m$), is much higher for electrons. Proton beams are easier to produce and squeeze, in order to obtain very high intensities, hence a pp collider has been preferred over a proton-antiproton collider, such as Tevatron.

The LHC is a high luminosity collider. The cross section σ of a physical process determines the event rate R according to the formula $R = \mathcal{L}\sigma$, where $\mathcal{L}$ is the instantaneous luminosity,

CERN's accelerator complex



European Organization for Nuclear Research | Organisation européenne pour la recherche nucléaire

© CERN 2008

Figure 2.1: A schematic representation of the CERN accelerator complex and the LHC [46].

representing the number of collisions per unit of time and cross section. The instantaneous luminosity is a parameter of the collider and can be expressed as: $\mathcal{L} = f \frac{n_1 n_2}{A}$, where f is the collision frequency of bunches composed of n_1 and n_2 particles and A is the overlapping cross-section area of the beams. To compensate for the low cross section of the interesting processes (new physics, Higgs boson production, etc.), the LHC must have a very high luminosity: the very short bunch crossing interval (25 ns or 40 MHz) and the high number of proton bunches accelerated by the machine (2808 per beam) allows to reach a design peak luminosity of $10^{34} \text{cm}^{-2} \text{s}^{-1} = 1 \text{nb}^{-1} \text{s}^{-1}$.

The design luminosity was actually surpassed in 2016, marking a success in the operational efficiency of the accelerator. On the downside, this very high luminosity causes a faster radiation contamination of the detector which negatively affects their efficiency and operation. The electronics and sensitive components of the detectors are designed to be as resistant as possible to radiation, however a degradation in the performance is inevitable, especially for the subsystems closer to the interaction point. The detector components most affected are being replaced during the long technical shutdowns happening every 2-3 years. Additionally, the large number of collisions happening at every bunch crossing (pileup) makes the reconstruction of the hard-scatter collision more challenging.

The LHC reuses the existing 27 km long tunnel that hosted the Large Electron-Positron Collider (LEP) to install the new accelerator. Here are reported the design parameters of the LHC:

- circumference: 26.659 Km,
- design centre of mass energy: 14 TeV per proton in pp collisions (as of 2017 the maximum reached is 13 TeV),
- dipole magnetic field: 8.3 T,

- number of particles per bunch: 2808 in pp collisions,
- bunch length: 53 mm in pp collisions,
- bunch crossing rate: 40.08 MHz,
- design Luminosity: $10^{34} \text{cm}^2 \text{s}^{-1}$ in pp collisions,
- beam radius at interaction point: $15 \mu\text{m}$.

The Large Hadron Collider posed new technological and engineering challenges to the scientific communities. In total, over 1,600 superconducting magnets are installed, weighing over 27 tonnes. Approximately 96 tonnes of liquid helium is needed to keep the magnets, made of copper-coated niobium-titanium, at their superconducting operating temperature of 1.9 K (271.3 C), making the LHC the largest cryogenic facility in the world at liquid helium temperature. Figure 2.2 shows the cross section of a superconducting dipole magnet.

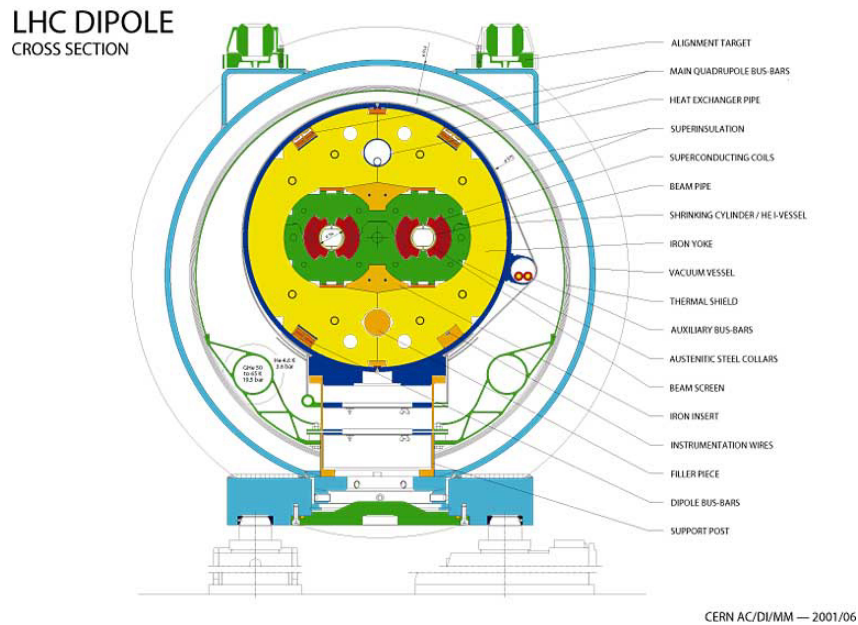


Figure 2.2: Cross section of a LHC dipole [47].

One aspect of LHC physics is the overwhelming rate of background processes compared to interesting physics processes: the Higgs production, for instance, has a cross section roughly ten orders of magnitude smaller than that of the total inelastic pp cross section, as shown in Fig. 2.3. Most of the events produced in pp collisions are composed by diffractive, low p_T scattering, where the proton collision is not central. These events are commonly referred to as minimum bias and they are generally considered background events for the high- p_T physics that is studied in this thesis.

The fact that the two partons, being part of the proton, interact with unknown energies implies that the total energy of an event is unknown. Most of the proton remnants, carrying a good fraction of the total proton energy, are scattered at very small angles and are mostly lost in the beam pipe, therefore almost impossible to detect. When reconstructing a collision event, it is thus not possible to define the total energy of the event, but only total and missing transverse energies. All the physics observables used for data analysis are therefore measured in the plane transverse to the beamline.

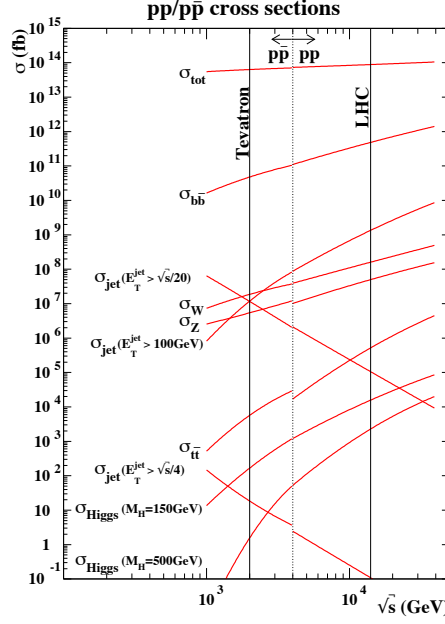


Figure 2.3: Cross section as a function of the center-of-mass energy at the LHC for some interesting physics processes [48].

The first beam was circulated through the LHC on 10 September 2008. The proton beams successfully circulated around the tunnel, however shortly after, on 19 September 2008, a sudden raise of temperature (“quench”) occurred in about 100 bending magnets causing a loss of approximately six tonnes of liquid helium, which was vented into the tunnel. The cause of the problem were faulty electrical connections between the magnets. A total of 53 magnets were damaged in the incident and were subsequently repaired or replaced, however more than a year passed before the LHC was put in function again. Due to this accident the LHC ran at a reduced energy of collisions $\sqrt{s} = 7$ TeV during 2011. In 2012 the center-of-mass energy was raised to 8 TeV, CMS collected almost 20 fb^{-1} of data. Between 2013 and 2014 lengthy additional repairs were performed on the magnets and their junctions to be able to run at a higher energy. During the year 2015 and 2016 the LHC successfully ran at an energy of 13 TeV, and CMS collected 2.7 and 36 fb^{-1} of data, respectively. It is planned to raise the energy to 14 TeV and, after an upgrade of the accelerator components, to significantly raise the instantaneous luminosity. This upgrade is called the high-luminosity LHC (HL-LHC). The various detectors installed at LHC will upgrade their subsystems to improve their performance in view of the higher luminosity conditions.

The Large Hadron Collider is the host of six different detector experiments, each experiment has a different composition and geometry of the subdetectors so that it is more specialized in a particular area of the research in particle physics. ATLAS (A Toroidal Lhc ApparatuS) and CMS (Compact Muon Solenoid) are two general-purpose, high-luminosity detectors, they are built with a cylindrical geometry around the beamline. Their subdetectors are optimized for the reconstruction of high energy objects with great efficiency and accuracy. The design of the two detectors presents significant differences, since different solutions were chosen for the configuration of the magnetic field and the technology used for the subdetectors. ATLAS uses a toroidal field produced by three sets of air-core toroids complemented by a small solenoid in the inner tracking region, while CMS uses a solenoidal field.

ALICE (A Large Ion Collider Experiment) is a detector optimized for ion collisions, in

particular for the study of the properties of matter at high temperature and high energy density generated by such collisions. LHCb is specialized in studies regarding the physics of heavy quarks and heavy mesons with a particular attention to the b quark and its mesons.

TOTEM and LHCf are smaller very-forward (small angle with respect to the beamline) detectors linked to CMS and ATLAS, respectively. They are placed ~ 100 m from the interaction points of the main experiments to study the diffractive physics happening in the very forward region of the collision. These detectors are placed far from the interaction point so that the products of very-forward inelastic or elastic collisions may exit the beampipe.

Figure 2.4 shows the position of the four major detectors installed at the LHC. Further details about LHC can be found in Ref. [47].

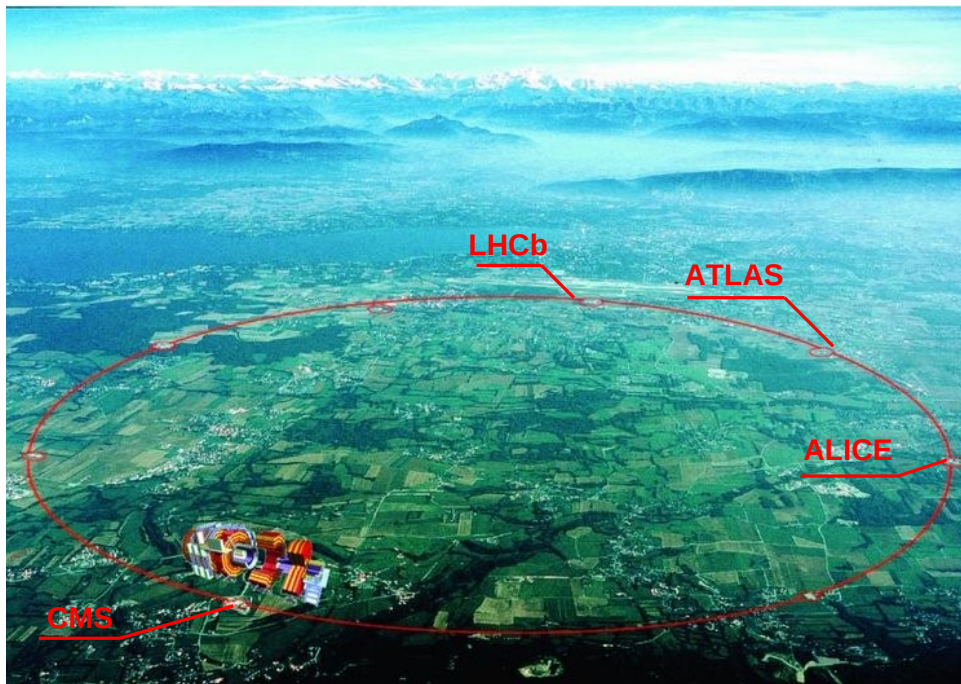


Figure 2.4: Aerial view of the LHC area and the four interaction points.

2.2 The Compact Muon Solenoid Experiment

This chapter describes the hardware details of the CMS detector. The various subdetectors are described and an overview is given of the data acquisition system and the computing infrastructure. To be able to reconstruct collision data in the high-energy and high-luminosity conditions of the LHC, and to fulfill the goals of the physics program of the LHC, a detector should have the following characteristics:

- good muon identification and momentum resolution over a wide range of momenta in the central region. The charge of muons should be determined without ambiguity for momenta up to at least 1 TeV; in particular the dimuon mass resolution should be excellent in the mass range of the Z boson.
- The tracking system, besides a good charged particle momentum resolution and reconstruction efficiency, should have very precise vertexing capabilities in order to be able to distinguish the hard scatter from the pileup interactions and to perform b jet identification.

- The electromagnetic calorimeter should allow a precise localization of photons and electrons and a measurement of their energy. π_0 rejection requires a good granularity of the calorimeter.
- In terms of hadron calorimetry, a good missing energy and dijet mass resolution is essential, requiring a wide geometric coverage and a fine segmentation.

Figure 2.5 shows a panoramic view of the open CMS detector during the first long-shutdown. On the left side is the barrel part of the detector with the muon chambers and the magnet clearly visible. On the right side is the endcap part of the detector. The main layout of the CMS detector is shown in figure 2.6.

The detector structure consists of a cylindrical barrel closed by two endcap disks. The overall length is 21.6 m, the diameter 14.6 m and the total weight about 12.5 thousand tonnes. The thickness of the detector is greater than 25 radiation lengths (X_0) for the electromagnetic calorimeter, and the thickness in interaction lengths varies from 7 to 11 λ_I for the hadronic calorimeter, depending on the detector region.

The core of the detector is the magnet, a superconductive solenoid, 13 m long, cooled with liquid helium, and providing a 3.8 T magnetic field. The magnet coil has a diameter of 5.9 m, it contains the tracker, and the electromagnetic and hadronic calorimeters. The iron return yoke of the magnet hosts the muon chambers. In the endcaps a lead/silicon pre-shower detector is installed to improve pion rejection.

The choice of a solenoidal topology for the magnetic field leads to a compact design for the detector with respect to, for example, the ATLAS detector design. This allows the calorimeters to be installed inside the magnet, with a considerable improvement in the detection of charged particles. In particular, the almost constant magnetic field inside the solenoid makes the momentum measurement easier and therefore more precise. The downside of this design is that muons passing through the return yoke of the magnet experience multiple scattering effects, affecting negatively muon reconstruction. The longitudinal view of one quarter of CMS and the transverse view of the barrel region are shown in Figure 2.7.

For a particle carrying 4-momentum (E, p_x, p_y, p_z) , the momentum vector p can be divided in two components: the longitudinal momentum p_z along the beamline and the transverse momentum defined as $p_T = \sqrt{p_x^2 + p_y^2}$. The rapidity is defined as:

$$y = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) . \quad (2.1)$$

Being invariant under boost of the centre-of-mass along the z direction, the rapidity is used for describing angular distribution of the momentum of particles. For an ultra-relativistic particle ($p \gg m$) the rapidity can be approximated by the pseudorapidity:

$$\eta = -\log \left(\tan \frac{\theta}{2} \right) , \quad (2.2)$$

where θ is the angle between the particle momentum p and the z axis. The other relevant variable in the detector is the azimuthal angle ϕ . Generally speaking, the main kinematic information of a particle in the detector are the momentum direction (η, ϕ) and the transverse momentum. More detailed technical information on the CMS detector and its subdetectors can be found in Ref. [50–53].

2.2.1 The solenoidal magnet

The magnet used by CMS is a 13 m long superconducting solenoid, the largest ever built. It is able to generate a uniform magnetic field of 3.8 T in the inner region, storing about 2.5 GJ of

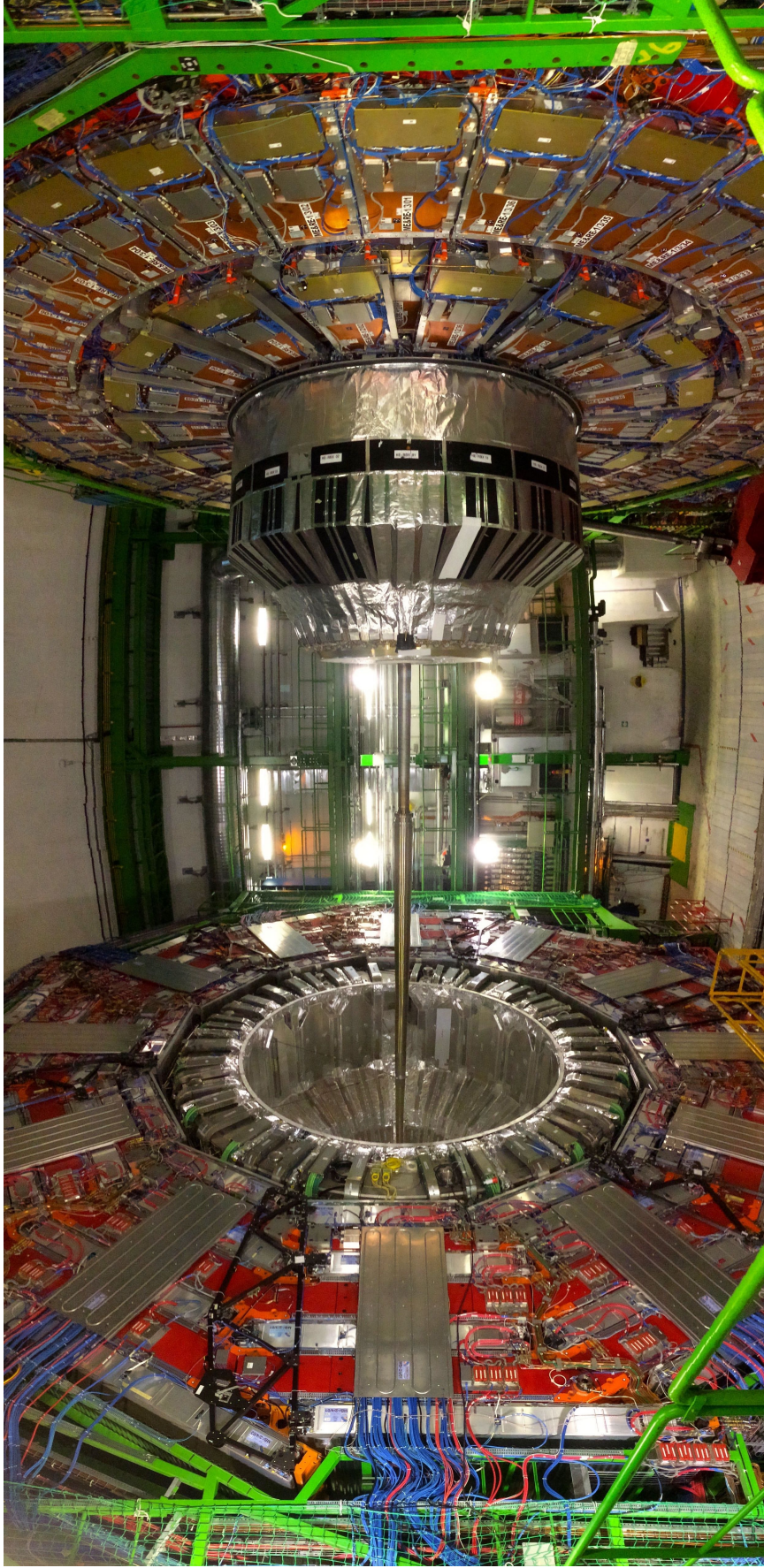


Figure 2.5: A view of the open CMS detector during the first long-shutdown.

CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel ($100 \times 150 \mu\text{m}$) $\sim 16\text{m}^2 \sim 66\text{M}$ channels
Microstrips ($80 \times 180 \mu\text{m}$) $\sim 200\text{m}^2 \sim 9.6\text{M}$ channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying $\sim 18,000\text{A}$

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips $\sim 16\text{m}^2 \sim 137,000$ channels

FORWARD CALORIMETER
Steel + Quartz fibres $\sim 2,000$ Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
 $\sim 76,000$ scintillating PbWO_4 crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator $\sim 7,000$ channels

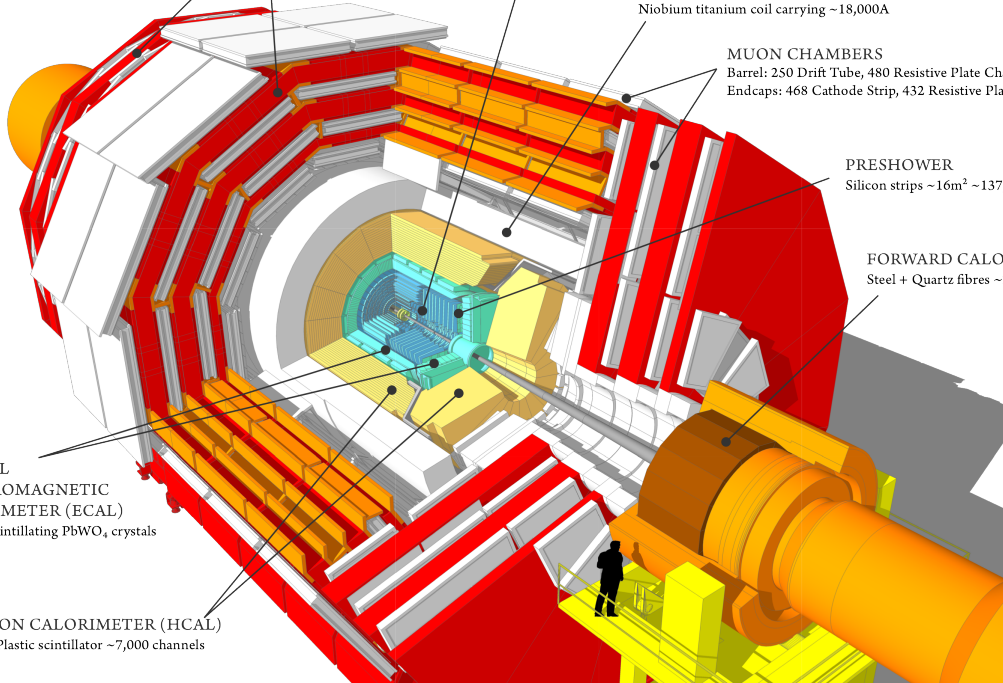


Figure 2.6: CMS detector structure [49].

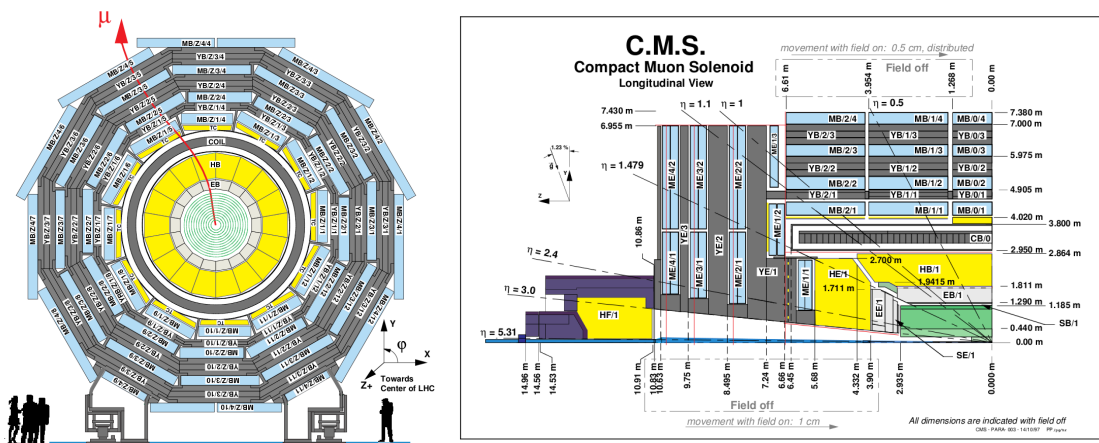


Figure 2.7: Section and side view of the CMS detector [50].

energy. It operates at a temperature of 4 K, ensured by a sophisticated helium cooling system. At the operating temperature, the cable made of NiTb alloy becomes superconducting, thus allowing a 20 kA current to flow with small losses.

The magnet is contained in a vacuum cylinder, isolated from the rest of the detector. A magnetic field is generated outside the magnet too: to avoid interferences and losses the return magnetic field must be guided back by an iron return yoke. The iron structure is composed by five barrel layers and three disks for each side. The magnet bending power allows a precise measurement of the transverse momentum of charged particles inside the solenoid, operated by the inner tracking system. For muons, a further and independent p_T measurement outside the solenoid is possible thanks to the iron yoke, which surrounds the muon chambers. Figure 2.8 shows a map of the magnetic field generated by the solenoidal magnet.

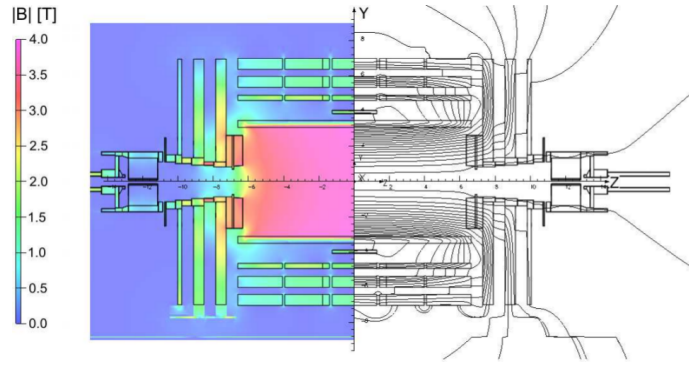


Figure 2.8: The CMS magnet and the generated magnetic field [50].

2.2.2 The tracking system

The tracking system is the core of the CMS detector and is essential for the reconstruction of charged particles produced by collisions. The subdetector is built using a silicon detector technology. The tracker covers the pseudorapidity region $|\eta| < 2.5$ with a radius ranging between 4.3 cm and 120 cm, and $-270 < z < 270$ cm, using the CMS coordinate system centered at the interaction point where z is the coordinate along the beamline. The innermost region is made of pixel detectors, while the outermost region uses strip detectors.

The pixel tracker consists of three 53.3 cm long barrel layers and two endcap disks on each side of the barrel section, as shown in Fig. 2.9. The innermost barrel layer has a radius of 4.4 cm, while for the second and third layer the radii are 7.3 cm and 10.2 cm, respectively. The layers are composed of modular detector units (called modules) placed on carbon fiber supports (called ladders). Each ladder includes eight modules, consisting of thin ($285\mu\text{m}$), segmented silicon sensors with integrated readout chips (ROC) connected by indium bump-bonds. The Barrel Pixel (BPIX) region is composed of 672 full modules and 96 half modules, each including 16 and 8 ROCs, respectively. The number of pixels per module is 66560 (full modules) or 33280 (half modules). The total number of pixels in the barrel section is almost 50 millions.

The Forward Pixel (FPix) endcap disks, extending from 6 cm to 15 cm in radius, are placed at $z = \pm 35.5$ cm and $z = \pm 48.5$ cm. Disks are split into half-disks, each including 12 trapezoidal blades arranged in a partially-overlapped geometry. Each blade is composed of two back-to-back panels. Rectangular sensors of five sizes are bump-bonded to arrays of

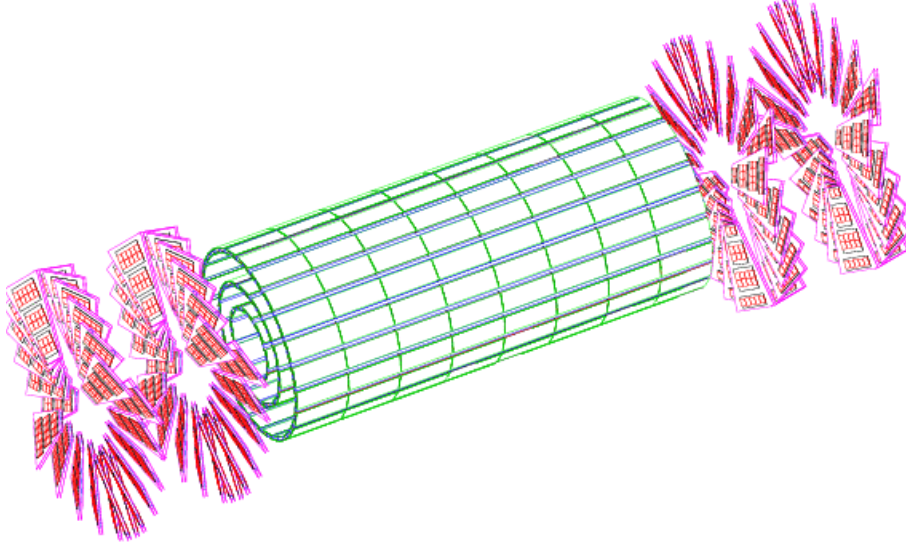


Figure 2.9: Schematic illustration of the pixel tracker [50].

ROCs. The endcap disks include a total of almost 18 millions pixels. The minimal pixel cell area is dictated by the readout circuit surface required for each pixel.

Outside the pixel detector, the outer tracker (see Figure 2.10) is composed of 10 layers of silicon microstrip detectors. The barrel region ($|\eta| < 1.6$) is divided into two parts: the Tracker Inner Barrel (TIB), covering $20 < r < 60$ cm and the Tracker Outer Barrel (TOB), covering $60 < r < 120$ cm. The TIB is composed by four layers of p-on-n type silicon sensors with a thickness of $320 \mu\text{m}$ and strips varying from $80 \mu\text{m}$ to $120 \mu\text{m}$. The first two layers are made with double sided modules, composed by two detectors mounted back to back with the strips tilted by 100 mrad . This kind of sensors provides a measurement in both $r - \phi$ and $r - z$ coordinates with a single point resolution between $23 - 34 \mu\text{m}$ and $230 \mu\text{m}$ respectively. The TOB is made of six layers. In this region the radiation levels are smaller and thicker silicon sensors can be used to maintain a good signal-to-noise ratio for longer strip lengths. The dimensions of the strips vary from $120 \mu\text{m}$ to $180 \mu\text{m}$. The TOB provides a single point resolution which varies from $35 \mu\text{m}$ to $52 \mu\text{m}$ in the r direction and $530 \mu\text{m}$ in z . The endcap region ($|\eta| > 1.6$) is covered by the Tracker Inner Disks (TID) and Tracker End Cap (TEC). The three disks of the TID fill the gap between the TIB and the TEC while the TEC comprises nine disks that extend into the region $120 < |z| < 280$ cm. Both subdetectors are composed of wedge shaped modules arranged in rings, centred on the beam line, and have strips that point towards the beam line (radial topology).

In general, the particle transverse momentum resolution can be written as:

$$\frac{\delta p_T}{p_T} = C_1 p_T \oplus C_2 , \quad (2.3)$$

where the term C_2 contains the effects of Coulomb scattering, while the coefficient C_1 depends on the detector geometry, in particular from the number of points used for the track reconstruction (n), its length (L), and the resolution on the single point measurement (σ_x):

$$C_1 \propto \frac{\sigma_x}{\sqrt{n} \cdot B \cdot L^2} . \quad (2.4)$$

For low energetic particles C_2 dominates. C_1 is minimized having a tracker composed by many layers, and a consistent number of points n in the track fit.

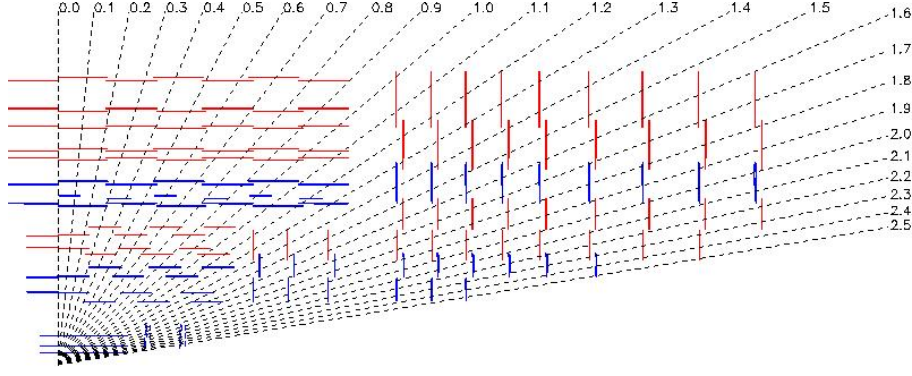


Figure 2.10: Schematic view of the strip tracker [50].

The CMS tracker detector has the following properties:

- reconstruction capability in the region $|\eta| < 2.5$ with an efficiency of at least 95% for charged tracks with $p_T > 10 \text{ GeV}$;
- high momentum resolution for isolated tracks:

$$\frac{\delta p_T}{p_T} = (1.5 p_T \oplus 0.5)\% \quad (2.5)$$

for $|\eta| < 1.6$, and

$$\frac{\delta p_T}{p_T} = (6.0 p_T \oplus 0.5)\% \quad (2.6)$$

for $|\eta| < 2.5$.

where the p_T is expressed in TeV;

- high resolution for transverse impact parameter, $\sigma(d_{xy}) = 35 \mu\text{m}$ and longitudinal impact parameter $\sigma(d_z) = 75 \mu\text{m}$.

The tracker has achieved an excellent momentum resolution: Figure 2.11 shows the χ_{c1} and χ_{c2} resonances, reconstructed through $\chi_c \rightarrow J/\psi \gamma$ with $J/\psi \rightarrow \mu^+ \mu^-$ and $\gamma \rightarrow e^+ e^-$. A mass resolution of less than 50 MeV is achieved, allowing the two resonances to be resolved clearly.

2.2.3 Calorimetry

The goal of the electromagnetic calorimeter is the accurate measurement of the energy and position of electrons and photons. The ECAL is composed of 75,848 finely segmented lead tungstate (PbWO_4) crystals, chosen because of their high density and excellent energy resolution. Lead tungstate is a fast, radiation-hard scintillator characterized by a small Moliere radius ($R_M = 19 \text{ mm}$) and a short radiation length ($X_0 = 8.9 \text{ mm}$), that allows a good shower containment in the limited space available for the detector. Moreover, the crystals are characterized by a very short scintillation decay time allowing the electronics to collect about 80% of the light within 25 ns. A pre-shower detector is installed in front of the endcaps, consisting of two lead radiators and two planes of silicon strip detectors, with a total radiation length of $3X_0$. It allows the rejection of photon pairs from π_0 decays and improve the estimation of the direction of photons. The coverage of the calorimeter extends up to $|\eta| = 3.0$, as shown

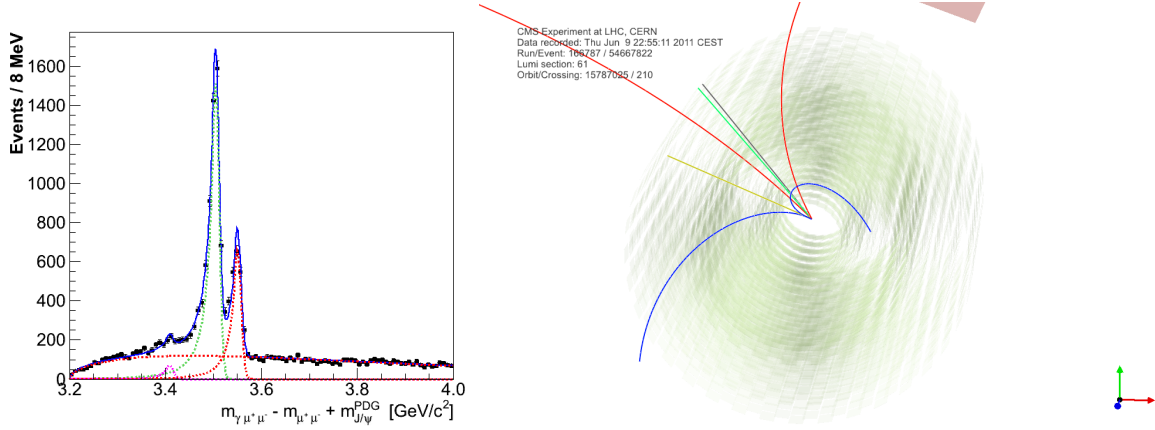


Figure 2.11: Left panel: mass distribution of χ_{c1} and χ_{c2} resonances, reconstructed through $\chi_c \rightarrow J/\psi\gamma$ with $J/\psi \rightarrow \mu^+\mu^-$ and $\gamma \rightarrow e^+e^-$. Right panel: event display with a candidate χ_c . Red tracks are muons, blue tracks are electrons from the photon conversion [54].

in Fig. 2.12. The crystals are arranged in a $\eta - \phi$ grid in the barrel and a $x - y$ grid in the endcaps and they are quasi-pointing to the interaction point: the axes are tilted a 3° in the barrel and at $2^\circ - 5^\circ$ in the endcaps with respect to the line from the nominal vertex position in order to avoid for particles to get lost in the space between the crystals.

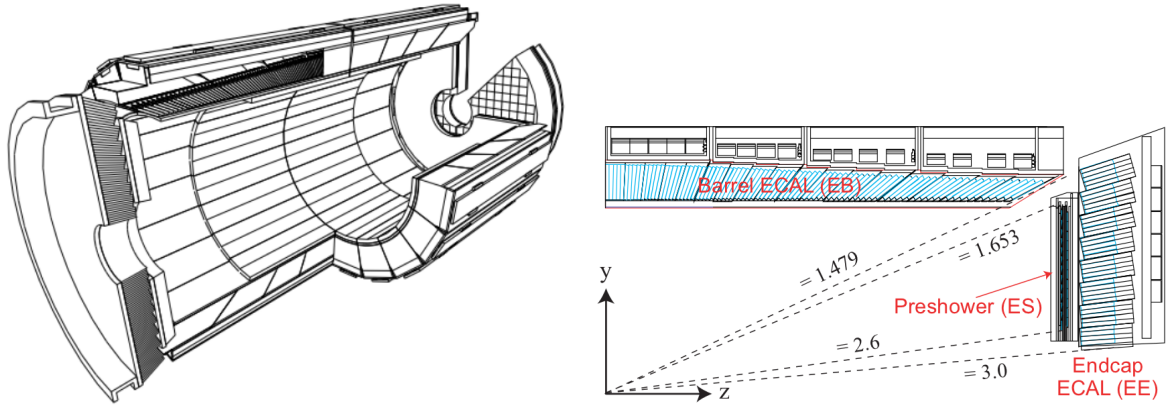


Figure 2.12: Schematic view of the electromagnetic calorimeter [50].

The energy resolution of a calorimeter can be parameterized as:

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

where a is the stochastic term and it includes the effects of fluctuations in the number of photo-electrons and in the shower containment, b is the noise from the electronics and pile-up, and c is a constant term related to the calibration of the calorimeter. The values of the three constants measured on test beams are reported in Table 2.1. The different contributions as a function of the energy are shown in Figure 2.13. Figure 2.14 shows the temperature stability of the ECAL barrel detectors during the early 2012 data-taking period. The temperature of the detector affects its performance, so it must be kept as uniform as possible.

The goal of the hadron calorimeter is to measure the position and energy of hadronic jets, the total transverse energy and the missing transverse momentum (MET) of the event.

Contribution	Barrel($\eta = 0$)	Endcap($\eta = 2$)
Stochastic term a	2.7%	5.7%
Noise (low luminosity) b	0.155 GeV	0.205 GeV
Noise (high luminosity)	0.210 GeV	0.245 GeV
Constant term c	0.55%	0.55%

Table 2.1: Different contributions to the energy resolution of ECAL.

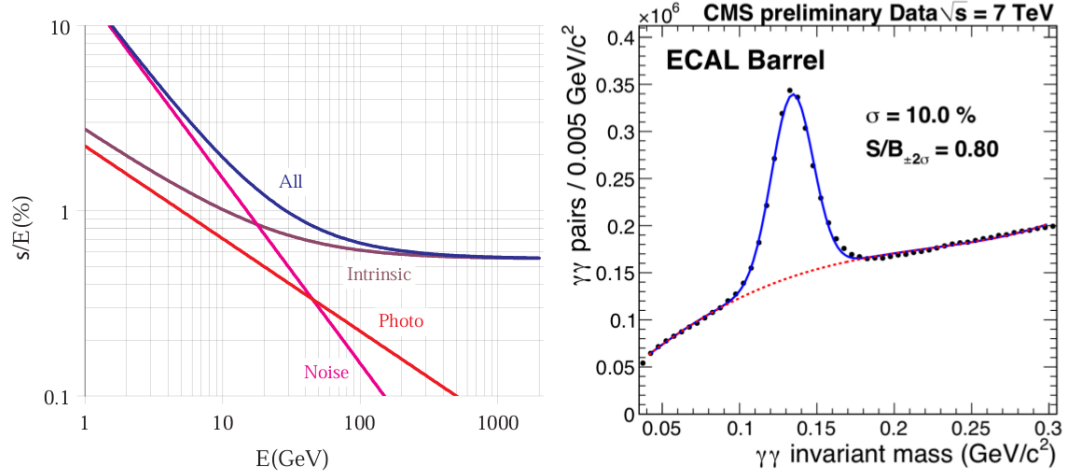


Figure 2.13: Left: Different contributions to the energy resolution of the ECAL. Right: diphoton invariant mass spectrum reconstructed by ECAL with about 250 nb^{-1} of data at 7 TeV. The π_0 peak is visible, the mass resolution is of the order of 10% [50].

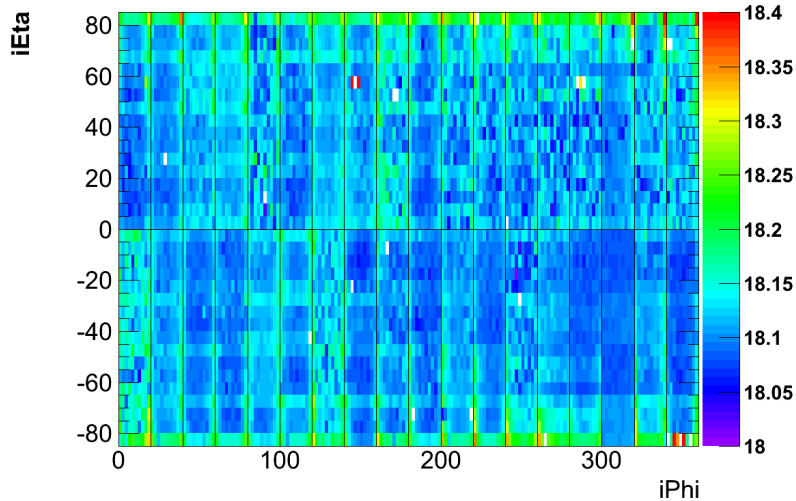


Figure 2.14: Temperature stability of the ECAL barrel detectors during the early 2012 data-taking period.

The detector should be as hermetic as possible for this purpose, which means the calorimeter must cover a portion of the solid angle as big as possible. For this reason, the barrel and endcap parts installed inside the magnet are complemented by a very forward calorimeter which is placed outside the magnet return yokes, with a total coverage of $|\eta| < 5.3$. The barrel and endcap HCAL cover the region $|\eta| < 3.0$. The detector is a sampling calorimeter, whose active elements are plastic scintillators interleaved with brass absorber plates and read out by wavelength-shifting fibers. The first layer is read out separately, while all others are read out together. The absorber material has been chosen for its short interaction length, and its non-magnetic property.

In the barrel, full shower containment is not possible within the magnet volume, therefore a tail catcher is placed outside the magnet, consisting of an additional layer of scintillators. The projective depth in terms of nuclear absorption length goes from $5.1\lambda_I$ at $\eta = 0$ to $9.1\lambda_I$ at $\eta = 1.3$ and $10.5\lambda_I$ in the endcap. The very forward calorimeter is placed outside the magnet yoke, 11 m from the interaction point. The active elements are quartz fibers parallel to the beam, inserted in steel absorber plates. The signal originated from quartz fibers is Cerenkov light. The expected energy resolution is $\sigma/E \simeq 65\%\sqrt{E} \oplus 5\%$ in the barrel, $\sigma/E \simeq 85\%\sqrt{E} \oplus 5\%$ in the endcaps and $\sigma/E \simeq 100\%\sqrt{E} \oplus 5\%$ in the very forward calorimeter.

2.2.4 The muon detection system

Muons are present in many physics processes of interest at the LHC (Z and Higgs bosons, top quarks, new heavy resonances). Therefore, the muon chambers must provide a reliable trigger system and an accurate measurement of the muon p_T and charge. The trigger function must also work without the contribution of the tracker. The muon system, shown in Fig. 2.15, is embedded in the iron return yoke of the magnet, this ensures that only muons, among charged particles, can reach this subdetectors (neutrinos will cross the detector undisturbed). The minimum value of the muon transverse momentum required to reach the system is $\simeq 5$ GeV. The muon spectrometer consists of three independent subsystems described in the following.

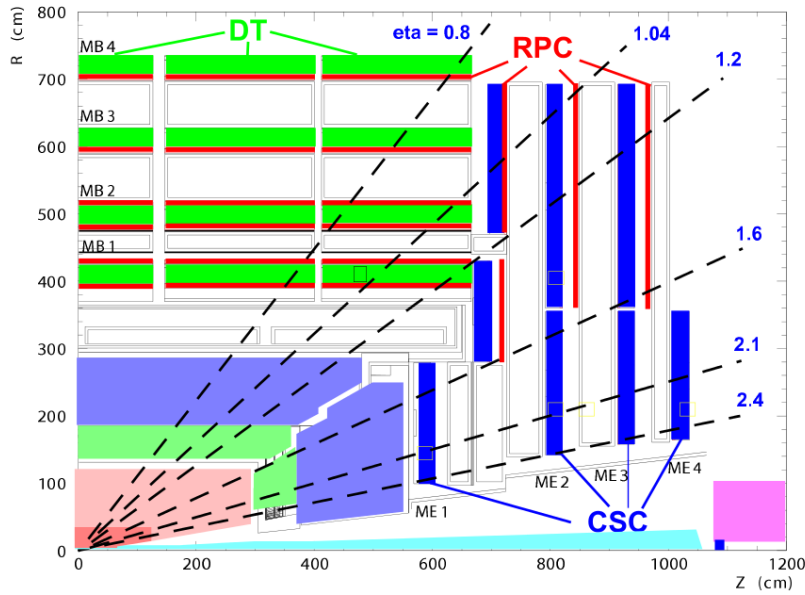


Figure 2.15: Schematic view of the muon system [50].

In the barrel ($|\eta| < 1.2$), where the track density and the residual magnetic field are low, four layers (stations) of drift tube chambers (DT) are installed. The chamber segmentation follows that of the iron yoke, consisting of five wheels along the z axis, each one divided into 12 azimuthal sectors. Each chamber has a resolution of about $100\mu\text{m}$ in $r - \phi$ and 1 mrad in ϕ .

In the endcaps ($0.8 < |\eta| < 2.4$), four disks (stations) of cathode strip chambers (CSC) are placed. This detector technology is more indicated for a region suffering high particle rates and large residual magnetic field between the plates of the yoke. The rings are formed by 18 or 36 trapezoidal chambers, which are stacked with a small overlap in ϕ . These chambers have a spatial resolution of about $200\mu\text{m}$ ($100\mu\text{m}$ for the chambers belonging to the first station) and 10 mrad in $r - \phi$. Redundancy is obtained with a system of resistive plate chambers (RPC), that are installed in both the barrel and the endcaps. RPCs have limited spatial resolution, but fast response and an excellent time resolution of a few nanoseconds, providing unambiguous bunch crossing identification. RPC detectors operate in avalanche mode, and are able to sustain higher rates. In the barrel, the RPC chambers follow the segmentation of DT chambers, and a total of six layers of RPCs are present. In the endcaps the chambers are trapezoidal and distributed on four disks. RPCs are also used to complement DTs and CSCs in the measurement of the transverse momentum. The RPC system covers the region $|\eta| < 2.1$.

In absence of collision data, cosmic muons were used to commission the detector: Figure 2.16 shows a cosmic muon crossing the detector and knocking off two electrons from the tracker material.

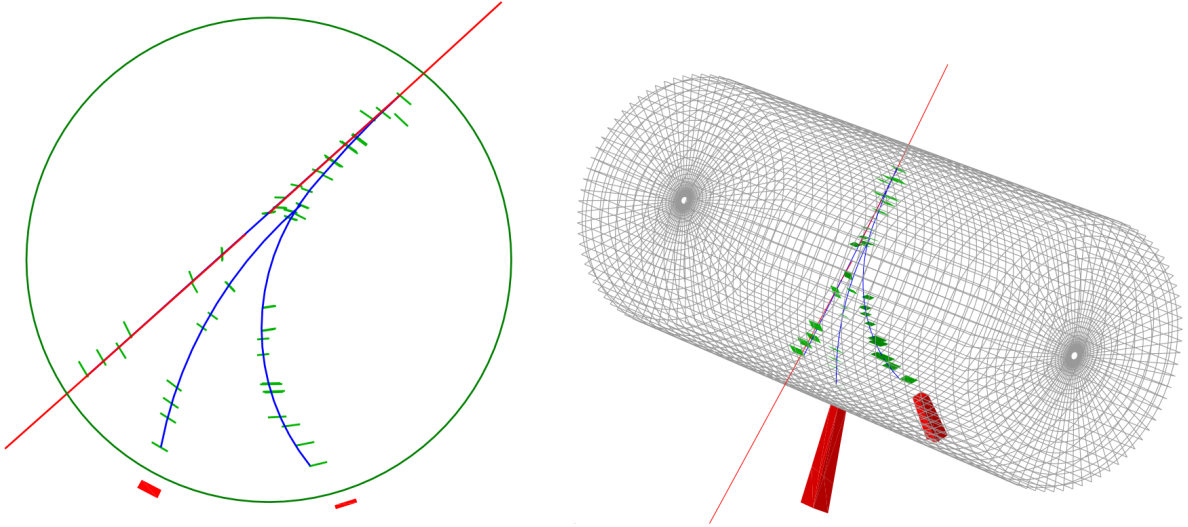


Figure 2.16: Event display of a cosmic muon crossing the detector and knocking off two electrons from the tracker material [55].

2.2.5 The data acquisition and trigger system

The huge amount of data produced due to the high interaction rate of the LHC cannot be sustained by any storage system presently available. Given the typical size of a raw event (1 MB), only a rate of ≈ 1 kHz can be stored for offline analysis, when the collisions rate is in fact 40 MHz. A reduction factor of several orders of magnitude is thus necessary: this is accomplished by the trigger and the data acquisition systems.

As previously mentioned, the event rate at the LHC is mainly composed of protons interactions with particles of low transverse momentum. A good triggering system should have a large rejection of the less interesting collisions and maintain at the same time a high efficiency on the potentially interesting events. This characteristic is achieved at CMS in two steps: a hardware-based Level 1 Trigger (L1) and a software-based High Level Trigger (HLT).

The Level 1 trigger consists of custom-designed, programmable electronics: it reduces the rate of selected events down to 100 kHz, assuming a standard, high-luminosity run. The full data are stored in cache pipelines, while waiting for the trigger decision. The maximum latency allowed is $3.2\mu\text{s}$: if the L1 trigger accepts the event, the data are moved to be processed by the HLT. The high interaction rate does not permit the full readout of the detector for the trigger decision, mainly because of the slowness of the tracking algorithms: only the calorimetric and muon chamber information are used, although efforts are underway to upgrade the trigger system and allow tracker information to be used. The calorimeter trigger identifies the best four candidates of each of the following classes: electrons, photons, central jets, and forward jets depending on the shape of the deposited energy. The information of these objects is passed to the global trigger, together with the measured missing p_T . The muon trigger considers separately each of the muon detectors. The information is then merged and the best muon candidates are transferred to the global trigger. The global trigger takes the decision to reject an event or to accept the event for further evaluation by the HLT. The decision is based on algorithm calculations and on the readiness of the sub-detectors and the data acquisition system (DAQ). The L1 trigger electronics is housed partly on the detectors, partly in the underground control room located at a distance of approximately 20 m from the CMS detector site.

The HLT is a software system implemented through a computing farm which reduces the output rate down to around 100 Hz. The idea of the HLT trigger software is the regional reconstruction on demand: only objects in the useful regions are reconstructed and uninteresting events are rejected as early as possible. The HLT has access to the high-resolution data in pipelined memories in the front-end electronics as well as the information from the silicon tracker: it can therefore perform complex calculations. The CMS Data Acquisition (DAQ) has the task to transport the data from about 650 data sources at the detector side, to the HLT farms for the processing of complete events. Each data source provides event fragments of about 2kB. The central DAQ runs software on about three thousands calculators used for buffering and processing of event data. A schematic representation of the DAQ system of CMS is shown in Fig. 2.17. The detector is read out through a network with a bandwidth of 100 GB/s by the so called Front-End Drivers (FED). The FEDs are located in the underground counting room ~ 70 m from the detector. Complete events are fed to the event filter systems at a maximum rate of 100 kHz. The large event rate fed to the filter systems stems from the design choice of CMS to build the full event already after the first software-level trigger instead of building partial events as in traditional multi-level trigger systems. This requires the read-out, assembly and forwarding of the full event data at the nominal level one trigger rate. The total rate of data produced by the online trigger system is ~ 230 MB/s. These data need to be stored for further processing and analysis.

2.2.6 The computing infrastructure

The CMS experiment poses new challenges in terms of the data volume and the necessary computing resources. Storage and computing resources requirements are significantly larger than in previous generation experiments such as LEP and Tevatron.

The CMS computing and storage requirements are impossible to fulfill in a single data center, both for technical and financial reasons. Therefore, the computing environment of

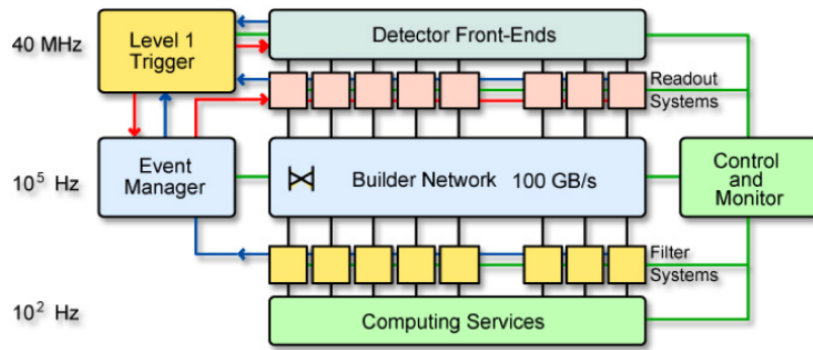


Figure 2.17: Scheme of the DAQ system [50].

CMS and the other LHC experiments has been constructed as a distributed system of computing resources communicating with each other spread all over the world at the local data centers of the collaborating institutions. This system is called the Grid. The resources available to the Grid are used for storage, data processing, and simulated events generation. Grid computing provides reliable and secure access to a large amount of widely scattered resources for users of the collaboration. The CMS Grid system is included in the larger Worldwide LHC Computing Grid Project (WLCG).

The computing centers available to CMS users through the Grid system are organized in a “tier” architecture. Figure 2.18 shows a schematic view of the CMS computing tier structure. Each of the four tier levels provides different resources and services:

- Tier-0: the main tier, for which there is only one site, CERN. Its tasks are to archive and distribute raw data collected from the CMS data acquisition system, derive calibrations, perform prompt reconstruction and transfer the results to Tier-1.
- Tier-1: there are seven Tier-1 (T1) sites, large centers in CMS-collaborating countries (large national laboratories or research institutes, such as DESY, INFN, or FNAL). Tier-1 sites archive and redistribute data samples to Tier-2 sites.
- Tier-2: a larger amount of smaller centers, but with substantial computing and storage resources. They are available for user data analysis and production of simulated events.
- Tier-3: small local clusters placed in collaborating institutions, they provide limited resources for local or remote users.

CMS data is arranged into a hierarchy of data tiers. Each collision event is written into each data tier, however different tiers contain different levels of information about the event. Figure 2.19 shows a schematic comparison of the content of the different data tiers used in CMS. The four main data tiers used in CMS are:

- RAW (GENSIM for simulated events): contains raw detector information (detector hits, trigger information), but no reconstructed event interpretation. This is not used directly for analysis. For simulated events this contains the simulated response of the detector to a collision event produced by a Monte Carlo generator.
- RECO: contains parts of the RAW information, and the reconstructed physics objects (tracks, calorimeter clusters, vertices) processed by the Tier-0. The event reconstruction procedure is structured in several steps:

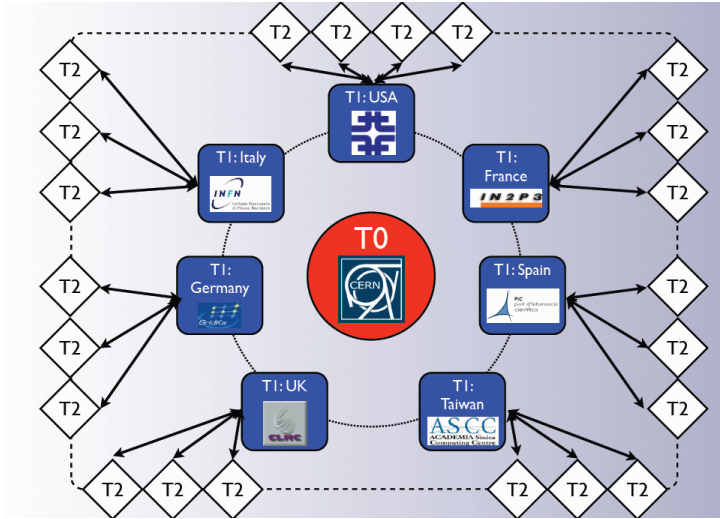


Figure 2.18: Schematic representation of the CMS grid tier structure [50].

1. Processing of detector-specific information: detector calibration constants, and reconstruction of clusters and hits.
2. Tracking: hits in the tracking system and the muon detector are fitted to reconstruct the tracks.
3. Vertexing: reconstructs primary and secondary vertex candidates.
4. Particle identification: produces the objects used in physics analyses. Using the particle flow reconstruction scheme (see Section 3.2), physics object candidates are created (electrons, photons, muons, and jets).

Reconstruction of the physics high-level objects is expensive in terms of computing resources and is dominated by the tracking algorithm. RECO datasets tend to be very large, therefore performing data analysis on these samples can be very time consuming.

- AOD (“Analysis Object Data”, AODSIM for simulated events): a reduced version of the RECO event content, most detector-level information is dropped as well as some reconstructed objects not commonly used in analyses.
- MINIAOD (MINIAODSIM for simulated events): a further skim of the AOD event content. Only high-level physics objects are kept. Despite the smaller amount of information contained in the MINIAOD format, it is still possible to re-run the jet clustering and b jet identification algorithms. This is a fundamental feature to be able to efficiently test and develop new high-level reconstruction algorithms.

The software framework used in CMS is called CMSSW, it contains all the software used to perform event interpretation from RAW to AOD and provides a powerful and efficient environment to develop data analyses. The CMSSW event processing model consists of one executable, called cmsRun, and many plug-in modules which are managed by the Framework. The CMS Event Data Model (EDM) is based on the concept of an event. An event is an object container for all raw and reconstructed data related to a particular collision. During processing, information about a collision are passed from one module to the next via the “event” container. All the modules composing a particular task are run on each event separately, making the parallelization of the task simpler. The CMSSW executable is configured at run time by a configuration file. The configuration contains information on:

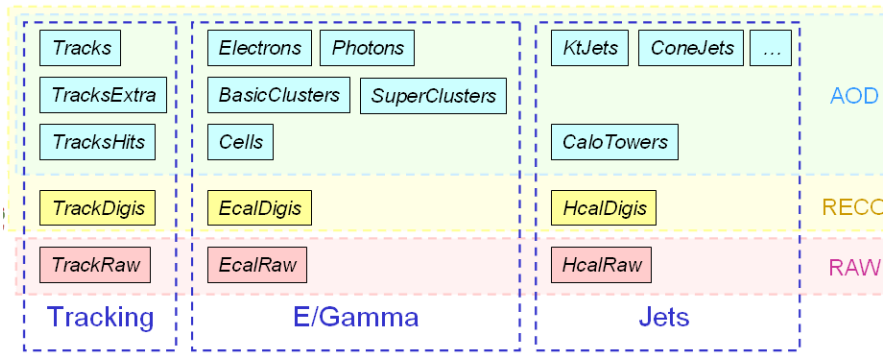


Figure 2.19: Schematic comparison of the content of the different data tiers used in CMS [50].

- which data to use,
- which modules to execute,
- which parameter settings to use for each module,
- the order of execution of the modules,
- and which events and what information about each event should be written to the output file.

2.2.7 Upgrades

The detector configuration described in this section is the one that has been used to collect the data analyzed in this thesis. However, as of 2017, a number of upgrades have been installed, collectively referred to as “Phase 1 upgrade”. The most relevant Phase 1 upgrade is the replacement of the pixel tracker with a new detector including a fourth pixel layer in the barrel and one additional forward disk. This will make the determination of the primary vertex more precise, increase b-tagging efficiency and tracking performance in general. Other improvements in the pixel include a reduced material budget and a new readout chip recovering inefficiencies in high-pileup data-taking scenarios. Upgrades in the other subsystems include improved photodetectors and more robust electronics in the hadronic calorimeter, improved muon triggering at L1 stage, and improved computing resources for the HLT and DAQ.

A more substantial upgrade, called Phase 2 upgrade [56], is planned for the high-luminosity LHC project (HL-LHC). The details of the upgrades are yet to be fully defined, but include replacements for most of the subsystems in order for the detector to be able withstand the very high radiation environment of the HL-LHC.

Event reconstruction

3.1 Luminosity calculation

It is essential for physics analyses at the LHC era to have a precise determination of the luminosity, this is true both for analyses performing precision measurement of Standard Model parameters (e.g. the cross section of a particular process) and for searches for physics beyond the standard model, to correctly normalize the simulated samples. The CMS experiment measures the luminosity both with “online” methods, that is during the data taking, and “offline” methods. The online methods provide a fast, but less precise determination of the instantaneous luminosity, which is useful for monitoring purposes. On the other hand, offline methods are available with some delay (at least a day), but have reduced systematic uncertainties, hence offering a precise determination of the luminosity that can be used in physics analyses. A method for measuring luminosity consists of a luminometer, that is a subdetector, an algorithm for counting the number the interactions, and a calibration of the luminometer. The calibration is a relationship between the number of counted interactions and the instantaneous luminosity. In general, the luminosity can be expressed as:

$$\mathcal{L} = \frac{R}{\sigma_{\text{vis}}}$$

where $\sigma_{\text{vis}} = \sigma_0 \mathcal{A}$ is the cross section of a know process times the luminometer acceptance (the “visible” cross section), and R is the rate measured with a particular luminometer. The absolute calibration of the luminometer counts is obtained using Van der Meer (VdM) scans with a special collider set up. In a VdM scan the two beams are gradually moved with respect to each other (“scanned”) in the transverse plane of the detector. This allows to measure the instantaneous luminosity in function of the machine parameters.

CMS uses five different subsystem to measure the luminosity [57]: the inner pixel tracker, the barrel drift tubes (DT), the forward hadronic (HF) calorimeter, the fast beam conditions monitor (BCM1f) and the pixel luminosity telescope (PLT). The method using the HF system is called “tower occupancy” and exploits the quasi-linear dependence of the instantaneous luminosity on the average energy deposited in the calorimeter towers. However, the method is sensitive to pileup, which makes the response behavior non-linear. Therefore the method does not yield a precise measurement and nowadays is used only as a crosscheck for other methods. The method using the inner pixel detector is called the “pixel cluster counting method” and it counts the average number of pixel clusters per event. The BCM1f is composed by eight diamond sensors placed around the beam pipe in the pixel detector volume. The subdetector allows a fast determination of the instantaneous luminosity that can be used

as a monitoring tool during data-taking. The PLT is another dedicated detector for luminosity measurement, newly installed in 2015, it is composed by eight 3-layered silicon pixel telescope detectors situated around the beam pipe just outside of the CMS coverage, at a small angle with respect to the beam pipe. This subdetector was designed to be able to provide a precise luminosity measurement in light of the very high pileup data taking conditions predicted for the 13 TeV data-taking era. An uncertainty of 2.6% is quoted for the integrated luminosity measurement of the 2012 dataset, while for the 2015 data run at 13 TeV an uncertainty of 2.3% was achieved.

3.2 The particle flow paradigm and object reconstruction

The algorithm used by CMS to reconstruct the physical objects from the detector output is called particle flow (PF) [58]. The output of the particle-flow reconstruction algorithm is used by most physics data analyses. Some features of the CMS detector, such as the high-resolution tracker and calorimeters, allow for this algorithm to provide an excellent efficiency in the reconstruction of high-level physics objects and pursue a global description of the collision event. This leads to a substantial improvement in jet performance and pileup mitigation.

The general idea of the particle-flow algorithm is to correlate detector objects (tracks, calorimeter clusters, etc.) and combining their information, to follow the path of a particle through the detector and measure its properties in the most complete and precise way. This leads to five categories of physical objects being reconstructed with this algorithm: muons, electrons, photons, charged hadrons, and neutral hadrons. The structure of the detector does not allow to distinguish between the different kind of hadrons. The first step is the reconstruction of intermediate-level objects such as tracks, calorimeter clusters and muon chamber segments from the raw detector output.

Tracks are reconstructed in the tracker using a combinatorial track finder derived from the Kalman-Filter (KF) algorithm [59]. The algorithm works in three steps: first the initial seed is formed with a few pixel hits, then the trajectory is build using hits from all tracker layers using a pattern recognition approach, and finally the track is fitted to determine the momentum, the origin, and direction. Several selection criteria are applied to the number of consecutive hits, the number of missed hits, track χ^2 fit, and others in order to avoid misreconstructed tracks. To increase the reconstruction efficiency, the track finder is iteratively run several times with gradually looser seeding criteria, but stronger track quality cuts to keep the misidentification rate at a low value. At each iteration step the hits from the previous iteration are masked to avoid double-counting. A total of 10 iteration steps are run, each targeting a class of tracks, i.e. displaced tracks or tracks inside high- p_T jets. This methods allows computationally efficient reconstruction of tracks with a momentum as low as 200 MeV. Nuclear interactions between a particle produced by the collision and the detector material are common and may change the trajectory of the particle and produce a number of secondary particles. A dedicated algorithm is implemented to reconstruct these interactions and the iterative tracking approach improves its efficiency. Electron crossing the tracker might radiate energetic photons and therefore substantially change their momentum. In such cases the KF algorithm might not be able to reconstruct the trajectory of an electron. The Gaussian-sum Filter (GSF) algorithm is better suited for this purpose as it allows higher energy losses along the trajectory. Generally speaking, the seed for the electron reconstruction include both KF and GSF tracks, and electromagnetic calorimeter clusters. Given the high efficiency for low- p_T tracks, the iterative tracking is beneficial to the reconstruction of prompt or bremsstrahlung photon conversions to electron-positron pairs. To collect the energy deposited by an electron and its bremsstrahlung emission, ECAL superclusters are defined as groups of clusters span-

ning a small η range and a wider window in ϕ surrounding the electron direction, to account for the electron bending.

Interaction vertices are reconstructed starting from a collection of tracks compatible with the beam spot. The deterministic annealing algorithm [60] is used to identify groups of tracks close to each other, comparing their point of closest approach to the beam line. These clusters become the reconstructed vertices. Among all the vertices, the one with the largest square sum of the p_T of the tracks is defined as the primary vertex. A precise determination of the position of the vertices is obtained via an adaptive vertex fit [61] of the tracks in the cluster to a common vertex.

In the muon system, hits from DT or CSC are clustered in track segments that are used as seed to form a muon track combining the hits from DT, CSC, and RPC. Muon segments can also be combined with tracks in the inner tracking system without a fully-reconstructed muon track.

Calorimeter clusters are used to:

- detect and measure the energy of photons and neutral hadrons,
- reconstruct electrons and their bremsstrahlung emissions,
- aid in the identification and energy measurement of charged hadrons,
- distinguish neutral particles from charged hadron deposits.

The procedure of clustering starts from a seed, defined as a single calorimeter cell with an energy higher than the four or eight adjacent cells. A “topological” cluster is then built by adding cells that have at least a corner in common with a cell belonging to the cluster and an energy above a certain threshold. The final clusters are determined by an iterative fitting algorithm where each seed is assigned a Gaussian distribution with parameters depending on its energy and position. The energy response of the calorimeter clusters is calibrated with simulated events, test beam data, radioactive sources, and early collision data events.

The core of particle flow is the link algorithm. Stable particles crossing the detector will generally leave a signature in different subdetector systems, therefore it is essential to link the objects from different subdetectors that come from the same particle. Pair of elements are produced by linking together nearest-neighbors in (η, ϕ) . The quality of the link is determined by the distance in the (η, ϕ) plane. A link between a tracker track and a calorimeter cluster is established by extrapolating the trajectory from the last measured hits of the track to the preshower, ECAL, and HCAL, up to a defined amount of radiation or interaction lengths. If the extrapolated track hits a calorimeter cluster, defined as the union of the cells belonging to the cluster plus one cell in each direction, then a link is established. If more tracks are linked to the same cluster or vice-versa, only the link with the smallest distance is kept. Links involving GSF tracks also take into account the effect of a potential bremsstrahlung photon on the calorimeter, by extrapolating tangents to the track. Electron-positron conversion and nuclear interaction finders exploit links between two tracks originating from a displaced vertex. Links between ECAL and HCAL (or preshower and ECAL) clusters are formed when the cluster position of ECAL (preshower) lies within the envelope of a cluster in HCAL (ECAL). Finally, a link between tracker tracks and muon tracks is found by extrapolating the tracks to a common plane and checking the compatibility between the parameters, while a link between a tracker track and a muon segment is established if the extrapolated track is compatible with a muon segment.

Given the PF blocks found by the link algorithms, the full event interpretation starts with the muons. Single muon tracks form “standalone muons”, tracker tracks linked to muon segments form “tracker muons”, while tracker tracks linked to muon tracks are refitted to form

“global muons”. Only global and tracker muons are considered by the PF algorithm. Isolated muons are identified by requiring the sum of p_T of the tracks and E_T of the calorimeter deposits within $\Delta R < 0.3$ of the muon direction to not exceed 10% of the muon p_T . Non-isolated muons are more difficult to identify and more stringent criteria are applied, such as requirements on the χ^2 of the global fit or on the number of matched muon segments. The muon momentum is chosen to be the one of the tracker track if the p_T is smaller than 200 GeV. Otherwise the p_T of the trajectory with the smallest χ^2 among various fits is used. The fits include only the tracker hits or the tracker hits plus a different selection of matched muon segments. The PF components used for the identified muons are not considered for the following step of the event reconstruction.

The next step involves electrons and isolated photons. The seed for isolated photons is an ECAL supercluster with $E_T > 10$ GeV and no link to a GSF track. Conversely, an electron seed is defined as a GSF track with a link to an ECAL cluster. Electron identification criteria consider fourteen different variables and are combined in a BDT discriminator. Isolated photons are identified by requiring the ECAL seed to be isolated from other tracks and calorimeter clusters. Additionally, the ratio between ECAL and HCAL energies is required to be compatible with that expected from a photon hitting the detector. All PF components used to reconstruct these objects are removed from the objects available for the next reconstruction steps.

At this point, only charged hadrons, neutral hadrons and nonisolated photons are left. Where tracking information is available ($|\eta| < 2.5$), ECAL and HCAL clusters not linked to any track are identified as photons and neutral hadrons, respectively. Outside of the tracker acceptance, ECAL clusters linked to an HCAL cluster are classified as hadrons (charged or neutral), and ECAL clusters with no link are again classified as photons. Remaining HCAL clusters are linked with one or more tracks that are in turn linked to ECAL clusters. If the calibrated calorimetric energy is much larger than the sum of track p_T , the PF block is split. Each track is assigned to a charged hadron with momentum taken from the track and energy computed assuming the mass of a charged pion, the excess calorimetric energy is then assigned to either a photon, a neutral hadron, or both. If the calibrated calorimetric energy is compatible with the sum of the p_T of the tracks, each track is associated to a charged hadron with a p_T obtained from a combined fit to the track and the calorimeters. In the remaining case a search for muons with relaxed identification criteria is performed.

Particle misidentification, albeit small, cannot be eliminated. Adjustments to the procedure have been introduced to account for identifiable misidentification cases. For example, specific criteria have been introduced to correct artificially large missing transverse p_T caused by cosmic muons crossing the detector.

Particles not interacting with the detector material leave an indirect signature in terms of missing transverse momentum (also referred to as MET). The simplest definition of MET is:

$$\text{MET} = - \sum_{i \in \text{PF comp.}} \vec{p}_{T,i}$$

However, a more precise measurement can be obtained by using corrected jets:

$$\text{MET} = - \sum_{i \in \text{PF comp.}} \vec{p}_{T,i} - \sum_{j \in \text{PF jets}} \vec{p}_{T,j}^{\text{corr}} - \vec{p}_{T,j}$$

The particle flow approach has been deployed in the high-level trigger. Due to the constraints given by the timing of the HLT, the algorithm must be simplified, however the modified version should have a performance as close as possible to the standard algorithm. Among the several modifications, the most important in terms of timing impact is the reduction of

the steps of the iterative tracking to only three iterations. Trigger paths using PF information have a superior performance compared to classic algorithms. Moreover, the availability of high-level PF objects at HLT level opens the possibility to use more sophisticated algorithms, such as b tagging and jet grooming to improve even further the trigger performance (see Section 4.5).

3.3 Hadronic jets

Jets are generated by the decay of unstable particles generated by the collision and the subsequent hadronization of the quarks produced in the decay. They are an ubiquitous feature of hadronic colliders. They are used in many physics analyses and their performance must be optimized for the LHC data-taking conditions.

The average jet is composed in terms of energy by: 65% charged hadrons, 25% photons, and 10% neutral hadrons [62]. Generally speaking, jets can be reconstructed starting from low-level objects such as calorimeter clusters and tracks, however using PF components as input for the jet clustering algorithms provides a vastly superior performance, therefore PF jets are the standard ones used in the CMS. Jet reconstruction is met with a problem at the LHC: the large number of pileup interactions. During the 8 TeV data-taking period an average of 20 pileup interactions occurred at every bunch crossing, the amount rose to 40 during the 13 TeV era and is expected to grow even further with the high-luminosity LHC upgrade. The number of pileup interactions can be determined from the number of primary vertices reconstructed using tracker tracks, or from measuring the instantaneous luminosity. The PF reconstruction cannot distinguish between particles coming from the hard interaction and pileup. However there are methods to mitigate the effect of pileup on jets. Each charged PF components can be assigned to a primary vertex. Among all reconstructed primary vertices the one with the largest sum of the squares of the p_T of the associated tracks is defined as the hard scatter vertex, and all the others are considered pileup vertices. Therefore, by simply not considering all the charged PF components associated to pileup vertices the effect of pileup coming from charged components can be greatly reduced. This technique is called Charged-Hadron Subtraction (CHS). Jet clustering algorithms can be run either on the complete set of PF component or on the PF components surviving the CHS procedure. However, the superior performance of the CHS jets, make the use of CHS jets the most adequate choice for most physics analyses. The CHS algorithm cannot take into account the effect of pileup coming from neutral PF components, however its behavior is predictable: a map of the average pileup neutral energy density can be extracted from minimum bias events in function of η and ϕ . Therefore the average neutral energy contribution from pileup can be subtracted from the jet depending on its position, jet area and transverse momentum per unit area (ρ). The effect of pileup on jets is not limited to energy scale and resolution, but includes spurious jets from overlapping low- p_T jets from different pileup interactions. CHS and jet quality criteria can greatly reduce the amount of pileup jets, for the remaining cases advanced techniques can be employed [63].

3.3.1 Jet clustering

The main clustering algorithm used in CMS is the anti- k_T algorithm [64], and it belongs to a class of clustering algorithms called sequential recombination jet algorithms. This class of algorithms can be described in a rather general way that will be detailed in the following.

The distance d_{ij} between entities i and j (PF components or proto-jets) is defined as:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} ,$$

while the distance between an entity i and the beamline d_{iB} is defined as:

$$d_{iB} = k_{ti}^{2p} ,$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti} is the transverse momentum, y_i the rapidity, and ϕ_i the azimuthal angle of entity i . The d_{iB} distance is computed for all the entities and the d_{ij} distance is computed for all the pair of entities. The minimum among computed distances is found. If the minimum is a d_{iB} distance, then the entity i is defined a jet and is removed from the list, else if it is a d_{ij} distance, entities i and j are merged together in a proto-jet. The procedure is iterated until no entities are left.

In CMS, during the 8 TeV run, the standard jets used a radius parameter R of 0.5, while in the 13 TeV run the size had been changed to 0.4 as it yields a better performance against pileup. Jets for t - and W -tagging use a larger radius of 0.8 or 1.5. The choice of the value for the p exponent determines the behavior of the algorithm. For $p = 1$ the algorithm is called k_T algorithm [65], for $p = 0$ the algorithm is called Cambridge-Aachen (CA) [66], while for $p = -1$ it is called the anti- k_t algorithm, the one used in CMS. The algorithm is infrared and collinear safe in all the three cases. In the k_T and CA cases, the algorithms first cluster softer particles together and later the softer clusters are joined with p_T -harder components. Conversely, in the anti- k_T case softer components are first clustered together with a nearby energetic component before joining softer isolated components. This leads to jet shape to be conical and more regular. A more regular shape for the jets makes jet calibration easier and more precise. However, in jet substructure algorithms the CA algorithm is often used as a starting point. Due to the order in which the proto-jets are merged, by undoing the clustering history of the algorithm, isolated soft clusters are easily separated from the p_T -hard core of the events. In large-cone jets ($R=0.8$ or 1.5) such isolated soft clusters might be coming from pileup, the underlying event, or the initial state radiation, degrading the jet performance. Soft and isolated clusters satisfying determined conditions can be removed: algorithms performing this task are called jet groomers.

3.3.2 Jet energy corrections

The calibration of jets in CMS follows a sequential approach where a number of corrections are applied on top of each other as a scaling of the 4-momentum of the jet in function of η , ϕ , and other variables. Figure 3.1 gives a schematic summary of the jet energy correction (JEC) procedure.

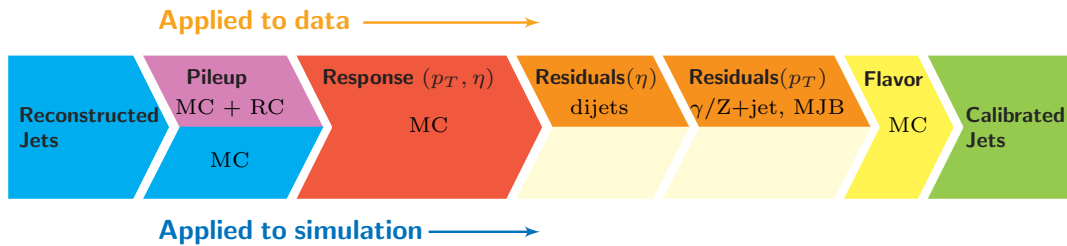


Figure 3.1: Schematic representation of the sequential jet energy correction procedure [62].

The first correction step, called L1, is the neutral pileup subtraction based on the jet area, as described in the previous section. It is applied to both data and simulated events. The second step, called L2-relative, aims at making the jet energy response linear in function of η . The correction is applied to data and simulation and is derived in simulated events by matching generator-level jets with their PF jets correspondents. The L3-absolute correction aims at correcting the nonlinearity of the energy response of the jets in function of p_T . This effect is caused by nonlinearity of the calorimeter response and is corrected in both data and simulation in a similar setup as the one used for L2-relative corrections.

After these steps of corrections a discrepancy in the behavior of jets is observed in data and simulation, therefore L2L3-residuals corrections are applied to data to correct for this discrepancy. The corrections are obtained in function of η and p_T and are measured in dijet data samples (η -dependent), or using a tag-and-probe technique on Z+jets and γ +jets samples (p_T -dependent). The tag-and-probe technique relies on the assumption that the momentum determination of the Z boson (decaying to a pair of muons) and the photon is more precise than that of the recoil jet, given that they are simpler objects to reconstruct. An optional flavor-based correction can be applied to both data and simulated events. The jet energy resolution is obtained by comparing the width of the response distribution $p_T^{\text{reco}}/p_T^{\text{gen}}$. Normally simulated events have a better resolution than data, therefore to make the simulation more similar to data the jet energy resolution is smeared in simulated events. The smearing factors are obtained from dijet and γ +jets samples.

3.4 B-jet identification

Jets coming from the hadronization of b quarks (b jets) play an important role in LHC physics, they appear in top quark decays, Higgs boson decays, and in the decay modes of many beyond-the-SM particles. B jets present a number of characteristics that allow them to be distinguished from jets coming from the hadronization of light quarks or gluons. The most prominent feature of b jets is that they originate from a secondary vertex displaced from the primary vertex of interaction, this is a consequence of the long decay lifetime of the B mesons contained in the hadronization history of a b quark. The long decay lifetime of B mesons is caused by the weak nature of their decays. Other features of b jets often include the presence of a soft lepton (coming from the weak decay process) and a larger jet mass. A special case are c jets, they present features somewhat similar to b jets, so it is more difficult to distinguish them from b jets, with respect to light quarks (u,d,s) and gluon jets.

These features, along with others are exploited by b-tagging algorithms. Two classes of discriminating variables are defined: track variables include a number of counting and kinematic variables on the particle-flow tracks associated to the jet via a jet-to-track association algorithm. The tracks used to compute the discriminating variable are required to pass quality criteria that depend on the different data-taking periods. The other class is the secondary-vertex variables: an algorithm to find secondary vertices inside the jet is run and a number of counting and kinematic variables are computed using the properties of the vertices found. The details of the variables are reported in Chapter 5. The main algorithm used in CMS is the combined secondary vertex (CSV) [67]. The algorithm uses a likelihood function to combine the information from track and secondary vertex variables, if a secondary vertex is found. If no secondary vertex is found, the algorithm tries to reconstruct a pseudo secondary vertex using only tracks with a good resolution of the impact parameter. If this also fails, then the algorithm uses exclusively information from the track variables. Over the years, several improvements have been implemented to the original CSV algorithm used in the 8 TeV era. The standard CSV version used for 13 TeV data collected in the year 2015 was called

CSVv2 [68] and used the inclusive vertex finder (IVF) [68] to reconstruct secondary vertices. The IVF uses all tracks in the events to reconstruct secondary vertices, therefore it is not limited by the jet-to-track association algorithm. This improvement is particularly important for close-by b jets, such as the ones generated by the boosted decay of a Higgs boson to a pair of b quarks. Moreover, the CSVv2 combines the variables using a neural network instead of a likelihood ratio. Other improved taggers have been developed such as the combined multivariate algorithm (cMVA2) [68], that combines several tagging algorithms (including CSVv2) using a boosted decision tree. Finally, the DeepCSV [69] algorithm is based on the same set of variables used by the CSVv2 tagger, but uses a deep neural network training to combine the information from the different observables.

The b-tagging algorithm is normally applied to standard AK5 (for 8 TeV) or AK4 (for 13 TeV) jets clustered using PF components. However b-tagging can be adapted to be used on larger-cone jets and, more importantly, on the subjets of large-cone jets. This helps in the identification of boosted hadronic decays of top quarks and Higgs bosons. A study of the performance of b tagging applied to subjets is described in Chapter 5. In the analysis described in Chapter 7, the likelihood-based CSV algorithm is used, while the analysis described in Chapter 8 uses the IVF-based CSVv2 algorithm.

Data-simulation correction factors (scale factors, SF) are needed to correct for the difference efficiency ϵ of the b tagging algorithms in data and simulated events. Scale factors are provided in function of η and p_T of the jet, for specific operating points of the algorithm. In addition, scale factors are provided both for b quark identification efficiency $SF_b = \epsilon_b^{data} / \epsilon_b^{sim}$, and for light quarks (misidentification probability) $SF_{light} = \epsilon_{light}^{data} / \epsilon_{light}^{sim}$. The measurement of SF_{light} is based on the concept of negative taggers. Negative taggers are equal to the actual b-tagging algorithms, except for using exclusively tracks with negative impact parameter and secondary vertices with negative decay lengths as input. Light jets are expected to have the same discriminator distributions for negative and positive taggers, while b jets rarely pass the selection of the negative taggers. Therefore, the misidentification probability can be derived from the number of events selected by the negative tagger (modulo minor corrections from the contamination of real b jets). The measurement of SF_b is a combination of methods covering different corners of the phase space. One method uses soft leptons to select a pure sample of b jets with a dijet topology. The efficiency is also measured in dileptonic and semileptonic $t\bar{t}$ events using the flavor tag matching (FTM) and flavor tag consistency (FTC) methods, respectively. The two methods perform a likelihood fit on the distribution of the multiplicity of b tags to extract the b tagging efficiency.

3.5 Event simulation

The proper simulation of collision events is an important aspect of LHC physics. Simulated samples are used in the development and validation of reconstruction algorithms and hardware upgrades. Additionally, they are used to model the expected distributions of SM processes and to determine the yield of signal events in searches of physics beyond the SM. Therefore they play a role in the statistical interpretation of the result of a search.

Given the aleatory nature of high-energy collision processes, Monte Carlo (MC) methods are used to generate draws from the probability distribution of a certain physical process to happen. Generating realistic collision events is a technically challenging and computationally intensive task. This section will summarize the main steps used to generate simulated samples of LHC collisions.

The first step is the simulation of the hard scatter process, for example $gg \rightarrow t\bar{t}$. The matrix element (ME) generator will compute the relevant matrix elements for the Feynman diagrams

contributing to the process, and calculate the differential cross sections. In hadron colliders, the full computation must take into account of the probability to find a determined parton within the proton. Parton Distribution Functions (PDF) are derived from global fits to experimental data. The HERA collider provided essential measurements for the determination of the PDFs used at the LHC [70].

The second step is the parton showering. In this step the particle shower coming from the radiation of the initial and final state particles of the hard scatter process is simulated. Generally speaking, showers are higher-order corrections of the hard process. Hard and large-angle radiation can be simulated by the matrix element generator, but the contribution from soft- and small-angle radiation cannot be calculated exactly. Therefore, parton showers algorithms employ statistical methods and parton splitting functions to model the showering. In parts of the phase space, higher-order emissions can be simulated both by the matrix element and the parton shower. To avoid double counting, a matching algorithm (such as MLM [71]) must be used to remove the overlapped emissions.

The third step is the hadronization, where final state partons are transitioned to colorless hadrons. This is a non-perturbative process, and there are different algorithmic approaches, such as cluster model and the string model [72]. The fourth step includes adding to the event the simulation of the underlying event (UE), that is the small momentum-transfer processes involving the remnants of the protons containing the hard-scatter partons. Additionally, for a realistic simulation, pileup (PU) interactions must be added to the event. These are low-energy pp interactions that occur in the background of the hard-scatter process. Pileup collisions can be traced back to the same bunch-crossing of the hard scatter (in-time PU) or to collisions occurring in bunch-crossings right after or before the collision of interest (out-of-time PU).

The fifth step is the detector simulation. The interactions of stable particles with the detector material is simulated, along with the electronic response of the detector (EM showers in ECAL, tracker hits, etc). This requires an accurate model of the geometry of the detector. Detector simulation is usually done with GEANT4 [73].

Jet substructure tools

The searches for new physics presented in this thesis contain highly-energetic all-jet decays of top quarks ($t \rightarrow bW \rightarrow bqq'$) or W bosons ($W \rightarrow qq'$). Focusing the discussion on the top quark, its all-jet decay is traditionally reconstructed using three standard-sized jets, however as the transverse momentum of the top quark increases, the angular separation of the decay products decreases (large Lorentz-boost). At a transverse momentum of around 200 – 300 GeV the decay products are at least partially merged together and cannot be reconstructed with standard techniques. The paradigm of jet substructure techniques consists in collecting all the decay products of the top quark in a single jet with a larger radius parameter (0.8-1.5 versus the standard 0.4-0.5) and then recovering the information on the particle and its decay products using dedicated algorithms. Figure 4.1 shows a schematic representation of the jet substructure paradigm.

A “top quark tagger” is an algorithm that, starting from a large-cone jet, decides whether the jet comes from the hadronization of a top quark, and provides some information such as the measured mass and p_T . Generally speaking, a top quark tagger is composed by a “grooming algorithm” that cleans the jet from components not coming from the decay of the top quark, and one or more algorithms combined together acting on the jet components surviving the grooming procedure. A combination of selection criteria using the aforementioned algorithms will decide if the jet is “top tagged” or not. As mentioned before these techniques are applicable not only to the top quark, but to W, Z, and Higgs bosons as well (referred to as “heavy SM particles” in the following). Figure 4.2 shows graphical representation of the decays where jet substructure techniques can be used. This chapter will describe some of these algorithms and taggers, with a focus on top quark taggers.

4.1 Jet grooming

Large-cone jets ($R=0.8$ or 1.5) are used to cluster the boosted hadronic decay of heavy SM particles. Given the large area collected by these jets it is likely that in an average busy LHC collision event, a number of particles coming from other decay processes (pileup, underlying event, or initial state radiation) will end up being collected by the jet. As explained above, jet groomers are a class of algorithms that act on large-cone jets with the aim of removing soft and wide angle components that might arise from processes outside of the decay of the particle the jet aims to reconstruct. Moreover some of the jet groomers also find a number of “subjets”, that is clusters within the jet that can be associated with the single decay products of heavy SM particles. Generally speaking, a jet groomer is composed by a first stage where subclusters are found, this can be achieved by either reclustering the jet components with a

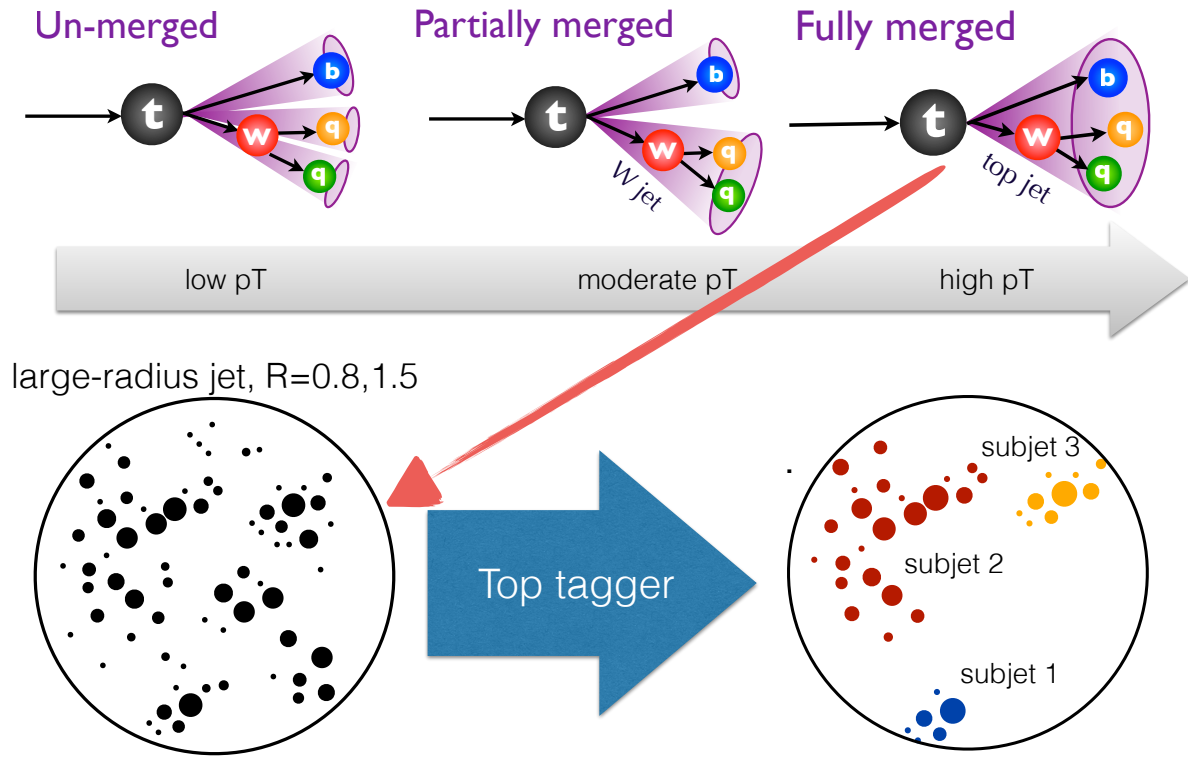
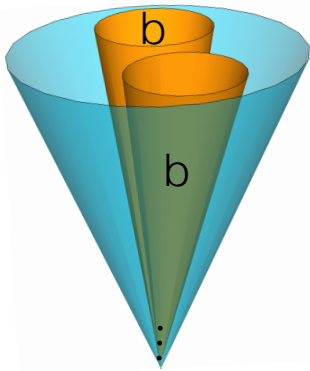
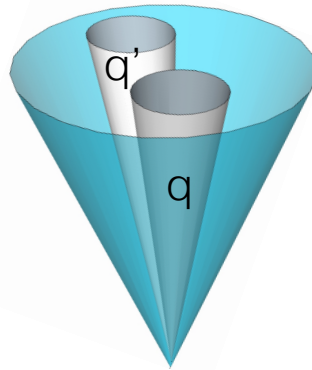


Figure 4.1: Schematic representation of the jet substructure paradigm.

$$H, Z \rightarrow b\bar{b}$$



$$W \rightarrow q\bar{q}'$$



$$t \rightarrow b q\bar{q}'$$

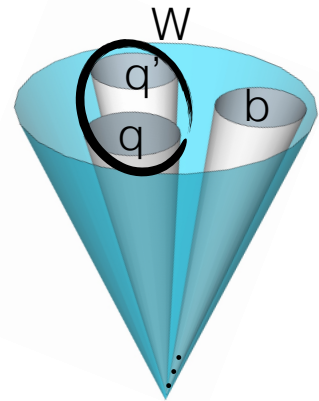


Figure 4.2: Graphical representation of the decays where jet substructure techniques can be used.

different algorithm or by iteratively undoing the clustering procedure of a specific jet algorithm. The subclusters are then checked against a number of requirements and rejected if the requirements are not met.

The final jet is usually composed by the surviving subjets or clusters, but sometimes further steps are employed. One of the main advantages of using a jet groomer is that the “groomed jet mass”, that is the jet mass computed using only the component that survive the grooming procedure, is a much more reliable quantity with respect to the ungroomed jet mass and is a powerful discriminator against QCD multijet, a background process to many all-jet analyses. Jets originating from light flavor quarks or gluons will have a small groomed jet mass compared to that of jets coming from the decay of heavy SM particles. The standard jet grooming algorithm used in the 13 TeV data taking era is the soft-drop groomer [74], described in Section 4.4.2. In the 8 TeV data taking era different algorithms were used, some of them are described in Sections 4.3 and 4.4.1. In this section the filtering grooming algorithm [75] will be described, which is used by the HLT paths described in Section 4.5. Figure 4.3 shows the schematic representation of some grooming algorithms.

The filtering algorithm [75] starts by feeding the jet components to a k_T algorithm with radius parameter R_{filt} . A typical value for the radius parameter is 0.3. The clustered jets are considered the subjets of the original jet. Only the subjet satisfying the following condition are kept:

$$p_T^{\text{subjet}} / p_T^{\text{jet}} > f_{\text{cut}}$$

where p_T^{subjet} is the transverse momentum of the subjet, p_T^{jet} is the transverse momentum of the original jet, and f_{cut} is a parameter that regulates the aggressiveness of the grooming. The final groomed jet is composed by the subjets that passed the f_{cut} condition.

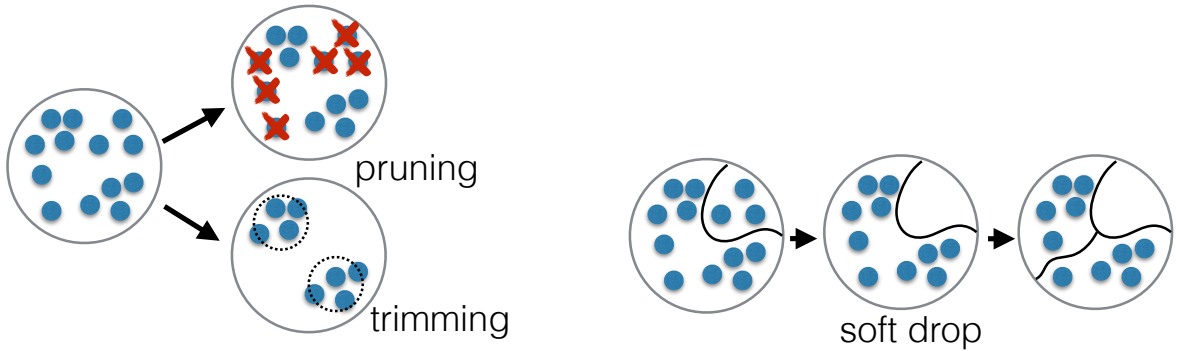


Figure 4.3: Schematic representation of some grooming algorithms.

4.2 The N-subjettiness algorithm

N-subjettiness is a jet shape variable designed to measure how consistent a jet is with having N or fewer subjets [76,77]. The variable is defined by:

$$\tau_N = \frac{\sum_{i=1}^{N_{\text{constituents}}} p_{T,i} \min\{\Delta R_{1,i}, \Delta R_{2,i}, \dots, \Delta R_{N,i}\}}{\sum_{i=1}^{N_{\text{constituents}}} p_{T,i} R} \quad (4.1)$$

Where the summation runs over all jet constituents and N represents the number of subjet axes that are being tested. $p_{T,i}$ is the transverse momentum of the i -th jet constituent.

$\min\{\Delta R_{1,i}, \Delta R_{2,i}, \dots, \Delta R_{N,i}\}$ is the minimum $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ distance between the jet constituent i and all the subjet axes. R is the original jet clustering radius parameter. The denominator is a normalization factor which ensures $0 < \tau_N < 1$.

The τ_N variable can be seen as the p_T -weighted sum of the angular separation between each jet constituent and the closest subjet axis. p_T -weighted means that the jet components with a higher transverse momentum have a higher weight in the summation. Small values of τ_N happen when the jet constituents are closely aligned with the subjet axes. Therefore, jets with a small values of τ_N can be interpreted as being consistent with having N or fewer subjets. The N -subjettiness variable becomes a more effective discriminator by taking the ratio of jet shapes: τ_N/τ_{N-1} . A large-radius jet containing the hadronic decay of a top quark is expected to have three prongs and thus τ_3/τ_2 has a good discrimination for such decays. Similarly a jet containing the hadronic decay of a W boson is expected to have two prongs, therefore τ_2/τ_1 is a powerful variable to distinguish W decays from background events.

Generally speaking, the axes of the subjets that enter the calculation of the N -subjettiness variable can be found by any clustering algorithm that finds subjets. The two most used techniques to find subjet axes to compute the N -subjettiness variable are the exclusive k_T algorithm [65] and one-pass optimization of the exclusive k_T algorithm [77]. The one-pass optimization differs from the base version of the algorithm by varying the direction of the subjet axes such that N -subjettiness is minimized. The one-pass optimization offers a better discrimination between real top quark decays and QCD multijet background.

The N -subjettiness variable is infrared (IR) safe [78], however the ratio of two N -subjettiness variables such as τ_N/τ_{N-1} is not IR safe. τ_3/τ_2 can be made IR safe by making a selection on τ_2/τ_1 [78]. This selection on τ_2/τ_1 is however almost 100% efficient when applied together with a τ_3/τ_2 selection and a requirement of the groomed jet mass to be inside an interval around the top quark mass.

4.3 The HEPTopTagger algorithm

This section will describe the HEPTopTagger (HTT) algorithm. The algorithm was introduced by Plehn et al. [79] and it is tailored for the reconstruction of hadronically-decaying top quarks with a medium boost, that is with a transverse momentum of the top quark between 200 GeV and 600 GeV. Commissioning of subjet b -tagging on HEPTopTagger jets is presented in Chapter 5. The algorithm optimization and its commissioning with 8 TeV data are described in Chapter 6. This algorithm has been used in the analysis presented in Chapter 8.

The HEPTopTagger (HTT) starts by clustering jets with the Cambridge/Aachen algorithm with a distance (or radius) parameter $R = 1.5$ (CA15). In the rest of the thesis, a “top jet” is defined as the large-radius jet clustering the hadronic decay of a top quark. The substructure of the jet is identified by reversing the clustering history of the jet and applying the mass-drop decomposition algorithm [80] (Fig. 4.4). The algorithm works as follows: the jet is decomposed according to the last clustering step into two subjets j_1 and j_2 , with $m_{j_1} > m_{j_2}$. If the mass drop condition $m_{j_1} < 0.8m_j$ is satisfied, then the algorithm attempts to further decompose j_1 and j_2 . If the mass drop condition fails, then j_2 is discarded and the algorithm attempts to further decompose j_1 . The procedure is iteratively repeated for every subjet found within the jet with a mass greater than 30 GeV. If the subjet has a mass smaller than 30 GeV or has only one constituent, then the subjet is saved for the next step of the HEPTopTagger algorithm. No limit is imposed on the number of subjets that can be found in this step of the algorithm, however if less than three subjets are found, the jet is rejected by the HTT. A flow chart of the mass drop procedure is shown in Fig. 4.4 and a cartoon is shown in Fig. 4.5.

HEP Top Tagger Mass drop decomposition

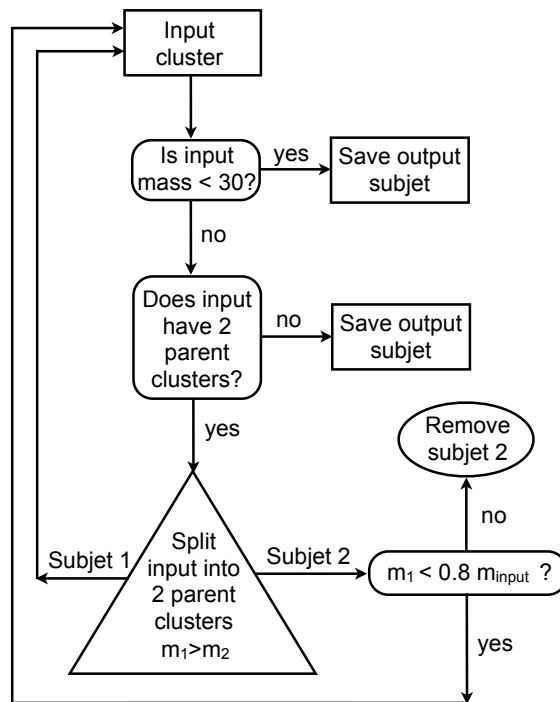


Figure 4.4: Flow chart for the HEPTOPTAGGER mass-drop decomposition algorithm [81].

The procedure continues by considering every possible combination of three subjects within the jet (triplet). For each of the triplets, a filtering algorithm is applied as shown in Fig. 4.6. The filtering algorithm reclusters the constituents of the three subjects with a variable distance parameter $R_{\text{filt}} = \min(0.3, \Delta R_{ij}/2)$, where i, j is the pair of subjects that minimizes the angular separation ΔR as depicted in Fig. 4.7. The filtering procedure finds a number of “reclustered subjects”, but only the five reclustered subjects with the largest p_T are retained, as shown in Fig. 4.8. These five subjects are referred to as the “filtered subjects”. The “filtered mass” is defined as the invariant mass of the 4-vector obtained by summing the 4-vectors of the five filtering subjects. Among all the possible triplets of mass-drop subjects, only the one with the filtered mass closest to the true top quark mass is retained, while all other combinations are discarded, as shown in Fig. 4.9. The constituents of the filtered subjects are then reclustered again using the exclusive Cambridge/Aachen algorithm forcing exactly three final subjects, as depicted in Fig. 4.10.

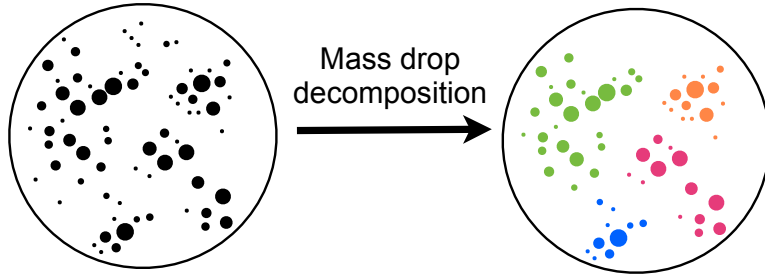


Figure 4.5: First step of the HEPTopTAgGER: the mass drop decomposition. The dots represent the particle-flow constituents. The components from the different subjects are drawn with different colors [81].

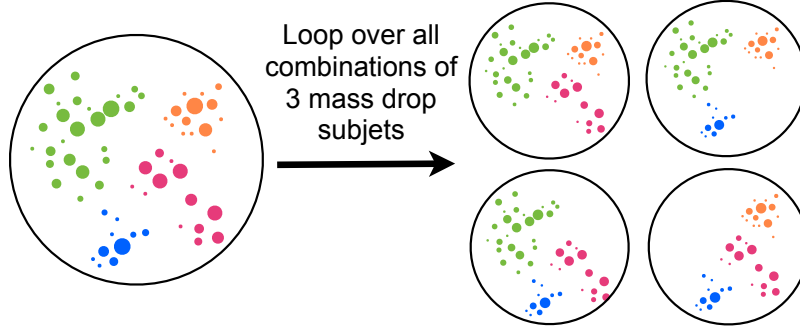


Figure 4.6: Second step of the HEPTopTAgGER: all the possible combinations of three mass-drop subjects (triplets) are considered [81].

The HEPTopTAgGER uses these three final subjects to tag jets by making selections on the following quantities:

- m_{123} – The invariant mass of the sum of the 4-vectors of the three final subjects.
- m_{12} – The invariant mass of the sum of the 4-vectors of the two highest p_T subjects.

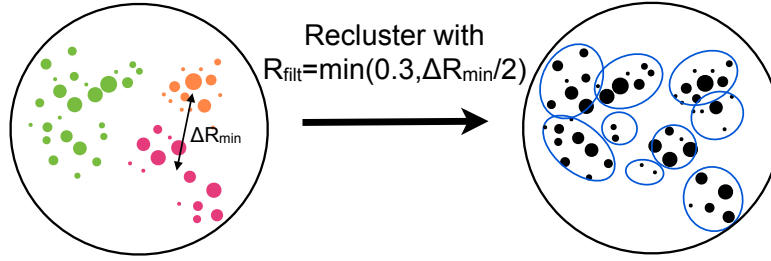


Figure 4.7: Third step of the HEPTopTAgGER: the constituents of each triplet are reclustered. The distance parameter is determined by the ΔR between the two closest mass-drop subjects in each triplet. New subjects are found by the reclustering procedure [81].

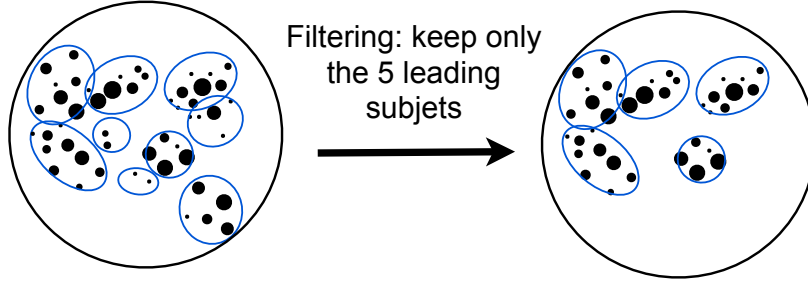


Figure 4.8: Fourth step of the HEPTopTAgGER: jet filtering. Only the five subjects with the highest p_T are retained [81].

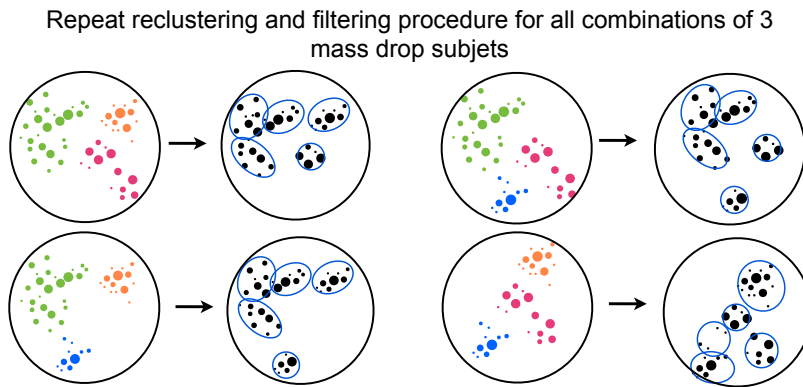


Figure 4.9: Fifth step of the HEPTopTAgGER: all the triplets of mass-drop subjects are passed through the filtering algorithm [81].

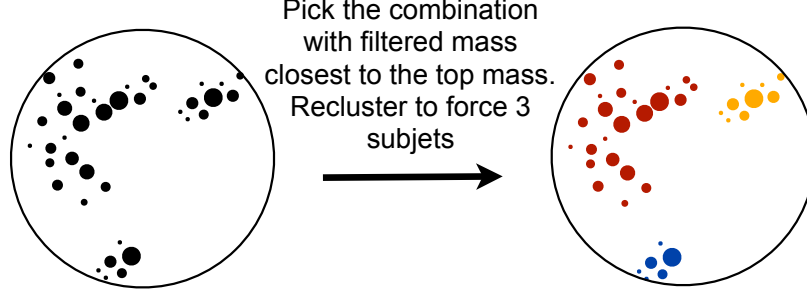


Figure 4.10: Sixth step of the HEPTopTAgGER: only the triplet with the filtered mass closest to that of the top quark mass is kept. Reclustering is performed to find exactly three subjects [81].

- m_{13} – The invariant mass of the sum of the 4-vectors of the highest p_T and the lowest p_T subjects.
- m_{23} – The invariant mass of the sum of the 4-vectors of the two lowest p_T subjects.

The HEPTopTAgGER identifies top jets by requiring the mass of the final three subjects (m_{123}), that is the reconstructed mass of the top candidate, to be within an interval around the real top quark mass. Additionally, a selection on the subject pairwise masses (m_{12}, m_{13}, m_{23}) is applied to identify the W boson within the top jet:

$$c_{\text{low}} < \arctan \frac{m_{13}}{m_{12}} < c_{\text{up}} \quad \text{and} \quad R_{\text{min}} < \frac{m_{23}}{m_{123}} < R_{\text{max}}$$

$$R_{\text{min}}^2 \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\text{max}}^2 \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right) \quad \text{and} \quad \frac{m_{23}}{m_{123}} > c_{\text{soft}}$$

$$R_{\text{min}}^2 \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\text{max}}^2 \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right) \quad \text{and} \quad \frac{m_{23}}{m_{123}} > c_{\text{soft}}$$

with $R_{\text{min}} = (1 - f_W) \times m_W / m_t$ and $R_{\text{max}} = (1 + f_W) \times m_W / m_t$. The requirements for the W boson identification are outlined in Fig. 4.11. These selections are motivated by three-body kinematics of the top decay, and can be obtained by writing the equation of the top invariant mass in function of the subject pairwise masses. The default values for the HEPTopTAgGER W identification parameters, as defined by the HEPTopTAgGER authors [79], are: $f_W = 0.15$, $c_{\text{low}} = 0.2$, $c_{\text{up}} = 1.3$, $c_{\text{soft}} = 0.35$. The parameter f_W is the width of the W mass selection region. The choice of this variable has the largest effect on the algorithm performance. This parameter is varied and optimized in Section 6.1.2. In the commissioning studies presented in Chapter 6 and in the analysis described in Chapter 8, the top quark mass selection window is fixed to be $140 \text{ GeV} < m_{123} < 250 \text{ GeV}$.

The HEPTopTAgGER algorithm is only able to identify top quarks with a transverse momentum greater than 200 GeV, at lower transverse momenta the decay products of the top quark are not expected to be merged in one single jet with a distance parameter of $R = 1.5$. In such cases, a classic fully-resolved strategy with a three-jets topology should be used. Alternatively, a two-jet topology might be used, where one is a standard jet associated with the b quark and the other is a large-radius jet associated with the W boson decay.

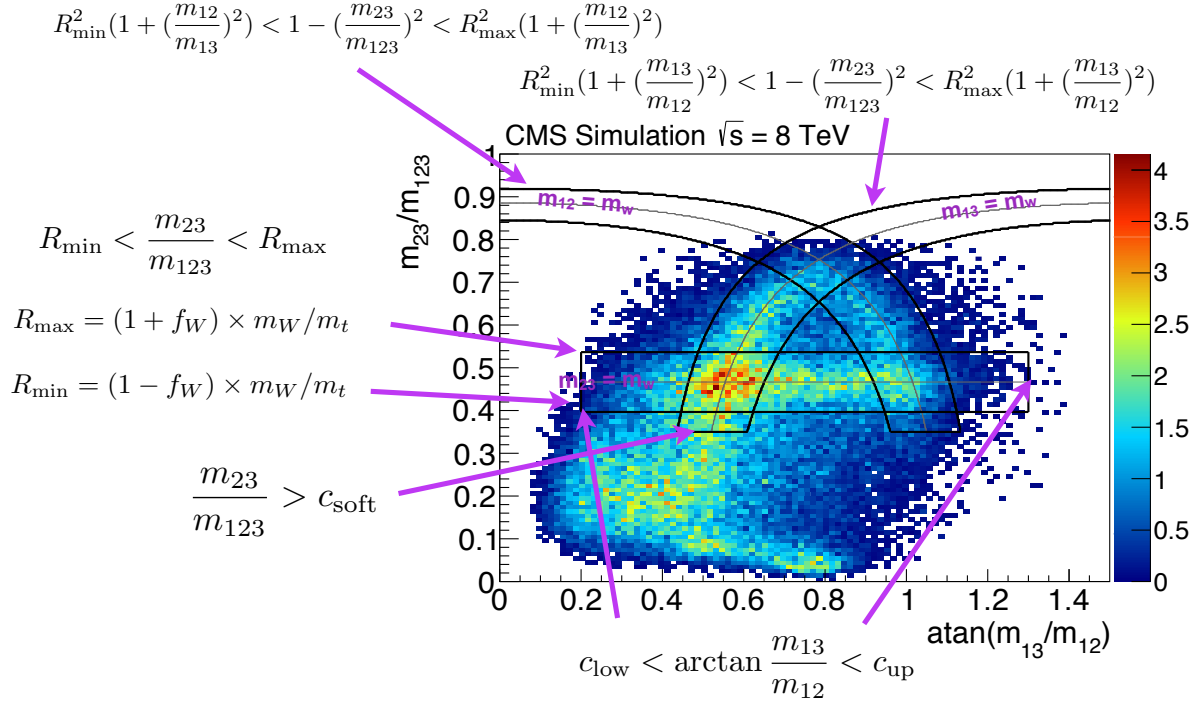


Figure 4.11: HEPTopTAgger W identification in the plane defined by m_{23}/m_{123} and the inverse tangent of m_{13}/m_{12} for a $t\bar{t}$ simulated signal sample (semileptonic decay mode only) [81].

4.4 The CMS Top Tagger algorithm

4.4.1 Version used with 8 TeV data

The CMS top tagging algorithm was used to reconstruct the decay of a top quark with a p_T larger than 400 GeV in the data-taking era at a center-of-mass energy of 8 TeV. The top tagger is based on the algorithm developed by Kaplan et al. [82].

The algorithm uses Cambridge-Aachen jets with a radius parameter $R = 0.8$ (CA8) and their constituents as a starting point. The algorithm has two steps: the primary decomposition, where the initial jet is split in two subclusters, and the secondary decomposition, where the clusters from the primary decomposition are further split if some conditions are met. The decomposition algorithm is applied to the initial CA jet in the primary decomposition, and to the subclusters found by the primary decomposition in the secondary decomposition. The decomposition algorithms works as follows:

1. The pairwise clustering sequence from the CA algorithm, which was used to assemble the jet, is unfolded in reverse order to find two subclusters.
2. Continue to the next step if the two subclusters satisfy $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.4 - A \times p_T^C$, where p_T^C is the transverse momentum of the input jet or subjet of the decomposition algorithm. The slope parameter $A = 0.0004$ is optimized using simulated events [83]. This selection is generally not satisfied if the subclusters are too close, in this case the decomposition fails and no subclusters are found. The primary decomposition uses the original CA jet p_T , while the secondary decomposition uses the p_T of the subclusters found by the primary decomposition.

3. The momentum fraction criterion $p_T^{\text{cluster}} > \delta_p \times p_T^{\text{hardjet}}$ is checked on each of the subclusters. If the condition is met by both the subclusters, then the decomposition succeeds. This condition removes low- p_T clusters from consideration. The value $\delta_p = 0.05$ was optimized in simulated events [83]. If only one of the subclusters satisfies the criterion, then the decomposition process is repeated on the cluster passing the requirement, removing the constituents from cluster failing the requirement. This step is repeated until both clusters pass or fail, or the cluster consists of a single constituent.
4. If, after this iterative process, none of the two subclusters pass the momentum fraction criterion, or the subcluster is a single constituent, the decomposition fails.

The primary decomposition recursively unclusters the hard jet until it finds two subclusters, well separated and containing a significant fraction of the transverse momentum of the initial jet. If the primary decomposition fails then the jet has only one subjet (the original jet). If the primary decomposition succeeds, then the secondary decomposition is applied to the two clusters found. If the two clusters can not be further decomposed then they become the jet's only two subjets. If both are successfully decomposed then the jet has four subjets. If the secondary decomposition succeeds on one cluster and fails on the other, then the jet has three subjets. The CMS Top Tagging algorithm requires at least three subjets to be present in the jet. The following jet variables can be used to provide further discrimination value:

- Groomed jet mass m_{jet} - The mass of the four-vector sum of the constituents of the subjets identified within the hard jet. These constituents, due to the jet grooming procedure, are a subset of those found in the hard jet.
- Number of subjets N_{subjets} - The number of subjets found by the algorithm.
- Minimum pairwise mass m_{min} - For all the possible pairs of the three highest p_T subjets the pairwise invariant mass is calculated. m_{min} is the mass of the pair with the lowest invariant mass ($m_{\text{min}} = \min[m_{12}, m_{13}, m_{23}]$). This variable is not defined for jets with less than three subjets.

The standard CMS top tagging working point is defined by:

$$140 < m_{\text{jet}} < 250 \text{ GeV} \quad (4.2)$$

$$N_{\text{subjets}} \geq 3 \quad (4.3)$$

$$m_{\text{min}} > 50 \text{ GeV}. \quad (4.4)$$

4.4.2 Version used with 13 TeV data

In the data-taking era at a center-of-mass energy of 13 TeV, the CMS Top Tagger V2 is used. In this new version of the algorithm, the constituents are initially clustered with the anti- k_T algorithm using a radius parameter $R=0.8$ (AK8 jet). The constituents of the AK8 jet are reclustered using the Cambridge-Aachen algorithm [66,84]. Then the modified mass drop tagger (MMDT) algorithm [80], also known as the “soft drop” [74] algorithm is used to remove soft, wide-angle radiation from the jet. The soft-drop jet grooming algorithm is composed of the following steps:

1. The jet is subdivided into two subjets (j_1 and j_2) by reversing the last step of the of CA clustering.

2. The soft drop condition is checked on the two subjets:

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta$$

with angular exponent $\beta = 0$, soft threshold $z_{\text{cut}} < 0.1$, and characteristic radius $R_0 = 0.8$. If the subjets pass the soft drop condition then the jet used as input is the final soft-drop jet.

3. Otherwise, the softer subjet is removed and the procedure is iterated on the other subjet.
4. If the jet has a single constituent then the jet used as input is the final soft-drop jet.

An important difference with respect to the previous version of the CMS Top Tagger is that this algorithm identifies exactly two subjets within the AK8 jet. Top jet candidates are selected by requiring the groomed jet mass to be within an interval around the top quark mass. Additionally, to identify the n-legged substructure of the hadronic top quark decay a selection the N-subjettiness variable $\tau_3/\tau_2 = \tau_{32}$ is done.

By varying the groomed jet mass window and the N-subjettiness intervals, different efficiency operating points for the tagger can be obtained. The working point used in Chapter 8, is defined by requiring that the groomed jet mass satisfies $110 < M_{\text{SD}} < 210 \text{ GeV}$ and the N-subjettiness variable satisfies $\tau_{32} < 0.86$, which corresponds to a misidentification rate in simulated events of 10%, and an efficiency greater than 70% for jets with a transverse momentum greater than 400 GeV [85]. Figure 4.12 shows an event display of a boosted top quark decay identified by the CMS Top Tagger algorithm.

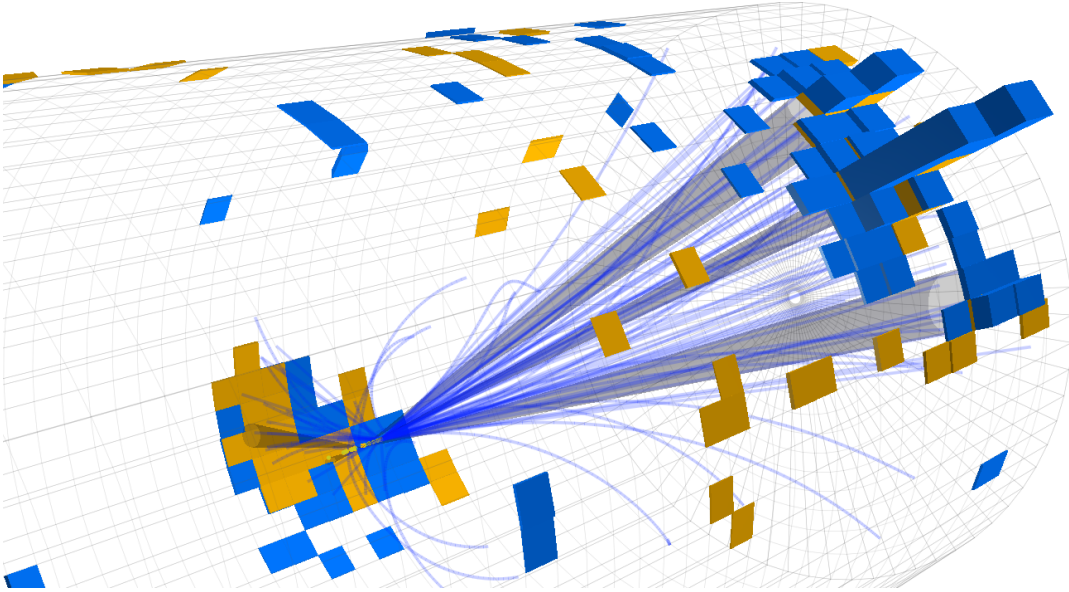


Figure 4.12: An event display of a boosted top quark all-jet decay. The yellow towers are ECAL clusters, the blue towers are HCAL clusters. The size of the towers is proportional to their energy. The tracks are charged particle-flow candidates.

4.4.3 Tagging W bosons

For the identification of W bosons decaying hadronically, several approaches have been developed [86, 87]. The procedures generally use a jet clustered with the anti- k_T algorithm, then a

jet grooming algorithm is applied to remove soft and wide-angle clusters. Finally a selection is applied on the groomed jet mass and the N-subjettiness variable $\tau_2/\tau_1 = \tau_{21}$

The W-tagging algorithm used in Chapter 8 uses the same reclustered jets generated by the CMS Top Tagger V2 algorithm, but with different selection criteria applied on the jet substructure variables. In particular, the interval on the groomed jet mass is $70 < m_{\text{jet}} < 100 \text{ GeV}$ while the N-subjettiness variable $\tau_2/\tau_1 = \tau_{21}$ is required to be smaller than 0.6. These criteria correspond to a misidentification rate of approximately 5% for multijet QCD, and an efficiency of approximately 60% for jets with a transverse momentum greater than 200 GeV [87].

4.5 Triggers using jet substructure techniques

4.5.1 Introduction and trigger description

Searches for physics beyond the standard model with boosted heavy SM particles in the all-jet final states pose a problem in terms of trigger selection. Traditionally, all-jet searches use a class of triggers called “ H_T triggers”. These trigger simply selects events with a certain amount of energy deposited in the detector. The H_T variable is calculated by summing the energy of standard-size ($R=0.5$ at $\sqrt{s} = 7, 8 \text{ TeV}$ and $R=0.4$ at $\sqrt{s} = 13 \text{ TeV}$) jets clustered using either calorimeter clusters or PF components as inputs. Due to the high rate of QCD multijet production at the LHC the H_T threshold for these triggers must be kept at a value large enough to avoid using up all the available HLT bandwidth. The H_T threshold was set to 750 GeV for the data taking conditions at a center-of-mass energy of 8 TeV in the year 2012, but was later raised to 800 GeV at 13 TeV in 2015-2016 and will be raised to over 1 TeV in 2017, due to the increased instantaneous luminosity. These high-energy thresholds do not allow events that might contain the decay of heavy resonances with masses below 1-1.2 TeV in all-jets final states to be collected with these trigger paths. A trigger path requiring 4 jets with $p_T > 50 \text{ GeV}$ was also deployed during the data-taking period at 8 TeV. However due to the high bandwidth required, the data-stream containing the trigger was reconstructed only at the end of the 2012 run. While this trigger path could in principle partially overcome the shortcomings of the H_T trigger, it is still not ideal, especially given the much higher luminosity predicted for the future. Figure 4.13 shows a schematic representation of the H_T and 4-jets triggers.

The studies of Chapter 6 show that the groomed jet mass is a very powerful discriminator against the predominant multijet QCD background, thus a solution to this problem is to introduce grooming algorithms on large-cone jets in the high-level-trigger. In this way interesting events can be selected likely containing the all-hadronic decay of a heavy SM particle already at the HLT level. However, jet grooming algorithms are computationally intensive, therefore the trigger path using them must be programmed to use the limited amount of resources available to the HLT. Figure 4.14 compares the structure of a traditional jet-based trigger with the structure of a trigger using jet grooming on a large-cone jet.

The thresholds for the selections of these triggers were optimized with the studies presented in this chapter, trying to balance the trigger efficiency for simulated signal events of singly-produced vector-like quarks and the exclusive rate of events selected by the trigger. The purpose of this section is to assess the performance of these triggers and prove that they outperform standard triggers in searches for heavy resonances in the all jet final state with a mass lower than 1 TeV. Two trigger paths using jet substructure will be examined:

1. “Single jet”: requires one $R=0.8$ jet with a trimmed mass greater than 50 GeV, $H_T > 600 \text{ GeV}$, one b-tagged $R=0.4$ jet with a “loose” operating point.

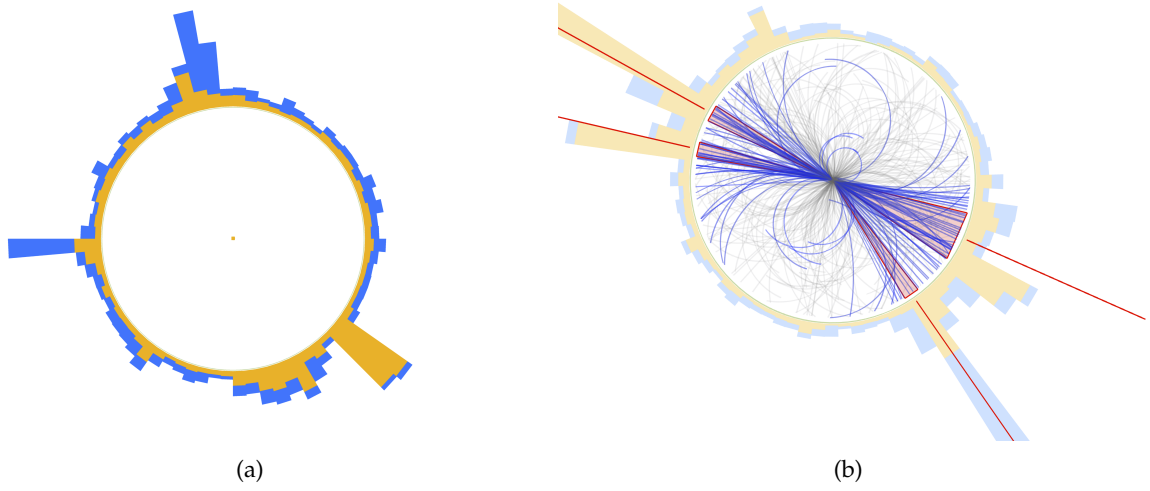


Figure 4.13: Schematic representation of the H_T trigger (a) and the 4-jets trigger (b).

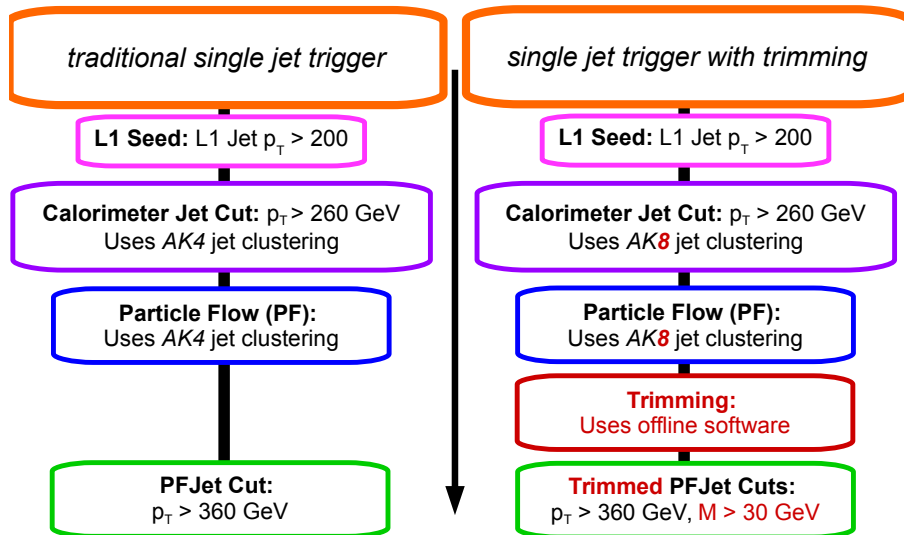


Figure 4.14: Comparison of trigger structure between a classic trigger and a trigger using jet substructure [88].

2. “Dijet”: requires two $R=0.8$ jets, one of them with a trimmed mass greater than 30 GeV, $p_T > 250, 200$ GeV for the two large-cone jets, one b-tagged $R=0.4$ jet with a “loose” operating point.

Figure 4.15 shows a schematic representation of the the jet substructure triggers considered in this section.

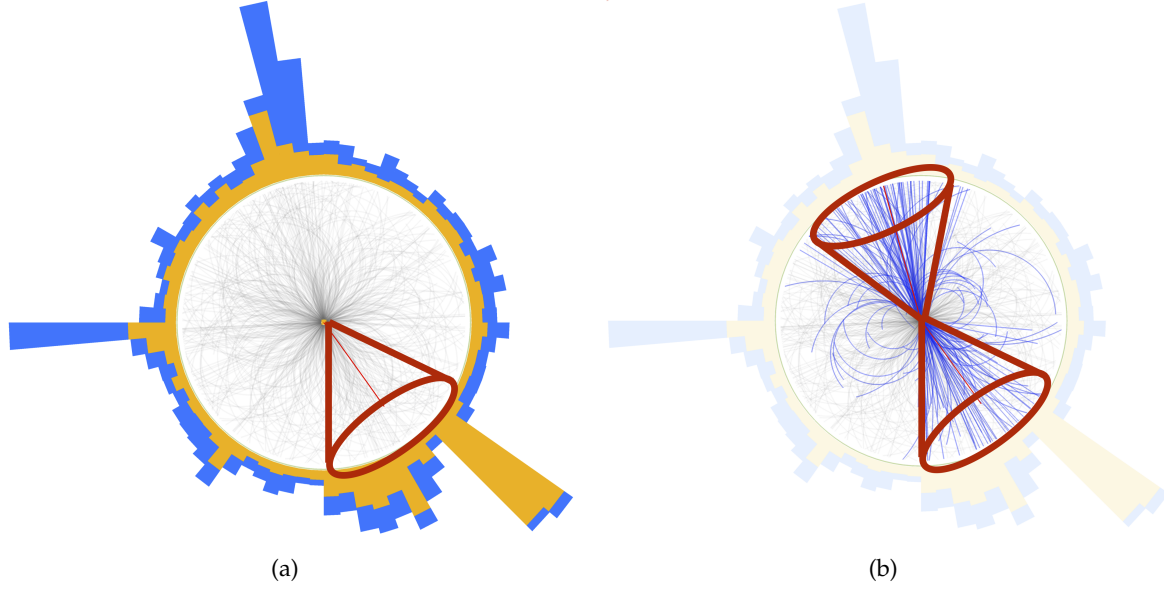


Figure 4.15: Schematic representation of the jet substructure triggers: single jet (a), and dijet (b).

4.5.2 Performance assessment

To assess the performance of the triggers, the case-study of searches for singly-produced vector-like quarks in the all-jets final state is considered. The mechanism of production of single vector-like quarks is shown in Figure 4.16 and includes in the final state, other than the VLQ itself, a forward light jet and a generally low transverse momentum b or top quark. Searches of vector-like quarks performed at the LHC exclude masses up to 800-900 GeV depending on the quark, its final state, and the coupling parameters. An example event display of a singly produced T quark decaying to a top quark and a Higgs boson, where both particles decay to jets is shown in Figure 4.17. The study consists in performing a basic analysis-like selection on simulated samples of signal events, and deriving the trigger efficiency comparing the standard H_T trigger with the triggers using jet substructure.

In particular this study will consider the case of a T vector-like quark decaying to a b quark and a W boson, and the case of a B vector-like quark decaying to a top quark and a W boson. In the former case, the decay topology consists of a b jet and a high- p_T W jet, while in the latter it consists of a boosted W jet and a boosted top jet. Some distributions of relevant kinematic variables are shown for the $T \rightarrow bW$ case for two different mass hypotheses of the T quark: 0.8 TeV and 1.2 TeV. From the distributions it can be seen that for the mass hypothesis of 800 GeV most of the signal events have H_T smaller than 800 GeV, while the situation is reversed for the mass hypothesis of 1200 GeV.

In the following, the mass of the vector-like quarks is assumed to be 700 GeV. Given the topology of the final states taken into account and the distributions of the relevant kinematic

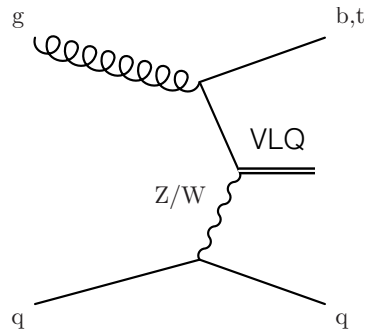


Figure 4.16: Diagram of the dominant production mechanism for a single vector-like quark at the LHC.

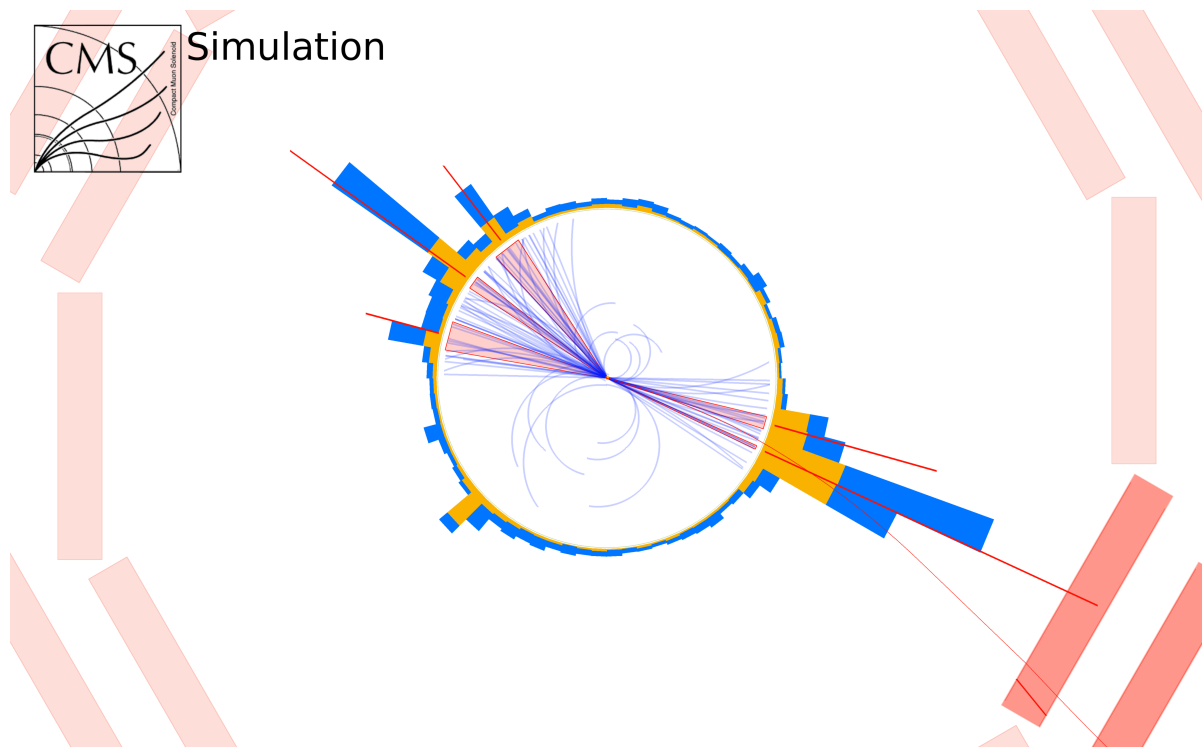


Figure 4.17: Event display of a singly-produced T quark decaying to a top quark and a Higgs boson in the all-jets final state.

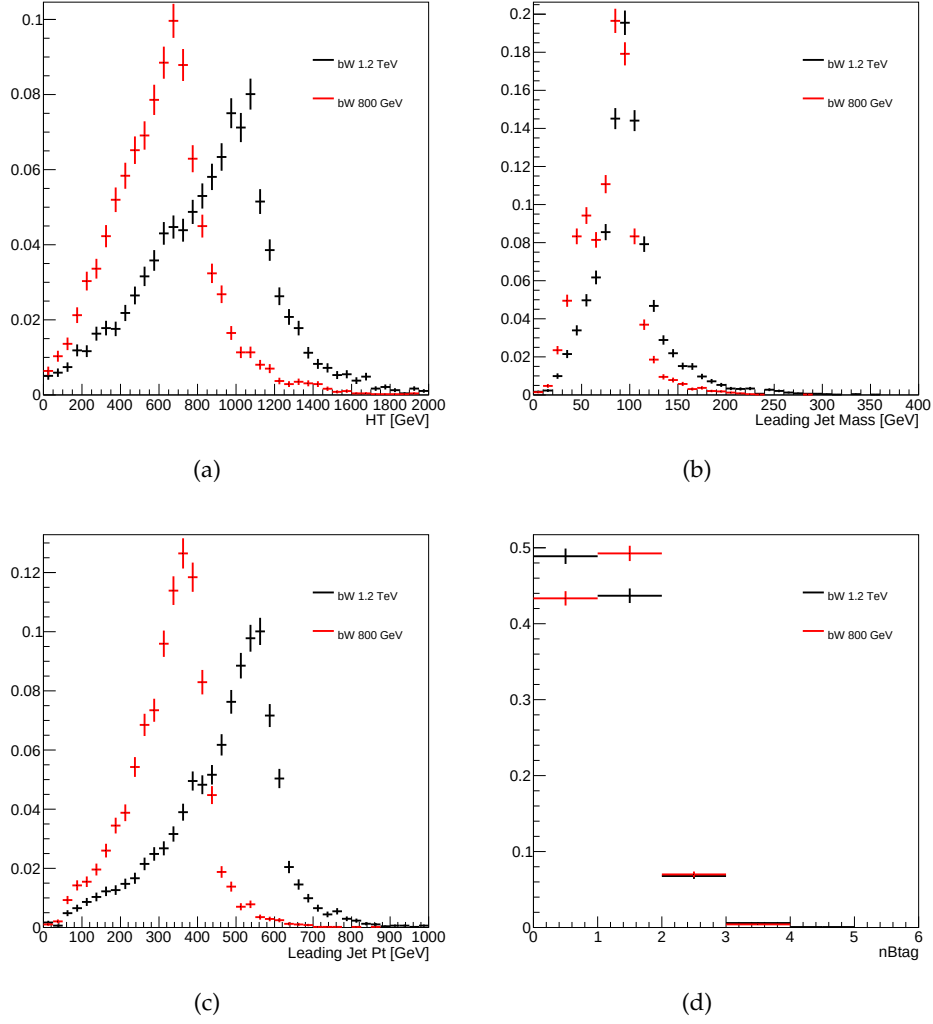


Figure 4.18: Distribution of H_T (a), leading AK8 trimmed jet mass (b), leading AK8 jet p_T (c), and number of b-tagged AK4 jets for a sample singly produced T quarks decaying to bW .

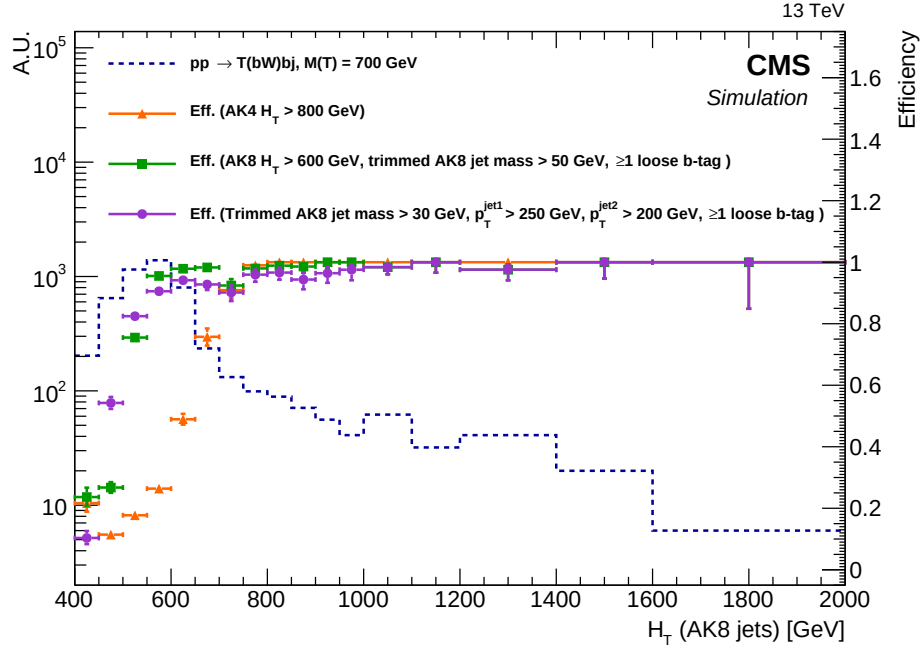
variables, the following selections are applied to select signal events. For the $T \rightarrow bW$ final state:

- At least one anti- k_T $R=0.8$ jet (AK8) with $p_T > 200$ GeV and groomed jet mass $m > 60$ GeV. The trimming algorithm is applied to compute the jet mass.
- At least one standard anti- k_T $R=0.4$ jet with $p_T > 50$ GeV passing the medium-efficiency requirements of the b-tagging algorithm.

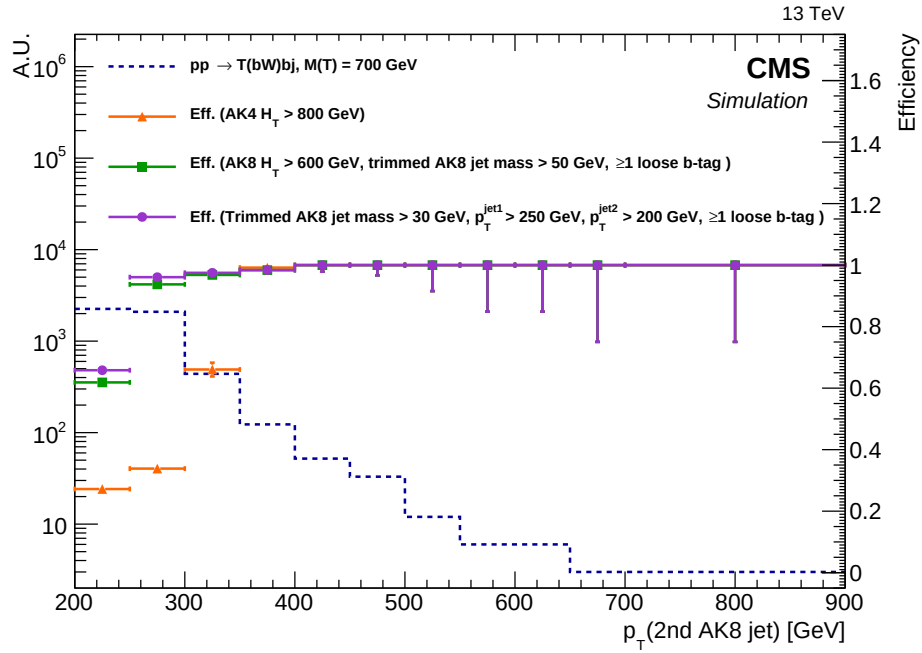
For the $B \rightarrow tW$ final state:

- At least two anti- k_T $R=0.8$ jet with $p_T > 200$ GeV and groomed jet mass bigger than 50 and 100 GeV, respectively. The trimming algorithm is applied to compute the jet mass.
- At least one standard anti- k_T $R=0.4$ jet with $p_T > 50$ GeV passing the medium-efficiency requirements of the b-tagging algorithm.

The trigger efficiency is computed as a ratio where the denominator is the number of signal events passing the selection, and the numerator is the number of signal events passing the selection and the trigger requirements. The efficiency is computed in function of the H_T variable, defined as the sum of the p_T of the AK8 jets in the event, and in function of the p_T of the AK8 jet with the second highest p_T . The trigger efficiency is shown in Figure 4.19 for the $T \rightarrow bW$ final state and in Figure 4.20 for the $B \rightarrow tW$ final state. Signal events are clustered in the region where the standard H_T trigger as a very low efficiency. Most of the signal events would not be collected by this trigger, on the other hand the single-jet and dijet substructure triggers are able to select most of the signal events.



(a)



(b)

Figure 4.19: Trigger efficiency for a sample of singly-produced T quarks decaying to bW for different all-jets triggers in function of H_T (a) and the p_T of the second most energetic AK8 jet (b). The dashed line represents the distribution of the signal, and the solid lines are the efficiencies for the trigger paths considered.

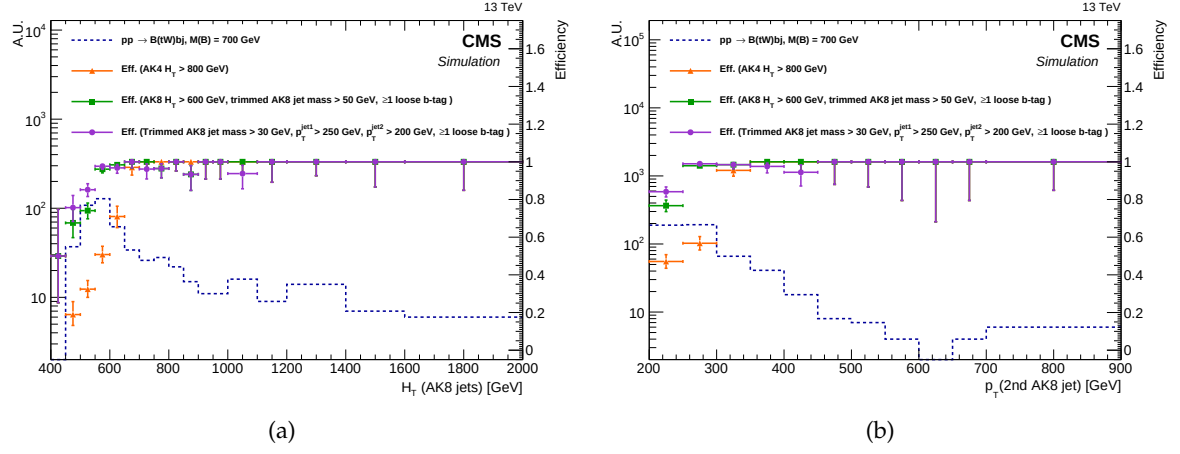


Figure 4.20: Trigger efficiency for a sample of singly-produced B quarks decaying to tW for different all-jets triggers in function of H_T (a) and the p_T of the second most energetic AK8 jet (b). The dashed line represents the distribution of the signal, and the solid lines are the efficiencies for the trigger paths considered.

Subjet b tagging commissioning on 8 TeV data

5.1 Introduction

The purpose of this chapter is to describe the procedure employed to validate the subjet b-tagging technique on 8 TeV data. “Validating” means checking that the kinematic variables used by the b-tagging algorithm are well-behaved in terms of data-simulation agreement and do not show any pathologic behavior. Some discrepancies between data and simulated events are expected due to a non perfect simulation of the distribution of the detector material. These discrepancies result in a different performance of the algorithms in terms of efficiency and misidentification rates in data and simulation. Dedicated data-simulation correction factors (also referred to as scale factors) are derived to take into account these differences. Figure 5.1 shows an illustration of subjet b tagging. The results of these studies are publicly available in Ref. [67]. Further documentation can be found in Ref. [89].

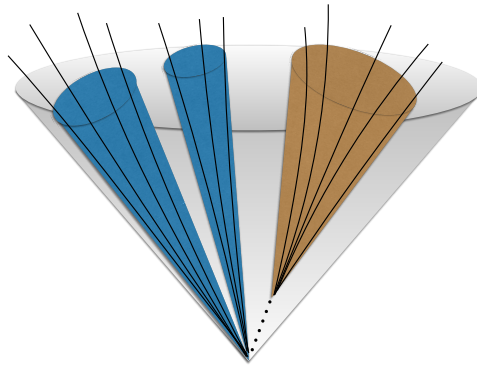


Figure 5.1: An illustration of the concept of subjet b tagging.

5.2 Data and simulation samples

This study uses the CMS framework software release 5.3. The validation studies are based on a semi-leptonic $t\bar{t}$ selection. In particular the study uses only muonic decays, offering a better reconstruction efficiency. The datasets from the single-muon data stream listed in Table A.1

are analyzed, corresponding to an integrated luminosity of 19.7 fb^{-1} collected at a center-of-mass energy of 8 TeV. The single-muon data stream is composed by a selection of high level trigger paths requiring at least a single muon to be present. The simulated samples used and their respective cross sections are listed in Table A.2.

The simulated $t\bar{t}$ production is based on MADGRAPH [90]. The single-top production and the di-boson production is simulated using POWHEG [91] and TAUOLA, while the Drell-Yan production is simulated using MADGRAPH. The W+jets samples are generated using MADGRAPH and the muon-enriched QCD samples are produced using PYTHIA. The cross sections used to normalize the different SM background simulation samples are the most updated available from phenomenology theory papers, according to what is reported in Ref. [92]. In particular, the NNLO inclusive $t\bar{t}$ cross section value of 245.8 pb is used [7]. The respective exclusive cross section for the fully leptonic, fully hadronic and semileptonic decay channels are obtained by multiplying by their relative branching ratio which is known to great precision.

5.3 Event reconstruction and selection

The validation of subjet b-tagging in the boosted top channel is based on the semileptonic $t\bar{t}$ production, following a tag-and-probe approach. This means that a relatively pure sample of boosted b jets in top decays is selected by operating a tight selection of a leptonic top and a top tagged jet on the opposite hemisphere. This selects semileptonic $t\bar{t}$ events with a small amount of other SM processes. The event selection requires exactly one isolated muon and at least one b-tagged AK5 jet in the same hemisphere of the event. These requirements account for the leptonic decay of the top. The hadronic decay of the top is selected requiring a top-tagged CA15 jet using the HEPTOPTAGGER algorithm in the hadronic hemisphere of the event. Figure 5.2 shows a schematic illustration of the semileptonic $t\bar{t}$ selection. The selection of semi-leptonic $t\bar{t}$ events is detailed in the following.

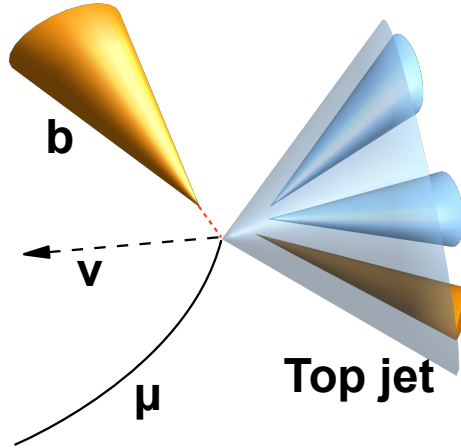


Figure 5.2: An illustration of a boosted semileptonic $t\bar{t}$ event.

Reconstructed muons are required to have $p_T > 45 \text{ GeV}$ and $|\eta| < 2.1$. The identification of the muon closely follows the reference selection from Ref. [93]:

- The muon is reconstructed using both the hits in the muon chambers and the inner tracking system;

- $\chi^2/N_{\text{DOF}} < 10$;
- $N_{\text{TrackerHits}} > 10$;
- $N_{\text{PixelHits}} > 0$;
- $N_{\text{MuonHits}} > 0$;
- $d_0 < 0.02$ cm (distance with respect to the hardest deterministic annealing filter primary vertex);
- $I_{\text{rel}} < 0.12$ (using a cone size of $R = 0.4$);
- Number of matched stations > 0 .

The I_{rel} isolation variable is defined as $(\sum E_T)/p_T$, where p_T is the transverse momentum of the muon, and the sum of the transverse energy E_T is performed on all the particle-flow components within a cone size $R = 0.4$ with respect to the direction of the muon.

Additional muons in the event are vetoed, as well as additional electrons with $p_T > 35$ GeV and $|\eta| < 2.5$. The identification of the electron also follows the reference selection from Ref. [93], omitting the isolation criteria.

At least two AK5 jets with $p_T > 30$ GeV and $|\eta| < 2.5$ are required, with one AK5 jet in the leptonic hemisphere of the event, b-tagged using the CSV algorithm at the medium-efficiency operating point. Exactly one CA15 jet with $p_T > 150$ GeV and $|\eta| < 2.5$ is required in the hadronic hemisphere of the event. This jet has to fulfill the top-tagging criteria of the HEPTOPTAGGER algorithm.

The missing transverse energy in the event has to be greater than 20 GeV and H_T^{lep} is required to be greater than 150 GeV, where $H_T^{\text{lep}} \equiv E_T^{\text{miss}} + p_T^\mu$.

The performance of the single muon trigger in data and simulation has been studied in Ref. [94]. The scale factors provided correct for the different performance of the trigger and of the muon identification between data and simulation [94]. Simulated events are weighted to account for discrepancies in the b-tagging efficiency and misidentification rates in data and simulation [95]. Similarly, a re-weighting of the p_T spectrum of the simulated $t\bar{t}$ signal is applied according to Ref. [96].

After all the selection requirements, a negligible amount of multijet QCD events is selected. The selected sample is largely enriched with $t\bar{t}$ semileptonic events, with major background coming from the fully leptonic $t\bar{t}$ production. Overall a good agreement between data and simulation is observed.

5.4 Commissioning studies

The b-tagging algorithms in the 8 TeV data-taking era have been developed and trained using regular anti- k_T jets with a cone size parameter of $\Delta R = 0.5$. The CSV algorithm considered for the studies here presented is based on the identification of tracks displaced from the primary proton-proton collision and on the reconstruction of secondary vertices inside jets. A general introduction to b-jet identification in CMS can be found in Sec. 3.4. When applying b-tagging to subjects, it is necessary to confirm that the properties of the tracks and of the secondary vertices associated to the subjects are described properly by the simulation, or at the same level of agreement observed on regular jets. A comparison between data and simulated events for all the relevant b-tagging observables for subjects is shown below. The plots have been produced for the semi-leptonic $t\bar{t}$ sample selected according to the description given in

Sec. 5.3. For each selected event, a CA15 jet produced by the HEPTOPTAGGER algorithm is identified, which is attributed to the hadronic decay of a top quark. All the three subjets of this jet are considered for the plots shown below.

Two categories of variables are considered by the b-tagging algorithm: track variables and secondary vertex variables. Track variables describe the properties of displaced tracks from the primary vertex, while secondary vertex variables describe the properties of the secondary vertex found by the vertex finder algorithm.

Track Variables

The tracks are associated to subjets using a matching based on the ΔR distance between the tracks and the axis of the subjets. This distance is required to be lower than 0.3. The tracks associated to subjets are further selected, in order to reject tracks from pile-up, according to the following criteria:

- number of tracker hits (including pixel) ≥ 8 ;
- number of pixel hits ≥ 2 ;
- track transverse impact parameter $|\text{IP}_{2D}| < 0.2$ cm, defined as the distance between the primary vertex and the the closest point of approach of the track with respect to the primary vertex, projected on the plane perpendicular to the beam line;
- track transverse momentum $p_T > 1$ GeV;
- track normalized $\chi^2 < 5$, computed from the fit to the hits in the tracker that form the track;
- track longitudinal impact parameter $|\text{IP}_z| < 17$ cm, defined as the distance between the primary vertex and the the closest point of approach of the track with respect to the primary vertex, projected on the beam line;
- track distance to jet axis < 0.07 cm, defined as the distance between the respective points of closest approach between the track and the subjet axis;
- track decay length < 5 cm, defined as the distance between the primary vertex and the closest point of approach on the track between the track and the subjet axis.

The previous list also describes the track variables used by the CSV algorithm. In Figs. 5.3-5.5 the relevant track-related b-tagging observables are shown. The subjet-to-flavor association is performed by selecting the flavor of the closest parton at generator level to the subjet. The distance is measured using the ΔR variable. When no parton is found within a radius of 0.5 of the subjet, or the flavor of the parton is undefined. The agreement between data and simulation is found consistent with what is observed for regular jets.

Secondary Vertex Variables

The secondary vertex variables used by the CSV algorithm are:

- multiplicity of secondary vertices: the number of secondary vertices found by the vertex finder algorithm,
- number of tracks associated to the secondary vertex,

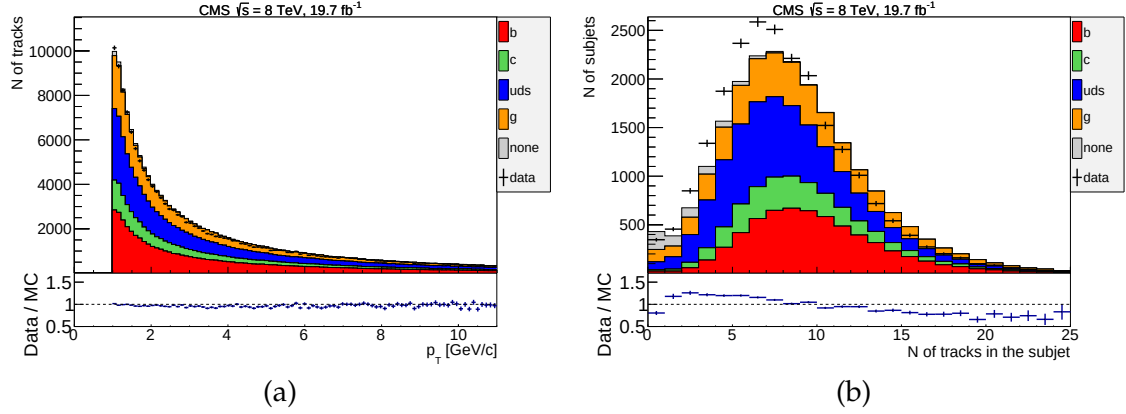


Figure 5.3: (a) Track p_T , (b) track multiplicity for all the subjects of HEPTOPTAGGER jets in semileptonic $t\bar{t}$ events. The contributions from subjects with different flavors are stacked: b quarks in red (the main target of these studies), c quarks in orange, light flavor quarks in blue, undefined flavor in gray. The light flavor contributions includes gluons as well. The simulated signal events and the selected backgrounds are also stacked. The plot at the bottom shows the ratio between the number of data and simulated events. Only statistical errors from data and simulated samples are included in the computation of the error bands.

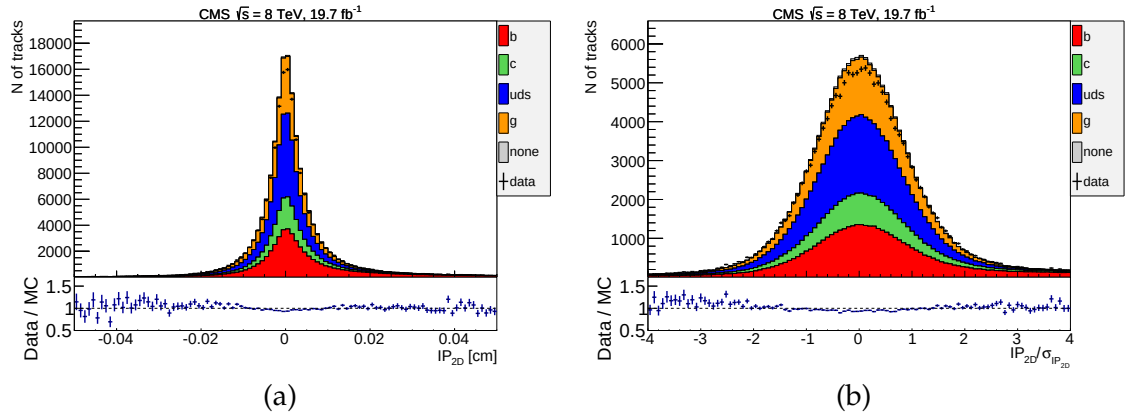


Figure 5.4: (a) Track transverse impact parameter, (b) track transverse impact parameter significance for all the subjects of HEPTOPTAGGER jets in semileptonic $t\bar{t}$ events. The contributions from subjects with different flavors are stacked: b quarks in red (the main target of these studies), c quarks in orange, light flavor quarks in blue, and undefined flavor in gray. The light flavor contributions includes gluons as well. The simulated signal events and the selected backgrounds are also stacked. The plot at the bottom shows the ratio between the number of data and simulated events. Only statistical errors from data and simulated samples are included in the computation of the error bands.

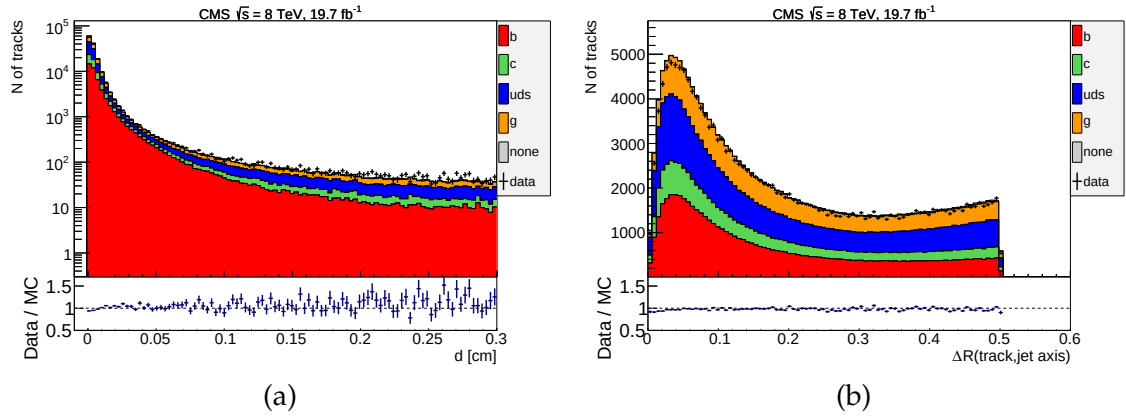


Figure 5.5: (a) Track distance from the subjet axis, (b) track ΔR distance from the subjet axis for all the subjets of HEPTopTAgGER jets in semileptonic $t\bar{t}$ events. The contributions from subjets with different flavors are stacked: b quarks in red (the main target of these studies), c quarks in orange, light flavor quarks in blue, and undefined flavor in gray. The light flavor contributions includes gluons as well. The simulated signal events and the selected backgrounds are also stacked. The plot at the bottom shows the ratio between the number of data and simulated events. Only statistical errors from data and simulated samples are included in the computation of the error bands.

- distance in ΔR between the subjet axis and the secondary vertex,
- distance between the primary and the secondary vertex,
- significance of the distance between the primary and the secondary vertex, that is: the ratio between the distance and its uncertainty,
- projection of the distance between primary and secondary vertex on the plane perpendicular to the beamline (transverse distance),
- significance of the transverse distance between between primary and secondary vertex,
- invariant mass of the secondary vertex,
- fraction of the energy of the subjet belonging to the secondary vertex,
- χ^2 of the secondary vertex fit.

In Figs. 5.6-5.8, the relevant b-tagging observables related to secondary vertices are shown. The prediction from simulated events is separated in the contributions from different quarks as for the track variables. The agreement between data and Monte Carlo is found consistent with what is observed for AK5 regular jets.

5.5 Conclusions

In Figs. 5.9, the Value of the CSV discriminator for all the subjets of HEPTopTAgGER jets is shown. All the variables contributing to the computation of the CSV discriminator for subjets have been checked for discrepancies between data and simulation. The level of agreement is comparable to the one observed for standard isolated AK5 jets. Residual discrepancies are corrected by dedicated data-simulation correction factors. The CSV b tagging algorithm

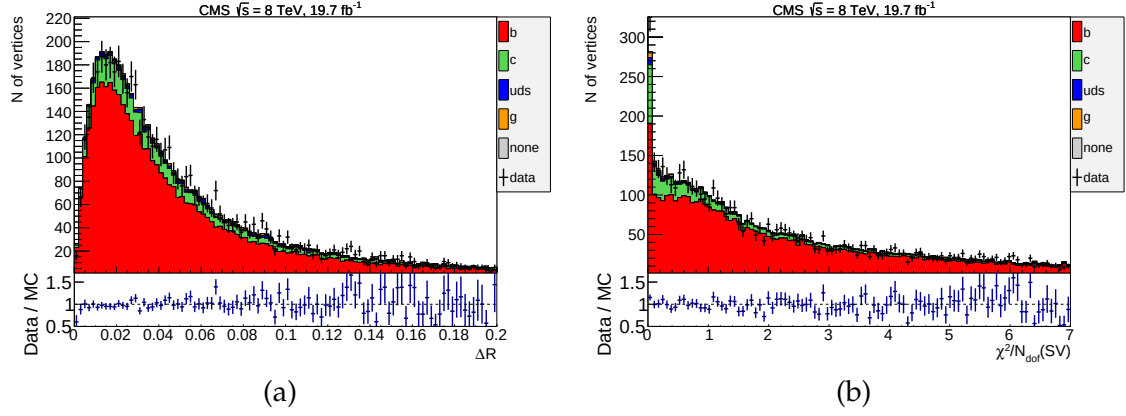


Figure 5.6: (a) Multiplicity of secondary vertices, (b) number of tracks at the secondary vertex, (c) ΔR distance between the secondary vertex the subjet axis, (d) χ^2 of the secondary vertex fit for all the subjets of HEPTopTAGGER jets in semileptonic $t\bar{t}$ events. The contributions from subjets with different flavors are stacked: b quarks in red (the main target of these studies), c quarks in orange, light flavor quarks in blue, and undefined flavor in gray. The light flavor contributions includes gluons as well. The simulated signal events and the selected backgrounds are also stacked. The plot at the bottom shows the ratio between the number of data and simulated events. Only statistical errors from data and simulated samples are included in the computation of the error bands.

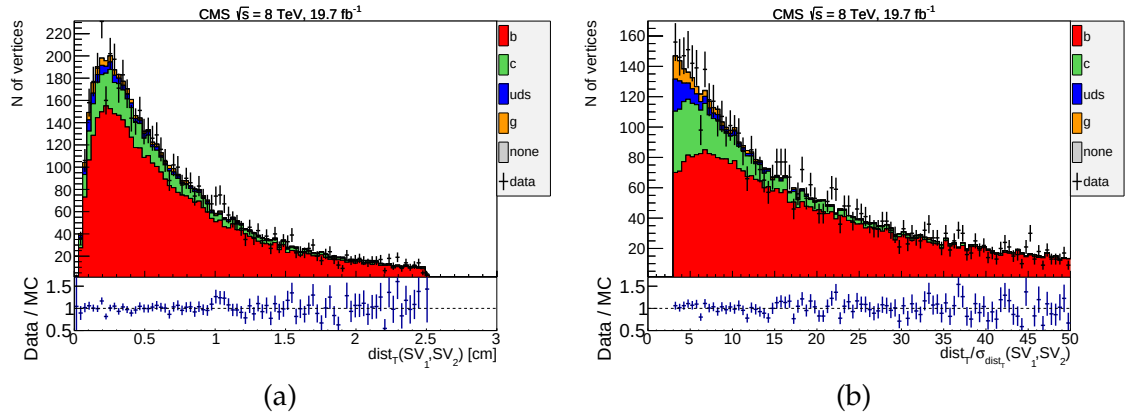


Figure 5.7: (a) transverse distance between primary and secondary vertex and (b) its significance for all the subjets of HEPTopTAGGER jets in semileptonic $t\bar{t}$ events. The contributions from subjets with different flavors are stacked: b quarks in red (the main target of these studies), c quarks in orange, light flavor quarks in blue, and undefined flavor in gray. The light flavor contributions includes gluons as well. The simulated signal events and the selected backgrounds are also stacked. The plot at the bottom shows the ratio between the number of data and simulated events. Only statistical errors from data and simulated samples are included in the computation of the error bands.

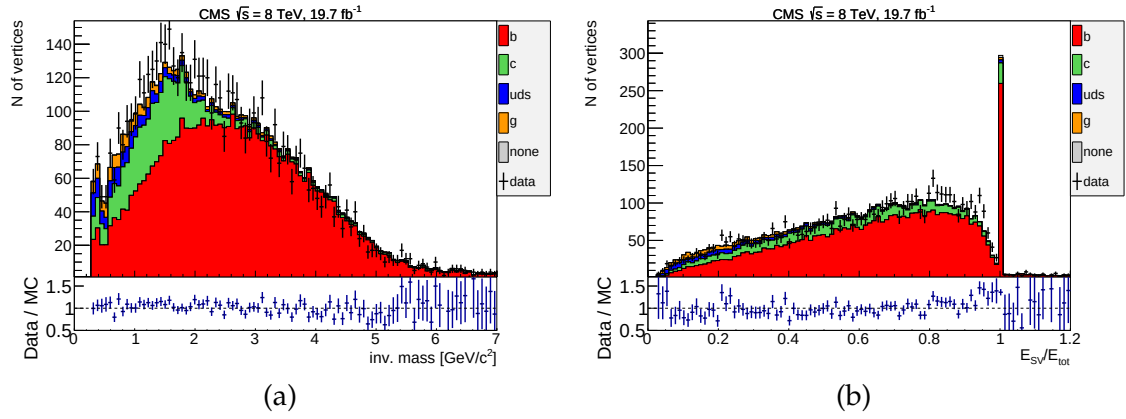


Figure 5.8: (a) Mass of the secondary vertex, (b) fraction of the subjet energy associated to the secondary vertex for all the subjets of HEPtopTAgGER jets in semileptonic $t\bar{t}$ events. The contributions from subjets with different flavors are stacked: b quarks in red (the main target of these studies), c quarks in orange, light flavor quarks in blue, and undefined flavor in gray. The light flavor contributions includes gluons as well. The simulated signal events and the selected backgrounds are also stacked. The plot at the bottom shows the ratio between the number of data and simulated events. Only statistical errors from data and simulated samples are included in the computation of the error bands.

behaves well on subjets of large-cone jets, this tool can therefore be safely used in physics analyses.

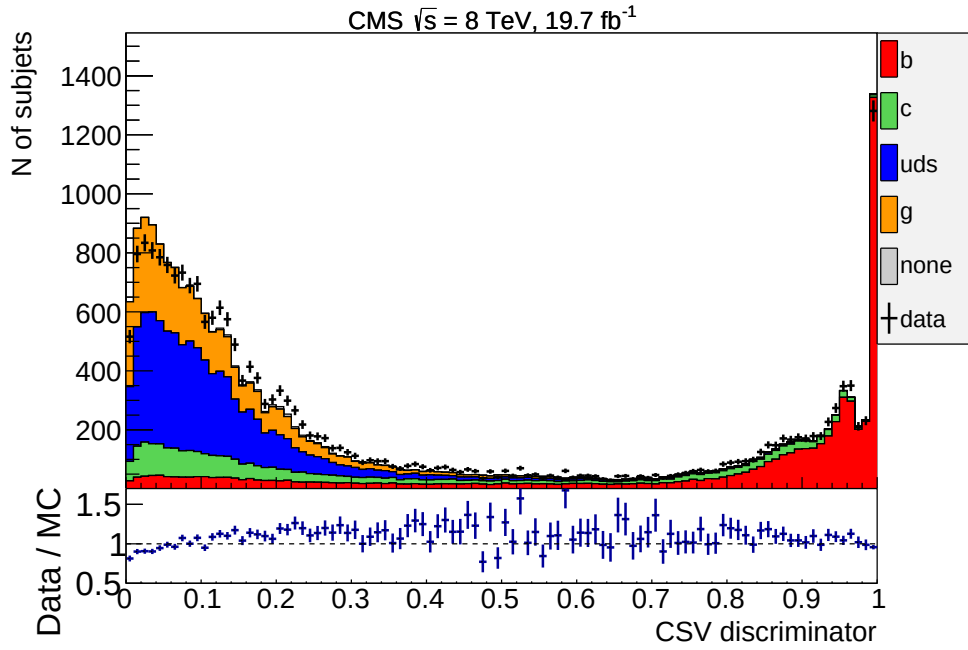


Figure 5.9: Value of the CSV discriminator for all the subjects of HEPTopTAgger jets in semileptonic $t\bar{t}$ events. The contributions from subjects with different flavors are stacked: b quarks in red (the main target of these studies), c quarks in orange, light flavor quarks in blue, and undefined flavor in gray. The light flavor contributions includes gluons as well. The simulated signal events and the selected backgrounds are also stacked. The plot at the bottom shows the ratio between the number of data and simulated events. Only statistical errors from data and simulated samples are included in the computation of the error bands.

Commissioning of the HEP_{TOP}TAGGER with 8 TeV data

The purpose of this chapter is to describe the procedure employed to commission the HEP_{TOP}TAGGER (HTT) algorithm on 8 TeV data. The performance of the tagger is optimized in simulated events in combination with other jet substructure tools such as subjet b tagging. The distribution of relevant kinematic variables for the tagger is compared in data and simulated events in a sample of events enriched in semileptonic $t\bar{t}$ decays. Finally, a tag-and-probe technique is used to measure efficiency corrections factors for simulated events. The results of these studies have been made public in Ref. [81], while further documentation can be found in Ref. [97].

6.1 Algorithm performance in simulated events

The first step for the commissioning of the tagger is the assessment of the performance of the algorithm in simulated events, additionally the observables used by top tagging algorithms to identify merged boosted top jets are studied. The simulated samples used for these studies are reported in Table A.3. The inclusive SM $t\bar{t}$ sample, which represents the “signal” for these studies, is generated with the software package POWHEG [91]. Hadron showering is treated with POWHEG too, additionally tau leptons are treated with TAUOLA. For each event in the simulated sample, only the two most energetic large-cone jets, matched to an hadronically-decaying top quark are considered. The matching between the jet and the generator level top quark is done by requiring $\Delta R(\text{jet}, \text{top}) < 0.8$. The second sample contains QCD multijet events and represents the “background” events the tagger algorithm attempts to reject. Generation and showering for this sample are performed with the PYTHIA package, version 6.

6.1.1 Discriminating variables

The most relevant variable of the tagging algorithm is the HEP_{TOP}TAGGER jet mass (m_{123}), which is shown in QCD and $t\bar{t}$ samples for jets with $p_T > 200 \text{ GeV}$ in Figure 6.1a. The $t\bar{t}$ sample peaks at the top jet mass, as expected, but has a large low mass tail due to unmerged tops. This tail becomes smaller when raising the p_T threshold, due to the increasing probability that the all-hadronic decay of a top quark is merged in a single jet. The distribution for the QCD multijet sample exhibits a peak at low mass, due to the hadronization of light flavor and gluon jets, but presents a secondary peak at the top quark mass. The top mass peak is due to the HEP_{TOP}TAGGER algorithm procedure in which the combination of mass

drop subjets with filtered mass closest to the top mass is chosen and all other combinations are discarded. The HEPTopTAgger top jet mass distribution (m_{123}) after the W mass selection is shown in Figure 6.1b.

The possibility of a combination of the tagger with the N-subjettiness variable is also explored. In this case the N-subjettiness is calculated using the constituents of the Cambridge-Aachen jet with a radius parameter $R=1.5$. The N-subjettiness jet shape ratio τ_3/τ_2 is shown for QCD multijet and $t\bar{t}$ samples in Figure 6.1c before the HEPTopTAgger is applied and in Figure 6.1d after the HEPTopTAgger is applied.

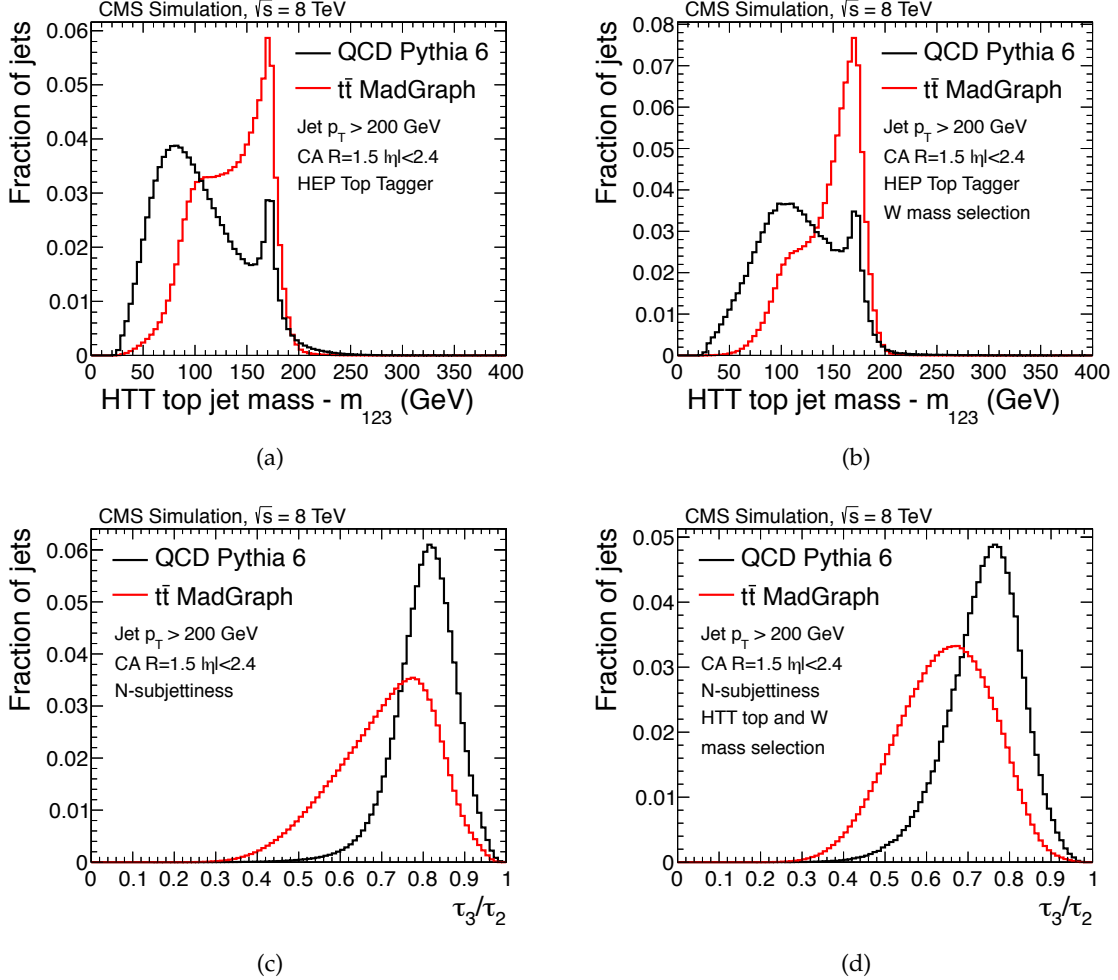


Figure 6.1: Top tagging variables for Cambridge-Aachen jets with $R=1.5$ after the grooming treatment of the HEPTopTAgger with $p_T > 200$ GeV from simulated $t\bar{t}$ and QCD multijet samples: (a) HEPTopTAgger jet mass (m_{123}), (b) HEPTopTAgger jet mass (m_{123}) after W mass tagging selections, (c) τ_3/τ_2 , (d) τ_3/τ_2 after the HEPTopTAgger top and W mass selections.

6.1.2 Algorithm performance in simulation

To optimize the algorithm, the parameters of the tagger are varied iteratively. For each configuration, the efficiency and misidentification rate (mistag) are measured. Given that different parameters of the tagger are varied together, it is possible that several configuration yield the

same efficiency. In such a case, only the best configuration is kept for a given signal efficiency, that is the configuration with the smallest misidentification rate. The configurations then are compared by plotting the minimum mistag rate versus the signal efficiency. A continuous line is drawn connecting all the configurations, this is called the performance curve. Working points (WP) are defined along the performance curve as example selections which provide the minimum misidentification rate for a given signal efficiency.

To compare the performance of different tagger algorithms it is important to have a common definition for the efficiency and the misidentification rate. Generally speaking, these quantities can be expressed as a ratio. The denominator in the efficiency calculation is defined as the number of jets matched to a generator-level top or anti-top quark that decays hadronically and passes the p_T selection. Similarly, the misidentification rate denominator is defined as the number of jets matched to a generator-level quark or gluon from the hard scatter passing the p_T selection. The numerator for both the efficiency and mistag rate is defined as the number of jets from the respective denominators passing the top tagging selection. The definition of the tagging rate denominators, based on the matched generator-level particle, is used in order to compare algorithms using different jet collections, such as the HEPTOPTAGGER and the CMS Top Tagger. The mistag rate is dependent on the jet flavor and the results presented here are applicable only to event topologies with a quark/gluon mixture similar to the selection defined in Section 6.1.2.

Using these definitions, the minimum misidentification rate for a given signal efficiency is shown in Figures 6.2 and 6.3 for various p_T ranges of the matched generator-level parton. The curves are determined by fixing the m_{123} selection ($140 \text{ GeV} < m_{123} < 250 \text{ GeV}$) and varying the width of the W mass selection (f_W). The curves using the N-subjettiness algorithm in addition are determined by varying also the τ_3/τ_2 selection. The curves using subjet b-tagging in addition are determined by varying also the selection on the maximum subjet CSV discriminant. For the curves where more discriminators are combined, all of the selections mentioned above are simultaneously varied and the minimum mistag rate for a given signal efficiency is determined. The combination of the HEPTOPTAGGER algorithm with N-subjettiness and subjet b tagging yields a considerable improvement in the performance of the tagger.

The performance of the HEPTOPTAGGER is compared to the CMS Top Tagger algorithm in Fig. 6.4 for various p_T ranges of the matched generator-level parton. The optimal top tagging algorithm depends on the p_T of the matched parton and on the chosen efficiency. In the lowest p_T range considered (jets matched to partons with $p_T > 200 \text{ GeV}$), only the HTT and HTT Combined Tagger (combining the HTT, subjet b-tagging, and N-subjettiness) are studied. Here the best performance is provided by the HTT Combined Tagger. The CMS Top Tagger and CMS Combined Tagger (CMS Top Tagger combined with subjet b-tagging and N-subjettiness) are not considered in this p_T range because the top decay products are rarely merged within $R=0.8$ jets. For jets matched to partons with $p_T > 400 \text{ GeV}$, the majority of boosted top quarks are fully merged into a $R=1.5$ jet, while a smaller fraction are reconstructed within $R=0.8$ jets. Therefore the HTT performs very well for high efficiency selections in this p_T range, especially when combined with N-subjettiness and subjet b-tagging. The CMS Combined Tagger results in the best performance for low efficiency selections in this range (Fig. 6.4a). For jets matched to partons with $p_T > 600 \text{ GeV}$ or $p_T > 800 \text{ GeV}$, the CMS Combined Tagger has superior performance. The performance loss for the HEPTOPTAGGER is due to the $R=1.5$ jet collecting the decay of other particles not belonging to the decay of the top quark.

Example selections with close to optimal performance for a given tagging efficiency are defined as working points (WP). These WP are defined in Table 6.1 for the HTT alone and

Working point	m_{123} selection	f_W selection	subject b-tag WP	τ_3/τ_2 selection
HTT WP0	140-250 (GeV)	0.495	none	none
HTT Combined WP1	140-250 (GeV)	0.495	CSV-loose	none
HTT Combined WP2	140-250 (GeV)	0.15	CSV-medium	none
HTT Combined WP3	140-250 (GeV)	0.15	CSV-medium	< 0.63

Table 6.1: Working points for the HTT and HTT combined tagger (HTT + N-subjettiness + subject b-tag)

the combination with other variables. One working point for the HTT alone, two working points for the HTT combined with subject b tagging, and one working point combining HTT, subject b tagging and N-subjettiness are provided. For the subject b tagging, standard operating points are used [67]. This simplifies the correction of data-simulation discrepancies possibly introduced by the usage of subject b tagging.

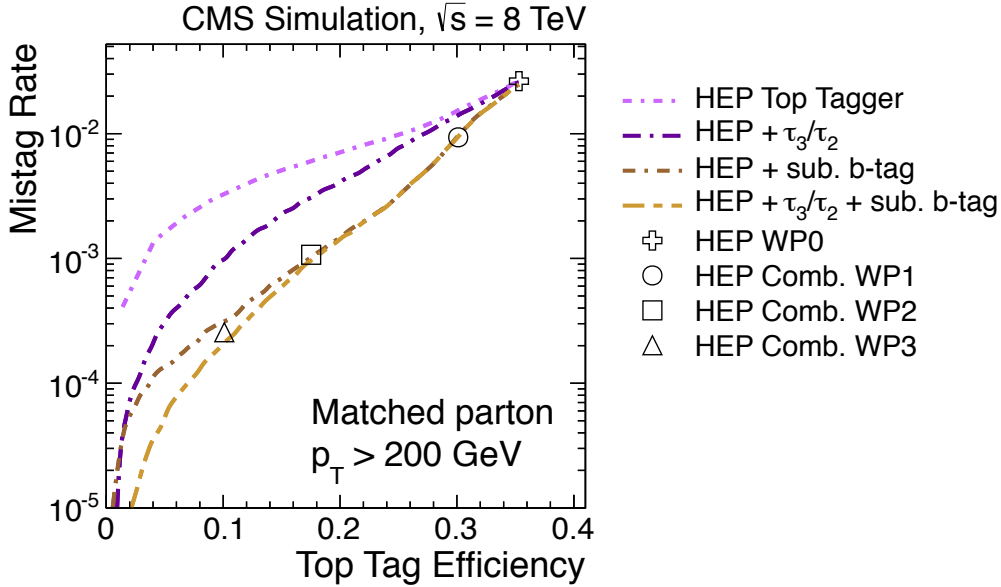


Figure 6.2: Mistag rate in function of the top-jet tagging efficiency for the HEPTOPTAGGER and HTT combined tagger for jets matched to generated partons with $p_T > 200$ GeV. The mistag rate is estimated from QCD multijet simulated events while the efficiency is obtained with $t\bar{t}$ simulated events. A mass cut of $140 < m_{\text{jet}} < 250$ GeV is required. Signal jets are matched to simulated all-hadronic generated top quarks, while background jets are matched to simulated partons from the hard scatter.

The efficiency and mistag rates of the HEPTOPTAGGER are studied in function of the multiplicity of reconstructed primary vertices (pileup). The performance of the ideal tagging algorithm has a small dependence on pileup, however this is difficult to achieve in the high-luminosity conditions of the LHC. The influence of pileup on top tagging efficiency and mistag rate is demonstrated in Figure 6.5 for the HEPTOPTAGGER alone and in Figure 6.6 in combination with subject b tagging. Numerical values for the slopes are given for two working

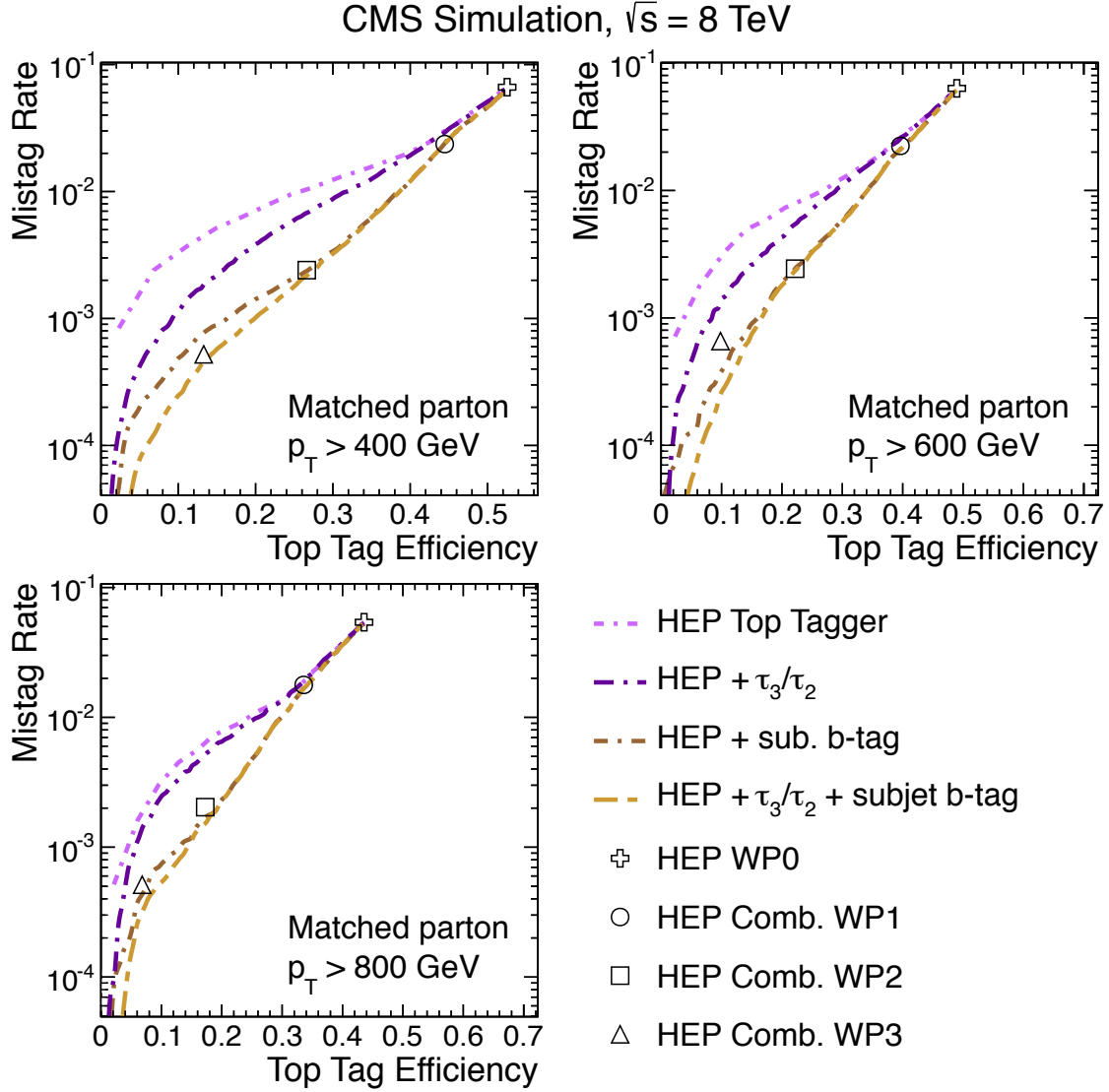


Figure 6.3: Selected curves comparing different combinations using the HEPTopTAgGER. Mistag rate in function of the top-jet tagging efficiency as obtained from QCD multijet and $t\bar{t}$ simulated samples, respectively. Signal jets are matched to simulated all-hadronic generated top quarks, while background jets are matched to simulated partons from the hard scatter. Distributions are shown for three p_T selections, where the p_T cut is applied to the matched generated parton.

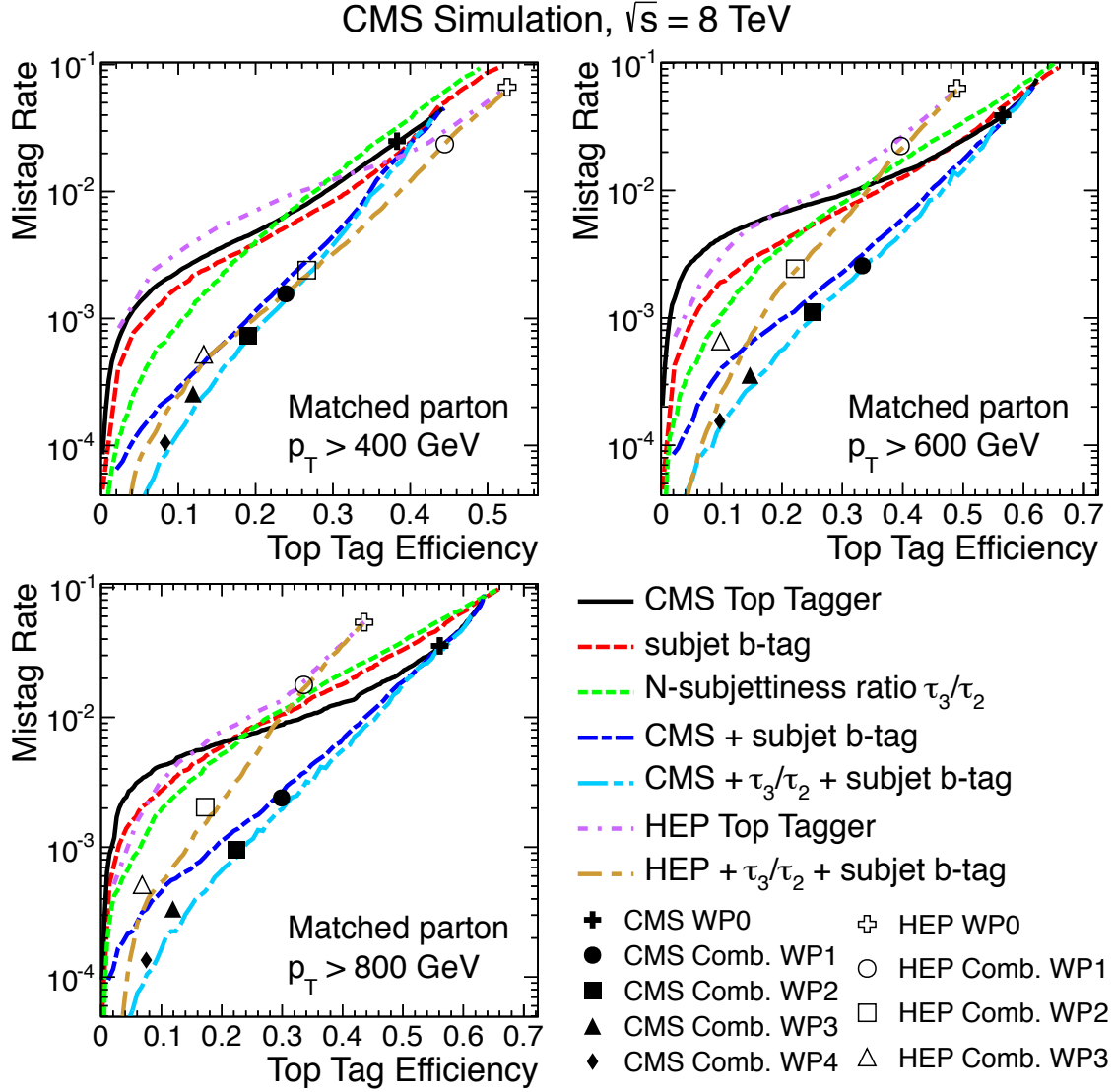


Figure 6.4: Mistag rate in function of the top-jet tagging efficiency as obtained from QCD multijet and $t\bar{t}$ simulated samples, respectively, comparing HEPTOPTAGGER and CMS Top Tagger alone or combined with subjet b tagging and N-subjettiness. Signal jets are matched to simulated all-hadronic generated top quarks, while background jets are matched to simulated partons from the hard scatter. Distributions are shown for three p_T selections, where the p_T cut is applied to the matched generated parton.

points in Table 6.2. It can be observed that the performance of the HEPTopTAgger alone has a moderate dependence on pileup, while when combined with additional jet substructure techniques (especially subjet b tagging) the effect is mitigated.

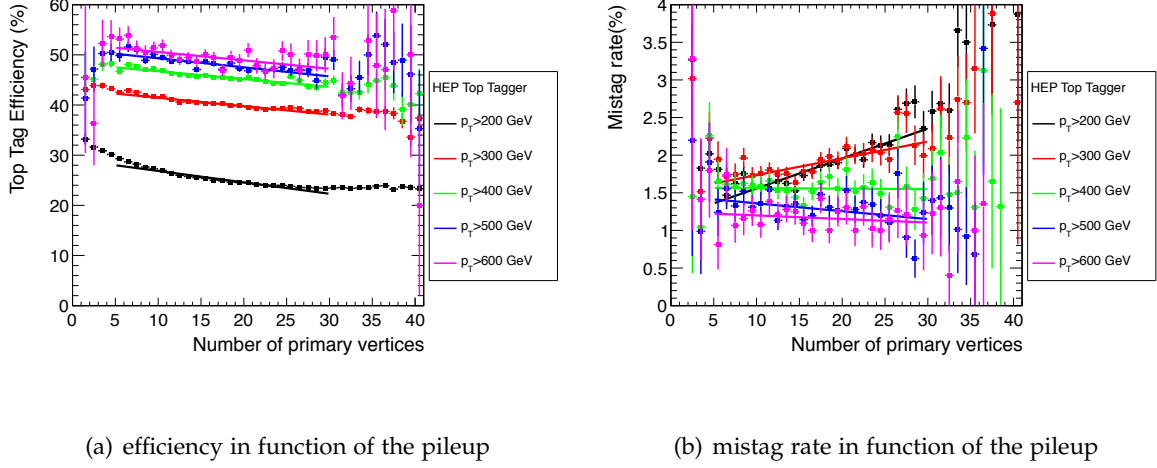


Figure 6.5: Efficiency and mistag rate as a function of pileup for the HEPTopTAgger (WP0).

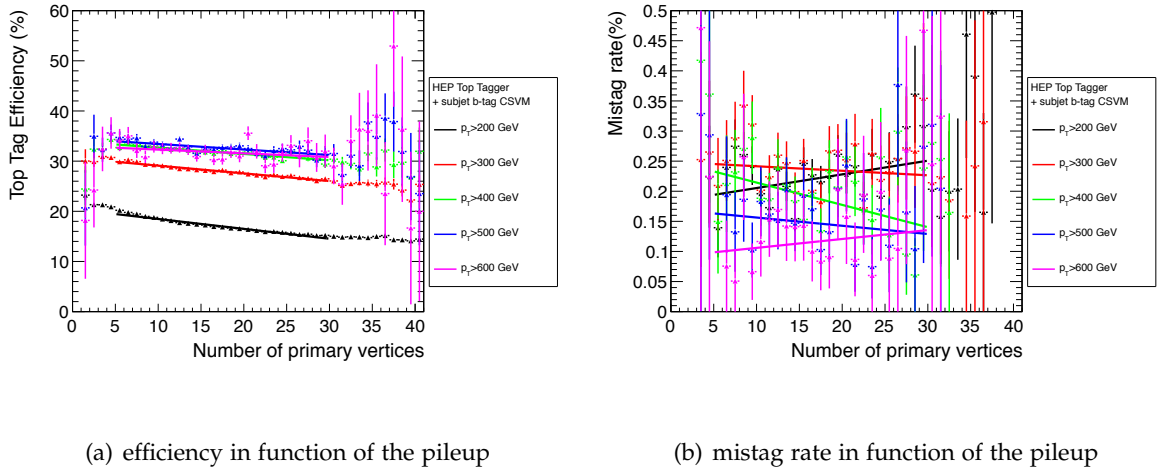


Figure 6.6: Efficiency and mistag rate as a function of pileup for the HEPTopTAgger in combination with a subjet b-tag (WP2).

6.2 HEPTopTAgger performance in data

6.2.1 Introduction

The purpose of this section is to study how the tagger behaves in data, this is achieved using a semileptonic $t\bar{t}$ selection. The same selection is applied to simulated events and is used to calculate data-simulation correction factors.

Tagging selection	Efficiency (%) vs N_{vtx} slope	Mistag rate (%) N_{vtx} slope
HTT combined tagger WP2	-0.180 ± 0.028	-0.010 ± 0.006
HTT combined tagger WP3	-0.463 ± 0.0236	-0.001 ± 0.002

Table 6.2: Slope of the tagging efficiency (%) in function of N_{vtx} , for different operating points and $p_T^{\text{jet}} > 500 \text{ GeV}$.

The measurement of the scale factor is performed in bins of p_T and η of the top-jet, and for different generator setups of the $t\bar{t}$ simulated events. To select top candidates, the leptonic part of the semileptonic $t\bar{t}$ event is used to tag the event, while the hadronic part is used to perform the measurement.

6.2.2 Samples

The dataset used is the same as described in Section 5.2: the single μ data-stream is used, corresponding to an integrated luminosity of 19.7 fb^{-1} of collisions at a center-of-mass energy of 8 TeV. The trigger path used is designed to collect events with one muon, passing minimum p_T cut of 40 GeV, $|\eta| < 2.1$, loose identification requirement, and isolation requirement.

The processed datasets and integrated luminosities are listed in Table A.1, while the simulated samples used are listed Table A.2 and described in Section 5.2. The event selection described in Section 6.2.3 includes a cut on the lepton p_T such that the triggers are nearly fully efficient. After the selection criteria described in Section 6.2.3, the amount of QCD multijet events selected is negligible, thus the QCD multijet simulated samples are not used further in the analysis.

6.2.3 Selection

A sample of boosted merged top-jets as pure as possible needs to be obtained. Semileptonic $t\bar{t}$ events are selected using a “tag and probe” approach; that means that a tight selection on the leptonic side is used to “tag” the event and “probe” the hadronic part where the measurement is performed. Figure 6.7 shows the display of a candidate boosted semileptonic $t\bar{t}$ event.

Exactly one isolated muon with $p_T > 45 \text{ GeV}$ and $|\eta| < 2.1$ and exactly zero electrons are required. The “tight” muon identification and isolation criteria are used, as described already in Section 5.3. The isolated muon divides the event in two hemispheres:

- the hadronic hemisphere: $|\phi - \phi_\mu| > \frac{2}{3}\pi$
- the leptonic hemisphere: $|\phi - \phi_\mu| < \frac{2}{3}\pi$

This division in two hemispheres is made to better associate the decay products of the leptonic top quark to the muon and to make sure that the selected top jet in the event is not coming from the decay products of the leptonic top quark. To further refine the leptonic top quark selection a missing transverse momentum of at least 20 GeV is required in the event. Given $E_W \equiv E_T^{\text{miss}} + p_T^\mu$, $E_W > 150 \text{ GeV}$ is required. These two requirements are made to select events containing a W boson that decayed leptonically. At least one b-tagged jet in the leptonic hemisphere with $p_T > 30 \text{ GeV}$ and $|\eta| < 2.5$ is required. This cut selects the b-jet produced in association with the W in the leptonic decay of the top quark. The combined secondary vertex algorithm is used at a medium-efficiency operating point. On the hadronic

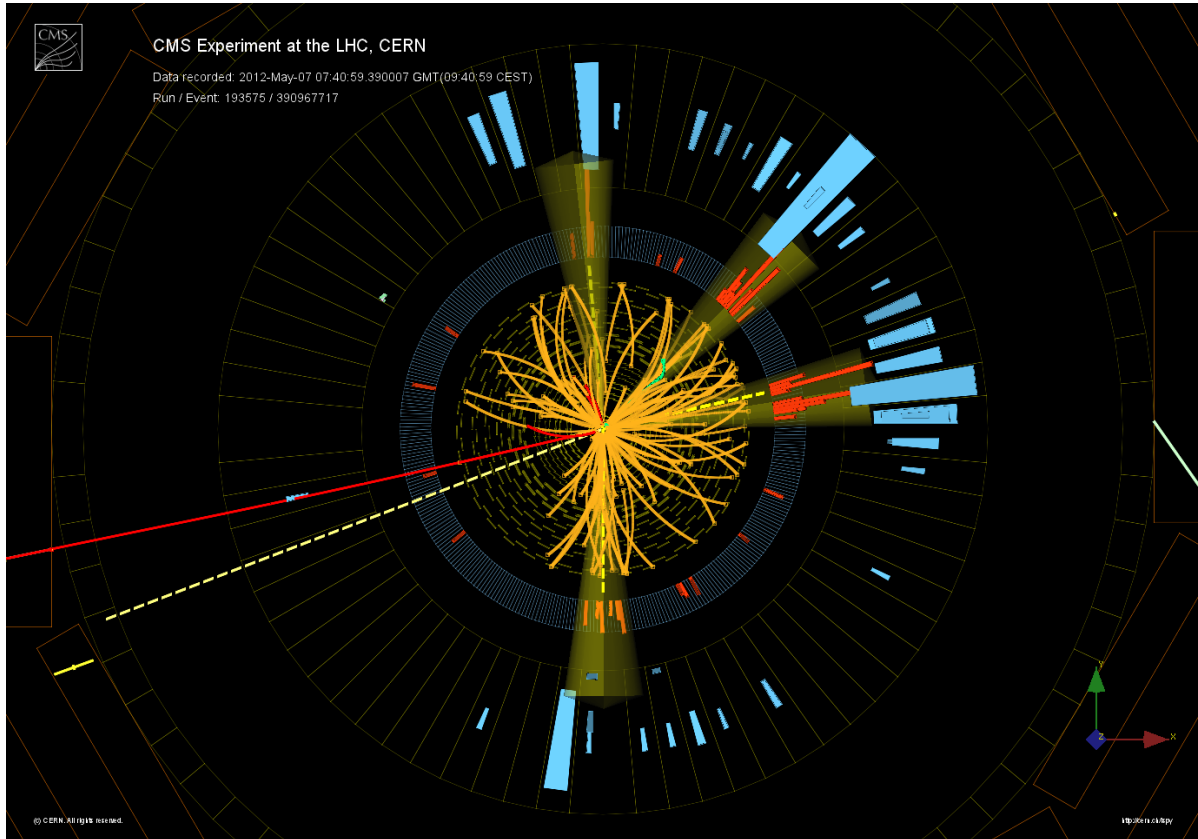


Figure 6.7: Display of a candidate boosted semileptonic $t\bar{t}$ event. Red towers are ECAL clusters, blue towers are HCAL towers, the tower size is proportional to the energy of the cluster. Yellow solid lines are particle-flow charged particles, the red line is the muon, and the yellow dashed line is the direction of the MET.

hemisphere, the event must contain exactly one CA15 jet with $p_T > 200 \text{ GeV}$, $|\eta| < 2.4$, and at least one b-tagged subjet (CSV algorithm, medium-efficiency operating point). This additional cut is necessary to reject events containing top decays that are not fully merged. The HEPTOPTAGGER algorithm is applied to the CA15 jet on the hadronic hemisphere.

Data-simulation correction factors for the muon identification, isolation, and trigger algorithms, are applied according to Ref. [94]. Data and simulation have a different distribution for the multiplicity of reconstructed primary vertices (pileup), this is corrected using a dedicated reweighting procedure as in Ref. [98]. In Fig 6.8 the number of primary vertices distribution can be seen for data and simulation. B-tagging efficiencies and misidentification data-simulation correction factors are applied to regular AK5 jets and subjets as in Ref. [95].

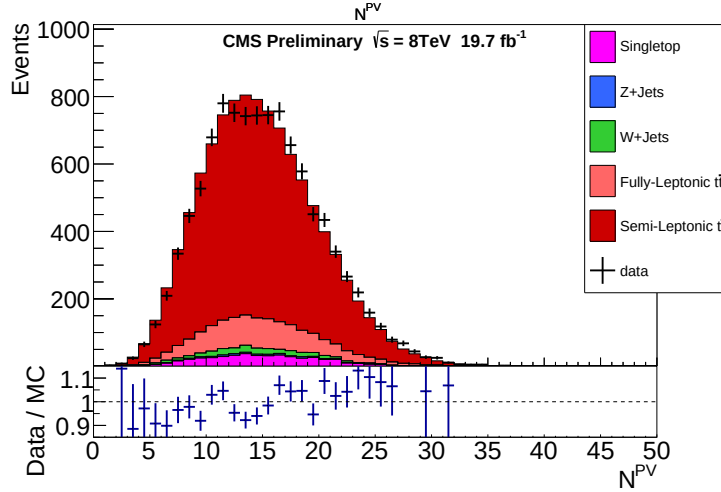


Figure 6.8: Distribution of the number of primary vertices for data and simulated events.

In Fig 6.9 the p_T distribution of the top jet candidate is shown, while Fig 6.10 shows the mass distribution (m_{123}) of the top candidates before applying the HEPTOPTAGGER W mass selection, for different p_T ranges. The mass distribution shows a low mass tail due to unmerged low- p_T top-jets, that gradually disappears when the p_T threshold of the top jet is raised. Further discussion of this is given in 6.2.4.

As described in Section 4.3, after the clustering step and finding the three subjets, the HEPTOPTAGGER performs a series of selections on two variables that are functions of the combined invariant mass of paired subjets (W mass selection). The plot in Fig 6.11 shows the bidimensional distribution of these two variables in data (a), $t\bar{t}$ simulated events (b) and background simulated events surviving the selection described in this section. The black lines drawn on the plot represents the three regions selected by the HTT. If an event lies in at least one of these three regions then the event passes the HTT W mass requirements. The distribution for $t\bar{t}$ signal events shows a peak outside of the three regions selected by the HEPTOPTAGGER, this is an effect on low- p_T top quarks not fully merged within a CA15 jet. The effect is more thoroughly studied in section 6.2.4. The distribution for the simulated background of the analysis (boson plus jets, diboson, single-top, $t\bar{t}$ all-hadronic, and $t\bar{t}$ leptonic production) is relatively flat and most of the background lies outside of the regions selected by the HEPTOPTAGGER requirements, as it is expected.

Figure 6.12 shows the distribution of the top jet mass, for different top jet p_T ranges, after the W mass selection of the HTT, but before the selection on the jet mass itself. After the W mass selection, the long low energy tail for $p_T > 200 \text{ GeV}$ jets disappears, showing that the HTT is able to reject unmerged top jets. However, the data-simulation comparison shows

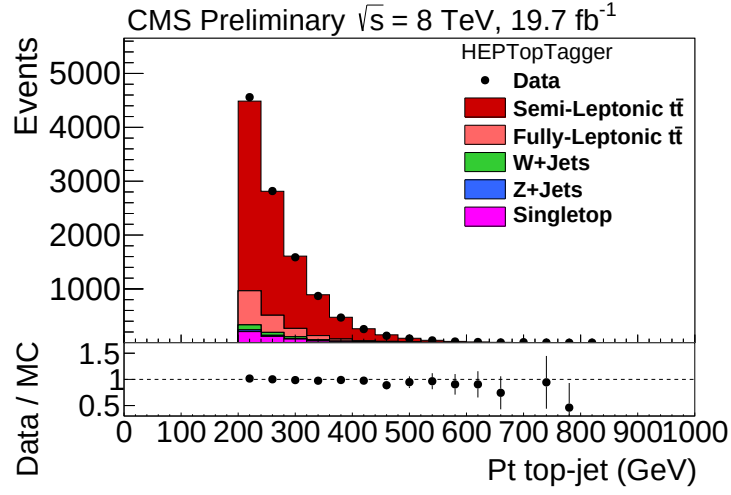


Figure 6.9: Distribution of the p_T of the top jet candidate after all the selection criteria, but before applying the HTT requirements.

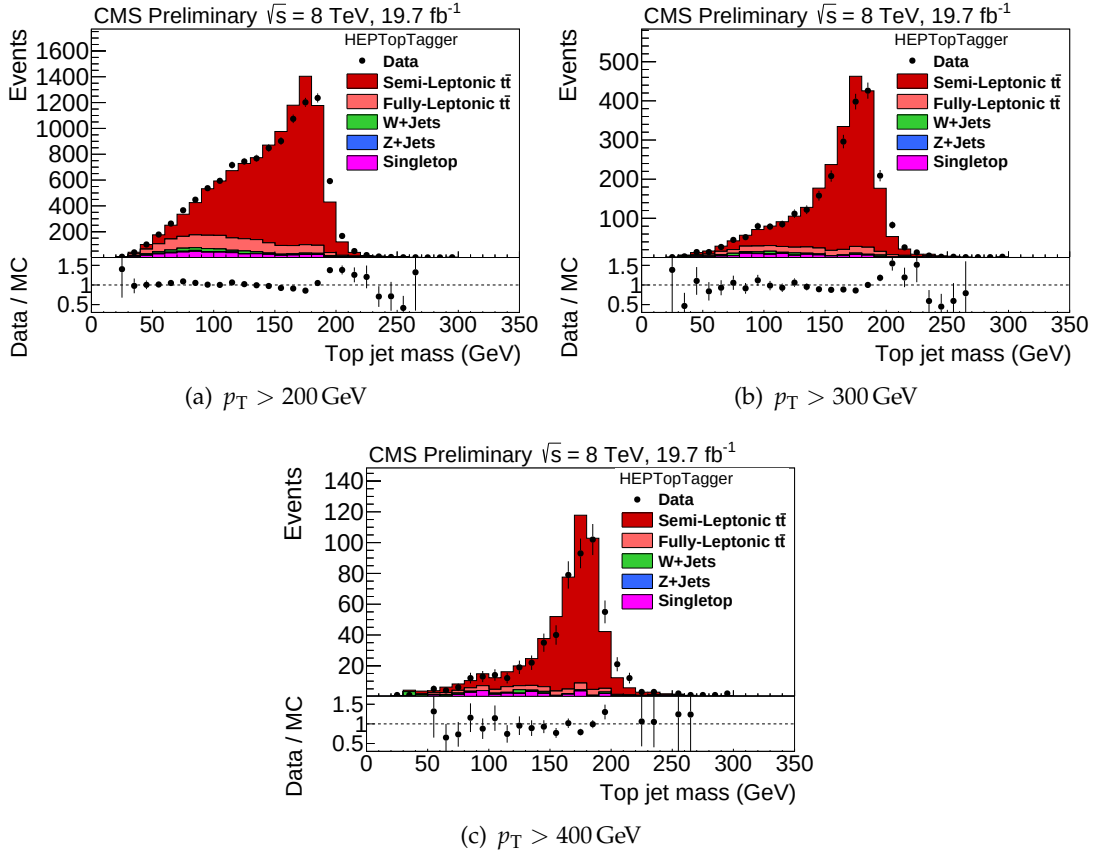


Figure 6.10: HTT jet mass m_{123} distribution with different p_T selections applied to the jet: (a) $p_T > 200$ GeV, (b) $p_T > 300$ GeV, (c) $p_T > 400$ GeV.

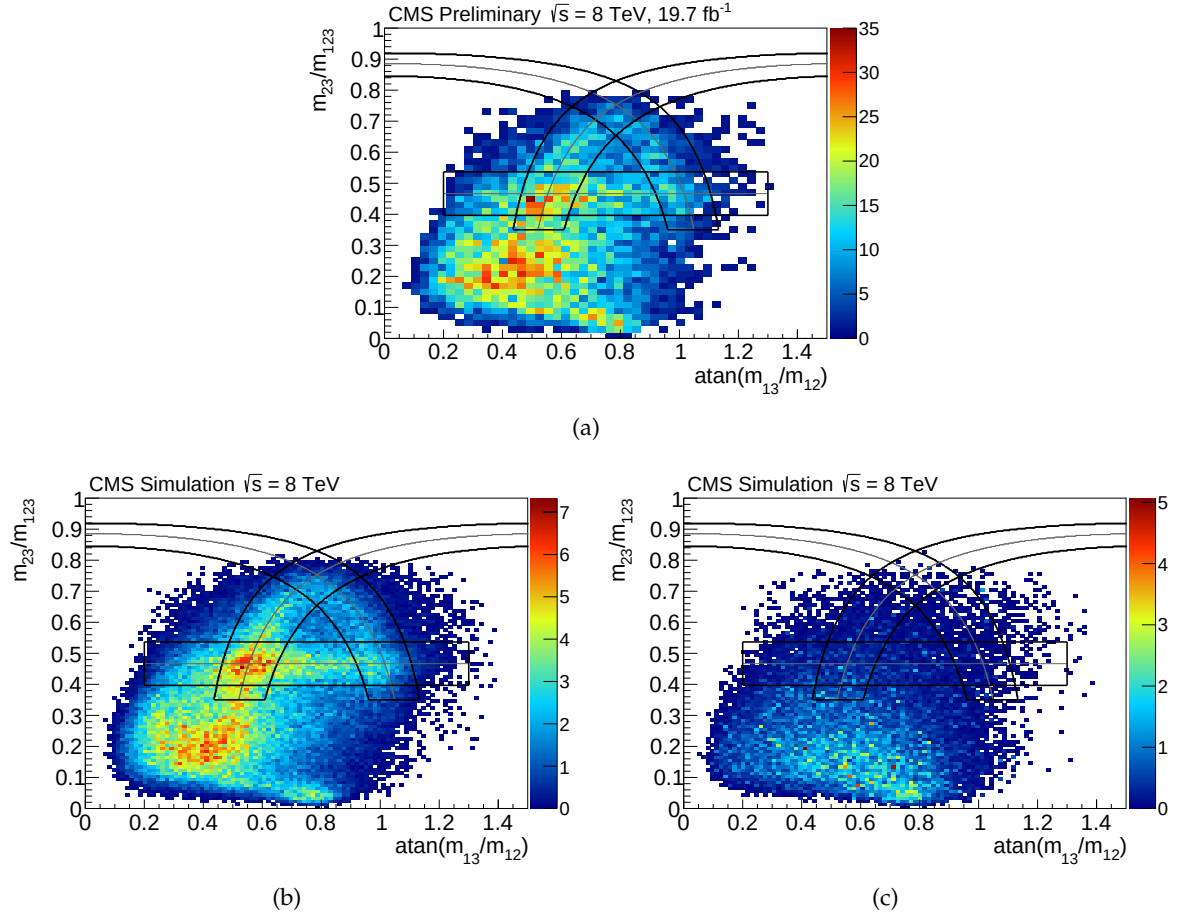


Figure 6.11: Bidimensional distribution of the variables the HEPTopTAgGER uses to select top quark candidates in data (a), semileptonic $t\bar{t}$ simulated events (b), and background events (c). The black lines delimit the regions of the phase space selected by the HEPTopTAgGER.

some disagreement that can be interpreted as a discrepancy between data and simulation in the jet energy resolution [99] or to a non-perfect simulation of the initial-state radiation. This disagreement is corrected by the correction factors that will be derived in this chapter.

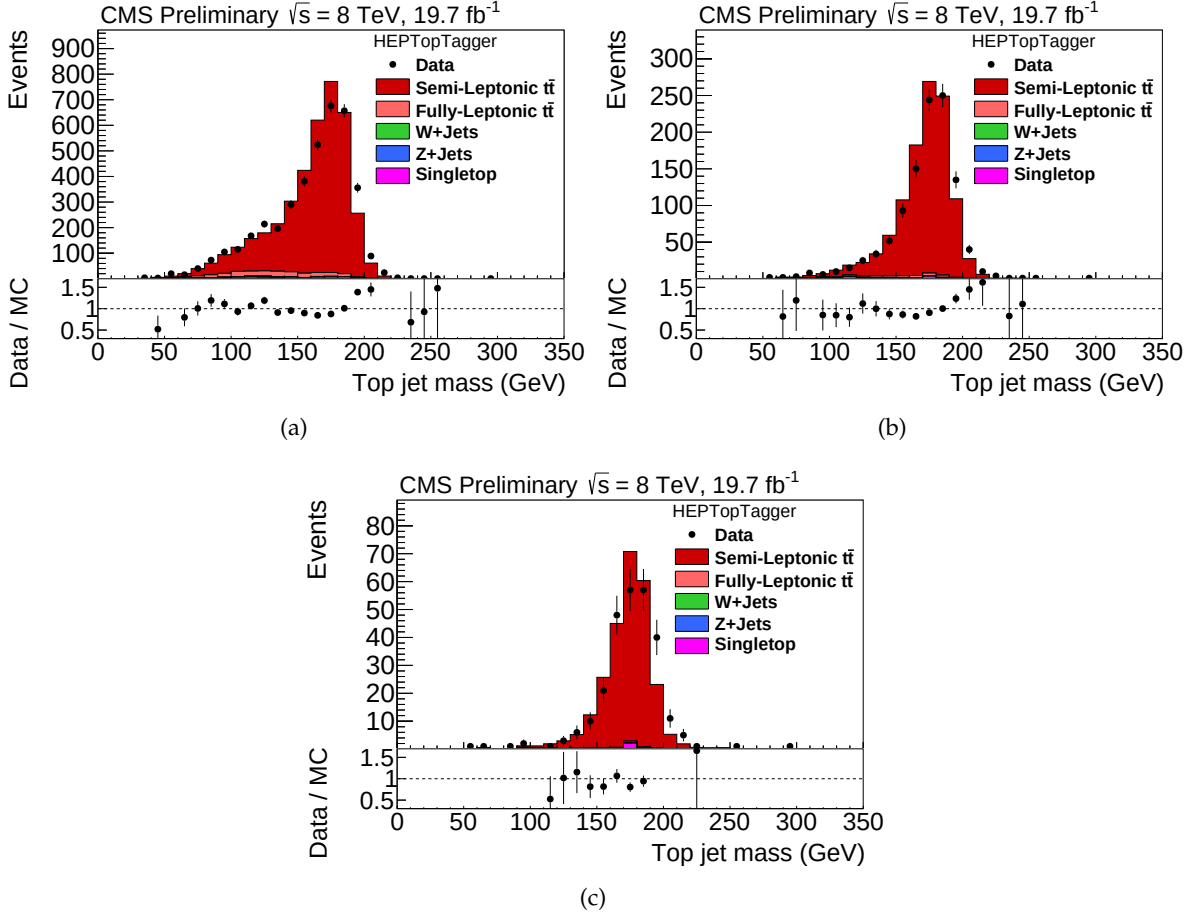


Figure 6.12: HEPTopTagger jet mass distribution (m_{123}) after the W mass selection but before the top mass selection. The distribution is shown with three jet p_T selections: (a) $p_T > 200$ GeV , (b) $p_T > 300$ GeV , (c) $p_T > 400$ GeV .

Figures 6.13 and 6.14 show the variables the HEPTopTagger uses to perform the W mass cuts in different regions of pseudorapidity. The level of agreement between data and simulation varies between the different η regions, thus the correction (or scale) factors will be derived for different pseudorapidity ranges.

6.2.4 Top-jet clustering at low p_T

As shown in the previous section, when the top-jet p_T decreases, the fraction of top decays that are fully merged within CA15 jets decreases. This results in a low mass tail in the HTT mass distribution and in some signal events clustering outside of the W mass selection regions.

Figure 6.15 shows the two dimensional distribution of the HTT variables used for the W mass selection, with a raised threshold for the p_T of the top jet. As expected, the cluster of events outside the regions selected by the HTT, due to non-merged top quark decays, gradually diminishes when raising the threshold. However, a smaller cluster of events at the

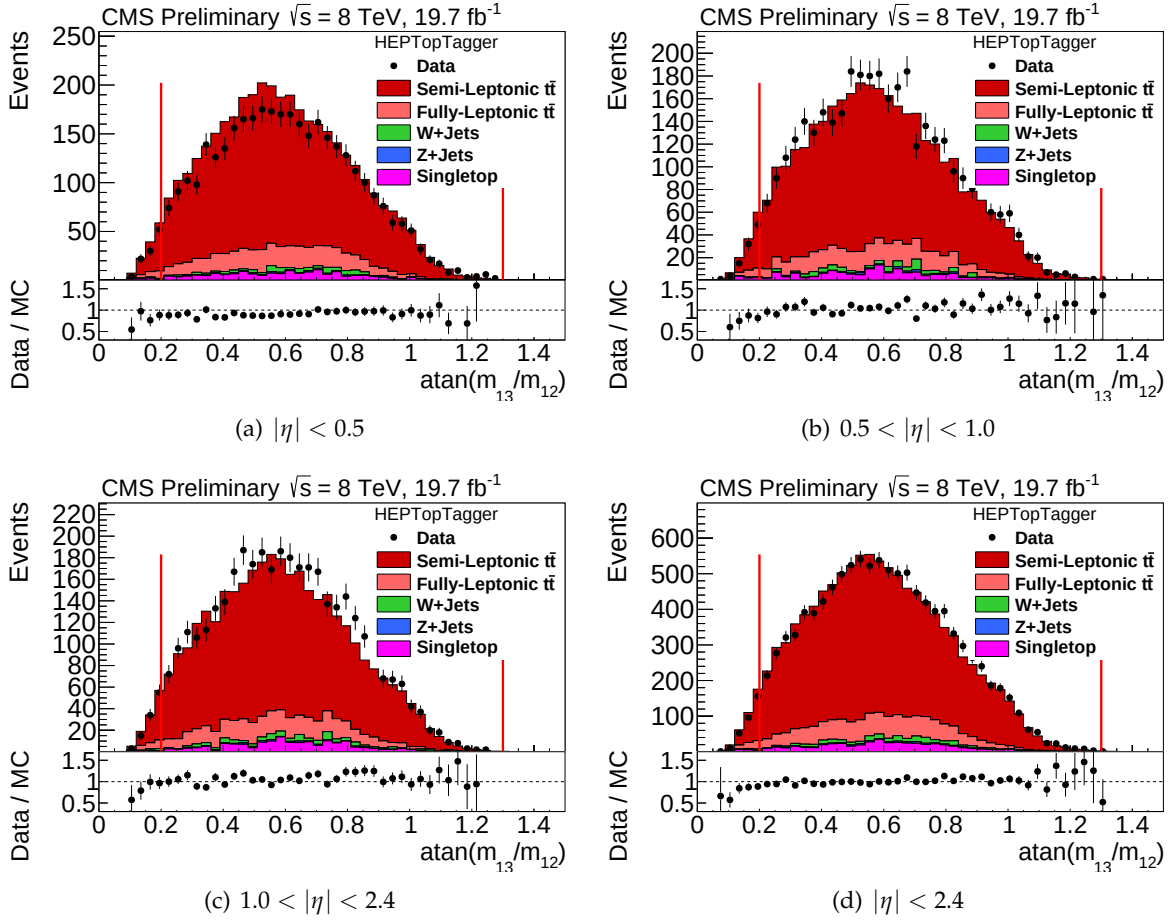


Figure 6.13: HEPTopTagger selection variable $\text{atan} \frac{m_{13}}{m_{12}}$ in various pseudorapidity ranges.

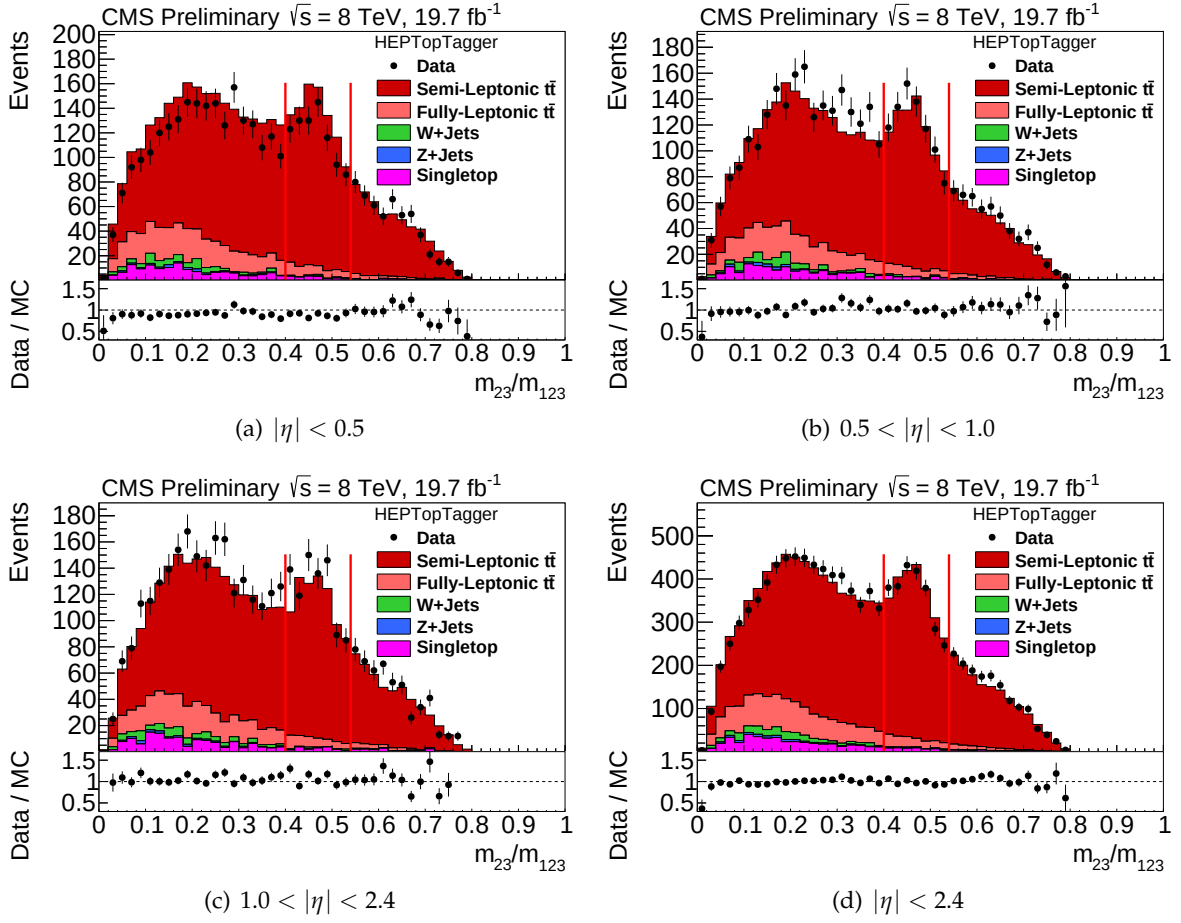


Figure 6.14: HEPTopTAgger selection variable $\frac{m_{23}}{m_{123}}$ in various pseudorapidity ranges.

bottom of the plot persists even at high p_T .

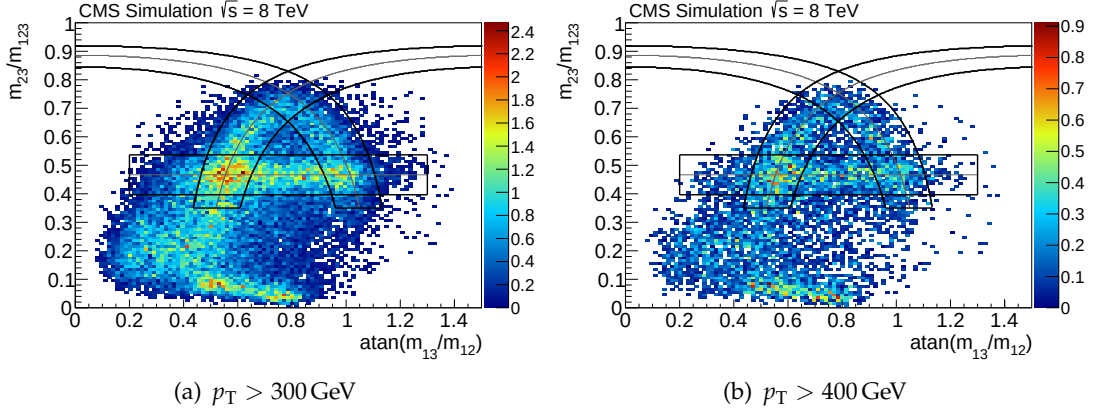


Figure 6.15: Bidimensional distribution of the variables the HTT uses for the W mass selection for (a) $p_T > 300$ GeV and (b) $p_T > 400$ GeV in $t\bar{t}$ events.

To further investigate this cluster of events outside the “A”-shaped region at high- p_T , the top jet is matched to its decay products at generator level. The generator-level matching is performed by requiring that the following particles are contained in a $R=1.5$ cone around the top jet axis:

- top quark;
- W boson from the top decay;
- b quark from the top decay;
- light quarks from the W decay;

Figure 6.16 shows the two-dimensional plots of the HTT variables for matched and not matched events for a semileptonic $t\bar{t}$ sample. The smaller cluster of events at the bottom of the plot is associated to unmatched top jets, and it can be shown that the cluster persists and becomes more prominent when raising the p_T threshold, thus it is not originated by low- p_T top decays. This effect can be interpreted as a failure of the clustering algorithm of the HEPTOPTAGGER to correctly separate the three subjects at high p_T , as the angular separation between the subjects might be too small at high energies.

6.2.5 Scale factor measurement

The scale factor measurement aims at correcting the discrepancies in data and simulation of the efficiency of the tagger. The measurement is performed using the selection outlined in Section 6.2.3. The scale factor is defined as:

$$SF = \frac{\epsilon_{\text{data-bkg}}}{\epsilon_{\text{MC signal}}} ,$$

where:

$$\epsilon_{\text{MC signal}} = \frac{N_{\text{after HTT}}^{\text{tt semi}}}{N_{\text{before HTT}}^{\text{tt semi}}} ,$$

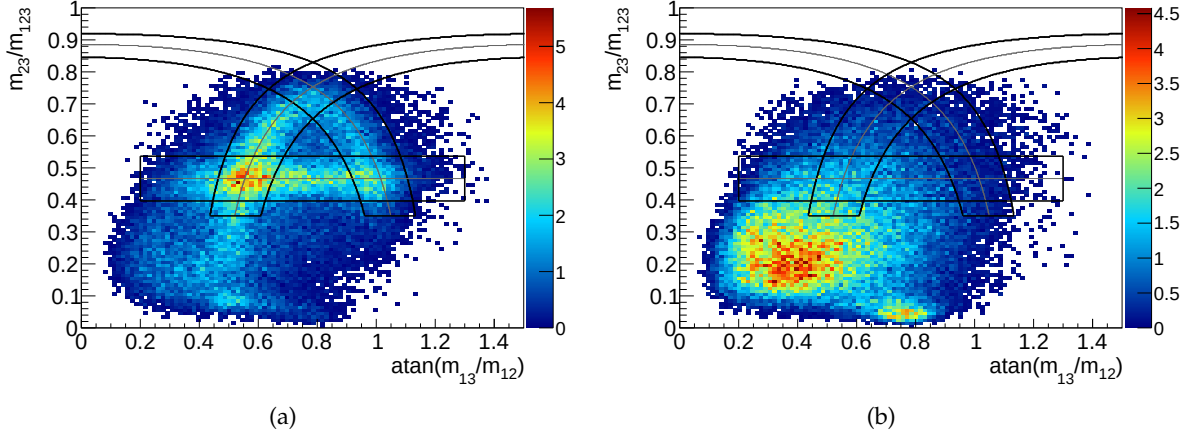


Figure 6.16: Bidimensional distribution of the variables the HTT uses for the W mass selection in $t\bar{t}$ events for gen-level matched (a), and not matched (b) top jets.

and similarly:

$$\epsilon_{\text{data-bkg}} = \frac{N_{\text{after HTT}}^{\text{data}} - N_{\text{after HTT}}^{\text{MC background}}}{N_{\text{before HTT}}^{\text{data}} - N_{\text{before HTT}}^{\text{MC background}}} ,$$

where $N_{\text{before/after HTT}}^{\text{tt semi}}$ is the number of weighted events in the semileptonic $t\bar{t}$ Monte Carlo sample before or after the HTT is applied, $N_{\text{before/after HTT}}^{\text{data}}$ is the number of events in data before or after the HTT is applied, and $N_{\text{before HTT}}^{\text{MC background}}$ is the number of weighted events from all the simulated background samples before or after the HTT is applied. The computation of the scale factor properly takes into account the contamination from the background processes, as the selection is not 100% pure in semileptonic $t\bar{t}$ events.

The scale factor is measured separately for the HTT alone (WP2) and the HTT in combination with N-subjettiness (WP3). The scale factor is measured in three different p_T bins and two η bins, for three different $t\bar{t}$ simulated samples and is shown in Table 6.3.

Additionally, the scale factor is calculated sequentially for each of the selection steps of the HTT WP3 in Table 6.4. The term “sequentially” means that the scale factor is measured for each selection with respect to the number of events passing the previous selection.

6.3 Conclusions

The HEPTopTagger algorithm and other jet substructure observables, including N-subjettiness, are studied in both data and simulation in 8 TeV data. A semileptonic $t\bar{t}$ selection is used to acquire a large, pure sample of merged boosted top jets and measure data-simulation scale factors for the top tagging algorithm. Studies of simulated events illustrate the performance of the HTT in separating top and QCD jets. Further sensitivity can be gained by utilizing the top tagger in combination with jet substructure variables to improve top tagging performance.

The optimal top tagging algorithm is the one that provides the smallest mistag rate for the desired signal efficiency. The HEP Combined Tagger, combining the discriminating power of the HEP Top Tagger, subjet b-tagging, and N-subjettiness performs best for low p_T selections (jets matched to partons with $p_T < 500$ GeV). For higher p_T selections, the CMS Combined Tagger, combining the discriminating power of the CMS Top Tagger, N-subjettiness, and subjet b-tagging is a more suitable algorithm.

$ \eta < 1.0$				
Tagger	Generator	$200 < p_T < 250 \text{ GeV}$	$250 < p_T < 400 \text{ GeV}$	$p_T > 400 \text{ GeV}$
Comb. WP2	MADGRAPH	0.91 ± 0.04	0.93 ± 0.03	1.15 ± 0.07
	POWHEG	0.92 ± 0.04	0.95 ± 0.03	1.36 ± 0.07
	MC@NLO	0.88 ± 0.04	0.93 ± 0.03	1.19 ± 0.07
Comb. WP3	MADGRAPH	0.86 ± 0.05	0.91 ± 0.04	0.98 ± 0.09
	POWHEG	0.86 ± 0.05	0.93 ± 0.04	1.10 ± 0.12
	MC@NLO	0.80 ± 0.04	0.87 ± 0.03	1.00 ± 0.10
$1.0 < \eta < 2.4$				
Selection	Generator	$200 < p_T < 250 \text{ GeV}$	$250 < p_T < 400 \text{ GeV}$	$p_T > 400 \text{ GeV}$
Comb. WP2	MADGRAPH	0.95 ± 0.05	0.91 ± 0.04	0.85 ± 0.11
	POWHEG	0.93 ± 0.06	0.95 ± 0.05	0.95 ± 0.15
	MC@NLO	0.93 ± 0.05	0.95 ± 0.04	0.99 ± 0.13
Comb. WP3	MADGRAPH	1.02 ± 0.07	0.90 ± 0.05	0.85 ± 0.16
	POWHEG	1.00 ± 0.08	0.97 ± 0.06	1.00 ± 0.22
	MC@NLO	0.96 ± 0.07	0.93 ± 0.05	0.99 ± 0.19

Table 6.3: Data-simulation scale factors for the HEP combined tagger WP2 and WP3. The scale factor is measured for three different generators in three p_T regions and in two η regions.

$ \eta < 1.0$				
Tagger	p_T bin (GeV)	MadGraph	powheg	mc@nlo
HEP top mass selection	$200 < p_T < 250$	0.92 ± 0.02	0.94 ± 0.03	0.92 ± 0.02
	$250 < p_T < 400$	0.93 ± 0.02	0.96 ± 0.02	0.94 ± 0.02
	$p_T > 400$	1.03 ± 0.03	1.07 ± 0.04	1.04 ± 0.04
HEP W mass selection	$200 < p_T < 250$	0.98 ± 0.03	0.98 ± 0.04	0.96 ± 0.03
	$250 < p_T < 400$	0.99 ± 0.02	0.99 ± 0.03	0.99 ± 0.02
	$p_T > 400$	1.11 ± 0.05	1.27 ± 0.07	1.14 ± 0.06
N-subjettiness selection	$200 < p_T < 250$	0.95 ± 0.04	0.93 ± 0.04	0.90 ± 0.03
	$250 < p_T < 400$	0.98 ± 0.03	0.98 ± 0.03	0.94 ± 0.03
	$p_T > 400$	0.85 ± 0.06	0.81 ± 0.07	0.84 ± 0.06
$1.0 < \eta < 2.4$				
Tagger	p_T bin (GeV)	MadGraph	powheg	mc@nlo
HEP top mass selection	$200 < p_T < 250$	0.89 ± 0.03	0.89 ± 0.04	0.90 ± 0.03
	$250 < p_T < 400$	0.92 ± 0.02	0.96 ± 0.03	0.97 ± 0.02
	$p_T > 400$	0.92 ± 0.07	0.98 ± 0.08	1.07 ± 0.08
HEP W mass selection	$200 < p_T < 250$	1.07 ± 0.04	1.04 ± 0.05	1.03 ± 0.04
	$250 < p_T < 400$	0.99 ± 0.03	0.99 ± 0.04	0.98 ± 0.03
	$p_T > 400$	0.92 ± 0.10	0.97 ± 0.13	0.92 ± 0.11
N-subjettiness selection	$200 < p_T < 250$	1.07 ± 0.05	1.07 ± 0.05	1.03 ± 0.05
	$250 < p_T < 400$	0.99 ± 0.04	1.03 ± 0.05	0.98 ± 0.04
	$p_T > 400$	1.01 ± 0.13	1.06 ± 0.17	1.00 ± 0.14

Table 6.4: Data-simulation scale factors for each sequential selection of the HEP combined tagger WP3. The scale factor is measured for three different generators in three p_T regions and in two η regions.

Search for heavy resonances decaying to $t\bar{t}$ with 8 TeV data

7.1 Introduction

As described in Chapter 1, many models of new physics predict heavy resonances with enhanced couplings to the third generation of the Standard Model (SM) [33, 41, 42, 100–105]. Thus, the study of the top quark can give an important insight into the validity of such models. This analysis studies top quark pair production to test for the presence of heavy resonances. Generic models where the resonance is labeled Z' are tested, as well as more specific models such as Randall–Sundrum KK gluon production.

This chapter presents the search for resonant $t\bar{t}$ production with 8 TeV data in the all-jets final state. In the high mass ranges accessible by the LHC at $\sqrt{s} = 8$ TeV, reconstructing the event topology of $t\bar{t}$ production requires special techniques. For example, if a hypothetical Z' resonance has a mass larger than 1 TeV and decays to $t\bar{t}$, the produced top quarks will be highly boosted and the subsequent $t \rightarrow Wb$ decays might merge together. Jet substructure techniques and top quark identification algorithms described and studied in Chapter 4 will be used. The topology is considered, where the hadronic decay products of the top quark are fully merged into a single jet. Given the extended p_T range of the boosted top quarks that will be reconstructed, differing reconstruction algorithms for the identification will be used. Figure 7.1 shows a schematic representation of an all-jet $t\bar{t}$ event.

For moderately boosted top quarks (jets with $200 < p_T < 400$ GeV) the HEPTOPTAGGER algorithm is used, along with subjet b-tagging. For top quarks with a large boost ($p_T > 400$ GeV) the CMS Top Tagger is used in combination with N-subjettiness and subjet b-tagging algorithms to identify top jets. The two top quark decay topologies are combined to search for boosted $t\bar{t}$ candidate events. This chapter will focus on the moderately boosted topology using the HEPTOPTAGGER, however a summary of the large boost analysis will be given. The results of the search will be presented as a combination of the two all-jets boosted topologies. The search for resonant $t\bar{t}$ production in the all-jets final state has been combined with the lepton-plus-jets and dilepton-plus-jets final states. A summary of the final combination will be given as well.

A search for $t\bar{t}$ resonances in all-hadronic final states was previously performed by CMS at $\sqrt{s} = 7$ TeV [106] and with a partial dataset at $\sqrt{s} = 8$ TeV [107]. Both of these searches utilize the CMS Top Tagger, but no b-tagging or subjet b-tagging. This search uses the full 8 TeV dataset with new top tagging tools and subjet b-tagging.

The dominant background for this analysis, given that the selection consists of dijet events,

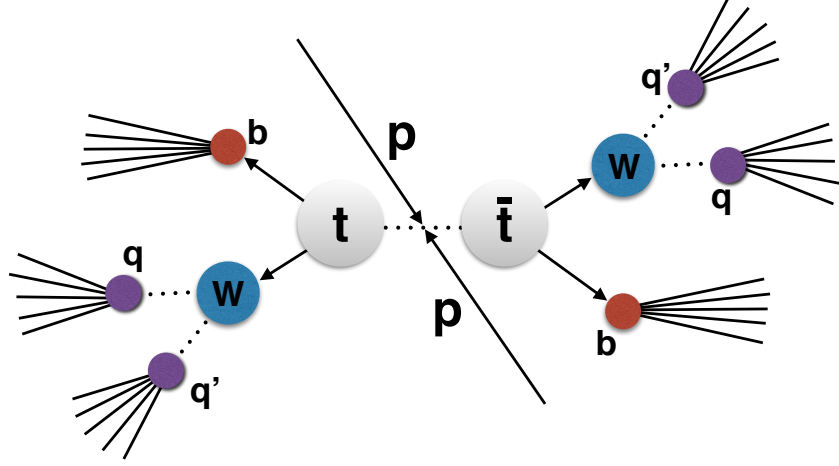


Figure 7.1: A schematic representation of an all-jet $t\bar{t}$ event.

is QCD multijet production. This contribution is determined from a data-based method where the top tagging misidentification rate is derived in a QCD-enriched sideband. Additional background contribution from standard model $t\bar{t}$ production is determined from simulation. The analysis presented here has been published in Ref. [108], and additional details of the analysis can be found in Ref. [109].

7.2 Samples

7.2.1 Simulated event samples

All simulated and data samples are reconstructed using the CMS framework version 5.3. Samples of simulated events are used to estimate the background from SM $t\bar{t}$ events. These samples were generated with POWHEG, and a cross section of $\sigma_{t\bar{t}} = 245$ pb [110] is used to determine the expected number of events. The $t\bar{t}$ sample used is inclusive in the top decay products. The analysis uses also simulated events of QCD multijet production for the validation of the background estimation procedure. The QCD multijet samples are split using the H_T variable calculated at generator-level. The samples are simulated with MADGRAPH and showered with PYTHIAv6. Table A.4 reports the simulated samples used in this analysis.

Different hypothesized models of new physics are used in this analysis to simulate events containing the decay of a heavy resonance in $t\bar{t}$ pairs. One physics model is that of Randall-Sundrum Kaluza-Klein (RS KK) gluon production [43]. In addition, a more generic Z' resonance production is studied as well [111], with two scenarios for the Z' width: $\Gamma/m = 0.1$ and $\Gamma/m = 0.01$. The signal samples are generated with PYTHIA8 for masses ranging from 0.75 to 3.0 TeV. Figure 7.2 shows the mass distribution of the $t\bar{t}$ system for some of the signal samples used in the analysis. Above 3 TeV, the resolution for these samples degrades significantly due to the drop in parton luminosities. Table 7.1 shows the simulated samples used in the analysis, full names of the samples used is reported in Table A.5. The cross sections listed for the RS KK gluon samples are obtained from the authors of Ref. [43].

7.2.2 Data samples

The data used in this analysis was collected during the year 2012, at a center-of-mass energy of $\sqrt{s} = 8$ TeV. Two different data streams are used for the analysis: the jet data-stream

Sample	Mass	Cross Section
RS KK Gluon	1 TeV	6.299 pb
RS KK Gluon	1.5 TeV	0.828 pb
RS KK Gluon	2 TeV	0.182 pb
RS KK Gluon	3 TeV	0.021 pb
RS KK Gluon	4 TeV	0.005 pb
Narrow Z'	1 TeV	—
Narrow Z'	1.5 TeV	—
Narrow Z'	2 TeV	—
Narrow Z'	3 TeV	—
Narrow Z'	4 TeV	—
Wide Z'	1 TeV	—
Wide Z'	1.5 TeV	—
Wide Z'	2 TeV	—
Wide Z'	3 TeV	—
Wide Z'	4 TeV	—

Table 7.1: Simulated signal samples used in the analysis.

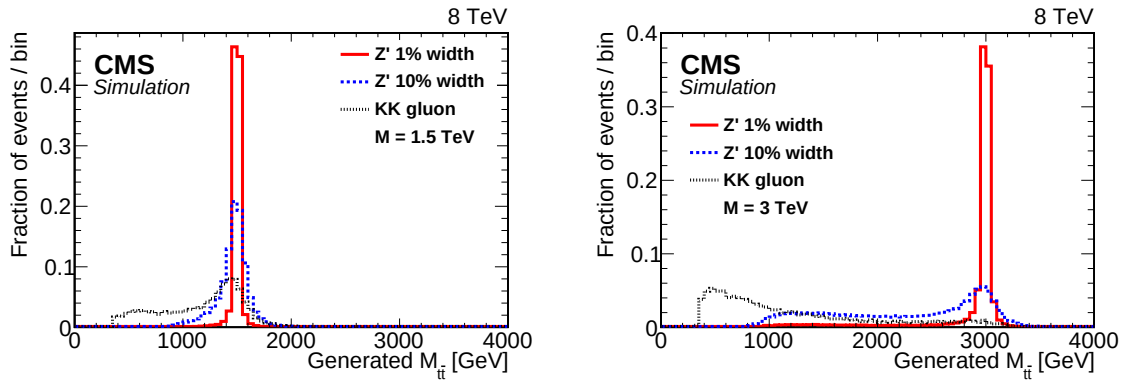


Figure 7.2: Mass distribution of the $t\bar{t}$ system for different signal models. The distributions are shown for masses of the resonance of (left) 1.5 TeV and (right) 3 TeV [108].

containing a selection of all-jets, high-threshold triggers, and the multijet parked data-stream containing a selection of all-jets, medium-threshold triggers. Due to the high inclusive trigger rate of the latter, the data stream was reconstructed at the end of the 2012 run, therefore saving computational time to promptly analyze the other data streams. This technique is called “dataset parking”. The total dataset considered in this analysis corresponds to an integrated luminosity of 19.7 fb^{-1} for the jet data-stream and 18.3 fb^{-1} for the multijet parked data-stream. The latter dataset has a smaller luminosity because the technique of data parking was not available during the early data-taking period. Table A.6 summarizes the data samples used for this analysis, and their corresponding integrated luminosities.

7.2.3 Trigger selection

The analysis uses an H_T trigger selecting events with a total energy above 750 GeV. The energy sum is calculated on the reconstructed calorimeter jets with a corrected transverse momentum greater than 40 GeV. This is the lowest-threshold unprescaled H_T trigger present in the 2012 trigger menus. Given that this analysis searches for moderately boosted top quarks ($p_T > 200 \text{ GeV}$), the H_T trigger alone would be limiting the strengths of the HEPTopTAgGER. Thus the 4-jets trigger, contained in the multijet parked data-stream, is used in addition. This trigger requires at least 4 jets with at least 50 GeV of p_T each.

The H_T trigger efficiency is parametrized using the sum of the p_T of all the reconstructed anti- k_t standard-sized jets, with $p_T > 50 \text{ GeV}$. The requirement of $H_T > 800 \text{ GeV}$ is then made to be in the region where the trigger is more than 95% efficient. The trigger efficiency is calculated for events in the kinematic region of interest of the analysis, requiring two HEPTopTAgGER jets, with $p_T > 200 \text{ GeV}$ (Sec. 7.3). Events selected by the analysis using the CMS Top Tagger are excluded, as described in Sec. 7.3 because the two analyses are meant to be combined. Figure 7.3(left) shows the H_T trigger efficiency curve parameterized as a function of H_T for a representative signal sample. For the 4-jets trigger, the efficiency is parametrized as a function of the p_T of the fourth most energetic AK5 jet, which is required to have a p_T greater than 70 GeV to stay in a region where the trigger is highly efficient. Figure 7.3(right) shows the 4-jets trigger efficiency curve parameterized as a function of the p_T of the fourth most energetic jet for a representative signal sample.

Given this trigger strategy, two different event categories are introduced:

- A boosted category, requiring $H_T > 800 \text{ GeV}$ and triggered by the H_T trigger.
- A moderate boost category, requiring $H_T < 800 \text{ GeV}$ and triggered by the 4-jets trigger.

The events with $H_T > 800 \text{ GeV}$, passing the 4-jets trigger but not the H_T trigger are lost in this setup, but it has been checked that this number is really small and poses no problem to the sensitivity of the analysis.

This analysis uses events close to the threshold cut of the triggers, thus a dedicated correction factor is needed to account for trigger efficiency discrepancies between data and simulation. The trigger efficiencies are measured using an uncorrelated muon trigger requiring the presence of at least one muon with $p_T > 40 \text{ GeV}$. For this study, the signal region is defined as events containing two jets with $p_T > 200 \text{ GeV}$ satisfying the requirements of the HEPTopTAgGER, at least one of the two jets must have a b-tagged subjet (“medium” operating point). The b-tagging requirement ensures that the selection is sufficiently enriched in $t\bar{t}$ events. For the correction factor measurement, the trigger efficiency is defined as:

$$\epsilon = \text{Numerator/Denominator},$$

where:

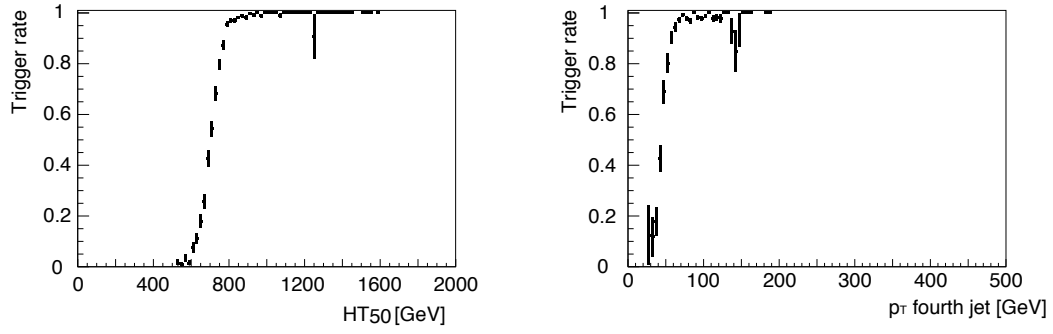


Figure 7.3: Trigger efficiency curve for the H_T (left) and 4-jets (right) triggers for a representative signal sample ($M_{Z'} = 1.5$ TeV, narrow width).

Denominator is the number of events:

- passing the muon trigger,
- with two top-tagged (HTT) jets with $p_T > 200$ GeV,
- with at least one b-tagged subjet.

Numerator is the number of events:

- passing the denominator requirements,
- passing the H_T trigger (or the 4-jets trigger, depending on the trigger being studied).

The data-simulation correction factor for the trigger efficiency is defined as $\epsilon_{\text{data}}/\epsilon_{t\bar{t} \text{ simulation}}$, where ϵ_{data} is the trigger efficiency measured in data, and $\epsilon_{t\bar{t} \text{ simulation}}$ is the trigger efficiency measured in a simulated $t\bar{t}$ sample. The SM $t\bar{t}$ simulation is used as the simulated sample, while for data the single-muon data-stream reported in Table A.1 is used. In Figure 7.4 a comparison of the trigger efficiencies in data and simulation is shown for the H_T (left) and 4-jets (right) triggers.

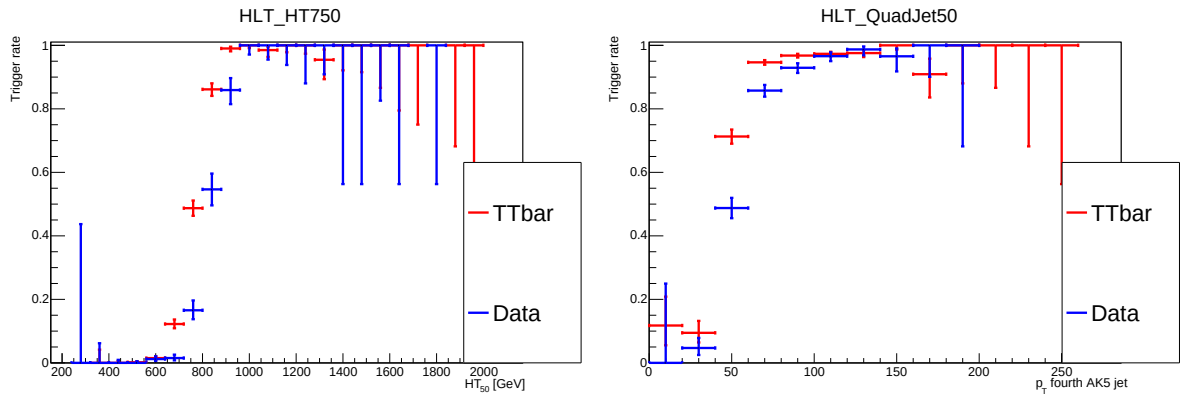


Figure 7.4: Comparison of the trigger efficiencies in data and simulation for the H_T (left) and 4-jets (right) triggers.

7.3 Selection of events and optimization

Some basic selection criteria are applied to select events with a hard scatter collision: the event must have a good primary vertex as computed by a deterministic annealing filter. The collinear distance with respect to the center of the detector is required to be $|z_{\text{Primary Vertex}}| < 24$ cm, and the number of degrees of freedom of the vertex fit must satisfy $N_{\text{DOF}} > 6$.

To select events containing the decay of a pair of highly-energetic top quarks, the event are required to have at least two CA15 jets with $p_T > 200$ GeV, $|y| < 2.4$. Additionally, loose particle-flow jet identification criteria, as described in Ref. [112], are applied to each jet. The two most energetic jets are considered further in the analysis. To these jets, the HEPTOPTAGGER grooming procedure is applied. To enhance the top-tagging performance of the analysis, subjet b-tagging is applied in addition to the HEPTOPTAGGER criterion. Subjet b-tagging is computed on the subjets obtained from a filtering procedure with a distance parameter of $R_{\text{filt}} = 0.3$ [75] applied to CA15 jets. CA15 filtered jets are then matched to jets produced by the HEPTOPTAGGER via a ΔR requirement, to avoid double-counting. Figures 7.5 and 7.6 show a display of a candidate boosted all-jets $t\bar{t}$ event.

Jet energy corrections are not applied to the subjets of HEPTOPTAGGER and filtered jets, and algorithmic procedures (jet filtering, HEPTOPTAGGER) are performed on uncorrected jet energies. However the corrections are applied to large-cone jets. The corrections are derived from studies using the anti- k_T $R = 0.7$ jet algorithm. The studies documented in Ref. [114] confirm that the anti- k_T -derived jet corrections ($R = 0.7$) are adequate for the CA $R = 1.5$ jet algorithm and can be safely used with an additional flat systematic uncertainty of 4%.

As previously mentioned, this search is meant to be combined with an analysis probing the very boosted regime, using the CMS Top Tagger instead of the HEPTOPTAGGER. The HEPTOPTAGGER and the CMS Top Tagger analyses investigate complementary kinematic regions, exploiting the region of best performance of the respective taggers. Therefore the event selections need to be exclusive: events that are part of the signal region or the sideband region used for the background estimate of the CMS Top Tagger analysis are removed. This is achieved excluding the events containing two CA $R=0.8$ jets with $p_T > 400$ GeV. The analysis is combined with searches for $t\bar{t}$ resonances in the lepton+jets and dileptonic final states. The overlapping between the selections of these three analyses has been checked and found to be negligible, therefore no muon or electron veto is applied to the selection. A graphical representation of the phase space areas covered by the HEPTOPTAGGER and CMS Top Tagger analyses can be seen in Figure 7.7; in Figure 7.8 the approximate Z' mass coverage of the analyses is shown.

Several options for the categorization of the events are explored, including the application of selections on N-subjettiness, subjet b tagging and rapidity difference between the jets. The analysis is optimized by computing expected cross section limits, in several different scenarios.

Concerning subjet b tagging, the CSV score is computed for all the subjets in a top jet, the variable β_{max} is defined as the maximum subjet CSV score for a top jet. A top jet is considered b tagged if $\beta_{\text{max}} > 0.679$, the threshold for the medium-efficiency CSV operating point [67].

No significant improvement in the sensitivity of the analysis is obtained when applying the selections on N-subjettiness or rapidity. The optimal classification is found to be the one categorizing events independently according to the number of b-tagged top jets found, according to the above definition. Taking into account the classification based on the trigger selection, there are a total of six independent event categories, summarized in Table 7.2.

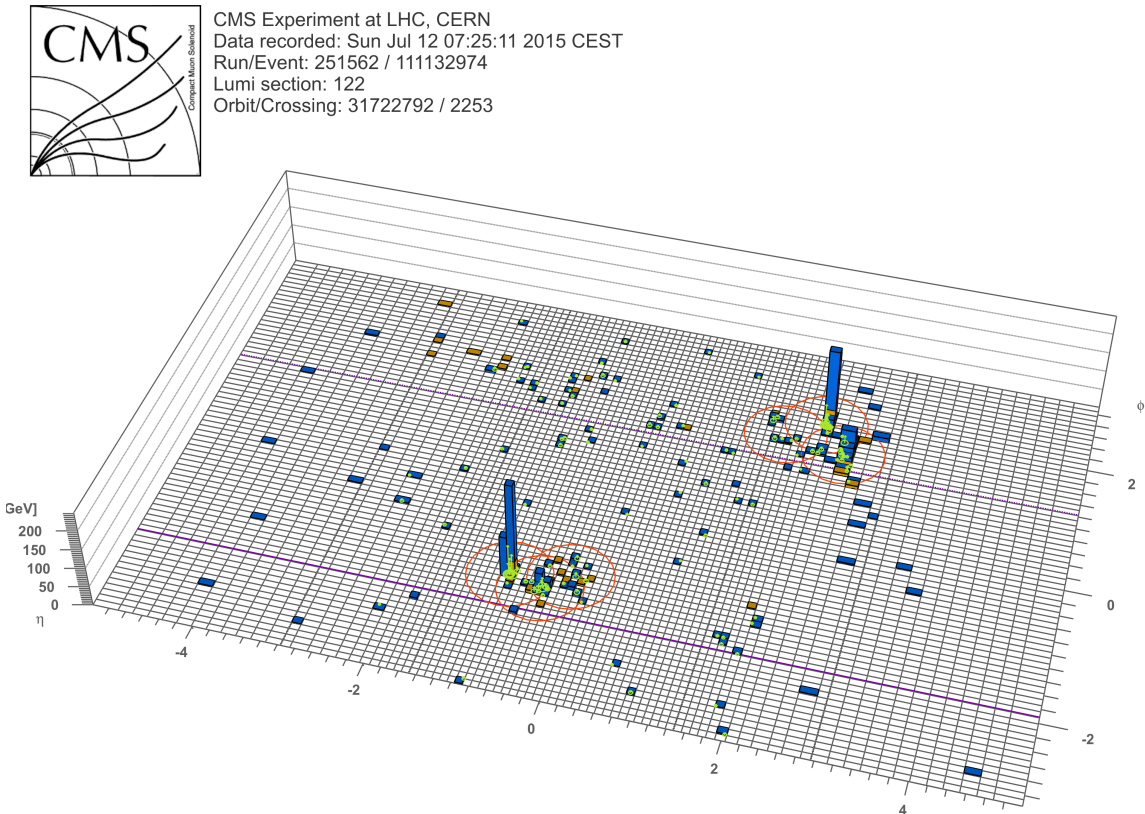
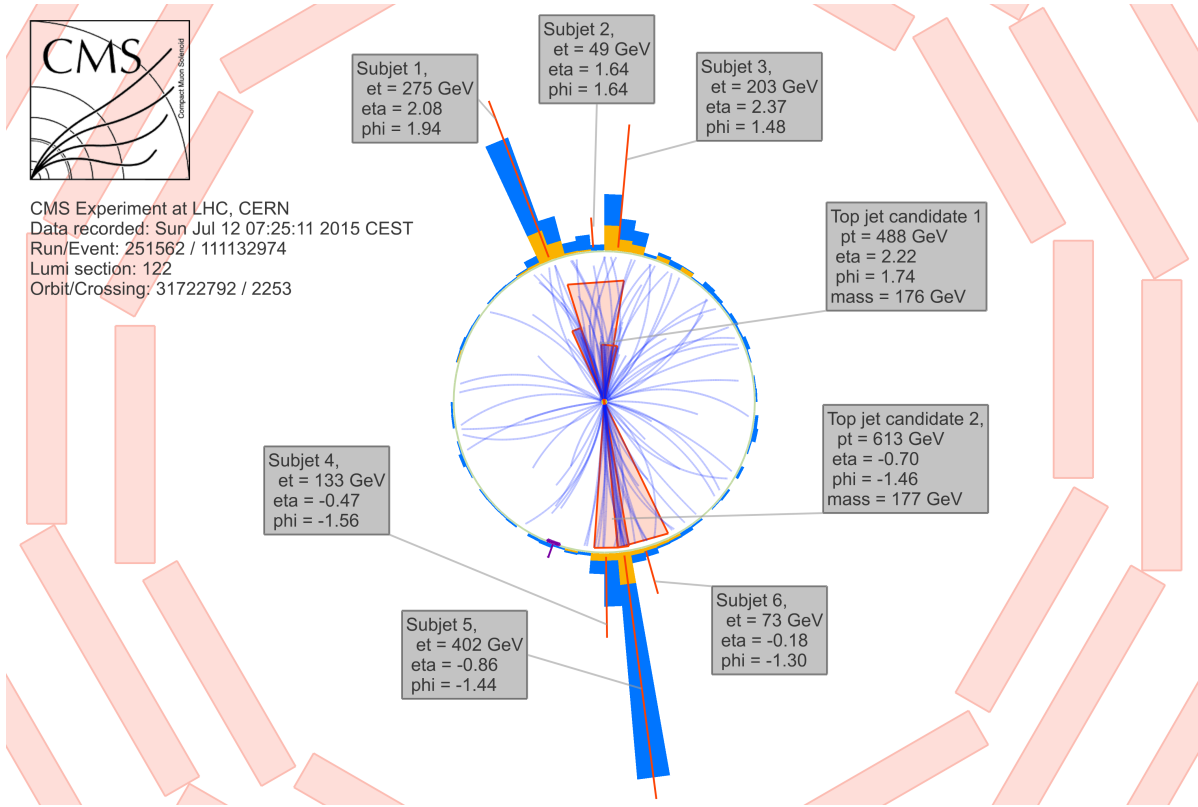


Figure 7.5: A display of a candidate boosted all-jets $t\bar{t}$ event. The yellow towers are ECAL clusters, the blue towers are HCAL clusters. The size of the towers is proportional to their energy. The tracks are charged particle-flow candidates [113].

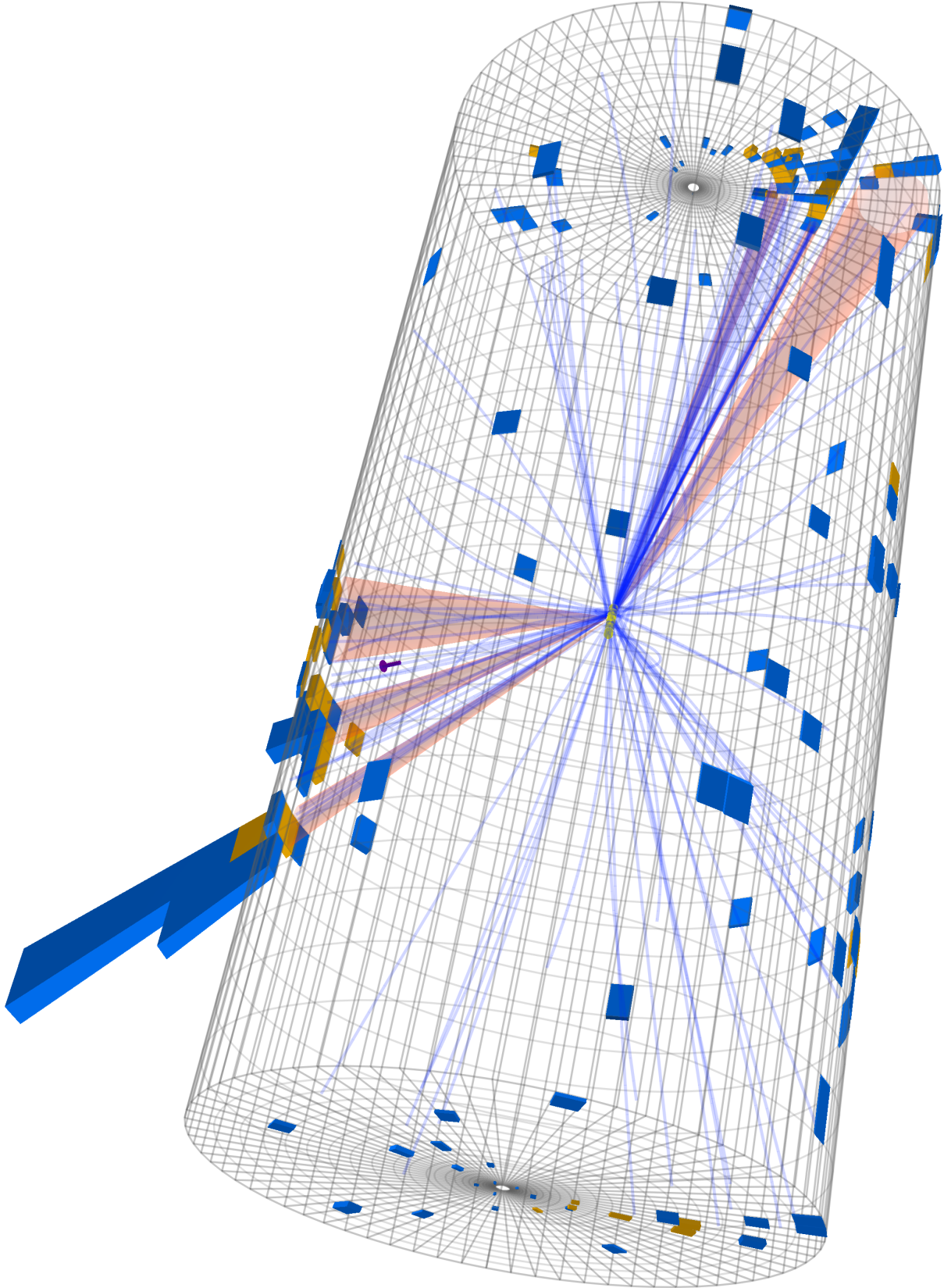


Figure 7.6: A display of a candidate boosted all-jets $t\bar{t}$ event. The yellow towers are ECAL clusters, the blue towers are HCAL clusters. The size of the towers is proportional to their energy. The tracks are charged particle-flow candidates [113].

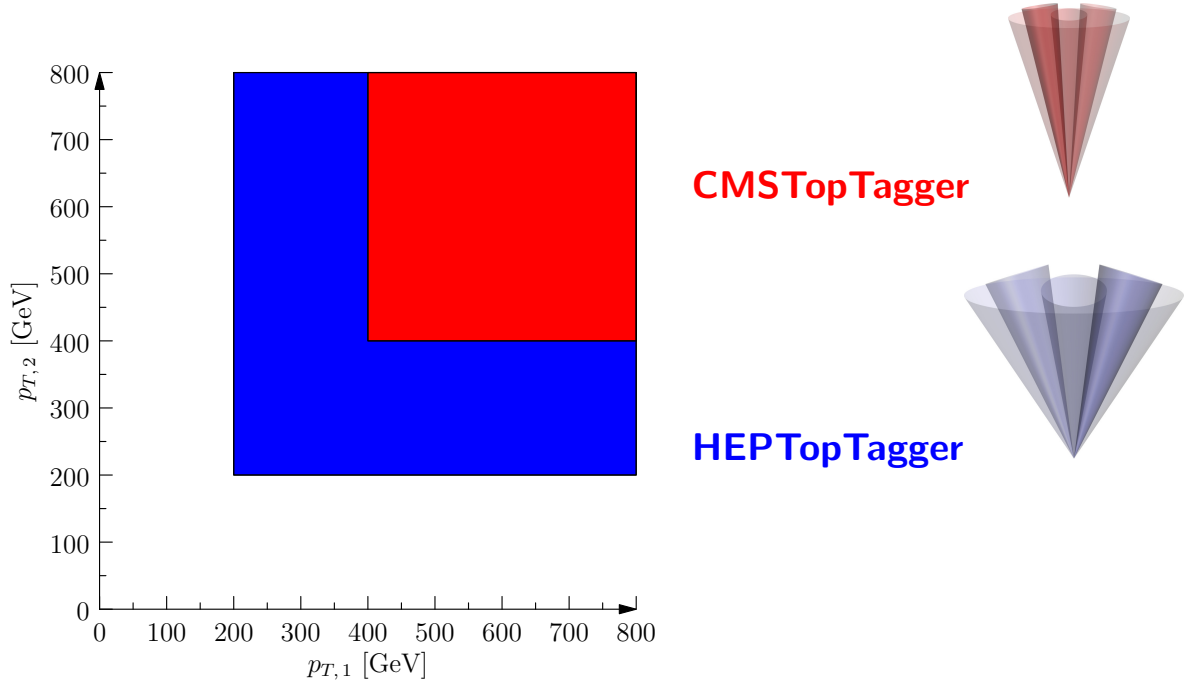


Figure 7.7: Graphical visualization of the areas of the phase space covered by the HEPTopTAgGER and CMS Top Tagger analyses.

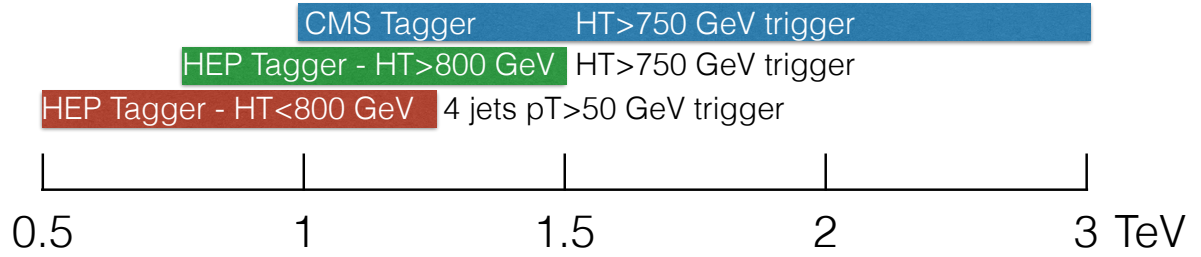


Figure 7.8: Graphical visualization of the approximate Z' mass coverage by the HEPTopTAgGER and CMS Top Tagger analyses, after the orthogonality requirement.

Event Category	HTT on both jets	Number of b-tags	H_T cut
A	✓	0	> 800 GeV (H_T trigger)
B	✓	1	> 800 GeV (H_T trigger)
C	✓	≥ 2	> 800 GeV (H_T trigger)
D	✓	0	< 800 GeV (4-jets trigger)
E	✓	1	< 800 GeV (4-jets trigger)
F	✓	≥ 2	< 800 GeV (4-jets trigger)

Table 7.2: Event selections defining the six categories used in the analysis.

7.3.1 Normalization of simulated samples

Simulated samples are used to estimate the $t\bar{t}$ background and to model the signal. To determine the correct normalization of signal and expected background events, several event weights are needed. Equation 7.1 shows the event weights applied to simulated events. These include scale factors for substructure methods, $t\bar{t}$ p_T -spectrum reweighting, trigger efficiencies, and normalization to cross section and integrated luminosity.

$$w_{\text{ev}} = \frac{\epsilon_{\text{filt}}}{N_{\text{gen}}} \cdot \sigma \cdot L \cdot \epsilon_{\text{trig}} \cdot f_{\text{sub}} \cdot f_{\text{top}} \cdot f_{\text{pileup}} \quad (7.1)$$

The term ϵ_{filt} is the filter efficiency of the simulated sample (the filter is applied to select a certain kinematic range in some samples), N_{gen} is the number of events in the sample, σ is the cross section of the process, and L is the integrated luminosity of the data samples analyzed. The $t\bar{t}$ sample is normalized to the cross section of $\sigma_{t\bar{t}} = 245$ pb [110]. The quantity f_{sub} represents the scale factor used for the top-tagging algorithm. Additionally, f_{pileup} is the weight to correct for the different distribution of the numbers of primary vertices in data and simulation, and ϵ_{trig} is the correction factor for the trigger efficiency.

Substructure scale factors f_{sub}

The efficiency of the HEPTOPTAGGER has been measured and compared to simulation to derive scale factors in Chapter 6. The scale factors are applied as an event weight and the uncertainties of these scale factor measurements are taken into account accordingly.

Top quark p_T reweighting f_{top}

This normalization factor is applied only to simulated $t\bar{t}$ events. Due to observed discrepancies in the $t\bar{t}$ p_T spectrum between simulated and observed data events, a recommended reweighting procedure is applied to improve the agreement in the p_T spectrum between data and $t\bar{t}$ events. Event weights are applied based on the p_T values of the two generated top quarks in the event, according to Equation 7.2.

$$f_{\text{top}} = \sqrt{(e^{0.156-0.00137p_T^t})(e^{0.156-0.00137p_T^{\bar{t}}})} \quad (7.2)$$

The weight is derived for p_T values up to 400 GeV. For $p_T > 400$ GeV, the value of f_{top} with $p_T = 400$ GeV is used. For the combination with the lepton-plus-jets and dilepton analyses, the top quark p_T reweighting is not applied to the analysis selection as it leads to a worse data-simulation agreement for the leptonic analyses.

7.4 Background estimation

This analysis has two sources of background: QCD multijet events, and SM $t\bar{t}$ events. The former background is determined from data-driven method, while for the latter simulated events are used. The simulated events are normalized according to Section 7.3.1.

The method to estimate the QCD multijet background involves selecting a sample with a low SM $t\bar{t}$ contribution. This is done by inverting the top tagging requirements on one selected jet (“anti-tag”) and determining the top tagging rate for the second jet (“probe”). This “anti-tag and probe” method yields a per-jet misidentification rate that is then applied to each jet in the event to determine the estimated number of QCD multijet background events. The misidentification rate is parameterized as a function of two quantities: jet p_T , and β_{max} . This

parametrization will be referred to as the “mistag matrix”. Figure 7.10 shows the mistag rate for different bins of the quantities used in parametrization. Because there can be a significant contribution from SM $t\bar{t}$ events, even after inverting the top tagging requirements on one jet, this contribution is first estimated from simulated events, then subtracted before computing the mistag matrix. The misidentification rate is determined by the following procedure.

1. Events are selected with at least two CA15 jets with $p_T > 200$ GeV.
2. One among the two most energetic CA15 jets is selected at random – this jet is “anti-tagged” to select an enriched sample of QCD multijet events. This jet must not satisfy the W mass selection applied by the HEPTopTAgGER. In order to be as close as possible to the kinematics of jets in the signal region, the jet is required to have a mass in the same window ($140 < m_{jet} < 250$ GeV) as the signal region and $\frac{m_{23}}{m_{123}}$ is required to be greater than 0.39. The anti-tag selection is depicted in Fig. 7.9: the events lying outside of the black region, and within the top mass window are defined as “anti-tagged”.
3. Once the anti-tagged jet is selected, the second jet is then used as the “probe” jet.
4. The top tagging rate of the probe jet is the misidentification rate for the top tagging algorithm, measured simultaneously in bins of jet p_T and jet β_{max} . Due to the inversion of the HEPTopTAgGER cuts, the data sample is dominated by multijet QCD in a kinematic regime close to that of the signal region, and is therefore an appropriate control region to extract the misidentification rate in different bins of the mistag matrix.
5. The contribution from continuum SM $t\bar{t}$ production is removed based on the expectation from simulated events.

The mistag matrix evaluated in data is shown in Fig. 7.10 separately for the low- H_T and high- H_T categories.

Once the mistag matrix is determined from the QCD-enriched control sample, it can be used to estimate the normalization and shape of QCD multijet events passing the final event selection. To do this, a selection that is looser than that of the signal region is used. A sample containing at least two CA15 jets is considered, one of the two leading top jet candidates is randomly chosen and required to pass the HEPTopTAgGER. If the randomly chosen jet is top-tagged, the event is used for the background estimate and an event weight is applied equal to the misidentification rate from the appropriate bin of the mistag matrix, using the p_T and β_{max} of the second jet. There is no further selection applied on this second jet, and if the randomly chosen first jet is not tagged, the event is not used to model the multijet background.

To model the individual signal region categories with differing numbers of b-tagged jets, the same selections are applied to the events weighted by the mistag matrix as those applied to the signal region. For example, in the 1 b-tag category, the event is required to have exactly 1 b-tagged jet. The b-tagged jet can be either the top tagged jet or the second jet used to determine the mistag rate weight. Analogous selections are applied for the remaining signal regions. Depending of the H_T category, two different mistag matrices are used.

Using a single-tagged control region without any requirements on the non-tagged jet leads to an overlap between the signal region and the region used to estimate the multijet background. To remove the effects of double-counting, the $t\bar{t}$ contribution is subtracted from the multijet estimate. This is done by evaluating the mistag weighting procedure described above on the simulated $t\bar{t}$ events, to find the amount of “mistagged SM $t\bar{t}$ ” events in the sideband region. This contribution is then subtracted from the data-derived contribution to obtain the final normalization and shape of events for the QCD multijet background. The size

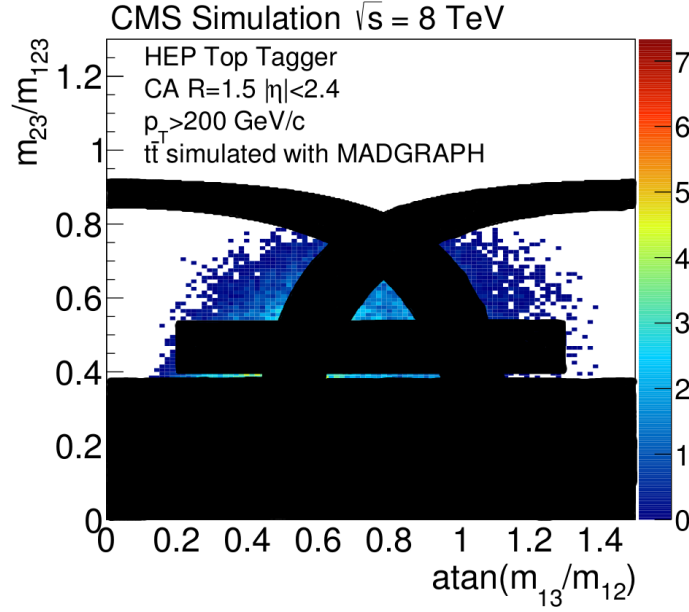


Figure 7.9: Graphical visualization of the inversion of the cuts on the subjet mass variables $\text{arctan} \frac{m_{13}}{m_{12}}$ and $\frac{m_{23}}{m_{123}}$ applied by the HEPTOPTAGGER. The black areas indicate the regions excluded in the “anti-tag”.

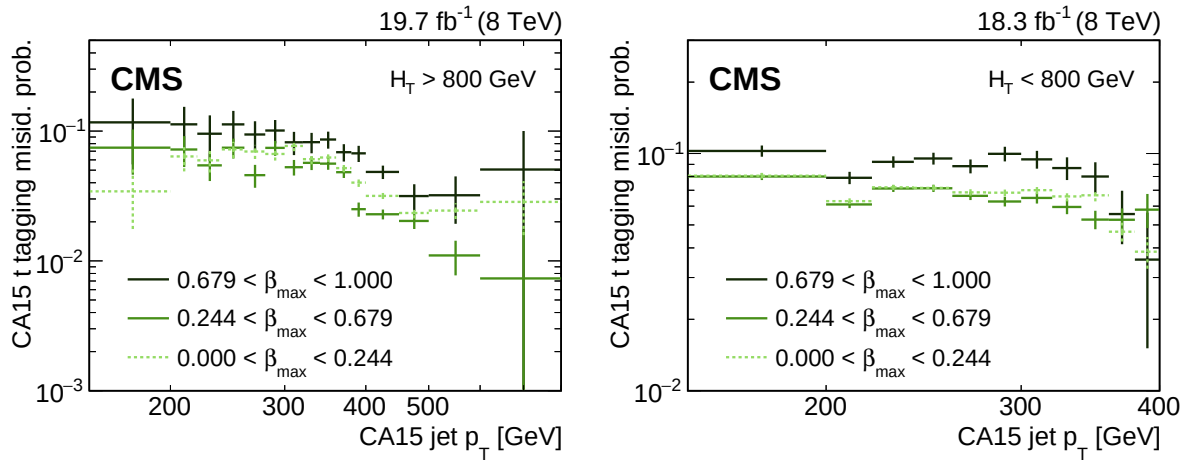


Figure 7.10: Misidentification probability for CA15 jets to be identified as top-quark jets by the HEPTOPTAGGER for different $\beta_{\max}$ values for $H_T > 800 \text{ GeV}$ (left) and $H_T < 800 \text{ GeV}$ (right).

of this overlap depends on the specific region, ranging from circa 5% for the 0 b-tag category, up to circa 15% for the 2 b-tag category.

As a final step to the shape determination of the QCD multijet estimate, the “mass-modified” procedure is used to account for the fact that the non-tagged jet, having no top tag selection applied, will have different kinematics than those jets in the signal region. To mimic the kinematics of the signal region, the mass of non-tagged jet is randomly sampled from a distribution of jet masses from simulated QCD multijet events in the same mass window used for the signal region selection, $140 < m_{jet} < 250$ GeV. Figure 7.11 shows a summary of the background estimation procedure.

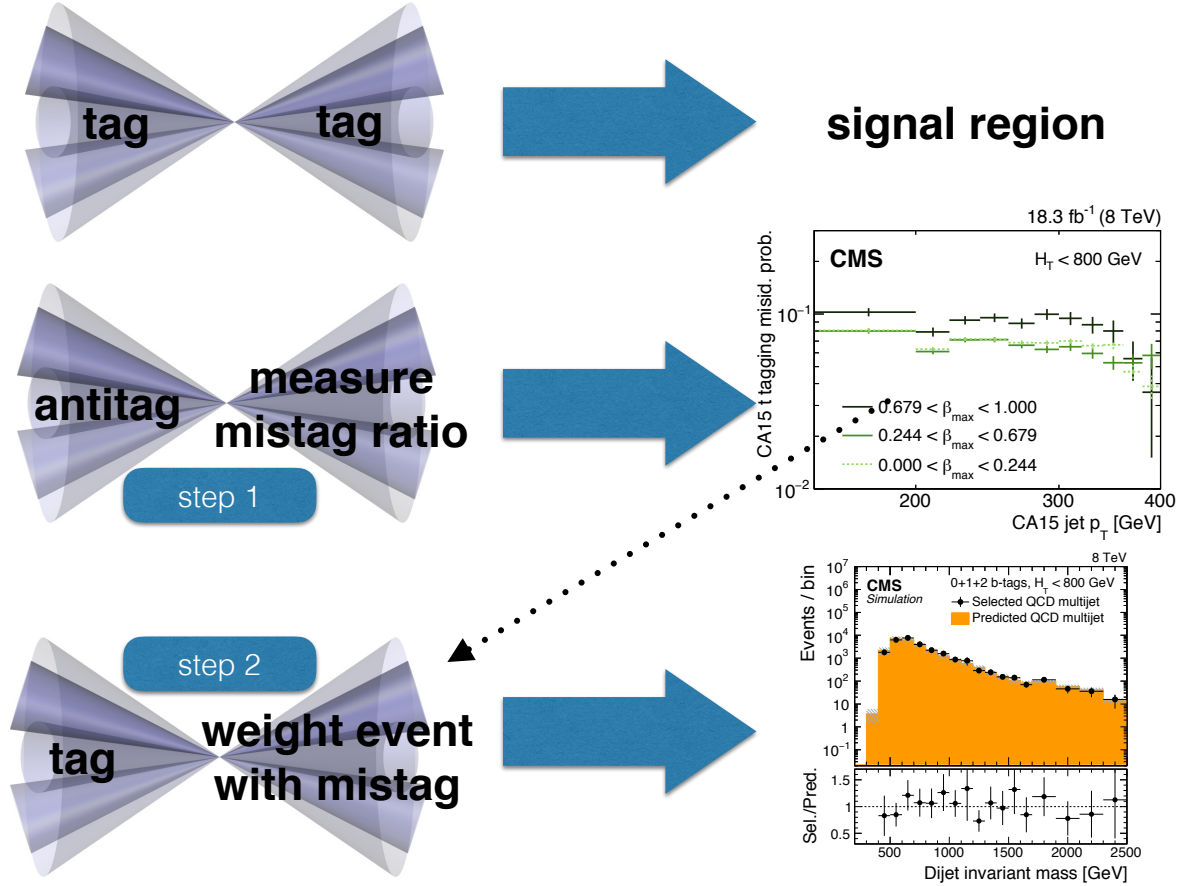


Figure 7.11: A schematic summary of the background estimation procedure.

To validate the method described above to estimate the QCD multijet background, a closure test using simulated QCD events is performed. The simulated events are split into two orthogonal samples: events with odd event numbers and events with even event numbers. One sample of events will be used to measure the mistag matrix according to the procedure outlined above. Once this misidentification rate is obtained, the analysis, including the background estimation procedure, is run on the second sample of simulated events in the usual way. The normalization and shape agreement are compared between the “measured” QCD events, which pass the nominal event selection of two top-tagged jets, and the “predicted” QCD events which are obtained by applying the mistag matrix to the same sample, again according to the procedure above. The mistag rates are measured in the QCD sample in the same bins of jet p_T , and β_{max} . The closure test is performed for each category used in the analysis, including the 0-tag, 1-tag, and 2-tag events, as well as low- H_T and high- H_T cate-

gories. The results are presented in Fig. 7.12, where the blue points represent the “measured” QCD events and red points the “predicted” QCD events using the mistag matrix measured in the orthogonal sample. The agreement between the selected QCD events and the predicted QCD events is found to be satisfactory. As an additional crosscheck it has been checked that the mistag matrix does not have a significant dependence on the maximum b-tagging discriminator of the anti-tagged jet.

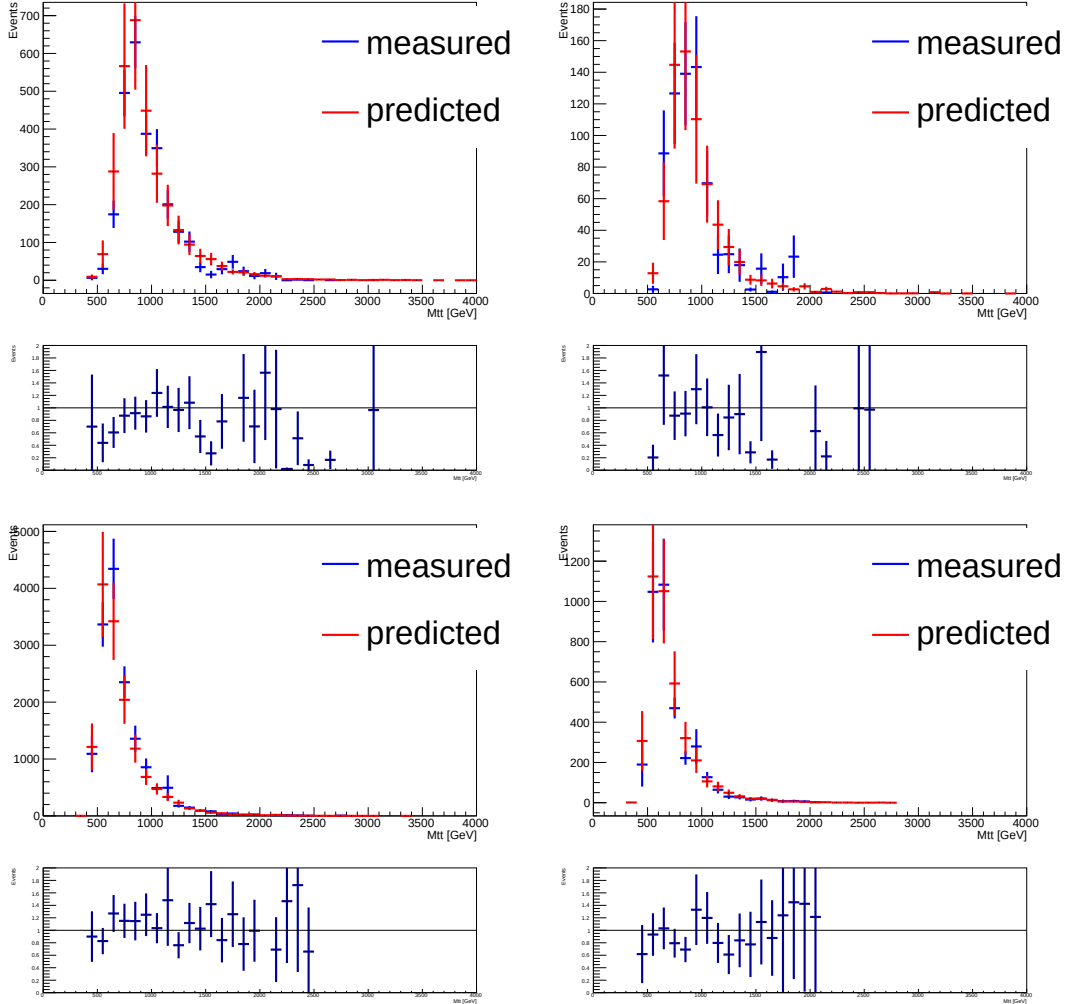


Figure 7.12: Closure test for the 0 b-tag high- H_T (top left), 1 b-tag high- H_T (top right), 0 b-tag low- H_T (bottom left), and 1 b-tag low- H_T (bottom right) event categories.

7.4.1 Kinematic distributions and event yields

Several kinematic distributions of the objects used in the analysis are shown in Figs. 7.13-7.18. In all of these plots the QCD multijet is estimated via the data driven procedure described in Sec. 7.4, and the $t\bar{t}$ background is estimated with simulated events. While the data-driven background estimation procedure is tailored to properly model QCD multijet background shape and normalization of the $m_{t\bar{t}}$ distribution, a generally good agreement between data and simulation is also observed for most of the kinematic variables shown. The gray error bands correspond to the statistical uncertainties alone.

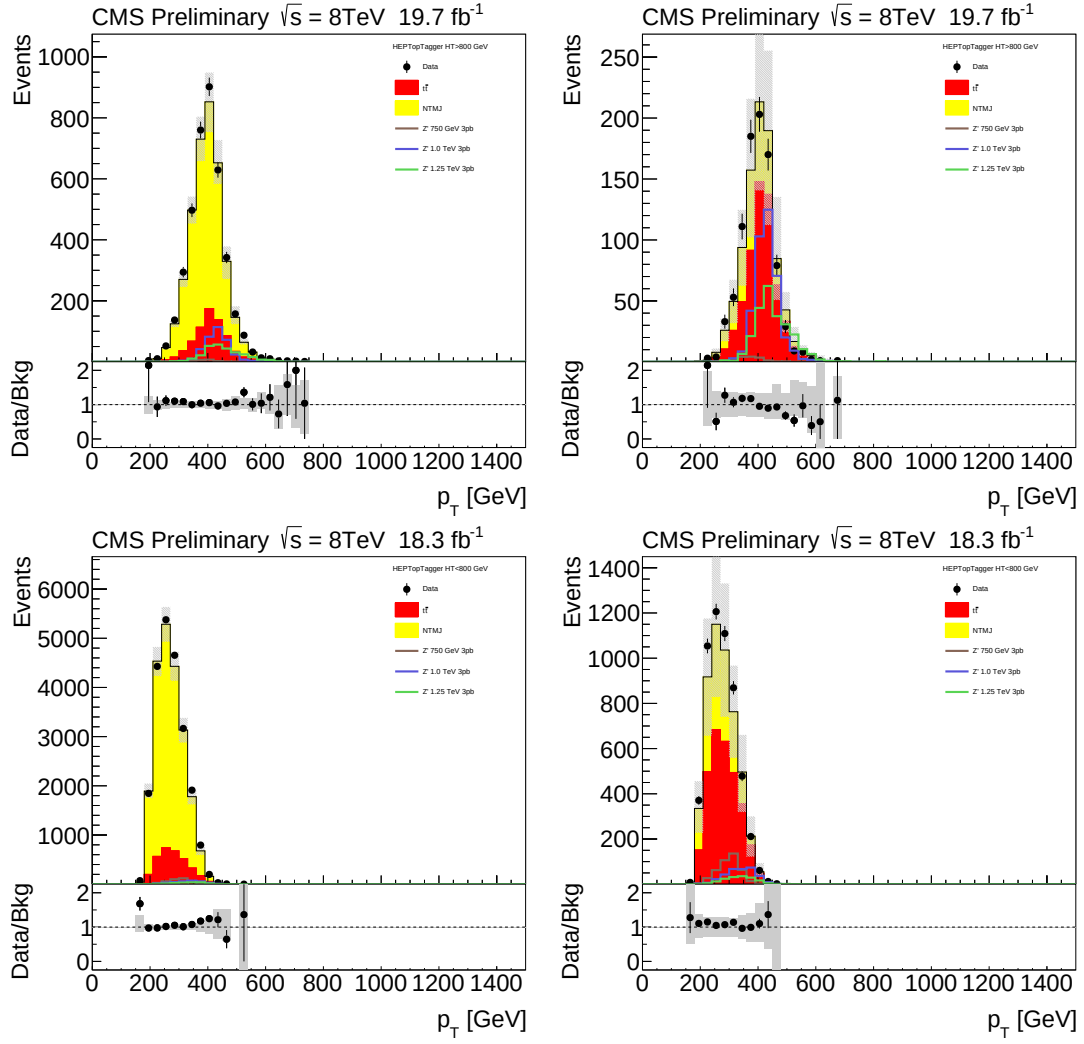


Figure 7.13: Leading top jet p_T distributions for the 1 b-tag $H_T > 800\text{ GeV}$ (top right), 2 b-tag $H_T > 800\text{ GeV}$ (bottom left), 1 b-tag $H_T < 800\text{ GeV}$ (bottom left), and 2 b-tag $H_T < 800\text{ GeV}$ (bottom right) categories.

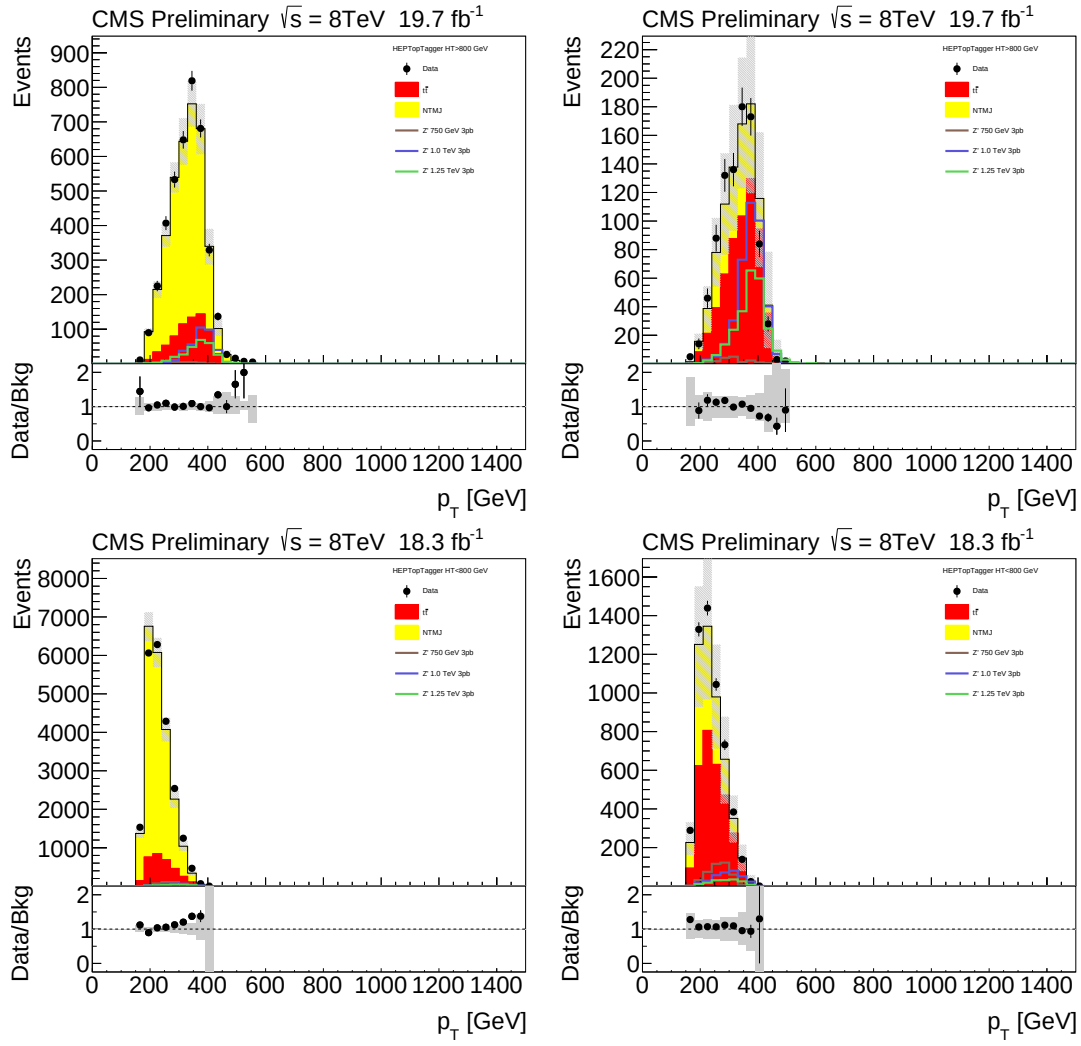


Figure 7.14: Distribution of the p_T of the second most energetic top jet for the 1 b-tag $H_T > 800$ GeV (top right), 2 b-tag $H_T > 800$ GeV (bottom left), 1 b-tag $H_T < 800$ GeV (bottom left), and 2 b-tag $H_T < 800$ GeV (bottom right) categories.

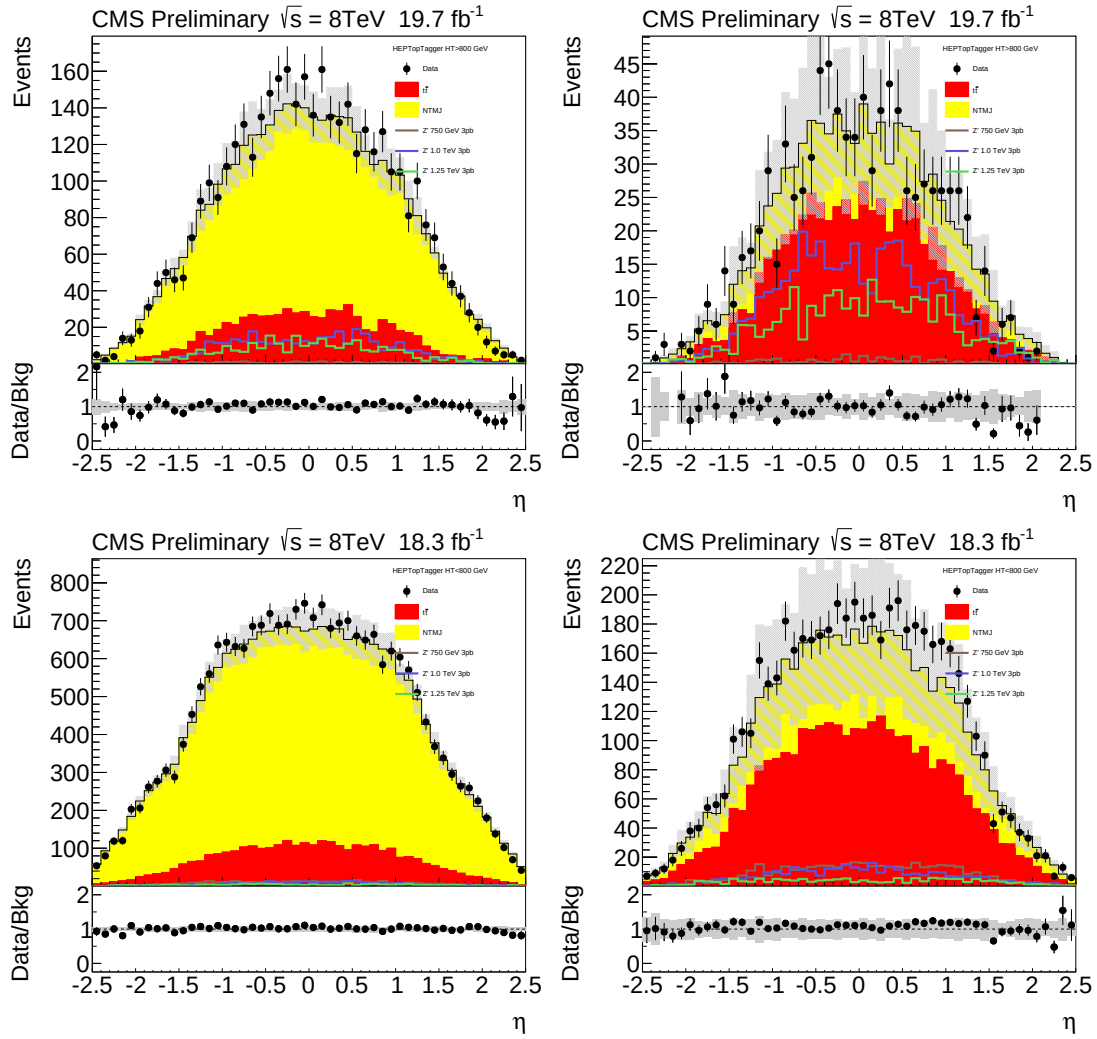


Figure 7.15: Leading top jet pseudorapidity distributions for the 1 b-tag $H_T > 800\text{ GeV}$ (top right), 2 b-tag $H_T > 800\text{ GeV}$ (bottom left), 1 b-tag $H_T < 800\text{ GeV}$ (bottom left), and 2 b-tag $H_T < 800\text{ GeV}$ (bottom right) categories.

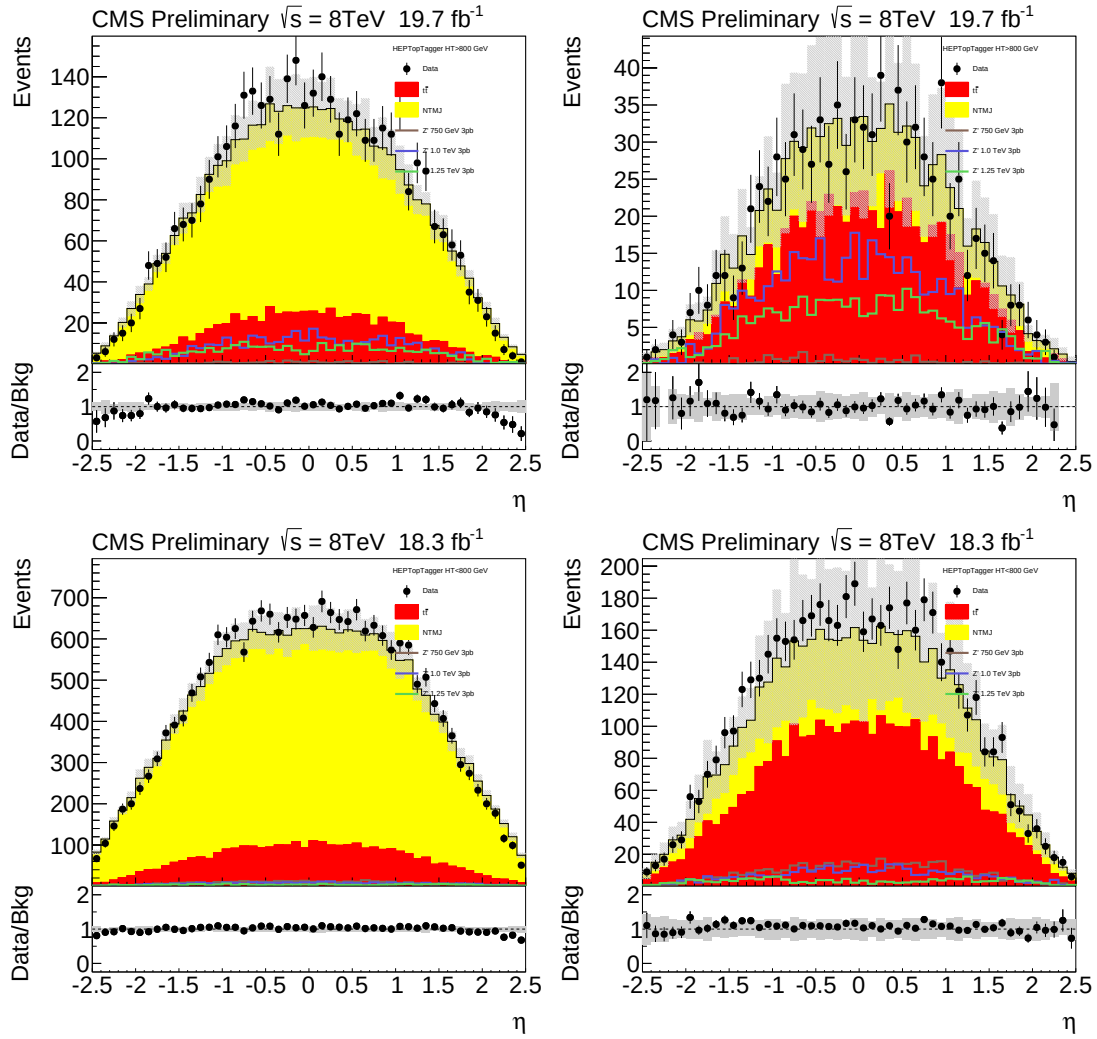


Figure 7.16: Pseudorapidity distribution of the second most energetic top jet for the 1 b-tag $H_T > 800$ GeV (top right), 2 b-tag $H_T > 800$ GeV (bottom left), 1 b-tag $H_T < 800$ GeV (bottom left), and 2 b-tag $H_T < 800$ GeV (bottom right) categories.

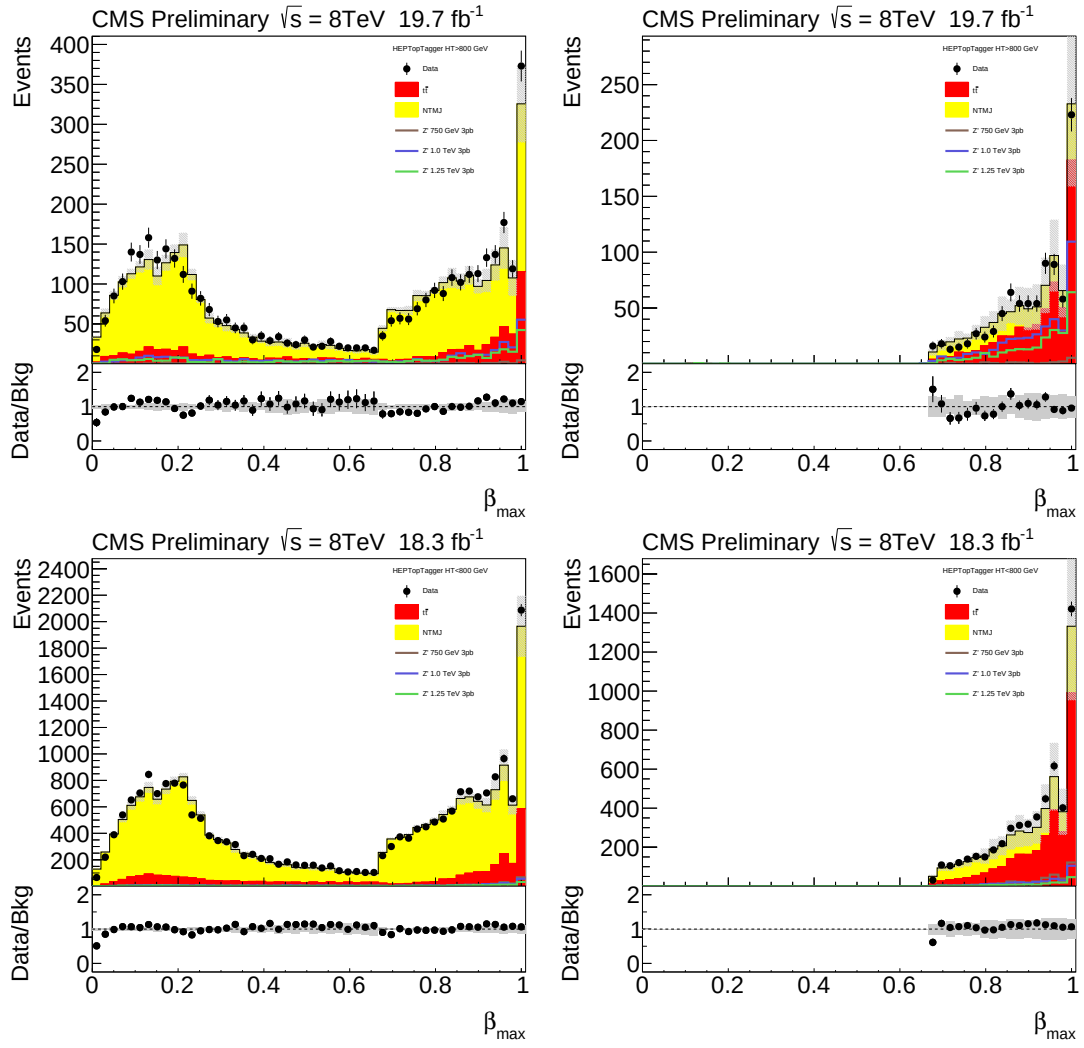


Figure 7.17: Leading top jet β_{max} distributions for the 1 b-tag $H_T > 800\text{ GeV}$ (top right), 2 b-tag $H_T > 800\text{ GeV}$ (bottom left), 1 b-tag $H_T < 800\text{ GeV}$ (bottom left), and 2 b-tag $H_T < 800\text{ GeV}$ (bottom right) categories.

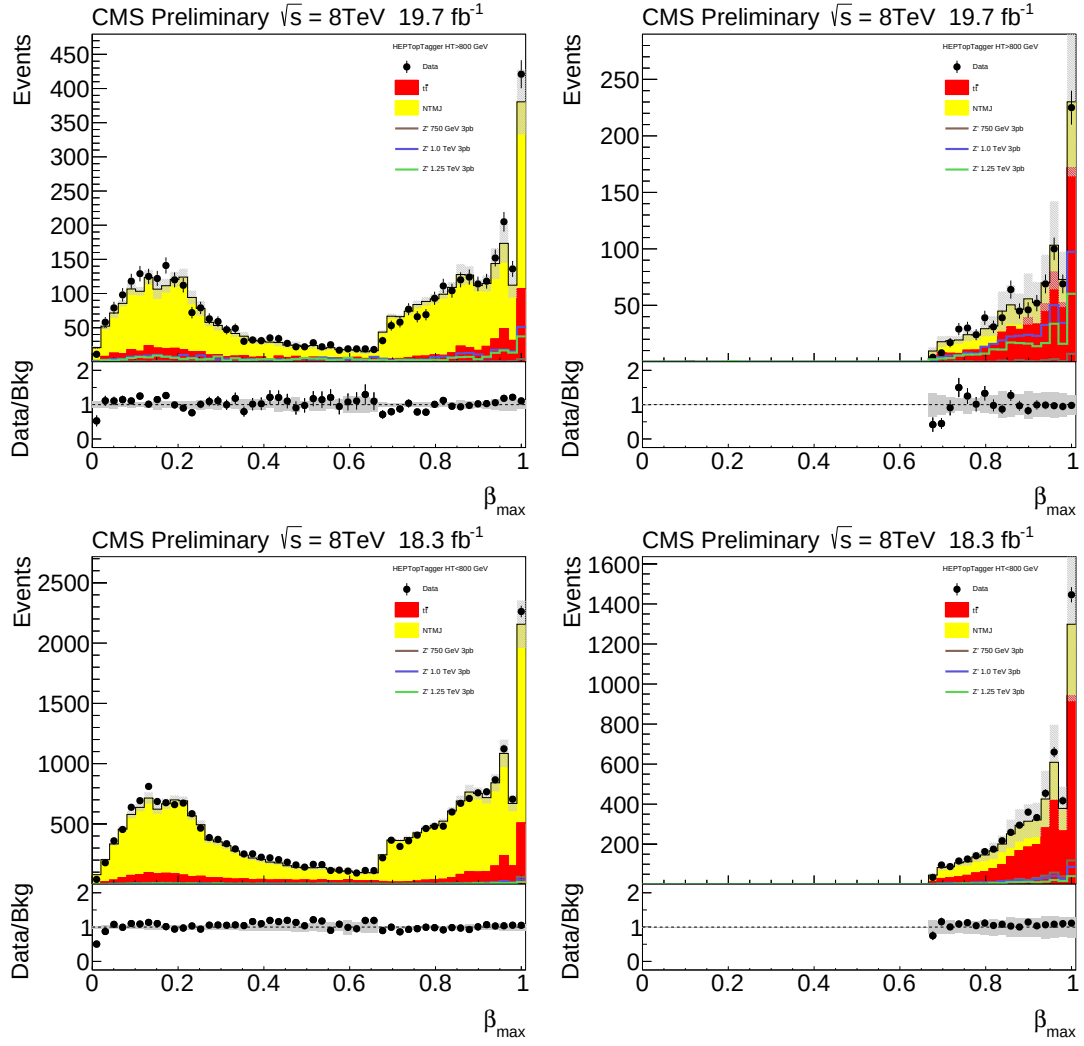


Figure 7.18: β_{max} distribution of the second most energetic top jet for the 1 b-tag $H_T > 800$ GeV (top right), 2 b-tag $H_T > 800$ GeV (bottom left), 1 b-tag $H_T < 800$ GeV (bottom left), and 2 b-tag $H_T < 800$ GeV (bottom right) categories.

	$H_T < 800 \text{ GeV}$			$H_T > 800 \text{ GeV}$		
	0 b-tag	1 b-tag	2 b-tags	0 b-tag	1 b-tag	2 b-tags
Efficiency						
Z' ($M = 0.75 \text{ TeV}, w = 1\%$)	0.17%	0.64%	0.70%	0.01%	0.04%	0.03%
Z' ($M = 1 \text{ TeV}, w = 1\%$)	0.13%	0.54%	0.56%	0.16%	0.61%	0.66%
Z' ($M = 2 \text{ TeV}, w = 1\%$)	0.04%	0.09%	0.07%	0.08%	0.26%	0.18%
Z' ($M = 0.75 \text{ TeV}, w = 10\%$)	0.15%	0.62%	0.64%	0.01%	0.06%	0.05%
Z' ($M = 1 \text{ TeV}, w = 10\%$)	0.12%	0.54%	0.54%	0.13%	0.49%	0.50%
Z' ($M = 2 \text{ TeV}, w = 10\%$)	0.04%	0.18%	0.15%	0.07%	0.27%	0.21%
g_{KK} ($M = 0.7 \text{ TeV}$)	0.11%	0.37%	0.42%	0.01%	0.04%	0.04%
g_{KK} ($M = 1 \text{ TeV}$)	0.11%	0.47%	0.49%	0.08%	0.40%	0.32%
g_{KK} ($M = 2 \text{ TeV}$)	0.06%	0.24%	0.19%	0.07%	0.23%	0.21%
Number of events						
$t\bar{t}$	851 ± 216	3238 ± 716	3009 ± 644	196 ± 65	698 ± 203	583 ± 165
QCD multijet	55932 ± 1598	18687 ± 613	1933 ± 92	8544 ± 331	3080 ± 168	311 ± 30
Total background	56781 ± 1633	21926 ± 984	4942 ± 655	8740 ± 341	3778 ± 268	894 ± 168
Data	57118	22485	5381	8920	3935	891

Table 7.3: Signal efficiency and event yields, in each signal region used in the analysis. The uncertainty includes the statistical uncertainty added in quadrature with the systematic uncertainties. The relative decay width $\Gamma_{Z'}/M_{Z'}$ is indicated by w .

The efficiency of the event selection is studied for the different signal models and individual mass points considered in the analysis. Table 7.3 summarizes the signal selection efficiencies in each of the different event categories. In the same table, the number of selected events for different signal hypotheses, SM backgrounds, and data is reported in each event category used in the analysis. The quoted uncertainties include the statistical and systematic uncertainties added in quadrature.

7.5 Systematic Uncertainties

Several sources of systematic uncertainty affect both the normalization of the event yields in the analysis, as well as the shapes of the $t\bar{t}$ invariant mass distribution for various processes. Systematic uncertainties are evaluated on both the simulated samples and the data-driven method used to estimate the QCD background contribution. This section will detail the individual contributions to the total systematic uncertainty of the analysis.

B-tagging is applied to subjects as discussed in Sec. 7.3. The agreement between simulation and measured efficiencies in data has been evaluated in Ref. [67] and correction factors have been derived [115]. The uncertainties on the correction factors are considered a source of systematic uncertainties. The scale factors for heavy flavors (b-quarks and c-quarks) are assumed to be fully correlated: the two scale factors are varied simultaneously $\pm 1\sigma$ when evaluating the systematics. The scale factors for light flavor subjects are varied independently with respect to the ones for heavy flavors and the uncertainty is added in quadrature to the latter. When the ΔR between the subjects is between 0.3 and 0.4 the uncertainties are doubled, due to the possibility of a large fraction of shared tracks between the different subjects.

The trigger efficiencies applied to simulated events are determined by running the trigger emulation on simulated events, and calculating the efficiency for events to pass the analysis

trigger selection, as a function of the H_T variable (for the H_T trigger) or the p_T of the fourth most energetic jet (for the four-jets trigger). The resulting trigger efficiency curve is shown in Figure 7.4. Trigger scale factors are derived from the ratio between the trigger efficiency curves in data and simulation. The scale factors are applied to the simulated samples and are considered as a source of systematic errors. The up/down variation is taken as $\pm|1 - SF|$, where SF is the scale factor.

The effect of the jet energy scale uncertainty is evaluated on simulated samples. This is done by varying the jet four-momentum up and down by the jet energy scale uncertainty [116], which is approximately 5%. The corrections are a function of the jet p_T and η . This results in a shape difference in the $m_{t\bar{t}}$ spectrum, which depends on the invariant mass itself. For low masses, the effect is large since the jets are close to the p_T selection threshold of 200 GeV. Varying the four-momenta causes jets to fall below the threshold, or be promoted above it – increasing or decreasing the acceptance. For higher invariant masses, the jets generally have higher p_T and this effect is smaller. This systematic uncertainty is applied to both the signal samples and the SM $t\bar{t}$ sample.

Differing jet energy resolutions between simulated events and data events are taken into account. The recommended η -dependent smearing is applied to jets, the errors on the smearing values are used to form systematic shape uncertainty templates. Additionally, it is assumed that the jet energy resolution for subjets is the same as that of standard jets. The variation of the subjet p_T is also included in these shape uncertainties. The efficiency of the HEPTOPTAGGER has been measured and compared to simulated to derive scale factors. The statistical uncertainties of the SF are taken into account.

Additional uncertainties applied only to SM $t\bar{t}$ events include two contributions: a rate-only component due to the uncertainty in the $t\bar{t}$ cross section, which is assumed to be 15%, and a second shape component for uncertainties in the factorization and normalization scale (Q^2). The shape uncertainties are implemented by generating additional $t\bar{t}$ samples with twice or one-half the nominal Q^2 scale used for the generation of default simulated events. To form the shape templates for the up/down variation, the larger of the difference between the nominal bin content and the varied bin content is taken and symmetrically applied to the bin for both the upward and downward variations of the uncertainty.

To measure the uncertainty on the method used to derive the QCD multijet background, the statistical uncertainty of the individual bins in the mistag matrix is used. For each bin in the $m_{t\bar{t}}$ distribution, there are contributions from several bins in the mistag matrix itself. Contributions within a given bin should have correlated uncertainties, while contributions coming from two independent bins in the mistag matrix can be uncorrelated. To compute the uncertainties for each bin in the $m_{t\bar{t}}$ distribution, the errors are propagated from each contributing bin in the mistag matrix, either correlating or uncorrelating them when summing over the errors from each individual bin in the mistag matrix. Because the mistag matrix has several bins which are sparsely populated with just a handful of entries, there is a relatively large error on the QCD estimate for this analysis. These sparsely-populated bins dominate the uncertainty. However, during the cross-section limit-setting procedure, this uncertainty is strongly constrained and so it does not significantly affect the sensitivity of the analysis.

The QCD multijet background is estimated by dividing the data into bins along three variables: jet p_T , β , and the variable of the distribution α (in this case $M_{t\bar{t}}$). The background yield for this three-dimensional bin is obtained by multiplying the events from a single top-tag sample in this particular bin, $N(M_{t\bar{t}}, p_T, \beta) = N_{\alpha, i, j}$ by the mistag rate $r(p_T, \beta) = r_{i, j}$.

Summing over the indices i , and j gives the total prediction for the bin α :

$$N_\alpha = \sum_{a=1}^{N_{\text{jets}}} N_{\alpha,i,j} r_{i,j},$$

where N_{jets} is the number of jets in the events in bin α . The uncertainty in each bin of the predicted mistagged distribution $\sigma(m_\alpha)$ has two parts: one arises from the uncertainty on the mistag matrix ($\sigma(r_{i,j})$), and the other from the number of jets in the three-dimensional bin ($\sqrt{N_{\alpha,i,j}}$):

$$\sigma(m_\alpha) = \sqrt{\sum_{a=1}^{N_{\text{jets}}} \left((N_{\alpha,i,j} \sigma(r_{i,j}))^2 + (\sqrt{N_{\alpha,i,j}} r_{i,j})^2 \right)},$$

The first term accounts for all the uncertainties in the i, j -th bin of the mistag matrix. The second term accounts for the statistical uncertainties in the jet sample. Each of these terms are individually added linearly when summing over bins, since they are fully correlated within each bin. While the two terms are added in quadrature since they are fully uncorrelated.

As the orthogonal medium-boost and high-boost channels using the HEPTOPTAGGER and the CMS Top Tagger are meant to be combined, the appropriate correlation between the systematic uncertainties of the two channel are taken into account during the statistical interpretation of the results. The various systematics are either fully correlated or fully uncorrelated, depending on the source of the uncertainty in question. Table 7.4 reports the list of the systematic uncertainties applied to the two channels, and their correlation. The shape templates for the main systematic uncertainties can be seen in Appendix B.

Systematic uncertainty	CMS Top Tagger	HEPTOPTAGGER
B-tagging scale factor uncertainty	✓	✓
CMS Top Tagger scale factor uncertainty	✓	
HEPTOPTAGGER scale factor uncertainty		✓
Jet energy corrections	☑	☑
Jet energy resolution	☑	☑
Luminosity	☑	☑
Mistag matrix uncertainty	✓	✓
PDF uncertainty	☑	☑
$t\bar{t}$ Q^2 uncertainty	☑	☑
$t\bar{t}$ p_T reweight uncertainty	☑	☑
$t\bar{t}$ cross-section	☑	☑
Trigger efficiency	✓	✓

Table 7.4: Summary of systematic uncertainties applied to the sub-channels of the analysis. Correlations between the channels are imposed when the symbol ☑ is shown.

7.6 Summary of the very-boosted decay category

This section presents a summary of the high-boost channel of the analysis, using the CMS Top Tagger algorithm. The selection criteria and the procedure for the estimation of the SM background processes are very similar to the ones used for the medium-boost analysis using the HEPTOPTAGGER. This channel select events using the H_T trigger and requiring two

CA8 jets with a transverse momentum greater than 400 GeV. On the two most energetic jets the CMS Top Tagger algorithm is applied in combination with an N-subjettiness selection ($\tau_3/\tau_2 < 0.7$). The events are categorized in function of the number of b-tagged jets and of the difference in rapidity between the two jets $|\Delta y|$. Figure 7.19 shows a schematic overview of the selection of the analysis. Figure 7.20 shows the distribution of the invariant mass of the $t\bar{t}$ system for the two most relevant event categories in terms of signal enrichment.

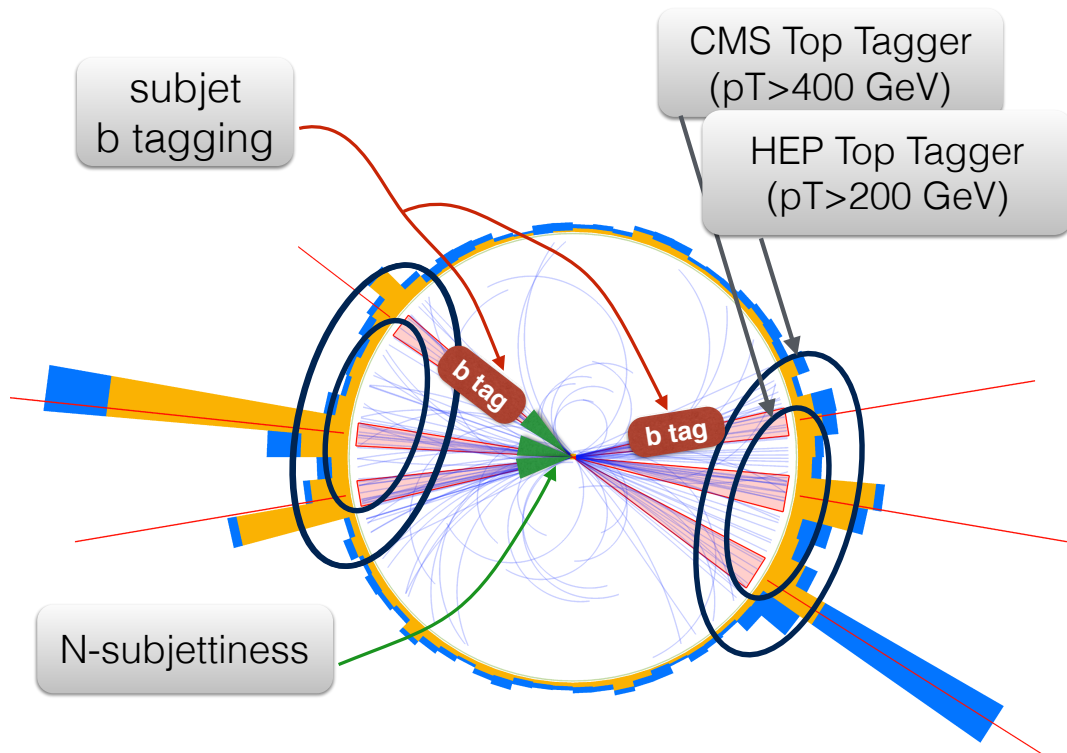


Figure 7.19: A schematic overview of the selection of the analysis.

Given that the selection of the CMS Top Tagger analysis uses the N-subjettiness variable the mistag matrix must be computed in different bins of this variable. The mistag matrix is shown in Figure 7.21, along with the closure test for the background estimation procedure for one of the event categories.

7.7 Statistical interpretation and results

Figures 7.22 and 7.23 show the final distribution of the invariant mass of the $t\bar{t}$ system for all the event categories considered in this analysis. The QCD multijet background has been estimated from simulation and all the systematic and statistical uncertainties considered are shown.

No excess of events is observed above the expected background from SM $t\bar{t}$ and QCD multijet. Therefore limits are set on the production cross sections of the three hypothesized signal models. A template-based shape analysis is performed, using the distributions of the candidate $t\bar{t}$ invariant mass. A Bayesian likelihood-based method is used, allowing the expected background model to fluctuate within the various systematic and statistical uncertainties to find the best fit to the observed data distribution.

The THETA software package [117] is used to produce 95% confidence-level (CL) upper

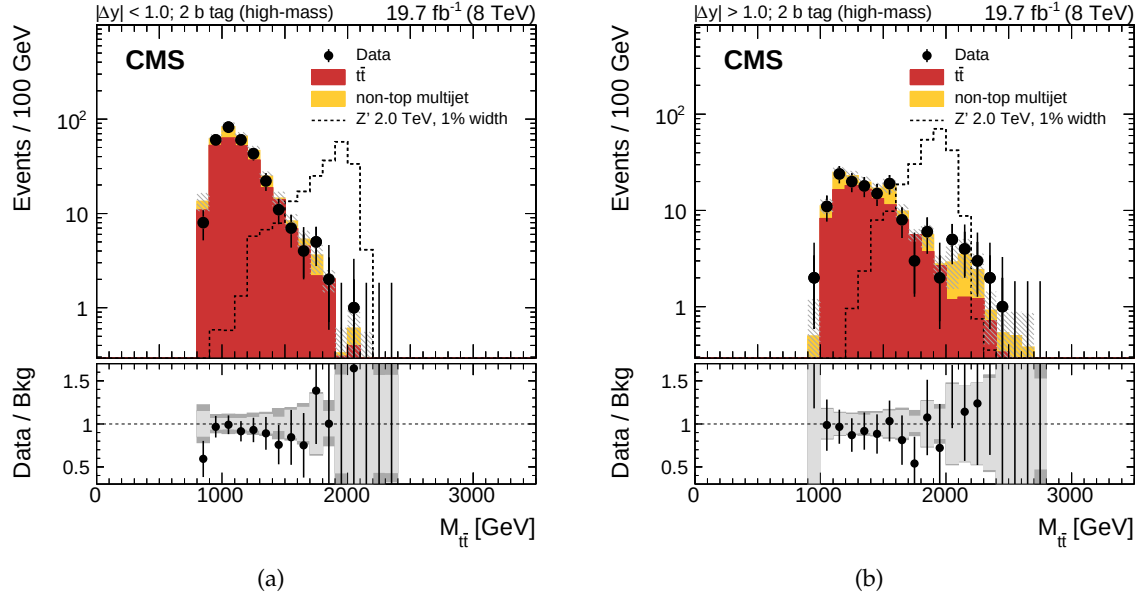


Figure 7.20: Distribution of the invariant mass of the $t\bar{t}$ system for the CMS Top Tagger analysis, 2 b-tags category, for $|\Delta y| < 1.0$ (a), and $|\Delta y| > 1.0$ (b) [108].

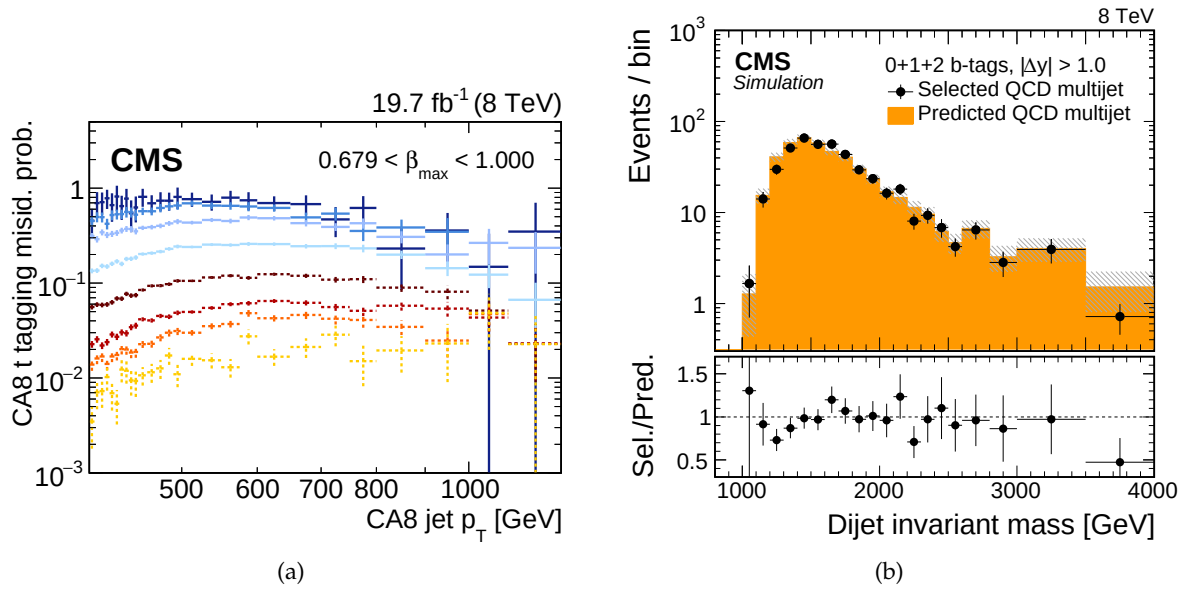


Figure 7.21: Left panel: top tagging misidentification rate distribution for the CMS Top Tagger category, $0.679 < \beta_{\max} < 1$. Right panel: closure test for the QCD multijet background estimation method in the CMS Top Tagger category, $|\Delta y| > 1.0$ [108].

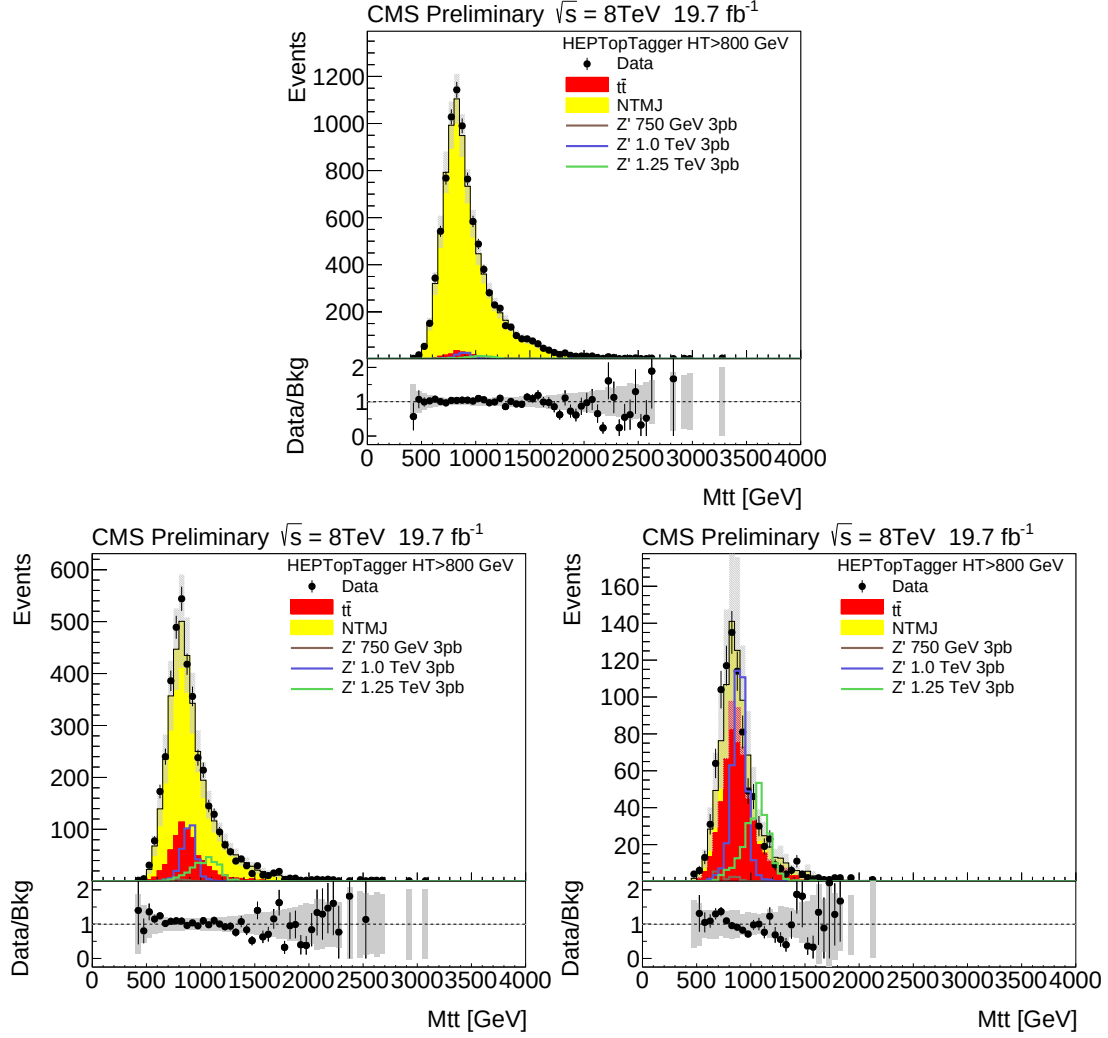


Figure 7.22: $m_{t\bar{t}}$ distributions for 0 (top), 1 (bottom left), and 2 (bottom right) b-tags in the high- H_T category.

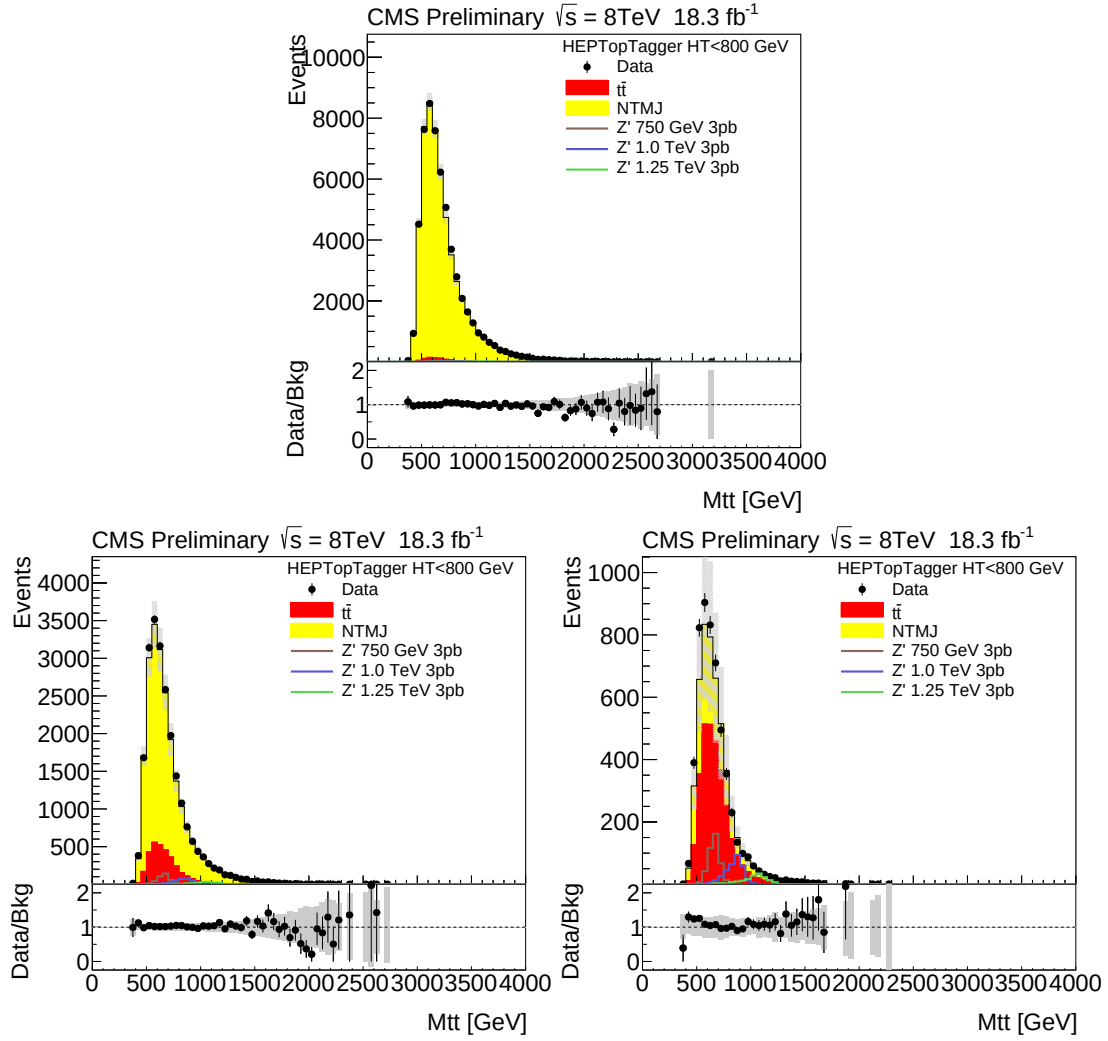


Figure 7.23: $m_{t\bar{t}}$ distributions for 0 (top), 1 (bottom left), and 2 (bottom right) b-tags in the low- H_T category.

limits on the cross sections for the three signal models, as a function of the hypothetical heavy particle mass. A Poisson model is used for each bin in the $m_{t\bar{t}}$ distribution, with the final likelihood composed by the product of these Poisson probabilities for each bin in the $m_{t\bar{t}}$ distribution. For a single bin, the mean of this distribution is shown in Equation 7.3.

$$\mu_i = \sum_k \beta_k \cdot T_{k,i} \quad (7.3)$$

Where the index k represents each physics process contributing to the bin i . T represents the number of expected events for the process k in bin i . The factor β is the Poisson mean for the individual process k . With this per-bin formula, the full likelihood can be formed as follows:

$$\mathcal{L}(\beta_k) = \prod_i \frac{\mu_i \cdot e^{-\mu_i}}{N_i^{\text{data}}!}. \quad (7.4)$$

For the Poisson parameters, β_k , different choices of prior distributions are used. A flat prior distribution is used for the QCD estimate and the signal cross section normalization parameters. For the other nuisance parameters, a lognormal prior distribution is used. Using these parameters, pseudoexperiments are performed to estimate the 68% and 95% CL (1- and 2- σ) expected limit bands. In the pseudoexperiments, the systematic and statistical uncertainties are accounted for through nuisance parameters. These nuisance parameters are randomly varied within their uncertainties, and the posterior is re-fitted for each individual pseudoexperiment. Through this method, the effect of the shape and normalization uncertainties on the sensitivity of the analysis can be evaluated.

The expected cross section limits obtained are shown in Table 7.5, using the combination of the 6 independent signal regions of the analysis. As the HEPTOPTAGGER analysis is using only events not selected by the CMS Top Tagger analysis, the sensitivity on the analysis in the high-mass region is greatly reduced.

7.7.1 Combination of HEPTopTagger and CMS Top Tagger analyses

A statistical combination of the analyses using the HEPTOPTAGGER and CMS Top Tagger is performed to improve the sensitivity relative to each individual analysis. As the HEPTOPTAGGER analysis is optimized for the medium-mass range, and the CMS Top Tagger analysis is optimized for the high-mass range, it is expected that the combination will gain sensitivity in the crossover region between the two analyses. This region lies in the range of masses between 1 TeV to 1.5 TeV. Below this range, the HEPTOPTAGGER is most sensitive, while above, the CMS Top Tagger analysis is most sensitive.

Figure 7.24 shows the expected and observed cross section limits on the production of a heavy resonance decaying to $t\bar{t}$, for the combination of the two analyses. The results are reported numerically in Table 7.6.

7.8 Combination with leptonic final states

The two analyses probing the all-jets final states are combined with searches for resonant $t\bar{t}$ production in final states containing one or more leptons [108]. More precisely, the combination includes an analysis requiring two leptons (dilepton analysis), therefore probing the final state where both of the top quarks decay leptonically, and an analysis requiring one lepton, probing the final state where one top quark decays hadronically and the other decays leptonically. The single-lepton analysis is further split in a “boosted” analysis using the CMS

Narrow Z' signal hypothesis

Mass (GeV)	Observed 95% CL limit (pb)	Expected 95% CL limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
750	1.650	0.873	1.202	1.704	2.496	3.330
1000	0.464	0.377	0.507	0.721	1.017	1.392
1250	0.916	0.298	0.412	0.601	0.875	1.228
1500	0.507	0.250	0.350	0.508	0.707	0.997
2000	0.366	0.197	0.269	0.384	0.586	0.920
3000	0.322	0.183	0.258	0.419	0.728	3.073

Wide Z' signal hypothesis

Mass (GeV)	Observed 95% CL limit (pb)	Expected 95% CL limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
750	2.444	1.410	1.943	2.758	4.088	5.596
1000	0.726	0.582	0.798	1.120	1.616	2.290
1250	1.071	0.442	0.588	0.823	1.187	1.700
1500	0.801	0.372	0.524	0.754	1.135	1.650
2000	0.720	0.309	0.432	0.618	0.907	1.353
3000	1.012	0.542	0.762	1.229	4.872	28.265

Randall–Sundrum gluon signal hypothesis

Mass (GeV)	Observed 95% CL limit (pb)	Expected 95% CL limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
700	5.602	2.662	3.772	5.578	8.404	12.458
1000	1.514	1.033	1.459	2.159	3.329	5.434
1200	1.874	0.879	1.220	1.843	2.932	4.816
1500	2.398	0.760	1.050	1.628	2.800	5.266
1800	1.767	0.729	1.028	1.777	4.474	25.609
2000	2.272	0.784	1.165	2.289	11.736	29.363
2500	1.297	0.799	1.202	3.325	23.210	34.248
3000	2.144	1.156	2.316	15.298	31.957	43.627

Table 7.5: Upper cross section limits on the production of a heavy resonance decaying to $t\bar{t}$, presented numerically for each of the three signal hypotheses considered in this analysis.

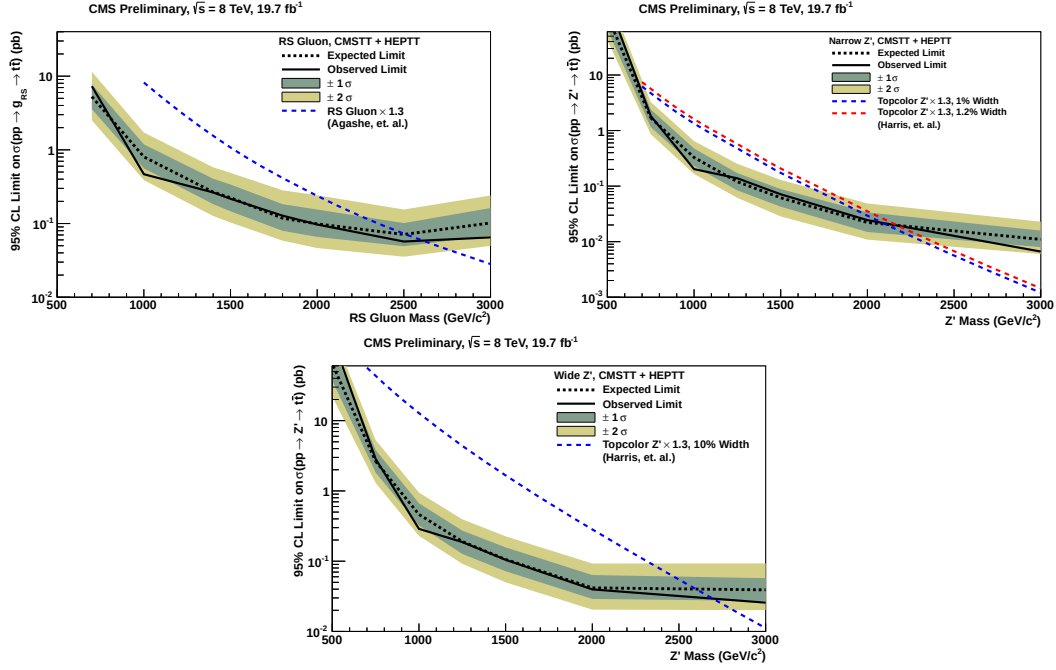


Figure 7.24: Expected and observed limits on the cross section for Randal–Sundrum gluon production (top left), narrow Z' production (top right), and wide Z' production (bottom). These results are presented for the combination of the CMS Top Tagger and HEPTOP Tagger analyses. The dashed black line represents the median expected limit, while the green and yellow bands represent the 1- and 2- σ bands, respectively, of the expected limits from the set of pseudoexperiments. The solid line shows the observed limit results using the observed data distribution.

Combined CMS Top Tagger +HEPTopTagger analyses, narrow Z' signal hypothesis						
Mass (TeV)	Observed 95% CL limit (pb)	Expected 95% CL limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
0.50	221.820	42.800	66.797	117.600	187.560	301.350
0.75	1.836	0.861	1.152	1.696	2.407	3.270
1.00	0.202	0.166	0.230	0.328	0.476	0.645
1.25	0.138	0.062	0.084	0.120	0.171	0.252
1.50	0.071	0.028	0.043	0.062	0.088	0.129
2.00	0.024	0.011	0.015	0.022	0.034	0.049
3.00	0.007	0.006	0.008	0.011	0.016	0.023

Combined CMS Top Tagger +HEPTopTagger analyses, wide Z' signal hypothesis						
Mass (TeV)	Observed 95% CL limit (pb)	Expected 95% CL limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
0.50	150.230	22.028	37.148	64.179	113.250	187.980
0.75	2.857	1.300	1.829	2.642	3.819	5.295
1.00	0.288	0.229	0.323	0.463	0.667	0.946
1.25	0.187	0.092	0.127	0.191	0.272	0.395
1.50	0.105	0.050	0.072	0.106	0.157	0.226
2.00	0.040	0.020	0.029	0.042	0.063	0.093
3.00	0.026	0.020	0.027	0.039	0.057	0.093

Combined CMS Top Tagger +HEPTopTagger analyses, Randall–Sundrum gluon signal hypothesis

Mass (TeV)	Observed 95% CL limit (pb)	Expected 95% CL limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
0.70	7.285	2.515	3.573	5.264	7.716	11.472
1.00	0.467	0.388	0.561	0.803	1.188	1.719
1.40	0.264	0.127	0.184	0.267	0.405	0.579
1.50	0.217	0.103	0.147	0.224	0.338	0.487
1.80	0.128	0.059	0.080	0.118	0.183	0.282
2.00	0.097	0.046	0.066	0.099	0.155	0.242
2.50	0.057	0.035	0.049	0.071	0.101	0.155
3.00	0.065	0.050	0.069	0.102	0.162	0.240

Table 7.6: Upper cross section limits on the production of a heavy resonance decaying to $t\bar{t}$, presented numerically for each of the three signal hypotheses, for the combined CMS Top Tagger +HEPTopTagger analyses.

Top Tagger to reconstruct the decay of the hadronic top quark, thus targeting high-mass resonances, and a “threshold” analysis assuming that the hadronic decay of the top quark is not merged in a single jet, optimized for low-mass resonances. All the leptonic analyses consider both the electrons and muons as lepton candidates. Figure 7.25 shows a graphical overview of the analyses combined.

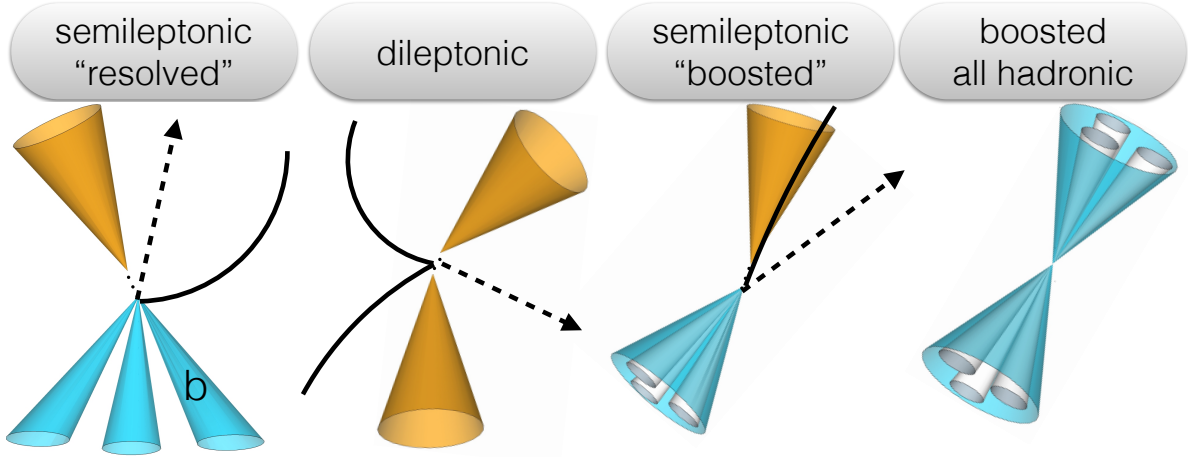


Figure 7.25: A graphical overview of the analyses combined.

As in the all-jets analysis, top-quark candidates are reconstructed and used to compute the $t\bar{t}$ invariant mass. For the single-lepton plus jet boosted analysis, the events are categorized by the type of lepton, electron or muon. For each lepton flavor, the events are further categorized in events with a top-tagged jet, events without a top-tagged jet but at least one b-tagged jet, and events with neither a top tag, nor a b tag. Distributions of invariant mass of the $t\bar{t}$ system are shown for the most sensitive categories of the single-lepton plus jet analysis in Fig. 7.26.

A comparison of the expected cross section limits obtained from the individual channels and their combination is shown in Fig. 7.27 for the signal hypotheses of a Randall–Sundrum gluon and a narrow Z' resonance. The boosted lepton+jet analysis is the most sensitive analysis, followed by the combined all-jets analysis.

Figure 7.28 shows upper limits cross section limits at 95% confidence-level on the production of a Z' boson decaying to $t\bar{t}$ with 1% or 10% relative width and a Randall–Sundrum gluon in the same final state. The vertical dashed line indicates the transition from a threshold analysis [118] to the combination. Left of the vertical dashed line, only the results of the low-mass “threshold” analysis are reported, while on the right the results from the combination of the channels optimized for high resonance masses are shown. The limits are compared to predictions for the cross section of a topcolor Z' boson with 1.2% and 10% width [111] and the prediction for the production of a Randal–Sundrum gluon [43]. The predictions are multiplied by a factor of 1.3 to account for the effect higher-order corrections in the cross section [119]. The numerical quantities relative to Figure 7.28 are reported in Table 7.7.

From the comparison between the expected and observed cross section limits and the theoretical prediction for the cross section of the heavy resonance in the different models considered, lower expected and observed mass limits can be extracted. Lower mass limits are shown in Table 7.8 for the different analyses separately and for the combination.

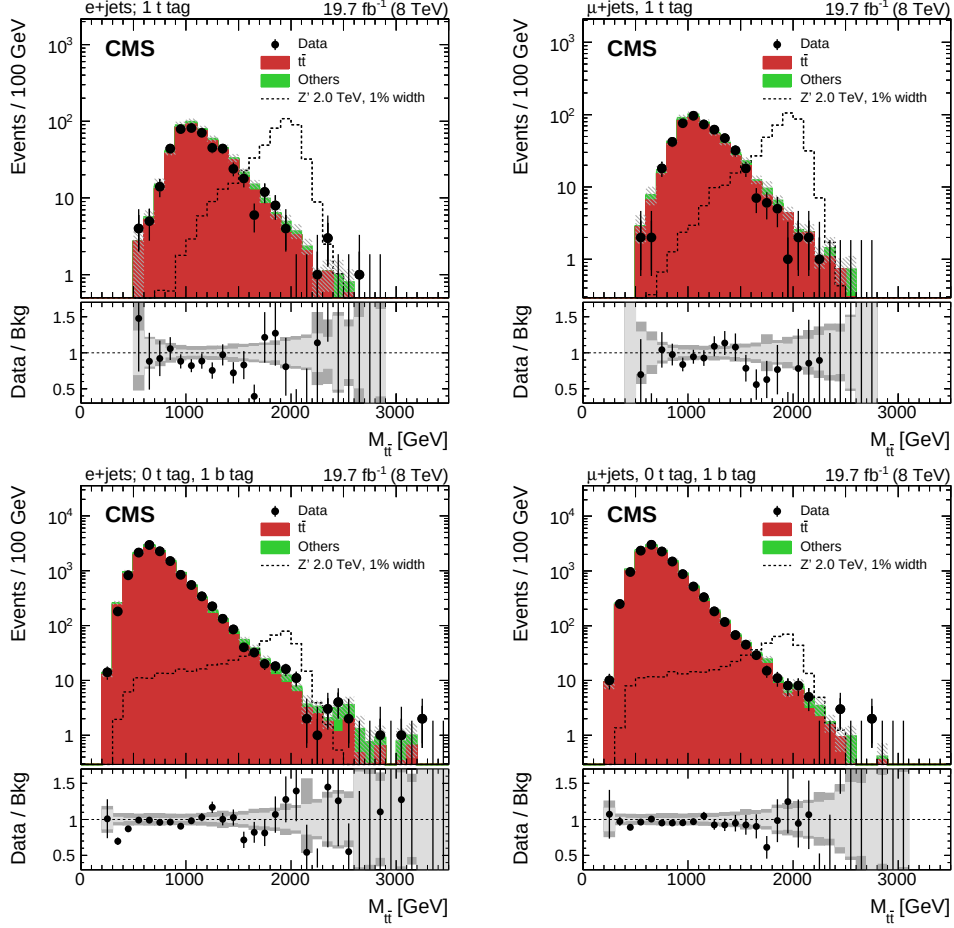


Figure 7.26: Invariant mass of the $t\bar{t}$ system in data and simulation in the electron+jets (left column) and muon+jets (right column) channels: events with a top-tagged jet in the top row, and events without a top-tagged jet but at least one b-tagged jet in the bottom row. The signal distribution is normalized to a cross section of 1 pb. The uncertainty associated with the background expectation includes all statistical and systematic uncertainties. For the ratio plot, the statistical uncertainty is shown in light gray, while the total uncertainty is shown in dark gray [108].

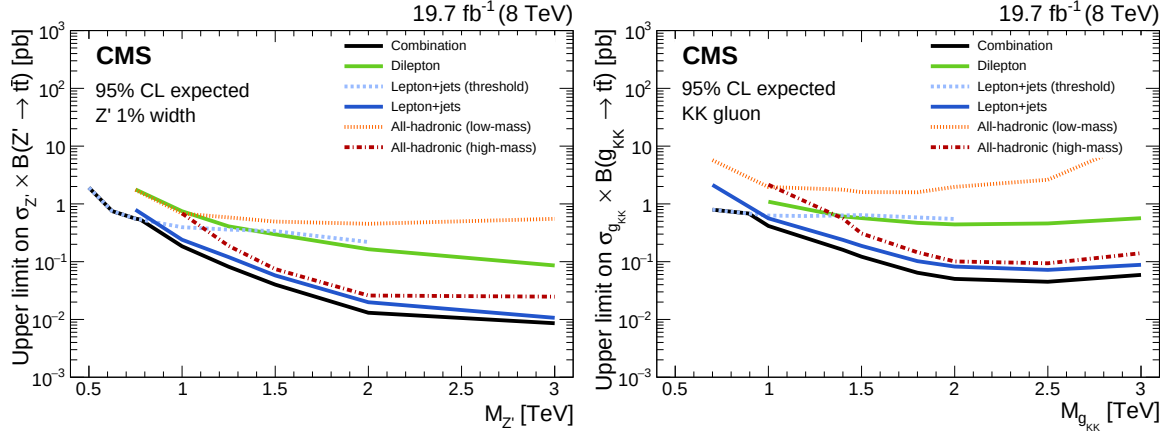


Figure 7.27: Expected 95% CL upper limits on the production cross section for a Z' boson decaying to $t\bar{t}$ with 1% width (left) and a KK gluon in the Randall–Sundrum model (right). The limits obtained from the individual channels are shown separately, together with the result from the combination [108].

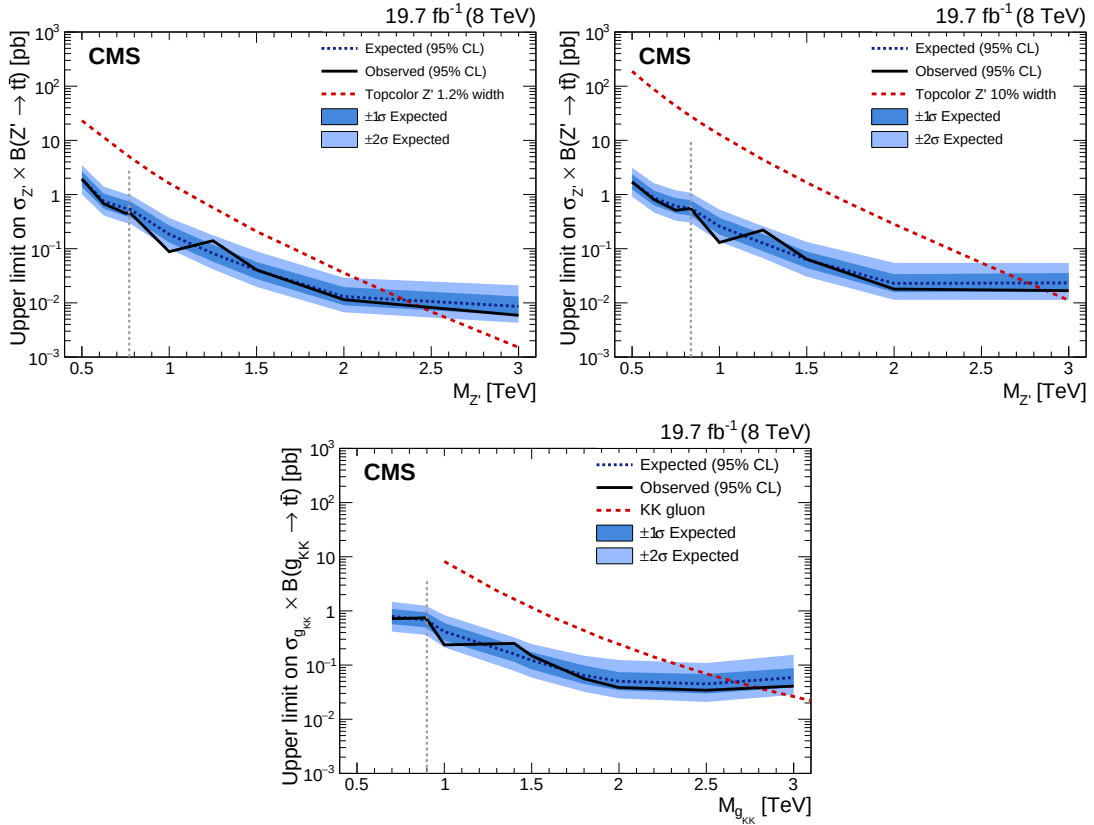


Figure 7.28: Upper limits at 95% CL on the production cross section times branching fraction for a Z' boson with 1% width (upper left), with 10% width (upper right), and a Randall–Sundrum gluon decaying to $t\bar{t}$ (bottom). The limits are compared to the theory predictions [108].

$Z', \Gamma_{Z'}/M_{Z'} = 1\%$							
$M_{Z'}$ (TeV)	Expected (pb)	Expected range ($\pm 1\sigma$) (pb)		Expected range ($\pm 2\sigma$) (pb)		Observed (pb)	
0.75	0.61	0.89	— 0.43	1.3	— 0.32	0.86	
1.0	0.18	0.27	— 0.13	0.37	— 0.099	0.088	
1.25	0.082	0.12	— 0.058	0.18	— 0.042	0.14	
1.5	0.04	0.057	— 0.028	0.089	— 0.02	0.041	
2.0	0.013	0.02	— 0.009	0.029	— 0.0067	0.011	
3.0	0.0086	0.013	— 0.0059	0.021	— 0.0043	0.0059	
$Z', \Gamma_{Z'}/M_{Z'} = 10\%$							
$M_{Z'}$ (TeV)	Expected (pb)	Expected range ($\pm 1\sigma$) (pb)		Expected range ($\pm 2\sigma$) (pb)		Observed (pb)	
0.75	0.83	1.2	— 0.57	1.8	— 0.42	0.89	
1.0	0.26	0.37	— 0.18	0.53	— 0.14	0.13	
1.25	0.13	0.19	— 0.09	0.26	— 0.067	0.22	
1.5	0.063	0.089	— 0.044	0.13	— 0.03	0.064	
2.0	0.023	0.034	— 0.016	0.055	— 0.011	0.018	
3.0	0.023	0.036	— 0.016	0.055	— 0.011	0.017	
RS KK gluon							
$M_{g_{KK}}$ (TeV)	Expected (pb)	Expected range ($\pm 1\sigma$) (pb)		Expected range ($\pm 2\sigma$) (pb)		Observed (pb)	
0.7	1.7	2.5	— 1.2	3.8	— 0.84	3.5	
1.0	0.42	0.6	— 0.28	0.84	— 0.21	0.24	
1.4	0.16	0.23	— 0.11	0.32	— 0.078	0.25	
1.5	0.12	0.17	— 0.083	0.24	— 0.059	0.15	
1.8	0.064	0.098	— 0.045	0.15	— 0.032	0.056	
2.0	0.05	0.074	— 0.034	0.12	— 0.024	0.038	
2.5	0.045	0.068	— 0.03	0.11	— 0.021	0.034	
3.0	0.059	0.088	— 0.039	0.15	— 0.028	0.041	

Table 7.7: Expected and observed limits for the production cross section of a Z' boson decaying to $t\bar{t}$ with a width of 1%, 10% and a Randall–Sundrum gluon. Limits are reported at 95% confidence-level [108].

Mass limit [TeV]								
	Dilepton channel		Lepton+jets channel		All-hadronic channels		Combined	
	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
$Z', \Gamma_{Z'}/M_{Z'} = 1.2\%$	1.4	1.5	2.2	2.3	2.1	2.1	2.4	2.4
$Z', \Gamma_{Z'}/M_{Z'} = 10\%$	2.1	2.2	2.7	2.8	2.5	2.5	2.8	2.9
RS KK gluon	1.8	2.0	2.5	2.5	2.4	2.3	2.7	2.8

Table 7.8: Expected and observed mass limits for the benchmark models. Mass limits are given separately for the two-leptons plus jets analysis, lepton plus jets analysis, the combined all-jets analysis, and the full combination of the analyses. Limits are given at 95% confidence-level [108].

7.9 Conclusions

A search for resonant $t\bar{t}$ production in the all-jets final state with 8 TeV data is described. The search targets the medium-boost region of the top quark. The HEPTopTAgGER algorithm along with subjet b tagging is used to reconstruct the decay of the top quarks. The analysis complements a search targeting highly boosted top decays and using the CMS Top Tagger algorithm. The two searches targeting the all-hadronic final state of the $t\bar{t}$ system are combined with searches in the dilepton and lepton+jets final states.

Thanks to the introduction of advanced techniques such as the combination of different jet substructure variables, the two all-jets analyses reach a sensitivity comparable to that of the lepton+jet analysis. Therefore, the combination of all the final states yields an improvement in the sensitivity with respect to the single analyses.

No excess over the SM background is observed in all the event categories considered, and limits are derived on the production of heavy resonances decaying to $t\bar{t}$. The cross section limits are compared to the predicted cross section of a narrow-width (1.2%) topcolor Z' , a large width (10%) topcolor Z' , and a Randal–Sundrum KK gluon, allowing to exclude these resonances with masses below 2.4, 2.9, and 2.8 TeV, respectively. This result represents the most stringent limit on the production of such resonances in the 8 TeV data-taking era.

Search for heavy resonances decaying to a top quark and a vector-like T quark with 13 TeV data

8.1 Introduction

As explained in Chapters 1, many models of new physics predict heavy spin-1 resonances (Z') which have enhanced couplings to the third generation of the Standard Model. However, these models often predict the existence of heavy vector-like quarks along with the heavy resonances. In such cases the decay modes of the spin-1 particle to vector-like quarks or a combination of vector-like quarks and third generation quarks can be enhanced as well.

Traditionally, heavy resonance searches focus on decay modes to SM particles, as the one described in Chapter 7. Depending on the coupling parameters and on the masses of the particles, the decay modes containing at least a vector-like quark might be the dominant ones. Hence, the need to explore decay modes of heavy resonances to non-SM particles with dedicated searches. These features are often included in theories with an extra dimension and composite Higgs models, as described in Chapters 1.

This chapter presents the first search for a neutral spin-1 heavy resonance, referred to as Z' , decaying to a top and a vector-like T quark with charge $2/3e$. This search focuses on the kinematic range in which Z' boson decays to a top quark and a T quark dominate over those to TT . Figure 8.1 shows a scheme of the relevant decay modes of the resonance in function of the T quark mass. Many VLQ models predict predominant decays of the T quark to third generation SM quarks and the three main decay modes: tH , tZ , and bW [20]. The search focuses on the decay of the T quark to a b quark and a W boson, but the other decay modes of the T quark are considered as well. The analysis is optimized for all-jets final states. Therefore, the top quark from the decay of the Z' and the W boson from the decay of the T quark are reconstructed through their all-hadronic decays. The tree-level Feynman diagram for the production of the Z' boson and the decay chain considered in this analysis is shown in Fig. 8.2.

In the kinematic scenarios considered the T quark does not have a large transverse momentum in relation to its mass, thus its decay products are angularly well separated. However, due to the large difference in mass between the W boson and the T quark, the former receives a high Lorentz-boost and its decay products are not well separated, leading to merged jets. Jet substructure algorithms are employed to reconstruct and identify the W boson coming

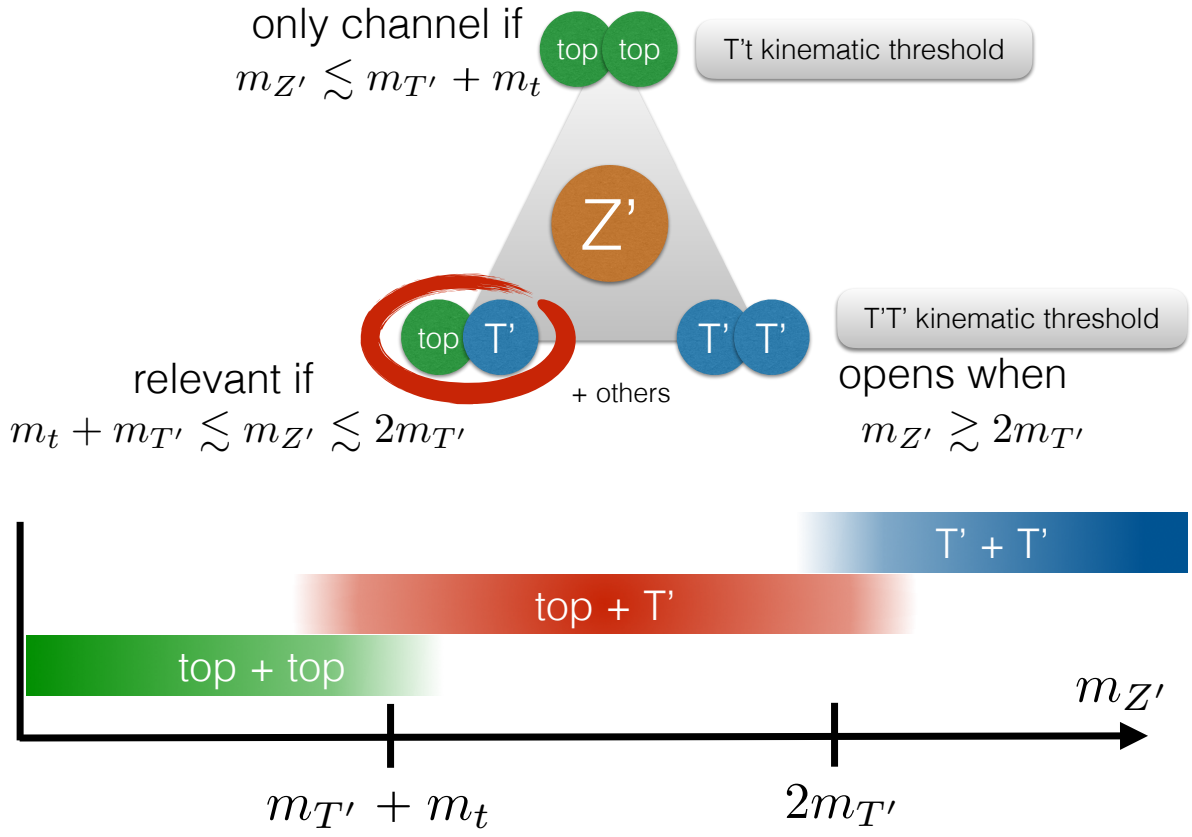


Figure 8.1: Scheme of the relevant decay modes of the resonance in function of the T quark mass.

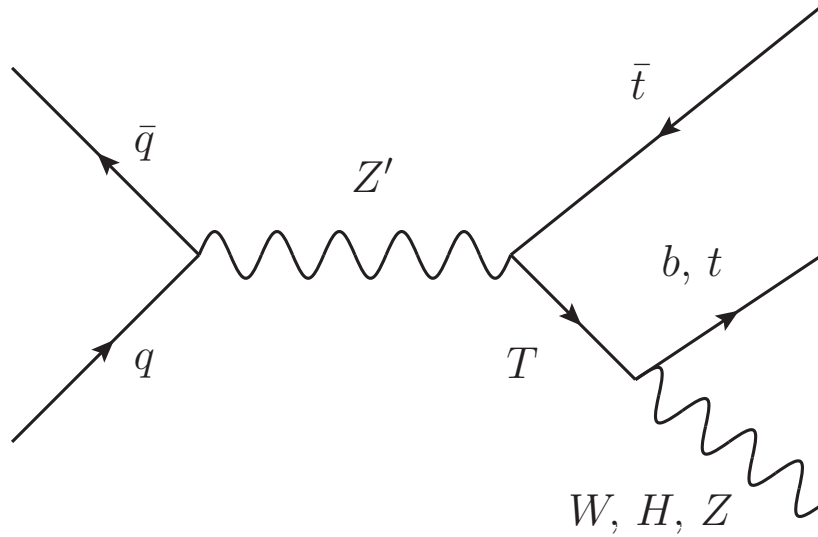


Figure 8.2: Feynman diagram showing the production mode of the Z' boson and the decay chain considered in this analysis.

from the decay of the T quark. If the mass difference between the Z' boson and the T quark is large enough, the top quark coming from the decay of the Z' has a large transverse momentum, and in that case jet substructure techniques and top tagging can be used to identify and reconstruct the hadronic decay of the top quark. Figure 8.3 shows a scheme of the analysis strategy. The main SM background to this search is QCD multijet production, followed

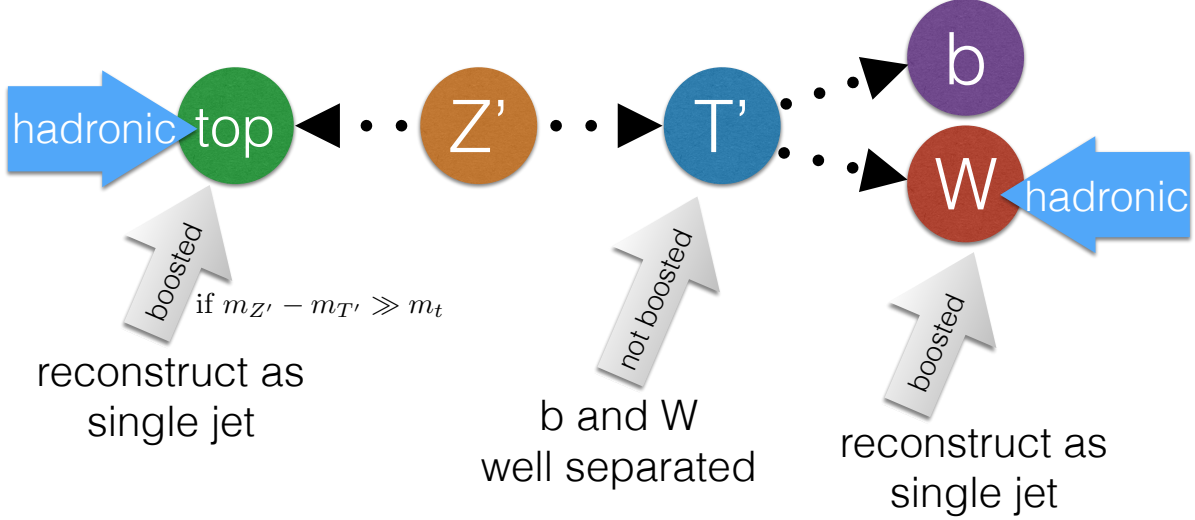


Figure 8.3: Schematic view of the analysis strategy.

by SM top quark production (including $t\bar{t}$ and single top quark). The SM top production is estimated using simulated events, while the QCD multijet production is modeled using a data-driven procedure. The analysis presented here has been published in Ref. [120], and additional details of the analysis can be found in Ref. [121].

8.2 Samples

All the data and simulated samples used in this analysis were processed and reconstructed using the CMS framework version 7.6. The analysis is based on proton-proton collisions collected during 2015 by the CMS experiment at the LHC, at a center-of-mass energy of $\sqrt{s}=13$ TeV. Recorded data from the jet data-stream are used, the samples used are reported in Table A.7. The total integrated luminosity of the data samples is $\mathcal{L} = 2.6 \text{ fb}^{-1}$.

For the modeling of the SM background production and the validation of the data-driven procedure samples of simulated SM processes are used. Table 8.1 lists the SM simulated samples used in this analysis, along with the integrated luminosity of the samples, used for their correct normalization. The SM $t\bar{t}$ sample is generated with POWHEG v2 [91, 122–125] at next-to-leading order precision and is normalized to the next-to-next-to-leading order cross section of $\sigma_{t\bar{t}} = 831.76 \text{ pb}$ [7]. Simulated samples of single top quark events produced in the tW channel are generated with POWHEG and are normalized to a cross section of 71.7 pb [126]. Production of single top quarks in the s-channel and t-channel is generated with aMC@NLO [127] and is normalized to a cross section of 10.32 and 216.99, respectively [128, 129]. All samples use PYTHIA 8.2 for the parton showering. QCD multijet simulated events are generated with PYTHIA 8.2, and are binned in the H_T of the hard interaction. These samples are used for the derivation and validation of the data-driven background estimate. Table 8.1 lists the simulated background samples and their integrated luminosity.

Sample	Int. Lumi. (pb^{-1})
$t\bar{t}$	117767.67
QCD multijet $500 < HT < 700$ GeV	612.6385
QCD multijet $700 < HT < 1000$ GeV	2276.089
QCD multijet $1000 < HT < 1500$ GeV	4183.32
QCD multijet $1500 < HT < 2000$ GeV	32853.02
QCD multijet $2000 < HT < \text{Inf}$ GeV	78495.56
Single top t-channel	18992240
Single top tW channel	28089.89
Single antitop tW channel	28073.03
Single top s-channel	1001494

Table 8.1: Simulated samples of SM processes used in the analysis.

Simulated samples have been produced for several hypotheses of the signal. Neutral spin-1 resonances (Z') are generated forcing the decay exclusively to a top quark and a up-type heavy vector-like quark (T quark). A 1% width for the Z' resonance and a 1 MeV width for the T quark is chosen. The T quark is generated with a left-handed chirality. The widths for the Z' boson and the T quark are chosen to be significantly below the experimental resolution and to follow the benchmark used for $Z' \rightarrow t\bar{t}$ searches. The T width was chosen to be consistent with direct VLQ searches targeting single or pair production of VLQs. While a vector-like quark can, for some configuration of the model parameters, have a larger width, as long as the width is much smaller than the detector resolution (20%), the effect of the width on the acceptance and sensitivity of the analysis is negligible. To test the impact of the resonance width on the analysis, for a single mass configuration, two samples are produced with the Z' boson width and the T quark width increased to 10%, respectively. Additionally, for the same test configuration, one sample is produced with a purely right-handed T quark.

Signal samples are generated for all the main final states of the T quark: a b quark and a W boson, a top quark and a Higgs boson, and a top quark and a Z boson. Several hypotheses for the masses of the Z' boson and the T quark are considered ranging from 1.5 to 2.5 TeV for the Z' and from 0.7 to 1.5 TeV for the T quark. The combination of the Z' and T masses is chosen so that the mass of the T is roughly half, two thirds, or five sixths of the mass of the Z' . The masses are rounded, with respect to the precise ratio, so that the same T quark mass is paired with different Z' mass hypotheses. Figure 8.4 shows a schematic view of the Z' and T masses for the generated signal samples. This kinematic range is the most interesting for the decay mode of the Z' to a SM quark and a vector-like quark, as opposed to decay modes to VLQ pairs or SM quark pairs. If the mass of the Z' is more than twice the mass of the T quark, then the decay mode of the Z' in T quark pairs is kinematically preferred with respect to Tt . So the range where the $Z' \rightarrow Tt$ is relevant and interesting lies between $m(Z')/2 < m(T) < m(Z') - m(t)$. The left-hand side is the threshold where $Z' \rightarrow TT$ becomes kinematically allowed, while the right-hand side is the threshold where the Tt decay mode becomes kinematically allowed.

The signal production is based on a simplified low-energy effective model describing the phenomenology of heavy vector resonances in minimal composite Higgs theories [38] (see also Sec. 1.4.1). The samples are generated using MADGRAPH v5.2 [127]. The decay of heavy resonances for signal samples is treated with MADSPIN [130] to correctly propagate the helicity through the decay products, more details are given in Section 8.2.1. Higher-order parton radiations up to one extra partons are included. PYTHIA 8.2 [131] is used to model

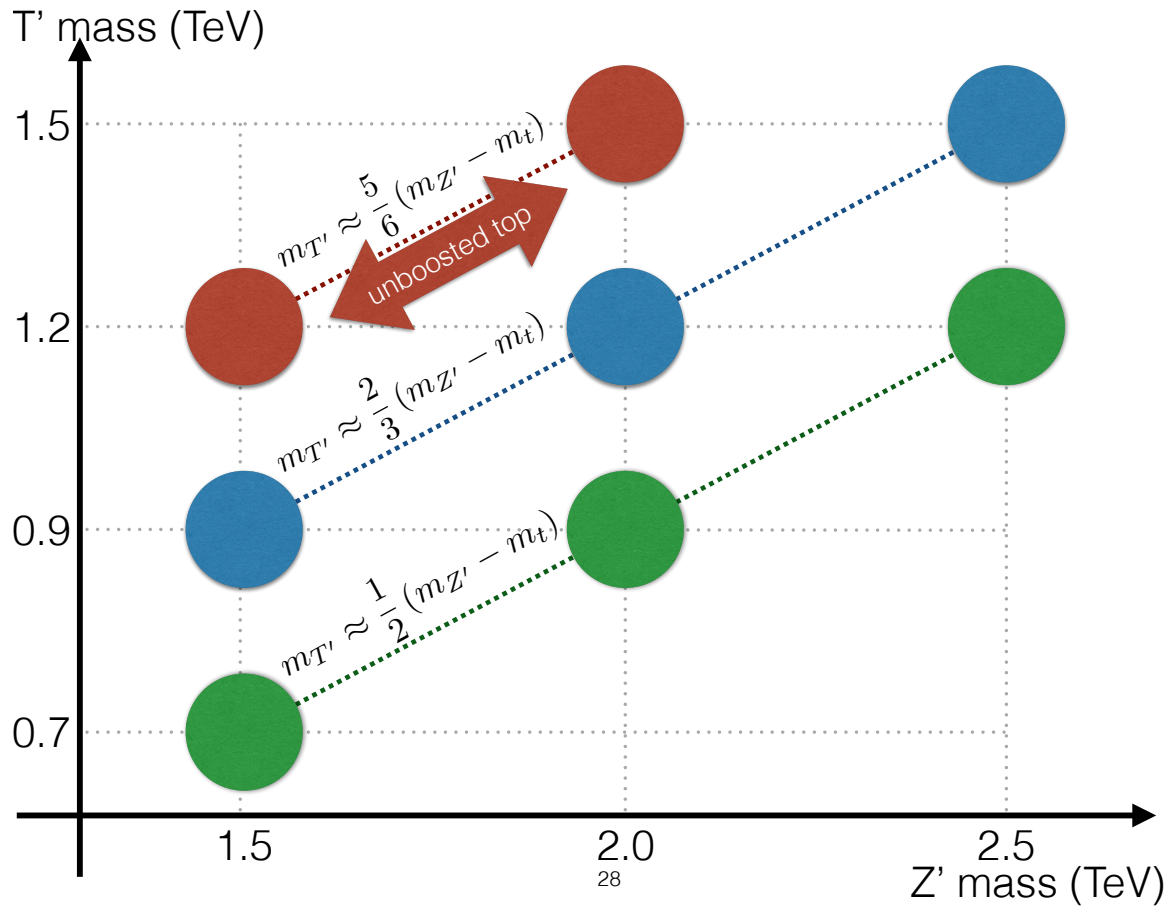


Figure 8.4: Scheme of the Z' boson and T quark masses for the generated signal samples.

the parton showering and the MLM algorithm [71] is used for the matching between the matrix element generator and the parton showering. Differential jet rates have been checked for smoothness to ensure that the matching scale is chosen correctly. The full list of signal samples used in the analysis is reported in Table A.8.

A selection of kinematic distributions of the signal samples at generator level is shown in Figs. 8.5, 8.6. The generator particles considered in the distributions are found by reconstructing the decay chain using daughter links, starting at the Z' boson. Features of general importance for the event selection are the transverse momentum of the top quark and the W boson and the angular separation of the decay products.

Simulated events are weighted as a function of the number of primary vertices so that the number of true pileup interactions in simulation matches the instantaneous luminosity profile measured in data. A minimum bias cross-section of 69 mb is assumed. The distribution of the number of primary vertices in data and simulation after the weighting procedure is applied can be seen in Figure 8.7.

8.2.1 Modeling of the spin correlations in signal samples

The chirality of a particle is propagated through its decay products and it affects the kinematics of the stable particles that are measured by the detector. Proper modeling of the spin correlations, that is, the correct propagation of the chirality of a particle through its decay products, is computationally intensive. Most simulated samples of SM processes used in CMS have spin correlation properly modeled. This is an essential feature to perform precise measurements of SM parameters. However, for simulated events containing beyond the Standard Model physics, the chirality is normally not propagated to the final decay products. The reason is that these kind of samples often present very long decay chains with respect to SM processes, making a full computation of spin correlations computationally very challenging. For some searches of new physics this level of precision is not essential, so this does not represent a problem. However, for decays of high-mass vector-like quarks, different chirality assumptions for the heavy quark might lead to different kinematics of the decay products. Such differences are lost if spin correlations are not properly modeled. The effect is visible the most in the distribution of the transverse momentum of the lepton coming from the decay of the VLQ, such as the muon coming from the decay of the top quark for a T quark decaying to tH or tZ , or the muon coming from the decay of a W boson for a T quark decaying to bW . This is of particular importance for analyses looking for VLQs in lepton plus jets final states. Figure 8.8 shows the difference in the distribution of the p_T of the lepton from the T quark decay, assuming $T \rightarrow tH$.

The simulated signal samples used for this analysis are generated with MADGRAPH at matrix-element level. While MADGRAPH is able to preserve spin correlations, extensive tests proved that the long decay chain makes the generation too computationally intensive. The MADSPIN package offers a solution to the problem, simplifying the calculation of the spin correlations assuming a narrow width hypothesis. The generation setup used is the following: the production of the Z' boson, and the decay of the Z' boson and the T quark are generated by MADGRAPH. The decay of the top quark, the W and Z bosons is treated with MADSPIN. The decay of the H boson and the showering is handled by PYTHIA. A schematic view of the setup is shown in Fig. 8.9.

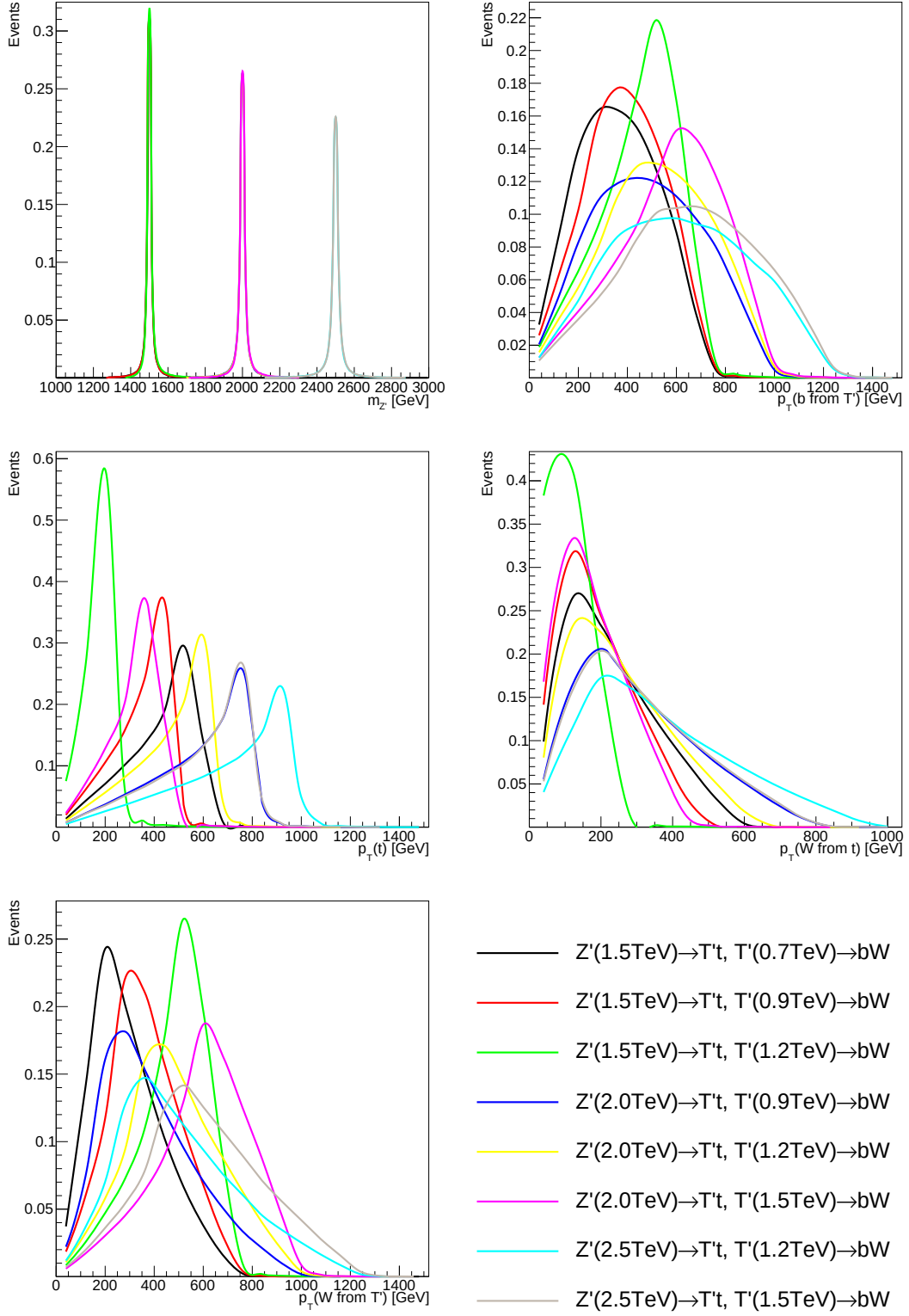


Figure 8.5: Generator-level kinematic distributions for signal samples. First row, from left to right: mass of the Z' boson, p_T of the b quark from the T quark, p_T of the top quark. Second row, from left to right: p_T of the W boson from the top quark, p_T of the W boson from the T quark.

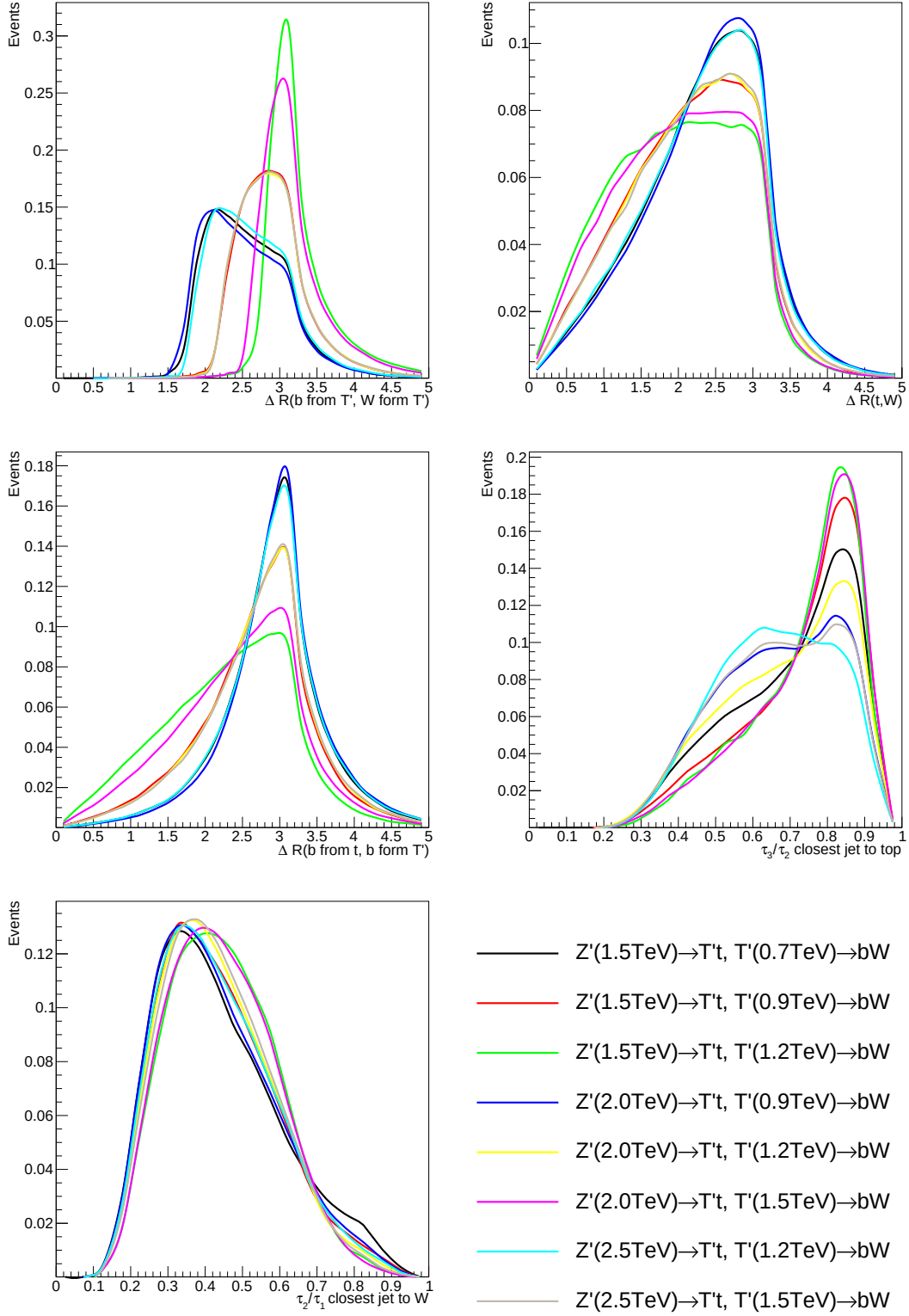


Figure 8.6: Generator-level kinematic distributions for signal samples. First row, from left to right: ΔR between the b quark from the T quark and the W boson from the T quark, ΔR between the W boson from the T quark and the top quark, ΔR between the top quark and the b quark from the T quark. Second row, from left to right: τ_3/τ_2 of the closest jet to the top quark, τ_2/τ_1 of the closest jet to the W boson.

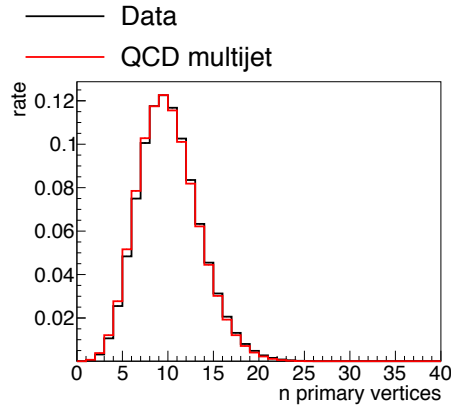


Figure 8.7: Distribution of the number of primary vertices in data and simulation after the pileup weighting procedure is applied.

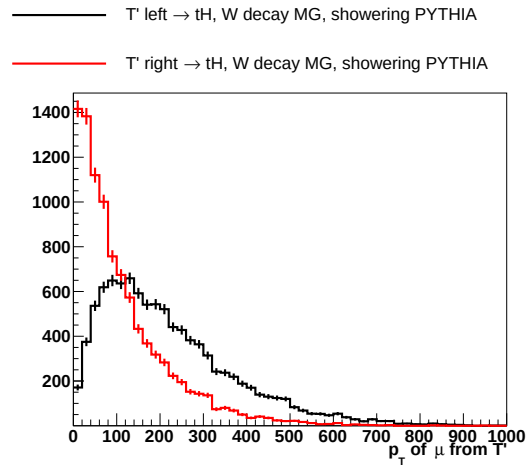


Figure 8.8: Distribution of the p_T of the muon coming from the decay of the T quark, assuming $T \rightarrow tH$.

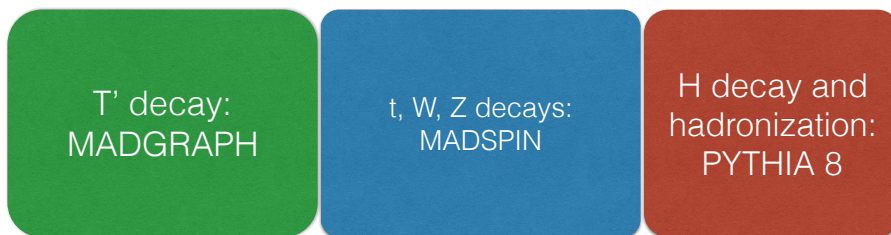


Figure 8.9: Generation scheme used to preserve spin correlations.

8.3 Selection

In this analysis, similarly to the one described in Chapter 8, full event reconstruction is pursued. The focus of the search is on the all-jets final state. The decays of the top quark and W boson are reconstructed using jet-substructure based taggers. B tagging is used to identify the b jet coming from the decay of the T quark. Subjet b tagging is used to further refine the selection of events containing a top quark.

The primary vertex is identified in each event by clustering tracks using a deterministic annealing algorithm. Only vertices passing the following requirements are considered:

- $\sqrt{x^2 + y^2} < 2 \text{ cm}$ and $|z| < 24 \text{ cm}$,
- $N_{DOF} > 4$, where N_{DOF} is the weighted number of tracks used to reconstruct that vertex.

The vertex with the highest $\sum_{tracks} p_T^2$ is selected as the primary vertex. Two different kind of jets are used: standard anti- k_T $R=0.4$ (AK4) jets, and anti- k_T $R=0.8$ (AK8) jets.

All AK4 jets are required to have a p_T larger than 30 GeV, must pass the loose particle-flow identification criteria, and have $|\eta| < 2.4$. Jet identification criteria aim at removing spurious jets due to detector noise effects by requiring that the neutral components contribute each less than 99% of the hadronic and electromagnetic energy in a jet. Level 1, 2 and 3 jet corrections are applied to simulated and data events, additionally Level 2 and 3 residual corrections are applied in data according to Ref. [132]. Jet energy resolution smearing is applied to simulated events, based on a matching algorithms between generator-level and reconstructed jets, according to the values reported in [133]. AK8 jets are used as a source for the reconstruction of top quark and W boson candidates, and are required to have $p_T > 200 \text{ GeV}$ and $|\eta| < 2.4$. The Soft-Drop (SD) grooming algorithm (angular exponent $\beta = 0$, and soft threshold $z_{cut} < 0.1$) is applied to the particle-flow components of the AK8 jet after the procedure of charged-hadron subtraction (CHS). No further jet identification criteria are applied to AK8 jets. Level 1, 2 and 3 jet corrections are applied to AK8 jets. For the subjets, level 2 and 3 AK4 jet corrections are used. The jet p_T is computed from the AK8 ungroomed jet components, while the jet mass is computed using SD subjets. The proper jet energy resolution smearing is applied to AK8 jets. The smearing factors derived for AK4 jets are applied to the subjets.

B-tagging is applied on AK4 jets and on the SD subjets of the AK8 jets, a medium-efficiency working point of the CSVv2-algorithm is used [134,135]. The correspondent b-tagging scale factors are applied to heavy flavour jets (b and c quarks) and light flavour jets (u, d, and s quarks, gluons) [95,115]. Similarly, dedicated scale factors for subjet b tagging are applied. This analysis uses W boson tagging to reconstruct the boosted hadronic decay of the W boson, coming from the decay of the T vector-like quark. The W tagger used in the analysis has been described in Section 4.4.3. The selection criteria of the W tagger are summarized in the following:

- AK8 jet with $p_T > 200 \text{ GeV}$,
- groomed SD mass: $70 < m_{SD} < 100 \text{ GeV}$,
- N-subjettiness ratio: $\tau_2/\tau_1 < 0.6$.

A flat rate correction factor of 0.98 ± 0.03 is used take into account data-simulation discrepancies in W-tagging efficiency. The value has been derived in Ref. [87]. The scale factor is applied only when the W-tagged jet is matched to a generator-level W boson, using a ΔR criterium. The analysis also uses a top quark tagging algorithm to reconstruct the boosted

hadronic decay of the top quark, coming from the decay of the Z' boson. The top tagger used is the CMS Top Tagger v2 and has been described in Section 4.4.2. The selection criteria of the top tagger are summarized in the following:

- AK8 jet with $p_T > 400 \text{ GeV}$,
- groomed SD mass: $110 < m_{SD} < 210 \text{ GeV}$,
- N-subjettiness ratio: $\tau_3/\tau_2 < 0.86$.

To increase the discrimination power against SM backgrounds, subjet b tagging is used in conjunction with the top tagger. The CSVv2 discriminator is computed for every subjet of a top-jet candidate. The maximum among these values is defined as $\beta_{\max}$. For a top jet to be b-tagged, the $\beta_{\max}$ value is required to be above the threshold of the medium efficiency operating point. Data-simulation correction factors for the top-tagging algorithm are derived in Ref. [87]. A p_T -dependent scale factor is applied to the top-tagged jet (0.85 for $p_T < 550 \text{ GeV}$, 1.08 for $p_T > 550 \text{ GeV}$). Figure 8.10 shows a summary of the tagging algorithms used in the analysis.

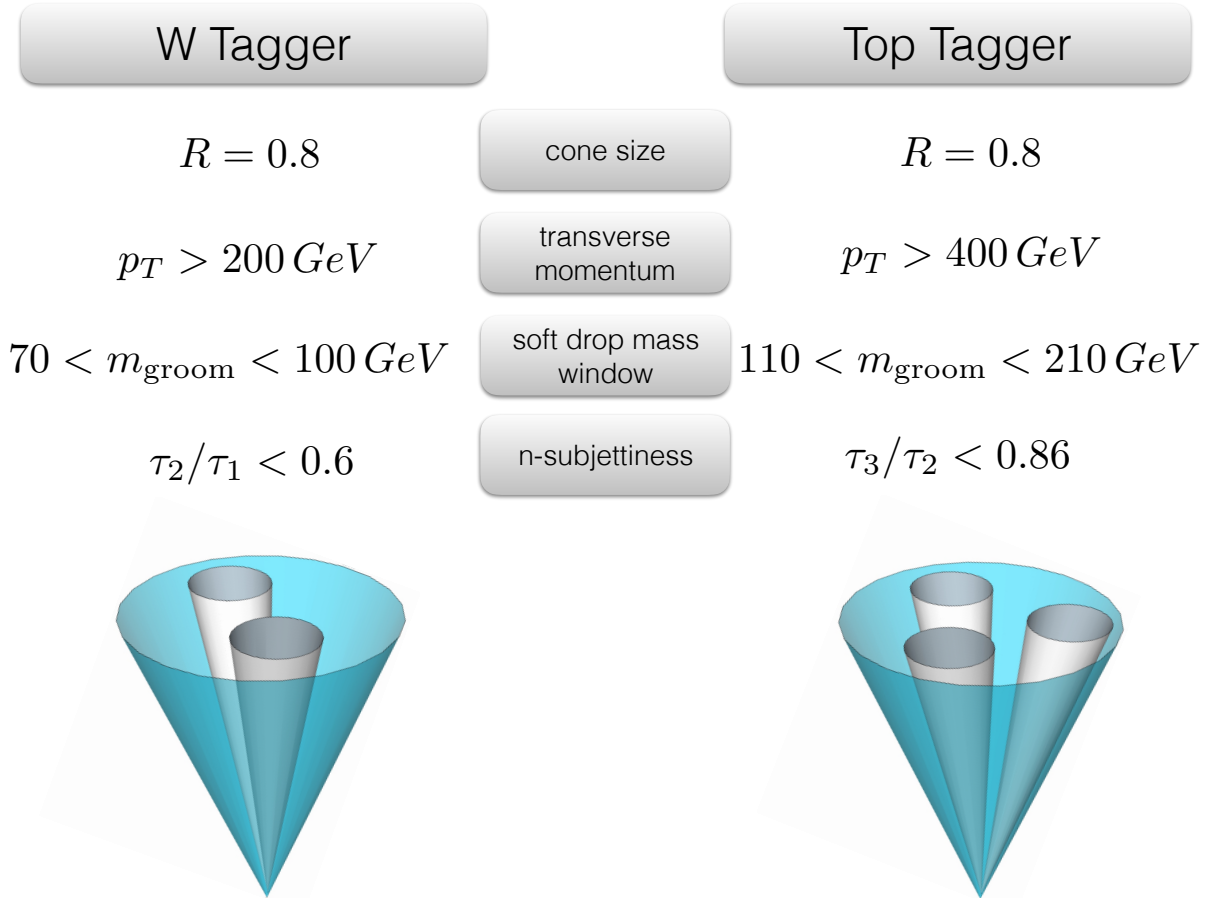


Figure 8.10: Summary of the tagging algorithms used in the analysis.

8.3.1 Trigger

The analysis uses an H_T trigger requiring H_T to be greater than 800 GeV . The H_T variable is defined as the sum of the p_T of AK4 jets clustered from trigger-level PF components. Given

that the masses of the Z' boson being probed start at 1.5 TeV, this condition is satisfied by the vast majority of the events for every signal configuration considered in the analysis. Hence jet substructure triggers are not considered in this analysis to complement the H_T trigger.

Trigger studies are performed to assess where to set the offline H_T selection, in order to reject the events where the trigger efficiency is not stable in function of the characteristic variable of the trigger (H_T in this case). A lower-threshold trigger is needed to pre-select the events in order to measure the trigger efficiency in data, referred to as “base trigger”. In this case an H_T trigger with a threshold set at 650 GeV is used. The trigger is prescaled, that is, not all the events firing the trigger are retained, in order to limit the rate of the trigger.

Trigger efficiency curves are measured using the final analysis selection defined in Section 8.3.2: one top-tagged jet ($p_T > 400$ GeV), one W-tagged jet ($p_T > 200$ GeV), one medium b-tagged jet ($p_T > 100$ GeV, $\Delta R(\text{b-jet}, \text{top-jet}), \Delta R(\text{b-jet}, \text{W-jet}) > 0.8$). The trigger efficiency is defined as a ratio where the denominator is the number of events passing the base trigger and the selection criteria. The numerator is defined by all events passing the criteria of the denominator and firing the trigger being studied. For this study data and simulated samples defined in Section 8.2 and Table 8.1 are used. The trigger efficiency curves are compared for data and simulation, where simulated events include SM $t\bar{t}$ production and QCD multijet weighted with their relative cross sections. The trigger efficiency is shown in function of the H_T variable in Fig. 8.11.

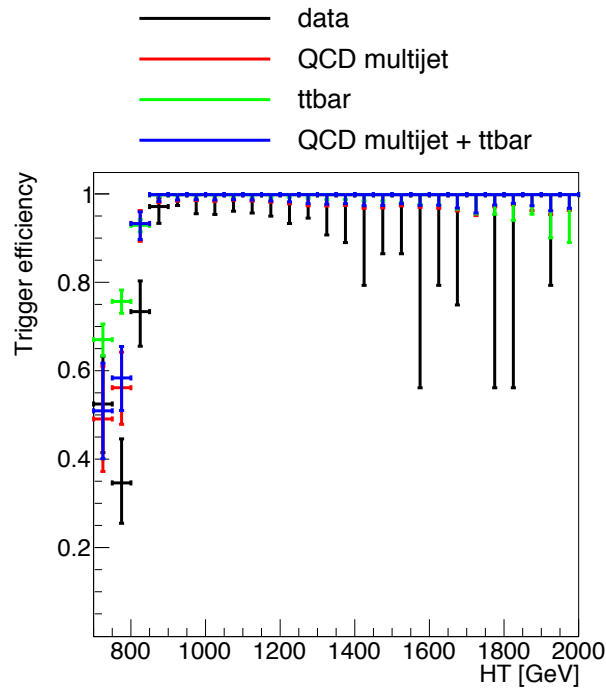


Figure 8.11: H_T trigger efficiencies in function of H_T for data (black), QCD multijet simulated events (red), $t\bar{t}$ simulated events (green), and QCD multijet plus $t\bar{t}$ simulated events (blue).

From these studies it can be concluded that requiring the H_T to be above 850 GeV is a good choice to reject the trigger efficiency turn on. No trigger scale factors are deemed necessary for this analysis since above this threshold the triggers are completely efficient,

except for the 850-900 GeV H_T bin in data. A flat 3% uncertainty is assigned to account for trigger efficiency discrepancy between data and simulation. The value is motivated by the maximum discrepancy observed between the data and simulation efficiency curves above $H_T = 850$ GeV.

8.3.2 Final selection

The selection is based on a three-jets topology: two AK8 jets corresponding to the decays of the top quark (from the decay to the Z' boson) and the W boson (from the decay of the T quark), and an AK4 jet corresponding to the b quark coming from the decay of the T quark.

The first step is to select events with a top-tagged AK8 jet and a W-tagged jet. The two jets are not required to be the two most energetic ones, imposing this condition might reject events where the b quark from the decay of the T quark has a p_T larger than the W boson or the top quark. Figure 8.12 shows the distribution of the substructure variables related to the top and W tagging algorithms before a selection of the variable itself is applied; all the other tagging selections are applied. These plots are called “N-1” plots. Figure 8.13 shows various distributions of the number of tagged objects (top-, W-, or b-tagged) in the event.

The next step is to select a candidate for the jet associated to the b quark from the decay of the T quark, in particular the jet must be angularly separated from the other AK8 jets to avoid double-counting. Given the generator-level angular distributions of the final-state particles for signal events, it is found that the optimal selection for the AK4 jet is: $p_T > 100$ GeV, $\Delta R(\text{top-jet, jet}) > 0.8$, $\Delta R(\text{W-jet, jet}) > 0.8$. Additionally, the jet must satisfy medium-efficiency b-tagging requirements. Figures 8.14, 8.15, and 8.16 show angular distributions between the objects selected in the analysis, and kinematic distributions associated to the AK4 and AK8 jets.

As a last step, the T quark candidate is defined as the sum of the 4-vectors of the W-tagged jet and the b-tagged jet is required to have a mass larger than 500 GeV. Similarly the Z' boson candidate is defined as the sum of the 4-vectors of the top-tagged jet and the T quark candidate. Figure 8.17 shows a summary of the selection criteria used in the analysis. Events are classified in two categories: the first requires that the top-tagged jet is not b tagged (referred to as “1 b tag category” in the following) and the second requires that the top-jet candidate is b tagged (“2 b tag category”). Figure 8.18 shows the event categories and their estimated background composition. A variety of distributions of relevant variables including kinematic distributions for the T quark and Z' boson candidates are shown in Figures 8.19–8.21, in particular Fig 8.21 shows the distribution of the Z' candidate mass in the two signal regions, which is the main observable of the analysis. Many of these plots show a poor agreement between data and simulation. The main component of the background is the QCD multijet production, estimated with a leading-order simulation in these plots. Simulated events of QCD multijet are not reliable in modeling both the normalization and the shape of distributions in some of the the areas of the phase space considered. A dedicated data-driven procedure to estimate the QCD multijet production in the signal region, for the distribution of the Z' candidate mass, will be developed in Section 8.4. The total efficiency for every signal hypothesis considered in this analysis is reported in Tab. 8.2.

Figure 8.22 shows the distribution of $M_{Z'}$ in the signal region for a single configuration of the masses of the Z' boson and the T quark, varying the Z' boson width, T quark width, and T chirality. In particular, the width of the Z' boson or the T quark is increased to 10%, and a pure right-handed T quark is simulated.

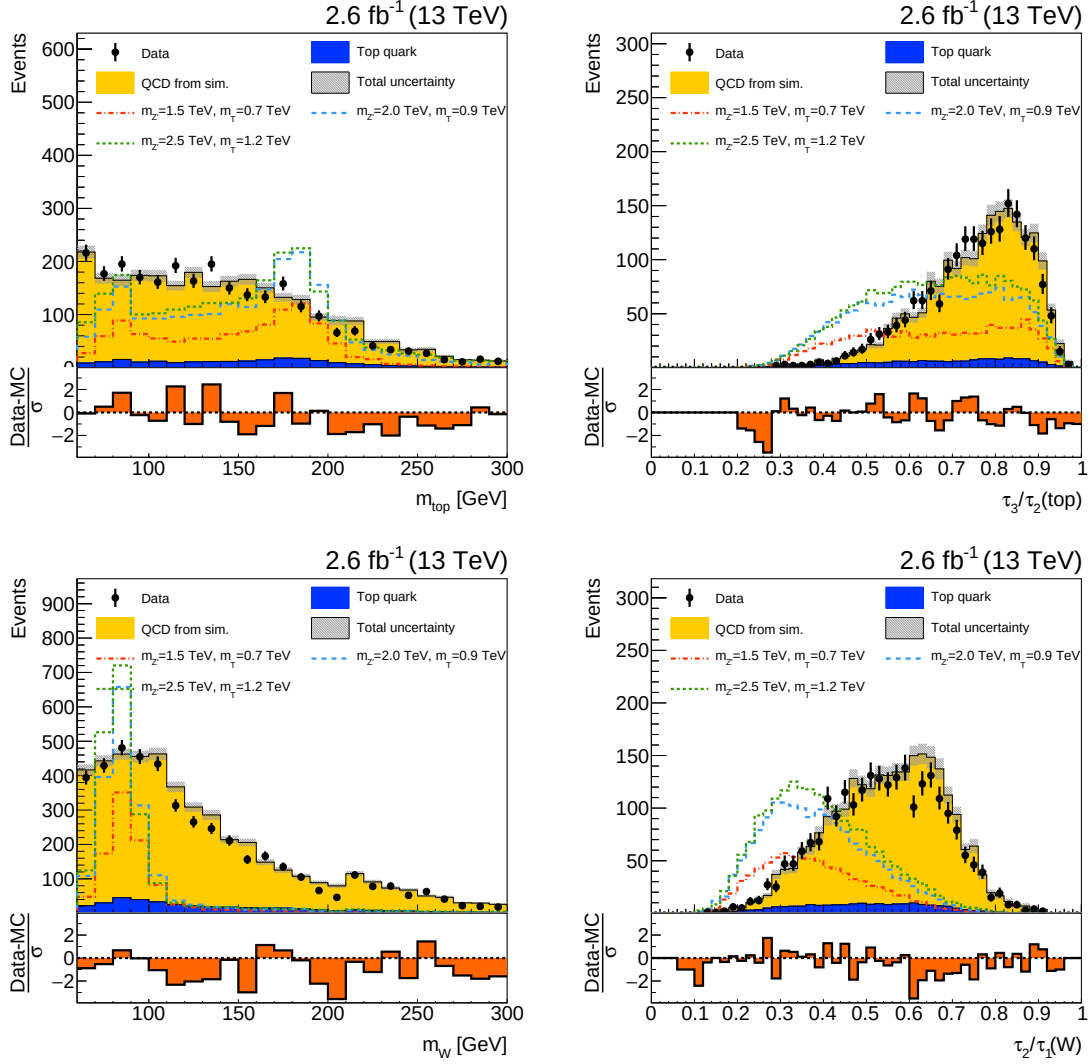


Figure 8.12: Various “N-1” distributions for the main selection criteria in the analysis for signal, data, QCD multijet, and top quark events. The distributions are plotted before the selection on the T quark mass. Signal events are normalized to a cross section of 10 pb. First row, from left to right: top-jet SD mass, top-jet N-subjettiness τ_3/τ_2 . Second row, from left to right: W-jet SD mass, W-jet N-subjettiness τ_2/τ_1 .

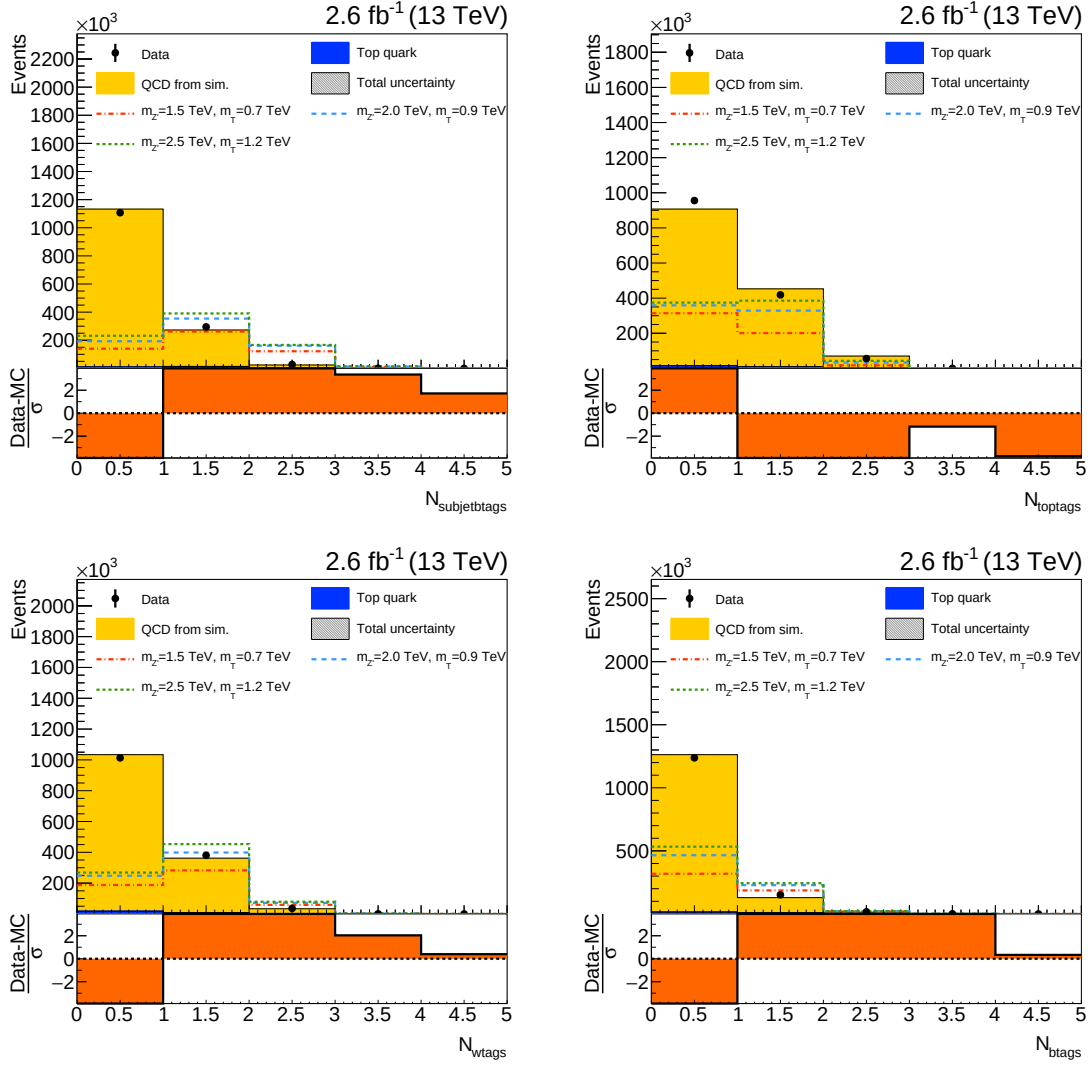


Figure 8.13: Various per-event counting distributions of the objects used in the analysis for signal, data, QCD multijet, and top quark events. The distributions are plotted before requiring a top-tagged jet and a W-tagged jet. Signal events are normalized to a cross section of 500 pb. For these plots, the QCD multijet background is normalized to the number of data events (top quark subtracted). First row, from left to right: number of b-tagged subjets, number of top-tagged jets. Second row: number of W-tagged jets, number of b-tagged jets satisfying the ΔR and kinematic requirements: $p_T > 100 \text{ GeV}$, $\Delta R(\text{b-jet}, p_{T\text{-leading CA8 jet})} > 0.8$, $\Delta R(\text{b-jet}, \text{second most energetic CA8 jet}) > 0.8$. The discrepancy between data and simulation is caused by the QCD multijet sample. Only statistical uncertainties are quoted.

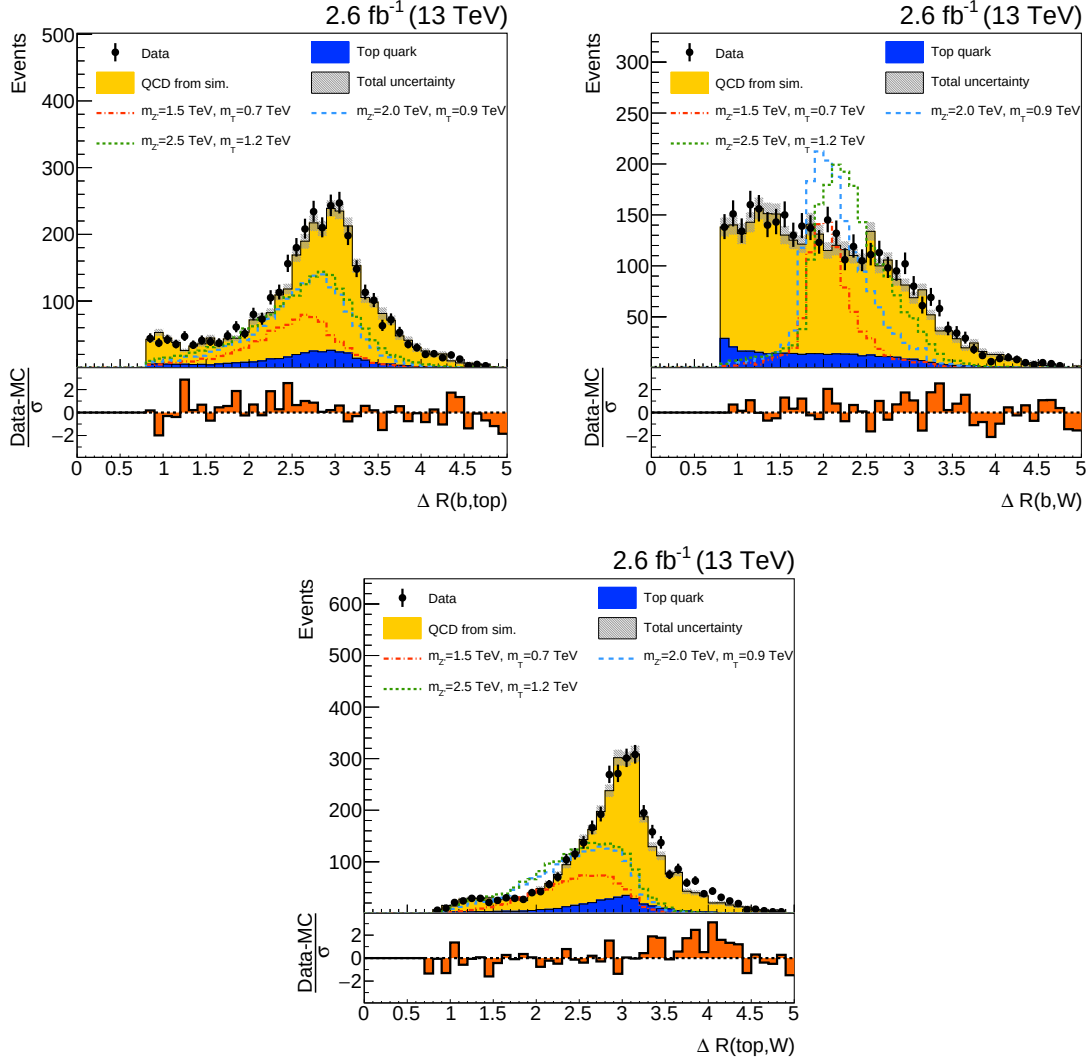


Figure 8.14: Various angular distributions between the selected objects in the analysis for signal, data, QCD multijet, and top quark events. The distributions are plotted before the selection on the T quark mass. Signal events are normalized to a cross section of 10 pb. First row, from left to right: $\Delta R(\text{b-jet, top-jet})$, $\Delta R(\text{b-jet, W-jet})$. Second row: $\Delta R(\text{top-jet, W-jet})$.

Table 8.2: Selection efficiency for the signal in the different categories used in the analysis [120]. Statistical uncertainties are quoted.

$\mathcal{B}(T \rightarrow bW) = 1$			
$m_{Z'}$ [GeV]	m_T [GeV]	Efficiency SR 1 b tag [%]	Efficiency SR 2 b tag [%]
1500	700	1.2 ± 0.2	1.9 ± 0.3
1500	900	0.74 ± 0.17	1.1 ± 0.2
1500	1200	0.23 ± 0.09	0.21 ± 0.09
2000	900	2.6 ± 0.3	3.6 ± 0.4
2000	1200	2.1 ± 0.3	3.0 ± 0.4
2000	1500	0.89 ± 0.18	0.87 ± 0.18
2500	1200	3.3 ± 0.4	3.9 ± 0.4
2500	1500	2.8 ± 0.3	3.6 ± 0.4
$\mathcal{B}(T \rightarrow tH) = 1$			
$m_{Z'}$ [GeV]	m_T [GeV]	Efficiency SR 1 b tag [%]	Efficiency SR 2 b tag [%]
1500	700	0.55 ± 0.15	0.75 ± 0.17
1500	900	0.65 ± 0.16	0.93 ± 0.19
1500	1200	0.26 ± 0.10	0.37 ± 0.12
2000	900	1.8 ± 0.3	2.6 ± 0.3
2000	1200	2.0 ± 0.3	2.9 ± 0.3
2000	1500	1.7 ± 0.3	2.2 ± 0.3
2500	1200	2.9 ± 0.3	3.9 ± 0.4
2500	1500	3.0 ± 0.3	4.1 ± 0.4
$\mathcal{B}(T \rightarrow tZ) = 1$			
$m_{Z'}$ [GeV]	m_T [GeV]	Efficiency SR 1 b tag [%]	Efficiency SR 2 b tag [%]
1500	700	0.62 ± 0.15	0.84 ± 0.18
1500	900	0.78 ± 0.17	0.98 ± 0.19
1500	1200	0.50 ± 0.14	0.54 ± 0.14
2000	900	2.4 ± 0.3	3.1 ± 0.4
2000	1200	2.8 ± 0.3	3.9 ± 0.4
2000	1500	2.3 ± 0.3	2.8 ± 0.3
2500	1200	4.3 ± 0.4	5.4 ± 0.5
2500	1500	4.5 ± 0.4	6.0 ± 0.5

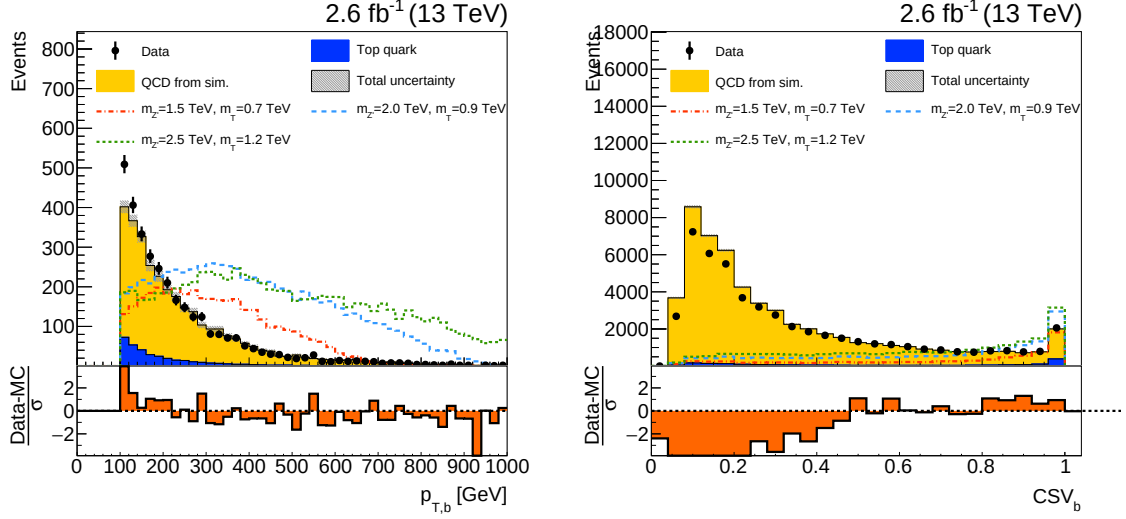


Figure 8.15: Various distributions of variables related to the AK4 jet satisfying the signal region selections except the b-tagging requirement. The distributions are plotted after requiring a top-tagged jet and a W-tagged jet. From left to right: jet p_T after requiring the jet to be b-tagged, CSVv2 discriminator of the jet with the highest CSVv2 discriminator. Signal events are normalized to a cross section of 40 pb. The discrepancy between data and simulation is caused by the QCD multijet sample.

8.4 Background estimate

8.4.1 Top-quark background estimate

The background in the signal region consists, in large part, of QCD multijet events, even in the 2 b tags category. The relatively small fraction of background events containing a real top quark ($t\bar{t}$ or single top production) is caused by the selection criterium applied to the invariant mass of the T quark candidate. It is estimated that the fraction of background events containing a real top quark amounts to 6% and 23% in the 1 b tag and 2 b tag categories, respectively. The SM top quark contribution is therefore estimated from simulated events using the $t\bar{t}$ and single top quark samples reported in Section 8.2.

Top- p_T reweighting, as explained in Section 7.3.1, is not applied to simulated $t\bar{t}$ events because the systematic uncertainties associated to the choice of the renormalization and factorization scales cover the possible shape variation due to the mismodeling of the p_T distribution.

8.4.2 Non-top background estimate

The major background is composed by the non-top component. In this all-jets analysis, the main component of this background is multijet QCD production. Simulated QCD multijet event samples are often not able to properly reproduce the data in the corners of the phase space probed by BSM searches, therefore a data-driven estimate of this background is needed. The first step is the definition of a sideband region with the following requirements:

- A top tagged jet and a W tagged jet selected with the same criteria as the signal region.
- A veto on b-tagged AK4 jets with the same kinematic criteria used for the signal region.

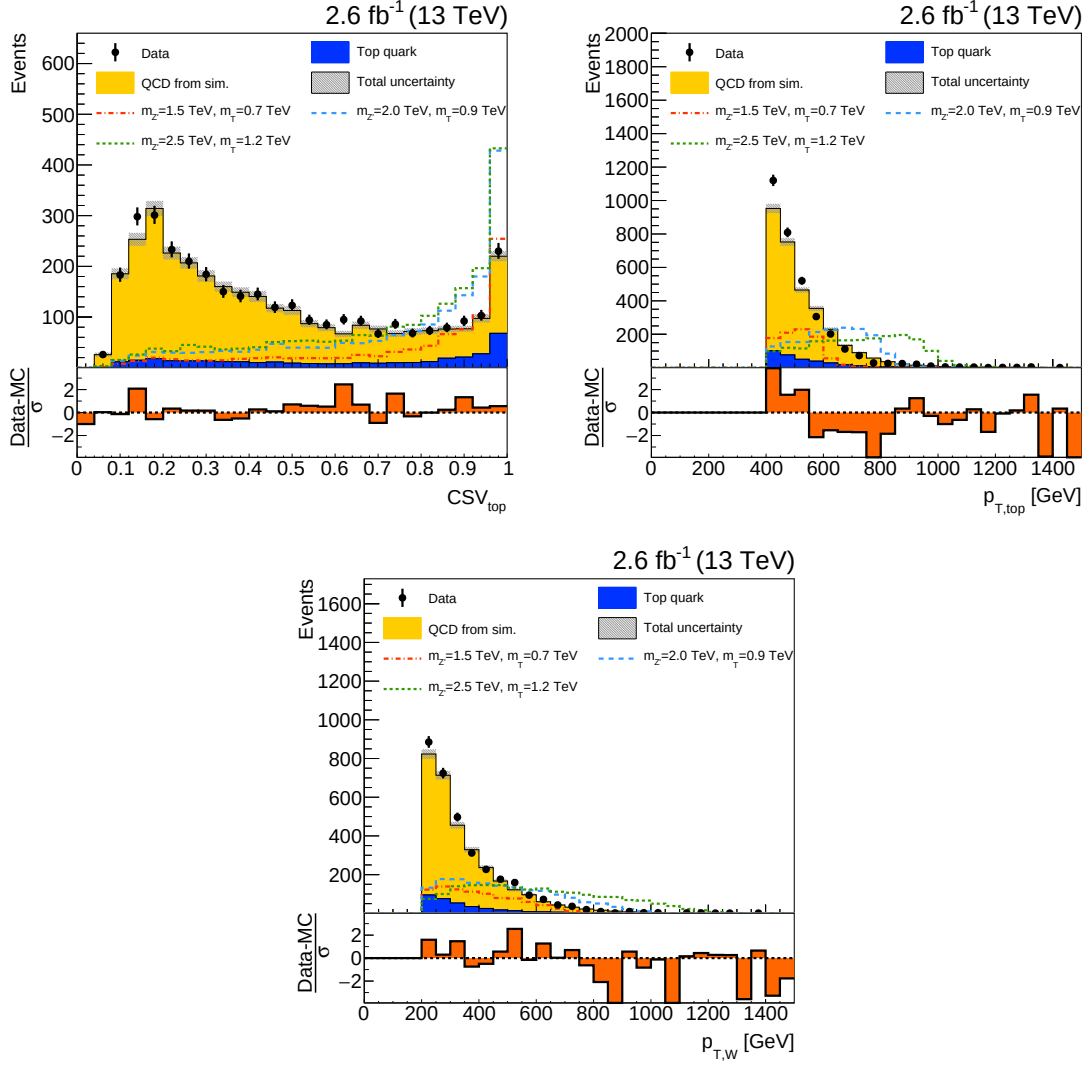


Figure 8.16: Distribution of various kinematic variables related to the top-tagged jet and the W-tagged jet, before the selection on the T quark mass. Signal events are normalized to a cross section of 10 pb. First row, from left to right: $\beta_{\max}$ of the top-tagged jet, top-tagged jet p_T . Second row: p_T of the W-tagged jet.

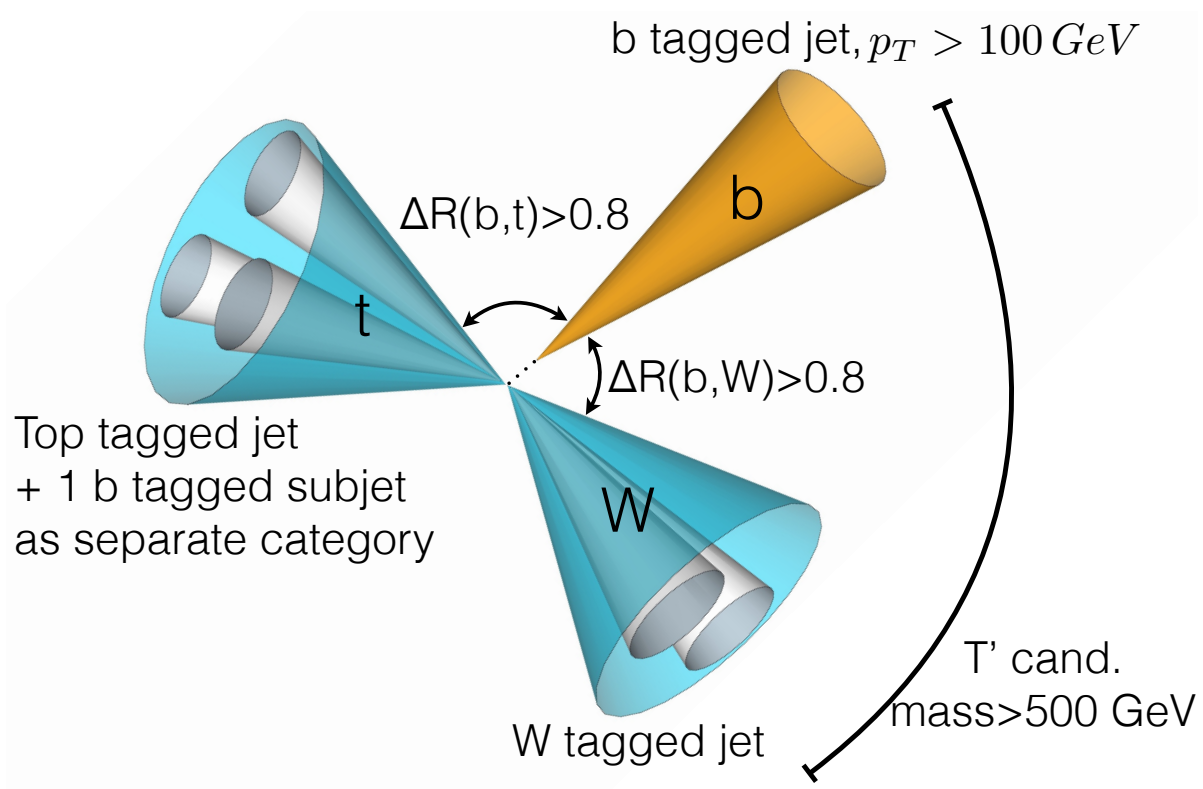


Figure 8.17: Summary of the selection criteria used in the analysis.

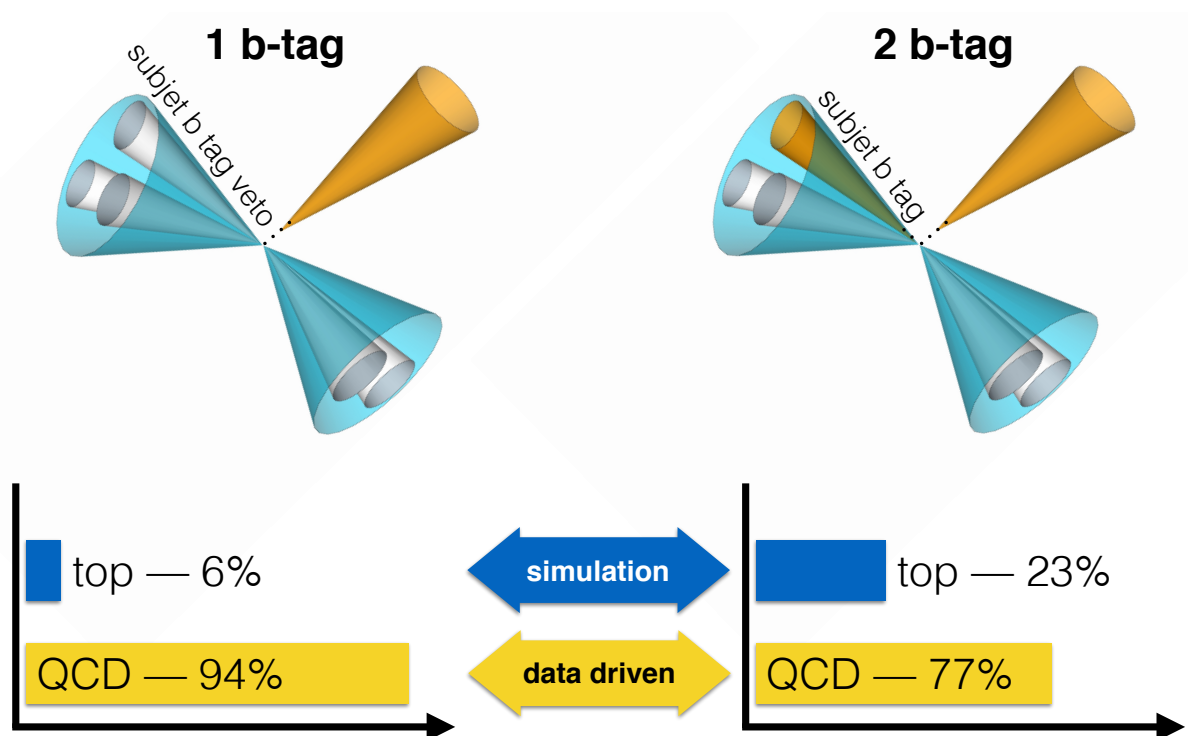


Figure 8.18: The event categories and their estimated background composition.

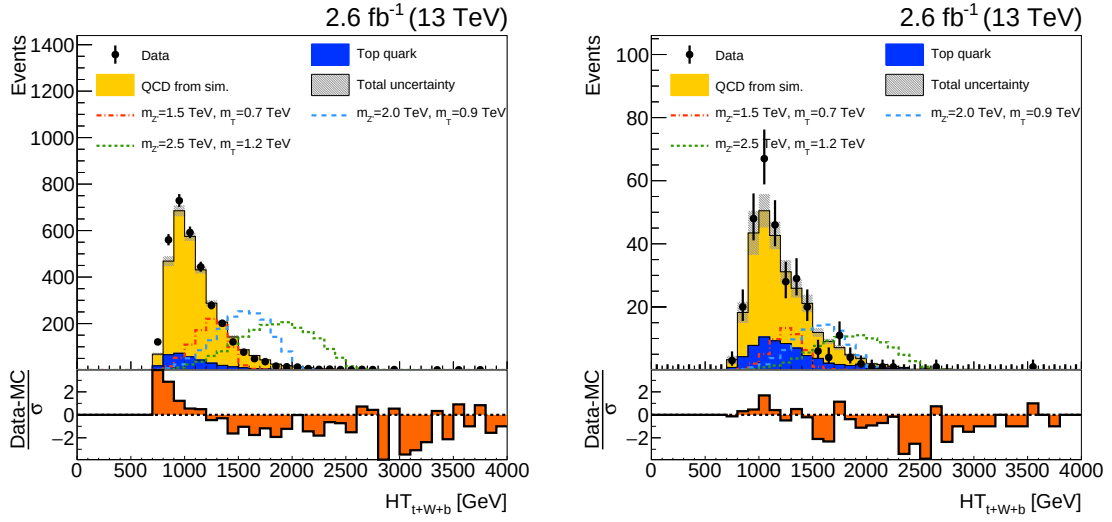


Figure 8.19: Distribution of $H_{T,twb}$ defined as the sum of the p_T of the top-tagged jet, the W-tagged jet and the b-tagged jet, before the selection on the T quark mass in the left panel (signal normalized to 10 pb), and after the selection in the right panel (with subjet b tagging, signal normalized to 1 pb).

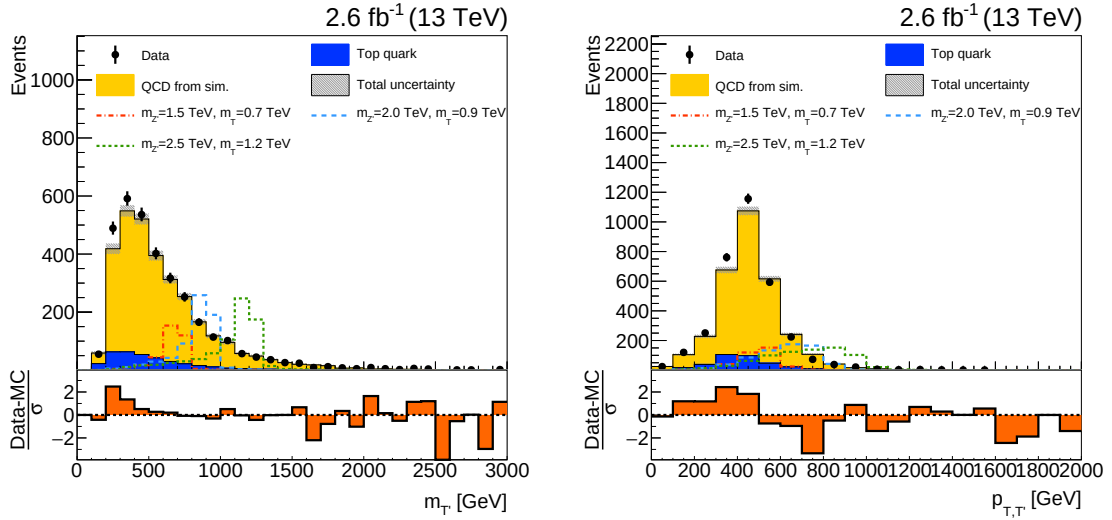


Figure 8.20: Distribution of various kinematic variables related to the T quark candidate. Signal events are normalized to a cross section of 4 pb. From left to right: T quark candidate mass, T quark candidate p_T .

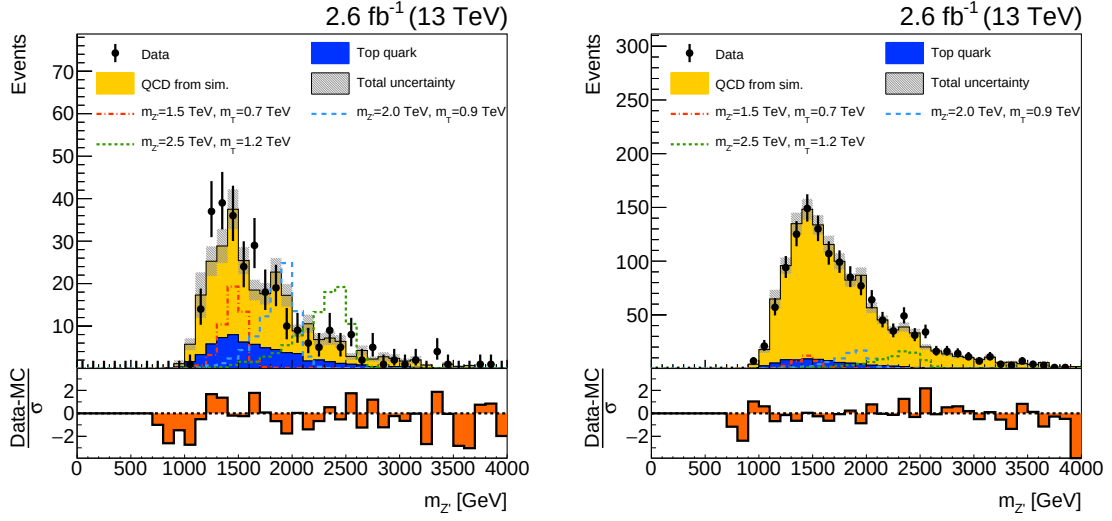


Figure 8.21: Distribution of the Z' candidate mass for the two event categories: 2 b tag category (left) and 1 b tag category (right). Signal events are normalized to a cross section of 1 pb. QCD multijet production is estimated from simulated events.

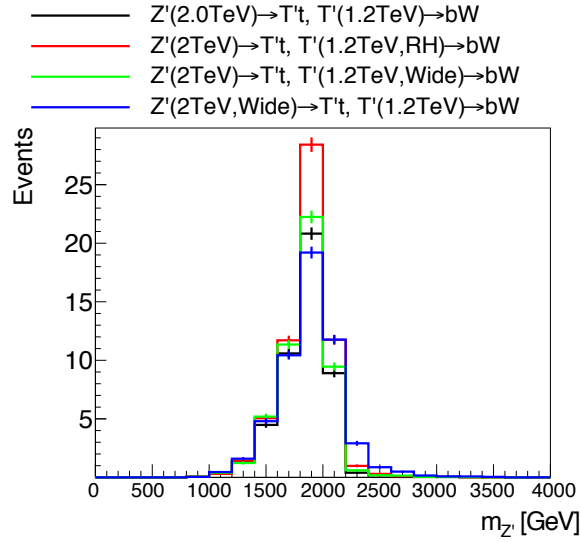


Figure 8.22: $M_{Z'}$ distribution in the signal region for a signal samples with different hypotheses for the T quark width, the Z' boson width, and the T quark chirality, for the bW final state of the T quark.

For the veto a loose-efficiency operating point of the CSVv2 algorithm is used. This operating point yields an efficiency of approximately 85%, and a rejection rate of light-flavor jets of 90% [68].

- An AK4 jet with $p_T > 100 \text{ GeV}$, $\Delta R(\text{top-jet, jet}) > 0.8$, $\Delta R(\text{W-jet, jet}) > 0.8$, the CSVv2 discriminator of the jet must be smaller than the loose operating point, that is: the jet is “anti b-tagged”.
- The T quark candidate reconstructed as the sum of the 4-vectors of the W tagged jet and the anti b-tagged jet is required to have a mass larger than 500 GeV.

The choice of this definition of the sideband region follow the criteria of selecting events with a high number of events and low signal contamination, but with kinematic properties as close as possible to the ones of the signal region. In particular, reverting the b-tagging requirement ensures the amount of signal in the signal region is reduced and that the sidebands have a high statistics. Keeping the substructure selections intact maintains the kinematics of the event very close to that of the signal region. Several other definitions of the sideband were considered (e.g. inverting the requirements of the taggers), but the one that was chosen fulfills the aforementioned requirements the best.

The b-tagging veto uses a loose operating point instead of the medium working point because it greatly reduces the signal contamination in the sideband region, while having a limited impact on the amount of events selected in the sideband region. Two different sideband regions are defined for the two event categories, differing by the b tagging requirement on the top jet. A summary and comparison of the selections of the signal and sideband region is shown in Table 8.3 and, schematically, in Figure 8.23.

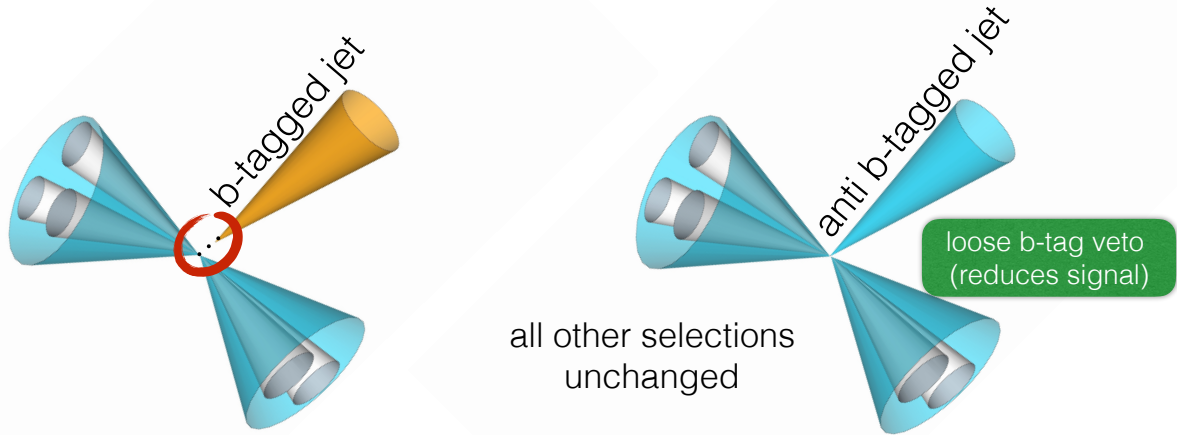


Figure 8.23: Graphical comparison of the selection criteria of the signal (left) and sideband (right) regions.

The Z' boson candidate in the sideband region is obtained by summing the 4-vectors of the top tagged jet, the W tagged jet, and the anti-tagged AK4 jet. The mass distribution of the Z' boson candidate can be seen in Figure 8.24. In these plots the simulated QCD multijet background has been scaled to match the number of events in data, taking into account the top background. The sideband regions contain about ten times more events than the signal regions. A good agreement in shape between the data and simulation distributions of the Z' mass candidate in the sidebands is observed. This is important for the solidity of the background estimate procedure.

Selection	SR 1 b tag	SB for 1 b tag	SR 2 b tag	SB for 2 b tag
1 t tag and 1 W tag	Yes	Yes	Yes	Yes
Subjet b tag on t-tagged jet	Veto	Veto	Yes	Yes
1 AK4 jet, $p_T > 100$ GeV, $\Delta R(t - /W - \text{jet}, \text{jet}) > 0.8$	Yes	Yes	Yes	Yes
b tag on AK4 jet	Yes	“loose” Veto	Yes	“loose” Veto
$m_T > 500$ GeV	Yes	Yes	Yes	Yes

Table 8.3: Comparison of the selection criteria for the signal region (SR) and the sideband region (SB).

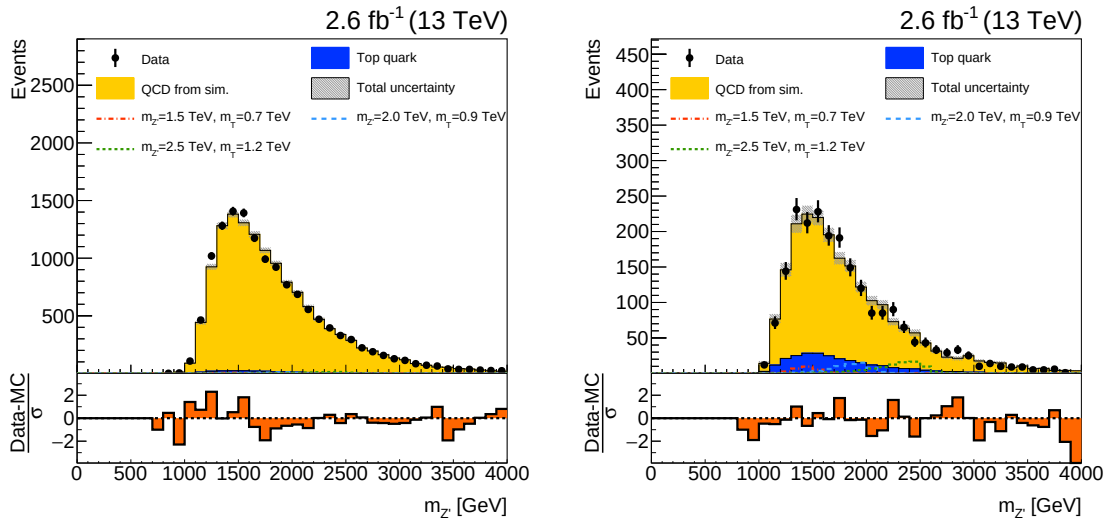


Figure 8.24: Distribution of the invariant mass of the Z' boson candidate in the sideband region. Left: top-jet not b-tagged (sideband for the 1 b-tag category), right: b-tagged top-jet (sideband for the 2 b-tag category). The QCD multijet background from simulated events has been scaled to match the number of events in data (taking into account the top background).

The goal of the procedure is to use the shape of $M_{Z'}$ in the sideband region to estimate the shape of the distribution in the signal region. A comparison of the two shapes, both normalized to one, in QCD multijet simulated events can be seen in Figure 8.25. The shapes of the signal region and the sideband region do not match perfectly, hence shape corrections must be derived.

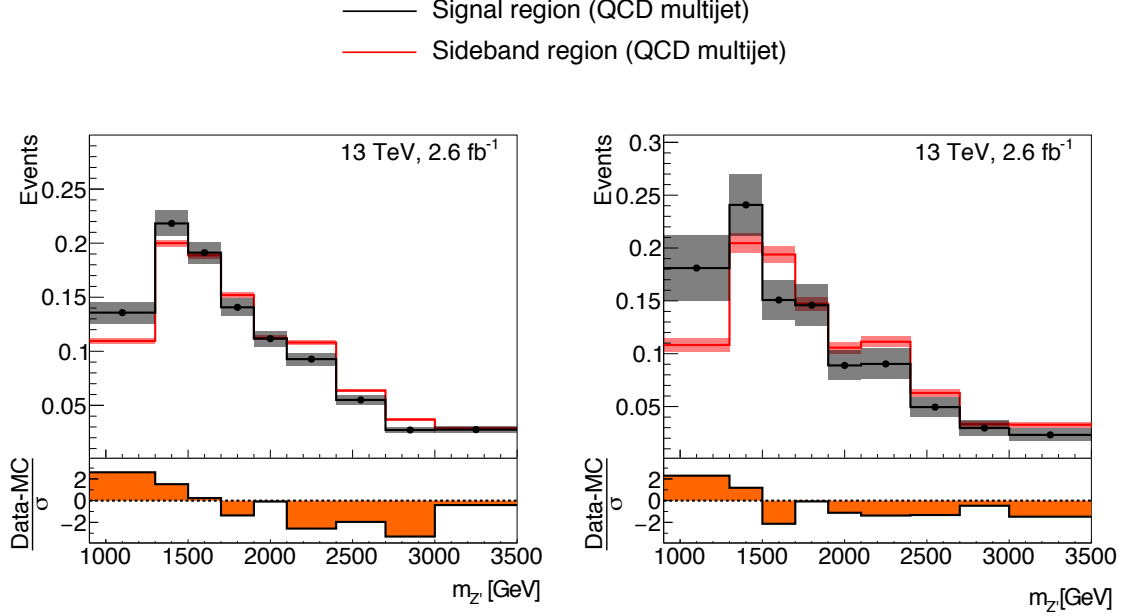


Figure 8.25: Shape comparison of the Z' candidate mass distribution in the signal and sideband region in QCD multijet simulated events. Left: top-jet not b-tagged, right: b-tagged top-jet.

The sideband-to-signal-region corrections are derived from simulated QCD multijet events. The correction will be applied to the sideband region obtained from data in order to predict the shape of the $M_{Z'}$ distribution in the signal region. Figure 8.26 shows the bin-by-bin ratio of the $M_{Z'}$ distribution in the signal region divided by the same distribution in the sideband region. The ratio is fitted with a parabola. The parabola fits the points well except for the first bin, which is however affected by large uncertainties due to the reduced statistics. The fitted function is used as the correction factor, that is, each event of the sideband region is weighted with a value corresponding to the appropriate value of the quadratic function. There are two correction functions for the two categories in the analysis.

Other fitting functional forms are explored. Fig. 8.27, for example, shows the same ratio fitted with a straight line.

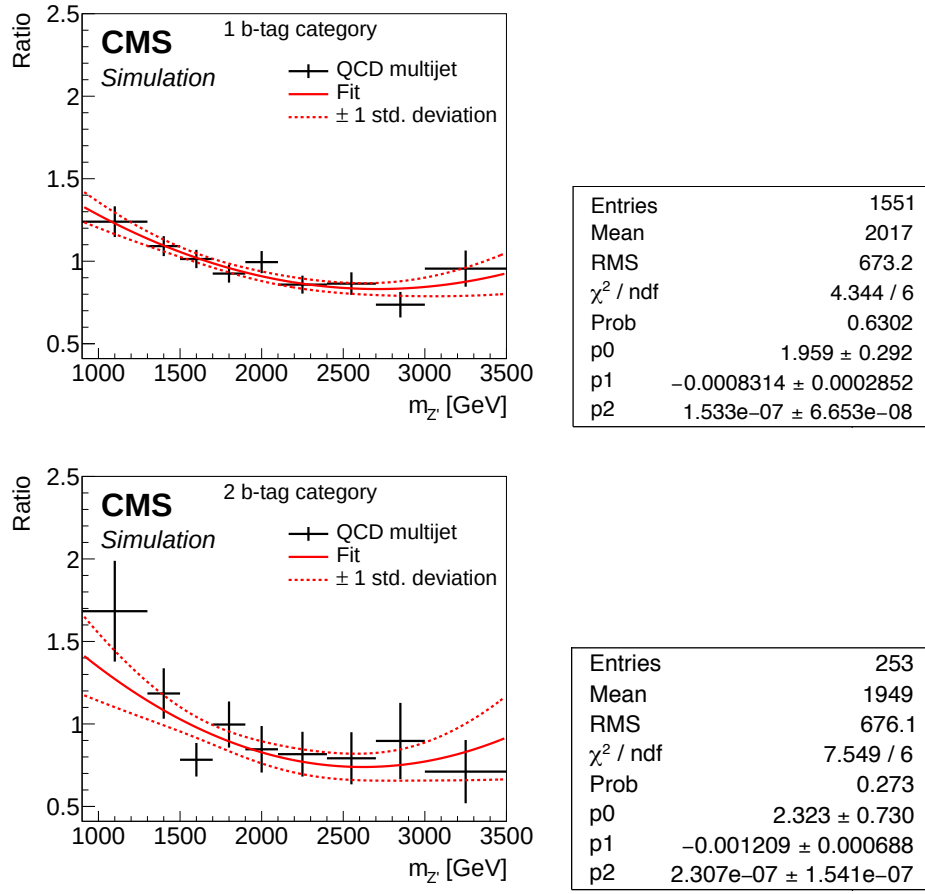


Figure 8.26: Bin-by-bin ratio of the number of events in the signal region divided by the number in the sideband region, as a function of the Z' mass, for simulated QCD multijet events. The distribution is fitted with a parabola, $w = p_0 + p_1m + p_2m^2$. Left: top jet not b-tagged, right: b-tagged top jet [120].

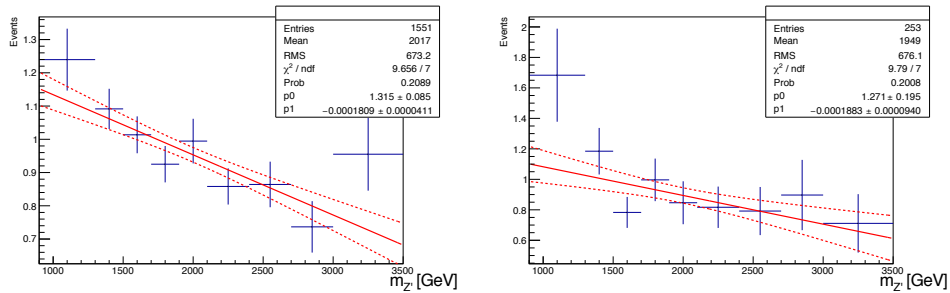


Figure 8.27: Bin-by-bin ratio of the number of events in the signal region divided by the number in the sideband region, as a function of the Z' mass, for simulated QCD multijet events. The distribution is fitted with a straight line, $w = p_0 + p_1m$. Left: top jet not b-tagged, right: b-tagged top jet.

The procedure is validated on multijet QCD simulated events. The weight function is applied to events in the sideband region in simulation and the resulting corrected shape of the $M_{Z'}$ distribution is compared with the shape in the signal region. The result is displayed in Fig. 8.28, showing a satisfying agreement between the shape of the $M_{Z'}$ distribution in the signal region and in the corrected sideband region.

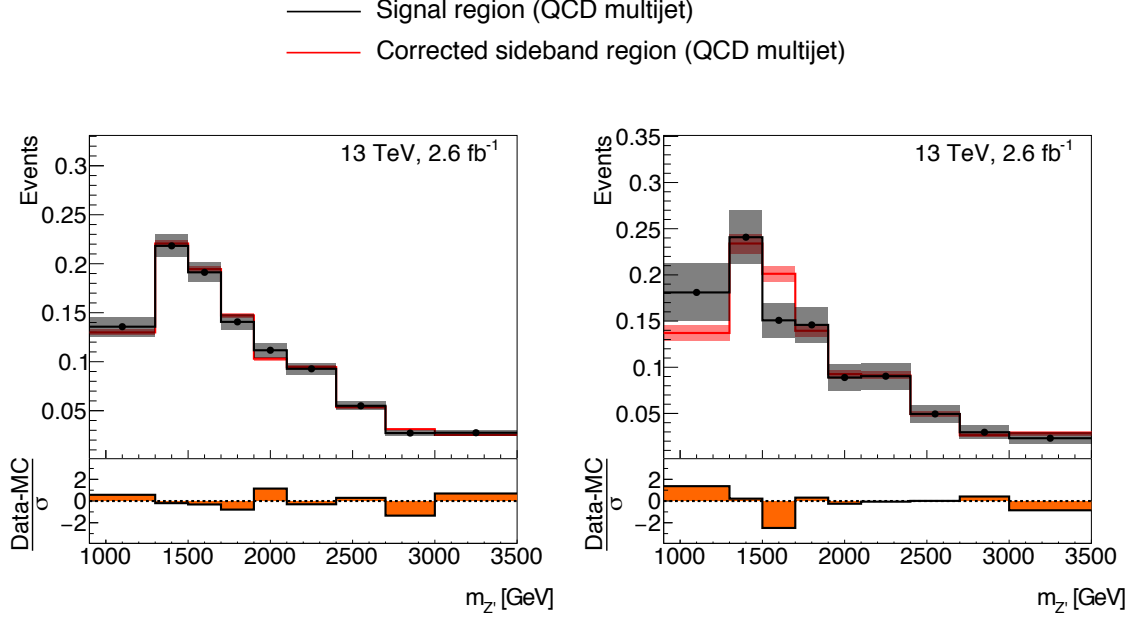


Figure 8.28: Validation test: comparison between the shape of the $M_{Z'}$ distribution in the signal region and in the corrected sideband region, in simulated events. Left: top-jet not b-tagged, right: b-tagged top-jet.

Additionally, Fig. 8.29 shows the bin-by-bin the ratio of the $M_{Z'}$ distribution between the signal region and the corrected sideband, fitted with a straight line. The slope of the line is compatible with zero.

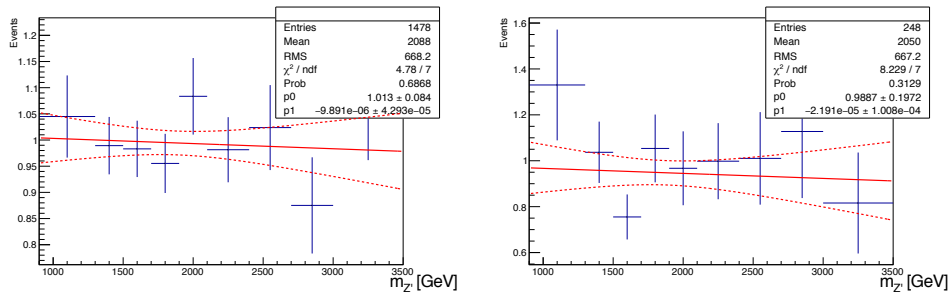


Figure 8.29: Validation test: bin-by-bin ratio between the shape of the $M_{Z'}$ distribution in the signal region and in the corrected sideband region, in simulated events. The distribution is fitted with a straight line, $w = p_0 + p_1 m$. Left: top jet not b-tagged, right: b-tagged top jet.

The impact of signal events in the background estimate is studied. Signal events normalized to a cross section of 1 pb are added to the sideband region and are propagated through

the background estimation procedure. Then a comparison is made between the background estimate with added signal events, the observed background in the signal region, and signal events observed in the signal region. This is shown in Fig. 8.30 for the two event categories and for different signal hypotheses.

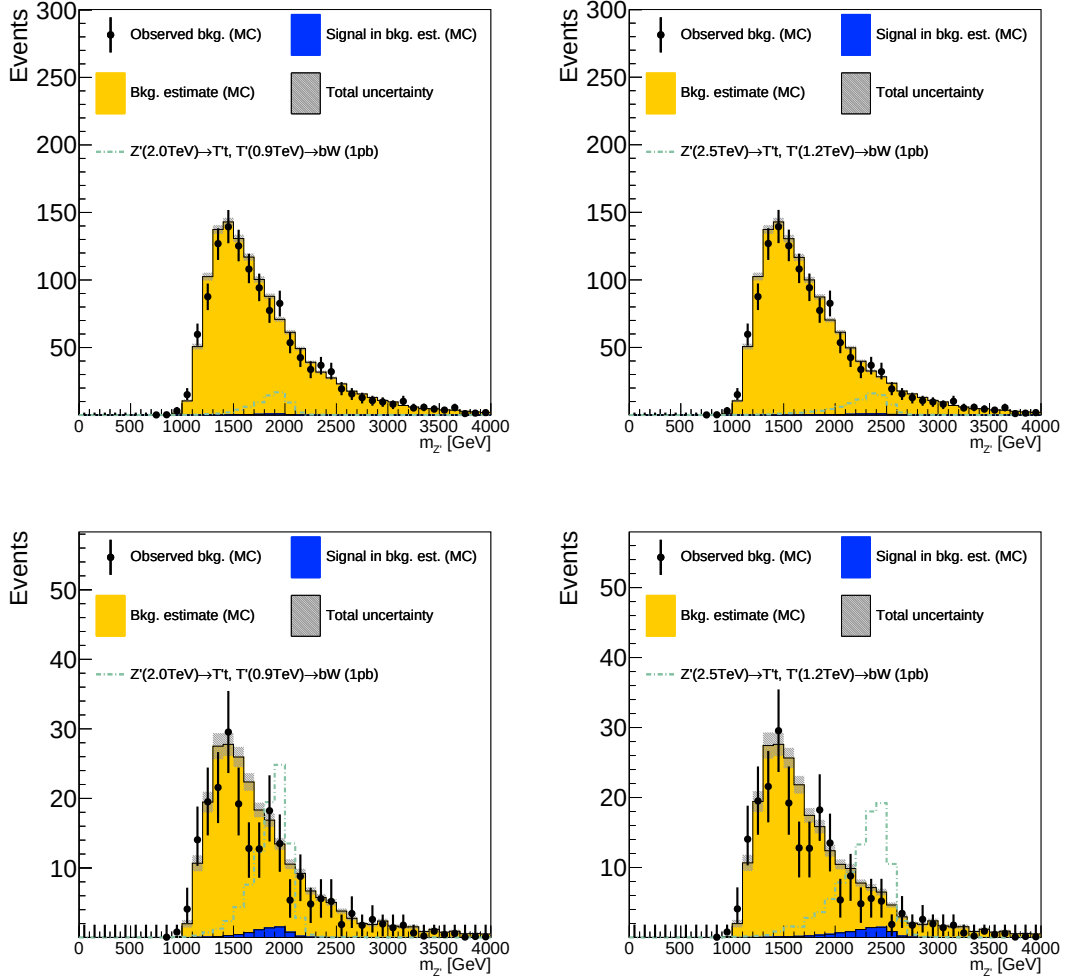


Figure 8.30: Study of the impact of signal in the background estimate, for the 1 b tag (upper panels) and 2 b tag category (lower panels). A signal with a cross section of 1 pb is added to the sideband region (different mass points in the different plots). The signal is propagated through the background estimation procedure. The yellow area is the predicted background component in the signal region coming from QCD multijet events, while the blue area is the one coming from the signal in the background estimate. The dots represent the observed background component coming from QCD multijet, while the overlaid green line is the observed signal in the signal region.

Table 8.4 reports the amount of signal events divided by the number of QCD multijet events in the sideband region used to estimate the background, for different signal hypotheses and assuming a signal cross section of 1 pb. Table 8.5 reports the ratio between the number of signal events in the background estimate and the number of signal events in the sideband region. The signal contamination is small and does not pose an issue to the background

Signal	Sig./QCD in sideband 1 btag	Sig./QCD in sideband 2 btag
$Z'(1.5 \text{ TeV}) \rightarrow Tt, T(0.7 \text{ TeV}) \rightarrow bW$	0.14%, 18 sig. ev.	1.59%, 30 sig. ev.
$Z'(2.0 \text{ TeV}) \rightarrow Tt, T(0.9 \text{ TeV}) \rightarrow bW$	0.34%, 45 sig. ev.	3.11%, 57 sig. ev.
$Z'(2.5 \text{ TeV}) \rightarrow Tt, T(1.2 \text{ TeV}) \rightarrow bW$	0.47%, 62 sig. ev.	3.80%, 60 sig. ev.

Table 8.4: Ration between the number of signal events and the number of QCD multijet events in the sideband region used to estimate the background, for different signal hypotheses and assuming a signal cross section of 1 pb .

Signal	Scaled Sig. SB / Signal SR 1 btag	Scaled Sig. SB / Signal SR 2 btag
$Z'(1.5 \text{ TeV}) \rightarrow Tt, T(0.7 \text{ TeV}) \rightarrow bW$	5.13%	7.19%
$Z'(2.0 \text{ TeV}) \rightarrow Tt, T(0.9 \text{ TeV}) \rightarrow bW$	5.99%	7.71%
$Z'(2.5 \text{ TeV}) \rightarrow Tt, T(1.2 \text{ TeV}) \rightarrow bW$	6.66%	8.71%

Table 8.5: Ratio between the number signal events in the background estimate and the number of signal events in the sideband region.

estimation procedure.

To obtain the QCD multijet estimate from the weighted sideband from data, the estimated contribution from the background due to top quark events in the sideband region is subtracted. The top quark background contribution in the sideband region is estimated using simulated events, as described in Section 8.4.1. Figure 8.31 shows a schematic overview of the background estimation procedure.

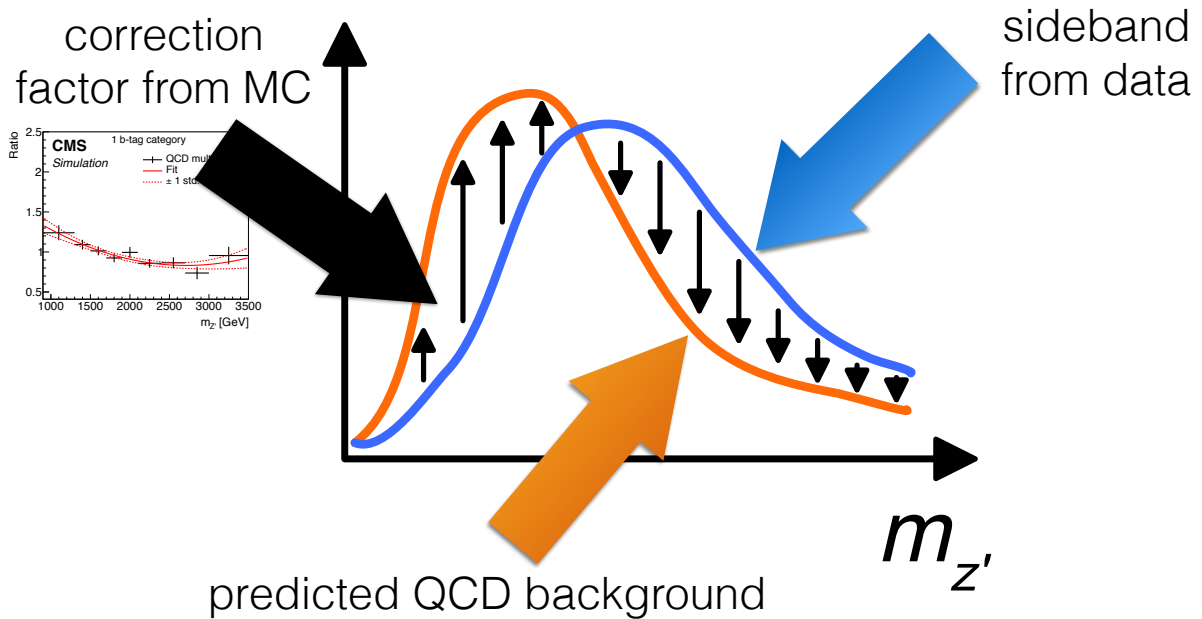


Figure 8.31: Schematic overview of the background estimation procedure.

The procedure described above predicts only the shape of the background distribution in the signal region, therefore the last point in the background estimate procedure is to determine the normalization. A maximum-likelihood estimate (MLE) fit is performed to extract the normalization. The data-driven background is fitted to data in the signal region, taking

into account the top quark background as well. The fit is performed on the distribution of $M_{Z'}$, assuming a no-signal hypothesis, and using a flat prior for the background normalization nuisance parameter. The THETA package [117] is used to perform the fit. The full scale factors obtained from the fit are $SF(1 \text{ b tags}) = 0.093 \times 0.996$ and $SF(2 \text{ b tags}) = 0.12 \times 0.996$. The fit uncertainty is 2.8%. The scale factors represent the event weight that must be applied to the events in the sideband region together with the shape-modifying weight function. As a crosscheck, the background normalization procedure is performed using the H_T variable rather than $M_{Z'}$. The differences are smaller than 1% and within the 2.8% MLE fit uncertainty. The MLE fit is also repeated in the signal hypothesis, for all the mass points available. The difference in the MLE scale factor is minimal, because the fitted signal strength is very small. To make the analysis more solid in presence of a signal, the non-top background normalization prior uncertainty is increased from 2.8% (the MLE fit uncertainty) to 50%. In this case, the presence of a signal will not affect the QCD multijet background normalization in the statistical interpretation of the results. It has been checked that this change has a negligible effect on the results of the analysis. This is due to the fact that the 1 b tag category, which is relatively signal-depleted, act as a control region helping in constraining the normalization of the background, while the 2 b tag category drives the sensitivity of the analysis to the presence of a signal.

The distribution of the reconstructed Z' mass in the signal region, using the data-driven background estimate is shown in Fig 8.32. The estimate of non-top background is able to model very well other kinematic variables in the signal region. In Fig 8.33 the distribution of the T quark candidate mass and the H_T calculated as the sum of the p_T of the three jets used for the interpretation of the event is shown. Additionally Fig. 8.34 shows the distribution of the reconstructed Z' mass in the signal region where the signal distributions are shown for events in the tH and tZ final state of the T quark.

8.5 Systematic uncertainties

Given the similarities with the analysis described in Chapter 7, this section will briefly remind the common systematic uncertainties already described in Section 7.5 and focus on the ones specific to this analysis.

- B tagging on AK4 and AK8 jets: variation of the light and heavy flavour scale factors by 1σ , applied to all simulated samples. Light and heavy flavours scale factors are correlated. The uncertainty is also correlated between AK4 and AK8 jets [68].
- Jet energy calibration: variation of $1\sigma \approx 1 - 5\%$ of the uncertainties on the calibration scale factors. The uncertainty is correlated between AK4 jets, AK8 jets, and AK8 subjets. When the calibration is varied, the jet mass is recalculated.
- Jet energy resolution (jet smearing): variation of $1\sigma \approx 2 - 8\%$ of the uncertainties on the smearing scale factors. The uncertainty is correlated between AK4 and AK8 jets.
- Top tagging scale factors: variation of $1\sigma = 0.09 - 0.19$ of the uncertainty on the top tagging scale factors.
- W tagging scale factors: variation of $1\sigma = 0.03$. The value has been derived in Ref [87]. For jets with $p_T > 300 \text{ GeV}$ the uncertainty is increased to 0.048.
- Factorization and renormalization scale (μ_F, μ_R): the two scales are varied by doubling or halving them. Six possible combinations of variations of the renormalization and

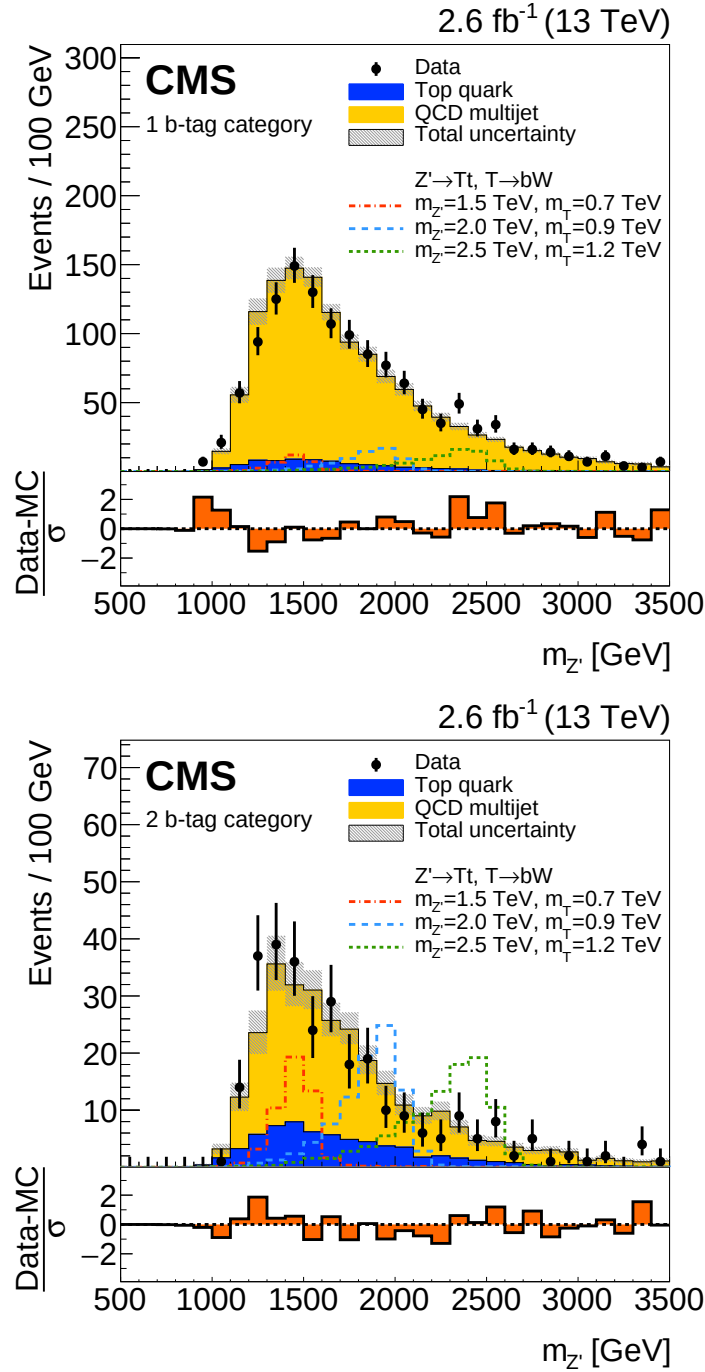


Figure 8.32: Invariant mass of the Z' candidate in the signal region with data-driven non-top background estimate. Upper panel: 1 b tag category, bottom panel: 2 b tag category [120]. The error bars shown for data are computed using Poisson statistics. The full impact of systematic and statistical errors is shown as gray bands for the distribution of background events. The combination of the total error for data and background is used to compute the uncertainty for the pull distribution in the bottom panel.

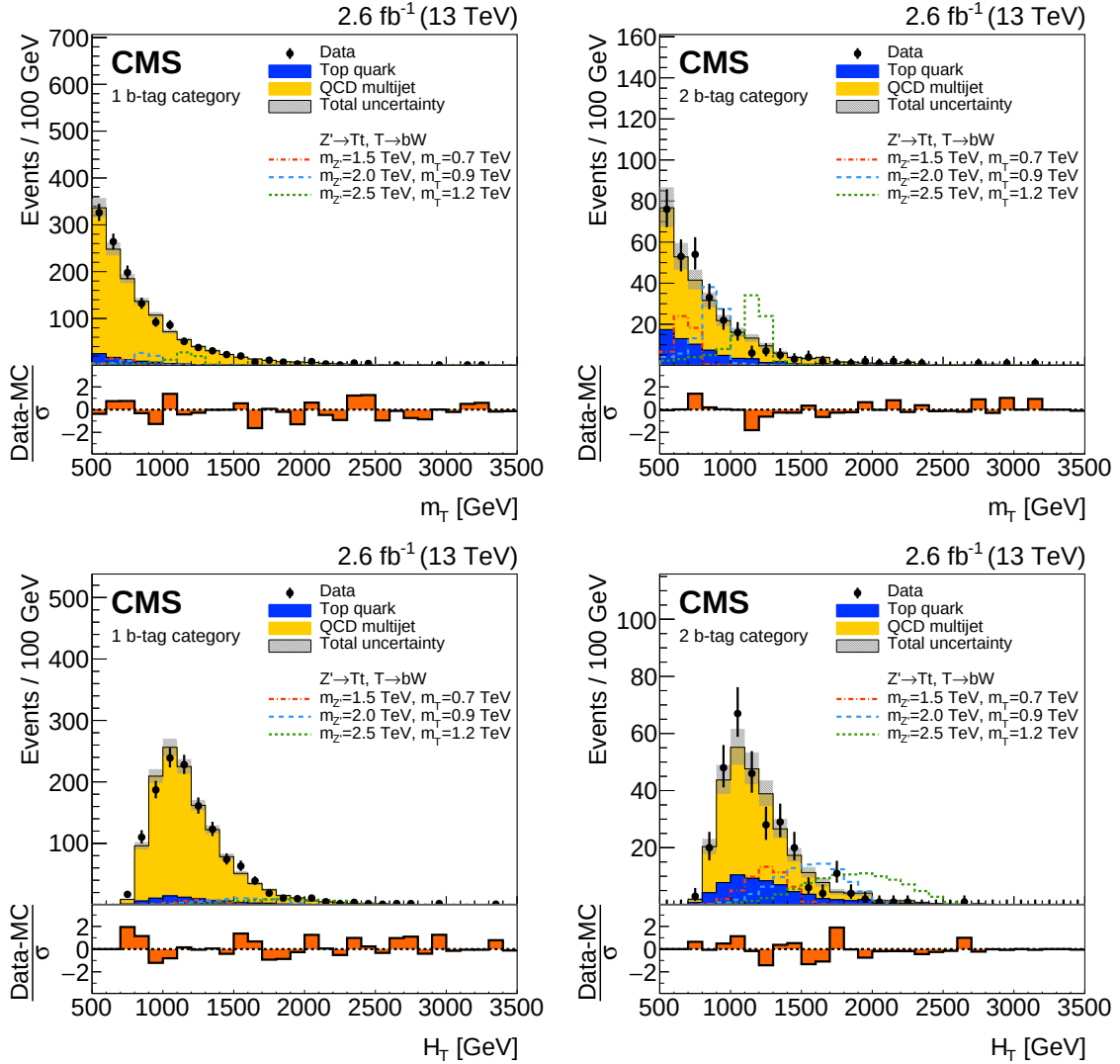


Figure 8.33: Plots of additional variables in the signal region using the data-driven background estimate [120]. First row: T quark candidate invariant mass in the 1 b tag category on the left, and the 2 b tag category on the right. Second row: H_T calculated as the sum of the p_T of the three jets used in the event in the 1 b tag category on the left, and the 2 b tag category on the right. The error bars shown for data are computed using Poisson statistics. The full impact of systematic and statistical errors is shown as gray bands for the distribution of background events. The combination of the total error for data and background is used to compute the uncertainty for the pull distribution in the bottom panel.

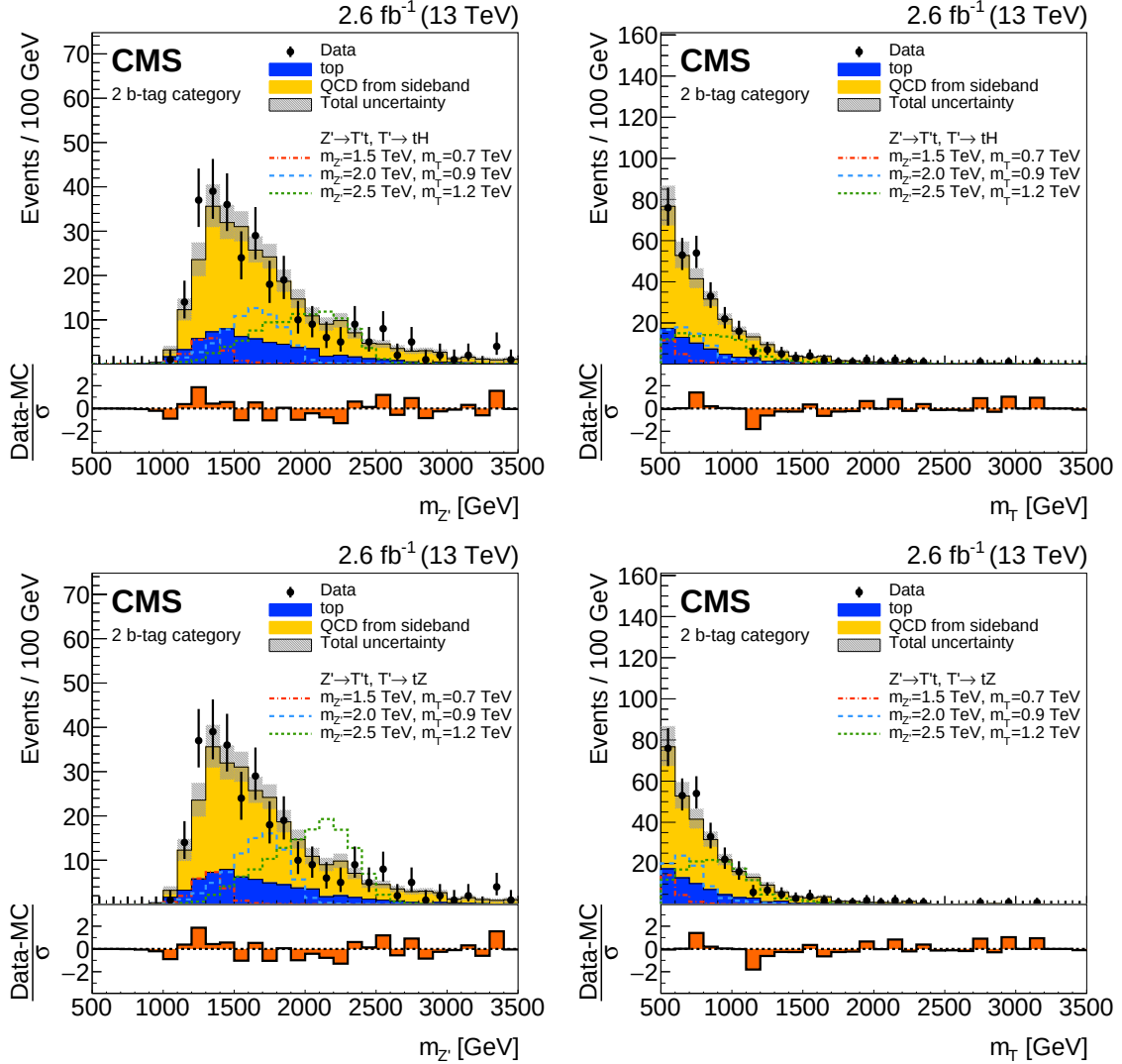


Figure 8.34: Invariant mass of the Z' candidate in the signal region with data-driven non-top background estimate. Upper panel: 1 b tag category, bottom panel: 2 b tag category. The signal distributions are shown for events in the tH and tZ final state of the T quark. The error bars shown for data are computed using Poisson statistics. The full impact of systematic and statistical errors is shown as gray bands for the distribution of background events. The combination of the total error for data and background is used to compute the uncertainty for the pull distribution in the bottom panel.

factorization scale ($\mu_{R,up} + \mu_{F,up}$, $\mu_{R,down} + \mu_{F,down}$, $\mu_{R,avg} + \mu_{F,up}$, $\mu_{R,avg} + \mu_{F,down}$, $\mu_{R,avg} + \mu_{F,up}$, $\mu_{R,avg} + \mu_{F,down}$) are considered. Subsequently, for every bin of every distribution, the envelope of the six variations is considered as the up and down variation.

- Pileup: the pileup cross section is varied from the base value of 69mb to 65.6mb and 72.5mb. The variation is propagated to the pileup reweighting procedure used for the simulated samples.
- Luminosity: a 2.3% flat rate uncertainty is applied to simulated samples. The value reflects the uncertainty on the measurement of the luminosity performed by the CMS.
- Trigger: a 3% flat rate uncertainty is applied to simulated samples (Section 8.3.1).
- Choice of the PDF values: the eigenvectors of the NNPDF 3.0 [136] set used in the simulation are varied. The variations are summed in quadrature to obtain the total uncertainty due to PDF effects.

Three different source of systematic uncertainties are associated to the non-top background estimation procedure.

- When fitting the ratio between the sideband and the signal region with a parabola, the statistical uncertainty of the fit parameters is considered as a source of systematic errors ($\approx 5 - 20\%$). Ultimately, this is related to the statistical uncertainty of the simulated samples in the procedure.
- The ratio between the sideband and the signal region is additionally fitted with a linear functional form. The resulting estimated background shape is compared with the one obtained with the standard quadratic fit, and the difference between the two distributions is taken into account as a systematic uncertainty.
- The normalization of the multijet background is determined during the limit setting procedure, therefore a prior uncertainty of 50% is assumed on the background normalization in the maximum likelihood fit to data. It has been checked the assigning such a large uncertainty to the normalization does not significantly affect the analysis results.

A schematic summary of the systematic uncertainties, the type of uncertainty (rate only or rate plus shape), and the processes affected by the uncertainty are summarized in Table 8.6. The rate impact of the uncertainties is also quoted in the table: the most relevant uncertainties are the ones associated with the QCD multijet background estimation, and with the choice of μ_R and μ_F .

The impact of the systematic uncertainties on the $M_{Z'}$ distribution for the most relevant uncertainties is shown in Appendix C.

8.6 Results

No excess of events is found in the signal region over the SM prediction, therefore limits are derived on the hypothetical production cross section of a heavy resonance decaying to Tt . The THETA [117] framework is used to set limits. The statistical model used is described in Section 7.7. The systematic uncertainties included are reported in Section 8.5 of the thesis. Table 8.7 reports the number of selected events in the signal region categories for data, and the SM backgrounds considered.

Systematic uncertainty	Rate 1 b tag	Rate 2 b tag	Type	Process
B-tagging efficiency	9–14%	12–17%	rate + shape	t bkg and signal
Top-tagging efficiency	8–14%	8–14%	rate + shape	t bkg and signal
W-tagging efficiency	0.1–6%	0.1–6%	rate + shape	t bkg and signal
Jet energy scale	0.4–10%	0.1–8%	rate + shape	t bkg and signal
Jet energy resolution	0–2%	0–2%	rate + shape	t bkg and signal
Integrated luminosity	2.3%	2.3%	rate	t bkg and signal
Trigger efficiency	3%	3%	rate	t bkg and signal
PDFs	3–9%	3–8%	rate + shape	t bkg and signal
Pileup reweighting	0–3%	0.1–2 %	rate + shape	t bkg and signal
μ_R, μ_F	3–44%	1–41%	rate + shape	t bkg and signal
Sideband corr. (fit unc.)	4%	9%	rate + shape	QCD multijet
Sideband corr. (fit form)	1%	2%	rate + shape	QCD multijet
Sideband norm.	50%	50%	rate	QCD multijet

Table 8.6: Summary of the systematic uncertainties considered in the analysis. The rate ranges refer to both background and signal events [120].

Sample	SR 1 b tag	SR 2 b tag
QCD multijet	1227^{+59}_{-59}	222^{+22}_{-22}
SM top quark	81^{+31}_{-23}	66^{+23}_{-18}
Total background	1308^{+67}_{-63}	288^{+32}_{-29}
Data	1307	289

Table 8.7: Number of selected events in the two event categories of the signal region of the analysis [120]. The uncertainties include both statistical and systematic components. The QCD multijet background is estimated with a data-driven procedure, and the SM top quark production is estimated is from simulation. The yields have been normalized to give the observed total numbers of events with a MLE fit. The integrated luminosity of the data analyzed amounts to 2.6 fb^{-1} .

Table 8.8 shows the expected and observed upper cross section limits on the production of a Z' boson that decays to Tt at 95% of confidence level for different Z' boson and T quark mass hypotheses. Three different hypotheses for the decay of the T quark are considered: 100% branching fraction into bW , tH , or tZ . Different configurations of the Z' boson and T quark widths (up to 10%), and T chirality are considered for a single value of the masses of the Z' boson and the T quark, as shown in Figure 8.22. Some differences can be observed in the distributions, however the impact on the cross section limits is found to be negligible in the three cases, because the detector resolution is close to 20%, and the chirality of the T quark has a reduced impact on the analysis acceptance in all-jet final states.

Observed and expected limits on the production cross section of $Z' \rightarrow Tt$ are provided at 95% of confidence level for all possible branching fraction combinations of the T quarks in Figs 8.35, 8.36. The two-dimensional plots show the cross section limit in function of the branching fraction of the T quark to bW on the x-axis and to tH on the y-axis. The observed limits vary from 0.13 to 10 pb, the observed and expected limits agree within two standard deviations. Figure D.1 shows the observed limits arranged on a bidimensional grid in function of the masses of the Z' boson and the T quark.

Fixing the mass of the T quark to 1.2 TeV, one-dimensional cross section limits are derived in Fig. 8.37 for branching fractions of the T quark to tH/tZ of 50/50% (left) or to $bW/tH/tZ$ of 50/25/25% (right). The limits are compared to the expectation of the composite Higgs boson model [38] (left) and to the warped-extra dimension model [44] (see also Sections 1.4.1 and 1.5.1). The models and the benchmark values for the model parameters used in this comparison are described in Sections 1.4.1 and 1.5.1. Two changes in slope in the theoretical cross sections can be observed around $m_{Z'} = 1.6$ and 2.4 TeV; they are caused by the Tt and T quark pair decay channels becoming kinematically allowed, respectively. Additionally, the warped-extra dimension prediction shows another change of slope around 2 TeV, caused by the $T_{5/3}T_{5/3}$ decay channel becoming kinematically allowed. The $T_{5/3}$ vector-like-quark is predicted to have a mass of circa 1 TeV for a T quark mass of 1.2 TeV and the model parameters considered for this benchmark. For some values of the mass of the heavy resonance, the width of the resonance is computed to be larger than 10% in the benchmark theory predictions. The increase of the total width of the resonance is caused by additional decay channels becoming kinematically allowed. In these cases the simulated samples do not reproduce the resonance width predicted by the theory models, therefore the cross section values cannot be reliably considered for the comparison and are drawn with a dashed line in Fig. 8.37. The theory predictions for the cross section and width of the G^* and ρ_L^0 resonances for all the mass configurations considered can be found in Table 8.9. It can be concluded that the analysis has no sensitivity to the composite Higgs model [38], but is sensitive to the extra-dimension model [44], however more data is needed to provide mass exclusion limits and to exclude specific scenarios. This is the first result obtained with LHC data on resonances decaying to a top quark and a vector-like T quark.

8.7 Conclusions

The first search for a massive vector resonances decaying to a top quark and a vector-like T quark is performed in the all-jet final state, using 2.6 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ collected by CMS during the year 2015. The analysis uses jet-substructure techniques, including top- and W -tagging algorithms, and subjet b tagging. The top- and W -tagging algorithms are based on the N -subjettiness variable and the Soft-Drop algorithm. The QCD multijet background is estimated in data from a sideband region modified with a simulation-based correction function. The top quark background, including $t\bar{t}$ and single-top production, is

$\mathcal{B}(T \rightarrow bW) = 1$							
$m_{Z'}$ [GeV]	m_T [GeV]	Observed [pb]	Expected [pb]				
			-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
1500	700	0.73	0.32	0.48	0.67	1.0	1.6
1500	900	1.5	0.64	0.94	1.5	2.2	3.7
1500	1200	8.6	3.7	5.2	7.8	13	22
2000	900	0.19	0.17	0.24	0.36	0.56	0.90
2000	1200	0.27	0.24	0.33	0.49	0.76	1.3
2000	1500	0.96	0.82	1.2	1.9	3.0	5.4
2500	1200	0.29	0.10	0.15	0.24	0.39	0.64
2500	1500	0.30	0.11	0.16	0.24	0.39	0.65

$\mathcal{B}(T \rightarrow tH) = 1$							
$m_{Z'}$ [GeV]	m_T [GeV]	Observed [pb]	Expected [pb]				
			-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
1500	700	4.0	0.98	1.4	2.1	3.3	5.8
1500	900	3.2	0.76	1.0	1.6	2.6	4.2
1500	1200	9.4	2.6	3.6	5.6	9.3	19
2000	900	0.53	0.39	0.55	0.84	1.4	2.3
2000	1200	0.53	0.36	0.52	0.79	1.2	2.2
2000	1500	0.60	0.50	0.67	0.99	1.6	2.9
2500	1200	0.24	0.24	0.34	0.52	0.83	1.5
2500	1500	0.23	0.21	0.31	0.49	0.81	1.3

$\mathcal{B}(T \rightarrow tZ) = 1$							
$m_{Z'}$ [GeV]	m_T [GeV]	Observed [pb]	Expected [pb]				
			-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
1500	700	3.1	0.84	1.2	1.8	2.9	4.7
1500	900	2.8	0.77	1.1	1.6	2.5	4.3
1500	1200	3.4	1.3	1.8	2.7	4.2	6.4
2000	900	0.37	0.30	0.41	0.61	0.97	1.8
2000	1200	0.30	0.23	0.34	0.50	0.80	1.3
2000	1500	0.32	0.26	0.37	0.55	0.85	1.7
2500	1200	0.16	0.14	0.21	0.31	0.52	0.92
2500	1500	0.13	0.12	0.17	0.27	0.45	0.77

Table 8.8: Table of expected and observed upper cross section limits on the production of a Z' boson that decays to Tt at 95% of confidence level, assuming the T quark decays exclusively to bW (top table), tH (middle table), or tZ (bottom table) [120]. The integrated luminosity of the data analyzed amounts to 2.6 fb^{-1} .

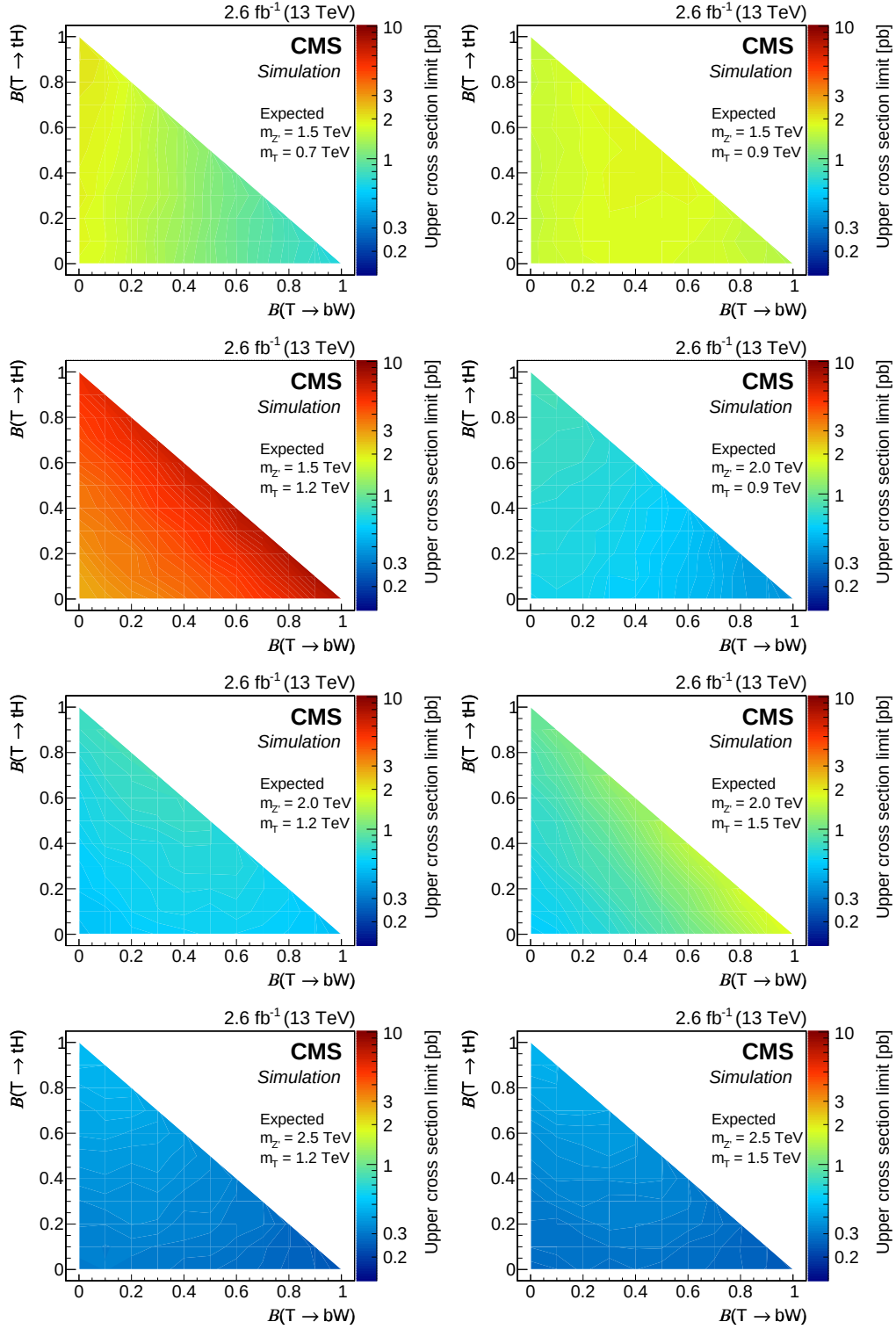


Figure 8.35: Expected cross section limits for $Z' \rightarrow Tt$ for different Z' boson and T quark mass hypotheses, as a function of the branching fraction of the T quark to bW and tH [120].

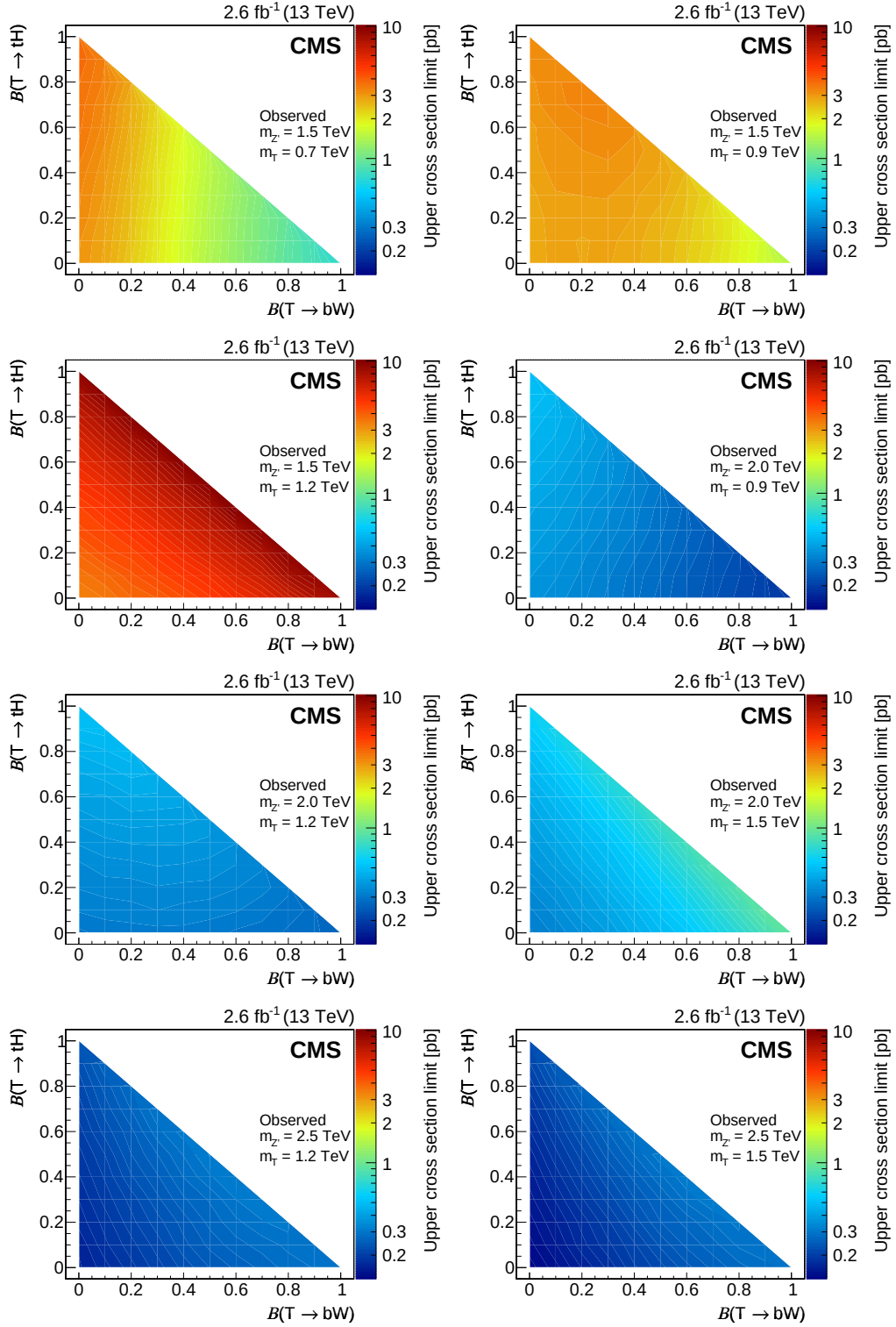


Figure 8.36: Observed cross section limits for $Z' \rightarrow Tt$ for different Z' boson and T quark mass hypotheses, as a function of the branching fraction of the T quark to bW and tH [120].

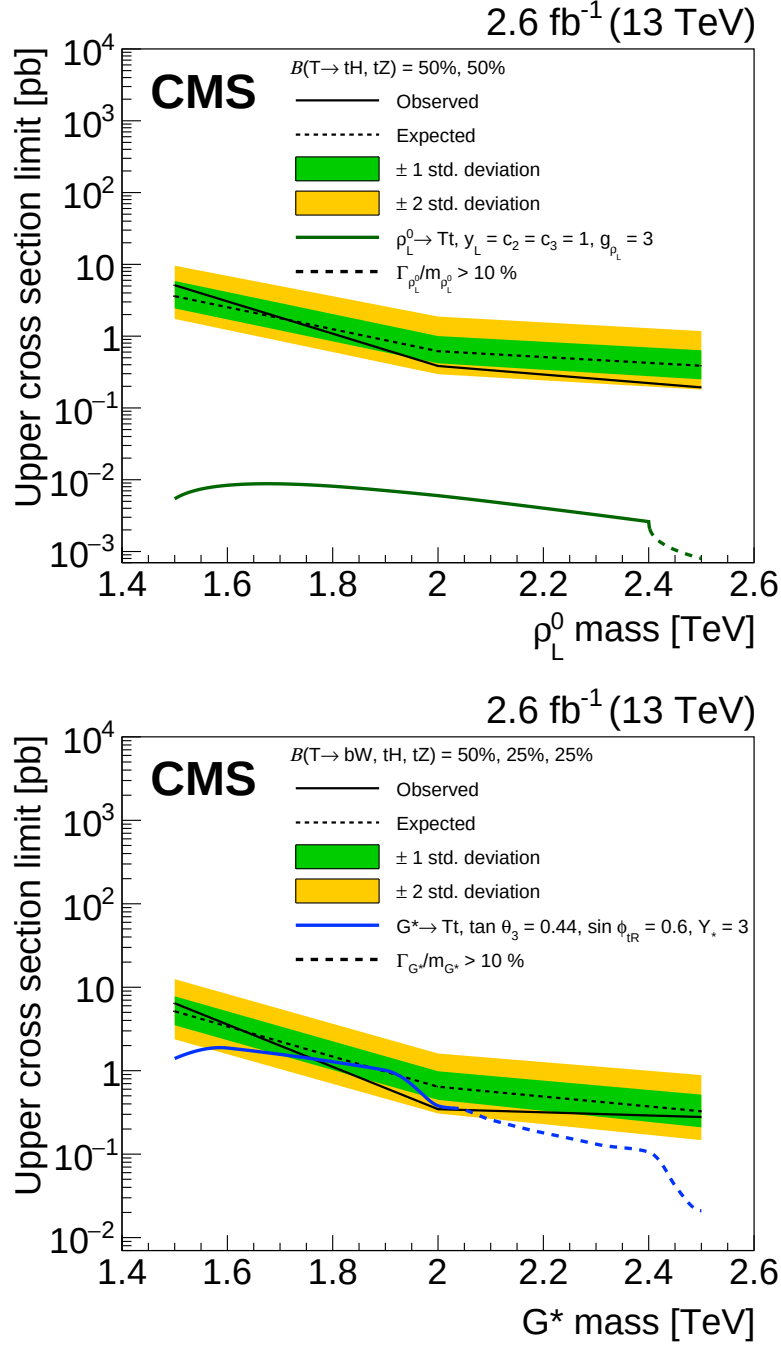


Figure 8.37: Cross section limits at 95% CL as a function of the heavy resonance mass for a mass of the T quark of 1.2 TeV, assuming branching fractions of the T quark to tH/tZ of 50/50% (left) or to bW/tH/tZ of 50/25/25% (right) [120]. The limits are compared to the cross section times branching fraction of a heavy resonance $\rho_L^0 \rightarrow Tt$ in a composite Higgs boson model [38] (left) or a heavy gluon $G^* \rightarrow Tt$ in a warped extra-dimension model [44] (right). The predictions from the theory models are shown as dashed lines where the width of the resonance is larger than 10% of its mass.

$m_{Z'}$ [GeV]	m_T [GeV]	$\sigma(pp \rightarrow G^* \rightarrow Tt)$ [fb]	$\Gamma(G^*)$ [GeV]	$\sigma(pp \rightarrow \rho_L^0 \rightarrow Tt)$ [fb]	$\Gamma(\rho_L^0)$ [GeV]
1500	700	280	740	12	410
1500	900	1600	130	28	91
1500	1200	1400	47	5.9	51
2000	900	65	1100	2.7	610
2000	1200	380	170	6.6	120
2000	1500	430	74	3.1	76
2500	1200	21	1000	0.88	560
2500	1500	110	200	1.9	150

Table 8.9: Theory predictions for the cross section and width of the G^* [44] and ρ_L^0 [38] resonances for all the mass configurations considered in the analysis. The models and the benchmark values for the model parameters used in this comparison are described in Sections 1.4.1 and 1.5.1.

estimated using simulated events.

No excess is observed in data with respect to the standard model background expectation, therefore upper limits are set on the production cross section of hypothetical heavy mass resonances. The cross section limits are compared to the cross sections of a spin-1 resonance in a composite Higgs boson model [38] and a Kaluza-Klein gluon in a warped extra-dimension model [44], for benchmark values of the model parameters. With the amount of data analyzed, the analysis has no sensitivity to the composite Higgs model, but the prediction for the warped extra-dimension model is just at the same level of the upper cross section limit for masses of the resonance around 1.8 TeV. From a model-independent point of view, a Z' boson decaying to a top and a T vector-like quark is excluded at 95% CL, with upper limits on production cross sections ranging from 0.13 to 10 pb, depending on the mass of the Z' boson and the T quark.

Conclusions

Two searches for physics beyond the Standard Model with data collected by the CMS Experiment at the LHC have been presented. A search for heavy resonances decaying to $t\bar{t}$ in the all-jet final state has been carried out with 20 fb^{-1} of proton-proton collisions at 8 TeV. The analysis is interpreted in the frameworks of topcolor Z' models and Randal-Sundrum theories. The combination of the analysis with other searches for resonant $t\bar{t}$ production allows to exclude resonances with masses between 2.4 and 2.9 TeV, depending on the model. The first search has been presented for resonances decaying to a top quark and a vector-like T quark, using 2.6 fb^{-1} of pp collisions at 13 TeV. A heavy resonance decaying to Tt is excluded with production cross sections from 0.13 to 10 pb, depending on the masses of the heavy particle and the T quark. The search is sensitive to theories with a warped extra-dimension, but more data is needed to exclude specific scenarios. Jet substructure techniques have been used to identify the high Lorentz-boost decays of the top quark and the W boson, greatly increasing the sensitivity of the searches. In particular, the HEPTOPTAGGER algorithm and subjet b tagging have been studied, and their performance has been assessed in data.

While the searches presented could not find signs of new physics, stringent limits are placed on the production of these resonances. Moreover, the rapidly increasing amount of data collected by the LHC makes it essential to extend the analyses presented to the larger dataset, in the future, in search for resonances with a higher mass and a lower cross section. The searches presented join the very broad corpus of searches for new phenomena in physics of the fundamental structure of matter. It is essential to continue the development of new detector and accelerator technologies to further increase the sensitivity to BSM phenomena, and eventually confirm one of the many theories predicting new sectors of particle physics.

Appendix

A

Tables of data and simulated samples

This section collects the samples used in the analyses presented in this thesis.

Dataset	Integrated luminosity (pb^{-1})
/SingleMu/Run2012A-13Jul2012-v1/AOD	854.7
/SingleMu/Run2012A-recover-06Aug2012-v1/AOD	82.6
/SingleMu/Run2012B-13Jul2012-v1/AOD	4828
/SingleMu/Run2012C-24Aug2012-v1/AOD	501.9
/SingleMu/Run2012C-PromptReco-v2/AOD	6796
/SingleMu/Run2012D-PromptReco-v1/AOD	6684

Table A.1: Processed datasets and integrated luminosities used for the analysis described in Chapter 5. The sample name is reported using the CMS dataset convention.

SM process	sample name	cross section (pb ⁻¹)
tt (all hadronic)	/TTJets_HadronicMGDecays_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A*-v1/AODSIM	115.92
tt (semi-leptonic channel)	/TTJets_SemiLeptMGDecays_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-ext-v1/AODSIM	103.7
tt (dileptonic channel)	/TTJets_FullLeptMGDecays_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM	26.18
Drell Yan $10 < m_{\ell\ell} < 50$ GeV	/DY*JetstOLL_M-50_TuneZ2Sstar_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	969.3
single t (s-channel)	/T_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3.79
single t (t-channel)	/T_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	56.4
single t (tW-channel)	/T_tW-channel_DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	11.1
single $\bar{t}$ (s-channel)	/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1.76
single $\bar{t}$ (t-channel)	/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	30.7
single $\bar{t}$ (tW-channel)	/Tbar_tW-channel_DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	11.1
WW	/WW_TuneZ2star_8TeV_pythia6-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	54.838
WZ	/WZ_TuneZ2star_8TeV_pythia6-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	33.21
ZZ	/ZZ_TuneZ2star_8TeV_pythia6-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	8.059
W+Jets	/W+JetstOLL_TuneZ2star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	9725
QCD μ -enriched $15 < p_T < 20$ GeV	/QCD_Pt-15to20_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM	7.022×10^8
QCD μ -enriched $20 < p_T < 30$ GeV	/QCD_Pt-20to30_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	2.87×10^8
QCD μ -enriched $30 < p_T < 50$ GeV	/QCD_Pt-30to50_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	6.609×10^7
QCD μ -enriched $50 < p_T < 80$ GeV	/QCD_Pt-50to80_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	8082000.0
QCD μ -enriched $80 < p_T < 120$ GeV	/QCD_Pt-80to120_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1024000.0
QCD μ -enriched $120 < p_T < 170$ GeV	/QCD_Pt-120to170_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	157800.0
QCD μ -enriched $170 < p_T < 300$ GeV	/QCD_Pt-170to300_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	34020.0
QCD μ -enriched $300 < p_T < 470$ GeV	/QCD_Pt-300to470_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1757.0
QCD μ -enriched $470 < p_T < 600$ GeV	/QCD_Pt-470to600_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	115.2
QCD μ -enriched $600 < p_T < 800$ GeV	/QCD_Pt-600to800_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27.01
QCD μ -enriched $800 < p_T < 1000$ GeV	/QCD_Pt-800to1000_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3.57
QCD μ -enriched $p_T > 1000$ GeV	/QCD_Pt-1000_MuEnrichedPt5_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	0.774

Table A.2: Standard Model background simulated samples and cross sections used for the analysis described in Chapter 5. The sample name is reported using the CMS dataset convention.

Dataset
/TT_CT10_TuneZ2star_8TeV-powheg-tauola/
/QCD_Pt-15to3000_TuneZ2star_Flat_8TeV_pythia6/

Table A.3: Processed datasets used for the analysis described in Chapter 6. The sample name is reported using the CMS dataset convention.

Sample	Dataset
$t\bar{t}$	/TT_CT10_TuneZ2star_8TeV-powheg-tauola/
$\text{QCD } 250 < H_T < 500 \text{ GeV}$	/QCD_HT-250To500_TuneZ2star_8TeV-madgraph-pythia6/
$\text{QCD } 500 < H_T < 1000 \text{ GeV}$	/QCD_HT-500To1000_TuneZ2star_8TeV-madgraph-pythia6/
$\text{QCD } 1000 < H_T < \text{Inf GeV}$	/QCD_HT-1000ToInf_TuneZ2star_8TeV-madgraph-pythia6/

Table A.4: Simulated samples used in the analysis presented in Chapter 7, using the CMS convention.

Sample	Dataset	Mass	Cross section
RS KK Gluon	/Pythia8_RSGLuonToTTbar_M_1000_8TeV	1 TeV	6.299 pb
RS KK Gluon	/Pythia8_RSGLuonToTTbar_M_1500_8TeV	1.5 TeV	0.828 pb
RS KK Gluon	/Pythia8_RSGLuonToTTbar_M_2000_8TeV	2 TeV	0.182 pb
RS KK Gluon	/Pythia8_RSGLuonToTTbar_M_3000_8TeV	3 TeV	0.021 pb
RS KK Gluon	/Pythia8_RSGLuonToTTbar_M_4000_8TeV	4 TeV	0.005 pb
Narrow Z'	/ZPrimeToTTJets_M1000GeV_W10GeV	1 TeV	–
Narrow Z'	/ZPrimeToTTJets_M1500GeV_W15GeV	1.5 TeV	–
Narrow Z'	/ZPrimeToTTJets_M2000GeV_W20GeV	2 TeV	–
Narrow Z'	/ZPrimeToTTJets_M3000GeV_W30GeV	3 TeV	–
Narrow Z'	/ZPrimeToTTJets_M4000GeV_W40GeV	4 TeV	–
Wide Z'	/ZPrimeToTTJets_M1000GeV_W100GeV	1 TeV	–
Wide Z'	/ZPrimeToTTJets_M1500GeV_W150GeV	1.5 TeV	–
Wide Z'	/ZPrimeToTTJets_M2000GeV_W200GeV	2 TeV	–
Wide Z'	/ZPrimeToTTJets_M3000GeV_W300GeV	3 TeV	–
Wide Z'	/ZPrimeToTTJets_M4000GeV_W400GeV	4 TeV	–

Table A.5: Simulated signal samples used in the analysis presented in Chapter 7, listed using the CMS convention.

Dataset	Luminosity
/Jet/Run2012A-22Jan2013-v1/AOD	1056 pb ⁻¹
/JetHT/Run2012B-22Jan2013-v1/AOD	4369 pb ⁻¹
/JetHT/Run2012C-22Jan2013-v1/AOD	7051 pb ⁻¹
/JetHT/Run2012D-22Jan2013-v1/AOD	7102 pb ⁻¹
Total integrated luminosity ("jet" data stream)	19.7 fb ⁻¹
/MultiJet1Parked/Run2012B-05Nov2012-v2/AOD	2942 pb ⁻¹
/MultiJet1Parked/Run2012C-part1_05Nov2012-v2/AOD	7051 pb ⁻¹
/MultiJet1Parked/Run2012C-part2_05Nov2012-v2/AOD	
/MultiJet1Parked/Run2012D-part1_10Dec2012-v1/AOD	
/MultiJet1Parked/Run2012D-part2_17Jan2013-v1/AOD	7102 pb ⁻¹
/MultiJet1Parked/Run2012D-part2_PixelRecover_17Jan2013-v1/AOD	
Total integrated luminosity (parked "multijet" data stream)	18.3 fb ⁻¹

Table A.6: Primary datasets used in the analysis presented in Chapter 7, including the corresponding integrated luminosity of each dataset. The datasets are listed using the CMS dataset convention.

Dataset name	Luminosity (pb ⁻¹)
/JetHT/Run2015C_25ns-16Dec2015-v1/MINIAOD	13.23
/JetHT/Run2015D-16Dec2015-v1/MINIAOD	2601.62

Table A.7: Data samples used in the analysis presented in Chapter 8.

Sample	Dataset
Z'(1500GeV) → Tt, T(700GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-1500Nar.MTp-700Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(900GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-1500Nar.MTp-900Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1200GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1200GeV, RH) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2000Nar.MTp-1200Nar.RH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1200GeV, 10% <i>width</i>) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1500GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2000Nar.MTp-1500Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(900GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2000Nar.MTp-900Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV, 10% <i>width</i>) → Tt, T(1200GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2500GeV) → Tt, T(1200GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2500Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2500GeV) → Tt, T(1500GeV) → Zt	/ZprimeToPrimeT.TprimeToZT.MZp-2500Nar.MTp-1500Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1200GeV, RH) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1500GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-2000Nar.MTp-1500Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(900GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-2000Nar.MTp-900Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV, 10% <i>width</i>) → Tt, T(1200GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2500GeV) → Tt, T(1200GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-2500Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2500GeV) → Tt, T(1500GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-2500Nar.MTp-1500Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(1200GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-1500Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1500GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-2000Nar.MTp-1500Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV, 10% <i>width</i>) → Tt, T(1200GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2500GeV) → Tt, T(1200GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-2500Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2500GeV) → Tt, T(1500GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-2500Nar.MTp-1500Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(1200GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-1500Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(700GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-1500Nar.MTp-700Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(900GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-1500Nar.MTp-900Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1200GeV) → Wb	/ZprimeToPrimeT.TprimeToWB.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1200GeV, RH) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-2000Nar.MTp-1200Nar.RH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(2000GeV) → Tt, T(1200GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-2000Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(1200GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-1500Nar.MTp-1200Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(700GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-1500Nar.MTp-700Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/
Z'(1500GeV) → Tt, T(900GeV) → Ht	/ZprimeToPrimeT.TprimeToHT.MZp-1500Nar.MTp-900Nar.LH.TuneCUETP8M1.13TeV-madgraphMLM-pythia8/

Table A.8: List of the simulated signal samples used in the analysis described in Chapter 8.

Systematic uncertainties templates for the search for resonant $t\bar{t}$ production at 8 TeV

In this section, the templates used to define the $\pm 1\sigma$ shape uncertainties are shown for the most relevant systematic uncertainties. The distribution of $m_{t\bar{t}}$ is shown for the QCD multijet or the SM $t\bar{t}$ backgrounds, depending on the uncertainty.

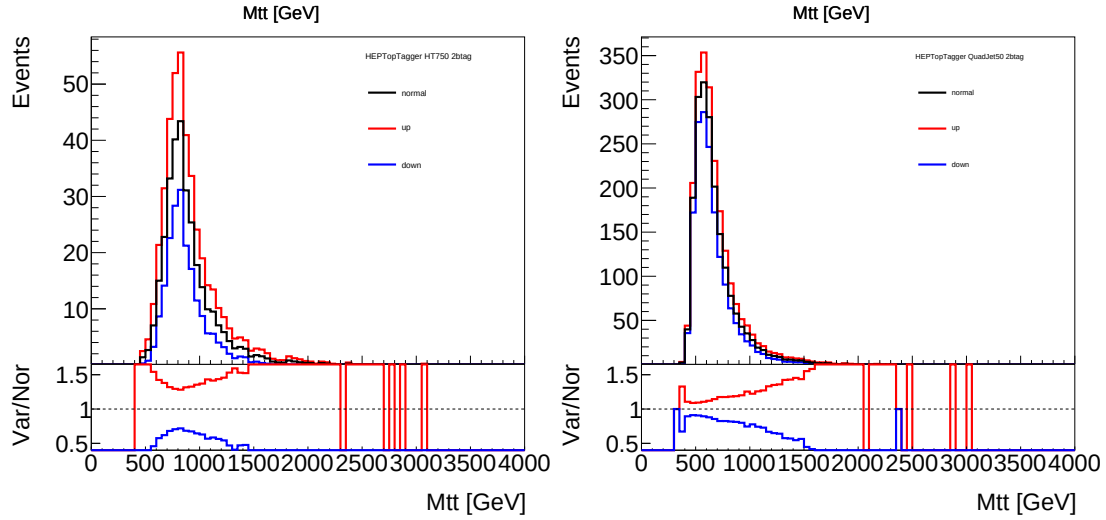


Figure B.1: Impact of the mistag matrix uncertainty on the QCD multijet predicted background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

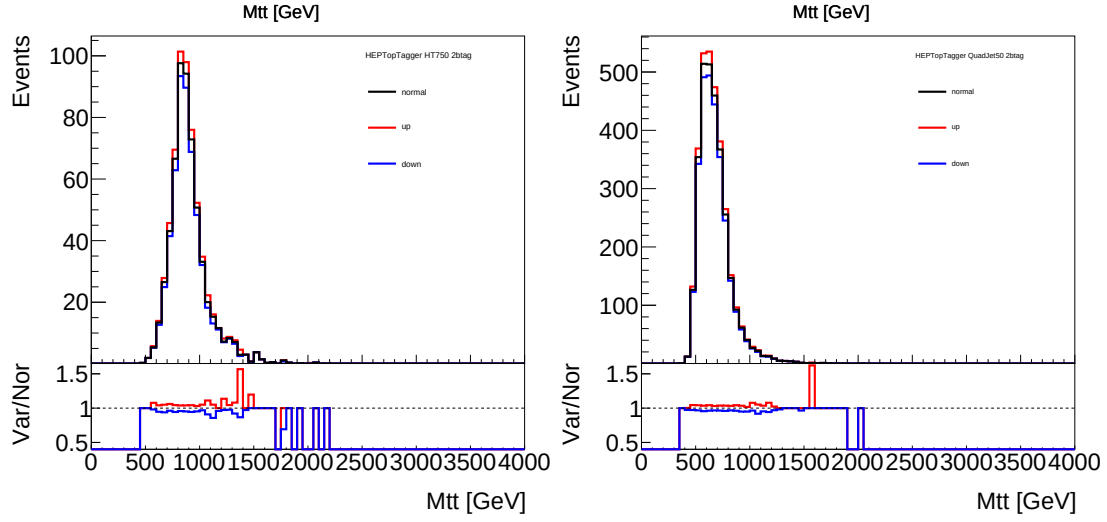


Figure B.2: Impact of the b-tagging efficiency uncertainty on the SM $t\bar{t}$ background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

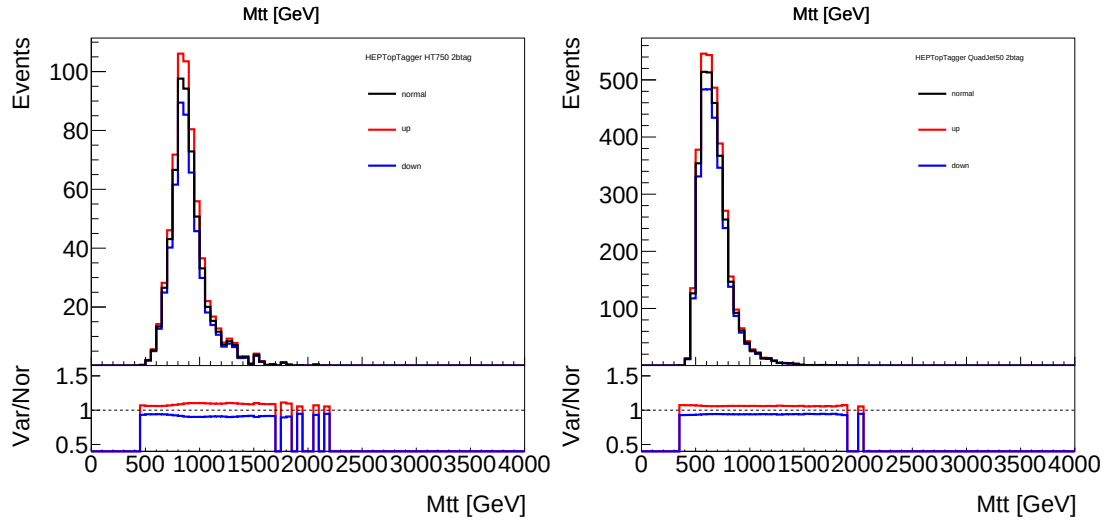


Figure B.3: Impact of the HEPTopTAgGER scale factor uncertainty on the SM $t\bar{t}$ background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

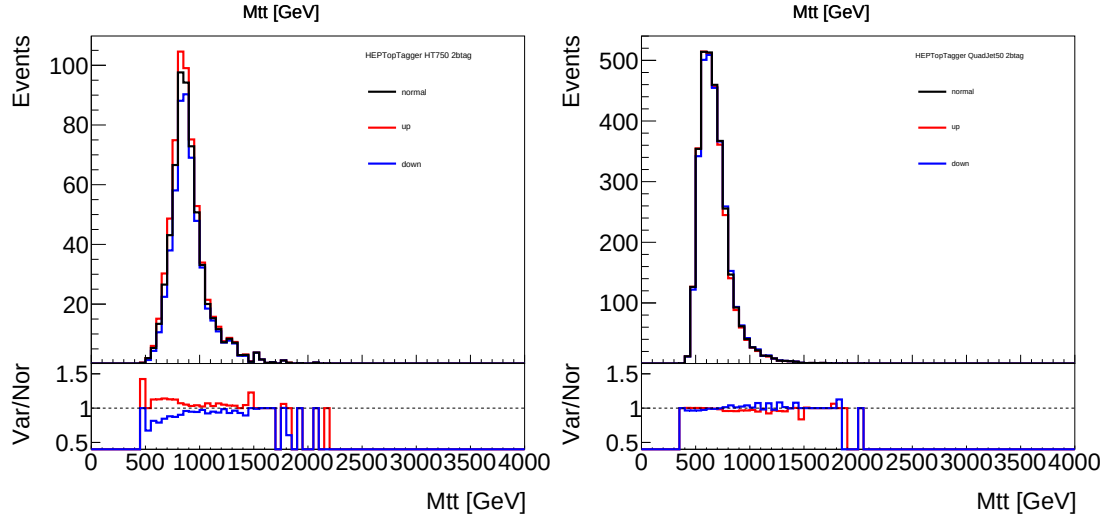


Figure B.4: Impact of the top-jet energy scale uncertainty on the SM $t\bar{t}$ background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

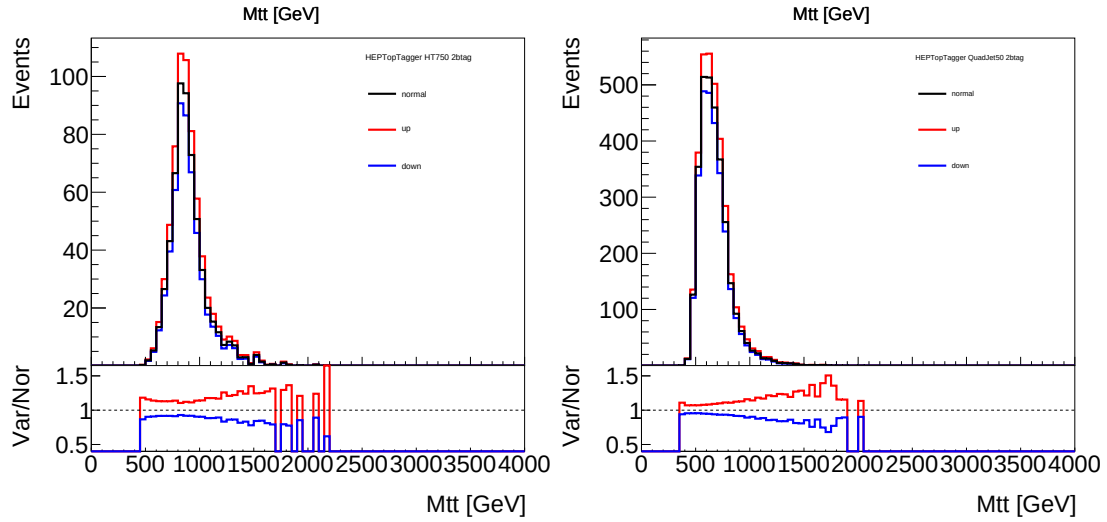


Figure B.5: Impact of the parton distribution functions uncertainty on the SM $t\bar{t}$ background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

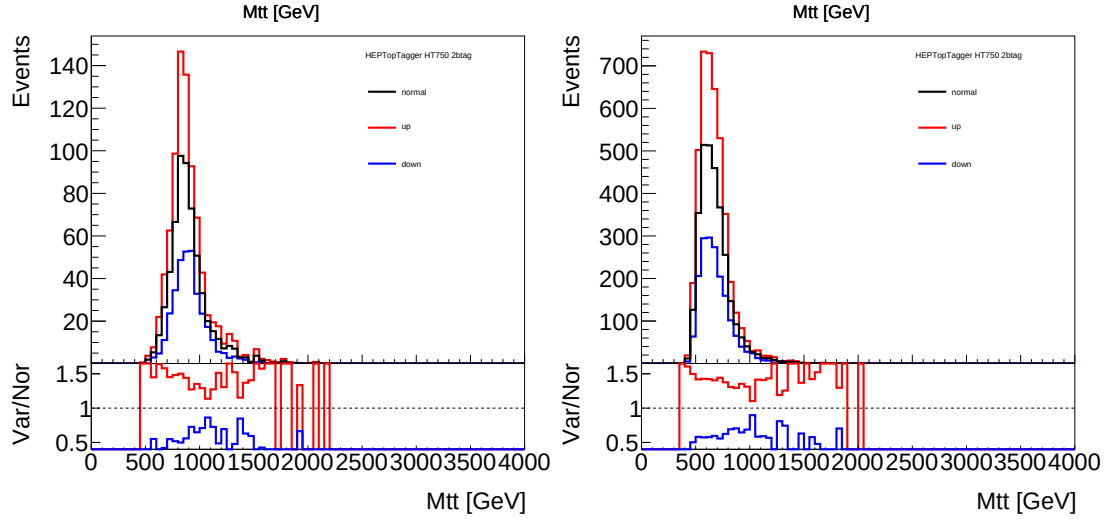


Figure B.6: Impact of the Q^2 scale variation on the SM $t\bar{t}$ background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

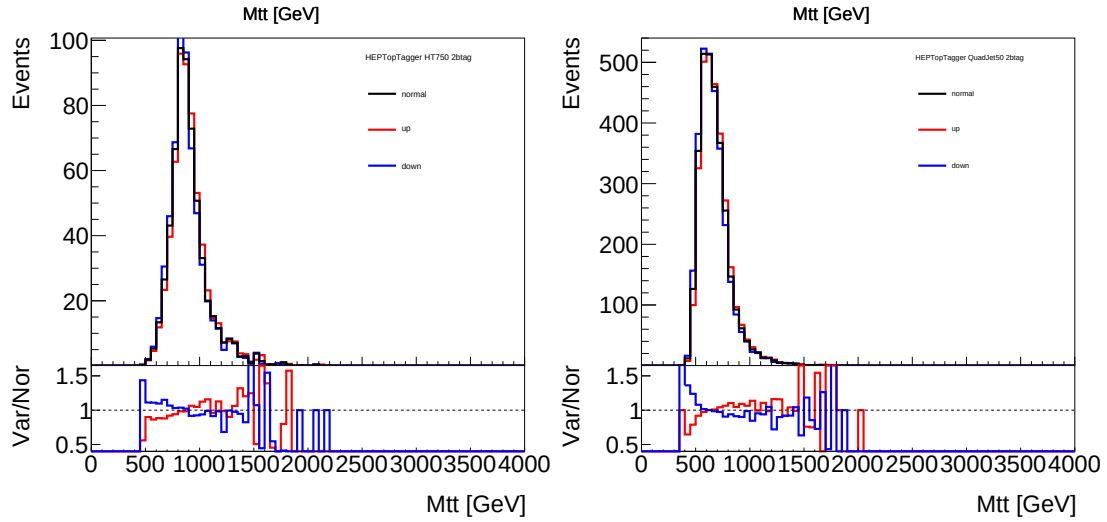


Figure B.7: Impact of the subjet energy scale uncertainty on the SM $t\bar{t}$ background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

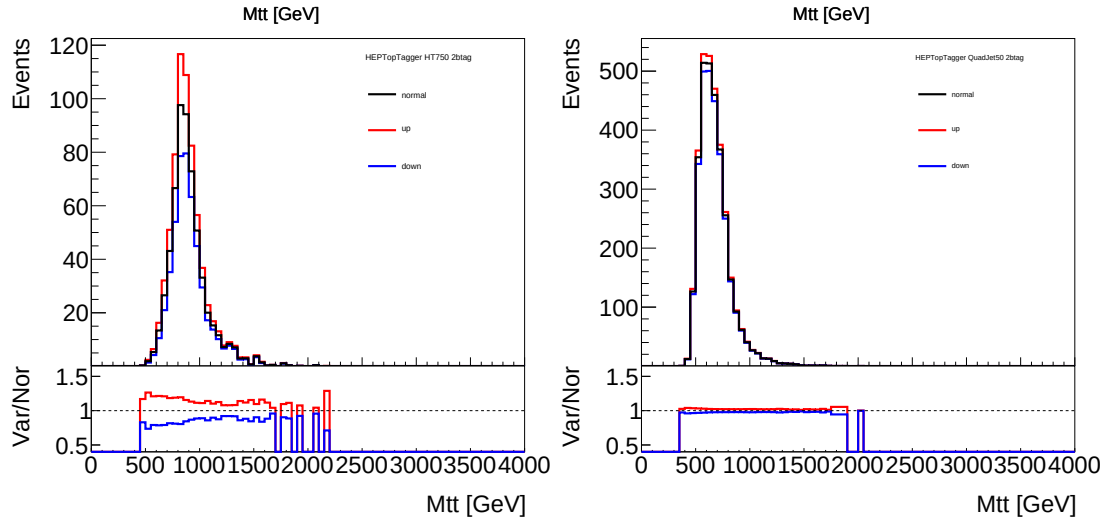


Figure B.8: Impact of the trigger efficiency uncertainty on the SM $t\bar{t}$ background for the 2 b-tags $H_T > 800$ GeV (left) and 2 b-tags $H_T < 800$ GeV (right) event categories.

Systematic uncertainties templates for the search for resonant top plus vector-like T quark production at 13 TeV

Figures C.1-C.10 report the templates used to define the $\pm 1\sigma$ shape uncertainties for the most relevant systematic uncertainties. The distributions are shown for the uncertainties affecting the $t\bar{t}$ simulated sample and the data-driven QCD background estimate.

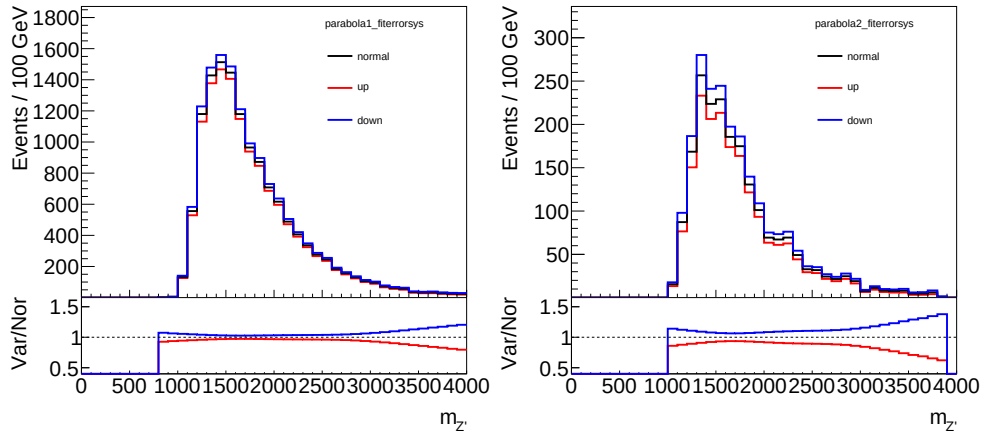


Figure C.1: Impact of the QCD background estimate systematic uncertainty (fit parameter error) for the 1 b-tag category (left) and the 2 b-tag category (right).

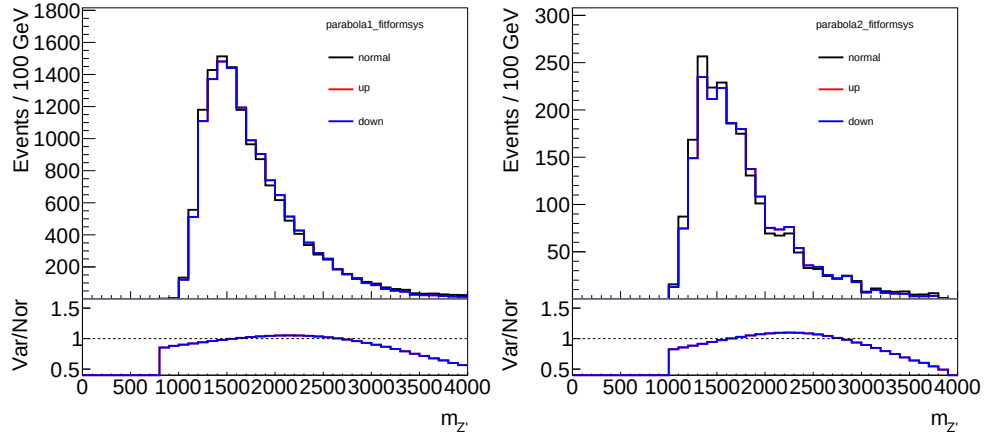


Figure C.2: Impact of the QCD background estimate systematic uncertainty (fit functional form) for the 1 b-tag category (left) and the 2 b-tag category (right).

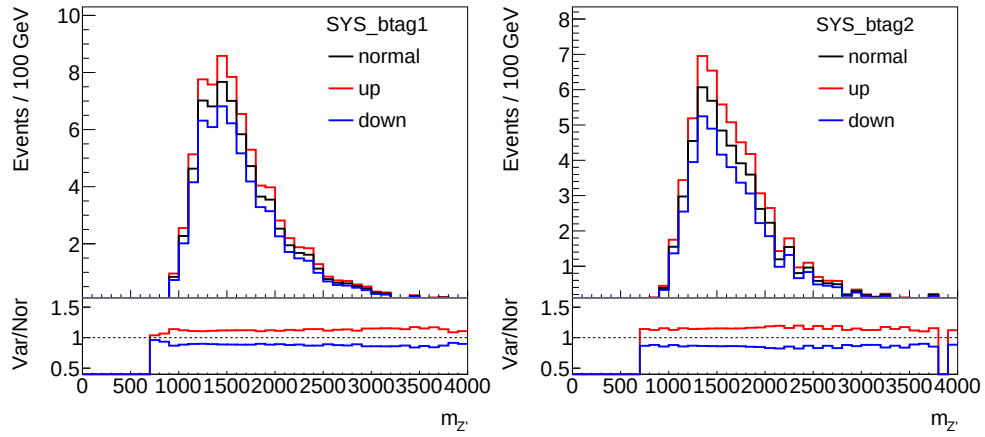


Figure C.3: Impact of the b-tagging systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

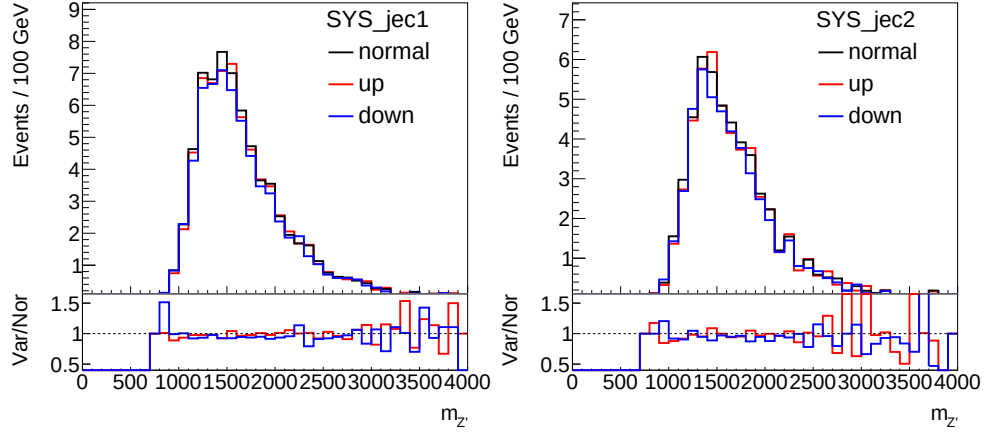


Figure C.4: Impact of the jet energy scale systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

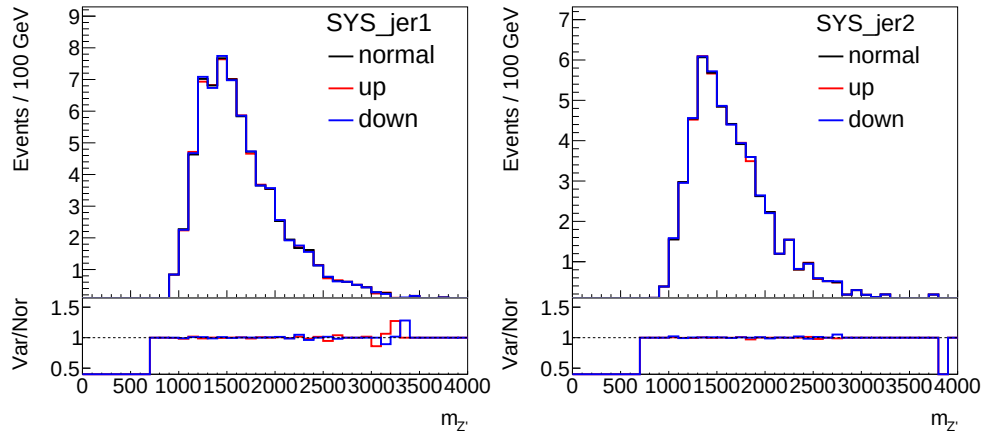


Figure C.5: Impact of the jet energy resolution systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

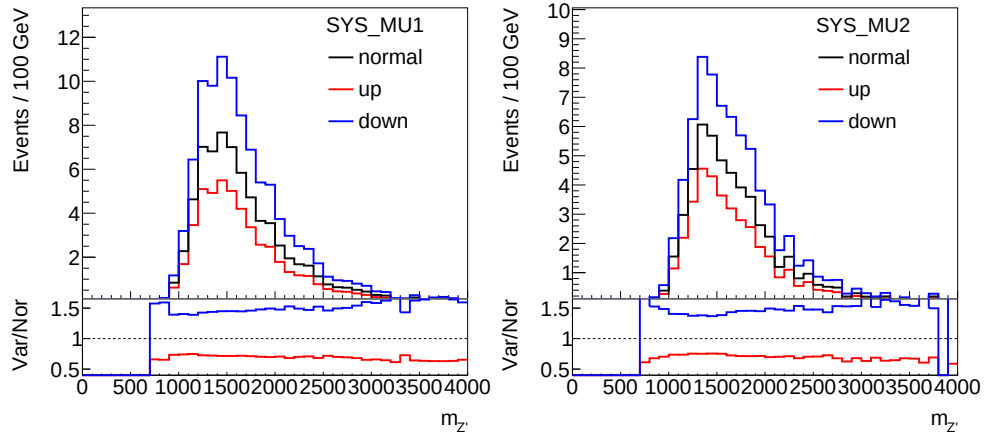


Figure C.6: Impact of the renormalization and factorization scale systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

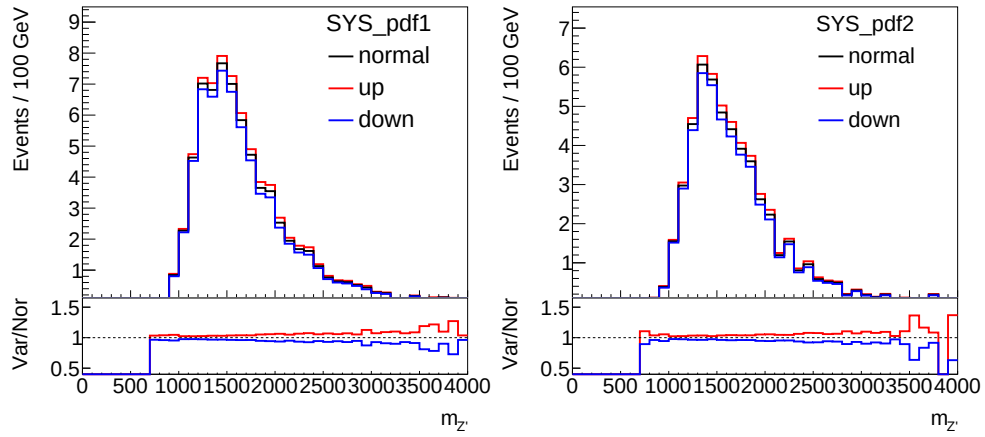


Figure C.7: Impact of the parton distribution function systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

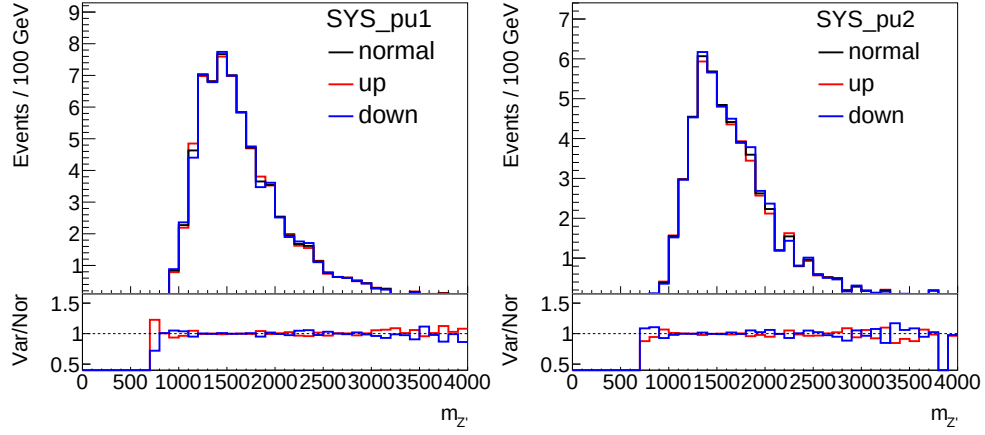


Figure C.8: Impact of the pileup reweight systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

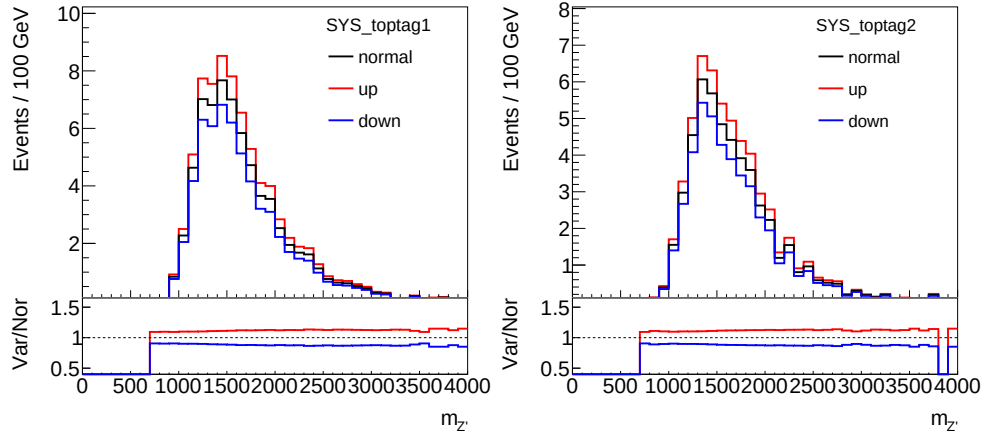


Figure C.9: Impact of the top-tagging systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

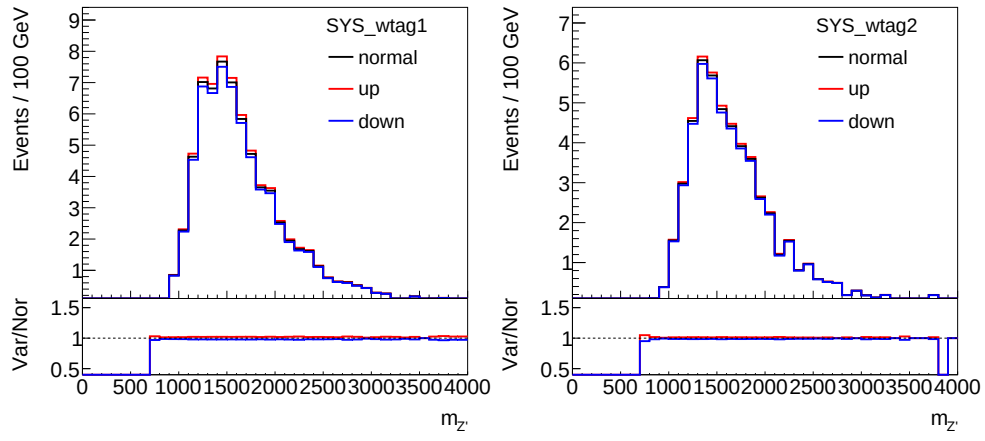


Figure C.10: Impact of the W-tagging systematic uncertainty on the $t\bar{t}$ simulated sample for the 1 b-tag category (left) and the 2 b-tag category (right).

Appendix D

Observed limits arranged on a bidimensional grid of masses

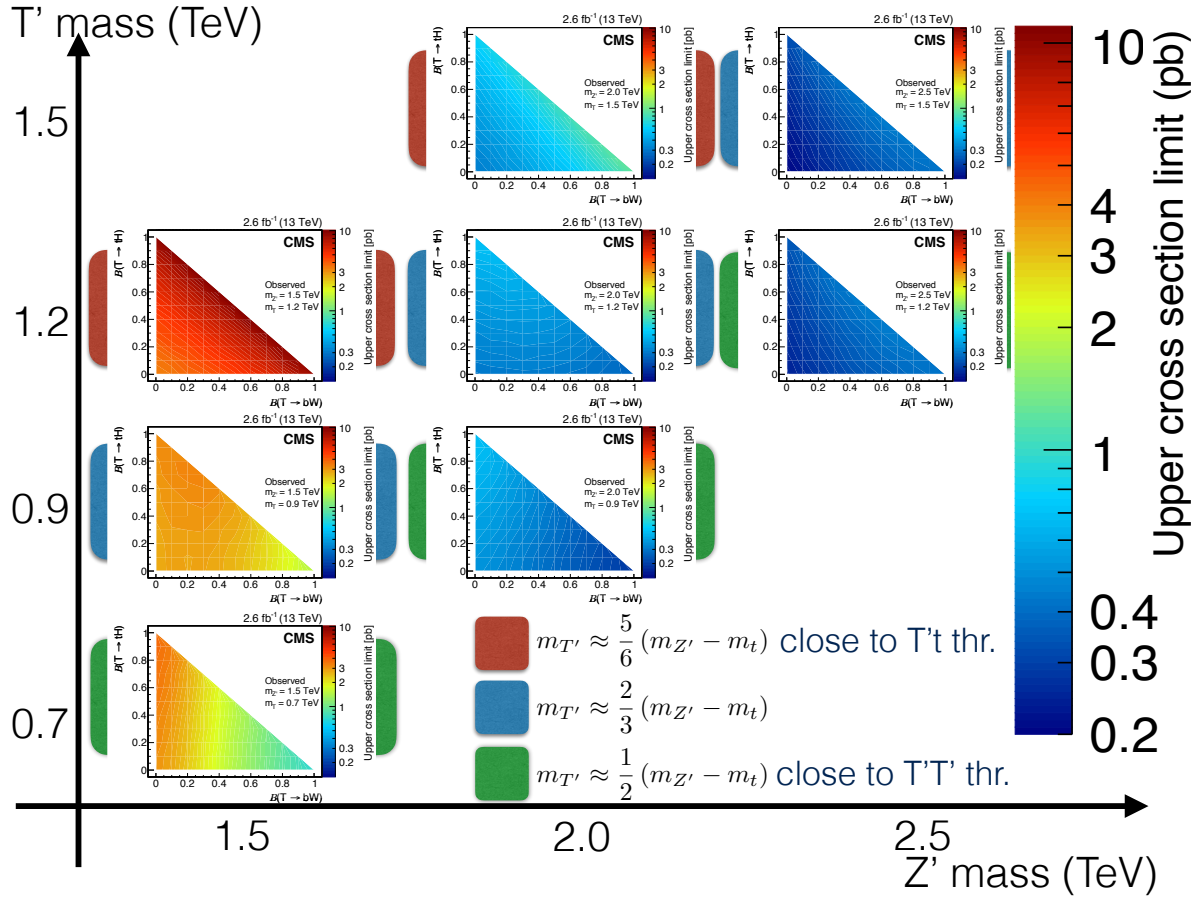


Figure D.1: Observed upper cross section limits on the production of a Z' boson decaying to Tt . The results are arranged on a bidimensional grid in function of the masses of the Z' boson and the T quark.

Bibliography

- [1] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [2] 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. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [3] “Standard Model”. https://en.wikipedia.org/wiki/Standard_Model. Accessed: 2017-06-23.
- [4] ATLAS, CDF, CMS, D0 Collaboration, “First combination of Tevatron and LHC measurements of the top-quark mass”, arXiv:1403.4427.
- [5] CDF Collaboration, “Observation of top quark production in $\bar{p}p$ collisions”, *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.
- [6] D0 Collaboration, “Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV”, *Phys. Rev. Lett.* **74** (1995) 2422–2426, doi:10.1103/PhysRevLett.74.2422, arXiv:hep-ex/9411001.
- [7] M. Czakon and A. Mitov, “Top++: a program for the calculation of the top-pair cross-section at hadron colliders”, *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [8] “Top quark Feynman diagrams”. https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html. Accessed: 2017-06-23.
- [9] M. E. Peskin and D. V. Schroeder, “An Introduction to quantum field theory”. Addison-Wesley, Reading, USA, 1995.
- [10] ATLAS, CMS Collaboration, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.
- [11] B. Zwiebach, “A first course in string theory”. Cambridge University Press, 2006.
- [12] D. Clowe et al., “A direct empirical proof of the existence of dark matter”, *Astrophys. J.* **648** (2006) L109–L113, doi:10.1086/508162, arXiv:astro-ph/0608407.

- [13] SNO Collaboration, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* **89** (2002) 011301, doi:10.1103/PhysRevLett.89.011301, arXiv:nucl-ex/0204008.
- [14] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.
- [15] F. Bezrukov, M. Yu. Kalmykov, B. A. Kniehl, and M. Shaposhnikov, “Higgs Boson Mass and New Physics”, *JHEP* **10** (2012) 140, doi:10.1007/JHEP10(2012)140, arXiv:1205.2893. [275(2012)].
- [16] J. R. Espinosa et al., “The cosmological Higgstory of the vacuum instability”, *JHEP* **09** (2015) 174, doi:10.1007/JHEP09(2015)174, arXiv:1505.04825.
- [17] G. F. Giudice, “Naturally Speaking: The Naturalness Criterion and Physics at the LHC”, doi:10.1142/9789812779762_0010, arXiv:0801.2562.
- [18] S. P. Martin, “A Supersymmetry primer”, doi:10.1142/9789812839657_0001, 10.1142/9789814307505_0001, arXiv:hep-ph/9709356. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [19] “Supersymmetry”. <https://en.wikipedia.org/wiki/Supersymmetry>. Accessed: 2017-06-23.
- [20] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria, “Handbook of vector-like quarks: mixing and single production”, *Phys. Rev. D* **88** (2013) 094010, doi:10.1103/PhysRevD.88.094010, arXiv:1306.0572.
- [21] A. De Simone, O. Matsedonskyi, R. Rattazzi, and A. Wulzer, “A First Top Partner Hunter’s Guide”, *JHEP* **04** (2013) 004, doi:10.1007/JHEP04(2013)004, arXiv:1211.5663.
- [22] CMS Collaboration, “Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure techniques at $\sqrt{s} = 13$ TeV”, arXiv:1706.03408.
- [23] CMS Collaboration, “Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state”, *Phys. Lett.* **B771** (2017) 80–105, doi:10.1016/j.physletb.2017.05.019, arXiv:1612.00999.
- [24] CMS Collaboration, “Search for single production of vector-like quarks decaying to a Z boson and a top or a bottom quark in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **05** (2017) 029, doi:10.1007/JHEP05(2017)029, arXiv:1701.07409.
- [25] ATLAS Collaboration, “Search for single production of a vector-like quark via a heavy gluon in the $4b$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett.* **B758** (2016) 249–268, doi:10.1016/j.physletb.2016.04.061, arXiv:1602.06034.
- [26] CMS Collaboration Collaboration, “Search for top quark partners with charge 5/3 in the single-lepton final state at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-B2G-17-008, CERN, 2017.

- [27] CMS Collaboration Collaboration, “Search for vector-like quark pair production $T\bar{T}(Y\bar{Y}) \rightarrow bWbW$ using kinematic reconstruction in lepton+jets final states at $\sqrt{s}=13$ TeV”, CMS Physics Analysis Summary CMS-PAS-B2G-17-003, CERN, 2017.
- [28] E. Farhi and L. Susskind, “Technicolour”, *Physics Reports* **74** (1981), no. 3, 277 – 321, doi:http://dx.doi.org/10.1016/0370-1573(81)90173-3.
- [29] E. Eichten and K. D. Lane, “Dynamical Breaking of Weak Interaction Symmetries”, *Phys. Lett.* **90B** (1980) 125–130, doi:10.1016/0370-2693(80)90065-9.
- [30] S. Dimopoulos and L. Susskind, “Mass Without Scalars”, *Nucl. Phys.* **B155** (1979) 237–252, doi:10.1016/0550-3213(79)90364-X. [2,930(1979)].
- [31] H. Georgi and A. Pais, “Vacuum symmetry and the pseudo-Goldstone phenomenon”, *Phys. Rev. D* **12** (Jul, 1975) 508–512, doi:10.1103/PhysRevD.12.508.
- [32] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson, “The Littlest Higgs”, *JHEP* **07** (2002) 034, doi:10.1088/1126-6708/2002/07/034, arXiv:hep-ph/0206021.
- [33] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, “Electroweak symmetry breaking from dimensional deconstruction”, *Phys. Lett.* **B513** (2001) 232–240, doi:10.1016/S0370-2693(01)00741-9, arXiv:hep-ph/0105239.
- [34] M. Schmaltz and D. Tucker-Smith, “Little Higgs review”, *Ann. Rev. Nucl. Part. Sci.* **55** (2005) 229–270, doi:10.1146/annurev.nucl.55.090704.151502, arXiv:hep-ph/0502182.
- [35] D. B. Kaplan, “Flavor at SSC energies: A New mechanism for dynamically generated fermion masses”, *Nucl. Phys.* **B365** (1991) 259–278, doi:10.1016/S0550-3213(05)80021-5.
- [36] Y. Nagashima, “Beyond the standard model of elementary particle physics”. Wiley-VCH Verlag, Weinheim, Germany, 2014.
- [37] M. Perelstein, “Little Higgs models and their phenomenology”, *Prog. Part. Nucl. Phys.* **58** (2007) 247–291, doi:10.1016/j.ppnp.2006.04.001, arXiv:hep-ph/0512128.
- [38] D. Greco and D. Liu, “Hunting composite vector resonances at the LHC: naturalness facing data”, *Journal of High Energy Physics* **2014** (2014), no. 12, 1–50, doi:10.1007/JHEP12(2014)126.
- [39] T. Kaluza, “Zum Unitätsproblem in der Physik”, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* (1921) 966.
- [40] O. Klein, “Quantentheorie und fünfdimensionale Relativitätstheorie”, *Zeitschrift für Physik* **37** (Dec, 1926) 895–906, doi:10.1007/BF01397481.
- [41] L. Randall and R. Sundrum, “A large mass hierarchy from a small extra dimension”, *Phys. Rev. Lett.* **83** (1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.
- [42] L. Randall and R. Sundrum, “An alternative to compactification”, *Phys. Rev. Lett.* **83** (1999) 4690–4693, doi:10.1103/PhysRevLett.83.4690, arXiv:hep-th/9906064.
- [43] K. Agashe et al., “CERN LHC signals from warped extra dimensions”, *Phys. Rev. D* **77** (Jan, 2008) 015003, doi:10.1103/PhysRevD.77.015003.

- [44] C. Bini, R. Contino, and N. Vignaroli, “Heavy-light decay topologies as a new strategy to discover a heavy gluon”, *JHEP* **01** (2012) 157, doi:10.1007/JHEP01(2012)157, arXiv:1110.6058.
- [45] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys. C* **40** (2016), no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.
- [46] C. De Melis, “The CERN accelerator complex. Complexe des accélérateurs du CERN”, general photo, Jul, 2016.
- [47] LHC Study Group Collaboration, “The Large Hadron Collider: Conceptual design”, technical design report, 1995.
- [48] LHC/LC Study Group Collaboration, “Physics interplay of the LHC and the ILC”, *Phys. Rept.* **426** (2006) 47–358, doi:10.1016/j.physrep.2005.12.003, arXiv:hep-ph/0410364.
- [49] “CMS detector design”. <http://cms.web.cern.ch/news/cms-detector-design>. Accessed: 2017-06-23.
- [50] CMS Collaboration, “CMS physics: Technical design report”, technical design report, 2006.
- [51] CMS Collaboration, “The CMS magnet project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [52] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [53] CMS Collaboration, “The CMS tracker system project: Technical Design Report”. Technical Design Report CMS. CERN, 1997.
- [54] E. Usai, “Measurement of the ratio $\sigma(pp \rightarrow \chi_{c2})/\sigma(pp \rightarrow \chi_{c1})$ with the CMS detector at LHC”. Master’s degree thesis.
- [55] E. Usai, “Preparazione alla presa dati del calorimetro elettromagnetico dell’esperimento CMS”. Bachelor’s degree thesis.
- [56] D. Contardo et al., “Technical Proposal for the Phase-II Upgrade of the CMS Detector”, Technical Report CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, Geneva, Jun, 2015.
- [57] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017.
- [58] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, arXiv:1706.04965.
- [59] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, *Transactions of the ASME—Journal of Basic Engineering* **82** (1960), no. Series D, 35–45.
- [60] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems”, in *Proceedings of the IEEE*, pp. 2210–2239. 1998.

- [61] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, “Adaptive vertex fitting”, *J. Phys.* **G34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.
- [62] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [63] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification”, *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.
- [64] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [65] S. Catani, Y. Dokshitzer, M. Seymour, and B. Webber, “Longitudinally-invariant k_T -clustering algorithms for hadron-hadron collisions”, *Nuclear Physics B* **406** (1993), no. 1, 187 – 224, doi:http://dx.doi.org/10.1016/0550-3213(93)90166-M.
- [66] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better Jet Clustering Algorithms”, *JHEP* **08** (1997) 001, arXiv:hep-ph/9707323.
- [67] CMS Collaboration, “Performance of b tagging at $\sqrt{s}=8$ TeV in multijet, ttbar and boosted topology events”, CMS Physics Analysis Summary CMS-PAS-BTV-13-001, CERN, 2013.
- [68] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, CERN, 2016.
- [69] CMS Collaboration, “Heavy flavor identification at CMS with deep neural networks”, technical report, Mar, 2017.
- [70] ZEUS, H1 Collaboration, “Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data”, *Eur. Phys. J.* **C75** (2015), no. 12, 580, doi:10.1140/epjc/s10052-015-3710-4, arXiv:1506.06042.
- [71] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching Matrix Elements and Shower Evolution for Top-Quark Production in Hadronic Collisions”, *JHEP* **01** (2007) 013.
- [72] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.
- [73] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [74] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.
- [75] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new Higgs search channel at the LHC”, *Phys.Rev.Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.
- [76] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **1103** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.

- [77] J. Thaler and K. Van Tilburg, “Maximizing Boosted Top Identification by Minimizing N-subjettiness”, *JHEP* **1202** (2012) 093, doi:10.1007/JHEP02(2012)093, arXiv:1108.2701.
- [78] G. Salam, “Jet substructure - back to basics”, 2013. CMS Jet Substructure Workshop, CERN, Switzerland.
- [79] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, “Stop Reconstruction with Tagged Tops”, *JHEP* **1010** (2010) 078, doi:10.1007/JHEP10(2010)078, arXiv:1006.2833.
- [80] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure”, *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.
- [81] CMS Collaboration, “Boosted Top Jet Tagging at CMS”, CMS Physics Analysis Summary CMS-PAS-JME-13-007, CERN, 2014.
- [82] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, “Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks”, *Phys. Rev. Lett.* **101** (2008) 142001, doi:10.1103/PhysRevLett.101.142001, arXiv:0806.0848.
- [83] CMS Collaboration, “Studies of Top Tagging Algorithms”, *CMS Note* **2010/080**.
- [84] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross sections in deep-inelastic scattering”, arXiv:hep-ph/9907280.
- [85] CMS Collaboration, “Top Tagging with New Approaches”, CMS Physics Analysis Summary CMS-PAS-JME-15-002, CERN, 2016.
- [86] CMS Collaboration, “Identification techniques for highly boosted W bosons that decay into hadrons”, *JHEP* **12** (2014) 017, doi:10.1007/JHEP12(2014)017, arXiv:1410.4227.
- [87] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, 2017.
- [88] D. Rankin, “CMS triggers with boosted objects or jet substructure”. BOOST 2015, <https://indico.cern.ch/event/382815/contributions/910647/>. Accessed: 2017-06-23.
- [89] CMS Collaboration, “Identification of b-quark jets in boosted topologies”, CMS Internal Analysis Note CMS-AN-13-185, CERN, 2013.
- [90] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.
- [91] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [92] CMS Collaboration, “Standard Model Cross Sections for CMS at 8 TeV”, <https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSectionsat8TeV>.
- [93] CMS Collaboration, “Top Quark Reference Selections”. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiTopRefEventSel>.

- [94] CMS Collaboration, “Reference muon id and isolation efficiencies”,
<https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs>.
- [95] CMS Collaboration, “Recommendation for b/c-tagging and mistagging Data/MC Scale Factors”,
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation>.
- [96] CMS Collaboration, “ p_T (top-quark) based reweighting of $t\bar{t}$ MC”,
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtRewighting>.
- [97] CMS Collaboration, “Boosted Top Jet Tagging at CMS”, CMS Internal Analysis Note CMS-AN-13-029, CERN, 2014.
- [98] CMS Collaboration, “Pileup Studies”,
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupInformation>.
- [99] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **6** (2011) P11002,
doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- [100] S. Dimopoulos and H. Georgi, “Softly Broken Supersymmetry and SU(5)”, *Nucl. Phys.* **B193** (1981) 150, doi:10.1016/0550-3213(81)90522-8.
- [101] S. Weinberg, “Implications of Dynamical Symmetry Breaking”, *Phys. Rev.* **D13** (1976) 974–996, doi:10.1103/PhysRevD.13.974.
- [102] L. Susskind, “Dynamics of Spontaneous Symmetry Breaking in the Weinberg- Salam Theory”, *Phys. Rev.* **D20** (1979) 2619–2625, doi:10.1103/PhysRevD.20.2619.
- [103] C. T. Hill and S. J. Parke, “Top production: Sensitivity to new physics”, *Phys. Rev.* **D49** (1994) 4454–4462, doi:10.1103/PhysRevD.49.4454, arXiv:hep-ph/9312324.
- [104] R. S. Chivukula, B. A. Dobrescu, H. Georgi, and C. T. Hill, “Top quark seesaw theory of electroweak symmetry breaking”, *Phys. Rev.* **D59** (1999) 075003,
doi:10.1103/PhysRevD.59.075003, arXiv:hep-ph/9809470.
- [105] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, “The hierarchy problem and new dimensions at a millimeter”, *Phys. Lett.* **B429** (1998) 263–272,
doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.
- [106] J. Conway et al., “Search for BSM $t\bar{t}$ Production in the Boosted All-Hadronic Final State”, CMS Note 2011/194, 2011.
- [107] CMS Collaboration, “Search for Anomalous Top Quark Pair Production in the Boosted All-Hadronic Final State using pp Collisions at $\sqrt{s} = 8$ TeV”, CMS Physics Analysis Summary CMS-PAS-B2G-12-005, 2013.
- [108] CMS Collaboration, “Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev.* **D93** (2016), no. 1, 012001, doi:10.1103/PhysRevD.93.012001, arXiv:1506.03062.
- [109] CMS Collaboration, “Improved Search for BSM $t\bar{t}$ Production in the Boosted All-Hadronic Final State using pp Collisions at $\sqrt{s} = 8$ TeV”, CMS Internal Analysis Note CMS-AN-13-397, CERN, 2014.

- [110] N. Kidonakis, “Differential and total cross sections for top pair and single top production”, [arXiv:1205.3453](#).
- [111] R. M. Harris and S. Jain, “Cross sections for leptophobic topcolor Z' decaying to top-antitop”, *Eur. Phys. J. C* **72** (2012) 2072, doi:10.1140/epjc/s10052-012-2072-4, [arXiv:1112.4928](#).
- [112] CMS Collaboration, “Jet Performance in pp Collisions at $\sqrt{s} = 7$ TeV”, *CMS PAS JME-10-003* (2010).
- [113] CMS Collaboration, “Event display of a boosted $t\bar{t}$ event at 13 TeV”, technical report, Jul, 2015.
- [114] CMS Collaboration, “Search for top-Higgs resonances in all-hadronic final states using substructure methods”, CMS Note 2013/129, 2014.
- [115] CMS Collaboration, “Jet-by-jet updating of the b-tagging status”. <https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods>.
- [116] CMS Collaboration, “Jet Energy Corrections and Uncertainties. Detector Performance Plots for 2012.”, http://cms.cern.ch/iCMS/jsp/openfile.jsp?type=DP&year=2012&files=DP2012_012.pdf.
- [117] T. Muller, J. Ott, and J. Wagner-Kuhr, “theta - a framework for template-based modeling and inference”, *CMS Internal Note IN-2010/017* (2010).
- [118] CMS Collaboration, “Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s}=8$ TeV”, *Phys. Rev. Lett.* **111** (2013) 211804, doi:10.1103/PhysRevLett.111.211804, [arXiv:1309.2030](#).
- [119] J. Gao et al., “Next-to-leading order QCD corrections to a heavy resonance production and decay into top quark pair at the LHC”, *Phys. Rev. D* **82** (2010) 014020, doi:10.1103/PhysRevD.82.014020, [arXiv:1004.0876](#).
- [120] CMS Collaboration, “Search for a heavy resonance decaying to a top quark and a vector-like top quark at $\sqrt{s} = 13$ TeV”, [arXiv:1703.06352](#).
- [121] CMS Collaboration, “Search for a heavy resonance decaying to a top quark and a vector-like top quark at $\sqrt{s} = 13$ TeV”, CMS Internal Analysis Note CMS-AN-16-019, CERN, 2014.
- [122] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, [arXiv:hep-ph/0409146](#).
- [123] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, [arXiv:0709.2092](#).
- [124] S. Frixione, P. Nason, and G. Ridolfi, “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”, *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, [arXiv:0707.3088](#).
- [125] E. Re, “Single-top Wt -channel production matched with parton showers using the POWHEG method”, *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, [arXiv:1009.2450](#).

- [126] N. Kidonakis, “Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^- ”, *Phys. Rev. D* **82** (2010) 054018, doi:10.1103/PhysRevD.82.054018, arXiv:1005.4451. Updates in arXiv:1311.0283.
- [127] 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* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [128] M. Aliev et al., “HATHOR: HAdronic Top and Heavy quarks crOSS section calculatoR”, *Comput. Phys. Commun.* **182** (2011) 1034, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.
- [129] P. Kant et al., “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”, *Comput. Phys. Commun.* **191** (2015) 74, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.
- [130] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”, *Journal of High Energy Physics* **2013** (2013), no. 3, 1–19, doi:10.1007/JHEP03(2013)015.
- [131] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [132] The CMS Collaboration, “Jet energy corrections guide”, 2016, <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- [133] The CMS Collaboration, “Jet energy resolution guide”, 2016, <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- [134] The CMS Collaboration, “b-Tagging Offline Guide”, 2014, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagging>.
- [135] The CMS collaboration, “Identification of b-quark jets with the CMS experiment”, *Journal of Instrumentation* **8** (2013), no. 04, P04013.
- [136] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium.

Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 26.07.2017

Unterschrift der Doktorandin / des Doktoranden

Emanuele Usai