Cross Section measurements of $t\bar{t}$ pair production in the boosted regime in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS experiment at the Large Hadron Collider

Dissertation

zur Erlangung des Doktorgrades

des Departments Physik

der Universität Hamburg

vorgelegt von

Daniela Dominguez Damiani

aus Cienfuegos, Kuba

Hamburg

2019

Gutachter der Dissertation:	PD Dr. Hannes Jung Prof. Dr. Elisabetta Gallo
Gutachter der Disputation:	Prof. Dr. Dieter Horn Prof. Dr. Elisabetta Gallo Prof. Dr. Günter Sigl PD Dr. Hannes Jung Dr. Paolo Gunnellini
Vorsitzende des Prüfungsausschusses:	Prof. Dr. Dieter Horn (Chair)
Datum der Disputation:	10. September 2019
Vorsitzende des Fach-Promotionsausschusses Physik:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs Physik:	Prof.Dr. Michael Potthoff
Dekan der Fakultär für Mathematik, Informatik und Naturwissenschaften:	Prof. Dr. Heinrich Graener

Hiermit erklare ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den 8 August 2019

Acknowledgements

I would like, first of all, to thank my supervisor Dr. Hannes Jung, who made possible this long journey starting from a so far away country, Cuba. I would like to thank to Hannes not only because of his supervision related to Physics topics, but also because of his support and consideration as a great person he is.

I would like to thank Dr. Paolo Gunnellini, who supported me and guided me in each step of this project. Furthermore, Paolo is one of the friends I gained during these last years working at DESY.

I would like to thank the rest of the members in the examination committee, Prof. Dr. Elisabetta Gallo, Prof. Dr. Dieter Horns, and Prof. Dr. Günter Sigl.

I would like to thank all my colleagues from the QCD group. Special thanks goes to Ola, who is one of my best friends (when one is so far away from home, friendship plays sometimes the role of even family).

I would also like to thank Yoana. She is completely outside DESY world, but without doubts she played an important role as best friend in this long path.

When I look back to all the phases of my studies, names of teachers and professors come to my mind, to whom I am certainly grateful. Without their contributions to my education, I probably wouldn't have reached this point. Their names are Gladys Bradley, Marta de Las Casas, Ismael Jesus Sanchez Cardenas (El Suso), Prof. Dr. Carlos Manuel Cruz Inclan and Prof. Dr. Fernando Guzman Martinez.

Last but not least, the greatest gratitude goes to my family. They have been present in all moments, though those eight thousand kilometers in actual distance. Specially to my mother Griselda Liliana Damiani Cavero and my father Rafael Dominguez Milan. Also thanks to my aunt Julieta Sonia Damiani Cavero and my cousin Claudia Alejandra Damiani Cavero.

To my new family, that made Hamburg also my home, goes the most honest gratitude. Specially to my husband, Tobias Schliemann, who was always incondicionally supporting me and giving me power. Thanks to Sieta Schliemann, Kai Schliemann, Marc Schliemann, Rebecca Schliemann and Jasper Schliemann.

Finalizing these words, I would like to dedicate this thesis to our son, who we will be welcoming to this world in few months.

Abstract

In this thesis measurements of the $t\bar{t}$ pair production at the very high transverse momentum are presented. When a top quark is produced with very high momentum the decay products are collimated in the direction of the quark, and they can be clustered within a single large cone size jet. The decay products of top quarks with transverse momentum larger than 400 GeV are likely having this boosted configuration. A definition of a particle-level top jet is developed which is independent of the partonic configuration. The measurements are performed with proton-proton collision data collected by the CMS experiment at the LHC, at the center of mass energy of $\sqrt{s} = 13$ TeV and they are presented differentially in the transverse momentum of the top jets and differentially in the azimuthal angle $\Delta\phi$ between the two top-jets. The results are compared to theory predictions provided from different Monte Carlo event generators: i.e POWHEG+PYTHIA8, POWHEG+HERWIGpp and aMC@NLO+PYTHIA8.

Zusammenfassung

In dieser Arbeit werden Messungen von $t\bar{t}$ -Paarproduktion bei sehr hohem Transversalimpuls vorgestellt. Wenn ein Top-Quark mit einem sehr hohen Impuls erzeugt wird, werden die Zerfallsprodukte in Richtung des Quarks kollimiert und können in einem einzigen Jet mit großen Jet-Cone gebündelt werden. Die Zerfallsprodukte von Top-Quarks mit einem Transversalimpuls von mehr als 400 GeV weisen diese besondere Konfiguration auf. Es wird eine Definition eines Top-Jets auf Teilchenlevel erarbeitet, die unabhängig von der zugrunde liegenden Partonkonfiguration ist. Die Messungen werden mit Proton-Proton-Kollisionsdaten durchgeführt, die durch das CMS-Experiment am LHC bei einer Massenschwerpunktsenergie von $\sqrt{s} = 13$ TeV aufgenommen wurden, und sie werden dargestellt differentiell im Transversalimpuls der Top-Jets und im azimuthalen Winkel $\Delta\phi$ zwischen den beiden Top-Jets. Die Ergebnisse werden mit theoretischen Vorhersagen verglichen, die von verschiedenen Monte-Carlo-Ereignisgeneratoren bereitgestellt werden: d. H. POWHEG + PYTHIA8, POWHEG + HERWIG pp und MC @ NLO + PYTHIA8.

Contents

Introduction

The European Council for Nuclear Research (CERN) hosts the largest accelerator facility in the world: the Large Hadron Collider (LHC). This machine has been developed to access high interaction scale energies and to reproduce the conditions given at the early stage when the universe was created. The Standard Model of Particle Physics is the theory predicting all the fundamental particles building up the visible matter of the universe, and also describing the fundamental interaction occurring between the particles. The LHC and the experimental facilities (i.e the Compact Muon Solenoid, CMS) open windows to test the Standard Model and to explore even beyond the Standard Model Physics. One of the fundamental blocks in the Standard Model are the quarks. They interact between each other through the strong interaction, described by Quantum Chromodynamics (QCD). The heaviest particle in the Standard Model is the top quark, having a particular and exciting role in the model, mainly because of its large mass and being the only quark decaying before the hadronization time scale.

The LHC is considered as the largest top factory ever built, given the amount of top quarks produced during proton-proton collisions at very high energies (i.e. $\sqrt{s} = 13$ TeV). Furthermore, the high interaction scales accessed in the proton-proton collisions at the LHC, brought to possibility of exploring easier a phase space region in top quark production. This thesis focuses on measuring the top-quark pair production, when both top quarks have very high transverse momentum. In this phase space region, the decay products of each of the top quarks are collimated enough in the direction of the originating quark, and can be clustered within a single large cone size jet. Those scenarios are known as boosted topologies. Top quarks produced with transverse momentum larger than 400 GeV are most likely having a boosted configuration.

Boosted top topologies probe a new interaction scale in which different interesting phenomena can occur. For instance, the quarks interact among each other through the electroweak and strong interactions. The former one predicts the mixing between different flavour quarks: the decay of one flavour quark into a different flavour quark. The latter one is a flavour-blind theory. Since the strong interaction is responsible for creating quarks, the flavour-blindness would indicate that the probability of producing any flavour specific quark would be the same. However, this statement is not valid due to the fact that different flavour quarks cover a broad range of masses, and lighter quarks have a considerable larger cross section production in comparison to heavier quarks. Nevertheless, by increasing the energy scale interaction, the kinematics limitations given by the mass might disappear. Boosted top topologies, can open the possibility of testing the QCD-blindness, since the transverse momentum of the quark is at least two times larger than the top quark mass scale ($m_{top} = 172.5$ GeV).

In QCD the hard interaction can be described with a perturbative approximation, in terms of the strong coupling (α_s). When the top quark-pair is produced at high transverse momentum, it will most likely be produced with a back-to-back azimuthal angle configuration ($\Delta\phi \sim \pi$), and with a small transverse momentum of the system ($q_T \sim 0$). In the perturbative expansion, the logarithmic terms of the form $\propto \ln(m/q_T)$ enhance the cross section. Resummations at all orders of soft gluon radiation are required to heal those divergences. Furthermore, the so-called factorization breaking phenomena, given by the fact that top quarks are colour-charged final state, appear. These phenomena connect the initial state radiation to the final objects, breaking the factorization theorem of the QCD. Therefore, accessing to boosted top topologies gives the possibility of studying resummation and factorization breaking effects.

Jet physics plays a fundamental role for studying boosted top quarks. The suggestion of jets, as useful objects to reconstruct the decay products of the top quarks, goes back in time even before the top-quark was discovered. The earliest suggestion of kinematically reconstructing the hadronic decay products of the top quarks using jet clustering algorithms was published in 1995 [1]. In boosted topologies, however, the use of jets as objects to reconstruct and to identify the top-quark becomes more challenging. The main reason is given by the large cone size jets, used to reconstruct the decay products. Large cone size jets are highly contaminated by constituents not necessarily coming from the heavy object, and usually, those contaminations are minimized using the so-called grooming techniques. In the past ten years this topic has considerably developed, by studying new observables sensitive to the jet substructure. One of the most powerful observables nowadays are the so called N-subjettiness variables, testing the hypothesis of having N multi-prong configuration inside the heavy jets.

The LHC collision energies allow to measure top jets having a transverse momentum up to approximately 1.2 TeV. However the future energies in the High Luminosity LHC conditions (HL-LHC) might allow to measure top jets with transverse momentum up 2 TeV. The studies presented in this thesis focus on measuring the $t\bar{t}$ cross section in the boosted regime differential in the transverse momentum and the azimuthal separation between the two top jets, in proton-proton collisions at $\sqrt{s} = 13$ TeV. In Appendix A, studies concerning high p_T jets (including top-jets) at the HL-LHC scenarios are provided [2]. I contributed significantly to those studies, providing the predictions for top-jets. Those studies were included in the Yellow CERN report and published in reference [3], together with other Standard Model predictions for the HL-LHC perspective.

This thesis is organized as follows:

- ✓ The theoretical framework for the data analysis is presented in Chapter 1 and Chapter 2. The first one focuses in general aspects (i.e the Standard Model, Monte Carlo event simulations and top-quark physics), while in the latter one the jet physics and jet substructure techniques used for studying boosted topologies are discussed.
- ✓ In Chapter 3, phenomenological studies are presented providing the particle level top-jet definition. The specific way used in this thesis, to deal with boosted top jets, is presented in this Chapter.
- ✓ In Chapter 4 and Chapter 5, the experimental setup (the CMS experiment) used to perform the measurements, and the event reconstruction considered in the CMS experiment, are discussed.
- ✓ In Chapter 6, the event selection is presented. The selection strategy leads to a signal over background ratio smaller than one, therefore an additional selection criteria is added to the strategy. This last selection criterion is based on a multivariate selection technique and is presented in Chapter 7.
- ✓ In Chapter 8, the background subtraction strategy is explained. The methodology is based on data driven methods.

- ✓ The distributions at detector level are then unfolded to a stable particle level, which is defined using the top-jet particle level definition. The unfolded results, and the methodology of the unfolding algorithms are presented in Chapter 9.
- ✓ In Chapter 10, the systematic uncertainties affecting the measurements are estimated.
- ✓ In Chapter 11, the measurements are provided at particle level and compared to theory predictions provided from different Monte Carlo event simulations.
- ✓ Finally, in Chapter 12, the summary of the results, and futures perspective for the topic of this thesis are presented.

Chapter 1

The Standard Model of Particle Physics

This chapter aims to provide the fundamental theoretical background needed for any measurement performed in Particle Physics. In section 1.1, a general overview of the model responsible for explaining all the fundamental particles observed in the nature, is discussed. This model started to be formulated since 1954, when the Yang-Mill theory for Abelian groups [4] was extended to describe the Quantum Electrodynamics [1]. In 1967 the model was continued to the form which is valid nowadays, with the incorporation of the Higgs mechanism [5] into the Glashow electroweak interaction theory [6]. The model grouping all the fundamental particles and their interaction is known under the name *Standard Model* since 1975. The experimental observation of each of the components of the model finalized in 2012 with the discovery of the Higgs boson [7] [8] by the CMS[2] and ATLAS[3] Collaborations at the Large Hadron Collider (LHC). There are significant amount of books dedicated to explain the complex theory of the model, therefore, only the general overview of the model is discussed in the section 1.1.

In section 1.2 a description of the basic ideas of modeling interaction processes occurring in hadron-hadron collisions is provided. They are named as *Monte Carlo simulations* since they are performed using Monte Carlo methods [11].

Finally, in section 1.3, the fundamental physics of the top-quark, one of the particles in the Standard Model, is discussed in detail. This thesis aims to measure the cross section of this particle in proton-proton collisions at the current LHC energies. Therefore a fundamental understanding of this component of the Standard Model is needed.

[1] The Quantum Electrodynamics is the theory describing the electromagnetic interaction.
[2] Abbreviation for the *Compact Muon Solenoid* [9].
[3] Abbreviation for *A Toroidal LHC ApparatuS* [10].

1.1 The general picture of fundamental particles and their interactions

The Standard Model (SM) of Particle Physics provides a general picture of the fundamental particles observed in nature, and how they interact between each other. They can be classified either as fermions or bosons. The latter are the ones mediating the interaction between all the particles predicted in the model. Figure 1.1 illustrates a picture of the model, where two main groups of particles can be seen. Information like the electric charge, the mass, the flavour-quark charge, and the spin of each of the particles is provided. The masses of the particles predicted in the SM cover a broad range of energy scale: from few eV ($1.60 \cdot 10^{-19}$ J) up to hundreds of GeV (10^{11} eV). In this section, a description of each of the fundamental blocks in the SM is provided.

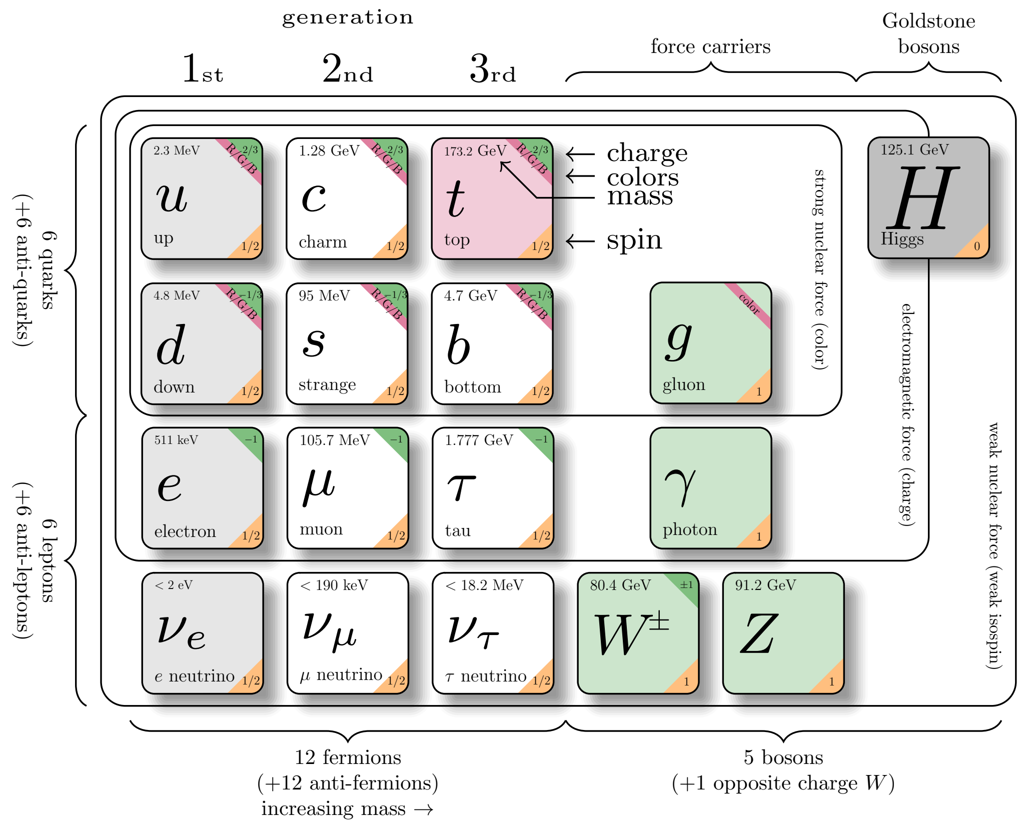


Figure 1.1: (figure modified using a template from [12]) Summary of particles building the Standard Model of fundamental particles in the nature. In the model, two main groups can be distinguished, the fermions (left) and bosons (right). The former ones are the particles building up matter while the latter ones are responsible for their interactions and masses. The fermions can be divided into two main groups: quarks and leptons, and in three generations, depending on their mass. The bosons are represented in the picture according to the interactions that they are responsible for. The gluons are responsible for the strong interaction, the photons for the electromagnetic interaction, and the bosons $W^{\pm}$ and Z for the weak interaction. The Higgs boson appears as a result of the spontaneous symmetry breaking and its field is responsible for giving mass to all the particles in the SM. The information like the electric charge, the colour charge, the mass and the spin of each particle is provided.

1.1.1 Fermions

The fermions have their name from Enrico Fermi[4]. They accomplish the Fermi-Dirac statistics [13], satisfying the Pauli Exclusion principle [14] that they cannot occupy the same quantum state at the same time. They are one of the two main blocks in the SM, and are those particles forming the visible matter of the universe. There are in total 12 fermions, and their respective antiparticles, divided into two main groups: quarks and leptons. They all have half as spin number ($s = 1/2$).

The quarks are subsequently divided into two groups: up-type and down-type, according to their electric charge (up-type quarks have positive charge, while down-type quarks have negative charge). The leptons are also divided into two groups: charged leptons and neutral leptons. There are three generations of fermions, according to their mass: the first generation consists of the lighter particles, while in the third generation the heavier particles are included. For instance, in the case of up-type quarks, the only physical quantity changing from one generation to another is the mass: the first generation up-type quark has a mass about 2.3 MeV, while the third generation (top quark) is approximately five orders of magnitude heavier (173.2 GeV).

There are six different quarks flavour. The group forming the up-type: up, charm and top (u-quark, c-quark, t-quark) have positive electric charge of $2/3$ times the electron charge, while the down-type: down, strange and bottom (d-quark, s-quark, b-quark) have negative charge of $1/3$ times the electron charge.

The quarks build up hadrons, which are stable subatomic particles. A hadron could contain either a combination of two quarks (a up-type and a down-type), or three quarks. The hadrons formed by two quarks are classified as mesons, while the hadrons formed by three quarks are baryons. For instance, protons are baryons with two up quarks and one down quark, while neutrons are also baryons, but with two down quarks and one up quark.

The leptons, which are the other type of fermions in the SM, do not have colour charge therefore they do not experience the strong interaction. They interact with other particles either by the electromagnetic force or by the weak interaction.

There are six leptons in nature, and their respective antiparticles, grouped in two main blocks: charged leptons and the neutral leptons. The electron (e), muon (μ) and tau (τ) leptons have negative charge equal to the electron charge[5]. The electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ) are neutral particles. There is a conservation law related to the leptonic number[2], implying that the leptonic number of a specific flavour (electron, muon or tau-like flavour) is conserved. This phenomenon can be violated due to the neutrino oscillation phenomena [15], as well as by others phenomena known as chiral anomalies [16]. The chiral anomalies can change the leptonic number only by a small number.

[4] [1901-1954], Enrico Fermi was a physicist winning a Nobel Prize in Physics on induced radioactivity. Inventor of the formulation know as Fermi-Dirac statistics.

[5] The charged antileptons are then positron(e^+), antimuon (μ^+) and antitau (τ^+).

[2] The leptonic number is given by the number of leptons minus the number of antileptons.

1.1.2 Bosons as mediator of the interactions in the SM

The bosons (second group shown in Figure 1.1) are the particles of the SM responsible for the force carrier and for the interactions between all the particles. They follow the Bose-Einstein statistics[17].

Three sources of carrier forces can be distinguished: the strong nuclear force, the electromagnetic force and the weak force; mediated by gluons (g), photons (γ) and massive bosons ($W^{\pm}$, Z), respectively. The strong force occurs between colour charged particles. The electromagnetic interaction occurs between electrically charged particles. The weak interaction occurs between particles carrying weak isospin.

The three fundamental interactions in the SM are described within the framework of the Gauge Field Theory. Hence they are known as Gauge bosons. In the SM an additional boson is included, the Higgs boson, which is responsible for giving mass to all the particles in the SM, through a phenomenon known as spontaneous symmetric breaking.

The electromagnetic interaction

The electromagnetic interaction is experienced by the particles in the SM carrying electric charge: the quarks and the charged leptons (electrons, muons, and taus). In the SM this interaction is described by a Quantum Field Theory (QFT) known as Quantum Electrodynamics (QED). The QED predicts that charged particles interact with each other by exchanging a virtual massless particle: the photon (γ). The mediators (γ) cannot interact with each other, since they are neutral particles. QED can be described with a local $U(1)$ symmetry group.

The running coupling of this theory [1] (α_{EM}) characterizes the strength of the electromagnetic force as function of the distance between the two interacting particles. At long distances (atomic scales), this value is approximately constant equal to the fine structure constant :

$$\alpha_{EM} = \frac{e^2}{4\pi\epsilon_0\hbar} = \frac{1}{137}\,, \tag{1.1}$$

where e is the electric charge of the electron, $\hbar$ the Max Planck constant, c the speed of the light, and ϵ_0 is the vacuum permittivity. At short distances ($\sim$ fm), or what is equivalent to say higher energy scale interaction, the coupling increases up to values approximately equal to $1/129$. The small values of α_{EM} allows to describe the interaction as a perturbative expansion in terms of the coupling for all possible energy interaction scale.

The strong interaction

The strong interaction is experienced by those particles carrying colour charge: the quarks. In the SM this interaction is described by a QFT theory known as Quantum Chromodynamics (QCD) and mediated by gluons, which are massless particles with spin equal to unity ($s = 1$). Since the gluons are colour-charged particles, they can couple additionally to themselves.

[1] In a non-Abelian gauge theory, the gauge coupling constant appears in the Lagrangian as a running parameter which characterizes the interaction.

The running coupling of the strong interaction (α_s) has a peculiar behavior, leading to two main observations:

- ✓ the confinement of quarks: quarks cannot be observed in nature as isolated particles,
- ✓ the asymptotic freedom: when increasing the energy to a certain scale, probing shorter distances of interaction, quarks behave as free particles.

Figure 1.2 shows the dependence of the QCD running coupling (α_s) as a function of the energy scale of interaction Q. The predicted values at different scales are compared to measurements from different experiments. The values of α_s at the scale of the mass of the Z-boson ($\alpha_S(M_Z)$), corresponding to a measurement performed by the CMS collaboration using inclusive 3-jet differential cross section [18] is compared to the world average of previous measurements at that scale.

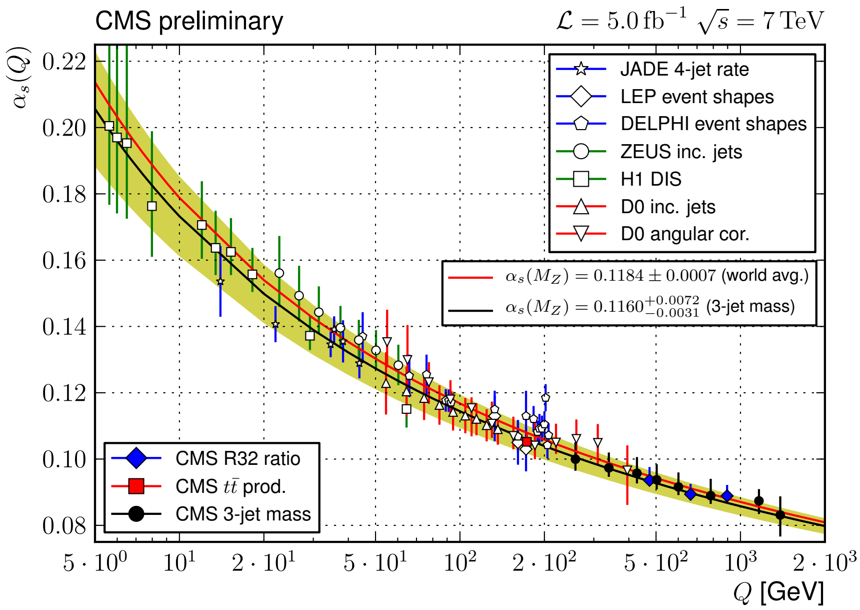


Figure 1.2: (Figure taken from [18]) Dependence of the strong coupling in QCD theory as a function of the interaction scale Q. Different measurements of this constant are illustrated in the plot, and specific value is provided for the measured α_s at the scale of M_Z (mass of the Z-boson).

The quark confinement as a result of the strong interaction can be interpreted in the following way: the increase of the distance between two colour-charged particles causes that the energy flow given by the exchange of gluons between the particles increases; after a certain distance, the energy is high enough to create a quark-antiquark pair.

When increasing the energy of the interaction (decreasing the distance between the particles), the running coupling decreases, allowing the quarks to behave as free particles.

The asymptotic freedom can be extensively explored by High Energy experiments. Contrarily, the quark confinement is more challenging to probe (i.e for energies around the GeV scale).

The weak interaction

The weak interaction is experienced by all the fundamental particles with weak charge: quarks, leptons[6], $W^{\pm}$ bosons and Z bosons. The mediator particles are the $W^{\pm}$ and the Z bosons. Those are massive particles (80.4 GeV and 91.2 GeV, respectively), with spin equal to unity ($s = 1$). The W-bosons can have either positive or negative electric charge ($\pm e$), while the Z-bosons are neutral particles. Therefore, the W bosons can additionally be coupled to the photons through the electromagnetic interaction.

The weak interaction is approximately two orders of magnitude smaller than the strong interaction. Its interaction distance is small, usually within the nuclear radius ($10^{-15} - 10^{-14}$ m), although larger interaction distances have been observed for example for the muon decay. This interaction is the unique force included in the SM which affects the neutrinos and is responsible for the quark and lepton decays.

A quark-antiquark pair can interact by exchanging neutral current (Z-boson), while the decay of the quarks changing their flavour is mediated by charged currents ($W^{\pm}$-bosons). The probability that a certain flavour quark decays into a different flavour quark is expressed by the Cabibbo Kobayashi Maskawa matrix (CKM) [19]. Each term in this matrix reflects the mixing probability of the decay among different flavour quarks. Decays in the same generation of quarks are more likely to occur, than mixing generations. Each term of the matrix has been estimated and can be written as follows [20]:

$$V_{CKM} = \begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97428 & 0.2253 & 0.00347 \\ 0.2252 & 0.97345 & 0.0410 \\ 0.00862 & 0.0403 & 0.99915 \end{bmatrix}. \tag{1.2}$$

The CKM matrix, for instance, gives a 99.9% of probability that a t-quark decays into a b-quark instead of decaying in a s or a d quark.

Charged leptons can emit or absorb charged current ($W^{\pm}$) decaying to their corresponding neutrino. In this case, given the conservation of the leptonic number, only interactions in the same family are allowed.

The electroweak interaction and the spontaneous symmetry breaking phenomenon

The electroweak theory is described by the SM as a combination of electromagnetic and weak interactions. This interaction can be considered as the superposition of two symmetry groups corresponding to each of the two interactions: $SU(2)_L \otimes U(1)_Y$. The $SU(2)$ symmetry group predicts three associated gauge bosons without mass (W^1, W^2, W^3), while in the $U(1)$ symmetry group one massless boson (B^0) is predicted. The $W^{\pm}$ and Z bosons in the SM are originating from mixing W^1, W^2, W^3 and B^0 bosons, and they acquire mass via a mechanism called the *Higgs mechanism*.

The Higgs field can be defined in terms of an additional field Φ , which can be expressed in terms of a scalar field $SU(2)$ doublet as:

$$\Phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.3}$$

[6] Leptons with chiral-odd symmetry: under the Poincare group transforming by the left-handed representation.

The Higgs potential can be written in terms of Φ as follows:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \tag{1.4}$$

where $\lambda > 0$. For the spontaneous symmetry breaking to occur, the minimum of this potential, corresponding to the vacuum state with the lowest energy, needs to be different from zero. For instance, if $\mu^2 > 0$, the minimum would occur when $\Phi = 0$, implying that the $W^\pm$ and Z bosons would be massless. Therefore, a necessary condition to find a minimum in which the Gauge bosons could be massive requires that $\mu^2 < 0$.

If $\lambda > 0$ and $\mu^2 < 0$, the Higgs potential expressed by the equation 1.4 has infinite non zero solutions (degenerated vacuum states), in which the norm of the Higgs potential is given by the following expression:

$$|\Phi|^2 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2\lambda} \tag{1.5}$$

The vacuum states are not invariant under the symmetry transformation of the group $SU(2)_L \otimes U(1)_Y$, but they are invariant under a subgroup, the $U(1)_{EM}$. This fact is known as spontaneous symmetry breaking, and is introduced via the Higgs field. The specific v value indicates the energy scale in which the electroweak symmetry breaking is introduced.

For instance, if one requires that ϕ_3 is the only non zero scalar field in equation 1.3, the Higgs field is expressed as:

$$\Phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.6}$$

The observed particles in nature are obtained by introducing small oscillations around the vacuum state, which can be parameterized as the Higgs field as follows:

$$\Phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} e^{i\frac{\vec{\kappa}\vec{\tau}}{v}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.7}$$

where the vector $\vec{\kappa}$ and the scalar $h(x)$ are small fields, giving to the Higgs potential four degrees of freedom. The three scalar fields contained in the vector $\vec{\kappa}$ are known as Goldstone bosons (massless particles).

Therefore, in the $SU(2)_L \otimes U(1)_Y$ symmetry group, once the Higgs field is introduced, there are 12 degrees of freedom given by four Gauge bosons (W^1, W^2, W^3, B^0) and four scalars introduced by the Higgs potential ($\vec{\kappa}$, $h(x)$). All of them, the Gauge bosons and the scalars, are massless particles.

Through the Higgs mechanism, an unitary transformation $U(\vec{\kappa})$ is applied to the field expressed in equation 1.7, such that the scalars fields contained in the $\vec{\kappa}$ vector disappear:

$$\Phi(x)' \Longrightarrow \frac{1}{\sqrt{2}} e^{-i\frac{\vec{\kappa}\vec{\tau}}{v}} \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.8}$$

If this transformation is applied to the Lagrangian of the SM, three of the four Gauge bosons acquire a mass in order to absorb the degrees of freedom removed by the Higgs field. This absorption is given by the coupling of the Gauge bosons with the Higgs field $h(x)$.

After considering the Higgs transformation, or equivalent to say, the symmetry breaking, there are three massive Gauge bosons ($W^\pm_\mu$, Z_μ), one Gauge massless boson (A_μ) and a scalar $h(x)$ with mass (summing up the 12 degrees of freedom). The mass of the new Gauge bosons ($W^{\pm}_{\mu}$, Z_{μ}) can be expressed as follows:

$$m_W = m_Z cos\theta_W = \frac{1}{2} g_W v \tag{1.9}$$

where θ_W is known as the mixing angle (or Weinberg angle) and g_W is the coupling constant of the weak interaction ($SU(2)_L$).

The Higgs mechanism expresses the masses of each of the fermions in the SM as follows:

$$m_f = \frac{1}{\sqrt{2}} g_f v, \tag{1.10}$$

where g_f is the Yukawa coupling constant of each fermion to the Higgs, and v is still the scale where the coupling occurs in the vacuum state.

Finally, the mass of the Higgs boson can be expressed as a function of λ and the vacuum scale of the symmetry breaking v as:

$$m_H^2 = 2\lambda v^2 \tag{1.11}$$

The vacuum expectation value has been measured to be $v = 246$ GeV. Therefore, this value sets the scale limit to the mass of the Higgs boson and the fermions, leaving λ and the g_f as free parameters to be determined experimentally, giving the exact mass of the Higgs boson and the fermions.

The Higgs boson was discovered by the CMS and ATLAS collaborations in 2012. Figure 1.3 illustrates the measurements in the $\gamma\gamma$ decay channel. The $\gamma\gamma$ invariant mass distribution is showing an excess of events near 125 GeV, indicating the discovery of this particle. The observed significance for this channel was $4.1 \cdot \sigma$. This measurement was performed by the CMS collaboration in proton-proton collisions at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. Similar results were obtained in other decay channels, from both Collaborations.

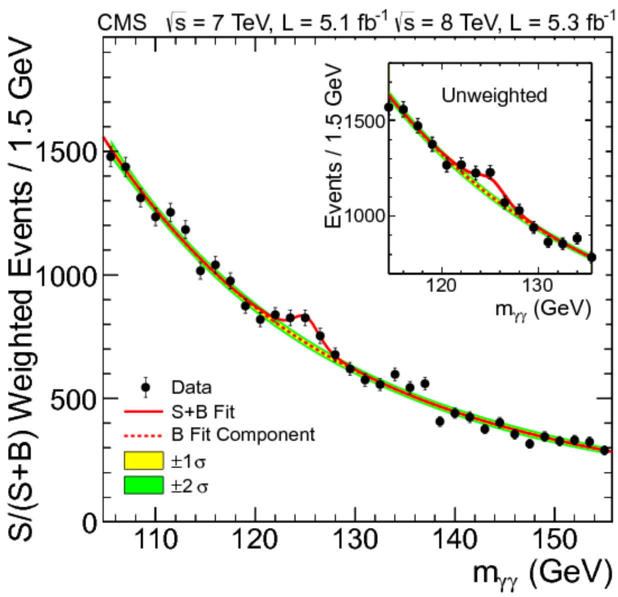


Figure 1.3: (taken from [7]) Discovery of the Higgs boson in 2012 by the CMS collaboration. The plot shows the diphoton mass distribution. The colour bands represent the ± 1 and ± 2 standard deviation uncertainties in the background estimate. The appearing peak indicates the Higgs boson.

1.2 Hadronic collision and the Monte Carlo event simulation

In this section, the theoretical framework for any experimental analysis is provided. The theoretical framework is based on the Standard Model and focuses on describing a hadron-hadron collision.

In a hadron-hadron collision, the partons inside the hadrons interact between each other through the strong interaction. The theory able to describe the ongoing processes is the QCD. These complex processes can be divided into two main parts, following the factorization theorem [21]:

- ✓ the hard process: where the hardest part of the collision, with the highest energy of interaction, takes place.
- ✓ the underlying event: counting for all the accompanying processes which underly the hard scattering and are not considered in the previous group.

The factorization theorem in the perturbative QCD theory [21] considers the differences between the soft (long distance) and hard (short distance) processes. The theorem settles boundaries between them. Assuming two incoming hadrons $h1$ and $h2$, two interacting partons a and b, and two outgoing partons c and d, the theorem factorizes the hadronic cross section as follows:

$$d\sigma^{h1h2\to cd} = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_{a/h_1}(x_a, \mu_F^2) f_{b/h_2}(x_b, \mu_F^2) d\sigma^{ab\to cd}(Q^2, \mu_F^2), \tag{1.12}$$

where $x_a(x_b)$ is the fraction of the longitudinal momentum of one of the colliding particles carried by the parton $a(b)$. The indices a and b run over all possible pairs of interacting partons from the incoming hadrons; f_a and f_b are the so-called Parton Distribution Functions (PDFs); Q is the energy scale of the interaction; μ_F is known as the factorization scale; and $\sigma_{ab\to cd}$ is the partonic cross section which gives the probability that two specific partons (a,b) interact between them and produce two outgoing partons (c,d).

In equation 1.12, the terms representing the Parton Distribution Functions count for the long-distance (low energy) processes. They are functions representing the momentum distributions of the partons in the colliding hadrons, giving the probability that a certain parton a carries a fraction of the momentum x_a of the incoming hadron. The processes described by those terms can be thought as independent processes, but they cannot be calculated in perturbation theory. Hence, they are not defined observables and must be extracted from measurements.

There are dedicated measurements performed to determine the PDFs. The PDFs are determined at a given scale interaction μ, and extrapolated to other scales by using evolution equations (i.e DGLAP [22]). Those measurements are mostly related to Deep-Inelastic-Scattering (DIS) collisions (for instance ep collisions), although they have been also determined using LHC data [23], e.g. the double differential inclusive dijet production cross section measurements at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV by the CMS collaboration [24]. Examples of the available PDFs set are: CTEQ1 PDFs [25], CT14 PDFs [26], HERAPDFs [27] [28], NNPDFs [29], among others. Figure 1.4 represents an example of PDFs included in the HERAPDF set [27], at $Q^2 = 10\,\text{GeV}^2$. The PDFs were obtained using a QCD analysis performed by the H1 [30] and ZEUS [31] collaborations. From the represented PDFs, it can be noticed that the valence quarks carry the largest amount of the longitudinal momentum (having high x values), while the gluons appear preferably at low x values.

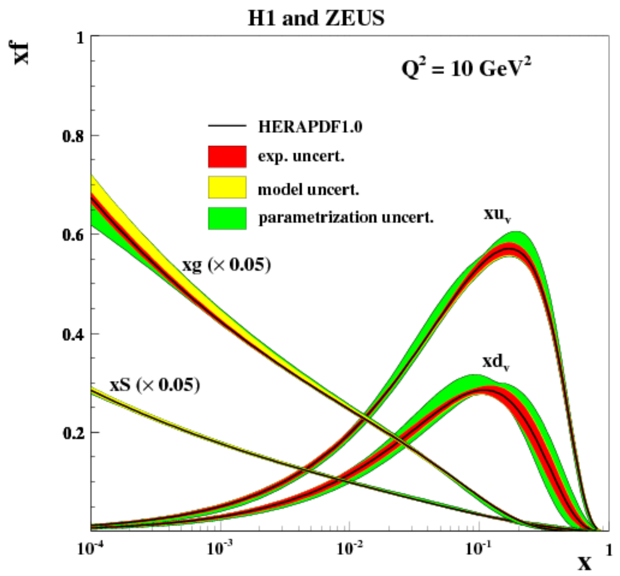


Figure 1.4: (Figure taken from [27]) The PDFs from HERAPDF1.9 set at $Q^2 = 10\mathrm{GeV}^2$. The functions are provided for up, down, strange quarks, and gluons.

The partonic cross section ($\sigma_{ab\to cd}$) expressed in equation 1.12, accounts for processes occurring at higher interaction energy (hard processes), and those can be estimated in perturbative QCD at a specific order of the Matrix Element expansion in terms of the strong running coupling α_s. The hard scattering contains only emissions above a certain energy μ_F, known as factorization scale, which sets the boundaries between hard and soft processes in a hadron-hadron collisions. The factorization scale is an arbitrary parameter which in theoretical calculations need to be considered. Usually, this value is taken as the hard scattering scale, although variations of its value might be considered to estimate its effects on the predictions.

Using perturbative QCD theory, the cross section can be expanded in terms of α_s as follows:

$$\sigma = \sigma^{LO}(\alpha_s^n) + \sigma^{NLO}(\alpha_s^{n+1}) + \sigma^{NNLO}(\alpha_s^{n+2})... \tag{1.13}$$

where σ^{LO}, σ^{NLO}, σ^{NNLO} are the cross sections corresponding to the Matrix Elements at leading order (LO), next-to-leading order (NLO), and next-to-next-to-leading order (NNLO), respectively. For instance, in pp collisions, for QCD processes $\sigma^{LO} \propto \alpha_s^2$, $\sigma^{NLO} \propto \alpha_s^3$, etc. The number of vertices in the representative Feynman diagrams [32] depend on the power of α_s, and the number of final partons after the hard scattering depend on how many real corrections are computed in the hard process calculations.

In order to have a more precise computation of the cross section, more terms in the expansion in the equation 1.13 need to be considered. However, each further term turns to be more challenging and nowadays there are only computations up to NNLO available. In order to improve a prediction at lower order of accuracy in the perturbative QCD expansion, a so-called *"K-Factor"* can be applied. For example, for correcting NLO predictions from NNLO calculations, this factor is expressed as follows:

$$K = \frac{\sigma^{NNLO}}{\sigma^{NLO}}, \tag{1.14}$$

in terms of the fixed-order cross sections, i.e NLO (σ^{NLO}), NNLO (σ^{NNLO}).

1.2.1 Monte Carlo event simulation: PYTHIA8 as a Monte Carlo at LO accuracy

A Monte Carlo (MC) event generator aims to model all the possible processes occurring during a hadron-hadron collision. In order to do that, the hadron collision is considered in its factorized way: first the hard scattering process and then the underlying events. The main differences between the available Monte Carlo event generators are the accuracy at which the ME in the hard process are estimated, and how the underlying events are simulated.

Figure 1.5 represents a schematic picture of a simulated *p*-*p* collision. In the simulation, first, the hard scattering is considered. In order to simulate the emissions produced by the interacting partons of the incoming protons, which occur before the hard process, a backward evolution is applied for simulating the *Initial State Radiation*. The *Final State Radiation* is simulated after the hard process. Finally, the resulting partons hadronise, building hadrons which consequently can decay into other particles. Additionally to the mentioned processes, the partons from the incoming protons that did not interact in the hard scattering can produce multi-parton interactions. Those last processes are also underlying the hardest process.

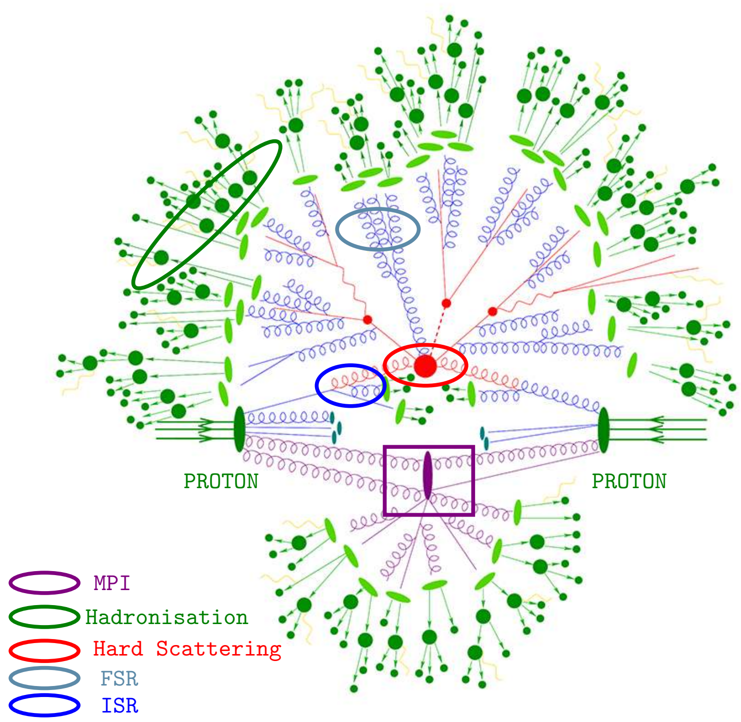


Figure 1.5: (figure modified from [33]) Schematic picture of a hadronic *p*-*p* collision and the subsequent interactions considered in a Monte Carlo event simulations.

In the following all of processes occurring within a *p*-*p* collision are described. The description is mainly based on how they are considered in the PYTHIA8 [34] MC event generator, which is one of the most used Parton Shower event generator to simulate *p*-*p* collisions [1].

[1]Other collisions like $p\bar{p}$, e^+e^- and $\mu^+\mu^-$ can also be simulated with Pythia8.

Hard Scattering
The hard process is considered in the perturbative QCD expansion at LO accuracy, meaning that only $2 \to 1$ and $2 \to 2$ processes are simulated. Extra radiations are only considered in the subsequent Parton Shower. The hard scattering considers different kind of processes: the hard QCD processes (i.e $qg \to qg$, $q\bar{q} \to q\bar{q}$, where q indicates u,d,c,s and b-quarks), the heavy flavour production (i.e $gg \to t\bar{t}$), among others.

Initial (Final) State Radiation
The incoming (outgoing) partons before (after) the main interaction can subsequently emit partons. Those emissions are simulated through Initial State Radiation (ISR) and Final State Radiation (FSR), depending on if they occur before or after the hard scattering. The radiations could also be considered within the hard scattering by real emissions not affecting the accuracy in the perturbative QCD expansion. However, in PYTHIA8 extra emissions are only considered as underlying events (Parton Showers) and not as part of the hard process.

A parton a can split into subsequent branches b and c with a certain probability characterized by a kernel $P_{a\to bc}(z)$, given by the expression $\int P_{a\to bc}(z)dz$, where z represents the fraction of energy from the initial parton that the emission carries. The probability that certain parton produces additional emissions is governed by the so-called virtuality scale (Q^2), which has different possible definitions. In PYTHIA8 the evolution of the emissions are p_T ordered, considering the virtuality parameter as $Q^2 = p_T^2 = z(1-z)m^2$. This means that the first emission is always producing a parton with higher p_T with respect to the subsequent emissions.

Multiple parton Interactions (MPI)
The multi-parton interactions describe processes taking place simultaneously to the hard interaction between two partons of the colliding particles which do not intervene in the main process. Since the protons are composed by multiple partons, the probability that those processes also occur is not negligible, playing an important role in pp collisions.

The occurrence of multiple-interactions can be predicted estimating the partonic cross section of the hard process. The latter has an increasing behavior towards the decrease of the exchanged transverse momentum among the interacting partons. Therefore, a minimum value for the exchanged transverse momentum needs to be defined ($p_{T_{min}}$). The cross section is having the following dependence:

$$\sigma_{hard}(p_{T_{min}}) \propto \int_{p_{T_{min}}} \frac{d\sigma}{dp_T^2} dp_T^2 \propto \frac{1}{p_{T_{min}}^2} \tag{1.15}$$

Considering $p_{T_{min}} \sim 3$ GeV, the σ_{hard} turns to be larger than the measured cross section for non-diffractive processes [35]. Such contradiction can be solved if multiple interactions occur at the same time. The number of the multiple interactions would correspond to the excess of the hard scattering cross section over the total cross section as follows:

$$n = \frac{\sigma_{hard}(p_{T_{min}})}{\sigma_{nd}} \tag{1.16}$$

However, the divergent behavior of σ_{hard} when $p_{T_{min}} \to 0$ cannot be explained by the existence of MPI. The gluon saturation phenomenon [36] can explain the need for a cut-off, implying that the exchange of transverse momentum between two partons with lower value than the cut-off cannot occur, given the decreasing of the coupling interaction α_s.

There are several models to describe multiple-interactions and the one used in PYTHIA8 is further discussed. The simulation of MPI relies on free parameters which need to be determined from fits to the measurements. There are several tunes [1] of the MPI parameters, based on different models and considering different sets of observables to perform the fits. The interactions are then ordered in p_T with a decreasing behavior, meaning that partons with higher p_T are first considered.

The partonic hard cross section can be regularized with the following criterion:

$$\frac{d\sigma_{hard}}{dp_T} \propto \frac{\alpha_s^2(p_T^2)}{p_T^4} \rightarrow \frac{\alpha_s^2(p_{T_0}^2 + p_T^2)}{(p_{T_0}^2 + p_T^2)^2} \tag{1.17}$$

where the definition of p_{T_0} depends on the center of mass energy as follows:

$$p_{T_0}(E_{CM}) = p_{T_0}^{ref} \cdot \left(\frac{E_{CM}}{E_{CM}^{ref}}\right)^{E_{CM}^{pow}} . \tag{1.18}$$

In the equation 1.18, $p_{T_0}^{ref}$ and E_{CM}^{pow} are free parameters, while E_{CM}^{ref} is usually taken as 7000 GeV. The MPI model also considers an impact parameter which differentiates the interactions occurring with different overlapping area (i.e a central collision is more active). An additional consideration is taken into account concerning the rearrangement of final-state colour connections. The idea is to reduce the overall length of the strings [1] connecting all the final particles. The probability that a system reconnects, introduces a new free parameter to the model, R. Such probability can be expressed as:

$$P = \frac{p_{T_R}^2}{p_{T_R}^2 + p_T^2} \text{ , } p_{T_R} = R \cdot p_{T_0} \tag{1.19}$$

The tunes of the MPI parameters are performed with measurements sensitive to the UE. The MPI parameters, in addition to the α_s in the Parton Shower, and hadronization parameters are considered in the fit. The observables which are most sensitive to MPI are the charged particles density and the p_T density, being the principal observables used to perform the fits. The following tunes (corresponding to the PYTHIA8 MC) are examples of the available tunes, and are further considered in order to simulate signal and background events in this thesis:

- ✓ CUETP8M1 Tune [37]: the tune is determine fitting data from UE at $\sqrt{s} = 0.9$, $\sqrt{s} = 1.96$ and $\sqrt{s} = 8$ TeV, simultaneously, and it is based on the Monash Tune [38]. The α_s and Parton Shower parameters are kept as in the Monash tune ($\alpha_s^{ISR,FSR} = 0.1365$).
- ✓ CUETP8M2T4 Tune [37] [39]: in this tune the Parton Shower parameters were optimized for $t\bar{t}$ events using collision data at $\sqrt{s} = 8$ TeV. The tune first considers the fit for determining α_s considered in the ISR using $t\bar{t}$ simulated events with POWHEG+PYTHIA8[1]. After this step α_s is tuned to lower values ($\alpha_s = 0.1108$). The second step consists of tuning the rest of the MPI parameters using UE and Minimum Bias measurements, keeping the α_s value from the previous step.
- ✓ Colour Reconnection Tunes [40]: based on the CUETP8M2T4 Tune, and considering two possible colour reconnection models, the QCD-inspired model [34] and the Gluon-Move

[1] A tune is a set of parameters determined from fits to the data.

[1] The hadronization process further explained is based on the Lund String model, connection all the gluons an quarks by coloured strings.

[1] Predictions obtained using POWHEG for simulating the hard process while PYTHIA8 for the Parton Shower and UE. It is further explained in more details.

model [34]. The tune is performed with UE and Minimum Bias observables using collision data at $\sqrt{s} = 13$ TeV.

Hadronization and Decays
The hadronization occurs after all the final colour-charged particles are produced, joining them together for building colourless particles: the hadrons. These processes take place when the interaction energy is of the order of ~ 1 GeV, not interfering neither with the Parton Shower, nor with the hard process. The colour-reconnection process could take place simultaneously though. There are several hadronization models used in Monte Carlo event simulations. In PYTHIA8, the Lund String model [41] is used. This model is a phenomenological approach considering all the highest energy gluons as lines binding each other due to the gluon self-interaction. The bindings are considered as they were strings. By considering the *breaking* of a string, a new quark-antiquark pair is created. This process continues until the interaction energy given by the string connection is below the energy needed to produce a new pair. Finally, after the hadronization process, the decays of the hadrons to stable particles are considered.

Parton remnants
The beam remnant considers those partons that did not contribute to the interaction, neither in the hard process, nor in the underlying events mentioned until now. However, those partons need to be considered for the colour rearrangement in the hadronization process.

1.2.2 Other LO Monte Carlos event generators: HERWIGpp and MADGRAPH5

HERWIGpp [42], as PYTHIA8, is a MC event generator taking into account LO accuracy in the ME calculations ($2 \to 2$ scattering processes), and considering the subsequent UE events previously described. The main difference relies on how the UE is simulated. The Parton Shower is evolved with the DGLAP equation but with angular ordering of the emissions. This means that the radiations are generated coherently rather than with increasing p_T values. The considered hadronization model is also different from PYTHIA8 using the Cluster Hadronization approach [43] instead. Additionally, for simulating the MPI, a fixed impact parameter is considered, and without colour-reconnection processes.

MADGRAPH5 is an event generator which only computes the Matrix Element at LO, and additional emissions need to be considered through a Parton Shower event generator like PYTHIA8 or HERWIGpp. On the other hand, this MC event generators considers additional real emissions within the hard process. Therefore, processes with $2 + n$ partons as outgoing particles are possible ($2 \to 2 + n$ scattering processes).

1.2.3 Monte Carlo event generators at NLO with Parton Shower

The Born process representing the hard interaction can be estimated with higher order accuracy. Two examples of MC event generators considering NLO accuracy are in the following discussed.

The idea behind both MC starts by writing the partonic cross section as follows:

$$d\sigma = B(\Phi_B)d\Phi_B\left[\Delta_{t_0}^{MC} + \Delta_t^{MC}\frac{R^{MC}(\Phi_R)}{B(\Phi_B)}d\Phi_r^{MC}\right] \tag{1.20}$$

where t is the radiated transverse momentum. The term $B(\Phi_B)d\Phi_B$, is the differential cross section at the Born level as function of the Born variables Φ_B. It is possible to estimate this term at higher order corrections in the perturbative QCD expansion. When the process is computed at NLO, both real and virtual emissions need to be added in order to have a non divergent result. The additional term in the previous equation represents two different probabilities: the $\Delta_{t_0}^{MC}$ factor gives the probability that no radiation occurs down to the scale cut-off t_0, while the second term: $d\Delta_t^{MC} = \Delta_t^{MC} \frac{R(\Phi_R}{B(\Phi_B)} d\Phi_r^{MC}$, gives the probability that no radiation is emitted down to the scale t. The factor Δ_t^{MC} is the so-called Sudakov Form Factor [44]. The two different approaches to consider the partonic cross section interfered with a Parton Shower MC are: the POWHEG method and the MC@NLO method.

The POWHEG method stands for *Positive-Weight Hardest Emission Generator* and is a method for interfacing next-to-leading order (NLO) calculations of the hard process with Parton Shower, following the ideas expressed by the equation 1.20. The peculiarity of this generator is that the hard process is estimated completely independent from the Parton Shower. That means that above a certain cut ($t > t_{\text{POHWHEG}}$), the POWHEG program generates all the radiations, while below that values ($t < t_{\text{POHWHEG}}$) all the radiations are generated by Parton Shower event generators (i.e PYTHIA8 or HERWIGpp).

In the POWHEG method, an important quantity is the so-called HDAMP parameter. This is a non-physical parameter, controlling through a damping function [1] the resummations of higher orders in the Sudakov from factor, without spoiling the NLO accuracy of the ME. This is preformed by rescaling the probability of a real-emission by a damping function $R \rightarrow D \cdot R$ in which the parameter HDAMP is considered. If the HDAMP parameter is not considered, the resummation is performed up to the scale of the hard process, leading to unrealistic results.

The MC@NLO method similarly as POWHEG, estimates the hard process at NLO accuracy, but the difference relies on the fact that the Parton Shower is not treated independently. This means that real emissions are considered within the hard process, merging the virtual and real emissions at NLO. Therefore, this method could lead to additional issues given that the MC Parton Shower need to reproduce exactly the soft collinear singularities of the radiation in ME.

Matching ME with Parton Shower events generators

The main problems with considering ME at higher order accuracy and Parton Shower is that double counting needs to be avoided. The matching procedure is responsible for summing correctly the ME and Parton Shower calculations. This procedure has different approaches, as for example the so-called CKKW [45] and MLM matching schemes [45]. The first one consists of a veto algorithm for the Parton Shower, meaning that the Parton Shower is truncated below the lowest scale of the emissions produced in the ME. The MLM method on the other hand, works as event by event rejection, where events are removed if partons are emitted by the Parton Shower with a higher scale than the ones considered in the ME. The POWHEG method counts with its own approach for the matching procedure, since the ME settles a limit for the hardest emission in the Parton Shower [46]. In the MC@NLO method, the matching procedure is performed by subtracting part of the contribution of the showering from the total cross section to avoid double counting [47].

[1] A damping function is a function taking values ± 1.

1.3 The top-quark physics

The top-quark cross section at the current LHC energies contributes approximately 10^{-9} to the total p-p cross section collision [1]. Even if this value seems a small contribution, the top-quark production is among the processes with sizeable cross section at the LHC energies. This section focuses on describing the main features of the top quark and its role for testing the boundaries of the SM.

1.3.1 Top quark production at the Tevatron and LHC scales

The top quark was discovered in 1995 [48] [49] at the Tevatron experiment (Fermilab) by the CDF [50] and D0 [51] collaborations. At the Tevatron experiment, proton-antiproton ($p\bar{p}$) collisions took place at the TeV scale (referring to the center of mass energy of the collisions, $\sqrt{s}$). Specifically, the top quark discovery was performed by colliding particles at $\sqrt{s} = 1.96$ TeV. Both collaborations published in 2014 a combined result for the measured cross section of top-quark pair production[2] of $\sigma_{t\bar{t}} = 7.60 \pm 0.41$(stat+syst) pb [52], which is consistent with the SM predictions: $\sigma_{t\bar{t}} = 7.35 \pm -0.27$(scale+pdf) pb [53]. The given SM prediction for the $t\bar{t}$ production[3] corresponds to calculations at next-to-next-leading order (NNLO) in perturbative QCD, considering soft gluon resummation at next-to-next-to-leading logarithmic accuracy (NNLO-NNLL).

The LHC, colliding protons at a center of mass energy of $\sqrt{s} = 13$ TeV (highest operating energy up to now), is considered as a top quark factory since the predicted SM cross section for top-quarks pairs is at least two orders of magnitude larger than at the Tevatron. A large amount of measurements have been performed using data collected by CMS [55] and ATLAS [56] experiments at lower ($\sqrt{s} = 1.96$ TeV,$\sqrt{s} = 5$ TeV, $\sqrt{s} = 7$ TeV, $\sqrt{s} = 8$ TeV) and higher collision energies ($\sqrt{s} = 13$ TeV). Some of the results of the $t\bar{t}$ pair cross section at different energies are summarized in Figure 1.6, where the SM predictions at NNLO-NNLL accuracy, in the energy range from 1 TeV up to 13 TeV (for pp and $p\bar{p}$ collisions) are compared to them. Details on some of the measurements can be found in the references [57, 58, 59, 60, 61, 62]. The measured cross section corresponding to the channel where the W-bosons from the $t\bar{t}$ pair decay into two $e\mu^+$ pairs [4] at $\sqrt{s} = 13$ TeV is $\sigma_{t\bar{t}} = 815 \pm 9(\text{stat})^{+38}_{-19}(\text{syst})$ pb, which is in agreement with the SM prediction of: $\sigma_{t\bar{t}} = 815^{+20}_{-29}(\text{scale}) \pm 35(\text{PDF})$ pb [63]. All the shown measurements agree with the SM predictions within uncertainties.

The top quark can be produced either through the strong interactions in pairs ($t\bar{t}$), or through the weak interaction as single top quarks. In hadron-hadron collisions the top quark production occurs mainly in pairs. The cross section of $t\bar{t}$ production is approximately four times larger than the single-top cross section (136 pb [5]) [65].

The $t\bar{t}$ pair is produced in two different ways: gluon fusion ($gg \to t\bar{t}$) or quark-antiquark annihilation ($q\bar{q} \to t\bar{t}$). The corresponding Feynman diagrams at LO are shown in Figure 1.7. At

[1] At $\sqrt{s} = 13$ TeV, the bottom quark production is approximately six order of magnitude higher, while the Z-boson and W-boson production two and three order of magnitude higher.

[2] top-quarks are produced in pairs.

[3] considering top mass 172.5 GeV, $\alpha_s(M_Z) = 0.118$ and the Parton Distribution Functions corresponding to the NNPDF3.0 set [54].

[4] The $t\bar{t}$ pair decays on the following way: $t\bar{t} \to bW \to be^-\mu^+$.

[5] The given cross section refers to the process represented in the second Feynman diagram in Figure 1.8, which is the largest single-top channel.

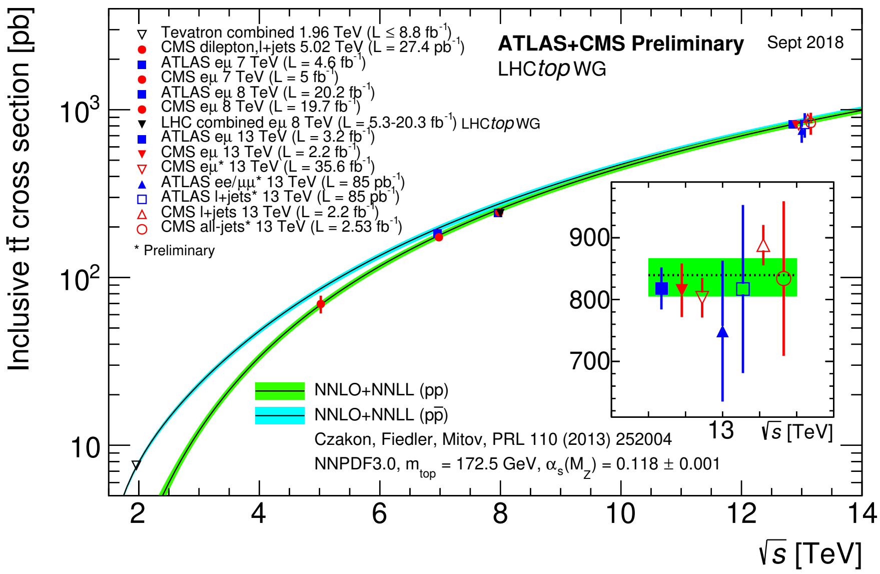


Figure 1.6: (taken from [64]) Summary Measurements of the inclusive $t\bar{t}$ cross section in pp and $p\bar{p}$ collisions compare to SM theory predictions. The SM predictions correspond to NNLO-NNLL accuracy of the QCD processes [53], while the coloured bands represent the theory uncertainties.

the Tevatron interaction energies, the predominant channel to produce top-pair was the quark-antiquark annihilation channel, while at the LHC energies, 85% − 90% of top-pairs are produced by the gluon fusion channels.

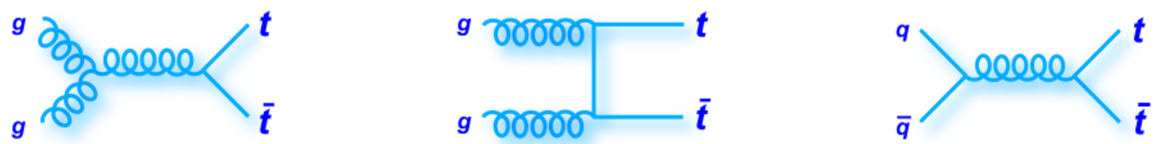


Figure 1.7: (taken from [66]) Feynman diagrams representing top-quark pair production at leading-order.

The single-top are produced through the exchange of charged-electroweak currents ($W^{\pm}$) mediated by the weak interaction. There are three main single-top production channels represented by their Feynman diagrams at LO in Figure 1.8. The two first diagrams illustrate channels where the production occurs by exchanging a W-boson. The third diagram is the so-called associated Wt production channel, where the W-boson is an on-shell particle (real) and produced together with a t-quark. The second diagram represents the channel with the relative highest cross section, at both Tevatron and LHC energies. A summary of the SM predictions and measurements for single top production can be found in reference [67].

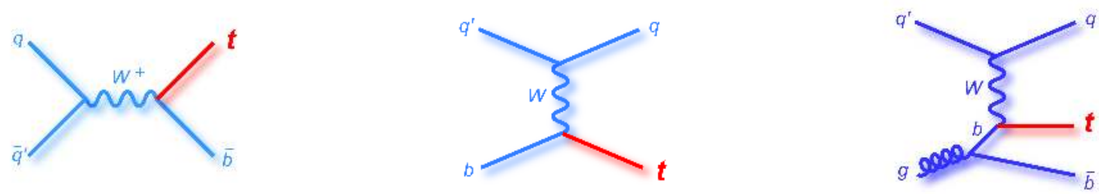


Figure 1.8: (taken from [66]) Feynman diagrams representing single top-quark production at leading-order.

1.3.2 Top quark decay channels

The top quark is the heaviest elementary particle in the Standard Model and the only particle decaying before the hadronization time scale. The life time of the top quarks is approximately 20 times shorter than the time when the hadronization occurs:

$$\tau_{had} \approx h/\Lambda_{QCD} = 2 \cdot 10^{-24}\text{s} \qquad \tau_{top} \approx h/\Lambda_{top} = 5 \cdot 10^{-25}\text{s} \tag{1.21}$$

The decay of a top quark is mediated by the exchange of charged electroweak current ($W^{\pm}$ bosons), hence, the fact of being able to study those decays provides the possibility of directly testing the electroweak interaction of the Standard Model. This property is used for fundamental researches in order to study general properties of the top quark, like mass, cross sections, spin correlations, charge asymmetries, Yukawa coupling to the Higgs boson, among others. Additionally, since the top quark is the only quark whose mass is larger than the W-boson, the latter is produced on-shell as a real particle.

When the top quark decays by exchanging a charged electroweak current ($W^{\pm}$), its flavour-charge changes preferably to the b-flavour. The CKM matrix predicts a probability of 99.9% that the top quark decays into a bottom quark with respect to the probabilities that for instance a down or a strange quark is produced. Therefore, the decay products of a top quark are a real W-boson and a b-quark.

The different decay channels are then defined by the respective W-boson decay channels. Since the top quark is mostly produced in pairs, for each $t\bar{t}$ pair there are two W bosons ($W^{\pm}$). The W bosons can decay either into a quark-antiquark pair, or into a lepton-antilepton pair. The lepton-antilepton pair is for example $e^+\nu_e$, $\mu^+\nu_\mu$, $\tau^+\nu_\tau$ (always a pair of charged-lepton with a neutrino). Figure 1.9 (right picture) represents all the possible combinatorial decay modes of a $t\bar{t}$ pair. The branching ratios for each of the channels (left picture) are also shown.

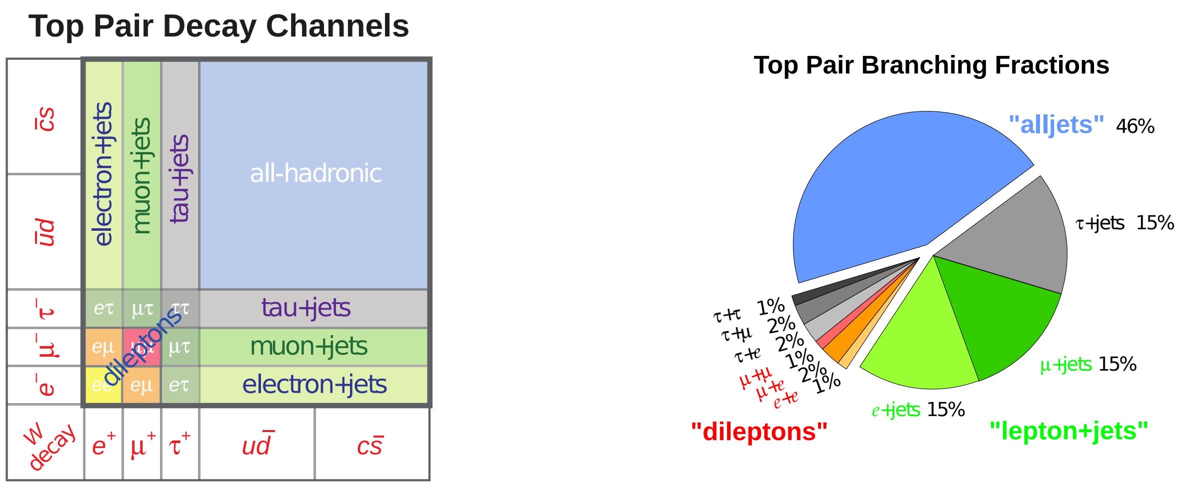


Figure 1.9: (taken from [66]) (Left picture) Possible decay channels for the $t\bar{t}$ pair given by the decays of the W-bosons. (Right picture) relative probabilities for each of the decay channels. In the case of the leptonic decays, the corresponding neutrino accompanying the leptons is not mentioned (i.e $W^+ \to e^+\nu_e$ is recognized as e^+ channel).

The possible channels previously illustrated are:

- ✓ all-hadronic: each of the W-bosons ($W^{\pm}$) decays in a quark-antiquark pair.
- ✓ semi-leptonic: one W-boson decays into a quark-antiquark pair, while the other into a lepton-antilepton pair,
- ✓ dileptonic: each of the W-bosons ($W^{\pm}$) decays into a lepton-antilepton pair.

The channel with highest branching ratio is the so-called *all-hadronic* channel, having $\sim$ 46% of relative probability. Even if this channel has the highest cross section, the final state signature are highly contaminated by QCD processes, whose cross section is at least two orders of magnitude larger than the $t\bar{t}$ pair production. Therefore it is considered a challenging channel to be studied.

In the Figure 1.9, the channels representing the semi-leptonic decay mode are named as *lepton+jets*, where the leptons could be either a τ, μ or e. Jets refer to a bunch of collimated particles flying in the same direction, in this case originated by quarks from the hadronic decay of one of the W-bosons. More details about jet physics are provided in Chapter 2. When a $t\bar{t}$ pair decays in the all-hadronic channel, up to six possible jets could be expected (two b-jets and four originating from each of the W-bosons decaying hadronically).

The specific channel studied in this thesis is illustrated in a Feynman diagram (at LO) in Figure 1.10, and corresponds to the all-hadronic decay channel. When a t-quark is produced with very high p_T, the decay products are boosted in the direction of the t-quark, and they can be clustered within a single jet. If each of the t-quarks are having high p_T, two jets (illustrated as cones in the figure) are originating from each of the quarks and are known as top-jets. This thesis focuses on those scenarios known as boosted topologies.

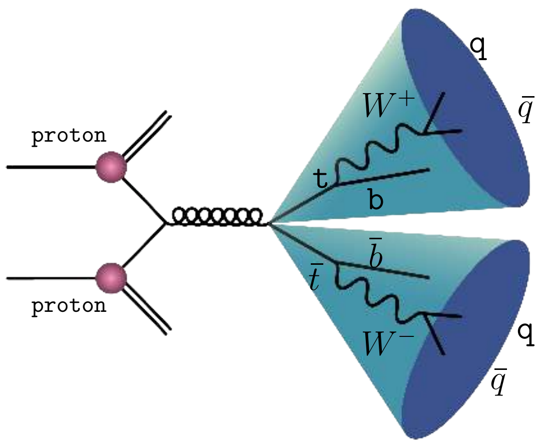


Figure 1.10: Feynman diagram representing the full hadronic decay channel of the $t\bar{t}$ pairs. Additionally, two jets are clustering all the decay products of each of the t-quarks, representing the boosted regime, occurring when each of the t-quarks has high transverse momentum.

1.3.3 Top quark mass measurements

The mass of the fundamental particles in the SM cover a range of approximately 12 orders of magnitudes (difference among the lightest and the heaviest mass). The top quark sets one of the boundaries of the mass phase space in the SM, being the heaviest particle. Furthermore, the Yukawa coupling between the Higgs boson and the top quark is the only coupling constant which is supposed to be near to unity. Therefore, the estimation of the mass of the top quark is crucial for precise tests of the SM predictions.

The QCD Lagrangian, however, does not predict directly the mass of the quarks, leaving those quantities as free parameters in the SM. Therefore, the mass of the particles is not an observable and needs to be estimated by the underlying theory, using experimental data. The estimated value usually depends on the definition of the renormalization schemes (i.e pole mass, $\overline{MS}$, etc.).

Other definition of the mass is the so-called Monte Carlo mass (m_t^{MC}), which is an effective parameter in the Monte Carlo simulation. This parameter is usually determined by calibrations performed in QCD predictions of some special defined observables, as well as by directly measuring it.

The method used to directly measure the m_t^{MC} parameter relies on the kinematic reconstruction of the top-decay products. An up-to-date summary of measurements for this parameter performed by the CMS collaboration is illustrated in Figure 1.11. The CMS measurements are also compared to combined Tevatron measurement and a world combined measurement including results from other experiments.

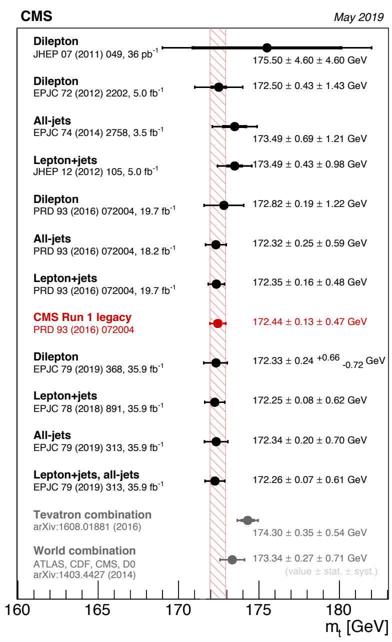


Figure 1.11: (taken from [68]) Summary of CMS measurements for the determination of the Monte Carlo top quark mass. The measurements are based on directly reconstructing the top decay products.

1.3.4 Resummation for top-quark pair production in the boosted regime.

The azimuthal angle separation ($\Delta\phi$) between the two jets originating from top-antitop pairs having high p_T can be experimentally measured. In the boosted regime, the $t\bar{t}$ pair is preferably having an azimuthal separation near to π, implying that the transverse momentum of the system is small ($q_T \sim 0$). This phase space is particularly interesting.

In the low q_T region, difficulties for fixed order theoretical calculations start to appear, given the enhancement of the cross section when $q_T \leq M$ from logarithms terms in the ME calculations. Those logarithms terms are in the form $\alpha_s^n \ln^{2n} M^2/q_T^2$, mainly caused from the quasi-collinear emissions of gluons. Figure 1.12 shows a Feynman diagram representing a $t\bar{t}$ process at NLO, where a gluon is emitted with transverse momentum q_T, originating an azimuthal separation between the two top-quarks different than π. The emission is given by a LO process, and hence its cross section diverges when $q_T \to 0$. Those calculations need resummation to all orders either by analytic methods or by considering Parton Shower.

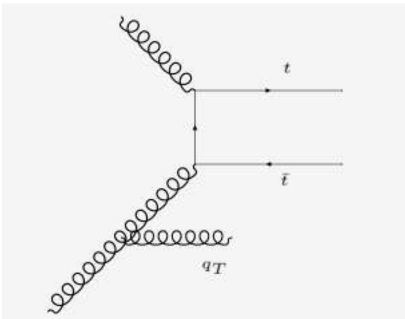


Figure 1.12: Feynman diagram representing a $t\bar{t}$ pair production process, in which, additionally, a gluon with transverse momentum q_T is emitted.

In Figure 1.13, the NLO calculation of the cross section differential in q_T for $t\bar{t}$ processes (illustrated in Figure 1.12) at $\sqrt{s} = 8$ TeV are compared to predictions where additionally resummation corrections at NLL are included [69]. As it might be noticed, the NLO predictions are significantly enhanced when $q_T \to 0$, while the resummed prediction heals the divergent behavior.

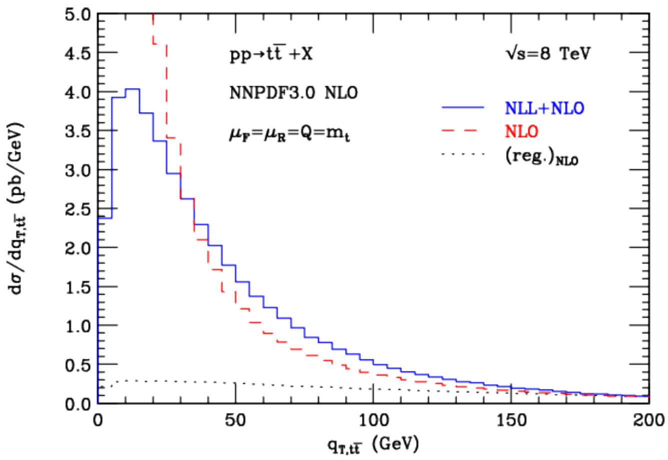


Figure 1.13: (taken from [69] Transverse momentum cross section of the $t\bar{t}$ pair, at $\sqrt{s} = 8$ TeV, computed at NLO accuracy in the ME expansion, compared to predictions where additionally resummation corrections at NLL are accounted (NLL-NLO).

An additional interesting aspect appearing in $t\bar{t}$ topologies is the so-called factorization breaking phenomenon [70]. This issue starts to appears when resummations at NLL accuracy are computed, in coloured final state objects, connecting the initial and final state radiation by an additional term in the factorization QCD theorem. Figure 1.14 illustrates schematically this phenomenon. The interaction process can be factorized on the following way:

$$W_{ab}^{F}(s,Q,b) \sim C_{ca}\left(\alpha_s(b_0^2/b^2), z1\right) C_{\bar{c}b}\left(\alpha_s(b_0^2/b^2), z2\right) \sigma_{c\bar{c}}\left(Q^2, \alpha_s(Q^2)\right) S_c(Q,b) \tag{1.22}$$

where the factors $C_{cx}(\alpha_s(b_0^2/b^2), z1)$ represents the collinear radiations of a parton x (carrying a fraction of momentum z_1 from the incoming proton i.e f_b) at scale $1/b$; $S_c(Q,b)$ is referred to the Sudakov form factor representing the soft and flavour conserving collinear radiation in the scale $1/b \leq q_T \leq M$, where M is the mass of the interacting partons; and $\sigma_{c\bar{c}}(Q^2, \alpha_s(Q^2))$ is the hard process occurring between parton c and $\bar{c}$, at a scale approximately of the top-quark mass. The factorization breaking introduces an additional term (Δ) to this formula, to account for the colour-connection between the initial state radiation (contained in the Sudakov factor) and the final state radiation:

$$W_{ab}^{F}(s,Q,b) \sim C_{ca}\left(\alpha_s(b_0^2/b^2), z1\right) C_{\bar{c}b}\left(\alpha_s(b_0^2/b^2), z2\right) \sigma_{c\bar{c}}\left(Q^2, \alpha_s(Q^2)\right) S_c(Q,b) \Delta \tag{1.23}$$

This Δ factor is caused from soft radiation at large angles with respect to the partons that are interacting.

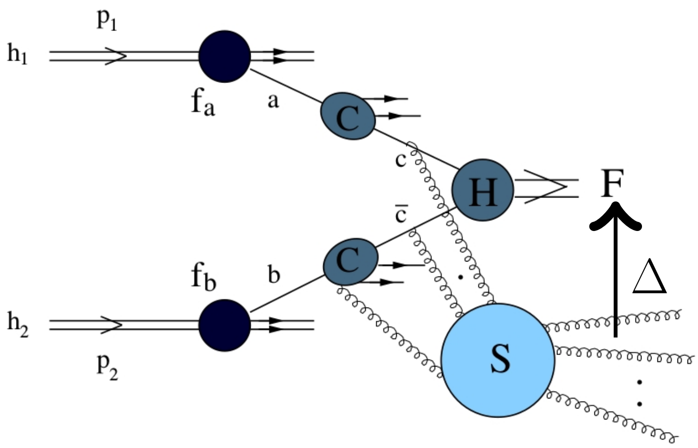


Figure 1.14: (modified from [70]) Factorization theorem of the hard process: the factors C represents the collinear radiations of a parton a or b at scale $1/b$; S is the Sudakov form factor representing the soft and flavour conserving collinear radiation in the scale $1/b \leq q_T \leq M$, M is the top-quark mass, and H represents the hard process calculated at a fixed order. The factorization breaking introduces an additional term (Δ) to account for the colour-connection between the initial state radiation (contained in the Sudakov factor) and the final state radiation.

The accuracy for available fixed order calculations considering on-shell top quark production (either as total cross section or differential distributions) correspond to the NNLO in the perturbative α_s expansion. Examples for those calculations are given in references[71, 72, 73, 74]. In the available Monte Carlo generators, from which predictions can be obtained with NLO accuracy in the ME, the resummations effects are considered at leading order through the Parton Shower, and additionally colour reconnection models are implemented. The measurements performed in this thesis test whether the MC approach correctly covers resummation and factorization breaking effects. If discrepancies are found, it could indicate the need of considering resummation effects at higher orders to cover those discrepancies.

Chapter 2

Jet clustering algorithms and jet substructure for boosted topologies

Quarks and gluons cannot be directly observed. Instead, final objects are defined, which are then measurable by an experiment. From those defined objects, the basic partonic information can be extracted. Those objects are called jets and they provide a broad window for accessing almost all QCD properties.

This chapter aims to cover the physics behind the origin of the jets, as well as to give the necessary tools to deal with any analysis where boosted heavy jets [1] are involved. The chapter is divided into two main parts. In the first section (section 2.1), jets are presented as observables for LHC physics and the most used jet finding algorithms are described. In the second section 2.2, jet substructure techniques are explained. Those substructure techniques are further considered in Chapter 3, in order to provide a definition for a hadronic top jet.

[1] jets containing the decay products of an heavy object, i.e t-quark.

2.1 Jet definition

This section focuses on presenting jets as defined measurable objects. The theory behind the origin of the jets and how they are defined by different jet clustering algorithms is discussed.

2.1.1 Jets as result of soft-collinear QCD emissions

A jet is defined as a bunch of particles flying in the same direction. The reason that a bunch of particles appears in a collimated configuration, is due to the soft-collinear nature of perturbative QCD.

In perturbative QCD approach, the probability that a certain parton (quark or gluon) radiates an additional gluon is described by the following expression [75]:

$$p \propto \int \alpha_s \frac{dE}{E}\frac{d\theta}{\theta} \tag{2.1}$$

where α_s is the QCD strong coupling, E is the energy of the emitted gluon, and θ is the angle between the initial parton and the emitted gluon. According to this equation, soft (low energy gluons) and collinear ($\theta \to 0$) emissions are enhanced. Those soft and almost collinear emissions imply the presence of considerable amount of radiation surrounding the initial parton, giving origin to jets.

Since jets are defined at particle level, and the hadronization of partons is a non-perturbative phenomenon, the relation between the initial partons originating the jet, and the particle forming the jet, might be sensitive to non-perturbative corrections. Furthermore the hadronization transforms partons to colour-singlet (colourless) hadrons, while partons have colour-charge. Nevertheless, those corrections are considered to be of the order of λ_{QCD}/E (E is the energy of the jet), and hence the hadrons forming a jet can be directly related with the initial parton originating the jet.

Precision measurements of inclusive jet production is a window for testing QCD predictions at different order of the perturbative calculation. For instance, the differential cross section with respect to the jet transverse momentum can be sensitive to the Parton Distribution Function, and to α_s. In order to perform QCD studies, jet measurements are a crucial point. Dedicated studies at different center of mass energies have been provided at the LHC, from the CMS collaboration (e.g $\sqrt{s} = 2.76$ TeV [76] , $\sqrt{s} = 7$ TeV [77] , $\sqrt{s} = 8$ TeV [78] , $\sqrt{s} = 13$ TeV [79]), and from other experiments (i.e ATLAS). Figure 2.1 shows the inclusive cross section measurements at $\sqrt{s} = 13$ TeV as function of the jet p_T [79]) compared to theory predictions. The predictions are able to describe the measurements with reasonable agreement, confirming that jets are related to the underlying partonic configuration.

Another observable sensitive to QCD effects is the azimuthal angular correlation ($\Delta\phi$) between the jets. Examples of measurements with respect to $\Delta\phi$ between the two leading jets are illustrated in Figure 2.2. The measurements are compared to predictions provided with the POWHEG+PYTHIA8 event generator, and considering different p_T regions. The theory predictions describe the measurements across the whole phase space. In Appendix B, details of this measurement are provided, which is published in The European Physical Journal C (Eur. Phys. J. C (2018) 78:566), and to which I contributed significantly.

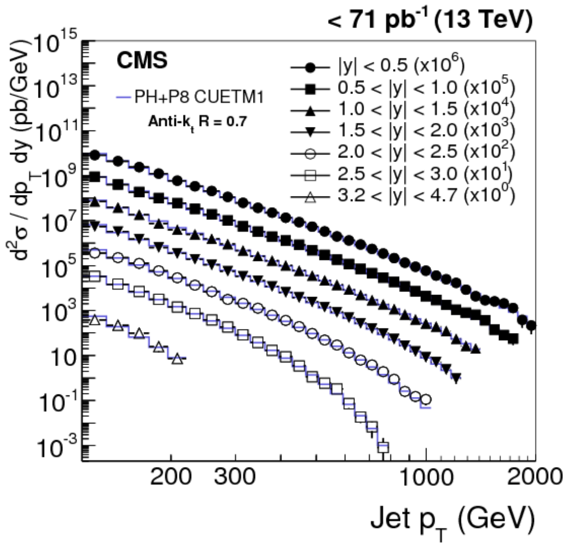


Figure 2.1: (taken from [79]) Double differential cross section as function of p_T of the jet. Jets are clustered with *anti-k_t* algorithm with radius parameter $R = 0.7$. The solid lines represent predictions at NLO produced with POWHEG+PYTHIA8.

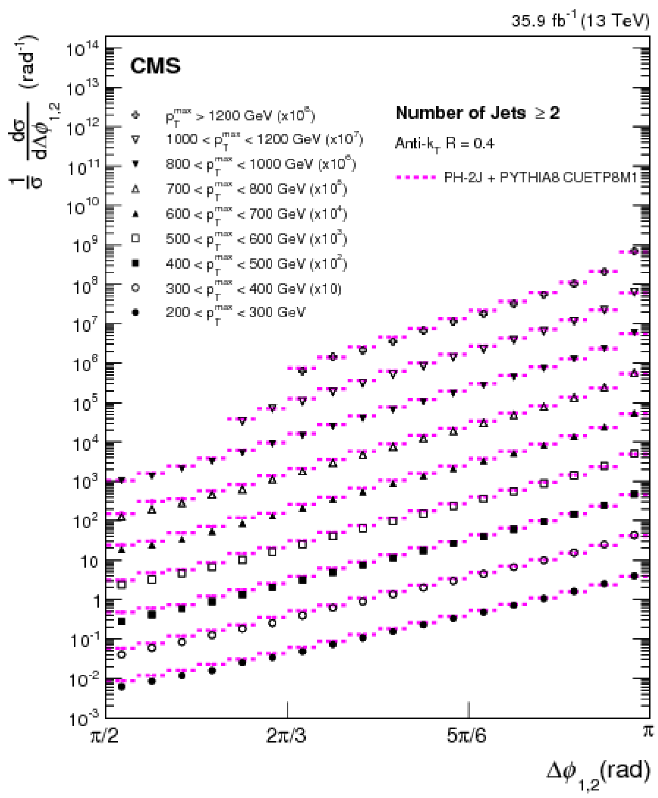


Figure 2.2: (taken from [80]) Normalized cross section differential in the azimuthal separation between the two leading jets in different p_T regions. The solid lines represent predictions for dijet calculations at NLO produced with POWHEG+PYTHIA8.

2.1.2 Jet mass

Jets originating from quarks are different from jets coming from gluons. The main difference relies on the energy distribution inside the jet. For instance, jets originating from gluons tend to have more constituents with larger opening angle radiation. This is due to the fact that the probability that a gluon emits is approximately twice larger than for a quark. There have been several developments of new techniques which allow to distinguish between those kind of jets [81]. Most of the algorithms are based on multivariate techniques taking as input variables observables sensitive to the energy distributions inside jets, like the jet mass and observables related to jet substructure.

The jet mass can be estimated at any order in perturbative QCD. For instance, at LO [1], this quantity has the following dependence as function of the jet radius (R) and p_T [82]:

$$m^2 \propto \frac{\alpha_s}{\pi} p_T^2 R^2 \tag{2.2}$$

Hence, at this order of accuracy, under the consideration that α_s is approximately constant, a linear dependence of the jet mass with respect to the p_T and the radius of the jet is predicted. Such behavior is valid for both, quark-jets and gluon-jets.

The linear behavior between the jet mass with respect to the jet radius and the p_T, is no longer valid when, for instance, further emissions in the parton shower are considered, and furthermore, different behaviors from quark-jets and gluon-jets arise [83].

Figure 2.3 shows, indeed, the behavior of the quantity defined as $\rho = \frac{m^2}{p_T^2 R^2}$, when the jet mass is estimated in simulated events with PYTHIA6 [84] [2], considering parton showering. The labeled plain mass is referred to the mass estimated by the MC simulation. Two plots are shown, one considering quark jets and the other one considering gluon jets.

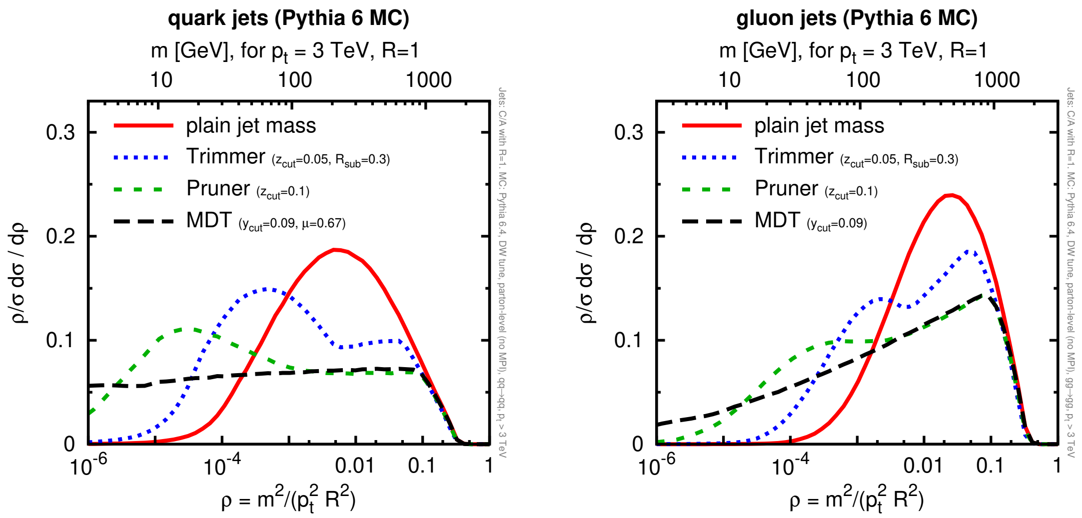


Figure 2.3: (taken from [83]) ρ distribution considering different groomer taggers (trimming, prunning, mass drop), and the plain jet mass estimated with MC for: (left) quark jets (right) gluon jets. The upper x-axis shows the values of the jet mass, specifically considering jets of radius $R = 1$, and $p_T = 3$ TeV.

[1] considering diagrams of real gluon emission from a partons.
[2] a previous version of PYTHIA8 MC.

As might be noticed, the ρ distribution is not flat, but having approximately a Gaussian distribution centered at a certain value, which is different for gluon jets with respect to quark jets. The gluon jets tend to have larger central values in the ρ distribution. This fact directly implies that, considering two jets originating from both sources (quark and gluons), and having the same radius R and p_T, the gluon jet will have larger mass.

In the presented plots, not only the plain mass was shown, but also the jet mass when grooming techniques are applied to the jet. Grooming techniques are cleaning techniques aimed to remove unwanted radiation (generally soft-wide radiation coming from uncorrelated processes). This topic will be discussed in section 2.2.3. However, it is worth to notice here, that for quark jets, when the Soft Drop Mass tagger (MDT) [85] as grooming technique is applied, the ρ distribution below certain limit has a flat behavior. Such tendencies are used to distinguish jets from different source.

2.1.3 Jet definition and clustering algorithms

The definition of jets as final objects has two main purposes: they are observables directly measurable, and they can be used for extracting specific QCD properties of the original partons. In order to define those final state objects, jet algorithms have been developed. A jet algorithm maps the momentum of the final particles into the momentum of a certain number of jets. The jet definition is an ambiguous concept depending on the specific selected algorithm to cluster their constituents. Figure 2.4 shows two recorded events by the CMS experiment, where one can clearly distinguish 3-jets and 4-jets event topologies.

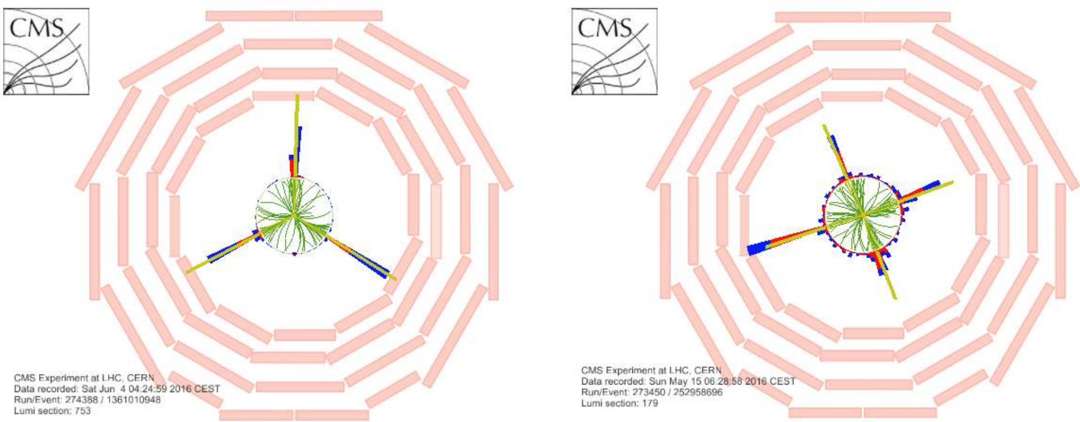


Figure 2.4: (taken from [86]) Display of events recorded by the CMS experiment in 2016 (left) for three jet topology and (rigth) for four jet topology.

Two assumptions for jet definitions need to be considered. The first one is related to the determination of how particles are grouped into jets (the jet clustering algorithm). The second assumption concerns how global jet magnitudes, i.e the transverse momentum of the jet, are defined (recombination schemes). For instance, the momentum of a jet is defined as the four-momentum sum of each of the constituent particles.

The main requirement that any jet algorithm must satisfy is that the defined objects should be invariant under infrared and collinear emissions, to ensure that perturbative QCD calculations at higher order for jet observables are not divergent. Furthermore, since detectors cannot resolve neither full collinear nor full infrared event structure, this requirement is also needed experimentally. Jet algorithms have free parameters which are selected in such a way that the final jet is as little as possible sensitive to non perturbative QCD effects (i.e: underlying event, hadronization), pileup and detector effects.

There are two groups of jet clustering algorithms: the sequential recombination algorithms, and the cone algorithms. Cone algorithms are based on finding regions where the deposited energy is higher. Stable cones are defined around the energy flow. Often, jet observables defined with these types of algorithms are not collinear infrared safe. There is only one type of cone algorithm, infrared and collinear safe defined: the SIS Cone algorithm [87].

The CMS and ATLAS Collaborations at the LHC use mainly the sequential recombination algorithms. The algorithms that are further described in this section belong to this group and are: the k_t algorithm [88], the Cambridge-Aachen algorithm [89], and the anti-k_t algorithm [90].

Illustrative pictures of the sequential clustering algorithms are shown in Figure 2.5. The main difference between the algorithms is in the projection in the azimuthal-rapidity (ϕ-y) plane, where the k_t algorithm and the Cambridge-Aachen algorithm are represented by irregular jet shapes, while by using the anti-k_t algorithm, jets are represented by circular cone shapes.

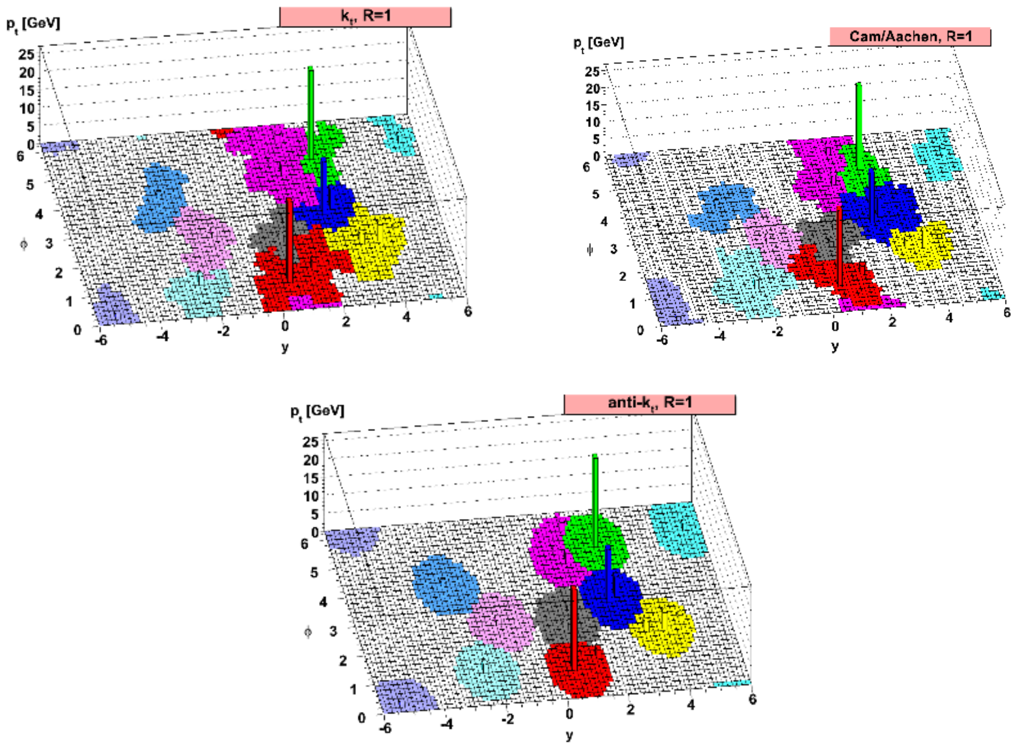


Figure 2.5: (taken from [90]) Illustrative picture of the three most used jet sequential clustering algorithm: (upper left) the k_t algorithm, (upper right) the Cambridge-Aachen algorithm and (down) the anti-k_t algorithm. The geometric jet areas for each algorithm is illustrated in the ϕ-y plane.

2.1.3.1 The k_t jet clustering algorithm

The k_t jet clustering algorithm was created for e^+e^- collisions [91] and later modified in order to be used for hadron-hadron collisions [88].

The physics behind the formulation of the algorithm relies on the concept that a jet is a result of successive parton branching. The QCD branching probability that one gluon splits into two gluons ($g \to g_i g_j$) can be described by the following equation [82]:

$$d\sigma_{g\to g_i g_j} \propto \frac{dE_j}{\min(E_i, E_j)} \frac{d\theta_{ij}}{\theta_{i,j}} \tag{2.3}$$

where E_i, E_j are the energies of the outgoing particles, and θ_{ij} is the angle between them. This equation is not infrared and collinear safe, since when one of the two outgoing particles is soft, or when $\theta_{i,j} \to 0$, the equation is divergent. In order to remove those divergent effects, the k_t algorithm assumes that those particles are just one single candidate and they are recombined together[1].

The algorithm is considered with the following sequential steps:

1. The particles in the event are taken as initial list of objects, and for each pair of particles the distance between them is computed as follows:

$$y_{ij} = \frac{2\min(E_i^2, E_j^2)(1-\cos\theta_{ij})}{Q^2} \tag{2.4}$$

   where Q^2 is the total energy of the event. The pair of particles giving the minimum value for y_{ij} are first considered.

2. A parameter y_{cut} is defined and compared to y_{ij}. If $y_{ij} < y_{cut}$ the two considered particles are combined together and the algorithm is repeated. If $y_{ij} > y_{cut}$ the algorithm declares all remaining particles as a jet.

The term $\min(E_i^2, E_j^2)$ in the equation 2.4 ensures that in case that one of the particles is soft, the distance value is also small and they are preferable clustered together. The latter statement is valid only if $\theta_{i,j}$ is small, otherwise they cannot be clustered together due to the multiplicative factor $(1-\cos\theta_{i,j})$. Hence, each soft particle is most likely clustered with the hardest particle closer to itself.

The initial k_t algorithm had to be modified in order to be used in hadron-hadron collisions [88], since for example, the scale interaction between partons (Q^2) is not a priori known[2]. In the redefinition of equation 2.4, the longitudinal boosted invariant quantities: ΔR ($\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) and p_T are instead used. In this case, the distance parameter is written as follows:

$$d_{ij} = \frac{2\min(p_{T,i}^2, p_{T,j}^2)\Delta R_{ij}^2}{R^2} \tag{2.5}$$

Additionally, another magnitude is considered for each constituent, known as beam distance parameter:

$$d_{iB} = p_{T,i}^2 \tag{2.6}$$

[1]This is in principle valid for all the jet clustering algorithm which faces such divergences.

[2]In principle the Q^2 depends on the longitudinal momentum distribution inside the protons among all the partons, the so-called PDF.

where in both previous equations, the R parameter plays the role of the previous y_{cut} threshold, and is known as jet radius.

The modified algorithm has now the following sequential steps:

1. For all possible pair of particles, the parameter d_{ij} is estimated and the pair with the smallest value is the starting point of the algorithm
2. This quantity is compared to d_{iB}
   - ✓ If $d_{ij} < d_{iB}$, i and j constituents are combined into a single constituent which will enter in the clustering algorithm as a new constituent, while the two individual particles are removed from the list of particles.
   - ✓ If $d_{ij} > d_{iB}$, i is a jet and it is removed from the list of particles.

The distance in ϕ-η between the pair of particles i, j is called ΔR_{ij}. Jets are defined by the radius R in ϕ-η plane, meaning that particles with $\Delta R_{ij} > R$ are never clustered together, independently of their p_T.

With this algorithm, two soft particles could in principle become a jet. Therefore, a lower cut-off for jet definition is needed (usually taken as 10 GeV). Additionally, since the algorithm has as starting point the smallest defined distance d_{ij} (soft particles), it tends to have as output irregular shape jets, sensitive to soft radiation. The geometric picture of the jets produced by this algorithm is illustrated in Figure 2.5 (upper left).

2.1.3.2 The Cambridge/Aachen (CA) jet clustering algorithm

The Cambridge-Aachen jet clustering algorithm [89] has the same fundamental idea as the k_t algorithm, but it uses a new definition for the distance parameter, given by the following formula:

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \tag{2.7}$$

Additionally, the beam distance parameter, whose values are compared to d_{ij}, is equal to unity ($d_{iB} = 1$). Under these new considerations, the same procedure than for the k_t algorithm is applied.

The main difference for this algorithm with respect to the k_t algorithm is that the procedure does not take into account the p_T ordering of the constituents, having the distance parameter d_{ij} as a pure geometric interpretation. Therefore, the jet definition is less affected by soft radiation compared to the sensitivity for the k_t algorithm. Furthermore, the jets defined by this algorithm preserve the angular ordering of QCD emissions given by the parton showers[1].

Figure 2.5 (upper, right) illustrates the geometric picture of this algorithm, which is similar to the one obtained by the k_t algorithm (irregular shapes).

[1] in the case of jet clustering in MC.

2.1.3.3 The anti-k_t jet clustering algorithm

The anti-k_t algorithm [90] is able to deal with the problem of soft-wide radiation from QCD parton shower. The fundamental idea is still the same as the k_t algorithm, but it is reformulated such that, instead of starting from the softer combinatorial pair of candidates, the hardest one is first considered. The new distance parameter d_{ij} is redefined as:

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{2.8}$$

while the threshold parameter is defined as $d = \frac{1}{p_{T,i}^2}$.

With this algorithm, softer particles are later clustered to an already defined hard jet, having less impact on the axis and momentum of the jet.

The resilience to soft emission leads to circular shapes in the ϕ-η plane, therefore this algorithm is considered as a good choice to experimentally access jet physics. Figure 2.5 (lower plot) shows the schematic picture of clustered jets with the anti-k_t algorithm.

2.1.3.4 Summarizing sequential jet clustering algorithm

After having discussed the main jet clustering algorithms, the following aspects can be summarized as common features:

- ✓ two particles cannot be clustered together if the distance ΔR (distance in the ϕ-η plane) between them is larger than a certain parameter R, defined as jet radius.
- ✓ jets are defined as infrared and collinear safe objects.

The three algorithms (k_t, CA and anti-k_t) have the same procedure, but considering in different ways the distance parameters d_{ij} and d_{iB}. A general formula can be written, valid for the three of them:

$$d_{ij} = \min\left(p_{T_i}^{2p}, p_{T_j}^{2p}\right)\frac{\Delta R_{ij}}{R}\ , d_{iB} = p_{Ti}^{2p} \tag{2.9}$$

where the parameter p takes the following values: $p = 2$ (for the k_t algorithm), $p = 0$ (for the CA algorithm) and $p = -2$ (for the anti-k_t algorithm).

The main difference between algorithms is how they take into account soft emissions. In the case of the k_t algorithm, since soft emissions are clustered as first candidates, jets are most sensitive to them, while for the anti-k_t algorithm soft candidates are considered at the end, being resilient to the soft constituents.

The selection of the exact value for the jet radius R depends on the type of physics which is being studied. Small radius jets are more sensitive to perturbative showering (gluon emissions) effects, since quite often they fall outside the jet cone size. If the jet radius is increased, the jet becomes less sensitive to these soft-wide gluon emissions, but paying the price of being more affected by underlying events and pileup effects. Therefore, the selection of R is a compromise of including as much as possible the physics of the process of interest, while adding sources of contamination which are not related to the main process.

2.2 Boosted jets and jet substructure

In section 2.1, the differences between quark-jets and gluon-jets were discussed. Among quark-jets two main groups can be further distinguished: light-quark jets, and heavy-quark jets. The former ones include the jets originating from the lighter quarks (i.e u, d and s), while for the latter ones those originating from heavier quarks (i.e c, b and t) are considered. The so-called heavy jets not only include the heavy-quark jets but also jets originating from massive bosons like W and Higgs. On the other hand, when light jets are referred, they include the light-quark jets and the gluon jets.

Another label further used in this thesis are QCD jets. Herein, QCD jets are referred to light jets, c-jets, and b-jets, while signal jets are the so-called top jets (jets originating from a t-quark). This definition is only meant for distinguishing signal and background jets.

The structure of QCD jets is, in some aspects, different from the structure of jets that are originating from a heavier object. The decay products of heavy objects, i.e t-quarks, under certain kinematics conditions, can be clustered as a single jet. Hence, we can define a *heavy jet* as a jet containing all the decay products of a certain heavy object. In the high energy regime, currently achieved by the LHC, those scenarios are often appearing (e.g. [92], [93]) .

This section is devoted to explain the basic tools needed to study boosted heavy jets. First, the kinematics requirements for boosted topologies are presented in subsection 2.2.1. In the following subsection (2.2.2), differences between QCD jets and top-jets concerning their mass distribution are discussed.

Generally, boosted jets are reconstructed by using large cone size jet (i.e with radius $R = 0.8$). As a consequence of the choice of large size radius for the jet reconstruction, additional sources of contamination affect the jet. Some jet observable are highly sensitive to those extra contamination sources. In order to remove unwanted contamination (usually appearing as soft-wide radiation), *cleaning* techniques are applied. They are known under the name of *grooming techniques* and are further discussed in the subsection 2.2.3.

For heavy boosted jets, multi-hard objects can be identified inside the jet (i.e for t-jets, three hard objects play a role: b-quark, and the two hadronic decay products of the W-boson). This aspect is known as multi-prong configuration of the jet. There are defined observables (N-subjettiness) sensitive to the multi-prong configuration, and they are, up to now, considered as one of the most powerful tools to distinguish boosted objects. They will be described in section 2.2.4.

2.2.1 Kinematics considerations for boosted topologies

In order to reach the boosted phase space region, some kinematics conditions need to be satisfied. We can start to infer, under which conditions those scenarios will occur, by writing the mass of the original object (parton or boson) as a function of the transverse momentum of the decay products, as follows [83]:

$$m^2 \approx 2p_1p_2 = p_{T,1}p_{T,2}\Delta R_{12}^2 \quad , \tag{2.10}$$

where $p_{T,1}$ and $p_{T,2}$ refers to the transverse momentum for each of the decay products and ΔR_{12} is the azimuthal-rapidity distance between them ($\Delta R = \sqrt{(\phi_1-\phi_2)^2+(\eta_1-\eta_2)^2}$). This magnitude can be redefined as a function of the transverse momentum of the decaying heavy object (p_T) as [83]:

$$m^2 \approx z(1-z)p_T^2\Delta R_{12}^2 \quad , \tag{2.11}$$

where z is the energy fraction of the heavy object carried by one of the decay products. The previous equations are valid only if the $p_{T,i}$ ($i = 1, 2$) are large enough, such that extra emissions can be neglected. If one assumes that $z \approx (1 - z)$ (the initial energy is equally distributed between the decay products), then the distance ΔR_{12} can be written as follows:

$$\Delta R_{12} \approx \frac{2 \cdot m}{p_T}. \tag{2.12}$$

If the p_T of the original object satisfies the condition $p_T > 2m/R_{jet}$, where R_{jet} is the radius of the clustered jet, then the two decay products can be clustered within the jet with radius R_{jet} (since $\Delta R_{12} < R_{jet}$).

Therefore, the kinematic condition to ensure that the decay products of a heavy object are inside a jet (boosted regime) can be directly obtained from 2.12, by requiring $p_T > 2 \cdot m/R$. This regime is achieved for t-quarks when $p_T \sim 400$ GeV (considering a jet radius of $R = 0.8$).

The higher the p_T of the original heavy object is, the more boosted configuration its decay products will have. With this statement, one can naively think, when increasing the p_T of the heavy objects that it is possible to consider a smaller radius parameters in the jet algorithm. For example, t-quarks with $p_T \sim 1.5$ TeV have their decay products within a cone with radius $R = 0.4$. Nevertheless, the replacement of large-cone size jet radius by small-cone size radius at the higher boosted regimes is not recommended since for small radius clustering algorithms, the jet becomes sensitive to soft emissions from the hard decay products. Therefore, it is better to define a large cone size jet radius, and use the jet substructure techniques to identify the components, rather than using smaller radius jets.

2.2.2 Jet mass as possible discriminating variable: QCD-jets vs top-jets

The jet mass is one of the potential observables to be used for distinguishing top-jets from QCD jets. In the case of top jets, the jet mass distribution is expected to be centered around the mass of top-quark ($m_j \sim 172.5$ GeV). Due to extra emissions and combinatorial effects, deviations around the central value are expected.

Starting from the boosted requirement (equation 2.12), the jet mass for top-jets can be expressed by the following formula, as function of the transverse momentum of the jet (p_T^j) and the distance between the W and b decay products ($R_{Wb} = \Delta R_{12}$):

$$m_j^{top} = \frac{1}{2} p_T^j R_{Wb} \tag{2.13}$$

In the case of QCD jets, the mass originates from gluon emissions during the parton cascade. The jet mass (m_j^{qcd}) depends on the transverse momentum of the jet and the radius as follows: $m_j^2 \propto \alpha_s p_T^2 R_j^2$. In principle, there is nothing that forbids a QCD jet having a mass value in the top-jet mass region.

Figure 2.6 shows the difference between the jet mass distributions between top jets and QCD jets. Both, normalized and absolute distributions are shown. The distributions correspond to signal and background simulated events. Signal events were simulated with POWHEG+PYTHIA8 MC, while QCD multijet events with MADGRAPH+PYTHIA8 MC. The distributions consider the leading jet in each event[1]. Events with at least two jets reclustered with the anti-k_t algorithm ($R = 0.8$), $p_T > 400$ GeV, $|\eta| < 2.4$ and with soft drop mass larger than 50 GeV are considered[2].

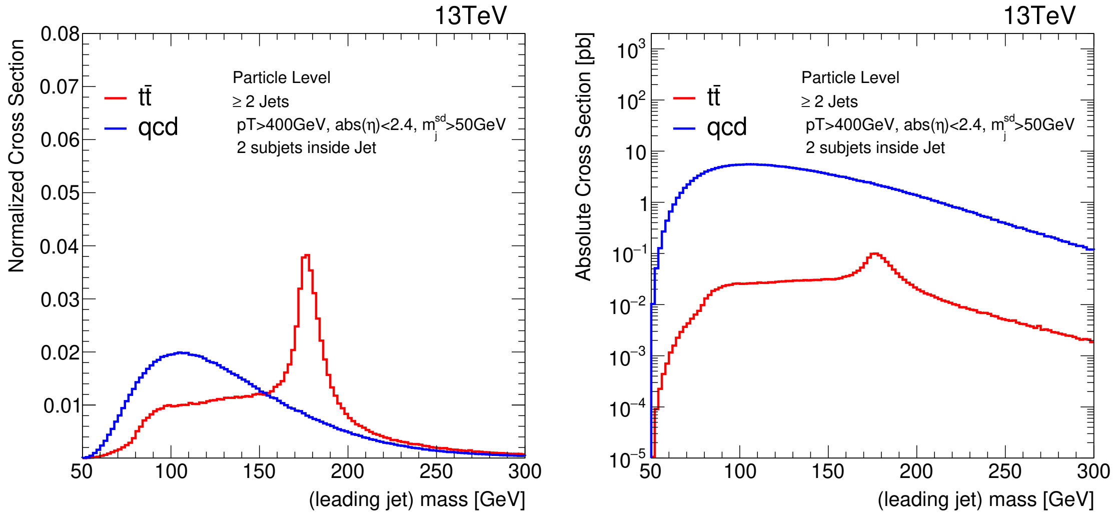


Figure 2.6: Jet mass distribution for the leading jet in QCD and $t\bar{t}$ simulated events: (left) normalized distributions (right) absolute cross section. Jets are reconstructed using anti-k_t with $R = 0.8$.

In figure 2.6 it can be seen that the top-jet mass is peaked at the expected value of approximately 172.5 GeV. The low mass region contribution, however, comes mainly from events where the leading jet is not properly reconstructing the t-quarks. Those events will be further discussed in section 3.2 (next chapter), when the definition of top-jets is discussed. The probability that a jet mass vanishes, for both, QCD-jets and top-jets is zero. On the other hand, the QCD jet mass distribution has no specific peak.

When comparing the absolute distributions it is observed that in the whole jet mass phase space the contribution from QCD-jets is approximately two orders of magnitudes larger than for top-jets. Considering the region around the top mass (150 GeV$< m_j <$ 200 GeV) the background over signal ratio becomes $\sim$ 60, meaning that the jet-mass is not enough for properly distinguishing top-jets.

2.2.3 Grooming techniques

Jets are collimated bunches of hadrons originated by at least one hard object (either quarks, gluons, or bosons) and their subsequent soft radiations (final state radiation). In hadron-hadron collisions, as a consequence of other phenomena occurring at the same time, jets can be contaminated by other constituents not specifically related to the hard process under consideration. Those additional sources could be:

- ✓ initial state radiation: before the hard process the partons produce soft radiation
- ✓ underlying events: secondary parton collisions in the same hadron-hadron collision are produced

[1] In $t\bar{t}$ events the two leading jets are considered as top-jets candidates, since they are the ones most likely associated to the t-quarks. This aspect will be further discussed in next chapter.

[2] Soft drop mass is referred to the jet mass after applying the Mass Soft Drop grooming technique (see section 2.2.3).

✓ pileup: several hadron-hadron collision occurring at the same time.

Those secondary phenomena have effects on jet observables. One of the most sensitive observable to those effects is the jet mass, having a shift in its distribution from the expected central value. If those effects would be exactly the same for each event, as result, the overall shift could be corrected by the calibration of the jets. But due to stochastic behavior of those phenomena, the effects vary in each event, and therefore an additional degradation on the mass resolution is expected. Grooming techniques aim to minimize the contamination of the jets. These sources are most likely soft contributions distributed uniformly inside the jet, therefore those grooming techniques are focused on removing mainly soft-wide constituents.

There are several grooming techniques currently available. The ones most used in LHC physics (jet trimming, jet pruning, jet filtering, and soft drop mass) are further discussed. Those grooming techniques are based on event by event corrections, actively removing the contaminating components.

2.2.3.1 Jet Trimming

Jet trimming [94] is a grooming technique which consists basically in the following steps:

1. a jet is clustered with a specific algorithm (i.e CA, anti-k_t, or k_t) and radius parameter R (large R values are normally considered),
2. a jet finder algorithm defines inside the fat jet a set of subjets with cone width R_{cut} ($R_{sub} < R$),
3. all subjets satisfying the condition: $p_t^{sub} > z_{cut} p_t$ are kept, while the ones below the threshold are removed. Here p_t^{sub} refers to the transverse momentum of the subjet, while p_t is the transverse momentum of the original jet and z_{cut} is a free parameter to be defined,
4. the trimmed jet is the sum of the kept subjets.

This algorithm has two free parameters: R_{sub} and z_{cut}, that must be chosen carefully to optimize the amount of constituents to be removed. Usually the R_{sub} parameter takes values in the interval 0.2-0.35, while $z_{cut} \sim 0.01$.

The trimming technique could be applied to a jet, which is clustered using any infrared-collinear safe algorithm. However, for this specific grooming technique, jets clustered with the k_t clustering algorithm are favorable. The k_t algorithm clusters soft radiation first and therefore the jet is sensitive to soft candidates. Soft-collinear constituents most likely coming from FSR are kept since they are supposed to be within a cone of radius R_{cut}. On the other hand, soft-wide radiation coming from the ISR, is likely removed.

Figure 2.7 shows a schematic picture of the jet trimming technique. In this specific case, only three subjets are passing the threshold z_{cut} (in picture meant as f_{cut} parameter).

2.2.3.2 Jet Pruning

The jet pruning algorithm is built according the following steps:

✓ first, the constituents of a jet, which was clustered with any jet clustering algorithm, are reclustered using the CA or k_t jet algorithms.

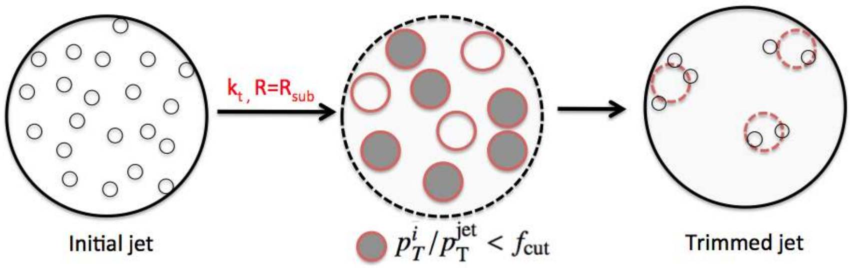


Figure 2.7: (taken from [95]) Jet trimming as grooming technique: as first step, a jet is clustered with k_t algorithm (taking as radius parameter R); as a second step, a set of subjets are defined with radius parameter R_{sub}, and only those ones with $p_{T,i} > f_{cut} p_T$ are kept to finally define the trimmed jet.

- ✓ from the jet with mass m, a parameter cut is defined: $R_{prune} = f_p 2m/p_T$, where f_p is an adjustable parameter and p_T is the transverse momentum of the jet.
- ✓ from the jet clustering algorithm, a step backward procedure is performed, such that for each splitting possible $P \to i, j$, the following conditions are checked:

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,(i+j)}} > z_{cut} \tag{2.14}$$

$$\Delta R_{ij} < R_{prune} \tag{2.15}$$

  where now $p_{T,i}$ and $p_{T,j}$ are the transverse momenta of the considered pair of constituents, and $p_{T,(i+j)}$ is the transverse momentum of the system i, j.
- ✓ if at least one of the previous conditions is satisfied, the constituents i and j are combined. If none of the conditions are satisfied, the softer of the constituents is removed, and the algorithm is repeated with the remaining constituents.
- ✓ the algorithm stops when the reclustered objects are hard enough to be considered as subjets: for any combinatorial pair of constituents the previous conditions are satisfied.

The free parameters f_p and z_{cut} must be optimized for each specific case, but values of 0.5 and 0.1 respectively, are often used. Ideally f_p is selected such that different hard prongs are assigned to different subjets. Figure 2.8 illustrates a schematic picture of the pruning technique.

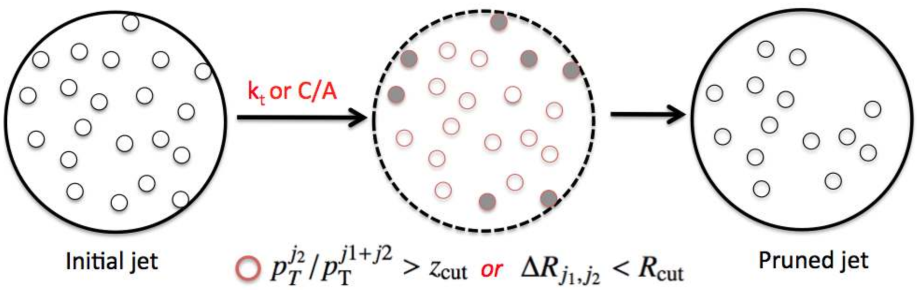


Figure 2.8: (taken from [95]) Jet pruning as grooming technique: as a first step, a jet is re-clustered with k_t(or CA) algorithm; as second step, for each pair of constituents two conditions are checked, for each pair not satisfying any of the conditions, the softer of them is removed from the set of constituents. Finally the remaining constituents define a pruned jet, where all the constituents are hard enough to be considered itself as subjets. R_{cut} in the figure has the same meaning that R_{prune} in the text.

2.2.3.3 Soft Drop

The Soft Drop mechanism [85] is one of the grooming techniques most used in heavy objects tagging. This algorithm is not only useful to remove the unwanted contaminating radiation in the jet, but it is also useful for defining subjets as observables, which can be further used to distinguish signal jets (i.e top jets) from background jets.

The Soft Drop algorithm uses jets clustered with the anti-k_t jet algorithm, but then the CA jet clustering algorithm is used to recluster the jet constituents of the initial jet. In that way angular ordering from the emissions are preserved[1], in the new reclustered constituents.

The sequential steps of the algorithm are:

1. The clustered jet is divided into two subjets j_1 and j_2 by undoing a step backward in the CA clustering algorithm
2. Considering m_{j1} as the mass of the heaviest declustered subjet, the following condition is checked:
$$\frac{m_{j1}}{m_{jet}} < \mu \tag{2.16}$$
where μ is one of the free parameters of the algorithm.
3. Additionally, the splitting needs to be symmetric, requiring that:
$$\frac{\min[p^2_{T,j1}, p^2_{T,j2}]\Delta R^2_{j1j2}}{m^2_{jet}} > z_{cut} \tag{2.17}$$
where now, $\Delta R_{j_1 j_2}$ is the usual R distance (opening angle) between the subjets j_1 and j_2, and z_{cut} is the additional free parameter, which reflects the energy sharing between the subjets j_1 and j_2 with respect the original jet.
4. If both previous conditions are satisfied, then a soft drop jet is defined by the combination of the two subjets.
5. Otherwise the softer subjet (j_2) is removed and the algorithm is repeated taking as initial clustered jet j_1.

This procedure is named as Mass Drop Tagger (MDT) [96]. By this grooming technique, hard substructures inside the fat jet are found, even if soft emissions influence the total jet mass.

In order to illustrate the Mass Drop Tagger, one can look to the case represented in Figure 2.9(a). The parton p_1 emits partons p_2 and p_3, such that the angular distance between all the partons satisfies $\theta_{13} \ll \theta_{12}$. This condition implicitly means that, the mass of the jet is governed by the p_3 emission ($m_{jet} \gg m_{12}$). The CA jet finder algorithm will most likely cluster p_1 and p_2 in one single subjet, while p_3 will be the second subjet. The condition expressed by equation 2.16, given the angular distribution, is always satisfied for the scenario represented in the figure. If the asymmetry condition, given by equation 2.17, is satisfied, then the whole jet is tagged by the soft drop algorithm. On the other hand, if the asymmetric condition is not satisfied, the MDT will keep the jet containing p_1 and p_2 since it is the one of the subjets with the harder emissions. Therefore the hard emissions are always kept in the tagged jet.

[1] In the case of jets from MC simulation the angular ordering of the Parton Shower is preserved.

The MDT, however, is not robust enough for the cases illustrated in Figure 2.9(b), where now one soft gluon emits partons p_2 and p_3. In this case, due to the angular ordering ($\theta_{23} \ll \theta_{12} \simeq \theta_{13}$), the parton p_1 will be contained in one subjet and p_2 together with p_3 will form the other subjet. It could happen, that the most massive subjet will be the one formed by the soft emission, and p_1 will be removed by the MDT. This is an unwanted feature of the tagger.

In order to avoid this unwanted behavior, the modified Mass Soft Drop tagger was developed [83]. The main difference is that, if the asymmetry condition of equation 2.17 is not satisfied, the subjet with larger transverse mass ($m_T = \sqrt{m^2 + p_T^2}$) is defined as the jet to be kept, and one proceeds from the beginning of the algorithm after removing the one with lower transverse mass. Herein, when the Mass Soft Drop Tagger is referred, the modified version is meant. With this version, for instance, if one reconsider the scenario represented in Figure 2.9(b), the p_1 subjet is the one most likely having the larger transverse mass, and the p_2 and p_3 constituents will be removed.

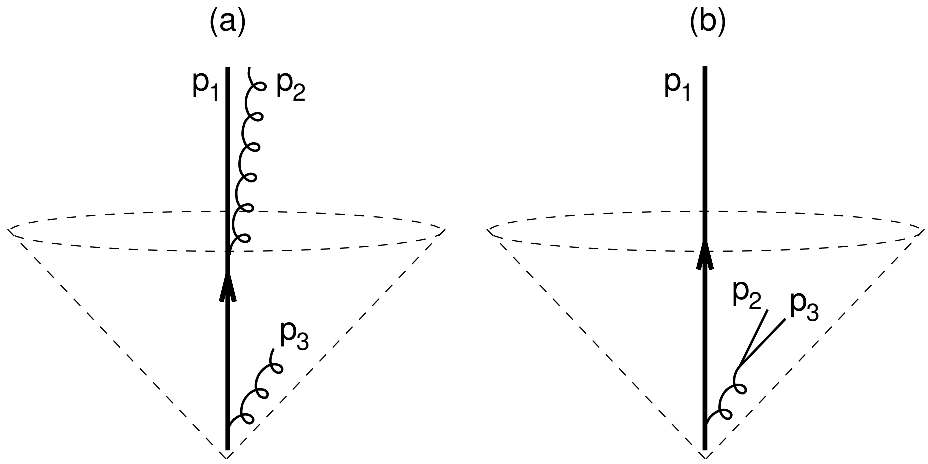


Figure 2.9: (taken from [97]) (a) a jet defined by a hard parton p_1 emitting two gluons p_2 and p_3 (b) a jet defined as a hard parton p_1 emitting a gluon, the gluon emits later two partons p_2, p_3.

2.2.3.4 Jet Filtering

The jet filtering grooming technique consists of finding the three hardest possible subjets and removing all the remaining soft constituents of the jet. Usually, it is applied after the Soft Drop Mass mechanism, where the two subjets are then reclustered. The standard algorithm for clustering the soft drop subjets is the CA algorithm with radius parameter:

$$R_{filter} = \min(0.3, \Delta R_{j1,j2}/2) \tag{2.18}$$

where $\Delta R_{j1,j2}$ is the spatial distance between the two subjets j_1,j_2b. With this parameter, it is guaranteed that $R_{filter} < R_{j1,j2}$. Then the three hardest subjets are resolved. The algorithm is represented schematically in Figure 2.6

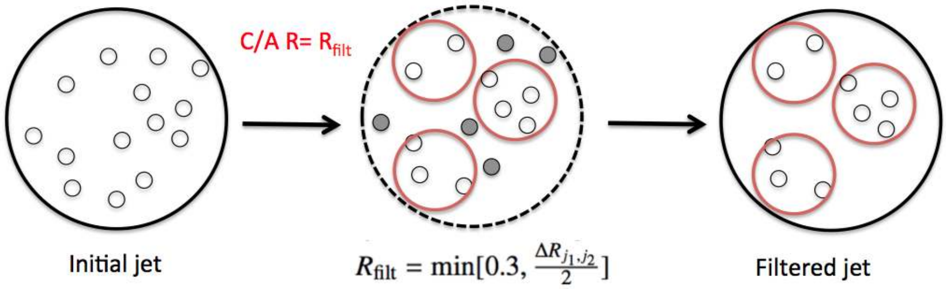


Figure 2.10: (taken from [95]) Jet filtering grooming technique: as first step, the jet obtained by reclustering the soft drop subjets with CA algorithm is formed; as second step, three hard subjets are defined with radius R_{filt}; and as last step all the remaining constituents which are not inside the three subjets are removed and the filtered jet is the combination of the three hard subjets.

2.2.4 Jet substructure observable definitions

One of the first developed jet substructure technique is known under the name of BDRS [1] algorithm [96]. The main idea behind the BDRS algorithm is to distinguish QCD jets from heavy jets using the energy sharing. In the case of QCD jets, the probability of a quark to emit a gluon is given by the DGLAP[22] splitting function as follows:

$$P(z) = C_F \frac{1+z^2}{1-z} \tag{2.19}$$

where C_F is the quark colour factor, and $1-z$ represents the fraction of momentum of the original quark carried by the gluon after the splitting. This expression gives singularities when $z \to 1$, or equivalent to say, the soft emission limit is enhanced. Therefore, a cut on the z variable would remove a big amount of the QCD jets. On the other hand, for boosted heavy jets, the sharing of the energy between the decay products has an approximately flat probability ($P(z) \propto 1$). Therefore, a cut on the z variable, would remove relative small amount of heavy jets. Hence, the energy sharing z variable can be used as discriminating observable to distinguish QCD and heavy jets, and this is considered in the BDRS substructure technique.

However, when extra emissions occur, the energy sharing criteria is not the most efficient way for distinguishing top-jets from QCD jets. Additional jet subtraction techniques have been developed dealing with the multi-prong configuration of heavy jets.

Heavy jets are formed mainly from the decay products of the heavy objects, and those hard substructures are not affected by soft-wide radiations. Therefore a multi-prong configuration can be observed. QCD jets, on the other hand, since the soft emissions are widely distributed, cannot be identified with any multi-prong configuration.

Observables to deal with the multi-prong configuration are the N-subjettiness variables [98]. Those variables test the hypothesis of being able to find inside jets a given number of subjets (jet axes), testing the multi-prong configuration.

[1] BDRS stands for the names of the author of paper which contains the algorithm: Jonathan M. **B**utterworth, Adam R. **D**avison, Mathieu **R**ubin, Gavin P. **S**alam.

The subjettiness can be defined by the following formula:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min[\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}] \tag{2.20}$$

where k refers to all particle constituents of the jet, $p_{T,k}$ is the transverse momentum of the constituent k, and $\Delta R_{j,k}$ with $j = 1, 2...N$, is the distance in the $\phi - \eta$ plane between each subjet axes and the constituent k. The factor d_0 is a normalization factor such that τ_N takes values between 0 and 1:

$$d_0 = \sum_k p_{T,k} R_0 \tag{2.21}$$

where R_0 is the jet radius of the original jet.

In order to understand the meaning of these variables, one can focus on the case of top jets. After the t-quark decay, there are three hard objects ($t \to bW \to bq\bar{q}$). For each jet constituent, hence, there will be a jet axis defined in the direction of the hard decay components, meaning that the variable τ_3, would take small values (since for all k components it is possible to find small distances $\Delta_{j,k}$, $j = 1, 2, 3$). If now one consider instead τ_2 (testing the hypothesis that only two axes are found), the constituents near by the third hard component of the decay products, would have large $\Delta_{j,k}$ values, meaning that τ_2 would have larger values.

Summarizing the main idea, lower values of τ_N ($\tau_N \to 0$) indicates that all the radiation and components are aligned with at least one of the subjet candidates, and this jet is most likely identified with a N-prong jet configuration. If $\tau_N \to 1$, it could indicate that a big fraction of the jet constituents are far away of the defined axes and probably the jet is identified with at least $(N + 1)$ subjets. Alternatively, the ratio between them could be also used as discriminating variables (i.e τ_3/τ_2).

In Chapter 3, details are provided on how the jet substructure techniques (i.e the Soft Drop tagger and τ_N variables) are considered in the boosted top jet topologies, studied in this thesis.

Chapter 3

Hadronic Top jet definition

This chapter focuses on presenting phenomenological studies on boosted top jet scenarios. The aim of this thesis is to measure the differential cross section of the boosted $t\bar{t}$ pair production with respect to the transverse momentum (p_T) of the top-jets, and with respect to the azimuthal separation ($\Delta\phi$) between the two top-jets. Therefore, a crucial point is to identify events where two jets can be certainly associated to the $t\bar{t}$ pair. This can be achieved if we define certain criteria, which allow us to classify a jet as a top-jet.

The chapter is divided into two main blocks. First, the section 3.1 presents the way in which the substructure techniques (discussed in Chapter 2) are considered in top boosted topologies. In the second part of the chapter (section 3.2), the hadronic top jet definition is then provided, using the jet substructure techniques, i.e the Mass Drop Tagger (MDT).

3.1 Jet substructure implemented for the top-jets

In Chapter 2, the jet substructure techniques used in the identification of heavy jets were discussed. In this section, they will be applied directly to the specific case of boosted top jets.

First, hadrons are clustered into jets using the anti-k_t algorithm [90] with a large cone size radius ($R = 0.8$). As grooming technique, in order to reduce soft and wide-angle contaminating radiation, the modified version of the Soft Drop tagger [85] is applied. Before applying this grooming technique, the constituents of the jet are reclustered with the Cambridge-Aachen (CA) algorithm [89], such that the new jet is sensitive to the softer constituents. The free parameters in the Soft Drop algorithm were taken as: $\beta = 0$ and $z_{cut} = 0.1$.

As result of this algorithm, not only the soft radiation is reduced, but also two new observables are defined: the two soft-drop subjets. The *first subjet*, herein is referred to the one with higher p_T value from the two soft-drop subjets.

In the case of top-jets, a 3-prong topology is expected, and therefore in order to study the jet substructure, the 3-subjettiness, 2-subjettiness, 1-subjettiness observables and the ratios between them are used. The filtering algorithm is applied to the soft-drop jet, trying to find those three jet axis given by the hard constituents coming from the t-quark decay products.

In figure 3.1, the steps for studying the jet substructure are schematically represented. Herein, *jets* (fat jets) are objects clustered using the anti-k_t algorithm with radius parameter $R = 0.8$, while subjets are the two subjets obtained from the Soft Drop algorithm.

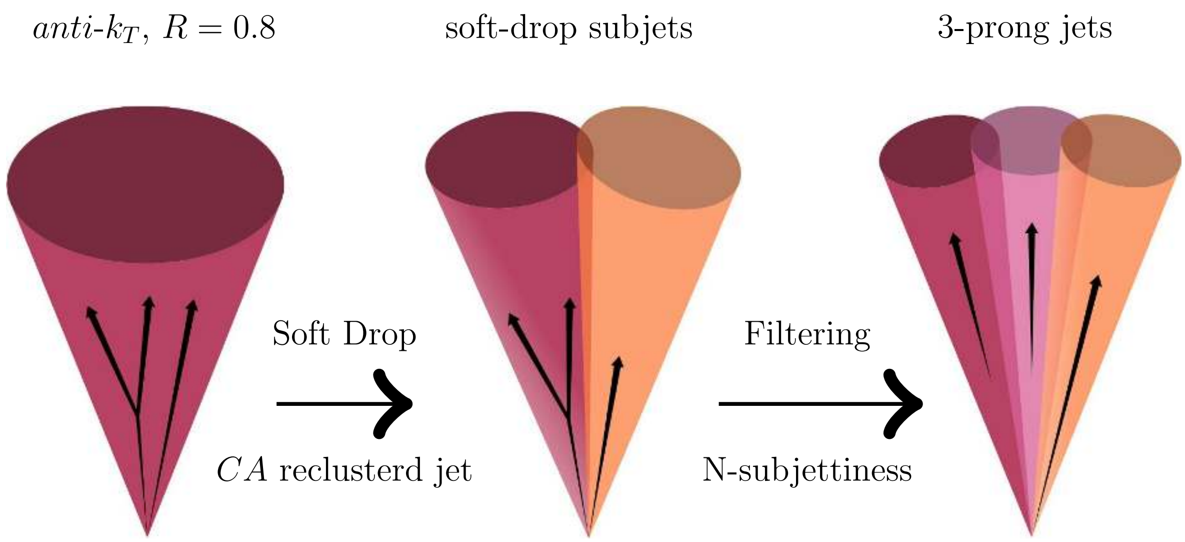


Figure 3.1: Schematic picture showing how the jet clustering and jet substructure techniques are used in this thesis (top jet tagging in the boosted regime).

After considering the Soft Drop algorithm and the filtering algorithm, observables like subjet mass, and N-subjettiness variables can be defined. In the following, those observables are presented, since they are further used for the top-jet hadronic definition.

Simulated $t\bar{t}$ and QCD multijet events are used to study hadronic top jets. Events are selected requiring at least two jets with $p_T > 400$ GeV, $|\eta| < 2.4$ and with soft drop jet mass larger than 50 GeV. Those selection criteria correspond to the baseline selection used in this thesis, and is named as the boosted dijet selection.

Figure 3.2 shows the jet mass and the soft drop jet mass distributions of the leading jet in simulated events ($t\bar{t}$ and QCD). In this figure the effect of the Soft Drop algorithm is shown. For QCD jets, as might be noticed, the jet mass distribution after the Soft Drop algorithm (soft drop jet mass) is shifted to lower values. In the case of top-jets, the peak around the top mass window appears more visible (better resolution around the t-quark mass). Additionally, for the latter ones, a peak around the W-boson mass appears, indicating the cases where the leading jet is not including all the t-quark decay products (i.e the b-hadron is not inside the jet, or it is soft and its contribution is removed by the Soft Drop algorithm). Those scenarios will be further discussed in this chapter.

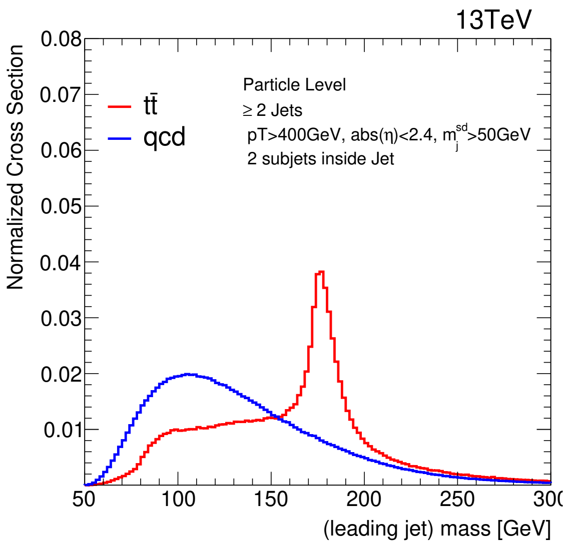


Figure 3.2: Jet mass distribution for the leading jet in QCD ant $t\bar{t}$ simulated events (left) the pure jet mass (right) the soft drop jet mass. Distributions correspond to the particle level after applying the boosted selection.

Figure 3.3 shows the distributions of the soft drop mass of the subjets in the leading jet. The distribution of the mass of the first subjet (left plot), for top-jets, has a peak at the W-boson mass. This contribution corresponds to $\sim$ 50% of the $t\bar{t}$ events, in which the first subjet includes all the subsequent decay products of the W boson. The distribution of the second subjet, on the other hand, does not show any specific shape, and furthermore, both contributions, QCD multijet and top-jets have similar behavior.

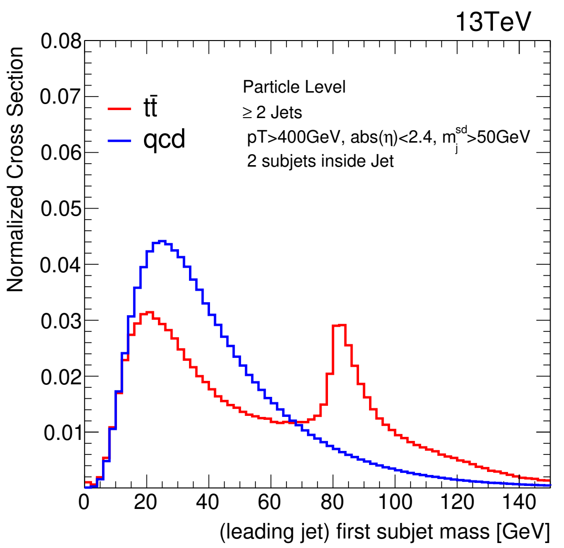


Figure 3.3: Jet mass distribution for the subjets in QCD ant $t\bar{t}$ simulated events (left) first subjet (right) second subjet. Distributions correspond to the particle level after applying the boosted selection.

After applying the filtering technique, up to three subjets are found and the N-subjettiness variables can be obtained. In Figure 3.4, the ratios τ_3/τ_2 and τ_3/τ_1 are shown. These variables are further used as input variables for a multivariate technique implemented to distinguish top-jets. in the analysis strategy.

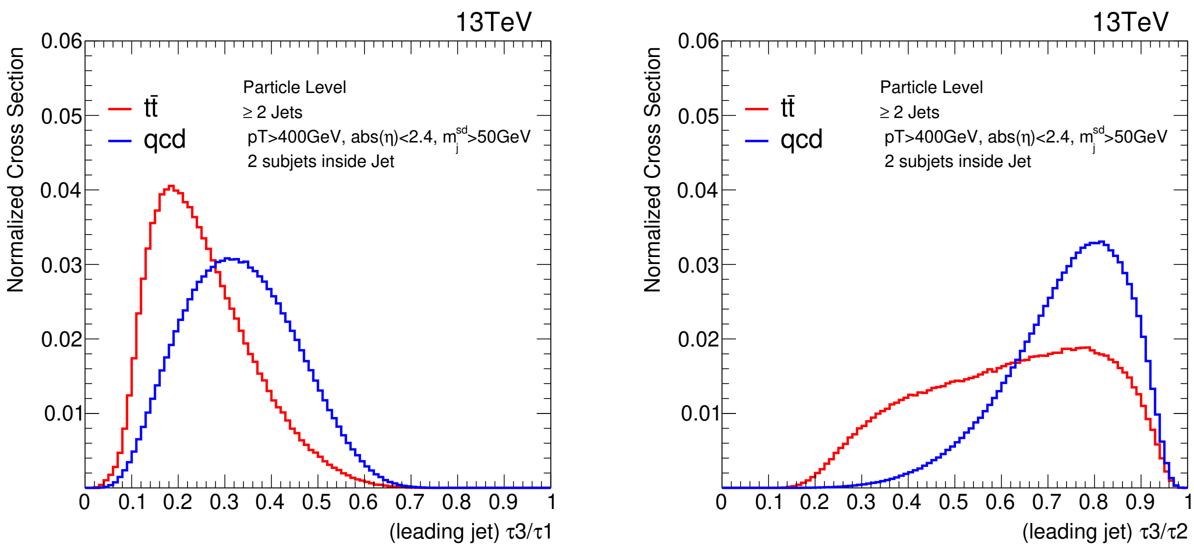


Figure 3.4: N- subjettiness observables in QCD ant $t\bar{t}$ simulated events (left) τ_3/τ_1 (right) τ_3/τ_2. Distributions correspond to the particle level after applying the boosted selection.

Figure 3.5 shows the relation between τ_3/τ_2 and τ_3/τ_1 for QCD jets and top-jets. Different patterns can be recognized distinguishing both kind of jets. Such pattern are taken on advantage for instance, in multivariate techniques for distinguishing signal jets.

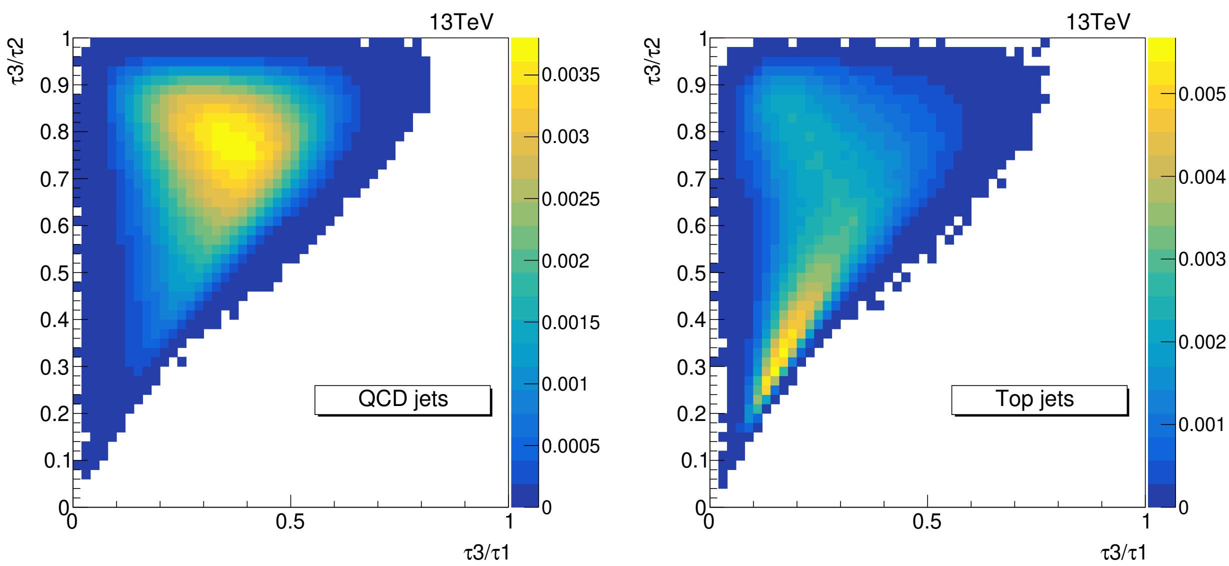


Figure 3.5: Correlation between N-subjettinness variables (τ_3/τ_1 and τ_3/τ_2) in (left) QCD jets (right) top jets. Distributions correspond to the particle level after applying the boosted selection.

3.2 Hadronic top-jet definition

Signal events are those having a $t\bar{t}$ pair with each t-quark produced at high p_T. In such scenarios, both t-quarks will be preferable balancing each other with opposite azimuthal angle ($\Delta\phi \rightarrow \pi$). Therefore, the jets in the event which would most likely be associated to the t-quarks are the leading and subleadig jets.[1].

However, there might be scenarios where one of the top-jets candidates is not related to any of the t-quarks, or not properly containing all the decay products of the t-quark. Those scenarios need to be identified and not considered as signal events.

One of those possible scenarios is shown schematically in Figure 3.6 (left picture). The $t\bar{t}$ system is boosted itself [2] and the leading jet might contain the recoiling QCD radiation, while the subleading jet and the third jet would be the top-jets. Those events have usually a third jet with relative high p_T and the azimuthal separation of the truth top-jets is taking values $\Delta\phi \ll \pi$. An important remark to notice is that in such cases the azimuthal separation between the two top-jets is different from the azimuthal separation between the two leading jets: $\Delta\phi_{t\bar{t}} \neq \Delta\phi_{12}$.

A signal event in which the two leading jets are the top-jets is schematically represented in Figure 3.6 (right picture). Those events have usually the third jet with lower p_T, in comparison to the p_T of the two top-jets candidates, and the azimuthal separation between the top-jets is closer to π.

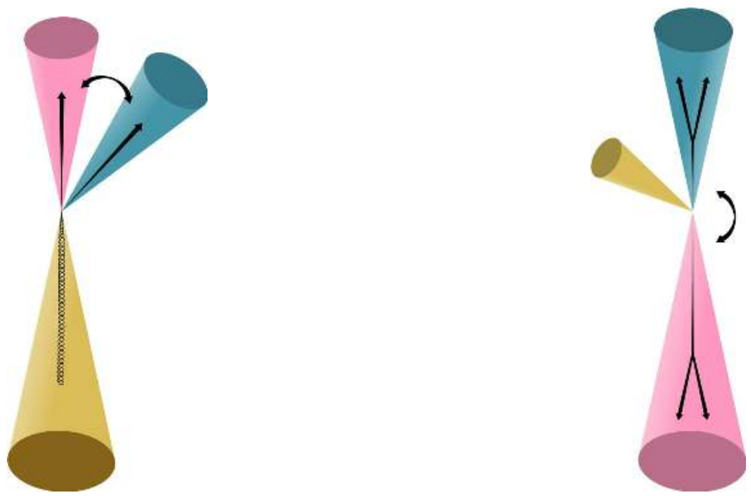


Figure 3.6: Sketch representing $t\bar{t}$ events. The top-jets are shown with blue and pink cones, and the additional jet with a yellow cone. Two scenarios are represented. (Left picture) the $t\bar{t}$ system is boosted itself, and the two top-jets are the leading and third jets, while the leading jet (yellow colour) contains the recoiling QCD radiation of the $t\bar{t}$ system. (Right picture) the two leading jets are the two top jets, while the third jet (yellow colour) contains additional radiation, having low p_T value. The decay products of the t-quarks (right picture) are contained within the top jets.

[1] The leading and subleading jets in events with both jets having high p_T, they are most likely with an azimuthal separation $\Delta\phi \rightarrow \pi$.

[2] Both t-quark flying in similar direction instead of in opposite direction.

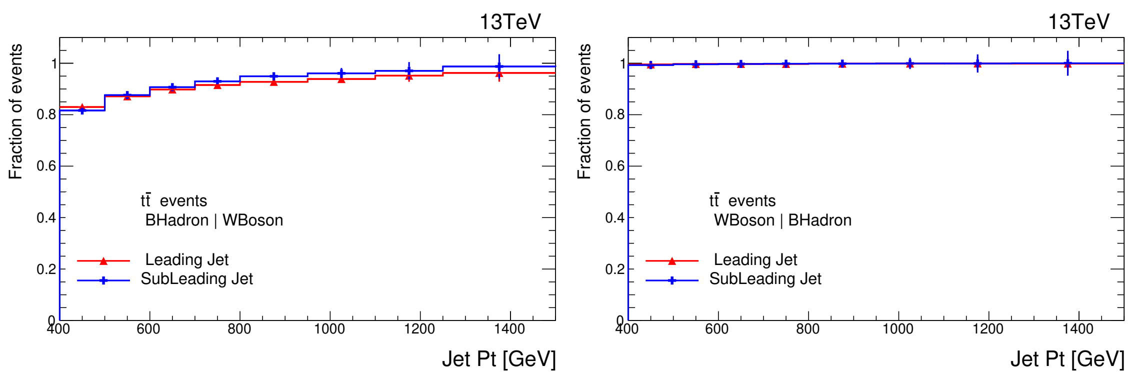
In addition to the requirement that the two leading jets must be associated to the $t\bar{t}$ pair, each of the t-quarks should have high p_T, such that, their decay products are contained within their respective top-jet. Figure 3.7a shows a schematic picture of one of the possibles scenarios where this criterion is not satisfied. The b-quark is not contained within the top-jet.

Since the b-quark is the lighter of the decay products, it is the most likely deviating from the t-quark direction. Figure 3.8b (left plot) shows the probability that either the leading or subleading jets contains a b-quark, given that the W-boson is within the jet. This probability increases to almost 100% for the higher p_T region, while in the lower p_T region goes down to 80%. On the other hand, the probability that the W-boson is within the top-jet, given that the b-quark is within is $\sim$ 100% (right plot) in the whole p_T phase space. Therefore, such scenarios, where the b-quark flies outside the top-jet, are affecting mostly the lower p_T region.

The other scenarios affecting the t-quark reconstruction are the cases where the W-boson is taking almost all the p_T fraction of the t-quark. Figure 3.7b shows the schematic representation of those scenarios. The b-quark would have very small p_T and therefore removed by the Soft Drop grooming algorithm. Furthermore, those events are difficult to experimentally deal with, since one of the selection criteria at detector level is related to the tagging of the b-hadrons[1]. These scenarios are affecting mostly in the higher p_T region.

(a) **(b)**

Figure 3.7: Schematic representation of the jets not properly reconstructing the t-quarks. The top jet candidate is represented, as well as subjets (the inner cones). Two cases are shown (a) non boosted topologies, the b-quark is outside the jet (b) the p_t of the b-quark has small p_t and is removed by the soft Drop Mass.



(a) **(b)**

Figure 3.8: Fraction of events which satisfied the following conditions: (a) given that the W-boson is inside the jet, the b-quark is inside also (b) given the b-quark is inside the jet, the W-boson is inside.

[1] neither the detector nor the b-tagging algorithms are sensitive to very low p_T constituents

An example of a jet which is properly reconstructing the t-quark is shown in Figure 3.9. The condition that the decay products are contained within the jet is satisfied. In the picture, furthermore, the first subjet is supposed to reconstruct the W-boson decay products.

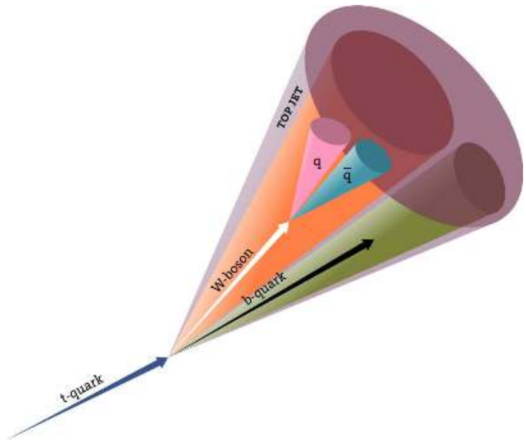


Figure 3.9: Schematic representation of the reconstruction of the t quark into the jet. The top jet candidate is represented, as well as the soft drop subjets (the inner cones). This case represents a fully merged scenario where all the decay products are inside the fat jet, and the hadronic products of the W-boson can be clustered in the first subjet, while the second subjet gathers all the remaining soft radiation and the decay products of the b-quark.

From the discussion on possible misidentification scenarios, where the two leading jets are not properly reconstructing the two t-quarks, the following criteria for top-jets are defined:

- ✓ a jet with $p_T > 400$ GeV and soft drop mass $m_j^{sd} > 50$ GeV
- ✓ the mass of the first subjet is larger than 40 GeV
- ✓ contains a B-hadron

Figure 3.10 shows the soft drop mass distributions, for the leading and subleading jets, in $t\bar{t}$ signal events, when events were selected, first, requiring two jets satisfying the boosted dijet criteria [1], and secondly (represented by the red colour area) when requiring two top-jets, using the hadronic top-jet definition. By applying the top-jet hadronic definition, the soft drop jet mass reconstructs better the peak around the t-quark mass.

The hadron definition is based on studies presented in Appendix C, were the mass of the first subjet was found as best variable to properly identify top jets. The specific value 40 GeV, as the proposed cut criterion, is based on efficiency studies also presented in Appendix C.

Selection efficiency

Figure 3.11 represents the selection efficiency as a function of p_T of the leading and subleading jets, requiring at least two top-jets. In order to estimate the efficiency, simulated $t\bar{t}$ events with POWHEG+PYTHIA8 generator are considered, at particle level. The efficiency can be written as follows:

$$\epsilon(p_{T,i}) = \frac{N(\text{top-jet}, p_{T,i})}{N(\text{dijet}, p_{T,i})} \tag{3.1}$$

[1] jet with $p_T > 400$ GeV and soft drop mass $m_j^{sd} > 50$ GeV.

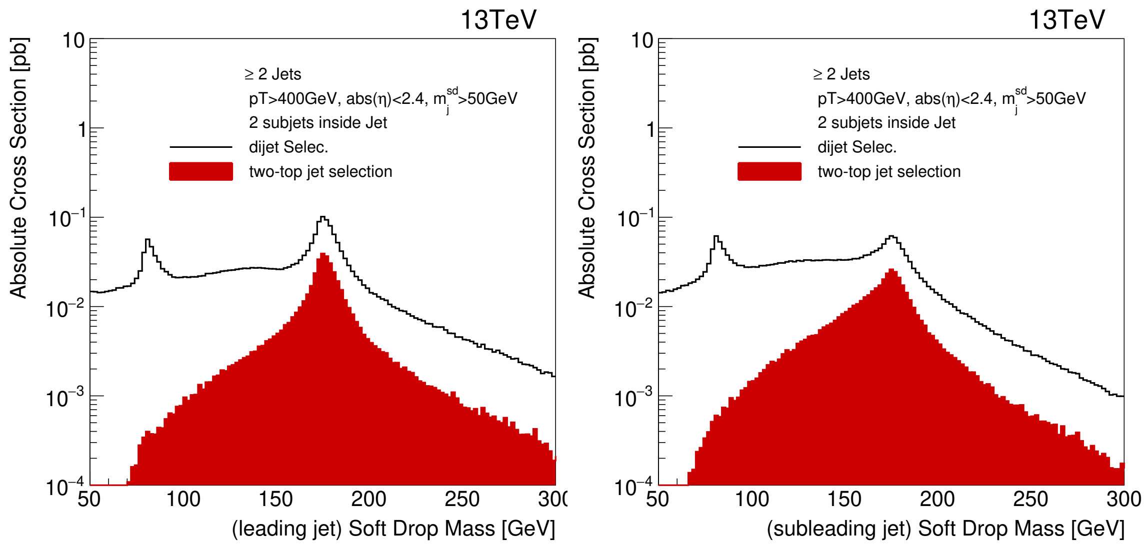


Figure 3.10: Soft drop jet mass distributions in $t\bar{t}$ simulated events for: (left) leading jet (right) subleading jet. Two selections are shown, the dijet boosted selection (black line curve) and when two top-jets are required (red shaded curve).

where $p_{T,i}$ stands for the transverse momentum of either leading or subleading jets, $N(\text{top-jet}, p_T)$ are the number of events selected by requiring two top-jets, in a certain p_T bin, and $N(\text{dijet}, p_T)$ are the number of events selected with the boosted dijet.

With this efficiency, one can determine the fraction of boosted $t\bar{t}$ events which satisfy the requirement that the two leading jets are properly reconstructing each of the $t\bar{t}$ quarks. The efficiency varies approximately in the range $10\% - 25\%$, having a decreasing behavior towards large p_T. Lower values of efficiency for the leading jet with respect to the subleading jet are observed. This fact is directly related to scenarios previously represented if Figure 3.6 (left picture), where exclusively, the leading jet is the one not properly reconstructing the t-quark.

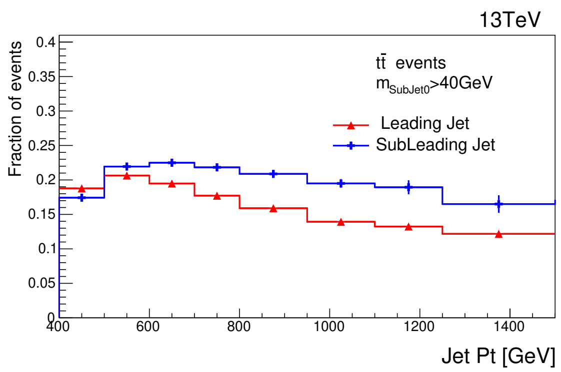


Figure 3.11: Event selection efficiency as function of the transverse momentum of the leading and subleading jets. As selection criteria two signal top jets are required, and as baseline selection, the boosted dijet selection is considered.

Chapter 4

Experimental setup at the CMS experiment

In this Chapter, the experimental setup used to perform the measurements is discussed. The first section (4.1) is focused on the accelerator facility used to collide protons: the Large Hadron Collider (LHC). The collision data are recorded by several experiments, one of them is the CMS experiment. A detail description of this experimental facility is provided in section 4.2, since this experiment recorded the data used in this thesis.

4.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a high energy accelerator facility located at the border between France and Switzerland hosted by The European Organization for Nuclear Research (CERN) [99]. The design of the accelerator was published in 1995 [100], and the construction of all the caverns and experiments lasted until 2008. The LHC was built in the same tunnel of a previous accelerator facility: the Large Electron Positron Collider (LEP) [101], which was colliding particles until 2000 with collision energies of hundred of GeV scales. Up to now, the LHC is the largest accelerator facility in the world.

The LHC is designed for colliding mainly protons with center of mass energy of up to $\sqrt{s} = 14$ TeV (each of the colliding protons beam with 7 TeV). Additionally heavy ions (Pb) with 2.8 TeV per nucleon can be collided. The beams of protons follow a tunnel 27 km long. In 2010 the first stable collisions at $\sqrt{s} = 7$ TeV took place, and in 2012, the energy collision was increased up to $\sqrt{s} = 8$ TeV. The collision period corresponding to the years from 2010-2013 are known as RunI period. After two years of shut down period, in 2015, the LHC started with RunII period, which aimed to reach energies up to $\sqrt{s} = 13$ TeV. Currently, the LHC is in shut down period, preparing conditions for the next run, which will reach an energy of $\sqrt{s} = 14$ TeV.

There are four main experiments around the LHC tunnel, designed with specific configuration and different purposes. Those experiments are:

- ✓ the Compact Muon Solenoid (CMS) [55, 102, 103]
- ✓ A Toroidal LHC ApparatuS (ATLAS) [56]
- ✓ the Large Hadron Collider beauty (LHCb) [104]
- ✓ A Large Ion Collider Experiment (ALICE) [105]

The CMS and ATLAS experiments, having different designs, are both multi-purpose machines, designed to cover the studies for a broad range of phenomena: from testing the Standard Model of Particles Physics, up to discovering particles Beyond the Standard Model. The LHCb, on the other hand, focuses on studying b-quark flavour and CP violations. The ALICE detector is focused on the studies of heavy-ion physics. Each of the LHC collaborations aimed to answer fundamental questions like how the fundamental particles in the SM acquire mass, why there is in the universe more matter than antimatter, among other open questions at the moment.

Additionally to the main experiments, there are three more facilities collecting data from the LHC. Those ones are very specific detectors with a unique final purpose. The *Total, Elastic and Diffractive Cross Section Measurement* (TOTEM) [106] measures the total pp collision cross section using very forward region (small angles), which is inaccessible by the other experiments. The Large Hadron Collider forward (LHCf)[107] is another of the relative small experiments at the LHC, and is also a forward detector but devoted to study the particles produced in the forward region in order to simulate cosmic rays. Finally, the Monopole and Exotics Detector at the LHC (MOEDAL) [108] detector is dedicated to search for an hypothetical particle with magnetic charge.

In order to achieve the goal of the input collision energy, the beam of protons undergoes by several accelerations procedures [100], which gradually increase the energy before entering to the LHC ring, where the collisions take place at the points where the main experiments are. Figure 4.1 shows a sketch of the acceleration procedure, together with the location of the experiments.

First, protons are obtained by extracting the electrons from hydrogen atoms, which are then injected into a linear accelerator facility (LINAC2). After this stage, the protons should have an energy of 50 MeV, and they are injected into the BOOSTER Proton Synchrotron, reaching energies up to 1.4 GeV. Afterwards the beam enters a larger circular accelerator facility: the Proton Synchrotron (PS). After this step, the protons already have 26 GeV and are injected into the Proton Supersynchrotron (SPS), where they are accelerated up to 450 GeV. Finally, the already energetic protons are injected into the LHC ring, which accelerates them up to the final collision energy (e.g 13 TeV, 14 TeV). The complete procedure, from when the protons are obtained, until they reach the LHC tunnel, lasts approximately 16 minutes. Additionally, before they collide in specific points where the experiments are located, tests of the beams need to be performed, lasting the whole procedure around 70 minutes. Instead of producing continuous beams of protons, beams are produced in bunches. A LHC beam has as standard value, 2808 bunches per ring with 288 bunches per injection coming from the SPS. Each bunch contains $1.15 \cdot 10^{11}$ protons in a beam size around 2.5 micrometers.

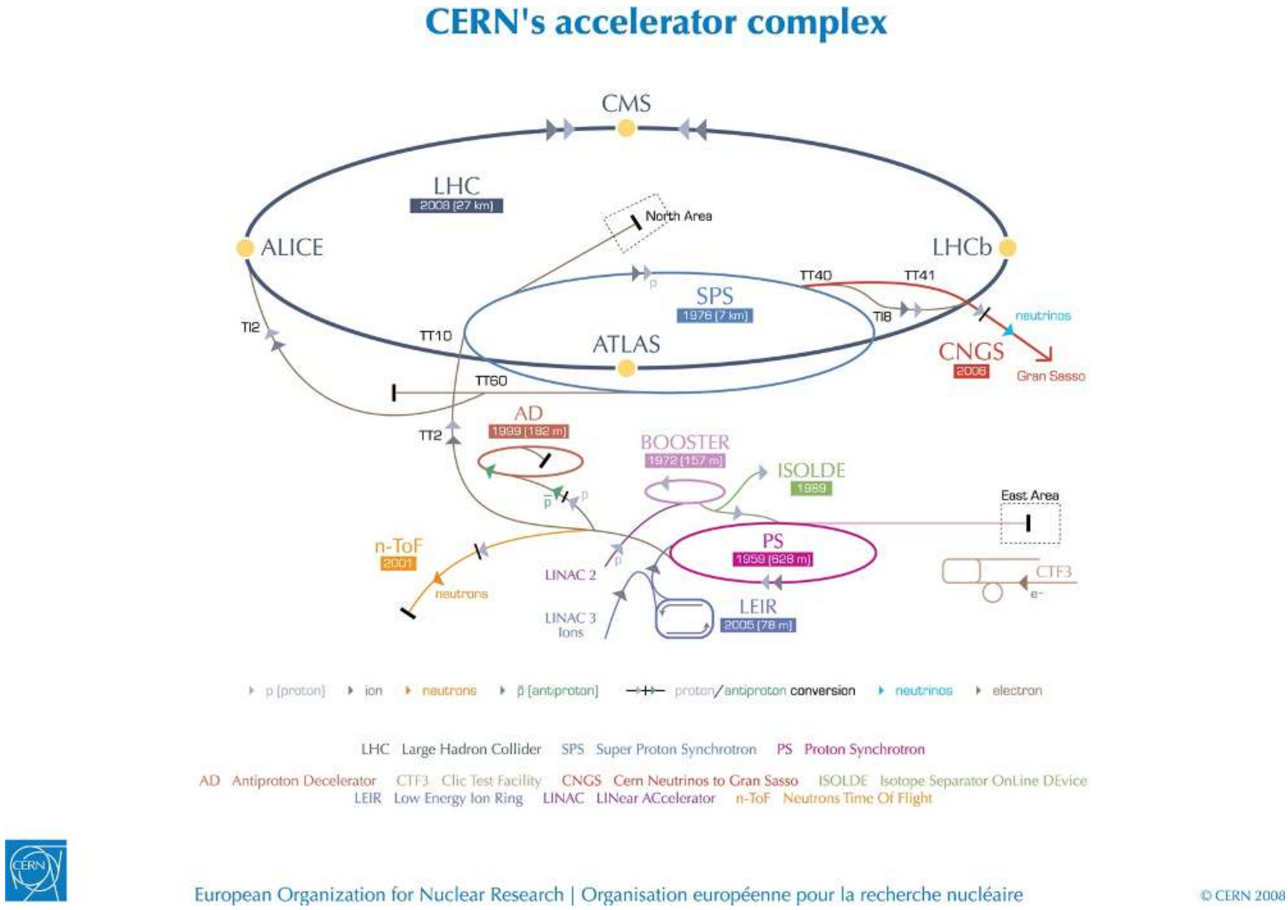


Figure 4.1: (taken from [109]) Sketch representing the CERN experimental facilities. Each of the experiments and accelerators facilities are illustrated. The LHC ring is 27 km long, and along the main ring, four main experiments are located: CMS, ATLAS, LHCb, and ALICE. In the figure, the direction of the beams, and which kind of particles are accelerated in each of the facilitates are presented.

One of the most important quantities is the instantaneous luminosity, which is meant to define the number of pp collisions per second per unity of area. The LHC is designed to produce a luminosity of up to 10^{34} per square centimeter per second. The number of events of a certain process occurring during pp collisions in a certain interval of time, can be estimated as function of the specific cross section and the integrated luminosity as follows:

$$N = \sigma \cdot \int \mathcal{L}(t)dt, \tag{4.1}$$

where $\mathcal{L}(t)$ is the instantaneous luminosity while $\int \mathcal{L}(t)dt$ is the integrated luminosity over a specific period of time. The luminosity is measured online, at the same time that the collisions are occurring, and also offline with a dedicated analysis, by all the experiments. The measurements of the luminosity performed by the CMS collaboration, corresponding to the 2016 data taking period can be found in reference [110]. The analysis performed in this thesis considers the collected data by the CMS experiment corresponding to the 2016 data taking period.

Figure 4.2 shows the integrated luminosity over the whole 2016 data taking period (left plot), which reached 40.99 fb^{-1} (while recorded by the CMS experiment 35.92 fb^{-1})[110]. The right plot shows the integrated luminosity, of both delivered by the LHC and recorded by the CMS experiment on a day-per-day basis.

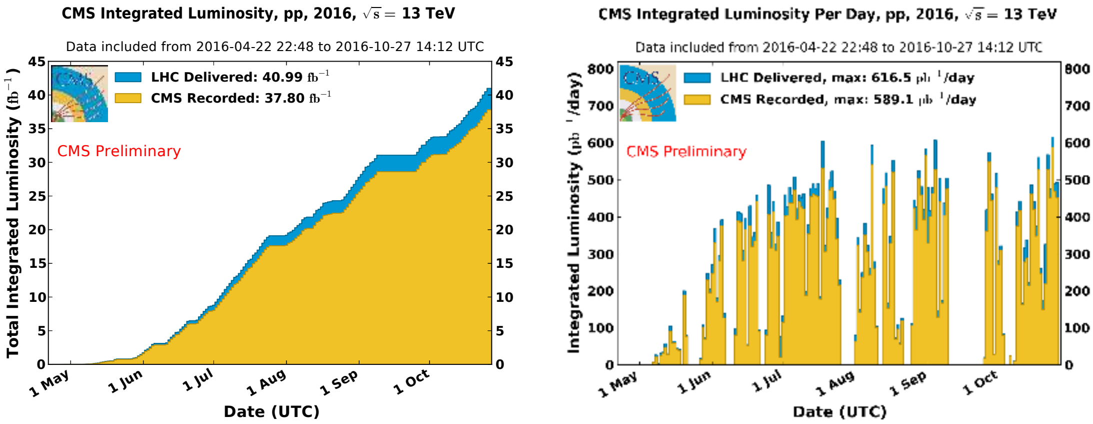


Figure 4.2: (taken from [111]) (left) integrated luminosity over the whole 2016 taking data period, (right) integrated luminosity on a day-per-day basis during the 2016 taking data period. Both, LHC delivery luminosity and CMS recorded luminosity are shown.

4.2 The CMS detector

The CMS detector is one of the two multi-purpose experiments at the LHC. The detector design consists of a cylindrical onion-like machinery covering almost all the interaction region. There are several subdetectors located from very few *cm* from the interaction point, towards the outside layer of the detector in a cylindrical configuration. Figure 4.3 shows a sketch of the CMS machinery of 15 m of diameter, representing the main components of the detector.

The part of the detector closer to the beam axis is the Silicon Pixel Tracker, followed by the silicon strip tracker, the electromagnetic and hadronic calorimeters. Those inner subdetectors are surrounded by the Solenoid Magnet and the Muon chambers. In each of the sides, forward calorimeters are also located. Each of the specific components will be further explained in this section.

The Solenoid Magnet [112] surrounding the inner detectors is a peculiar feature of the CMS experiment, providing a magnetic field of 3.8 Tesla. This allows the particles to deviate from their original direction after the collisions according to their momentum and electric charges.

The CMS coordinates system

The following coordinate system is used by the CMS collaboration in order to provide consistent results from different measurements:

- ✓ the origin of the coordinates system is assumed in the nominal interaction point,
- ✓ the x-axis looks towards the center of the LHC ring,
- ✓ the y-axis points vertically towards the surface,
- ✓ the z-axis is guided in the rotational axis of the cylinder with direction clockwise proton beam.

The azimuthal angle (ϕ) is defined in the x-y plane, while the distance towards the radius of the cylinder is denoted by r. Other quantity often used is the pseudorapidity (η), which is defined as

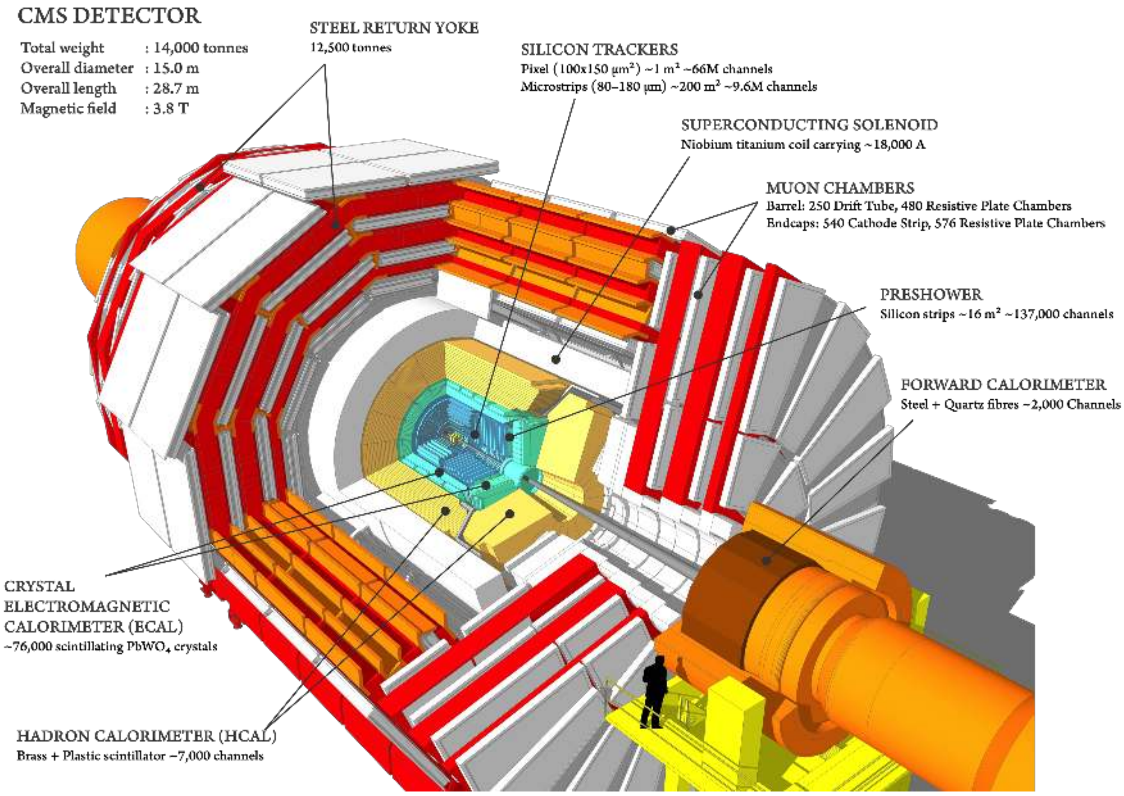


Figure 4.3: (taken from [113]) Sketch representing the CMS detector with the corresponding subdetector. Some features of the detector are additional provided.

function of the polar angle θ as follows:

$$\eta = -\ln\big(\tan(\frac{\theta}{2})\big), \tag{4.2}$$

The quantity η is represented geometrically in the Figure 4.4, and in the case of particles without mass, this observable coincides with the rapidity defined as function of the energy E and longitudinal momentum p_L as follows:

$$y = \frac{1}{2}\ln(\frac{E+p_L}{E-p_L}). \tag{4.3}$$

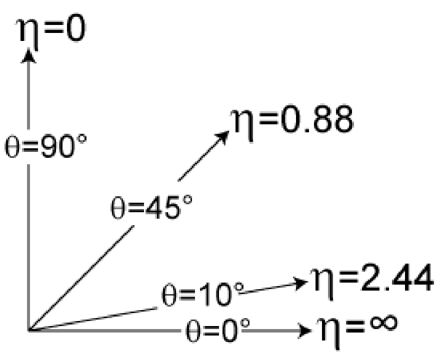


Figure 4.4: (taken from [114]) Relation between the polar angle and the variable η defined in equation 4.2. The lower values of η correspond in the CMS coordinates to the most central region, while the larger values to the most forward region.

4.2.1 The tracker system.

The tracker system [115] is the most inner part of the detector, responsible for reconstructing the momentum of the particles, measuring the trajectory of the charged particles affected by the magnetic field. Details of how this part of the detector intervene on the event reconstruction is provided in section 5 of the next chapter. A sketch of the design of the CMS tracker system is shown in Figure 4.5. Each of the layers of the tracker is made of silicon.

At the very inner part the Pixel detector is placed, which by the time when the data used in this thesis was collected, was formed by 3 layers (3 coaxial barrels, BPIX, located from the interaction point between 4.3 cm and 10.4 cm). Additionally, there are two rings (FPIX) at ±35.5 cm and ±46.5 cm from the interaction point along the z-axis, covering a total range of $-2.5 < \eta < 2.5$ phase space region. Later, in 2017, the Pixel detector was upgraded incorporating one coaxial layer more and a third disk, to reach even better resolution in the reconstruction of the tracks of the charged particles. In the 2016 version, the pixel was made up by 1440 pixels modules with a total of 65 millions of pixels of size of 100x150 μm^2.

The Silicon Strip Tracker is the following subdetector which is also part of the tracker system. The outer strip detector covers an area of up to 200 m^2 (the pixel detector covers 1 m^2). It consists of four main subdetectors: the Tracker Inner Barrel (TIB), the Tracker Outer Barrel (TOB), the Tracker Inner Disks (TID), and the Tracker EndCaps (TEC). The two former ones, TIB and TOB, cover the most central rapidity region with barrel layers (parallel to the beam axis). The latter ones, TID and TEC, consist of perpendicular layers to the beam axes covering the remaining part of the η phase space region of the whole tracker system.

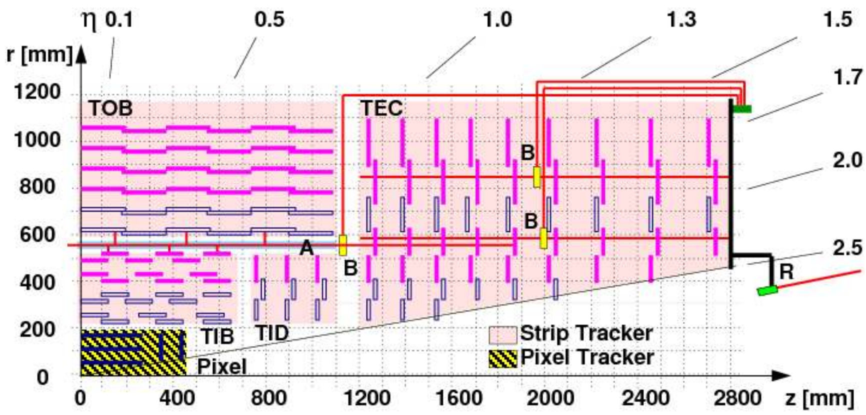


Figure 4.5: (taken from [116]) Sketch showing the tracker system in the r-z plane: the Pixel detector (inner part) and the Strip Tracker located in the outer part composed by four subdetectors: TIB, TID, TOP and TEC.

4.2.2 The electromagnetic calorimeter

The Electromagnetic calorimeter (ECAL) [117] is in charge to measure the energy of the electrons and photons. It consists of an hermetic and homogeneous design, sketched in Figure 4.6. The subdetector is made of lead tungsten crystals ($PbWO_4$) having high response and granularity. It consists of two main parts: the barrels, orientated through the z-axis (EB) and the endcaps (EE), which are oriented in a direction perpendicular to the beam axis. The whole system covers pseudorapidities up to $|\eta| < 3$. In addition, a preshower is located in the most forward region: $1.65 < |\eta| < 2.6$.

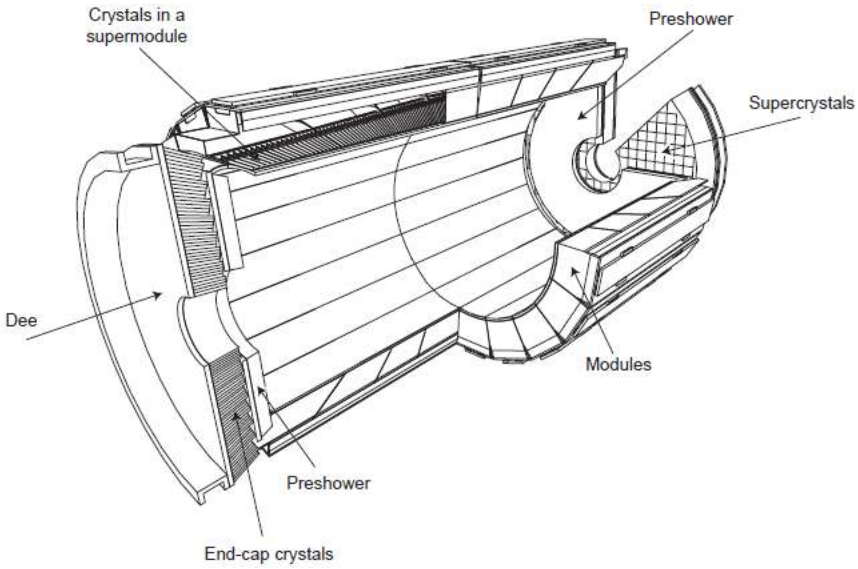


Figure 4.6: (taken from [117]) Sketch showing the Electromagnetic calorimeter (ECAL), the part of the detector responsible to measure the energy of the electrons and photons.

4.2.3 The Hadron Calorimeter

The Hadronic Calorimeter (HCAL) [118] completes the calorimeter detector, which is measuring the energy of all the charged and neutral hadrons. It surrounds the ECAL, and is also inside the Magnetic Solenoid. A sketch of this detector is shown in Figure 4.7, consisting of four main parts: the barrels, the endcaps, the hadronic outer and the hadronic forward (which is not represented in the Figure) located in the most forward region. The Barrel HCAL covers an η phase space within $|\eta| < 1.4$, while the EndCap HCAL extends the region up to $|\eta| < 3$. The Forward HCAL covers the region $2.9 < |\eta| < 5$. The Outer HCAL is responsible for ensuring the energy reconstruction missed by the Barrel and End Cap, and has a ring structure as shown in the picture.

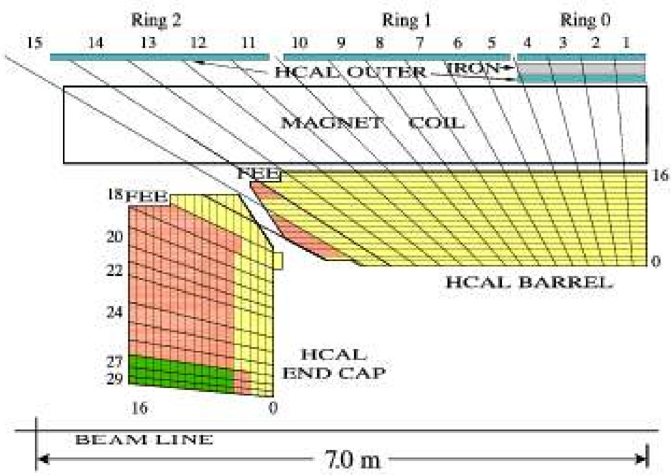


Figure 4.7: (taken from [118]) Schematic view of part of the HCAL in the CMS detector. The Forward HCAL is not shown which is located in the most froward region.

4.2.4 The muon system

The Muon detector [119] is located in the outer part of the detector, and only muon and neutrinos can reach the subdetector. This detector is devoted to identify and reconstruct muons. It is divided in three main parts: the drift tubes (DT) located in the most central region $|\eta| < 1.2$, the cathode strip chambers (CSC) (in $0.9 < |\eta| < 2.4$) in the endcap, and the resistive plate chambers (RPC). Figure 4.8 shows a schematic picture of the muon system and its parts.

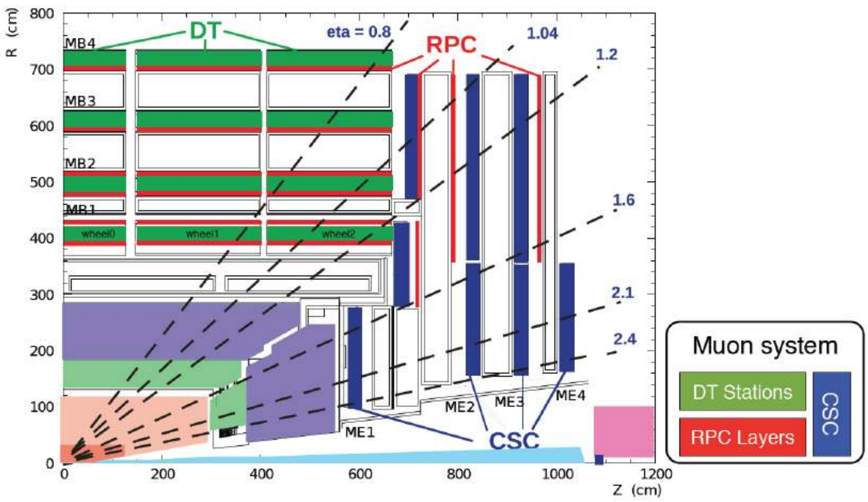


Figure 4.8: (taken from [119] Sketch showing the Muon system of the CMS detector located in the outer part of the CMS machinery.

4.2.5 The trigger system

The data collected by the CMS detector pass an online selection performed in the Trigger System [120]. The main idea is to select signal processes not occurring that often, and for those processes having a huge contribution, only a fraction of events are stored. The trigger procedure consists of two levels:

- ✓ the Level-1 Trigger (L1T): is a selection procedure performed at hardware level. The detectors are designed to accommodate 100 kHz from the total 40 MHz in just only 3.2 μs. It consists of using the energy deposited in the calorimeters and in the muon system to perform the validation of the events in less than $3.2\mu s$. The Global Trigger is the last L1 step, where the decision is made. After this L1 trigger, objects like the photons, electrons, and jets are reconstructed.
- ✓ the High Level Trigger (HLT): using the information received from the L1 Trigger system, the HLT triggers start to reconstruct the events and they are stored or remove depending if they satisfied any of the path triggers. There are signal path triggers, which are selecting events possible coming from some interesting processes, and storing all the event passing the path. There are other trigger paths with higher rates, and those only select a fraction of events. These HLT trigger normally have lower-thresholds p_T.

Prefiring issue in the L1 trigger system

During the 2016 and 2017 data taking period there was a problem with the L1 trigger system, affecting mainly events taking place in the forward detector region. At this level, a mistimed readout caused loss of events and hence inefficiencies in the data taking. The problem was detected during 2017 and ways to consider those inefficiencies to properly performing measurements are implemented by the CMS collaboration. However, those inefficiencies affected high p_T jets in the $2.5 < |\eta| < 3$ region, and since the analysis presented in this thesis is performed in the central region ($|\eta| < 2.4$), this issue did not affect the measurements.

4.2.6 Data Quality Monitor

The Data Quality Monitoring system (DQM) [121] is responsible for cross checking the hardware and software intervening in the data processing at the early stage of the data taking. The procedure is divided in two parts: one online, which is preformed as soon as the data is collected by the Trigger system, and one offline, which cross checks the results from the online DQM. In the online DQM, the Graphical User Interface (GUI) provides histograms of crucial features corresponding to different subdetectors, where problems for instance on the detector or trigger systems can be noticed. The data is monitored at this stage directly where the CMS is located. Afterwards, already considering the event reconstruction (it will be discussed in the next Chapter), the offline DQM is performed.

In physics analysis carried by the CMS collaboration, normally only the data classified by the DQM experts as good is used, to ensure that no bias is introduced because of obvious inefficiencies in the detector or trigger system.

Chapter 5

Event Reconstruction in the CMS experiment

This chapter aims to describe the event reconstruction at the CMS experiment. The information obtained from each of the components of the CMS detector, described in Chapter 4, is gathered and the signature left by particles passing trough the detector is used to reconstruct events in a hadron-hadron collision.

The Monte Carlo event simulation is a crucial point in any experimental analysis. The Monte Carlo predictions presented in Chapter 1 correspond to the level where a hadron-hadron collision occur (particle level), which is not comparable with the reconstructed data from the experiment. In order to bring both to a comparable level the final stable particles after a proton-proton collision, obtained from Monte Carlo event simulation, are then considered as input information to perform a further simulation of how they interact with each of the components of the detector. Similarly as it is performed in the data, reconstruction algorithms provide as output information stable objects which then can be compared to the measurements.

Figure 5.1 represents schematically how each step is considered in Monte Carlo event simulations. Four basic stages are emphasized (represented by each row of the picture):

- ✓ first, based on the Standard Model and using a Monte Carlo event generator, final state particles are obtained from the simulation of hadron-hadron collisions,
- ✓ secondly, the interactions of those outgoing stable particles with the detector are simulated, using Geant4[122], which is a Monte Carlo detector simulation tool. This tool simulates the hits of the particles in each part of the detector,
- ✓ those hits are considered through a detector response simulation, adding realistic detector responses like pileup, alignment calibration, noise, voltage signal. With those considerations more realistic Monte Carlo simulations at detector level are obtained.
- ✓ Finally, in the last step, those tracks are gathered together with reconstruction algorithms where final objects are obtained. Those reconstruction algorithms are applied in both, data and Monte Carlo simulations.

This chapter focuses on describing the event reconstruction at the CMS detector. After those considerations are implemented in the analysis, and the selection of events are applied, generally, one could perform a comparison of the theory predictions with the data at reconstructed level. However, this is not what it is aimed in this thesis, additionally, the reconstructed data need to be unfolded to the particle level reverting the detector effects accounted in the Monte Carlo simulation. This procedure is known as *Unfolding* and will be further discussed in Chapter 9.

The chapter is divided in four main blocks covering the following topics:

- ✓ First, in section 5.1, the algorithm responsible for reconstructing the particles and the collision vertices is presented. This algorithm is the so-called Particle Flow, used as well in many other experiments.
- ✓ Secondly, in section 5.2, the hadronic jet reconstruction as measurable objects, in the CMS experiment, is presented. There, the jet energy resolution related to the Particle Flow and detector performances are discussed.
- ✓ In section 5.3, the calibration of those hadronic objects (jets) is presented. This is a crucial point in the event reconstruction, in order to provide observables with precise measured quantities, i.e their transverse momentum.
- ✓ Finally in section 5.3, a crucial point for this thesis is covered: the identification of b-jets. Signal $t\bar{t}$ events, as was already discussed, is characterized by the presence of b-quarks, and the identification of jets containing those objects is an important step for the event selection, presented later in Chapter 6

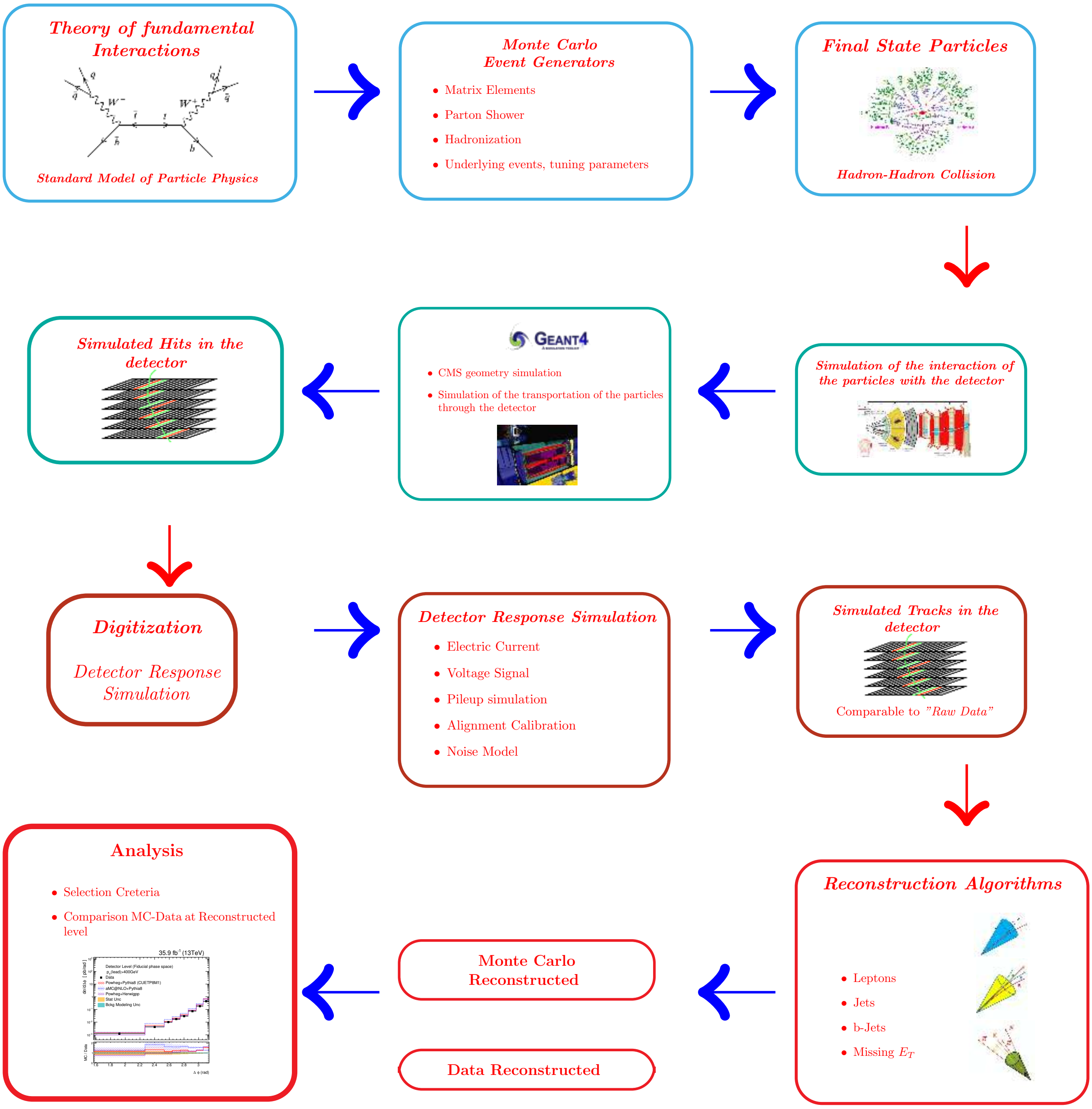


Figure 5.1: Sketch representing the steps considered at the CMS experiment for bringing to the same level Monte Carlo predictions and reconstructed data.

5.1 Particle Flow algorithm and track reconstruction.

The Particle Flow (PF) algorithm [123] is the algorithm implemented at the CMS experiment in order to reconstruct from the collision data stable objects. Hence, the PF algorithm is responsible of assigning signatures left by a bunch of particles to specific types of reconstructed objects. The basis of the algorithm relies on gathering the information provided by each of the subdetectors. This algorithm was developed by the ALEPH [124] Collaboration at LEP. Other collaborations have used the key points of this algorithm, i.e the ZEUS [125], H1 [30], and CDF [126] Collaborations.

An illustrative picture for the basic principles of the PF algorithm at the CMS experiment is given in Figure 5.2. The trajectories of different types of particles depend on how they interact with the specific part of the detector that they pass through. The general features of each part of the detector were already discussed in Chapter 4.

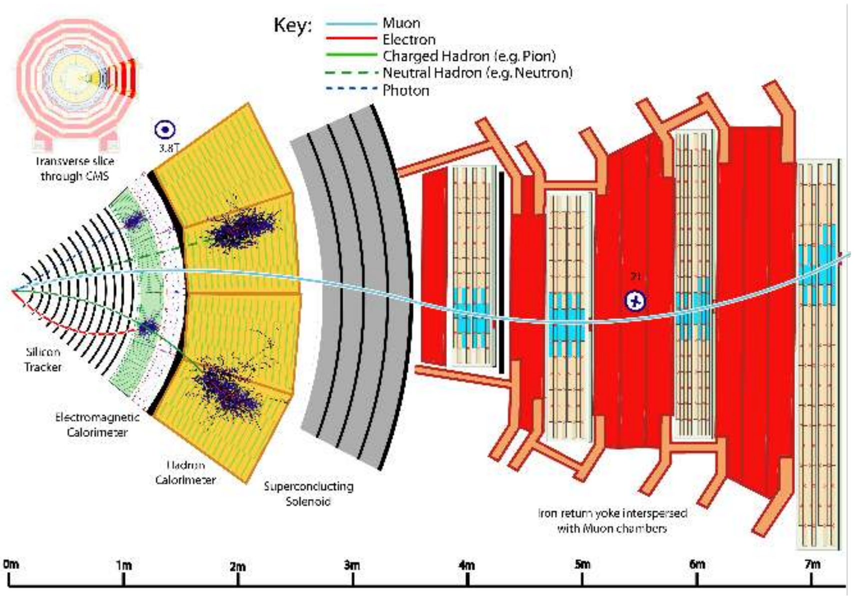


Figure 5.2: (taken from [123]) Schematic picture representing the interaction of particles with different parts of the CMS detector: five types of stable particles are illustrated: muons, electrons, charged hadrons, neutral hadrons, and photons. The Tracker System, Electromagnetic Calorimeter, Hadron Calorimeter, and Muon System are represented.

The stable particles reconstructed using the PF algorithm are:

- ✓ *Neutral and Charged Hadrons*: they are reconstructed from the deposited energy in the ECAL and HCAL, and in the case of charged hadrons, the information from the tracker system is used for better efficiency and resolution reconstruction. For instance, charged hadrons with p_T up to large p_T are preferable reconstructed with the tracker information, while for higher p_T, the measured energy in the calorimeters offers better resolution in comparison to the reconstructed energy from the tracker information. The Tracker Silicon Detector can reconstruct tracks of charged particles very precisely, even for particles with low transverse momentum down to 150 MeV and $|\eta| < 2.5$. Neutral hadrons are detected as an energy excess on the deposited energy by the charged hadrons in the same calorimeter cells.
- ✓ *Electrons*: they are reconstructed combining the information of the tracker system and the ECAL, where they deposit all the remaining energy. Due to their interactions with the material, often they emit Bremsstrahlungs photons. The signature of those photons is matched to

the corresponding electrons using the advantage that their energy deposition in the calorimeters is similar as for the electrons. This information is used to fully reconstruct the energy of the electrons. In the ECAL, in order to identify the clusters related to these particles, a technique named Gaussian Sum Filter [127] is applied.

- ✓ *Muons*: *isolated muons* can be reconstructed by combining the tracker system and the muon chamber information. There are two ways of reconstructing them. The simplest method takes into account only the tracker information and extrapolates it to the muon chambers considering the possible ways of losing energy caused by the interaction of the muons with the detector. The other method, called *isolated global muons*, consists of matching the tracking and muon information using the Kalman-filter technique [128]. This second method is more CPU consuming, but has better efficiency for reconstructing muons with larger p_T. The reconstruction of *non isolated muons* needs more sophisticated selection procedures [129].

- ✓ *Photons*: they are reconstructed using only the information from the ECAL. This part of the detector has a very fine granularity to distinguish between charged particles from photons. A photon candidate with transverse energy larger than 10 GeV is seeded in the ECAL if there is no track associated to the particle, in order to distinguish it from Bremsstrahlungs photons.

Identified and reconstructed charged hadrons, neutral hadrons and photons are clustered in hadronic jets and hadronic taus. With the precise reconstruction of jets, the exact missing traverse energy (E_T^{miss}) can be estimated indicating the presence of neutrinos or other possible signatures of particles that are not interacting with the detector.

Track reconstruction and linking algorithm for the PF implementation

The Silicon Tracker Detector is the part of the complex CMS apparatus responsible for the efficient track reconstruction of charged particles.

The first step of the tracking reconstruction is based on a simple vertex reconstruction algorithm named as Kalman Filtering (KF) [128], where three basic steps are performed: first, few hits compatible with the charged particle trajectory are detected; then the trajectory is built using *pattern recognition* techniques combining hits from different tracker layers; and finally a *fit* is performed in order to determine the charged particle properties (i.e which kind of particle and momentum).

However, using this reconstruction method, the efficiency for particles with $p_T < 10$ GeV decreases considerable. More complex algorithms [130] [128] have been developed in order to improve the trajectory reconstruction, based on iterative steps. Those iterative approaches allow the reconstruction with good efficiency ($\epsilon > 70\%$) for particles with p_T from 1 GeV up to 100 GeV. The tracking efficiency decrease, however, towards increasing p_T, and for instance for particles with $p_T > 400$ GeV, it takes values $\epsilon < 50\%$ [123].

After the tracks have been reconstructed, a linking algorithm is applied in order to determine which tracks belong to the same particle. The PF algorithm performs the next subsequent steps in order to gather the information and assign specific particles to the signatures:

- ✓ the information from the Muon and ECAL detectors are used to directly identified electrons and muons,

- ✓ if the deposited energy in the ECAL significantly excesses the transverse momentum reconstructed by the tracker system, a photon is associated to the signature in the ECAL,

- ✓ if the deposited energy in the HCAL is much larger than the transverse momentum reconstructed by the tracking, a neutral hadron is associated to the signature,

- ✓ if the transverse momentum reconstructed in the tracker system is similar to the energy reconstructed by the ECAL and HCAL, then a charge hadron is assigned,
- ✓ if the energy measured in the calorimeters is smaller than the transverse momentum reconstructed, then steps to associate muons or fake tracks to the signatures are applied.

The reconstruction of neutral particles (i.e. photons and neutral hadrons) relies exclusively on the calorimeters (ECAL, HCAL). In case of charged hadrons with $p_T < 100$ GeV, the tracker system is the one with better reconstruction efficiency, however, for higher p_T, the calorimeters play a major role in their reconstruction. For instance the energy resolution has the following dependence [131]:

$$(\frac{\Delta E}{E}) \propto \frac{1}{\sqrt{E}} \tag{5.1}$$

5.2 Hadronic jet reconstruction.

Stable particles reconstructed by the Particle Flow algorithms ($c\tau > 1cm$) are the input to the jet clustering algorithm. The jet clustering algorithm preferable used by the CMS Collaboration, is the anti-k_t sequential recombination algorithm [90], through the implementation in the FastJet[132] software. The standard choice of the jet cone size radius parameter in the clustering algorithm is $R = 0.4$. For boosted toplogies, however, the standard jet radius considered by the CMS collaboration is $R = 0.8$ (fat jets).

During RunI period, grooming techniques were used by the CMS Collaboration for pileup suppression, together with charged hadron subtraction techniques. The pruning algorithm was the one performing better [133]. During RunII period new grooming techniques were implemented based on Soft Drop Mass methods [85], together with a new algorithm to remove pileup (PUPPI) [134]. Efficiency studies comparing Soft Drop mechanism with the standard grooming algorithms performed by the CMS Collaboration can be found in the references [135, 136].

In the simulation, the clustered jets are associated to a specific flavour definition. However, this is an ambiguous definition since a jet is an object constituted for different kind of particles. The jets are classified as: light jets, b-jets and c-jets. The basic idea followed to classify them is assigning to all partons (or hadrons) a weight such that they have very small contribution, and recluster them with the jet itself. These *scaled* partons (hadrons) are the so-called *ghost* applied for the definitions (depending which method is used, parton or hadron reconstruction):

- ✓ a jet is classified as b-jet if there is at least one b-*ghost* parton (hadron) clustered inside the jet,
- ✓ a jet is classified as c-jet if there is at least one c-*ghost* parton (hadron) constituent inside the jet, with the condition that it was not b-jet classified,
- ✓ a jet is classified as $light$-jet if it was not previously classified either as b-jet or c-jet.

A jet must be classified from either the parton or hadron definition in the same category. If it is not the case, the assumed definition is the one using the hadron information (which is considered more stable).

5.2.1 Particle Flow jet composition

The charged hadron forming the jet have the advantage that they are reconstructed mainly with the tracker information which provide very precise resolution. However, non-charged particles also forming jets are fully reconstructed with other parts of the detector (i.e ECAL, HCAL), and thus they don't have this advantage. The jet resolution, hence, depends on their composition.

Figure 5.3 illustrates the PF jet composition measured in data in the CMS experiment, using proton-proton collisions at center of mass energy $\sqrt{s} = 13$ TeV. The measurement are performed using predictions obtained in dijet simulated events, since in the latter ones, truth (particle) level information is a priori known (i.e the jet composition). More details on this specific measurement can be found in reference [137]. The PF jet composition is illustrated with respect to p_T of the *tag* jet[1].

From the figure it can be observed that the charged hadrons are $\sim$ 65% of the jet constituents, for $p_T < 400$ GeV. For higher p_T values, this contribution decreases, while the contributions given by the photons and neutral hadrons constituents are slightly increased. The pileup contribution, however, is considerably reduced in the higher p_T region.

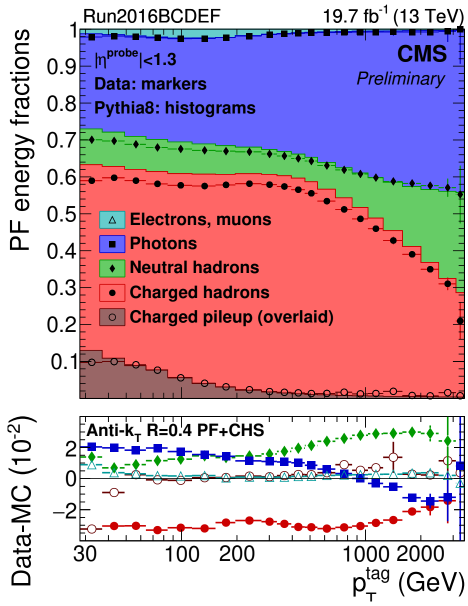


Figure 5.3: (figure taken from [137])PF jet composition (for data and simulation) as a function of the p_T of the "tag" jet which is in the barrel region ($|\eta| < 1.3$).

5.2.2 Particle Flow performance for jet reconstruction: the jet energy resolution

As it was previously shown, jets are composed of charged hadrons, photons, electrons, and neutral hadrons. Since neutral hadrons can be only reconstructed by the calorimeters, the jet energy resolution will be affected from limited resolution of this part of the detector. The jet energy resolution in the calorimeters is better however when more energetic jets are considered ($\sigma/E \propto 1/\sqrt{E}$).

The performance of the PF algorithm related to the jet reconstruction, has been studied in Monte Carlo simulation [138] and in data [139] [137] by the CMS collaboration. Those studies observed a considerably improvement of the resolution when jets are reconstructed with the PF algorithm with respect, for instance, when only the calorimeters information are used for the reconstruction. A brief discussion on the last statement is provided in the following.

Figure 5.4 shows an event display in a simulated event illustrating three different kind of reconstructed jets: *gen*-jet containing all the stable particles except neutrinos; the *PF*-jets containing all the reconstructed particles by the PF algorithm; and *Calo*-jets, which are reconstructed jets using exclusively just the information from the Calorimeters (HCAL, ECAL).

[1] the tag and probe method is used in the measurement, and the tag jet is in the barrel region ($|\eta| < 1.3$)

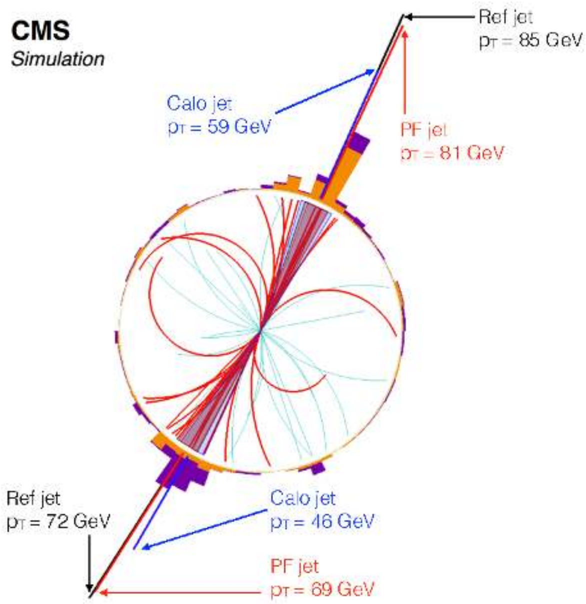


Figure 5.4: (from [123]) jet reconstruction in a simulated QCD dijet event. The *PF*-Jet is compared with the p_T of the reference jet (*gen-jet*) and the calorimeter jets (*Calo*-jets).

The *PF*-jets and *Calo*-jets are matched to the closer *gen*-jet, with the ΔR separation in the ϕ-η plane. For each matched pair of reconstructed jets (either *PF* or *Calo*-jets) the resolution response is estimated as follows:

$$\sigma = \frac{p_T^{rec} - p_T^{gen}}{p_T^{gen}} \tag{5.2}$$

where p_T^{rec} is the transverse momentum of the reconstructed jet, and p_T^{gen} is the transverse momentum of the generated jet. Normally, the resolution response is estimated in different p_T phase space regions.

Figure 5.5 (left plot) shows the estimated response resolution in the p_T^{gen} range 40 GeV$< p_T^{gen} < 60$ GeV, and $|\eta| < 1.5$. The difference between *Calo*-jets and *PF*-jets is noticeable: the resolution response is centered around approximately zero, confirming that the PF reconstructs very precisely the *gen*-jet. A completely different behavior is observed for the *Calo*-jets. A Gaussian function is fitted to the distribution and the width of the distribution over the central value (σ/μ) is considered as the jet energy resolution (JER) in the specific phase space region.

Figure 5.5 (right plot) shows the JER resolution as a function of P_T in the barrel region ($|\eta| < 1.3$), for both *Calo* and *PF* jets, showing the improvement of this value for the latter ones. The *PF* improved the JER, specially in the lower p_T region ($p_T < 200$ GeV) by a factor of about three. As might be noticed, the JER decreases as function of p_T, and this behaviour is given by the improvement of the calorimeters response for more energetic jets.

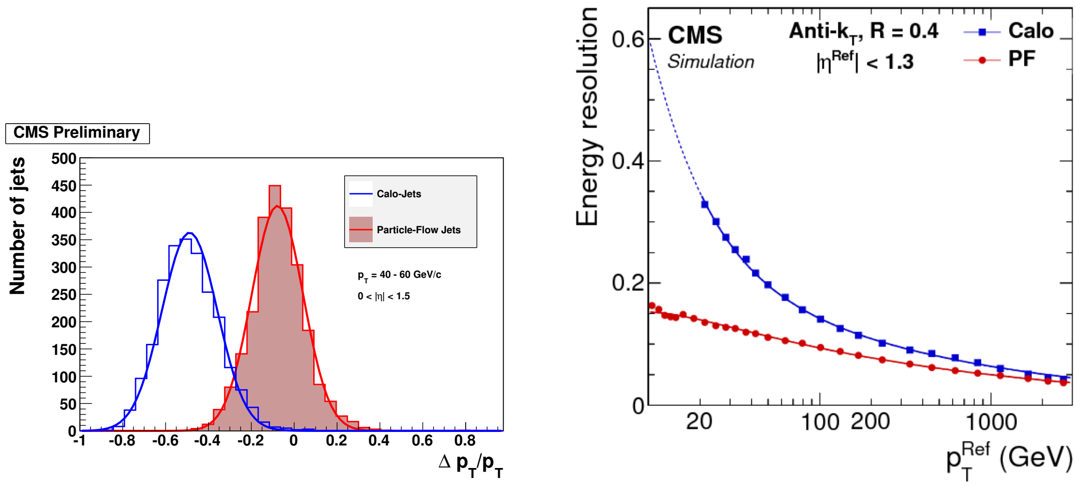


Figure 5.5: (taken from [123]) (left) Distribution of $(p_T^{rec} - p_T^{gen})/p_T^{gen}$ in the 40 GeV$< p_T^{gen} <$ 60 GeV region (right) JER as a function of p_T of the *gen*-jet in the barrel region. Both plots compare the performance of *PF* and *Calo* jets.

5.3 Jet Calibration.

The good reconstruction of the jet is a major key point in many analysis at the LHC, where jets are involved as final state objects. In real scenarios, the accuracy in which a jet is being reconstructed is affected for instance by pileup, detector noise, non-linear calorimeters response effects, among others. Therefore, corrections need to be implemented to those reconstructed jets.

This section describes a set of implemented calibrations in order to remove those bias effects mentioned before. They are results of dedicated studies performed by the CMS collaboration. Those corrections have uncertainties, since they are estimated by the combination of simulation and data driven methods. In many analysis they are one of the largest source of uncertainties and the effect of the uncertainties need to be estimated. The official recommendation for implementing both, the jet energy corrections (JEC) and their uncertainties can be found in reference [139].

The procedure is illustrated in Figure 5.6. It is characterized by a factorization approach where different level of corrections are implemented. Almost all the steps are applied to both, Monte Carlo and data, except the residuals corrections that are only applied to the data.

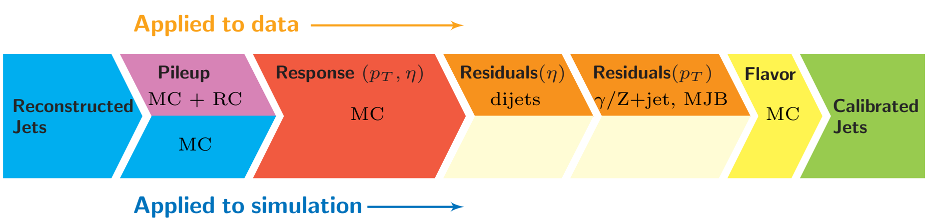


Figure 5.6: (taken from [139]) Graph showing the steps of JEC for Data and MC. All the correction that are referred with MC is because they are derived by MC simulation, while RC stands for random cone and MJB refers to multijet event were used for the simulation.

The set of corrections are:

- Level1 (L1) Corrections: *Offset Corrections*, concern the pile-up and noise suppression.

- Level2 and Level3 (L2L3) Corrections: *Response Calibration*, where from Monte Carlo the Jet Energy Resolution is derived as a function of p_t and η of the generated jet (explained in the last section). Here effects of non uniformity in η and non-linearity in p_t are the main concern. The jet is corrected at particle level.
- Level4 (L4): *Residual*, only applied to the data, in order to take into account differences between the responses in Data and Monte Carlo. Two kind of residuals are considered: relative (related with η) and absolute (related with p_t).
- Level5 (L5): flavour compositions of the PF jets.

Then factorized formula for correcting the p_T of jet can be expressed by the following equation:

$$p_t^{\text{corr}} = C_{\text{offset}}(p_t^{\text{raw}}) \cdot C_{\text{MC}}(p_t^{\text{raw}}, \eta) \cdot C_{\text{res}}(\eta) \cdot C_{\text{res}}(p_t) \cdot C_{\text{flavour}} \cdot p_t^{\text{raw}} \tag{5.3}$$

where, p_t^{corr} is the corrected p_T and p_t^{raw} is the p_T of the jet before the corrections, and each factor counts for each of listed corrections.

Generally, those corrections are centrally implemented, however the uncertainties related to those corrections need to be estimated in teach analysis separately.

5.3.1 Total JEC and uncertainties

The JEC uncertainties are summarized in Figure 5.7, as a function of p_T of the reconstructed jet, and as a function of η for jets with $p_T > 30$ GeV. Each of the individual sources previously mentioned are considered, as well as the total uncertainty.

As might be noticed, in the case of the uncertainties as a function of p_T, a decreasing behavior toward p_T is observed. In the lower p_T regions, values of up to 3%, for the total uncertainty, are reached, having the major contribution the uncertainties related to pileup and jet flavour. For jets with higher p_T (phase space region where boosted topologies are studied), the total JEC uncertainty is less than 1%.

From the η dependence, one can see that in the central region, for $p_T = 30$ GeV, values of up to $\sim 2\%$ are expected, while in the forward region, the total uncertainty increases up to $\sim 5\%$.

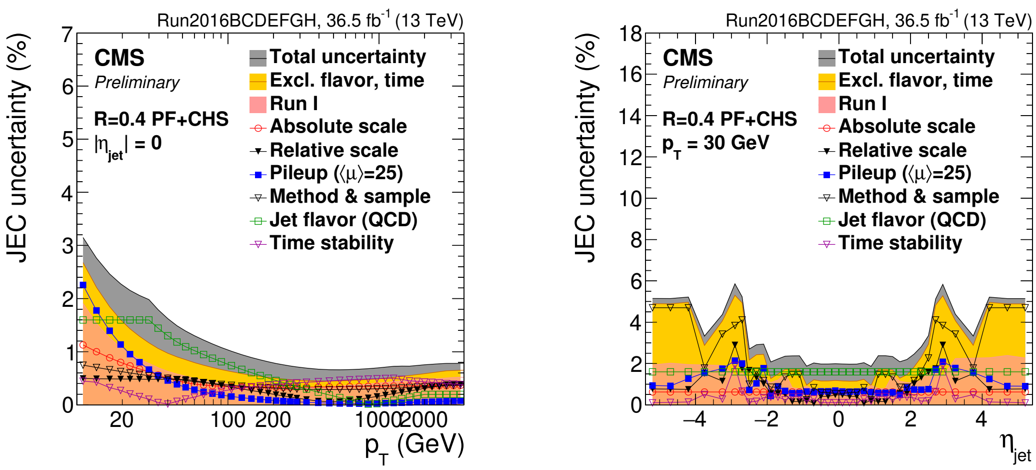


Figure 5.7: (from[140])Uncertainty of Jet Energy Scale with sum of all uncertainties in quadrature and the sources (left) as a function of p_t, (right) as a function of η.

5.4 Identification of b-like Jets

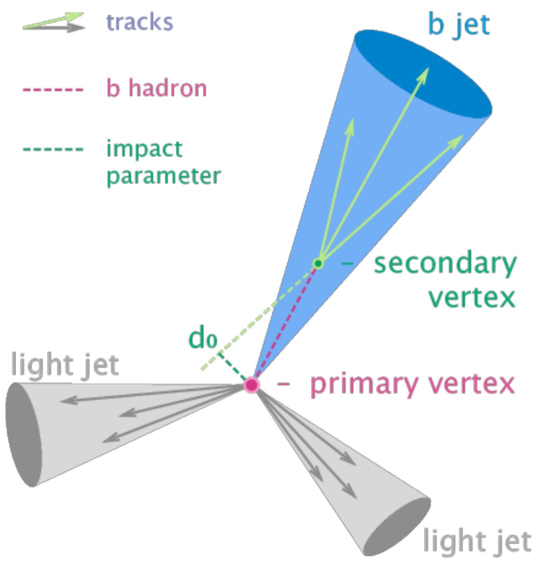
One of the mean features of $t\bar{t}$ events is the presence of $b\bar{b}$ pairs originating directly from the top and antitop quarks. In general, in pp collisions, b-quarks are mostly produced in pairs, either via s-channel or gluon-splitting. The identification of those b quarks is crucial for many data analysis performed at the LHC, and most of the analyses measuring top quark cross sections rely on this step. To achieve this goal, the CMS Collaboration has developed tools that allow to identify quite precisely jets containing b quarks.

The identification of those b-jets not only assures to identify the signature of $t\bar{t}$ events, but also to discriminate background events interfering with the signal. For the standard QCD multijet production, in principle, a considerable contribution contain also b-quarks, but nevertheless, the presence of b-jets in those events is governed by the b quark cross section which is much lower than the cross section for lighter quarks, i.e u-s-d jets or c-jets. That implies immediately that the fraction of removed events by requiring the presence of a b-jet in the latter one, is much larger than for the signal $t\bar{t}$ events.

The *B-hadrons* are final state particles that contain a b quark. These final state particles have peculiar features that can be taken at advantage for their signature identification in the detector. They can be observed in the detector because their long lifetime, and hence their travel distance is enough to be measured. Their lifetime is of the order of 1.5 ps (10^{-12}s) and the traveled distance before decaying, at average, is $\sim$ 1.8 mm (considering a mass of the B-Hadron $\sim$ 4.2 GeV and a transverse momentum of $\sim$ 20GeV). The distance covered before the decay allows to reconstruct a secondary vertex at the point where the decay occurs, and combining that information to the impact parameter of charge particles, algorithms to identify B-Hadrons have been developed and reported in the references [141, 142, 143, 144]. Figure 5.8 (left) illustrates a schematic picture of the secondary vertex reconstruction.

Figure 5.8 (right) shows the fraction for light-jets, c-jets, and b-jets, for three different vertex reconstruction categories obtained by the Inclusive Vertex Finder [145]. This classification is performed using the CSVv2 tagger [143] that will be further explain in this section. Events containing $t\bar{t}$ pairs with $p_T > 20$ GeV where simulated. The three considered categories were: *recovertex*, where the jet contains at least one secondary vertex; *novertex*, sorting jets not assigned to one of the other two categories; and an intermediate classification labeled as *pseudoVertex* (which is not further considered). By these probabilities, one can see that a secondary vertex is predomentinatly reconstructed in the case of b-jets, while the probability that in a b-jet, no secondary vertex is found, is small.

The performance of the developed algorithms is extremely challenging due to their dependence on the efficiency of the track reconstruction by the detector, as well the resolution of those tracks. The efficiency of the algorithms can be directly measured in data. The task of developing those algorithms that are able to actually identify the signature of b-quarks is carried by a dedicated group [147] in the CMS Collaboration, which has provided several tagging techniques through many years.



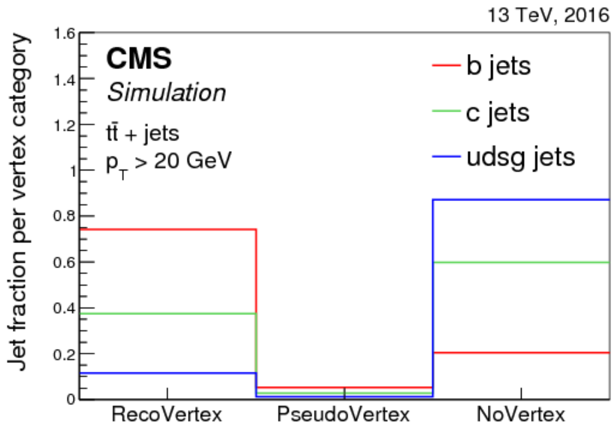


Figure 5.8: Secondary Vertex Reconstruction: **(left)** schematic picture of the second vertex reconstruction in a jet containing a B-Hadron (b-jet), Two light jets are also represented, which are originated directly from the primary vertex, and been distinguished from the one containing a secondary vertex (taken from[146]) , **(right)** probability of each vertex category for secondary vertices reconstructed with the Inclusive Vertex Finder [145] algorithm for light, c, and b-jets (take from [143])

B-tagging algorithms have as input information, variables related with the track impact parameters and the vertex reconstruction. The idea is to estimate as output a single discriminating variable that contains all the input information. For all the algorithms, there are three working points defined: the light, the medium and the tight working points (LWP , MWP, TWP respectively) and they represent the points with a nominal misidentification probability (probabilities that light jets are b-tagged) of 10%, 1%, and 0.1% respectively for an average of jet p_T of 80 GeV. The efficiency of b-tagging is generally p_T and η dependent, having a predominant behavior of decreasing when higher p_t of the jets are considered. This p_t dependence behavior is an important feature that newer algorithms are trying to improve, such that at higher p_t the efficiency keeps constant instead of decreasing.

The b-tagging algorithms used during RunI period at the LHC [141] are classified in three groups depending on the given input information:

- *Algorithms using track impact parameters*: those algorithms sort tracks in a jet by decreasing values of the impact parameter significance.
- *Algorithms using secondary vertices*: the Simple Secondary Vertex (SSV) uses the significance of the decay length.
- *Algorithms combining track parameters and second vertex*: the **(CSV)** algorithm is similar to the SSV algorithms but using in addition the track parameters, in this way the maximum possible efficiency is not limited to the fact of been able to reconstruct the secondary vertex.

More information about the performance of the b-tagging methods used during RunI period by the CMS experiment at the LHC can be found in [141].

During the preparation for the RunII period of data taking, a new generation of algorithm was developed based on the CSV algorithm. The first algorithm of this generation is the CSVv2 [143]. The main difference is that the information of tracks and impact parameters are combined using

Neural Network tools instead of a simple likelihood ratio approach, and optimized for RunII conditions at the LHC. In addition the secondary vertex is found with the Inclusive Vertex Finder algorithm[145].

In this thesis, the b-tagging is performed following the recommendations reported in [143] for the specific case of boosted topologies, where the CSVv2 algorithm is used in the subjets by the Soft Drop Mass mechanism. The expected general performance for this b-tagging algorithm for each working point are given in Table 5.1. The working point used in the analysis was the MWP.

Figure 5.9 shows the performance of tagging top jets using the CSVv2 algorithm on three different approaches: b-tagging the AK8 jets, b-tagging the AK4 jets matched to the Ak8 jets, and b-tagging the subjets in the Ak8 jets. The performance is estimated using QCD simulated events, and the best one is in the case of b-tagging the subjets, for jets in the p_T phase space region between 300 GeV up to 500 GeV. A considerable improvement on the selection efficiency can be noticed when applying the b-tagger in the subjets.

Working Point	$\varepsilon_b(\%)$	$\varepsilon_c(\%)$	$\varepsilon_{udsg}(\%)$
CSVv2LWP	81	37	8.9
CSVv2LMWP	63	12	0.9
CSVv2LTWP	41	2.2	0.1

Table 5.1: CSVv2 Tagger corresponding efficiency for the three working points, taking into account b jets with $p_T > 20$ GeV in simulated $t\bar{t}$ events. The numbers in this table are for illustrative purposes since the b jet identification efficiency is integrated over the p_T and η distributions of jets.

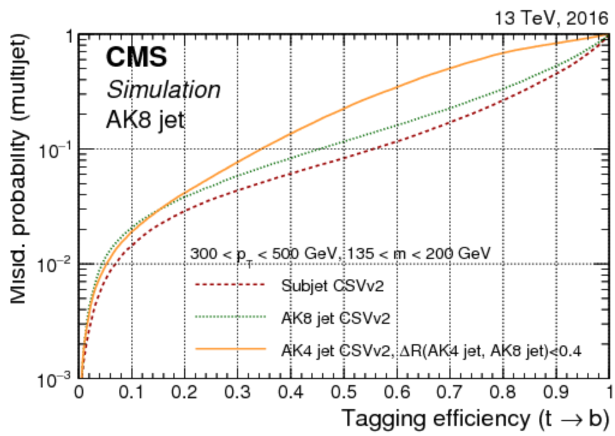


Figure 5.9: (taken from [143]) Efficiency performance of correctly identifying top jets versus misidentification probability of tagging jets in an inclusive multijet sample. The CSVv2 algorithm is applied to three different types of jets: AK8 jets, their subjets, and AK4 jets matched to AK8 jets. Two different p_t ranges are illustrated in left and right picture.

Chapter 6

Monte Carlo event simulation and event selection

In this chapter, the first steps of the analysis presented in this thesis are explained. Monte Carlo event simulations of signal and background processes allow us to control the kinematics selection, and furthermore they are used to estimate the systematic uncertainties on the final measurements. An additional (not less important) concern of any data analysis is related to the event selection. The data collected by the CMS experiment (or any other experiment) contain a large amount of events coming from different processes. Therefore, selection criteria need to be defined in order to enhance the signal contribution in the large data sample.

In this analysis, the signal events are those originating from $t\bar{t}$ pairs, decaying in the all-hadronic channel. The phase space which is of interest is the one where each of the t-quark has high transverse momentum (p_T), such that, the decay products of each of them can be clustered within a single large cone size jet. Therefore, the default (naive) selection would be to require two jets with high p_T.

This chapter is divided into five main blocks. First, in section 6.1, the Monte Carlo simulation of different sources of events, relevant for this analysis, are presented. In section 6.2, the used data sample is briefly described, including the preliminary online trigger selection. In section 6.3, the selection procedure is discussed in detail. After the event selection is applied, a reduced data sample is obtained, where the signal events have an important contribution. Control plots comparing the data distributions to the Monte Carlo predictions are provided in section 6.4, for those differential distributions that are measured. Finally, some corrections considered in the Monte Carlo predictions at detector level are presented in the last section of this chapter (6.5).

6.1 Monte Carlo event samples

Events are simulated at the level in which the collision occurs (generator level) using different Monte Carlo event generators. Each of the simulation for different processes are presented in this section. To obtain simulated events at detector level, the full CMS simulation performed with the GEANT4 [122] package is used.

Monte Carlo event simulation of $t\bar{t}$ signal events

The signal events are simulated using the POWHEG BOX v2 [148, 149, 150, 46, 151, 152], where the hard process is considered up to next to leading order (NLO) in the QCD matrix element expansion. The PYTHIA8 [34] event generator is used for simulating the underlying events (UE). This is the default simulation for signal events and, herein, the combination of those generators will be identified as the POWHEG+PYTHIA8 simulation. Variations of the default parameters are considered through the thesis, for instance, in order to estimate the modeling systematic uncertainties. A summary of all the simulated samples is provided in Table 6.1, with the most relevant parameters for each sample, and highlighting those that are changed with respect to the default set of parameters. The considered variations are:

- ✓ variations of the parameters corresponding to the CUETP8M2T4 UE tune [37].
- ✓ variations on the t-quark mass parameter (± 1 GeV)
- ✓ variations on the Colour Reconnection (CR) UE tune.
- ✓ variations on the `hdamp` parameter, for matching POWHEG and PYTHIA8.

Additionally to samples generated with POWHEG+PYTHIA8, two others are used:

- ✓ $t\bar{t}$ events are simulated at NLO in the perturbative QCD expansion with the POWHEG V2 box, as in the default sample, but for simulating the underlying event, the HERWIG++ [43] event generator is used (instead of PYTHIA8), with the EE5C Tune [37]. This sample is labeled as POWHEG+HERWIGpp.
- ✓ $t\bar{t}$ events are simulated at NLO in the perturbative QCD expansion with MADGRAPH5AMC@NLO [153] and PYTHIA8 for the UE modeling, labeled as MADGRAPH+PYTHIA8.

The total cross section obtained in each sample is normalized to the cross section predicted at NNLO accuracy in the QCD matrix element expansion [154, 155] ($\sigma_T = 831.76$ pb). This estimated value is the most precise available calculation for the signal process.

Monte Carlo event samples for QCD multijet events

QCD multijet events are simulated at Leading Order (LO) with Madgraph5aMC@NLO [153], considering the MLM [156] matching algorithm. The UE are simulated with PYTHIA8 [34], using the CUETP8M1 [37] tune. Processes including up to four outgoing partons are considered in the computation for the Born process: ($2 \to 2$, $2 \to 3$, $2 \to 4$). Samples are generated in different intervals of the scalar sum of the jet transverse momentum of the process: H_T. All the individual samples are combined considering the cross section in each H_T bin.
Alternatively, QCD multijet events are simulated with the PYTHIA8 [34] event generator, also at LO in the perturbative expansion of the QCD calculations. In the hard processes only $2 \to 2$ scatterings are considered, while additional partons are obtained in the Parton Shower. The samples are divided in bins of the minimum p_T in the Born process ($\widehat{p_T}$). All the generated samples for QCD multijet events are listed in Table 6.2, with their respective cross section, for each considered bin (H_T or $\widehat{p_T}$ bins).

Monte Carlo event samples for W+jets events.
Throughout the analysis, the contribution from W+jets[1] processes are taken into account. In order to estimate this contribution, events are simulated at LO with Madgraph5aMC@NLO [153] considering the MLM [156] matching algorithm. The UE are simulated with PYTHIA8, using the CUETP8M1 [37] tune.

6.2 Data samples

The data used for this analysis were recorded by the CMS detector, corresponding to the 2016 data taking period. The maximum instantaneous luminosity of the accelerator was 15.30 Hz/nb . The integrated luminosity during the whole period delivered by the accelerator was 41.07 fb^{-1}, while the CMS detector recorded 37.82 fb^{-1}. The total integrated luminosity of the used data is 35.91 fb^{-1}.

During this period, the mean number of interactions per bunch crossing (pileup) varied between 10 and 50, with an average of 20 pileup events. Figure 6.1 shows this behavior of the pileup for different run periods, and for all the data used in this analysis. The high-level triggers (HLT) considered to record the data are further described in 6.2.1. The data samples considered in this thesis, and their respective integrated luminosity are listed in Table 6.3.

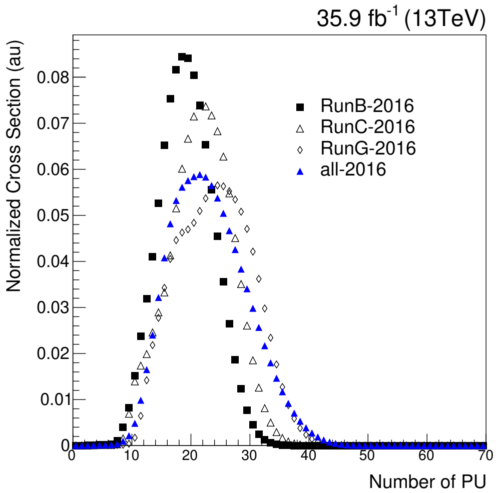


Figure 6.1: Pileup distributions measured in data during different runs periods. The blue points represent the whole data recorded during the year 2016.

6.2.1 Online trigger selection

The trigger system and its functionality was discussed in the Chapter 4, (section 4.2.5). The first online selection is performed at the first-level trigger (L1), by requiring at least one jet with transverse momentum larger than 180 GeV. At the high-level trigger (HLT) the jets are then reconstructed from the Particle Flow (PF) candidates using the anti-k_T jet algorithm with radius $R = 0.8$. At this level, different triggers are used, considering two main regions:

- ✓ **Signal region**: since one of the event selection criterion is related to the CSVv2 [143] tagger, an online trigger in which the b-tagging is already implemented has advantages. The trigger used for the event selection is a dijet trigger:[1]. This HLT trigger requires a jet with $p_T > 280$

[1] events where a W boson is produced in addition to at least one jet. In this case, up to two additional jets are considered

[1] $HLT\_AK8DiPFJet280\_200\_TrimMass30\_BTagCSV$

GeV and at least one additional jet with with $p_T > 200$ GeV. Additionally the trimming mass [2] of both jets should be larger than 30 GeV, and at least one of them should be b-tagged by the CSV tagger. This trigger is not prescaled, meaning that all the events satisfying the criteria are stored.

✓ **Control region**: in the background subtraction procedure, a control region is defined by requiring that none of the two leading jets is b-tagged. Therefore the previous trigger is not applicable. For this region, two HLT triggers are considered, depending on the p_T of the leading jet:

  ✓ If 400 GeV$< p_T^{lead} <$ 550 GeV, the $HLT\_AK8PFJet320$ trigger is used. This trigger selects events with at least one jet of radius $R = 0.8$ with $p_T > 320$ GeV. This trigger is prescaled, meaning that only a fraction of events are stored.

  ✓ If $p_T^{lead} > 550$ GeV, the $HLT\_AK8PFJet450$ trigger is used. This trigger requires at least one jet of radius $R = 0.8$ with $p_T > 450$ GeV. This trigger is not prescaled.

All the used triggers are fully efficient in the phase space region in which they are considered. The turn-on points[3] for the triggers corresponding to the control region are 362 GeV and 492 GeV for $HLT\_AK8PFJet320$ and $HLT\_AK8PFJet450$ respectively [158]. For the trigger used in the signal region, the efficiency has been estimated with the trigger emulation method, used in previous measurements by the CMS collaboration (i.e [159]). This method considers a reference trigger, with a looser selection requirement than the trigger whose efficiency is determined. Additionally, the reference trigger need to be fully efficient in the phase space region of interest. The efficiency is estimated then by the following expression:

$$\epsilon = \frac{N(trigg_{ref}, L1(X), HLT(X))_{p_T}}{N(trigg_{ref})_{p_T}} \tag{6.1}$$

where N refers to the number of events satisfying the required conditions. The numerator counts for those events selected by the reference trigger, by the $L1$ trigger and by the HLT trigger whose efficiency is being estimated. The denominator accounts for the number of the events that are selected by the reference trigger. The efficiency is estimated as a function of p_T of the leading jet in the event. The used reference trigger in this case was the trigger $HLT\_AK8PFJet140$[4]. Figure 6.2 shows the efficiency curve, and the blue line limits the phase space in which the trigger is used.

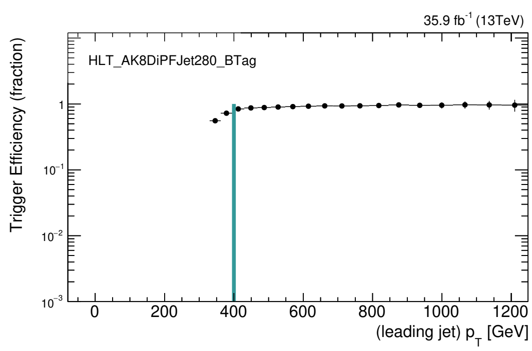


Figure 6.2: Trigger efficiency as a function of the leading jet p_T for the signal trigger.

[2] jet mass after applying the Trimming grooming technique
[3] p_T value where the trigger efficiency reaches 99%.
[4] This trigger requires at least one jet of radius $R = 0.8$ with $p_T > 140$ GeV.

Table 6.1: Summary of the simulated event samples used to model signal $t\bar{t}$ events. The simulations corresponds to pp collisions at $\sqrt{s} = 13$TeV and integrated luminosity of $\sim 36\text{fb}^{-1}$. For all the samples, the POWHEG V2 BOX was used in order to estimate the matrix elements (ME) of the hard process, at NLO precision. The renormalization and factorization scales are chosen as $\mu_R = \mu_F = \sqrt{p_t^2 + m_{top}^2}$, where m_{top} is the mass of the top quark (172.5 GeV). This parameters are changed in some of the samples listed bellow. The Parton Distribution Functions (PDF) in the ME and Parton Showers correspond to the NNPDF3.0 NLO set. The two first samples are the default sample, while the second sample is used in the background subtraction procedure. The other samples are used in the estimation of the systematic modeling uncertainties.

Generator UE Tune	Powhegv2+Pythia8 CUETP8M2T4	Powhegv2+Pythia8 CUETP8M1
gen.par.	`hdamp` $1.581 * m_{top}$ LowISR α_S 0.1108	`hdamp` m_{top} LowISR α_S 0.1108

Generator UE Tune	Powhegv2+Pythia8 CUETP8M2T4 **(UP/DOWN)**	Powhegv2+Pythia8 CUETP8M2T4	Powhegv2+Pythia8 CUETP8M2T4
gen.par.	hdamp $1.581 \cdot m_{top}$	hdamp $1.581 \cdot m_{top}$ $m_{top} = 173.5$**GeV**	hdamp $1.581 \cdot m_{top}$ $m_{top} = 171.5$**GeV**

Generator UE Tune	Powhegv2+Pythia8 CUETP8M2T4 **QCD based** **CR tune** [1]	Powhegv2+Pythia8 CUETP8M2T4 **Gluon Move** **CR tune**[2]	Powhegv2+Pythia8 CUETP8M2T4
gen.par.	hdamp $1.581 \cdot m_{top}$	hdamp $1.581 \cdot m_{top}$	hdamp $1.581 \cdot m_{top}$ α_S**, ISR** **(up/down)** [3]

Generator UE Tune	Powhegv2+Pythia8 CUETP8M2T4	Powhegv2+Pythia8 CUETP8M2T4	Powhegv2+Pythia8 CUETP8M2T4
gen.par.	**hdamp (UP)** $2.23 \cdot m_{top}$	**hdamp (DOWN)** $0.99 \cdot m_{top}$	hdamp $1.581 \cdot m_{top}$ α_S**, FSR** **(up/down)** [4]

Table 6.2: Summary of the simulated event samples for QCD multijet events. The simulations correspond to pp collisions at $\sqrt{s} = 13$ TeV and integrated luminosity of ~ 36 fb^{-1}. The prediction are provided at LO accuracy in the QCD perturbative expansion. The Parton Distribution Function (PDF) considered is the `NNPDF3.0` [157] at LO

Monte Carlo Generator	Diagrams in the Born process	$\sigma[pb]$
`Madgraph+ Pythia8` `MLM Matching Scheme, Tune CUETP8M1`	$2 \to 2, 2 \to 3, 2 \to 4$	
$H_T \in 300 - 500$ GeV		347700
$H_T \in 500 - 700$ GeV		32100
$H_T \in 700 - 1000$ GeV		6831
$H_T \in 1 - 1.5$ TeV		1207
$H_T \in 1.5 - 2.0$ TeV		119.9
$H_T \in 2.0$TeV-7 TeV		25.24
`PYTHIA 8` `Tune CUETP8M1`	$2 \to 2$	
$\widehat{p_T} \in 300 - 470$GeV		7823
$\widehat{p_T} \in 470 - 600$GeV		648.2
$\widehat{p_T} \in 600 - 800$GeV		186.9
$\widehat{p_T} \in 800 - 1000$GeV		32.293
$\widehat{p_T} \in 1.0 - 1.4$TeV		9.4183
$\widehat{p_T} \in 1.4 - 1.8$TeV		0.84265
$\widehat{p_T} \in 1.8 - 2.4$TeV		0.12163
$\widehat{p_T} \in 2.4 - 3.2$TeV		0.00682981
$\widehat{p_T} \in 3.2$TeV-7 TeV		0.000165445

Table 6.3: List of data sample used for the $t\bar{t}$ analysis in the full hadronic channel. For each subsample, the official name on the CMS storage , the run range and the integrated luminosity are provided.

Data Sample	Run Range	Integrated Luminosity (pb^{-1})
/JetHT/Run2016B-07Aug17_ver2-v1/MINIAOD	273150-275376	5748
/JetHT/Run2016C-07Aug17-v1/MINIAOD	275656-276283	2572
/JetHT/Run2016D-07Aug17-v1/MINIAOD	276315-276811	4242
/JetHT/Run2016E-07Aug17-v1/MINIAOD	276831-277420	4021
/JetHT/Run2016F-07Aug17-v1/MINIAOD	277932-278808	3104
/JetHT/Run2016G-07Aug17-v1/MINIAOD	278820-280385	7575
/JetHT/Run2016H-07Aug17-v1/MINIAOD	281613-284044	8650

6.3 Selection criteria.

In this section, the selection of the events at detector level is explained in detail. With the selection procedure we aim to distinguish from the data those events that most likely are originating from $t\bar{t}$ pairs, and satisfying the condition that the decay products of each of the t-quarks are within each of the two leading jets, respectively. The main goal is to measure the cross section of the $t\bar{t}$ pair production, differential with respect to the p_T of the leading and subleading jet, as well as, with respect to the azimuthal separation $\Delta\phi$ between them.

When both t-quarks, in a $t\bar{t}$ pair, have large p_T, they will be most likely appear opposite in direction in the transverse plane and balancing each other ($p_T^{t\bar{t}} \sim 0$). In such scenarios, the jets that are most likely associated to each of the t-quarks are the two leading jets. Additionally, in order to satisfy the condition previously mentioned, the decay products (W-boson and b-quark) need to be boosted enough within each of the leading jets. In the phase space where the t-quarks have $p_T > 400$ GeV, the decay products can be often clustered within a jet of radius $R = 0.8$ (using the anti-k_T jet clustering algorithm). The jets are named as *fat jets*, herein. The fully hadronic channel (where each W-boson decay in quark and anti-quark pair) is considered for better p_T resolution.

Selecting events with at least two fat jets and relatively high p_T is not enough, since the selected sample will be highly contaminated by QCD multijet events. The main challenge is to suppress the overwhelming QCD background. But one needs to keep in mind that, the tighter criteria are considered, the larger amount of signal events are rejected. Therefore, the selection strategy is a compromise between rejecting as less as possible signal events while suppressing as much as possible background events.

The selection procedure described in this chapter leads to a ratio of signal over background events of about $f = 0.5$. Therefore additional criteria are needed, in order to achieve a better signal to background ratios. In the next chapter (Chapter 7), an additional criterion is discussed, based on multivariate analysis techniques.

The selection criteria considered in this chapter has a twofold purpose. The first purpose is to discriminate the overwhelming background, while the second purpose is to identify those signal events, in which, the t-quarks are properly identified with the two top-jets candidates. They are explained in the following three subsections. First, in subsection 6.3.1, the baseline selection of this analysis is given. In the subsection 6.3.2, the additional selection criteria related to the top-jet definition, are explained. In subsection 6.3.3, the selection criteria related to the identification of b-jets, which highly suppress the QCD background, are presented.

A summary of the selection strategy is then provided in the subsection 6.3.4, where the number of events after each selection steps, for different Monte Carlo samples and in the data sample, are given.

6.3.1 Boosted dijet selection

An event is selected if the two leading *fat jets* satisfy the following conditions:

✓ $p_T > 400$ GeV

✓ $|\eta| < 2.4$

✓ soft drop jet mass (m_j^{SD}) larger than 50 GeV

Additionally, events with at least one lepton, either muon or electron, are removed in order to select events in the fully hadronic top decay channel. These selection criteria are referred as the *boosted dijet selection*. The required large transverse momentum for the two leading jets ensures, if they are properly identified with t-quarks, that the boosted regime (all the decay products are within the fat jet) is reached. By applying the selection criterion related to the soft drop jet mass, a considerable amount of QCD jets are removed, since those jets are most likely having lower soft drop jet masses. However, this criterion is not enough to suppress this background, but it is rather a starting point which already removes a considerable amount of background jets.

By applying the boosted dijet selection criteria, the background and signal events ratio is of the order of ~ 70. Figure 6.3 shows the number of events selected, after the boosted dijet criteria, as a function of the soft drop jet mass for the two leading jets. The data distribution is compared to the Monte Carlo predictions given by the sum of three different sources of events.
The Monte Carlo predictions for the observables shown in Figure 6.3 describe well the data dis-

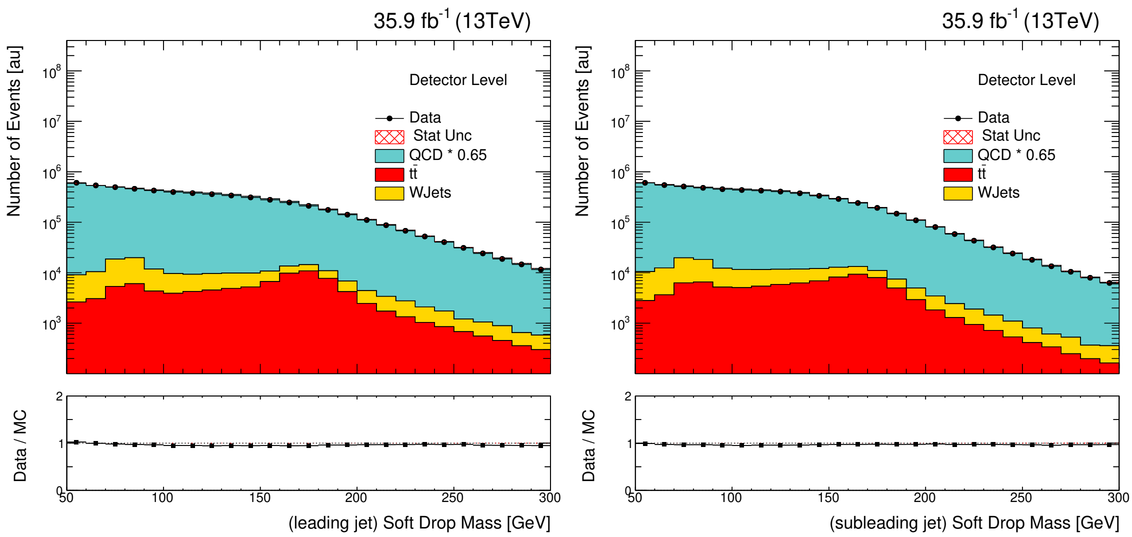


Figure 6.3: Data distributions representing the number of events as function of the soft drop jet mass for: (left) leading jet, (right) subleading jet. The data are compared to simulated signal ($t\bar{t}$) and backgrounds (QCD multijet, W+Jets) contributions. Events are selected by requiring the boosted dijet selection.

tributions. Three contributions are considered for the predictions: QCD multijet events ($\sim 97\%$), W+Jets events ($\sim 1.7\%$), and $t\bar{t}$ signal events ($\sim 1.3\%$). The QCD multijet contribution is scaled by a factor of $k = 0.65$, in order to obtain good agreement with the data.

6.3.2 Selection of events with properly defined top jets

A signal event can be imagined as in the sketch shown in Figure 6.4. Those events are originating from a $t\bar{t}$ pair, and the leading and subleading jets can be properly identified with each of the t-quarks. Furthermore all the decay products are contained within the fat jets.

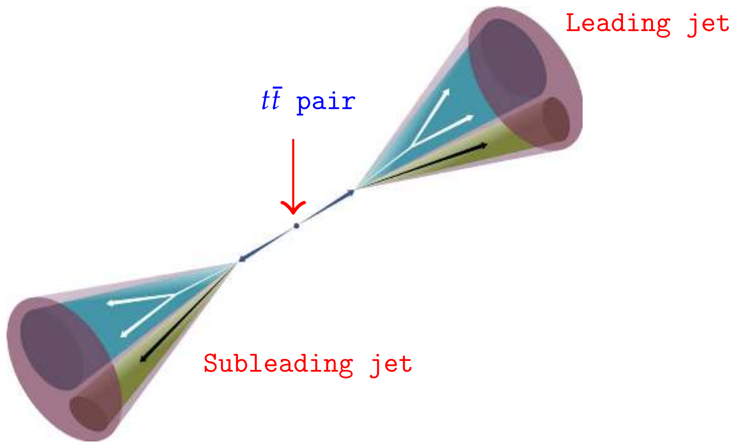


Figure 6.4: Schematic picture of an event where both leading jets are defined as signal top jets.

The condition that the two leading jets are the top jets candidates, in $t\bar{t}$ events, is not always satisfied, even in the most boosted phase space region. There are two main scenarios, which could affect the *ideal* picture of signal events represented in Figure 6.4. First, if the $t\bar{t}$ system is boosted itself (both t-quarks flight in the same direction, rather than in opposite configuration), it is rather unlikely that both leading jets are associated to the t-quarks. In those scenarios, the leading jet is the one that most likely contains the recoil QCD radiation from the $t\bar{t}$ system. Secondly, even if a jet can be associated to a t-quark, the decay products are not necessarily within one single fat jet.

Following a similar strategy as in Chapter 3, in order to properly distinguish those events containing two top-jets in $t\bar{t}$ scenarios, a new selection criteria is derived, using the detector level information. The specific studies can be found in Appendix D. This selection criteria are not aimed to discriminate QCD multijet events, but rather to remove those $t\bar{t}$ events where the two leading jets are not properly identified with each of the t-quarks.

Following the studies presented in Appendix D, an event is selected if the mass of the first subjet, in both leading jets, satisfies the following requirements:

- ✓ $m_{sub0}^{lead} > 59$ GeV (in the leading jet)
- ✓ $m_{sub0}^{sublead} > 55$ GeV (in the subleading jet).

These selection criteria are referred herein as *top-jet selection*. Figure 6.5 shows the soft drop jet mass distributions corresponding to the leading and subleading jets, in $t\bar{t}$ simulated events, first, when the boosted dijet selection is considered, and then, when the top-jet selection criteria are applied (the red filled area). As can be noticed, after this selection criteria is applied, the remaining events have two leading jets properly reconstructing the t-quark mass. For instance, the peak appearing around the W-boson mass is removed, meaning that those events in which the B-hadron is outside the fat jet, are not further considered.

The selection efficiency of the events, as function of p_T for the leading and subleading jets are shown in Figure 6.6. The efficiency is estimated with respect to the baseline selection given by the

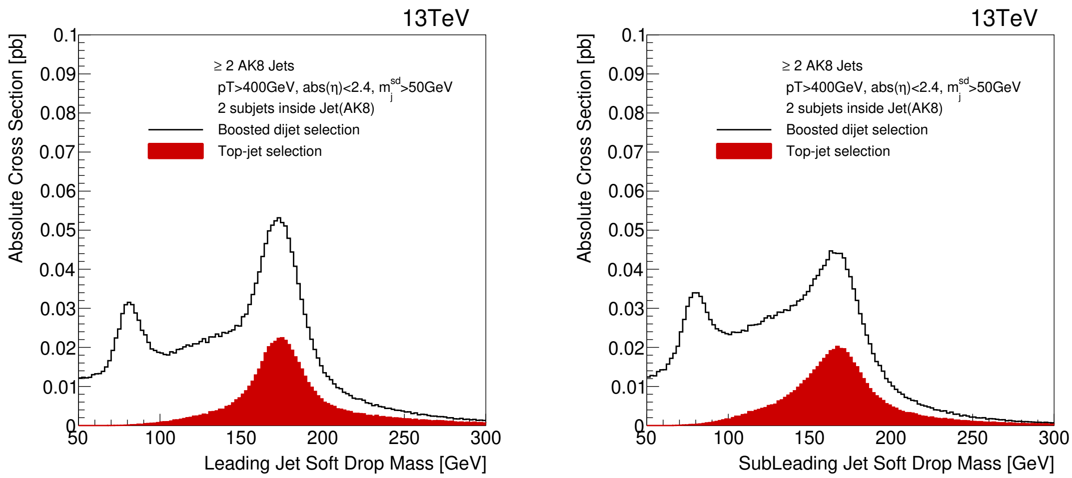


Figure 6.5: Soft Drop jet mass distributions in $t\bar{t}$ simulated events, after applying the boosted-dijet selection criteria, and after applying in addition the top-jet selection for: (left) leading and (right) subleading jets.

boosted dijet criteria. The efficiency varies from 20% up to $\sim$ 40%, from the lower p_T region to the higher p_T region. The increasing behavior is a direct consequence from the fact that at higher p_T, the boosted configuration is easier satisfied.

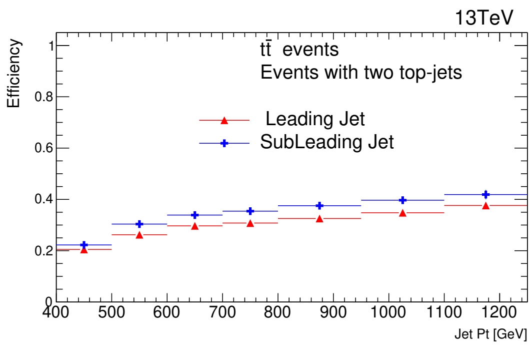


Figure 6.6: Selection efficiency as a function of p_T for leading and subleading jets, after applying the top-jet selection criteria. The efficiency is estimated with respect of the events selected by the boosted-dijet criteria.

After implementing the new selection criteria, 95.5% of the event are coming from QCD processes. The background over signal ratio is of the order of $\sim$ 31 (the QCD multijet contribution is still dominant). However, as a benefit from this step, those selected $t\bar{t}$ signal events satisfy the assumption that the two leading jets are the top jets candidates.

Figure 6.7 shows the number of events selected as function of the soft drop jet mass for both leading jets after applying the top-jet selection criteria. The Monte Carlo predictions from three sources of events are shown. The QCD multijet processes contribute about 95.5%, while the subdominant contributions coming from $t\bar{t}$ signal events and W+jets events are of the order of $\sim$ 3.7% and $\sim$ 1.5%, respectively.

The distributions shown in Figure 6.3 can be compared to the ones obtained after the top-jet selection, shown in Figure 6.7. The better reconstruction of the t-quark mass can be clearly seen.

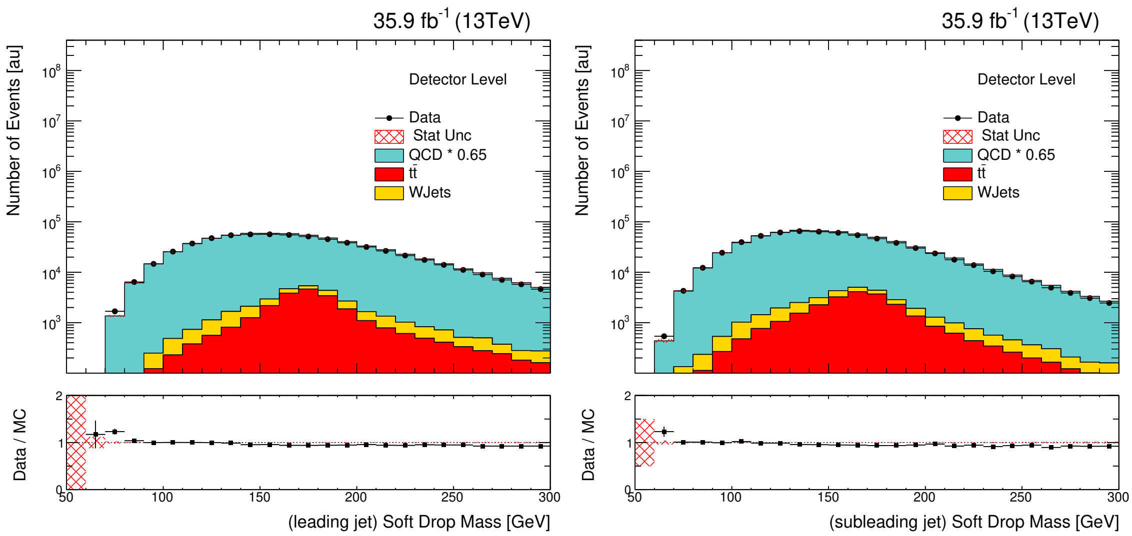


Figure 6.7: Data distributions representing the number of events as function of the soft drop jet mass for: (left) leading jet, (right) subleading jet. The data are compared to simulated signal ($t\bar{t}$) and backgrounds (QCD multijet, W+Jets) contributions. Events are selected by requiring the boosted dijet selection and the top-jet selection.

6.3.3 b-tagging selection criteria

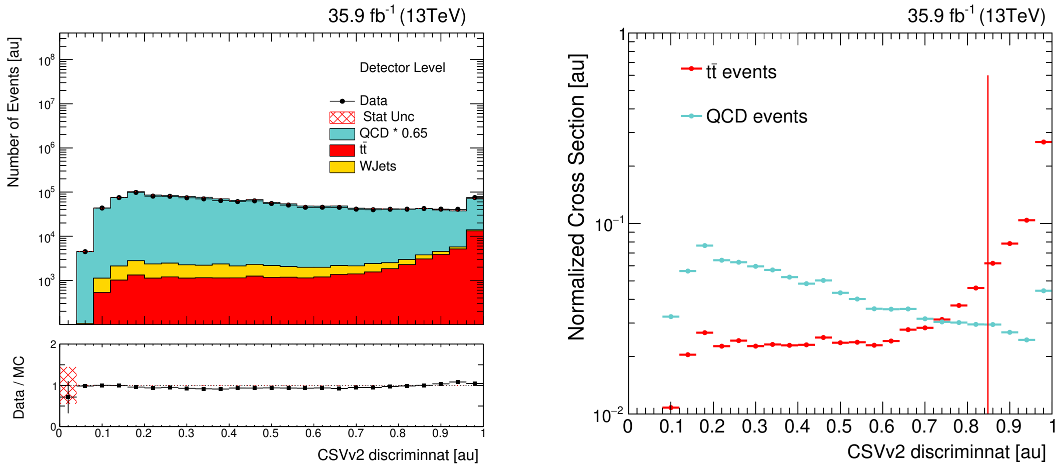
One of the potential information to be used, in order to effectively distinguish events originating from $t\bar{t}$ pairs, is the identification of one of their decays products: the B-hadron. This information can be exploited by using the b-tagging techniques provided by the CMS collaboration. The b-tagger algorithm is applied following the recommendation given by the BTV CMS group [160], for the specific case of boosted topologies. The considered algorithm is the CSVv2 [143] tagger, using the medium working point (defined as a criterion applied to the output discriminant and giving 1% of probability that a light-jet is wrongly classified as b-jet).

The tagging algorithm is applied to the subjets, instead of the fat jet. The benefits from that relies on the fact that inefficiencies are introduced by estimating the axis of the fat jet using the three prong fragmentation configuration, instead of the b-quark itself. Those inefficiencies can be partially removed by applying the algorithm to the subjets as alternative [161].

Studies of the performance of the b-tagging algorithm are provided in the Appendix E.

Events are selected if for both leading jets, at least one of the two subjets is classified by the CSVv2 tagger as a b-jet ($d_{btag} > 0.8484$). This selection is referred in the following as the *b-tagging selection criteria*. Figure 6.8a shows the data distribution representing the number of events (before appling the b-tagging criteria) as function of the b-tagging discriminant (d_{btag}). By requiring $d_{btag} > 0.8484$, the lower phase space region for this variable is removed, which is the part in which QCD multijet have the major contribution. In figure 6.8b, the normalized distribution of the b-tagging discriminant for QCD multijet events and for $t\bar{t}$ events are shown. The red line represents the selection cut. From this discrimination power of this observable can be seen.

After considering the b-tagging selection criteria, the contribution coming from W+jet events is negligible. Therefore, at this point, the only remaining background to deal with are the events originated by QCD multijet processes. The background over signal ratio is reduced up to ~ 1.9



(a) **(b)**

Figure 6.8: (left) Output discriminant of the CSVv2 in data, before the b-tagging selection criteria is applied, with the respective contributions of different sources of events, estimated by Monte Carlo simulations. (right) Normalized distribution as function of the CSVv2 tagger, in $t\bar{t}$ and QCD multijet events.

with $\sim$ 33.7% of the events coming from $t\bar{t}$ pairs. Figure 6.9 shows the soft drop mass distribution in data for both leading jets after the new selection criteria are applied, with the respective background and signal contributions estimated from the Monte Carlo simulations. The enhancement of the signal contribution is remarkable.

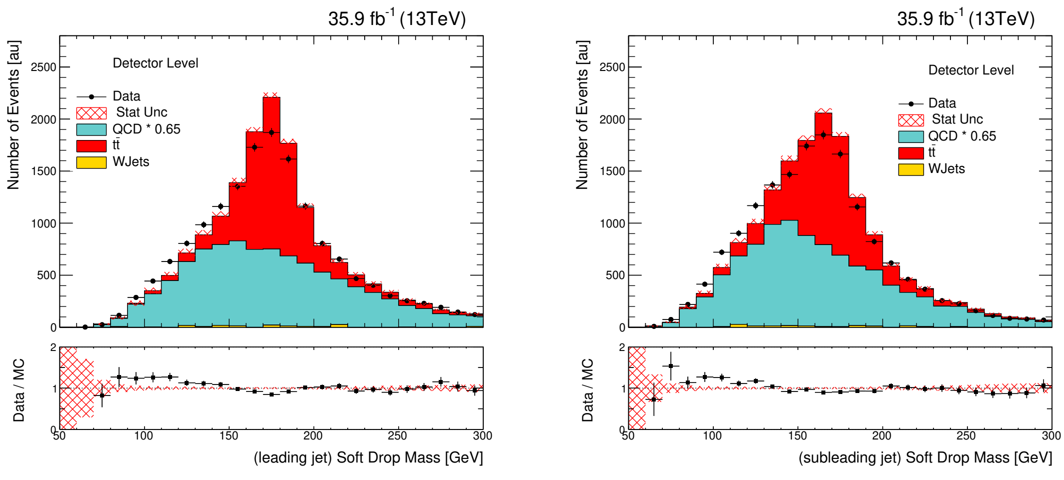


Figure 6.9: Data distributions representing the number of events as a function of the soft drop jet mass for: (left) leading jet, (right) subleading jet. The data are compared to simulated signal ($t\bar{t}$) and background (QCD multijet, W+Jets) contributions. Events are selected by requiring the boosted dijet selection, the top-jet selection, and the b-tagging selection criteria.

These criteria affect the signal selection efficiency. This can be noticed in 6.8b, where it is shown that roughly half of the signal events are removed. The main advantage is that the QCD background is strongly suppressed. Figure 6.10 shows the efficiency of these selection criteria estimated in signal events, as function of the p_T of the two leading jets. The considered baseline selection, in order to estimate the efficiency in each of the illustrated plots is different:

- ✓ in Figure 6.10a, the efficiency is estimated with respect to those events satisfying the boosted dijet and the top-jet selection criteria.
- ✓ in Figure 6.10b, the efficiency is estimated with respect to those events satisfying the boosted dijet selection criteria.

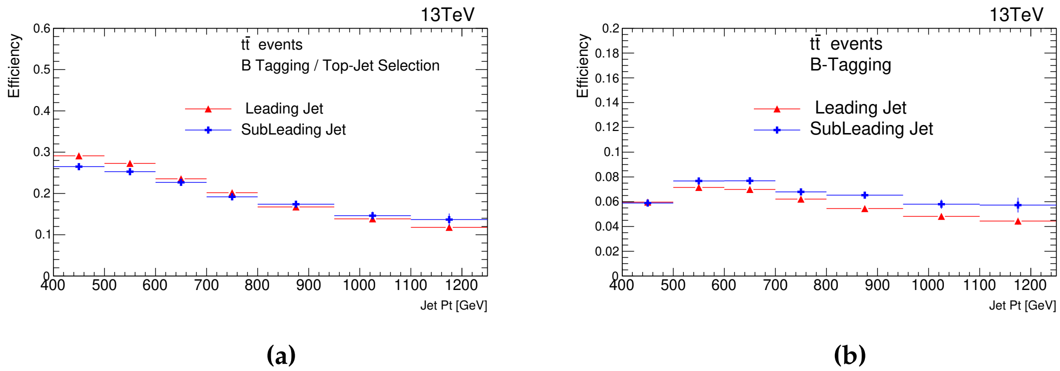


Figure 6.10: Efficiency selection as a function of p_T of the leading and subleading jets after the b-tagging selection criteria: (a) with respect of the events selected after top-jet selection (b) with respect to the boosted-dijet selection.

The first plot (6.10a) allow us to directly infer the efficiency of the b-tagging algorithm. The decreasing behavior of the efficiency of the b-tagger when increasing p_T is expected as a general performance of this algorithm. This is because higher p_T jets are more collimated, therefore the charged particles are closer to each other and quite often hits are overlapping; furthermore, the light flavor misidentification probability increases when the p_T of the jet is larger. The latter effect is illustrated in the studies presented in the Appendix E. The b-jet identification efficiency (the fraction of true b-jets, that are b-tagged by the algorithm) is approximately $\sim$ 56% (see details in the Appendix E, when the medium working point is applied). Therefore, if the criterion is applied for both leading jets, the expected efficiency would be $\sim$ 30%, which is the value obtained for the event selection efficiency, in the lower p_T region, illustrated in Figure 6.10a.

The total selection efficiency, illustrated in 6.10b, with respect to the events satisfying the boosted dijet selection criteria varies from 6% up to 8%.

6.3.4 Summary of the Event Selection

In order to select $t\bar{t}$ events the following selection criteria are applied:

- ✓ **boosted dijet selection:** the two leading fat jets are required to have $p_T > 400$ GeV, $|\eta| < 2.4$ (contained within the tracker region) and a soft drop jet mass larger than 50 GeV. Events with at least one lepton are vetoed.
- ✓ **top-jet selection:** the first subjet in the leading and subleading fat jet has a mass larger than 59 GeV, 55 GeV, respectively.
- ✓ **b-tagging selection:** at least one of the subjets, in both leading jets is b-tagged ($CSVv2$ using the MWP).

After these selection steps are applied, around $\sim$ 6.2% of signal events are selected (with respect to the boosted-dijet selection), while the background over signal ratio reaches values up to $\sim$ 1.9. In the data, 43894 events are selected after the criteria are applied.

6.4 Control Plots: Data-MC comparison at detector level

In this section, data distributions are compared to Monte Carlo predictions at detector level. The Monte Carlo predictions are the sum of all the contributions relevant in each specific phase space mainly given by three sources: QCD multijet events, $t\bar{t}$ events, and W+jets events. Specifically, distributions differential in p_T of the leading and subleading jets and in the azimuthal separation between the two leading jets ($\Delta\phi$), are presented.

The distributions are presented in two different phase space regions:

- the boosted region: the boosted dijet selection criteria are applied.
- the signal region: all selection criteria are applied (boosted dijet selection, top-jet selection, and b-tagging selection)

Figure 6.11 shows the measured distribution differential with respect to the mentioned observables. The left plots correspond to the boosted region, while the right plots to the signal region.

In the boosted region, only the QCD multijet contribution plays a role, since the others are approximately two order of magnitudes smaller. In this region, differences between the QCD multijet prediction and the data can be observed.

For instance, in the case of the p_T distribution of the leading jet, differences up to 12% are visible. In the p_T distributions for the subleading jet, differences of up to $\sim$ 20% in the higher p_T region are seen. Furthermore, in the absolute cross section as function of $\Delta\phi$, differences of $\sim$ 20% appear. A main conclusion is that the available QCD multijet predictions are not able to properly describe the main observables, and, in order of further subtracting this contributions, alternative methods need to be considered for estimating and subtracting this background contribution (i.e a data driven method).

In the signal phase space region, the differences in the p_T distributions are reduced (although also affected by the statistical uncertainties). The differences in the $\Delta\phi$ distribution, are still $\sim$ 20%.

6.5 Corrections applied to the Monte Carlo predictions at detector level

In the experimental analysis we are facing several challenges towards performing the measurements and comparing with theory predictions in a reliable way. One of these issues is that some algorithms are not performing the same in the simulation and in the measurements. Those differences have to be taken into account. This section is divided in four main blocks, in which different corrections taken into account throughout the analysis are explained.

Re-weighting of the events due to discrepancies in the pileup profile.
In the Monte Carlo simulations, the pileup profile is generated from a Poisson distribution around an expected value. The number of of pileup interaction (n_μ) can be estimated in the data by using the information of the instantaneous luminosity and the total pp collisions cross section[162]. The estimated value depends on different factors, i.e the beam conditions and the instantaneous luminosity. Therefore the simulation will rarely describe the distribution observed in the data.

Following the recommendations in [162], the re-weighting of the simulated events are applied, in order to correct the pileup profile in the Monte Carlo simulations.

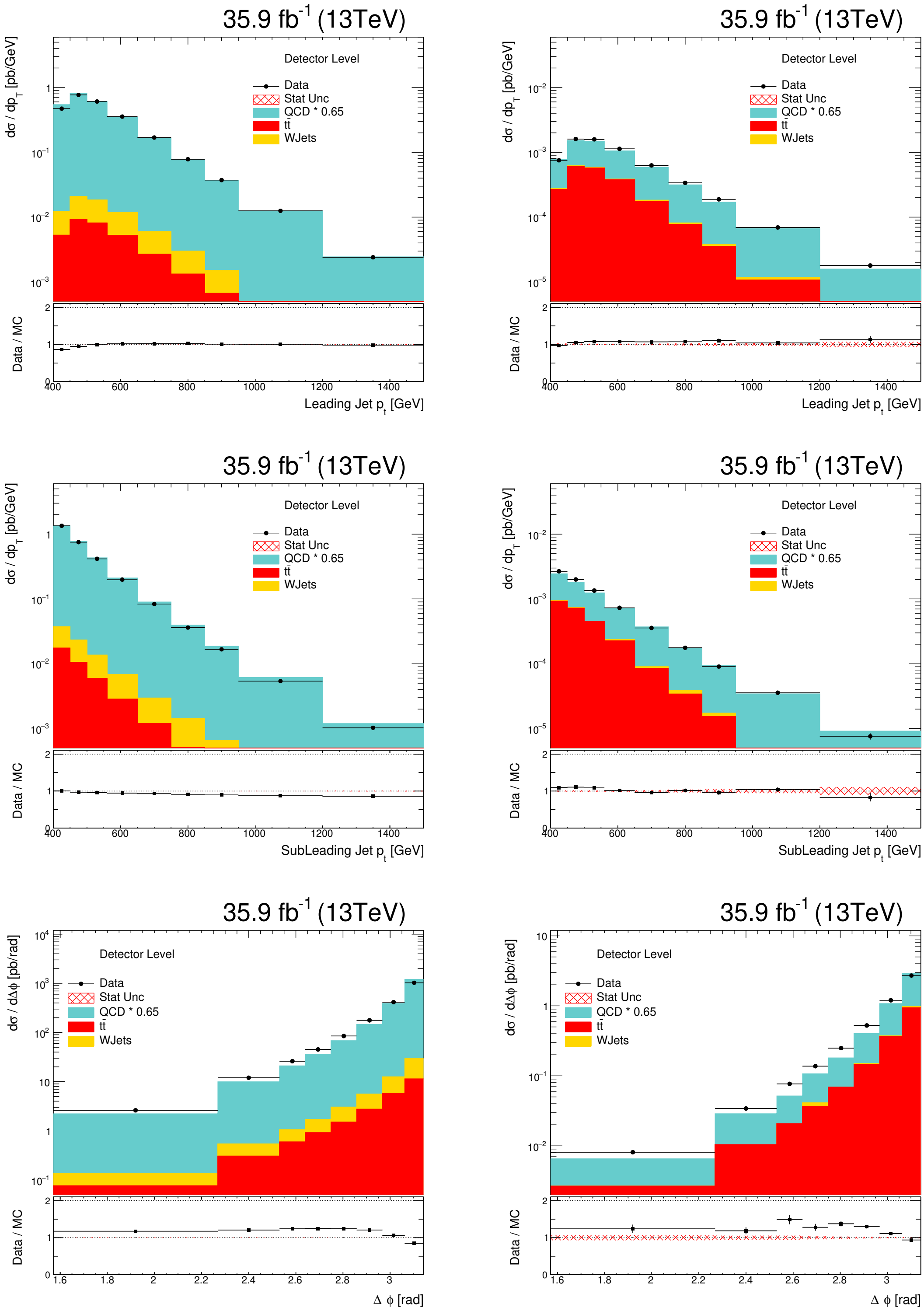


Figure 6.11: Cross section distributions differentially as function of: (upper plots) p_T of the leading jet (middle plots) p_T of the subleading jet, (lower plots) $\Delta\phi$ between the two leading jets. Two phase regions are illustrated: (left) after boosted dijet selection, (right) signal region

Figure 6.12 shows the comparison of the pileup distributions estimated in the data with respect to the one obtained in a Monte Carlo simulation. The ratio *Data*/*MC* gives the scales factors which are applied in an event-by-event re-weighting in the simulations. The second plot illustrates how the simulated profile is corrected after applying the re-weighting of the events.

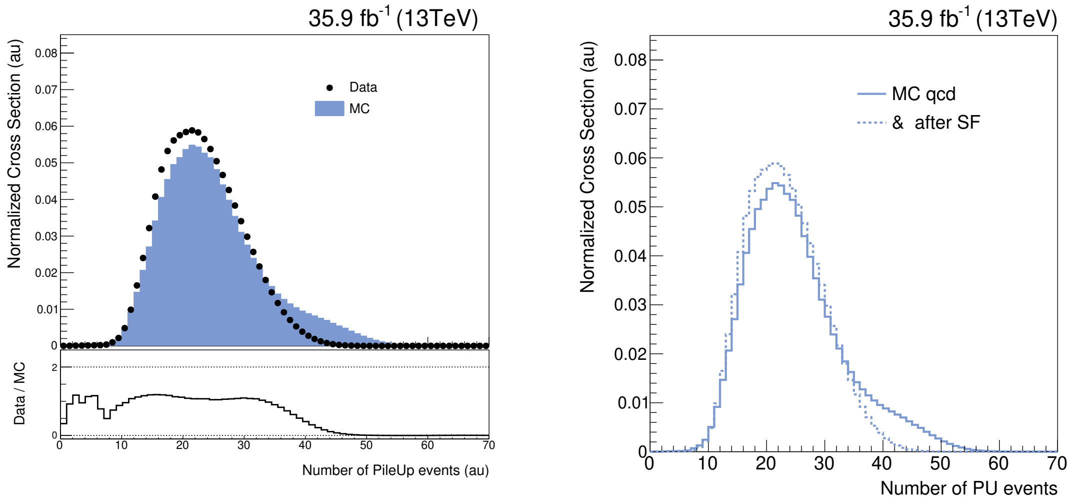


Figure 6.12: (left) Pileup distributions in a QCD Monte Carlo sample, compared to the pilup distribution in data collected during 2016. (right) Pileup distribution in Monte Carlo sample before and after applying the scale factors for re-weighting of the events.

Reweighting of the events using the b-tagging scale factors

Usually, the performance of the b-tagging algorithm behaves in somehow different in data and in the Monte Carlo simulations. In order to take those differences into account, corrections are applied, using scale factors provided by the experiment. Those are a result of dedicated measurements of the b-tagging efficiency in data [147].

The scale factors are provided for each flavour jet (i.e. b-jet, c-jet, and light-jet), and in different p_T (of the subjets) bins. They can be expressed by the following ratio:

$$SF_i = \frac{\varepsilon^{Data}(p_T)}{\varepsilon^{MC}(p_T)} \tag{6.2}$$

where i refers to the jet flavour (b, c or $light$).

These corrections are considered through reweighting of the events in the Monte Carlo samples. This method uses the provided scale factors in combination with the efficiencies in Monte Carlo samples, which need to be estimated before. The Monte Carlo efficiencies are specific to the analysis and might depend on the nature of the event (i.e. QCD multijet or $t\bar{t}$ events) and on the phase space region where the analysis is performed.

The probability that in a certain amount of b-jets, there are i-jets properly b-tagged, while j-jets non b-tagged (in Monte Carlo and in data) can be written by the following expressions:

$$P(MC) = \prod_{i=tagged} \varepsilon_i \prod_{j=nontagged} (1 - \varepsilon_j) \tag{6.3}$$

$$P(Data) = \prod_{i=tagged} SF_i \varepsilon_i \prod_{j=nontagged} (1 - SF_j \varepsilon_j) \tag{6.4}$$

where ε_i are the MC b-tagging efficiencies and SF_i are the standard scale factors.

The event weights are estimated as the ratio between both probabilities as follows:

$$w = \frac{P(Data)}{P(MC)} = \prod_{i=tagged} SF_i \prod_{j=nontagged} (1 - SF_j \varepsilon_j)/(1 - \varepsilon_j) \tag{6.5}$$

In order to compute the MC efficiencies used in equation 6.5, simulated QCD multijet and $t\bar{t}$ events are considered. Events are selected by requiring at least two fat jets with $p_T > 400$ GeV, $|\eta| < 2.4$, and soft drop mass larger than 50 GeV. The MC efficiencies are estimated for each flavour, in p_T bins of the b-tagged subjets by the following expression:

$$\varepsilon_f(p_T) = \frac{N(f, btagged)}{N(f, total)} \tag{6.6}$$

where f stands for the flavour jets (c,b or light jets), $N(f, total)$ are the total number of jets of a certain flavour f, and $N(f, btagged)$ are the total number of those jets that are b-tagged. Therefore ε_b corresponds to the b-tagging efficiency (percentage of truth b-jets that are b-tagged), while ε_c and ε_{light} correspond to the contamination (the probability that non-b-jets are b-tagged). The jet flavour is defined using the true Monte Carlo information with Jet Flavour Identification [163].

The b-tagging efficiencies for each jet flavour (b-jets, c-jets and light jets) as function of p_T are illustrated in Figure 6.13. The CSVv2 tagger with the medium working point is considered. As might be noticed, the b-tagging efficiency decreases as function of p_T, while an increase rate of the light-jet misidentification is observed.

Corrections due to the jet energy resolution (JER)

It was previously discussed that due to detector resolution effects the p_T of the jet is not necessarily the same at generated and reconstructed levels. The relative resolution can be estimated in Monte Carlo samples as:

$$\Delta R^{MC}_{p_t} = \frac{p_T^{reco} - p_T^{gen}}{p_T^{gen}} \tag{6.7}$$

The estimated resolution in Monte Carlo, given by equation 6.7, is different from the one measured in data. The differences between both are considered by changing (*"smearing"*) the p_T of the jet at reconstructed level in the Monte Carlo simulations.
First of all, a matching criteria is required for the jets at reconstructed and generator levels:

- $\Delta R_{jj} < R_{cone}/2$, where ΔR_{jj} refers to the distance ($\sqrt{\Delta\phi^2 + \Delta\eta^2}$) between the two jets at both levels, and R_{cone} is the R parameter of jet clustering algorithm ($R = 0.8$).
- $\Delta R^{MC}_{p_t} < 3\sigma_{JER}$, where σ_{JER} is the measured resolution in data (provided by the experiment).

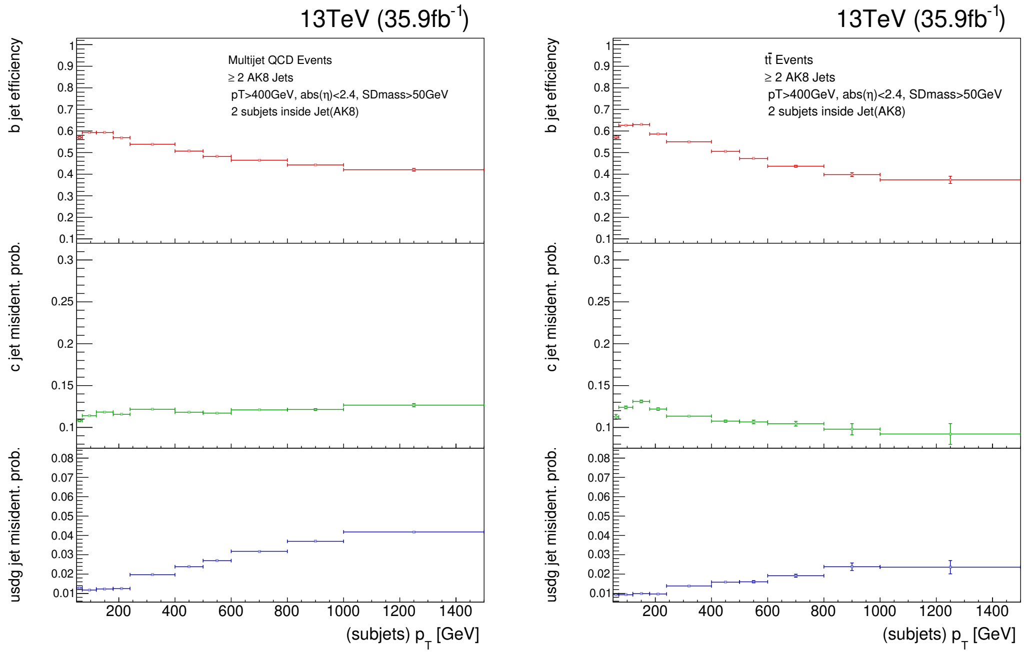


Figure 6.13: Monte Carlo b-tagging efficiencies as a function of p_T of the subjets, for each jet flavour: *b-jets*, *c-jets*, *light-jets*: for QCD events (left) and $t\bar{t}$ events (right). The events are selected by applying the dijet boosted selection criteria.

The resolution which is obtained from the Monte Carlo ($\Delta R^{MC}_{p_T}$) has to be corrected using the provided scale factors (SF_{res}). The resolution in data can be expressed as function of the scale factor as follows:

$$\Delta R^{data} = SF_{res}\Delta R^{MC}_{p_T} \tag{6.8}$$

These scale factors (SF_{res}) depend on the η observable. The reconstructed p_t can be expressed as:

$$p_t^{reco} = p_T^{reco}(1 + \Delta R^{MC}_{p_T}) \tag{6.9}$$

Finally, the reconstructed p_T smeared is given by the following formula then:

$$p_T^{reco} = p_T^{reco}(1 + SF_{res}\Delta R^{MC}_{p_T}) \tag{6.10}$$

Reweighting of the $t\bar{t}$ events depending on the p_T of the t-quark

The correction treated in this subsection is not directly related to inefficiencies in the detector simulation, but rather to differences observed in the measurements with respect to the Monte Carlo predictions, of $t\bar{t}$ events, at LO and NLO. The CMS recommendations [164] to take into account those differences are implemented. The method consists of an event-by-event reweighting considering the following factor:

$$\omega_t = \sqrt{SF(t, p_t)SF(\bar{t}, p_t)} \quad \text{where } SF(p_t) = e^{0.0615 - 0.0005 p_T} \tag{6.11}$$

where p_T refers to the transverse momentum of the parton-level top quark (after radiation and before its decay) in the PYTHIA8 Monte Carlo simulation. The p_T top re-weighting of the events is applied when the detector level information for the Monte Carlo predictions is used, but not for providing the Monte Carlo predictions at particle level.

Chapter 7

Identification of the $t\bar{t}$ boosted system with a multivariate approach.

In this Chapter the additional selection criteria needed to have a signal over background ratio larger than one, is implemented.

The implemented technique to achieve the mentioned goal is based on Multivariate methods. Multivariate techniques exploit correlation among the input variables, in order to produce a discriminating output to which a selection can be applied for distinguishing signal events more effectively than when a rectangular cut[1] is applied to the input observables.

The tool used in this thesis, in order to perform the Multivariate Analysis (MVA), is the TMVA method [165]. Several algorithms are tried out using different classifiers:

- ✓ the Likelihood estimator,
- ✓ the Fisher estimator,
- ✓ the Likelihood estimator with decorrelation of input variables,
- ✓ the boosted Fisher estimator,
- ✓ Multilayer perceptron (MLP),
- ✓ Boosted Decision Trees (BDT),

The technicalities of how those methods are implemented can be found in Ref. [165], while more theoretical explanation of the methods and their applications to LHC physics, is given in Ref. [166].

The MVA methods are based on a classification problem, where the first step is the training procedure. For the training procedure input variables, which are sensitive to signal events, need to be provided. Each of the methods evaluates in different ways the training, learning from signal or background simulated events. After the training is performed, other samples of simulated events are tested, by classifying the event either as signal or background.

[1] Rectangular cut is referred by a subsequent cuts on the input observables. For instance, having two observables x and y, a rectangular selection would be the region defined by Δx-Δy.

The input variables to the MVA must satisfy certain criteria. First the data distributions need to be well understood, by the Monte Carlo simulation of background and signal processes. Secondly, they need to be observables sensitive to the signal topology which is being studied, with already good discrimination power between signal and background events. Several input variables are tried out, finding that the ones with higher power of discrimination are the observables related to the multiprong configuration: the N-subjettiness variables. Specifically the ratios τ_3/τ_1 and τ_3/τ_2, for the leading and subleading jets are selected.

The MLP and BDT methods are found to be the ones performing best for suppressing QCD multijet background events. Figure 7.1 shows the comparison among all the implemented methods, representing the so-called Receiver-Operator-Characteristic (ROC) curves. These curves give the relation between signal and background rejection efficiencies. From the figure one can notice an improvement of the Neural Network (BDT, MLP) with respect to the simplest methods based for instance on the Likelihood and the Fisher estimators. Details on how the ROC curves are obtained, i.e training procedure and simulated samples, are provided in section 7.1.

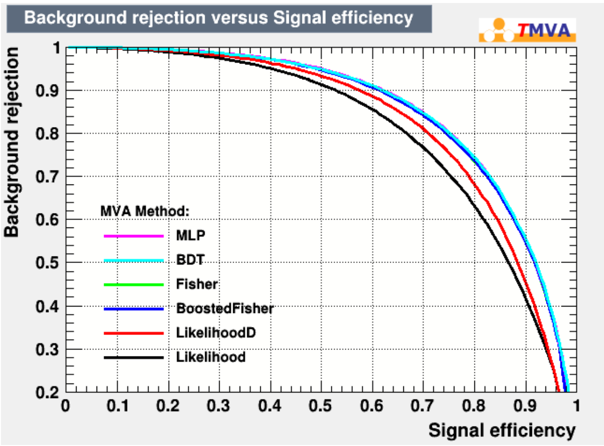


Figure 7.1: Performance of different multivariate discriminants represented by the so-called Receiver-Operator-Characteristic (ROC) curves

The method for further considering the selection of the events was the MLP. The Neural Network considered in the algorithm is represented in Figure 7.2: the network considers one internal layer with 10 nodes.

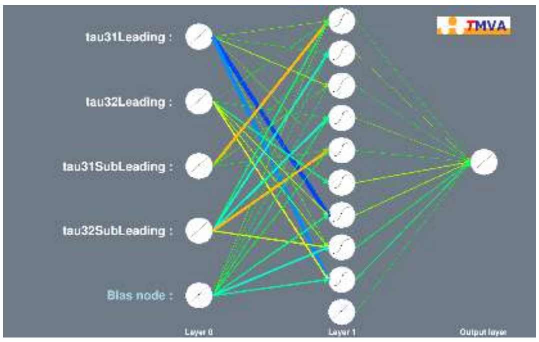


Figure 7.2: Neural network architecture for the Multilayer perceptron (MLP) algorithm. This Neural network is the one used to perform the training and testing procedure and provide the output discriminant for the event selection. τ_3/τ_2, τ_3/τ_1 for the leading and subleading jets are the input variables considered for the MLP architecture.

This chapter is divided into three main parts. In the section 7.1, the training and testing results from the MLP multivariate technique are presented. In section 7.2, the selection criterion derived from the output discriminant of the MVA is discussed. Finally, in section 7.3, the selection strategy completed by the MVA selection criterion is presented.

7.1 Multivariate training and results

The training of the algorithm is performed with simulated $t\bar{t}$ signal events and QCD multijet background events. The former are obtained using the POWHEG+PYTHIA8 sample, while the latter one is obtained using the MADGRAPH+PYTHIA8 QCD sample. Those simulations are further considered in this chapter when referring to signal or background events.

The training is performed on pre-selected events. In Chapter 6, a set of selection criteria was provided. However, after they are applied, a considerable amount of QCD multijet contribution is removed. In order to have enough statistics for the background input information in the training procedure, and given the fact that the N-subjettiness variables behave in a similar way for different b-tagged categories, the training is performed only considering the boosted dijet selection and the top-jet selection (without considering the b-tagging).

Figure 7.3 shows the data distributions of the N-subjettiness observables: τ_1, τ_2 , τ_3, τ_3/τ_2, τ_3/τ_1, for the leading and subleading jets. Those are the input information to the MVA algorithm. The data is better described for the ratios τ_3/τ_2, τ_3/τ_1, than when the other observables are considered. Therefore, the set of four variables given by the ratios for the leading and subleading jets are selected as the input information for the algorithm.

Figure 7.4 shows the comparison of the shape of the distributions, between signal and background events, where one can notice the discrimination power of those observables. Furthermore, the MPL technique exploits additionally the correlation between the input observables. Figure 7.6 shows the correlation among all the observables, for signal and background events. The correlation is behaving with the same trend for both contributions: the leading and subleading observables are completely uncorrelated, while the τ_3/τ_2 and τ_3/τ_1 shows a certain correlation. However, the absolute value of this correlation is quite different for signal ($\sim$ 68%) and background ($\sim$ 36%) contributions. This fact is taken as an advantage for the better performance of the algorithm.

Finally, Figure 7.7 shows the output discriminant for background and signal events of the MLP algorithm (MLP response), further recognized as d_{MVA}. In the figure, the training and testing results are shown, to verify the reliability of the training procedure. Furthermore, a very clear separation power is seen: the d_{MVA} for background events takes preferable lower values, while a different behavior can be distinguished for signal events, being more spread in the d_{MVA} values. This discrimination power is further used to define a selection cut.

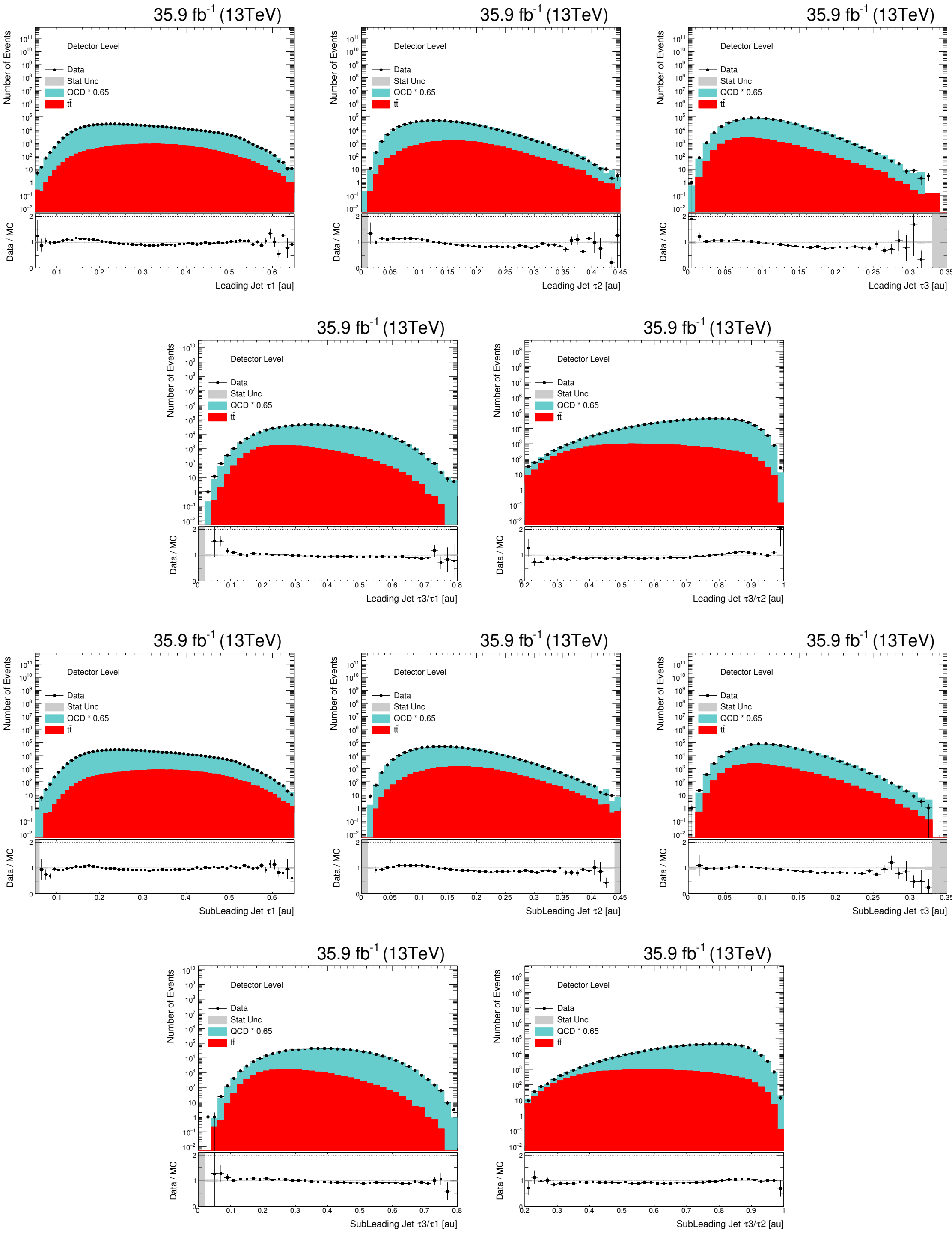


Figure 7.3: Data distributions representing the number of events as a function of the N-subjettiness variables for: (first and second rows) leading jet, (third and forth rows) subleading jets. In the first and third rows τ_1, τ_2 and τ_3 are represented, while in the second and forth rows the τ_3/τ_1, τ_3/τ_2 ratios are represented. The data distributions are compared to simulated signal ($t\bar{t}$) and background (QCD multijet) contributions. Events are selected by requiring the boosted dijet selection and the top-jet selection.

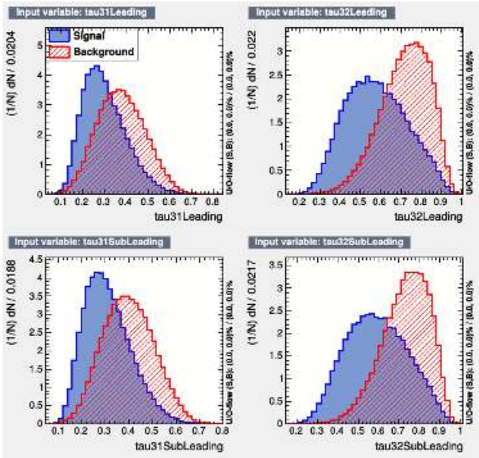


Figure 7.4: Input variables to the MLP algorithm, comparison between the shapes of the observables. The symbols *tau*31 and *tau*32 mean the ratio τ_3/τ_1 and τ_3/τ_2 respectively.

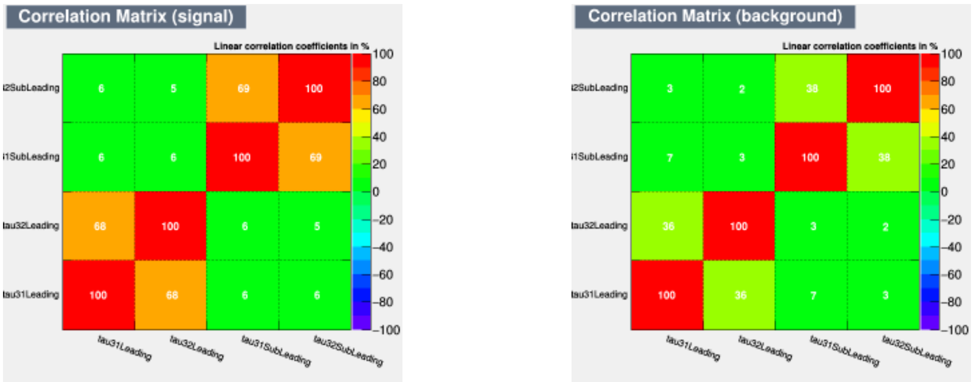


Figure 7.5: Linear correlation of the four substructures variables provided as input information for the MVA algorithm for: (left) signal events, (right) background events. The symbols *tau*31 and *tau*32 mean the ratio τ_3/τ_1 and τ_3/τ_2 respectively.

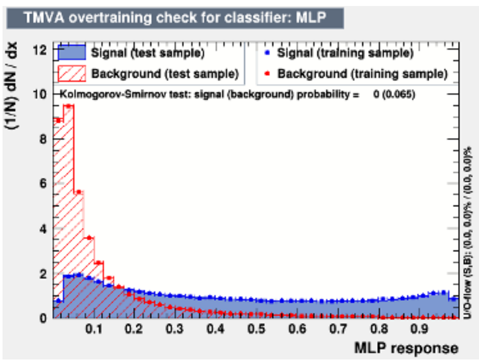
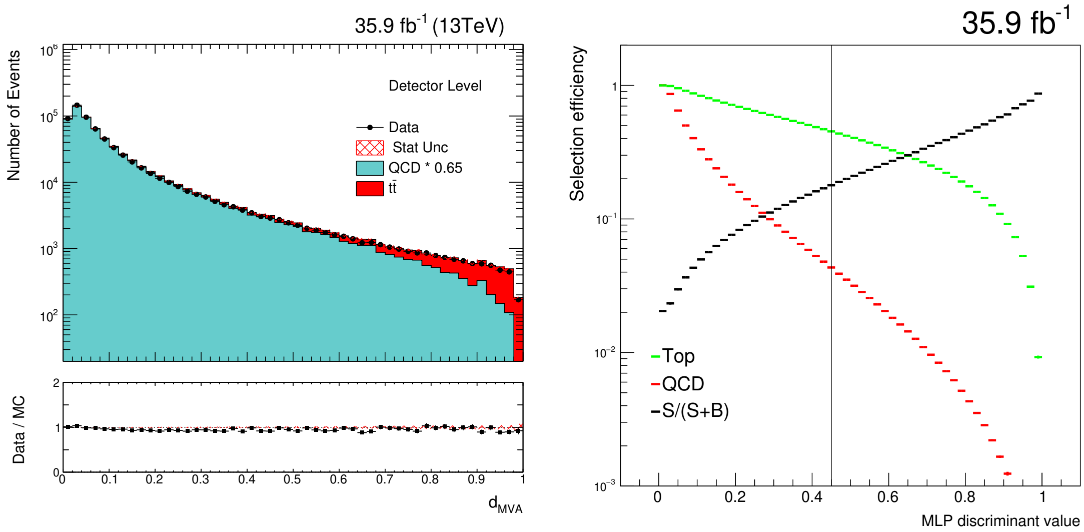


Figure 7.6: Test and training samples for background and signal events of the output of the MLP algorithm. The output discriminant for signal events is represented in blue while for background events in red.

7.2 Multivariate selection criterion

Once the multivariate technique provides the discriminant output (d_{MVA}), one needs to consider a cut on the discriminant in order to define the selection criteria. The distribution of the d_{MVA} observable in the data is illustrated in Figure 7.7a, compared to the Monte Carlo predictions provided by signal ($t\bar{t}$) and background (QCD multijet) contributions. A clear enhancement of the $t\bar{t}$ contribution towards the increasing of d_{MVA} is observed.



(a) **(b)**

Figure 7.7: (a) Data distribution representing the number of events as a function of the d_{MVA} discriminant after requiring the boosted dijet selection and the top-jet selection. The data distribution is compared to simulated signal ($t\bar{t}$) and background (QCD multijet) contributions. (b) Performance of the d_{mva} MLP discriminant for (green) signal efficiency selection (red) background efficiency selection (black) $S/(S+B)$. The efficiencies are estimated in stimulated events by requiring a cut on the discriminant: $d_{mva} > x_{cut}$.

In Figure 7.7b studies concerning the discrimination power of the d_{mva} observable are provided. Those studies help us to define a selection criterion on the d_{mva} discriminant. Three different curves are shown as function of the d_{mva}[1]: (green) the signal selection efficiency, (red) the background selection efficiency (which is the fraction of the background events remaining after considering the specific cut), (black) the ratio given by the following formula:

$$r = \frac{S}{S+B}, \tag{7.1}$$

where S is the signal contribution and B is the background contribution. For instance, applying a certain cut $d_{mva} > x_{cut}$, where x_{cut} takes larger values, the ratio r is increased, and hence the fraction of signal events in the selected sample is larger. However, toward increasing x_{cut}, the fraction of signal events decreases. Therefore a compromise might be assumed between rejecting as much as possible background events while not affecting that much the signal efficiency. The exact value of x_{cut} is arbitrary, and a value of $x_{cut} = 0.45$ has been considered. For this value, approximately 50% of the signal events are selected, while $\sim$ 96% of the background events are rejected, leading to the ratio $r \sim 0.2$ (or equivalent to say $S/B \sim 0.25$). After this selection criteria, one still needs to apply the b-tagging selection criteria, discussed in the previous chapter.

[1]As function of a cut requiring $d_{mva} > x$, where x is the values shown in the x-axis

In Figure 7.8, the data distribution of the d_{MVA} observable after applying the cut, is illustrated. The distribution is compared to the signal and background contributions obtained from the Monte Carlo predictions, where a good description from the Monte Carlo is observed.

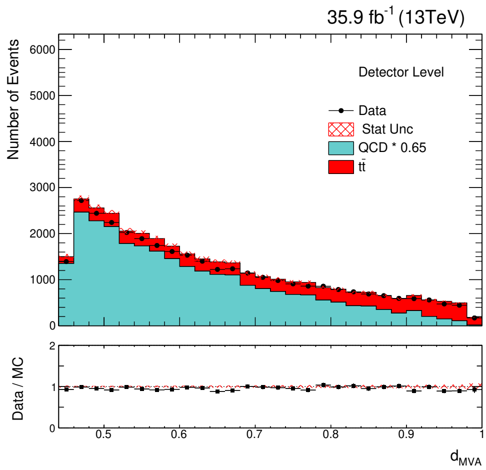


Figure 7.8: Data distribution representing the number of events as a function of the d_{MVA} discriminant after requiring the boosted dijet selection and the top-jet selection, and additionally applying the cut on the d_{MVA}: $d_{MVA} > 0.45$. The data distribution is compared to simulated signal ($t\bar{t}$) and background (QCD multijet) contributions.

In the following, studies on selection efficiency by considering either individual observables or different set of input variables for MVA method, are provided[1]. Figure 7.9 shows the relation between efficiency of correctly selecting signal events and the probability of misidentifying background events. The illustrated curves consider different observables to perform the selection cut for distinguishing signal events. Among those observables one can find all the possible N-subjettiness variables, for leading and subleading jets, and their respective ratio. Additionally, the output discriminant from two MVA techniques are taken as observable to perform the cut.

The two MVA approaches correspond to a MLP training procedure, but considering different set of variables as input information:

- ✓ Set 1: τ_1, τ_2 and τ_3, for the leading and subleading jets (in total six input variables).
- ✓ Set 2: τ_3/τ_1, τ_3/τ_2, for the leading and subleading jets (in total four variables).

As might be seen, there is almost no difference between the two MVA approaches: their performance are similar. The selection strategy further considered in this thesis, is the one corresponding to the second set. The main motivation to select the ratio of the N-subjettiness is given by the better agreement of the prediction with the data observed in Figure 7.3.

The cut on the d_{MVA} is performed when the signal efficiency is $\sim 50\%$. At this efficiency point, by performing a cut in any individual N-subjettiness observable, the misidentification probability of background events increases at least two times. This immediately shows the advantage of considering the multivariate technique instead of a simple cut on the sensitive observables.

[1]The idea is to illustrate how much the selection strategy is benefited by considering a multivariate strategy, with respect to performing simple cuts in the observalbes.

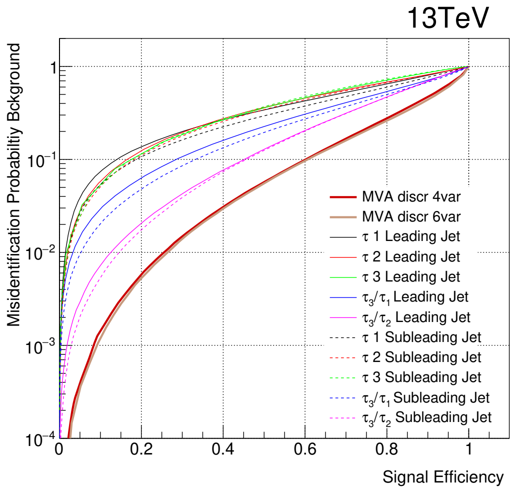


Figure 7.9: Signal efficiency selection as function of the misidentification probability of selecting background events, considering different observables to perform a selection cut. The *MVA discr 4var* refers to the MVA discriminant obtained training the algorithm with τ_3/τ_1, τ_3/τ_2 ratios for leading and subleading jets, while *MVA discr 6var* is referred when τ_1, τ_2 and τ_3 for the leading and subleading jets are used instead.

7.3 Final selection strategy.

The selection strategy was extended in this chapter by adding a criterion derived from a multivariate analysis. The event selection further considered is built by the following steps:

- ✓ **boosted dijet selection:** the two leading fat jets are required to have $p_T > 400$ GeV, $|\eta| < 2.4$ (contained within the tracker region) and a soft drop jet mass larger than 50 GeV. Events with at least one lepton are vetoed,
- ✓ **top-jet selection:** the first subjet in the leading and subleading fat jet has a mass larger than 59 GeV, 55 GeV, respectively,
- ✓ **MVA selection**: requiring that the MVA discriminant $d_{MVA} > 0.45$,
- ✓ **b-tagging selection:** at least one of the subjets, in both leading jets is b-tagged ($CSVv2$ using the MWP).

After the selection approximately 80% of the total number of selected events originate from $t\bar{t}$ pairs, meaning that the signal over background ratio is considerable larger than one. The total number of selected events in the analyzed data are 3094. Figure 7.10 shows the soft drop mass distribution in data for the leading and subleading jets after considering the final selection. The $t\bar{t}$ and QCD multijet contributions are shown. As might be noticed, the QCD multijet contribution has considerable statistical fluctuations caused from the tight selection criteria. In Chapter 8, this contribution is subtracted, using data driven methods, in order to define a measurement where only signal events are considered.

The selection efficiency as a function of p_T for the leading and subleading jets are shown in Figure 7.11. The boosted dijet selection is used for the estimation of the efficiency. Efficiency values varying from 4% up to 8% are observed. This means that approximately 4%-8% of the total number of events selected by the boosted dijet criteria are kept after applying the final selection. The highest selection efficiency is reached for jets having ~ 600 GeV.

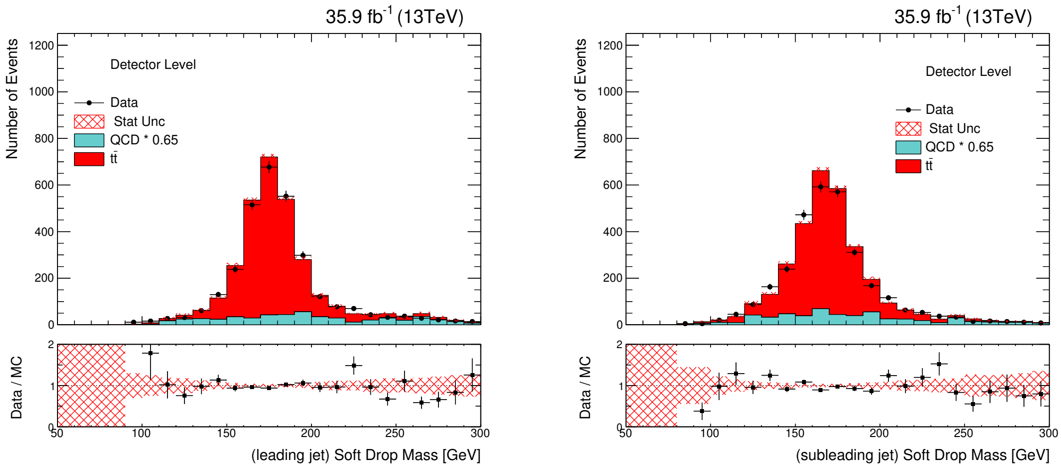


Figure 7.10: Data distributions representing the number of events as a function of the soft drop jet mass for: (left) leading jet, (right) subleading jet. The data are compared to simulated signal ($t\bar{t}$) and background (QCD multijet) contributions. Events are selected by requiring the boosted dijet selection, the top-jet selection, the MVA selection and the b-tagging selection criteria.

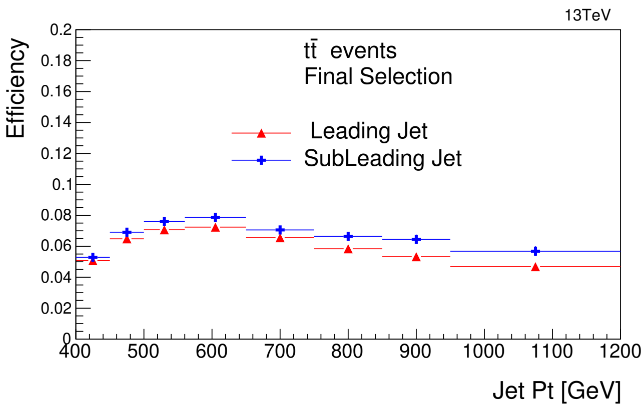


Figure 7.11: Efficiency selection as a function of p_T of the leading and subleading jets after the final selection strategy. The baseline selection considered for the efficiency is the boosted dijet criteria.

Chapter 8

Background subtraction and measurements at detector level.

In the previous chapters (Chapter 6 and Chapter 7), details on the event selection were presented. In order to have a major contribution originating from $t\bar{t}$ events, a selection based on a multivariate technique was implemented (Chapter 7). The selection criteria are a compromise of rejecting signal events, with the final goal of suppressing considerable amount of events originating from the background processes. After all the selection steps are applied, the transverse momentum phase space, for the two leading jets, is reduced to the interval 400 GeV $< p_T <$ 1.2 TeV.

In this chapter, the fiducial cross section measurements at detector level are presented. The fiducial phase space is the region where the measurement is performed. The cross section is estimated differentially in $\Delta\phi$ (azimuthal angular separation between the two leading jets), and also as a function of the transverse momentum, p_T, of either the leading or subleading jets.

The cross section at detector level is referred to the ratio of number of events in data and the integrated luminosity, and it can be expressed differential with respect to a certain observable x, by the following expression:

$$\frac{d\sigma_i^{det}}{dx_i} = \frac{N_i}{\mathcal{L}\ \Delta x_i} \tag{8.1}$$

where $\mathcal{L}$ is the integrated luminosity and N_i is the number of events in the corresponding bin in the Δx_i range. In the measurements presented in this chapter, the detector effects are not considered[1].

The normalized distributions are additionally provided. By normalizing the cross section, the

[1] Detector effects and corrected measurements are presented in next Chapter.

relevant feature of a certain distribution is the shape. Hence, the Monte Carlo predictions become independent of the global normalization factor, and therefore the comparison with the measurements can be better understood. The normalized differential distributions can be estimated as follows:

$$\frac{1}{\sigma_T^{det}}\frac{d\sigma^{det}}{dx} = \frac{1}{\sigma_T}\frac{N_i}{\mathcal{L}\ \Delta x_i}\,, \tag{8.2}$$

where σ_T is the total cross section in the fiducial phase space given by the following expression (where N_T is the total number of events):

$$\sigma_T^{det} = \frac{\sum N_i}{\mathcal{L}} = \frac{N_T}{\mathcal{L}}\,. \tag{8.3}$$

In the signal region, the selected events are not only originating from $t\bar{t}$ processes. About 20% of the events are coming from QCD multijet processes. In order to have well defined cross section measurements of $t\bar{t}$ processes, those background events need to be subtracted. The method implemented to estimate and subtract the background contribution in the signal region is discussed in Section 8.1.

The method used for background estimation could introduce some additional model dependence uncertainties, i.e, it uses a combination of theoretical and data-driven techniques. The complexity of accurately estimating the background contribution, in the signal region, arises from the fact that background events have a similar signature as $t\bar{t}$ processes. Furthermore, since the background subtraction is performed before the unfolding procedure, those sources of uncertainties are difficult to revert. As an alternative, an additional measurement is presented, in which, rather than focusing on finding events originating from $t\bar{t}$ processes, they are defined in terms of measurable stable objects identified as *inclusive top jets*. Those final state objects are defined by the selection criteria in the signal region.

The final results related to the differential distributions as a function of the p_T of the two leading jets are summarized in Figures 8.1 and 8.2, respectively, where the measured cross sections at detector level, given by formula 8.1 [1], before and after the QCD background subtraction are shown. The first plot in each figure represents the measurements in the fiducial phase space, when the QCD events have been considered as part of the signal, while the second plot shows the measurements after this contribution is subtracted. The inclusive fiducial phase space is divided into two $|\eta|$ regions: the most central region ($|\eta| < 0.5$), and the remaining fiducial phase space ($0.5 < |\eta| < 2.4$). The relative statistical uncertainty ($\epsilon = \sqrt{N}/N = 1/\sqrt{N}$) is shown for each bin, reaching up to 40% in the higher p_T region.

Analogously, figure 8.3 shows the differential distributions with respect to the azimuthal separation ($\Delta\phi$) between the two leading jets. The inclusive fiducial phase space is divided in two regions depending on the transverse momentum of the leading jet: 400 GeV$< p_T <$ 600 GeV, $p_T >$ 600 GeV. As can be noticed, the uncertainty due to limited statistics reaches 25% in the lower $\Delta\phi$ region. Figure 8.4 shows the measurements performed in the back-to-back dijet configuration ($170^\circ < \Delta\phi < 180^\circ$). A fine binning has been considered in order to explore better this region. The aim of the latter measurement is to explore this region where interesting effects might appear, i.e this region could be sensitive to effects from soft gluon resummation.

[1] Cross section without any corrections, which will be later unfolded by detector effects. The unfolding is performed in Chapter 9.

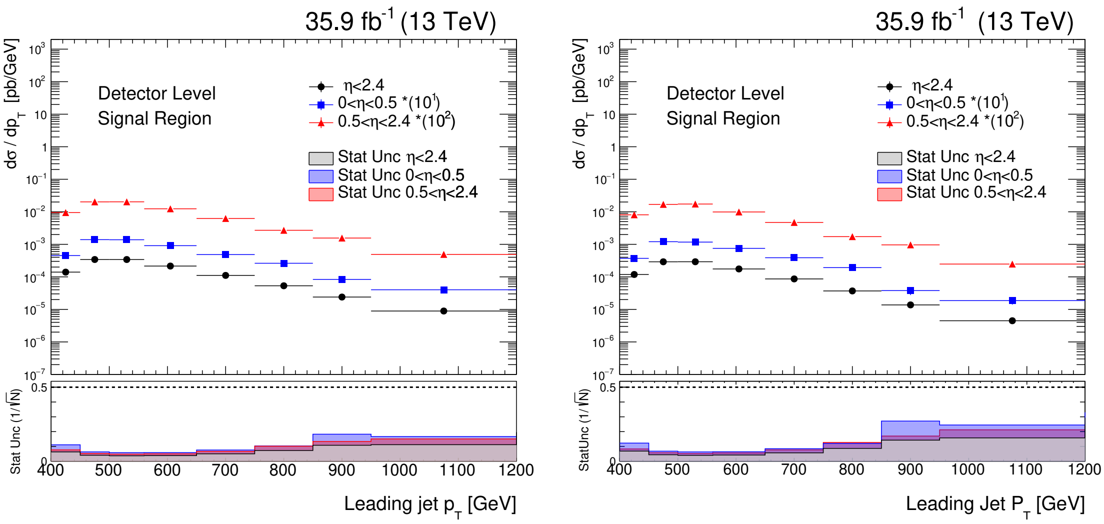


Figure 8.1: Cross section in the fiducial phase space (detector level) differential in the transverse momentum p_T of the leading jet: (left) inclusive top jet measurements (right) $t\bar{t}$ measurements

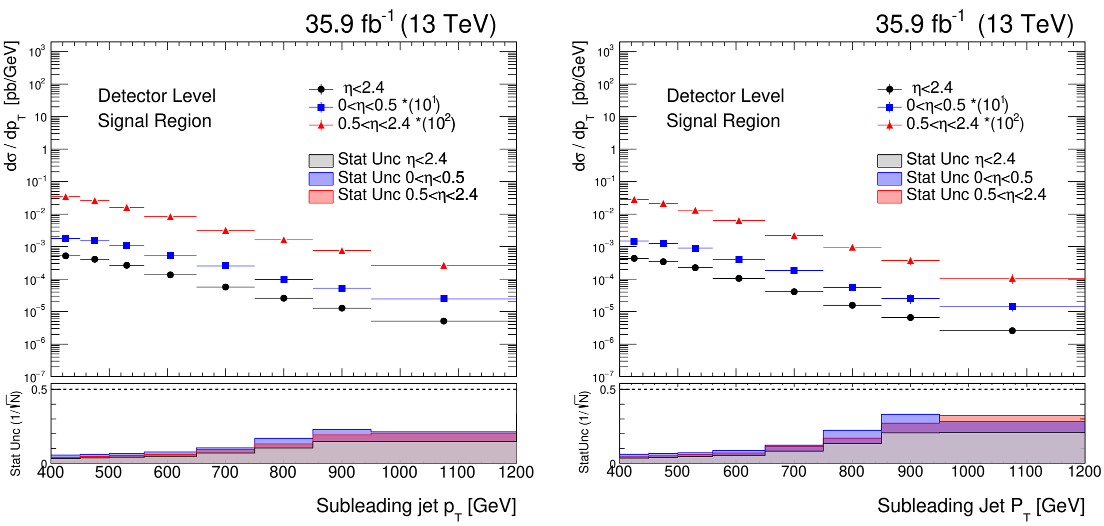


Figure 8.2: Cross section in the fiducial phase space (detector level) differential in the transverse momentum p_T of the subleading jet: (left) inclusive top jet measurements (right) $t\bar{t}$ measurements.

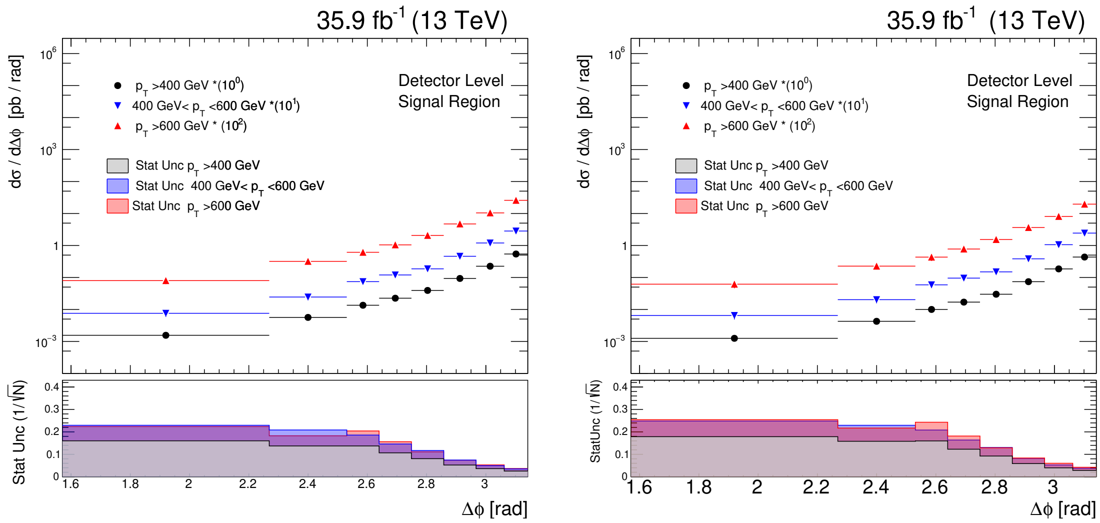


Figure 8.3: Cross section in the fiducial phase space (detector level) differential in the azimuthal separation between the two leading jets $\Delta\phi$: (left) inclusive top jet measurements (right) $t\bar{t}$ measurements .

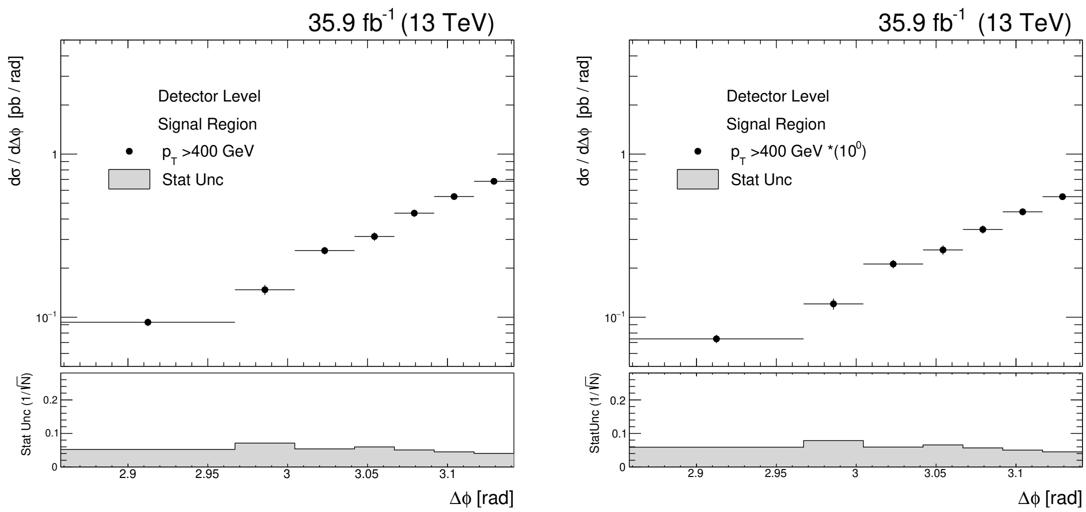


Figure 8.4: Absolute cross section in the fiducial phase space (detector level) differential in the azimuthal separation between the two leading jets $\Delta\phi$: (left) inclusive top jet measurements (right) $t\bar{t}$ measurements . The fine binning in the most back-to-back region is considered.

This chapter is further divided in two sections. In Section 8.1, the background subtraction procedure is explained in details, and in Section 8.2, the fiducial measurements previously presented (for both cases, when the QCD multijet contribution has been considered as part of the signal, and when it has been subtracted) will be compared with theory predictions at detector level.

8.1 Background Estimation and Signal Extraction

In this section, the procedure of background subtraction is described. After the final selection two main processes can be distinguished in the selected sample: events associated to $t\bar{t}$ processes, and QCD multijet events. The latter are considered as background events since the goal is to measure the former ones.

There are two main reasons why the background contribution cannot be estimated using the available Monte Carlo predictions for QCD multijet events. The main reason is related to the fact that the predictions are not able to accurately describe all the observables in the whole phase space region. This statement could be associated with the fact that the predictions are produced at LO in QCD. The second reason is related to the limited statistics of the available Monte Carlo samples. The distributions of the measured observables, in the signal region, are strongly affected by this source of uncertainty, due to the tight selection criteria considered in the analysis.

The background estimation is performed using data-driven techniques. The main steps in the procedure are the following:

1. A control region (CR) is defined, in which the dominant contribution is given by QCD multijet events. Similar kinematic requirements as in the signal region are considered. In this way, the distributions of the observables sensitive to the kinematic variables (i.e jet mass) are not significantly changing from the control region to the signal region. This point is further discussed in 8.1.1.

2. A background template (normalized distribution) is defined from the data. The contributions coming from $t\bar{t}$ and W+jets events are subtracted. The background template can be written as:
$$B^{CR}(x) = \frac{1}{N}(D^{CR}(x) - \sum_i N_i B_i^{CR}(x)), \tag{8.4}$$
where N is a scale factor which guarantees that $B^{CR}(x)$ is a normalized distribution in the considered phase space of the observable x, D^{CR} is the distribution taken from the data, and the last term is the sum of all other possible contributions in this region, scaled by the N_i factors. This point is further discussed in 8.1.2.

3. A transfer function ($f^{MC}(x)$), taking into account possible kinematic changes from the control region to the signal region is estimated using the QCD Monte Carlo predictions:
$$f^{MC}(x) = \frac{Q^{SR}(x)}{Q^{CR}(x)}, \tag{8.5}$$
where $Q^{SR}(x)$ and $Q^{CR}(x)$ are normalized distributions of an observable x in the signal and control regions respectively. This point is further discussed in 8.1.3.

4. The background templates in the signal region are estimated, by applying the transfer function to the background templates in the control region $B^{CR}(x)$ as follows:
$$B^{SR}(x) = \frac{1}{N_f} f^{MC}(x)\; B^{CR}(x), \tag{8.6}$$
where N_f is a new scale factor which guarantees that the new template in the signal region is a normalized function over the x phase space. By normalizing the background templates, the role of the transfer function is to only modify the shape of the background template in the signal region: $B^{CR}(x)$. This point is further discussed in 8.1.4.

The background templates are estimated for the following observables:

- ✓ $\Delta\phi$ in the inclusive phase space ($p_T^{lead} > 400$ GeV), and in the regions 400 GeV $< p_T^{lead} < 600$ GeV, $p_T^{lead} > 600$ GeV.
- ✓ $\Delta\phi$ corresponding to the fine binning in the back to back region, in the inclusive phase space ($p_T^{lead} > 400$ GeV).
- ✓ p_T leading jet, in the inclusive phase space ($|\eta| < 2.4$), and in the two regions $|\eta| < 0.5$, $0.5 < |\eta| < 2.4$.
- ✓ p_T subleading jet, in the inclusive phase space ($|\eta| < 2.4$), and in the two regions $|\eta| < 0.5$, $0.5 < |\eta| < 2.4$.
- ✓ soft drop jet mass of the leading and subleading jets in the inclusive phase space, and in all the exclusive η and p_T^{lead} regions.

The distributions of the observables associated to the first four items correspond to the final distributions which are measured. The observables corresponding to the last item, are functions which will be used for fits in the signal region, in order to estimate the scale factors for both contributions: QCD and $t\bar{t}$. Since the background templates are normalized functions, the scale factor obtained represents the absolute cross section for the QCD multijet contribution in a specific phase space. Therefore, the templates need to be provided in each of the considered phase spaces defined by η and the p_T of the leading jet.

In the following, each of the four steps previously mentioned are discussed. After the background templates are estimated, the signal is extracted from the measurements. This last step is described in the subsection 8.1.5.

8.1.1 Definition of the Control Region.

For the definition of the control region, the multivariate discriminant (d_{MVA}) and the b-tagging information are used. As base line selection, the boosted dijet selection criteria and the top jet selection ($m_{sub0}^{lead} > 55$ GeV, $m_{sub0}^{sublead} > 59$ GeV) are considered. Six different regions are defined. In Figure 8.5, the schematic representation of the different regions are shown. In the sketch, the signal region is named as *SRD*. This region is defined by the multivariate selection criterion and the condition that the two leading jets are b-tagged. By reverting the b-tagging requirement (none of the two leading jets are b-tagged), the $t\bar{t}$ contribution is reduced, and a control region can be defined. Hence, the control region, used for the background template definition, is the one labeled as *CRC* in the sketch. As can be noticed, the kinematic requirement given by the d_{MVA} criterion is the same than one for the signal region.

The additional defined regions (*SRA*, *VRE*, *VRF*, *CRB*) are further used in the analysis as validations regions and to deduce the transfer function. The regions labeled with *VRE* and *CRB*, are regions where the QCD contribution is considerably enhanced with respect to the $t\bar{t}$ contribution. Nevertheless, these regions cannot be used as control regions for defining the background templates, since the kinematic requirement differs from the signal region, given by the d_{MVA} criterion. General regions, only considering the b-tagging information, are defined as *SR*, *VR*, and *CR* (in those regions there is no cut on the d_{MVA} observable, for instance *SR* is the sum of the two regions labeled as *SRA+SRD*).

Figure 8.6 shows the soft drop jet mass distributions of the leading jet. In this figure the percentage of each contribution, predicted by the Monte Carlo simulations, can be seen. Each row represents

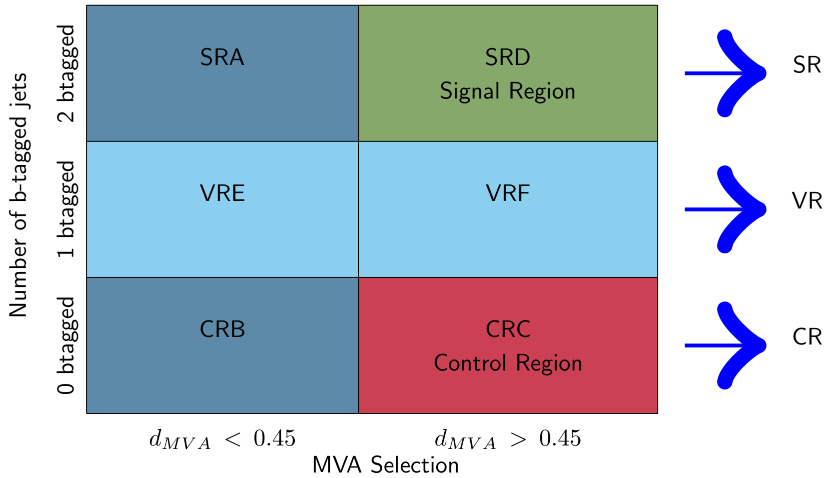


Figure 8.5: Schematic picture of different regions defined by using the multivariate selection and the b-tagging requirements. A base-line selection is previously assumed: the boosted dijet selection and the top jet selection. The Signal Region (SRD) is represented in green color, while the Control Region (CRC) is represented in red color.

each of the b-tagging categories, while the first column of plots corresponds to the general regions (*SR*, *VR*, CR).

The two main contributions (QCD multijet, and $t\bar{t}$) are scaled by normalization factors in order to have good agreement with data. The scale factors for the absolute cross sections were estimated requiring that good data-Monte Carlo agreement is reached in several regions simultaneously (by considering the same scale factors). In the case of the $t\bar{t}$ contribution, with scaling factor $N_{t\bar{t}} = 0.654$, this requirement is satisfied for all the regions simultaneously. In the case of the QCD multijet contribution, the respective scaling factors were defined for each b-tagging category: $N_{QCD} = 0.60, 0.67, 0.74$ (CR, VR, and SR respectively). A more rigorous method to estimate the scaling factors is discussed in section 8.1.5. In Figures F.1-F.2 (in the Appendix), the data-Monte Carlo predictions comparison considering other observables, i.e the multivariate discriminant (d_{MVA}) and the mass of the first subjet (W-subjet candidate), are provided.

In Table 8.1, the percentages of each contribution in all regions are given. In the signal region (SRD), the $t\bar{t}$ events are $\sim 77\%$ of the total amount of selected events. In the control region (CRC), the dominant contribution is the QCD multijet events, with $\sim 86\%$ of the total selected events.

Region	VR	VRE	VRF	CR	CRB	CRC	SR	SRA	SRD
$t\bar{t}$ (%)	4.83	2.91	32.50	0.80	0.54	6.36	25.70	13.22	77.22
QCD (%)	92.67	94.80	62.24	96.79	97.26	86.38	72.58	85.05	21.10
W+jets (%)	2.5	2.29	5.26	2.41	2.19	7.25	1.71	1.73	1.68

Table 8.1: Percentage of the main contributions in each of the regions represented in 8.5. Other contributions refer to $W + jets$ and Drell-Yan events.

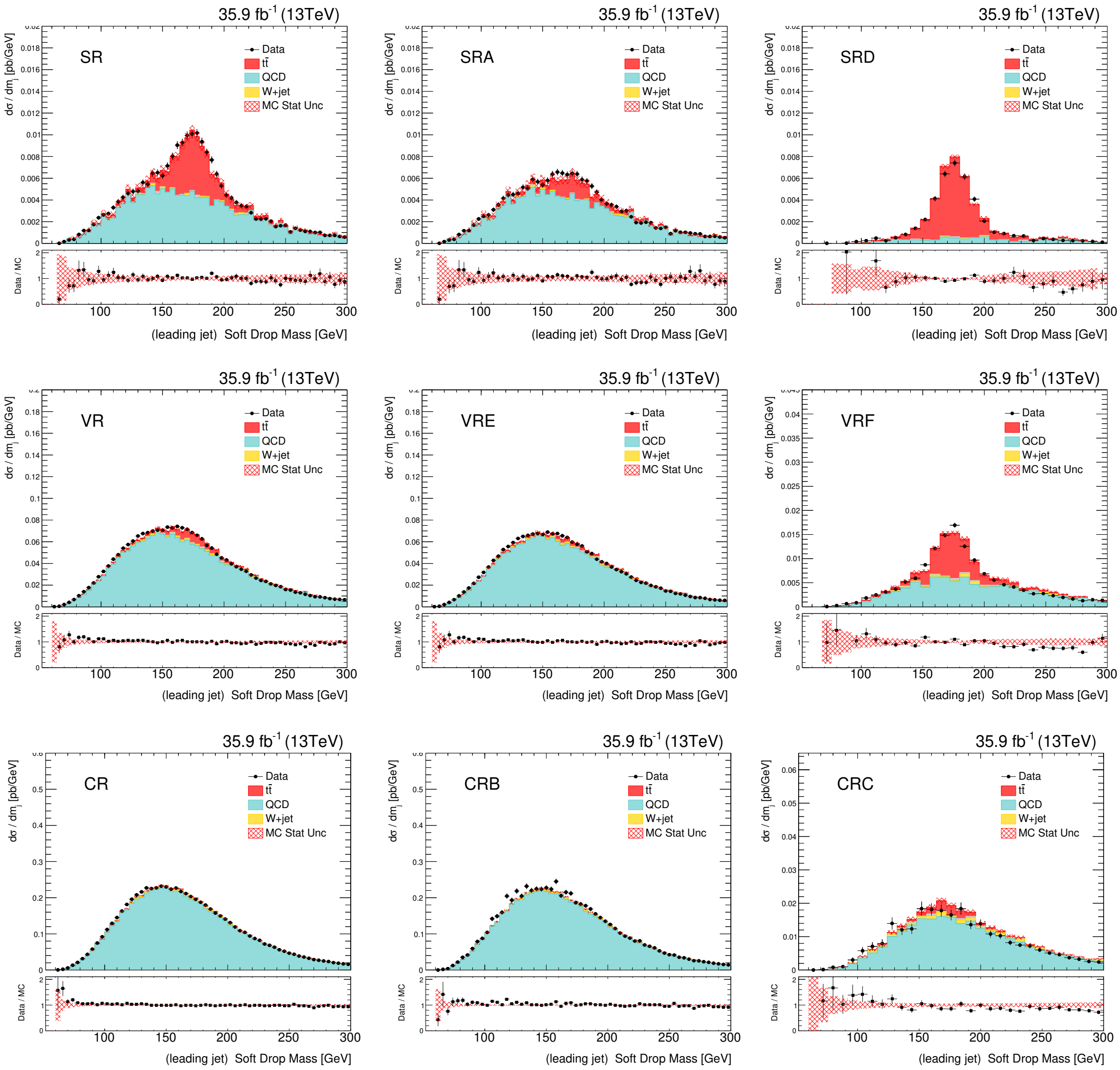


Figure 8.6: Soft Drop Mass distributions for the leading jet in different regions defined in Figure 8.5. The normalization factor for the $t\bar{t}$ contribution has been taken as $N_{t\bar{t}} = 0.654$, while for the QCD multijet contribution: $N_{QCD} = 0.74$ in the first row (2-btagged jets), $N_{QCD} = 0.67$ in the second row (exclusively 1-btagged jet), $N_{QCD} = 0.6$ in the third row (zero btagged jet).

While a good agreement with data has been found, for the jet mass and the *MVA* discriminant observables, considerable disagreement ($\sim 20\%$) is observed in the $\Delta\phi$ spectra. Figure 8.7 shows the $\Delta\phi$ distribution, considering the regions where the QCD multijet contribution is the dominant (CR, CRB). This effect was already observed after applying the boosted dijet selection presented in the previous Chapter (Figure 6.11).

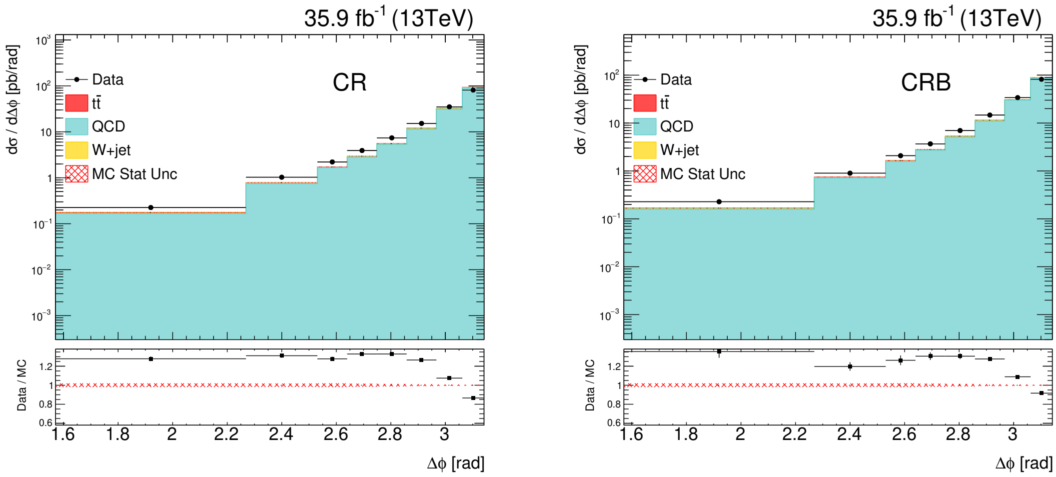


Figure 8.7: Comparison of the $\Delta\phi$ distribution in data with the Monte Carlo predictions in two different regions (left) CR (right) CRB. The only non-negligible contribution in this region is the QCD multijet events.

8.1.2 Background templates in the Control Region

As a second step, the background templates are defined in the control region (CRC). The main difficulty to deal with it is the contamination of about 14% from $t\bar{t}$ and $W + jets$ contributions.

The background templates, in the control region $(B^{CR}(x))$ of a certain observable x, can be estimated as follows:

$$B^{CR}(x) = \frac{1}{N}(D^{CR}(x) - N_{t\bar{t}}S^{CR}(x) - W^{CR}(x)) \tag{8.7}$$

where $D^{CR}(x)$ refers to the absolute cross section distribution from the data, $S^{CR}(x)$ is the $t\bar{t}$ distribution given by the Monte Carlo, and $W^{CR}(x)$ refers to the W+jet contribution, also obtained from Monte Carlo predictions. The parameter N is a scale factor such that the background template is defined as a normalized distribution. The factor $N_{t\bar{t}}$ is applied to the $t\bar{t}$ absolute cross section predicted in Monte Carlo in order to have reasonable agreement with data. This scale factor is taken as the factor previously mentioned ($N_{t\bar{t}} = 0.654$). A variation of this parameter *up* and *down* is then considered as systematic uncertainty on the background modeling. The *up* variation is taken as: $N_{t\bar{t}} = 1$, while the down variation is $N_{t\bar{t}} = 0.50$.

Since the major contribution is given by the QCD multijet events, variations of the second and third terms in the equation 8.7 should play a minor role in the background template distributions.

As additional source of uncertainty for the background modeling, the reweighting of the events in $t\bar{t}$ simulated sample, referred in the Section 6.5 (p_T top reweighting), is considered. An alternative background template is estimated when the re-weighting of the events is not taken into consideration, and the difference with the nominal background template is considered as uncertainty. Figure 8.8 shows the effect of reweighing on the $\Delta\phi$ and p_T spectra, for simulated $t\bar{t}$ events in the control region. As can be seen, in the case of the $\Delta\phi$ observable, the effect is less than 1%, while in the p_T spectra up to $\sim$ 30% of difference for higher p_T is observed. However, since in the control region, the $t\bar{t}$ contribution is about $\sim$ 6%, the effect in the background template is expected to be small (estimated as $\sim$ 1.8%).

Figure 8.9 shows the background templates of the three observables: p_T of the leading jet, $\Delta\phi$, and

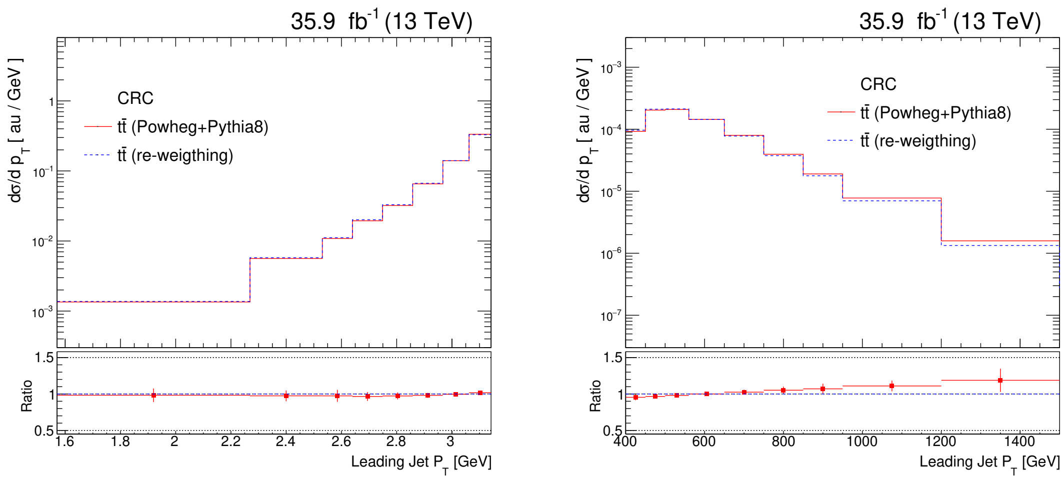


Figure 8.8: (left) $\Delta\phi$ and (right) p_T spectra for $t\bar{t}$ simulated events, in the control region, considering and not considering the re-weighting of the events mentioned in Section 6.5.

the soft drop jet mass of the leading jet. The model-systematic uncertainties are also illustrated for each observable in the additional plots. As systematic uncertainties the following variations are considered:

- ✓ Scaling the $t\bar{t}$ contribution by $N_{t\bar{t}} = 1$ (*up* variation)
- ✓ Scaling the $t\bar{t}$ contribution by $N_{t\bar{t}} = 0.5$ (*down* variation)
- ✓ Not considering the reweigthing of the $t\bar{t}$ simulated events (see Section 6.5)

For each of the three variations, a background template is estimated, and the difference with the nominal template is taken as uncertainty. The largest deviation considering the three mentioned cases, with respect to the nominal template, is taken as the global model background uncertainty for the templates in the control region. As can be noticed, in the case of p_T and $\Delta\phi$, the predicted model uncertainties have a maximum value of $\sim 2\%$ in the highest p_T and the lowest $\Delta\phi$ regions respectively. For the soft drop jet mass templates, the effect is up to $\sim 10\%$ in the bin corresponding to the t-quark mass (the region most sensitive to the variation of the scaling factor, $N_{t\bar{t}}$). The dominant uncertainty is the one obtained by scaling up the $N_{t\bar{t}}$ factor to one. The effect of applying the reweighting of the $t\bar{t}$ events is less than 1% over all the p_T phase space, while for the other observables it almost vanishes.

In Figure 8.10, the background templates are compared with the Monte Carlo predictions of simulated QCD multijet events at detector level in the control region. Two Monte Carlo predictions are considered in the comparison: MADGRAPH+PYTHIA8 and PYTHIA8.

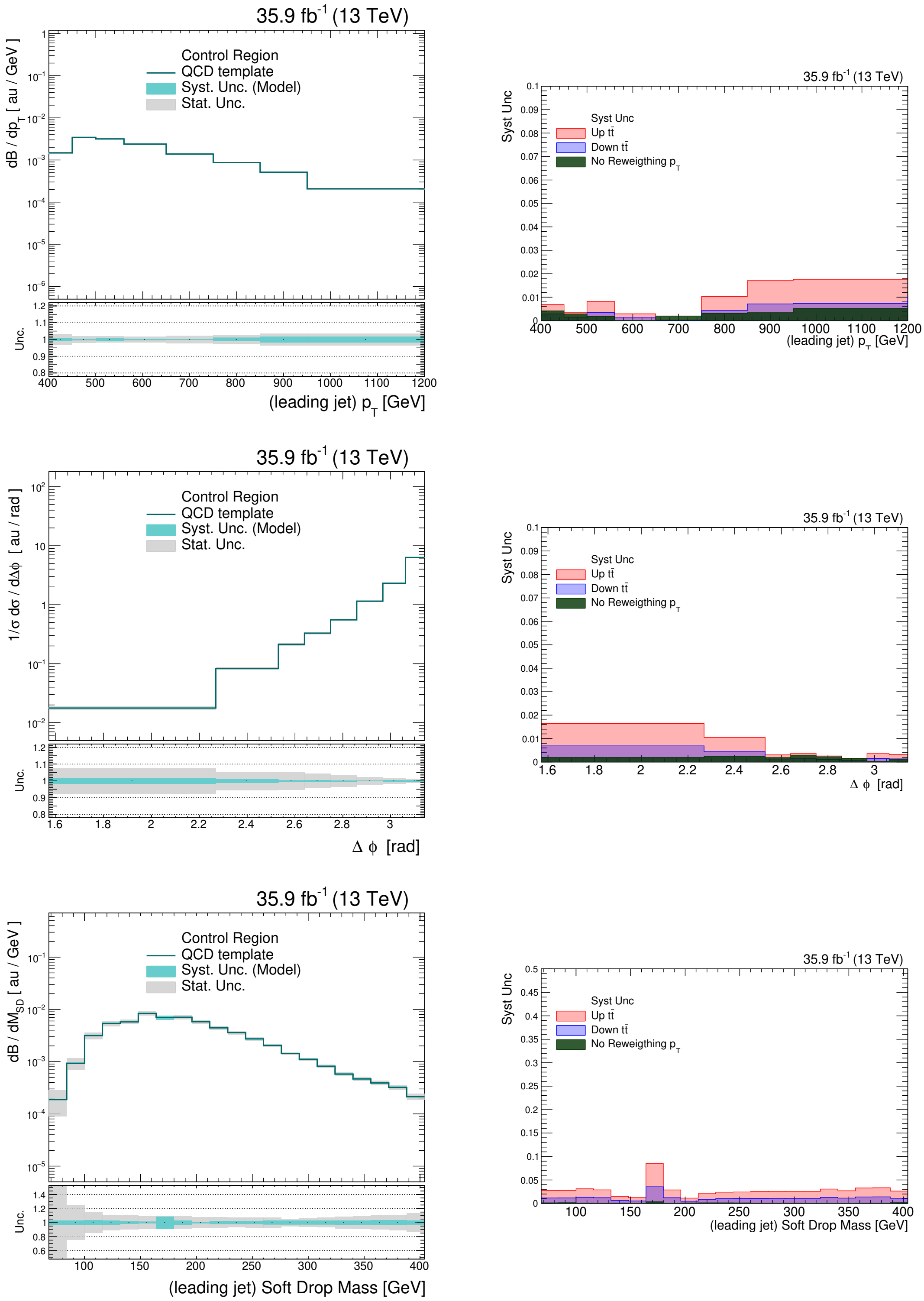
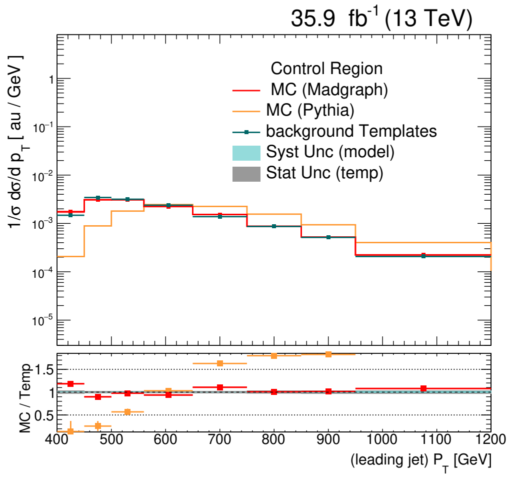
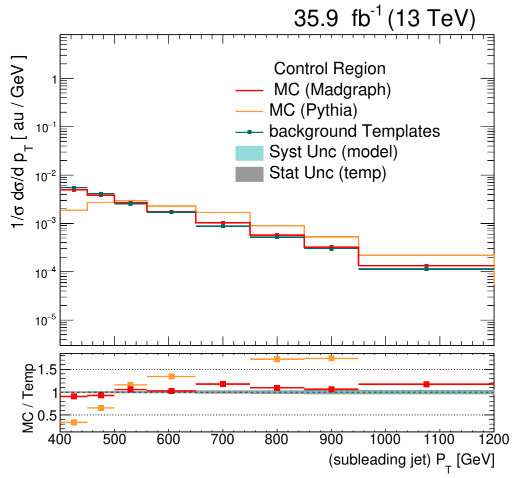


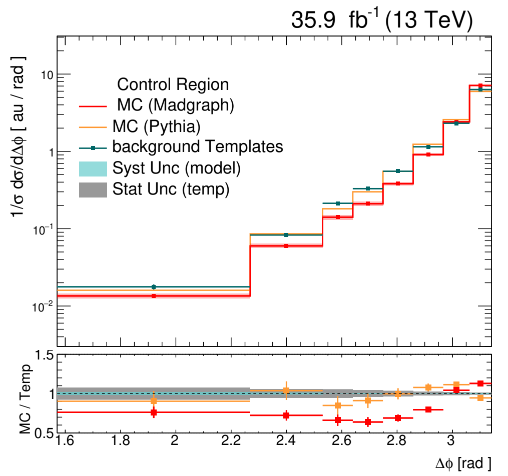
Figure 8.9: (left) Background template distributions in the control region and (right) systematic model-dependence uncertainties, for three observables: (upper) p_T of the leading jet, (middle) $\Delta\phi$, (lower) the Soft Drop mass of the leading jet.



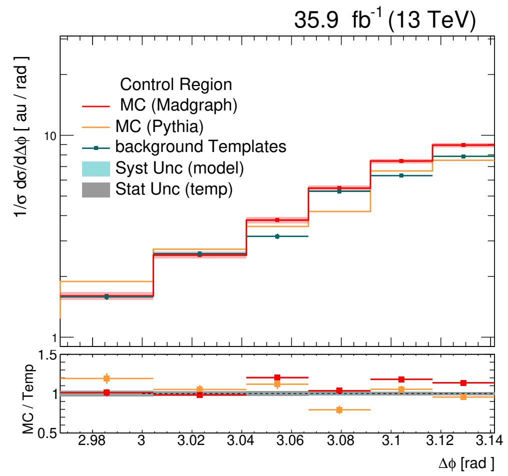
(a)



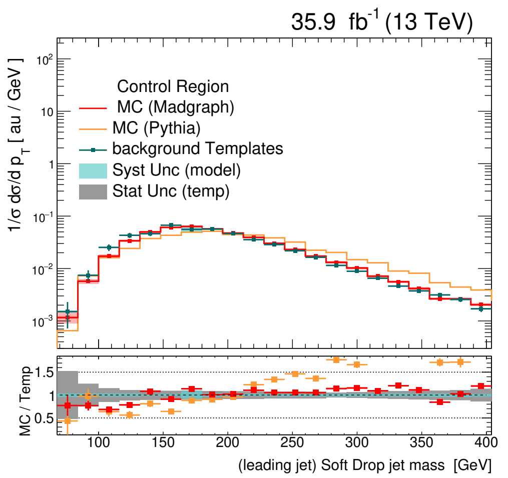
(b)



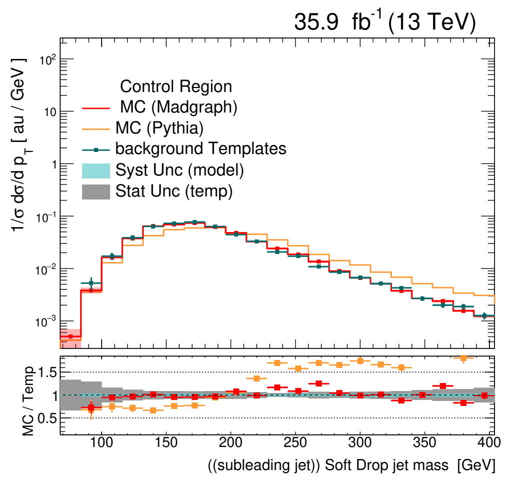
(c)



(d)



(e)



(f)

Figure 8.10: Comparison of background templates obtained with a data-driven method with respect to the Monte Carlo prediction in the Control Region. The plots corresponds to the following observables: (upper left) leading jet p_T (upper right) subleading jet p_T, (middle left) $\Delta\phi$ between the two leading jets (middle right) $\Delta\phi$ in the back to back region (lower left) soft drop jet mass of the leading jet (lower right) soft drop jet mass of the subleading jet.

In the case of the p_T spectra (Figure 8.10a, 8.10b), the background templates are found to be in reasonable agreement with the prediction provided by the MADGRAPH+PYTHIA8 simulations. The predictions from PYTHIA8, are not in good agreement with the background templates.

In Figures 8.10c, 8.10d, the templates for the $\Delta\phi$ observable are compared to the Monte Carlo predictions. The latter ones correspond to the back-to-back region. The PYTHIA8 sample describes better this observable. For the MADGRAPH+PYTHIA8 predictions, a differences of $\sim$ 10% up to $\sim$ 40% is observed.

In the case of the soft drop jet mass distributions (Figures 8.10e and 8.10f), a good agreement of the background template with the MADGRAPH+PYTHIA8 simulated sample is noticed, especially in the region of the t-quark mass window (150 GeV-200 GeV).

8.1.3 Transfer Function

The third step is to determine, for each observable, the function which characterizes the kinematic changes between the signal region (*SRD*) and the control region (*CRC*). This function is estimated using the Monte Carlo predictions obtained with the MADGRAPH+PYTHIA8 simulated sample.

The transfer function is obtained as the ratio between the normalized distributions in the two different regions:

$$f^{MC}(x) = \frac{Q^{SRD}(x)}{Q^{CRC}(x)}, \tag{8.8}$$

One disadvantage to deal with is that the statistics in the *SRD* region using the available QCD Monte Carlo samples is not enough to perform an accurate estimation of the function $f^{MC}(x)$. Nevertheless, since the change from the control region (CRC) to the signal region (SRD) is given by the b-tagging selection criterion, one could expect that similar changes might appear when the regions *CRB* (zero b-tagged jets) and *SRA* (two b-tagged jets) are applied. The previous condition can be summarized by the following expression:

$$f^{MC}(x) = \frac{Q^{SRD}(x)}{Q^{CRC}(x)} \approx \frac{Q^{SRA}(x)}{Q^{CRB}(x)}, \tag{8.9}$$

Figure 8.11 shows the comparison of the normalized distributions differential in p_T of the leading jet, considering the three possible b-tagging categories: none of the two leading jets has been b-tagged (0-btag), only one of the two leading jets has been b-tagged (1-btag), and the two leading jets have been b-tagged (2-btag). Two plots are shown, one where the region corresponding to the kinematic cut opposite to the signal region ($d_{MVA} < 0.45$) is considered, and the other when the respective kinematic cut to the signal region ($d_{MVA} > 0.45$) is considered. The transfer function is estimated in the first case (left plot), considering the regions *SRA* (two b-tagged) and *CRB* zero b-tagged). The transfer function is then plotted in the figure on the right (corresponding to the kinematic cut defining the signal region), to validate the assumption expressed by equation 8.9. In the case of the p_T spectra, a kinematic change of up to $\sim$ 40% is noticed, in the high p_T region.

Figure 8.12 shows the analogous comparison for the soft drop jet mass of the leading jet. In this case, the transfer function predicts almost no change in the t-quark mass window, while a noticeable change in larger mass range can be noticed (up to $\sim$ 50%).

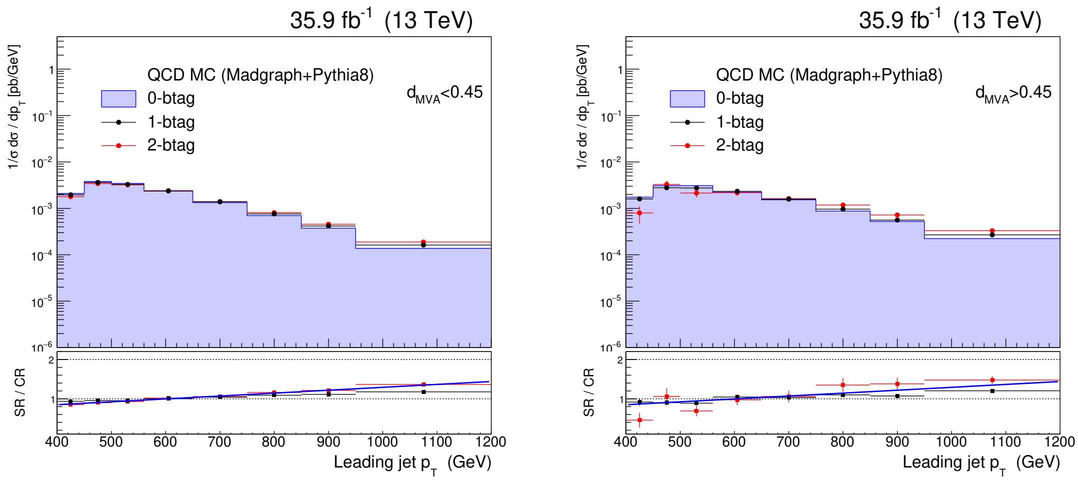


Figure 8.11: Transfer function with respect to the p_T of the leading jet (right) when $d_{MVA} < 0.45$ (left) when $d_{MVA} > 0.45$. The plot at the left give the ratio of the distributions between, i.e *SRA* and *CRB* , while the plot at the right corresponds to the ratio between the distributions in *SRD* and *CRC*.

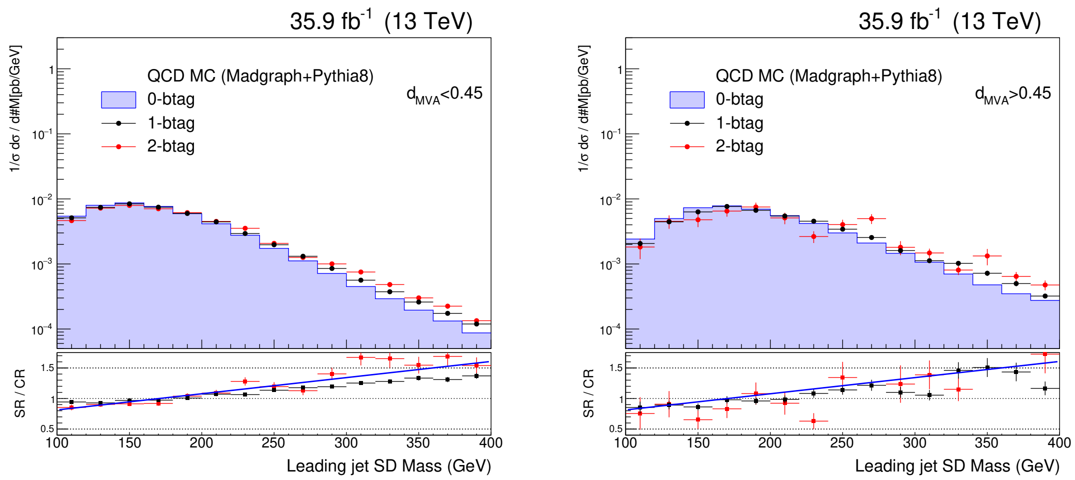


Figure 8.12: Transfer function with respect to the p_T of the leading jet (right) when $d_{MVA} < 0.45$ (left) when $d_{MVA} > 0.45$. The plot at the left give the ratio of the distributions between, i.e *SRA* and *CRB* , while the plot at the right corresponds to the ratio between the distributions in *SRD* and *CRC*.

Figure 8.13 shows the comparison corresponding to the $\Delta\phi$ observable. In this case, no change is observed when different b-tagging regions are considered. Both kinematic regions ($d_{MVA} > 0.45$ and $d_{MVA} < 0.45$) are similar with respect to this assumption. Hence, no transfer function is applied (the background templates obtained in the control region are supposed to be the same in the signal region).

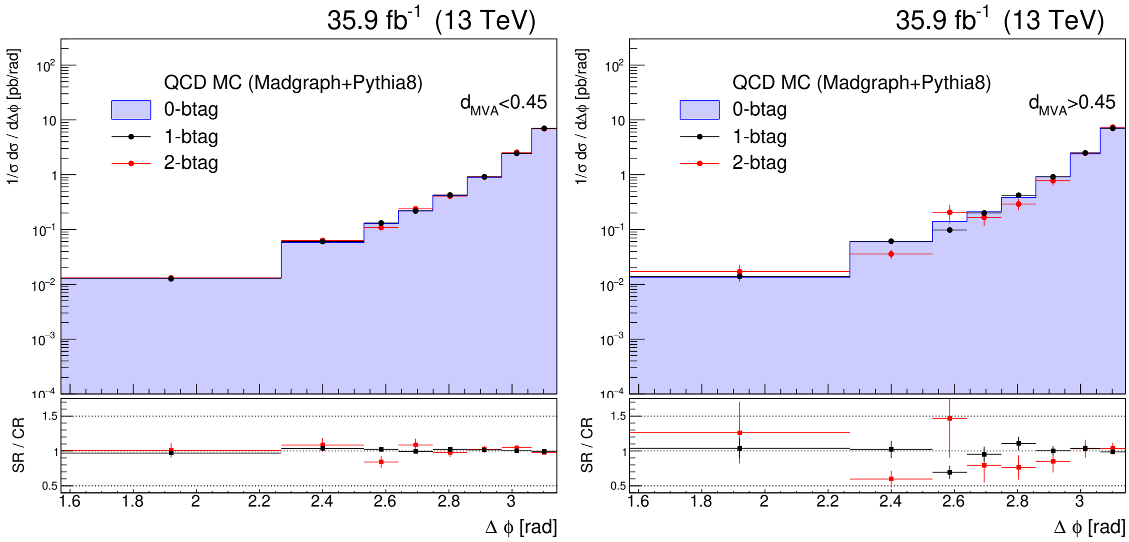


Figure 8.13: Transfer function with respect to the $\Delta\phi$ between the two leading jets (right) when $d_{MVA} <$ 0.45 (left) when $d_{MVA} > 0.45$. The transfer function in this case have been considered as a flat line (no change expected). Transfer function with respect to the p_T of the leading jet (right) when $d_{MVA} < 0.45$ (left) when $d_{MVA} > 0.45$. The plot at the left give the ratio of the distributions between, i.e *SRA* and *CRB* , while the plot at the right corresponds to the ratio between the distributions in *SRD* and *CRC*.

8.1.4 Background Templates in the Signal Region

As last step, the background templates in the signal region are estimated by applying the transfer function to the background templates in the control region.

Figure 8.14 shows the comparison of the background templates with the Monte Carlo predictions, obtained with the MADGRAPH+PYTHIA8 sample, in the signal region. The shown results are similar to the ones observed when the comparison was performed in the control region. But in this case, the most remarkable message is that the background templates are better populated in statistics terms, while huge fluctuations might be noticed in the Monte Carlo predictions due to the limited statistics. Additionally, the Monte Carlo predictions might be not optimal due to the LO accuracy in the QCD calculations. The previously described modeling dependence uncertainties are extrapolated to the signal region.

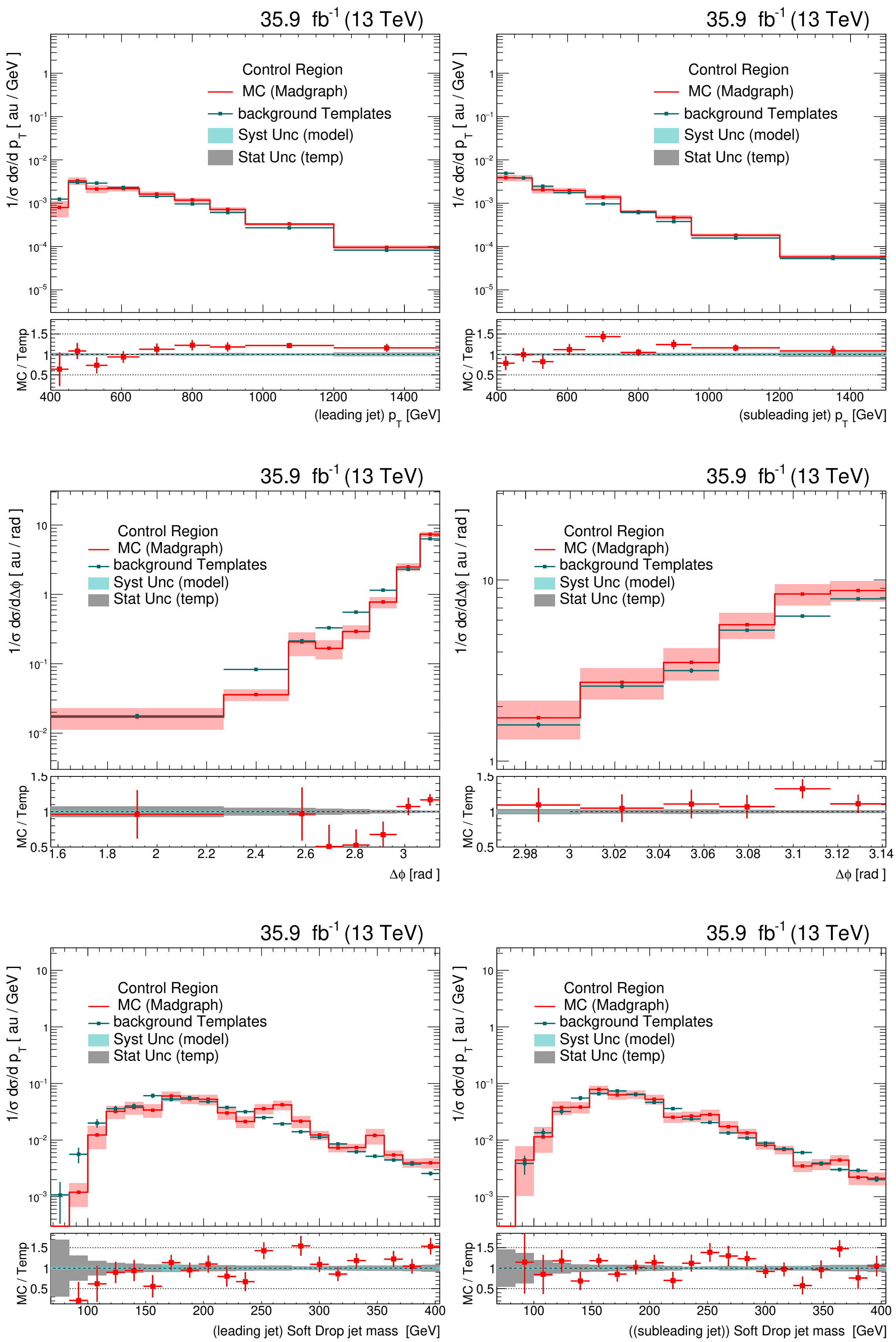


Figure 8.14: Comparison of background templates obtained by data-driven method and the Monte Carlo prediction in the Signal Region. The plots correspond to the following observables: (upper left) leading jet p_T, (upper right) subleading jet p_T, (middle left) $\Delta\phi$ between the two leading jets (middle right), $\Delta\phi$ between the two leading considering the back to back configuration, (lower left) soft drop jet mass of the leading jet, (lower right) soft drop jet mass of the subleading jet. The coloured band represented for the MC Madgraph predictions count for the statistical uncertainties

8.1.5 Signal Extraction

After all background templates have been determined, a fit of the two main contributions (QCD multijet and $t\bar{t}$) is performed in the signal region, in order to determine their respective normalization factors.

The fitting procedure is performed using the *Roofit* package [167], as main tool. Two probability density functions (pdf) are used, the Gaussian distribution and the Crystal Ball distribution given by the following expressions respectively:

$$G(x;\sigma,\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \tag{8.10}$$

$$C(x;\sigma,\mu,n,\alpha) = N \begin{cases} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ A(B - \frac{x-\mu}{\sigma})^{-n} & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha \end{cases} \tag{8.11}$$

where in the last equation N, A and B are parameters which guarantee that $C(x)$ is normalized to unity.

The following steps are then considered:

- ✓ the shape (normalized function) of the soft drop jet mass distribution predicted by simulated $t\bar{t}$ events, $S(m)$, is fitted to the sum of a Gaussian function and a Crystal Ball function as follows:

$$S(m;\sigma_G^t,\mu_G^t,\sigma_C^t,\mu_C^t,n_C^t,\alpha_C^t,f_G) = f_G G(m;\sigma_G^t,\mu_G^t) + (1-f_G)C(m;\sigma_C^t,\mu_C^t,n_C^t,\alpha_C^t), \tag{8.12}$$

  where $G(m;\sigma_G^t,\mu_G^t)$ is the Gaussian component with free parameters labeled as: σ_G^t, μ_G^t (standard deviation and central value of the Gaussian function); and $C(m;\sigma_C^t,\mu_C^t,n_C^t,\alpha_C^t)$ is the Crystal Ball component with the respective free parameters: σ_C^t, μ_C^t, n_C^t, α_C^t. The factor f_G is the fraction of the Gaussian component (also a free parameter to be determined from the fit).

- ✓ The background template (in the signal region) corresponding to the jet mass observable $B(m)$, is fitted to a Crystal Ball pdf:

$$B(m;\sigma_C^b,\mu_C^b,n_C^b,\alpha_C^b) = C(m;\sigma_C^b,\mu_C^b,n_C^b,\alpha_C^b), \tag{8.13}$$

  where now σ_C^b, μ_C^b, n_C^b, α_C^b are the free parameters of the Crystal Ball function to be determined in the fit.

- ✓ From both distributions, $S(m)$ and $B(m)$, are then fitted to the normalized distribution of the corresponding observable m, obtained in the data, $D(m)$. The parameters in S and B functions obtained in the previous steps, are fixed, except two of them: σ_G^t, μ_G^t, which correspond to the Gaussian component of the signal template ($S(x)$). For those parameters, the values previously determined are considered as initial values, but small changes in the fit are allowed, to account for possible disagreement between data and simulation (detector resolution effects), since the $S(x)$ is estimated by Monte Carlo simulation. This step can be written as follows:

$$D(m;\sigma_G^t,\mu_G^t,f) = fS(m;\sigma_G^t,\mu_G^t) + (1-f)B(m), \tag{8.14}$$

  where f is a free parameter to be determined, representing the fraction of signal events in the final sample.

Figure 8.15 illustrates the steps of the fitting procedure considering the leading and subleading jets in the inclusive fiducial phase space. As result, a fraction of signal events of 81.7% (80.0%) has been obtained in the fit using the soft drop mass distribution in the leading (subleading) jet. Table 8.2 shows the obtained parameters of the fits: σ_G^t and f. The parameter σ_G^t is shown first after the fit in the signal template (S), and then when the fit is performed in the data. The differences in the result of the fraction for signal events (f), when either the leading or the subleading jets are considered, are taken as source of uncertainties for the background estimation, as well as the error of the parameter.

Table 8.2: Value of the free parameters considered in the fitting corresponding to equation 8.14, for the leading and subleading jets. The parameters σ_G^t and μ_G^t are additionally shown after the fitting performed considering the signal template (equation 8.12).

Observable m^{SD}	$\sigma_G^t \pm \Delta\sigma_G^t$ (GeV) (S)		$\sigma_G^t \pm \Delta\sigma_G^t$ (GeV) (D)		f	Δf
leading jet	174.51	±22.22	175.40	±22.66	0.817	±0.054
subleading jet	167.69	±16.38	166.52	±15.58	0.800	±0.047

The scale factors for QCD multijet contribution (N_{QCD}) and for the signal contribution ($N_{t\bar{t}}$) can be directly estimated as function of the fraction of signal events (f), as follows:

$$N_{QCD} = (1-f)\frac{\sigma_{data}}{\sigma_{QCD}}\,,\, N_{t\bar{t}} = f\frac{\sigma_{data}}{\sigma_{t\bar{t}}}, \tag{8.15}$$

where σ_{data} is the integrated cross section in the data, σ_{QCD} in the integrated cross section for QCD multijet process (from MC), and $\sigma_{t\bar{t}}$ is the integrated cross section for the $t\bar{t}$ contribution (from MC). The latter two are deduced from the respective Monte Carlo simulations.

In the case of the *QCD* multijet contribution, the scale factor (N_{QCD}) represents the absolute cross section (since the background templates are normalized distributions). Therefore, in order to perform the measurements in each of the exclusive regions of the fiducial phase space, the fitting procedure is done in these regions independently.

The scale factors for both contributions are summarized in Table 8.3. The results are reported in each of the phase space regions and using the fit to the soft drop mass for the leading and subleading jets. In the case of the inclusive fiducial phase space , and also in the different phase space regions, the scale factors for the QCD contribution obtained by either the leading and subleading jets are expected to be the same. Therefore, the scale factor considered for the background subtraction is taken as the mean value, and the respective variations are considered as systematic uncertainty of the normalization factor. For the different η regions, the leading and subleading jets are independently analyzed and the scale factors are not expected to be the same.

Finally, the signal distribution for each final observable (p_T, $\Delta\phi$) is extracted in the following way:

$$S(x) = D(x) - N_{QCD}B(x), \tag{8.16}$$

where x refers the observable (p_T or $\Delta\phi$), $D(x)$ is the measured cross section distribution in the fiducial phase space, N_{QCD} and $B(x)$ are the corresponding normalization factors and background templates.

The distributions obtained by the equation 8.16 are the final measurements after background subtraction. The systematic uncertainty values associated to the background modeling are obtained by considering the following variations:

Table 8.3: Scales factors for QCD and $t\bar{t}$ contributions, obtained in the signal region after fitting both contributions to the data. The lower values of $N_{t\bar{t}}$ are given by discrepancies found for the $t\bar{t}$ MC simulations. The values of N_{QCD} are applied to the background templates, which are normalized distributions, hence, they give the absolute cross section for this contributions in the specific phase space region were they are being determined.

Observable m^{SD}	$N_{t\bar{t}} \pm \Delta N_{t\bar{t}}$		$N_{QCD} \pm \Delta N_{QCD}$	
inclusive fiducial phase space				
leading jet	0.695	0.0456	$1.57 \cdot 10^{-2}$	$1.03 \cdot 10^{-3}$
subleading jet	0.680	0.040	$1.72 \cdot 10^{-2}$	$1.05 \cdot 10^{-3}$
$\|\eta\| < 0.5$				
leading jet	0.688	0.052	$6.75 \cdot 10^{-2}$	$5.08 \cdot 10^{-3}$
subleading jet	0.669	0.045	$6.32 \cdot 10^{-2}$	$4.237 \cdot 10^{-3}$
$0.5 < \|\eta\| < 2.4$				
leading jet	0.678	0.051	$1.02 \cdot 10^{-2}$	$7.769 \cdot 10^{-3}$
subleading jet	0.678	0.040	$1.14 \cdot 10^{-2}$	$6.821 \cdot 10^{-3}$
400GeV$< p_T^{lead} <$ 600 GeV				
leading jet	0.678	0.051	$1.02 \cdot 10^{-2}$	$7.769 \cdot 10^{-3}$
subleading jet	0.678	0.040	$1.14 \cdot 10^{-2}$	$6.82 \cdot 10^{-3}$
$p_T^{lead} >$ 600 GeV				
leading jet	0.678	0.051	$1.02 \cdot 10^{-2}$	$7.77 \cdot 10^{-3}$
subleading jet	0.678	0.040	$1.14 \cdot 10^{-2}$	$6.821 \cdot 10^{-3}$

✓ *Up-Down* variation of the normalization factor N_{QCD}: a measured distribution is obtained by subtracting the background with a scale factor $N_{QCD} \pm \Delta N_{QCD}$. In case that both leading jets can be considered simultaneously[1], the result of their respective fits are combined:

$$N_{QCD} = \frac{N_{QCD}^{lead} + N_{QCD}^{sublead}}{2}\ , \Delta N_{QCD} = \frac{\sqrt{\Delta N_{QCD}^{lead} + \Delta N_{QCD}^{sublead}}}{2} \tag{8.17}$$

✓ Considering the *up-down* variation of the background templates, the fits are performed taking into account the *up* and *down* variations of the background templates in the signal region. The background is subtracted considering the new scale factor and the background templates.

The model uncertainty is taken as the maximum deviation from the nominal measurements from the considered variations.

[1]For example, when considering the $\Delta\phi$ observable.

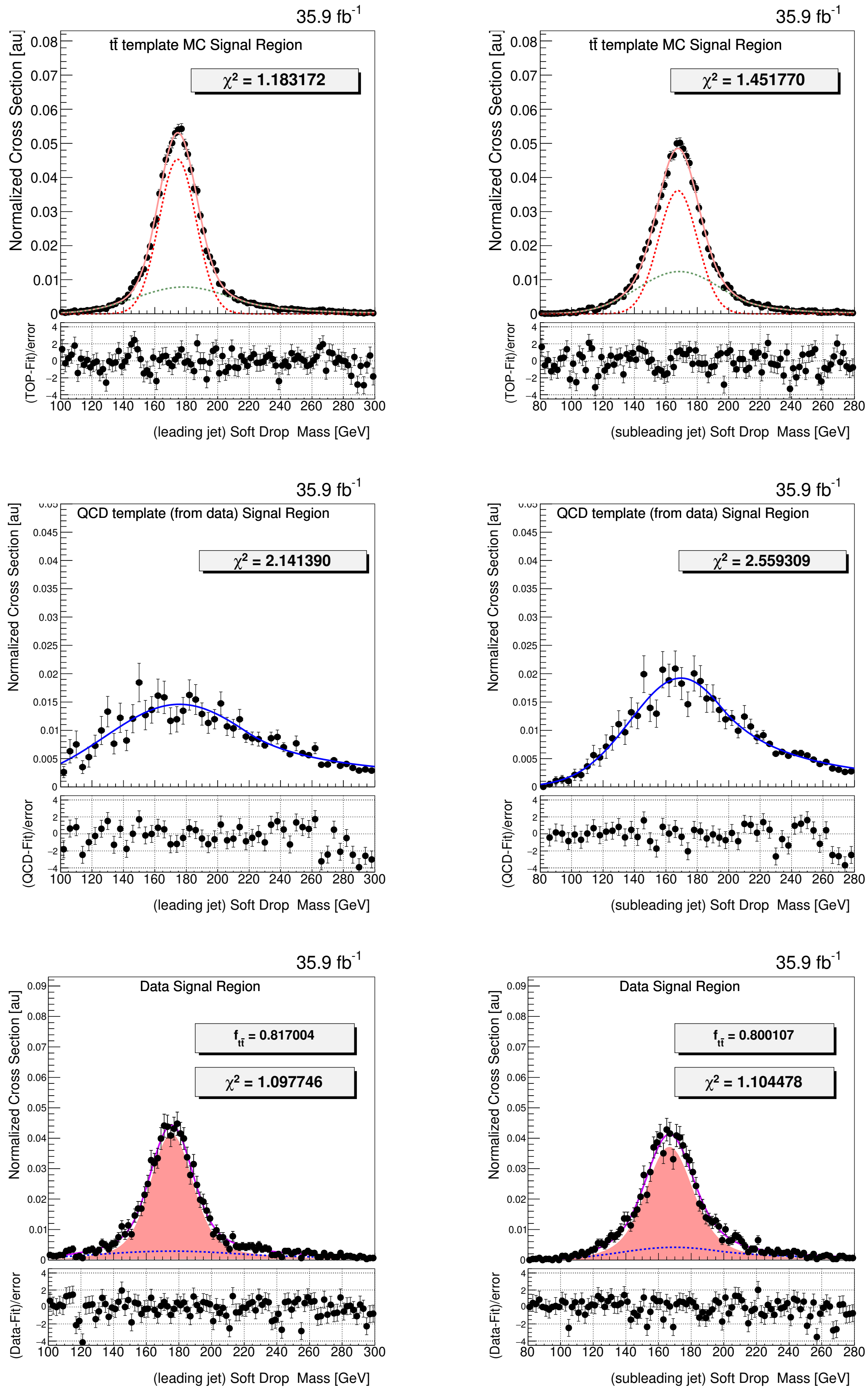


Figure 8.15: Different steps in the fit procedure in the signal region considering the Soft Drop Mass distribution of: (left) the leading jet, (right) the subleading jet. The upper plots show the fit of the signal template, $S(m)$, the middle plots show the fit of the background templates ($B(m)$) and the lower plots show the fit of the data distributions ($D(x)$)

8.2 Comparison of the fiducial measurements with theory predictions at detector level

In this section, the measurements in the fiducial phase space are compared to theory predictions at detector level. The inclusive top jet measurements and the $t\bar{t}$ measurements are presented. Additionally to the cross section measurements, the normalized distributions are as well compared to the shape of the predictions.

The theory predictions for the $t\bar{t}$ events are obtained from POWHEG+PYTHIA8, POWHEG+HERWIGpp and MADGRAPH+PYTHIA8. The details about each simulation can be found in Section 6.1. In the case when the QCD multijet contributions need to be considered for the inclusive top jet measurements, this contribution from predictions obtained with the MADGRAPH+PYTHIA8 event generator is added.

The comparison of measurements with theory predictions at detector level is, however, affected by detector effects which could bias the results. In Chapter 11, the comparison are performed when the measurements have been unfolded to the particle level.

8.2.1 Measurements differential in p_T of the two leading jets.

Figure 8.16 and 8.17 shows the cross section and the normalized distributions differential in p_T of the leading and subleading jet, respectively. The upper plots correspond to the inclusive top jet measurements, while the lower plots to the $t\bar{t}$ measurements. Both, the absolute (left plot) and normalized (right plot) distributions are shown. The results of both measurements are equivalent to each other, meaning that no bias has been introduced through the background subtraction procedure.

As a general conclusion all the predictions describe (within uncertainties) the shape of the measured distributions. In general, if a scale factor is applied to all the theory predictions, a good data-theory agreement would be observed over all the p_T phase space. The predictions are $\sim 40\%$ above the measurements, however, the differences are obtained in the whole p_T phase space region, meaning that, if one applies a scaling factor to the predictions, there would be a good agreement with measurements.

8.2.2 Measurements differential in $\Delta\phi$.

Figure 8.18 shows the cross section and the normalized distributions differential in $\Delta\phi$ (azimuthal angle between the two top jets candidates). Figure 8.19 represents the analogous results obtained when a fine binning is considered for the most back-to-back region. The upper plots correspond to the inclusive top jet measurements, while the lower plots to the $t\bar{t}$ measurements. The results of both measurements are equivalent to each other.

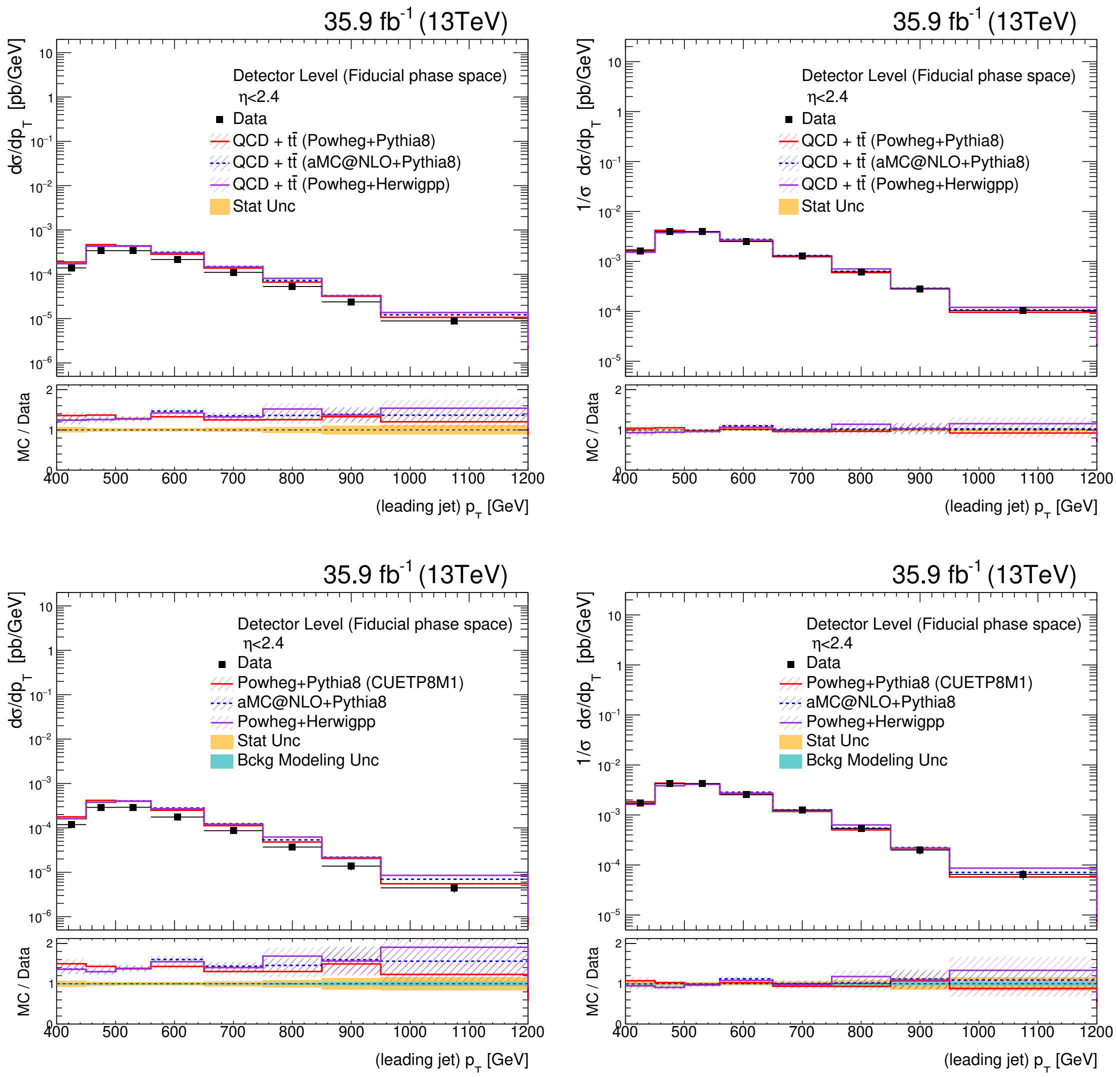


Figure 8.16: Measurements in the inclusive fiducial phase space differential in p_T of the leading jet: (left) cross section (right) normalized distribution, (upper plots) corresponding to the inclusive top jet measurements (lower plots) and to the $t\bar{t}$ measurements. The data points in the upper panel only include the statistical uncertainties. The shadow area in the Monte Carlo predictions correspond to the statistical uncertainties.

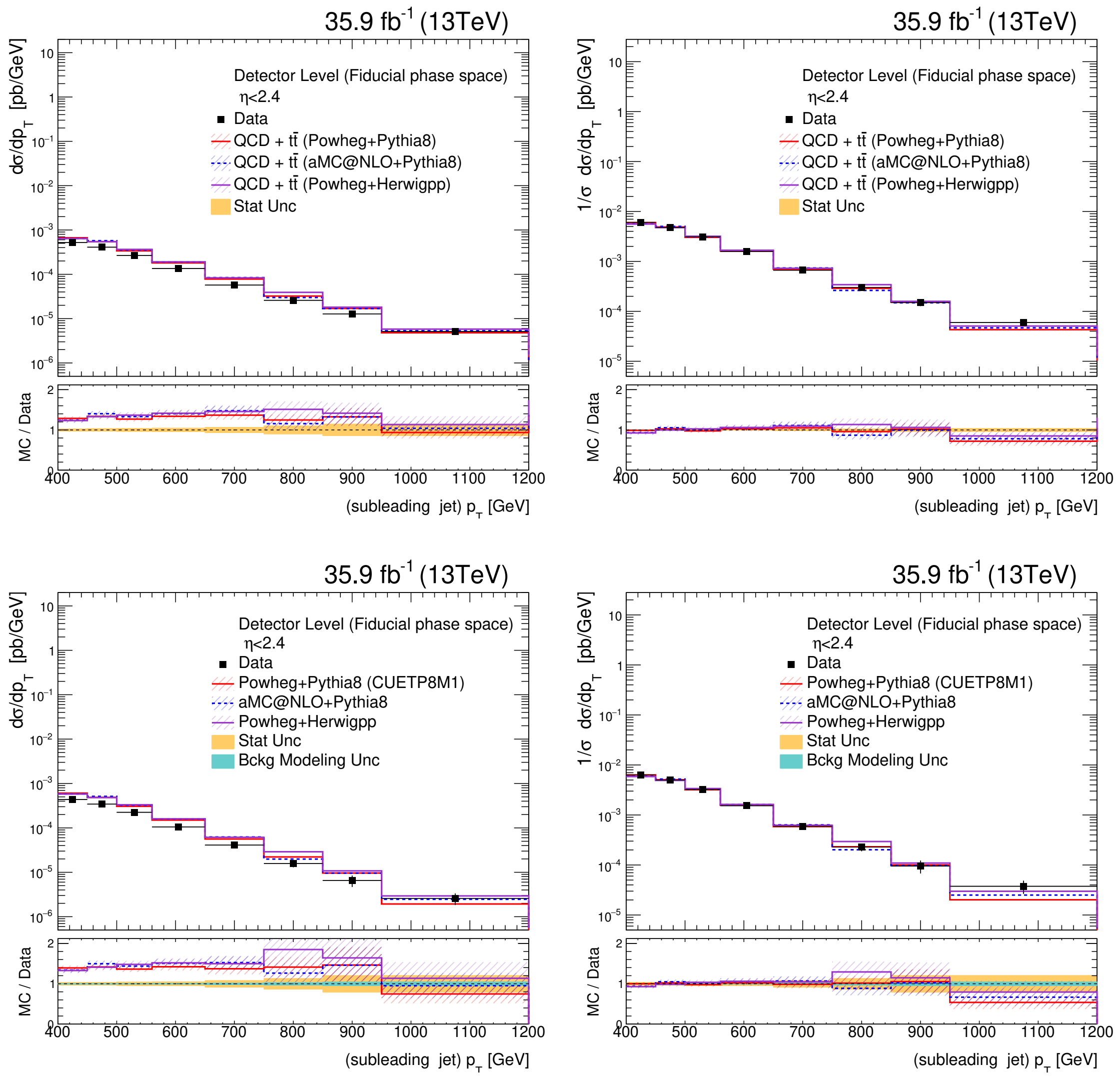


Figure 8.17: Measurements in the inclusive fiducial phase space differential in the p_T of the subleading jet: (left) cross section (right) normalized distribution, (upper plots) corresponding to the inclusive top jet measurements (lower plots) and to the $t\bar{t}$ measurements. The data points in the upper panel only include the statistical uncertainties. The shadow area in the Monte Carlo predictions correspond to the statistical uncertainties.

In this case interesting observations can be conclude:

- ✓ the shape of the distributions from MADGRAPH+PYTHIA8 differ with the ones given from POWHEG+PYTHIA8 and POWHEG+HERWIGpp. The distributions from the latter ones are equivalent, meaning that the sensitivity of this observable to the Parton Shower is not large, while in comparison to MADGRAPH+PYTHIA8, the distributions are having narrower opening angle.
- ✓ the shape of the measured distributions are better described by the MADGRAPH+PYTHIA8, meaning that, this observable is sensitive to the extra radiations considered in the hard processes (MADGRAPH+PYTHIA8 consider real emissions in the calculations of the ME). The differences on the shape of the measured distributions with respect to the POWHEG+PYTHIA8 and POWHEG+HERWIGpp are $\sim 20\%$.
- ✓ concerning the absolute cross section, and differently to what was observed for the distributions differential in p_T, in the case of POWHEG+PYTHIA8 and POWHEG+HERWIGpp predictions, the differences are up to $\sim 60\%$ and they are not an overall factor, rather they gradually increase towards the $\Delta\phi$ (the largest differences are observed when $\Delta\phi \to \pi$).
- ✓ when comparing measured distributions in the most back-to-back region considering the fine binning (Figure 8.19), a tendency of increasing the predicted cross section from POWHEG+PYTHIA8 and POWHEG+HERWIGpp is specially noticed when $\Delta\phi > 3.05$ rad ($\Delta\phi > 175°$).

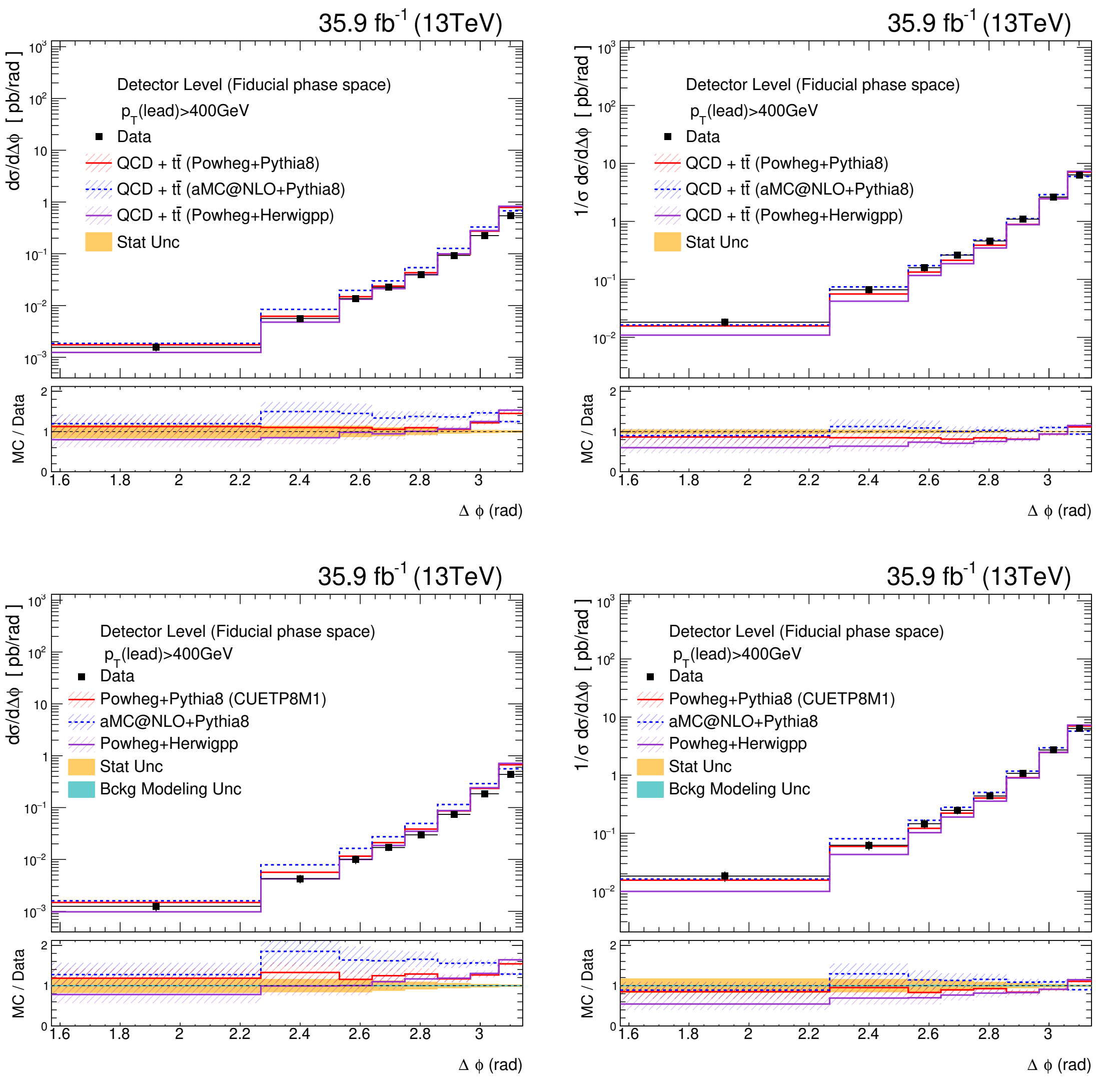


Figure 8.18: Measurements in the inclusive fiducial phase space differential in $\Delta\phi$: (left) cross section (right) normalized distribution, (upper plots) corresponding to the inclusive top jet measurements (lower plots) and to the $t\bar{t}$ measurements. The data points in the upper panel only include the statistical uncertainties.

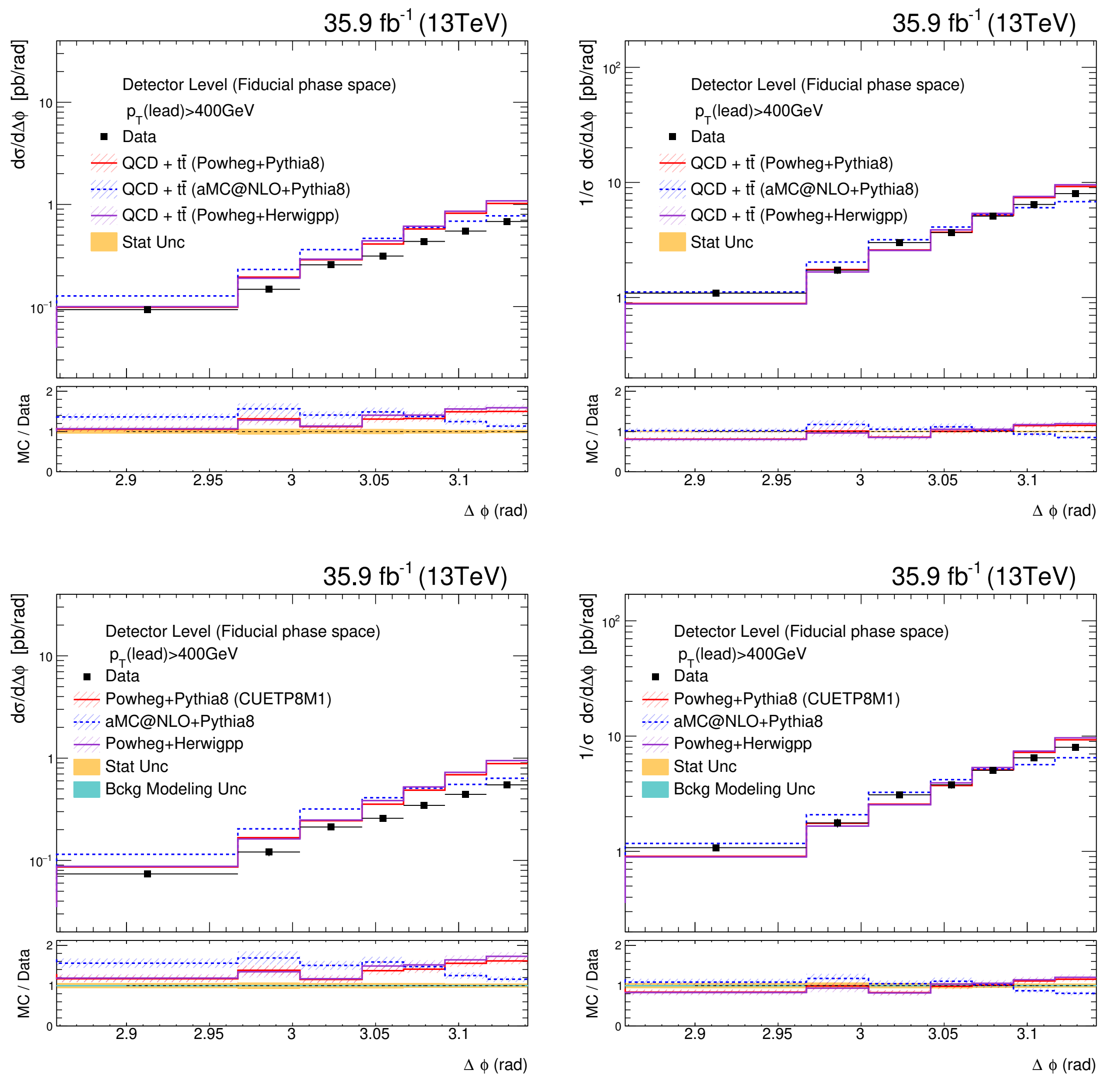


Figure 8.19: Measurements in the inclusive fiducial phase space differential in $\Delta\phi$: (left) cross section (right) normalized distribution, (upper plots) corresponding to the inclusive top jet measurements (lower plots) and to the $t\bar{t}$ measurements. The plots correspond when a fine binning is considered in the most back-to-back region. The data points in the upper panel only include the statistical uncertainties.

Chapter 9

Detector effects and unfolding procedure.

The differential cross section distributions, presented in Chapter 8, correspond to measurements at detector level in the fiducial phase space. Hence, the measured distributions might be affected by different sources of detector effects and they cannot be directly compared with theory predictions which do not consider the detector simulation. The theory predictions corresponding to the level where the particles are considered without any detector interaction are known as "*stable particle level*"[1]. This chapter is devoted to study those detector effects, and to correct the measured distributions in order to have an equivalent measurement at the particle level. The method to correct the measured distributions to generator level is known as *unfolding* procedure.

Figure 9.1 illustrates the results after the unfolding procedure has been applied. The upper plots correspond to the differential distributions with respect to the p_T (of the leading jet), while the lower plots to the $\Delta\phi$ observable. Two measurements are illustrated: the inclusive top jet measurements and the $t\bar{t}$ measurements. The plots show the comparison of the theory prediction and the measured distributions at detector and particle level. The measured distributions at particle level have been obtained after applying the unfolding procedure discussed in this chapter. The theory predictions for the $t\bar{t}$ measurements are provided with simulated $t\bar{t}$ events, while in the case of the inclusive top jet measurements the QCD multijet contribution is added to the latter ones. The aim of this chapter is to explain the steps of obtaining the measured distribution unfolded to the stable particle level.

[1]Stable particle level refers to those particles observed after the collision time corresponding to $c\tau > 10$ cm

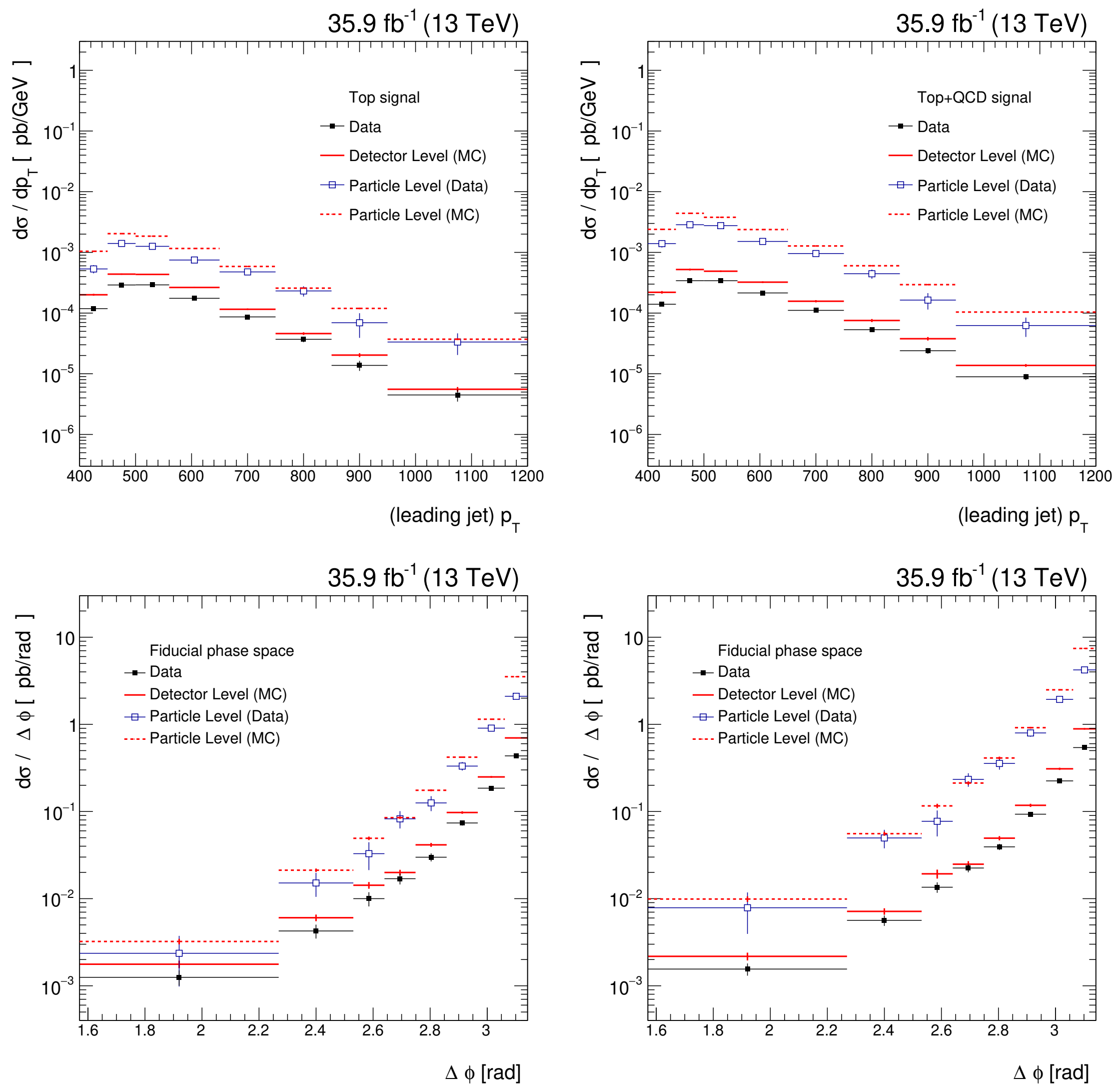


Figure 9.1: Monte Carlo and data comparison at particle and detector level, for (left) $t\bar{t}$ measurements, (right) inclusive top jet measurements. The unfolded distributions have been estimated with the iterative D´Agostini method [168]

This chapter is organized in two main blocks. Section 9.1 describes the studies of the detector effects which might affect the measured distribution, and to provide tools to quantify them. In Section 9.2, some of the available unfolding methods are briefly discussed, which are the most used in the current scenarios of High Energy experiments. One of the described methods (the iterative D´Agostini method [168]) is applied to the measurements at detector level. This method is an iterative procedure (the output information of one step is provided as input information for the subsequent step). Dedicated studies for a correct estimation of the number of iterations are presented.

9.1 Studies of detector effects

In general, a physical quantity can be measured only with finite precision. Detector effects, like noise, non-linearity response and calibration uncertainties cause a certain shift of the measured magnitude with respect to its true value.

The effects can be studied using Monte Carlo simulations of $t\bar{t}$ events, provided at both levels:

- stable particle level (*particle*): given by the theory prediction of the proton-proton collisions
- detector level (*det*): given by the theory prediction of the proton-proton collisions followed by the interaction of the outgoing particles with the detector.

In subsection 9.1.1, the resolution responses corresponding to the observables of interest are presented. In subsection 9.1.2, variables are defined, in order to quantify the detector effects on the specific measured distributions (i.e purity, stability, acceptance, background). Finally in subsection 9.1.3, the response matrices are presented. Those matrices contain the information of the correlation between the different levels, and are the crucial input information for the unfolding procedure.

9.1.1 Resolution studies

The distribution of a measured observable, if the detector is well calibrated, is centered in its true value, having a certain spread around it. This distribution response can be characterized by a Gaussian function with a standard deviation σ, also known as detector resolution related for the specific observable. Due to calibrations effects, the centered value of this distribution is in most of the cases different than zero. The smaller values the σ parameter takes, the better the resolution of the detector is.

The Gaussian distribution describes a stochastic phenomenon, while in reality there might be correlation effects, affecting the expected Gaussian pattern. This deviation from the Gaussian distribution might be seen in the low response region (tails of the distribution). Nevertheless, in the core of the distribution, a Gaussian shape is expected. Often the effects in the tails are of the order of $\mathcal{O}(10^{-2} - 10^{-4})$. Then, the resolution response can be described by combining a Gaussian distribution with additional function, i.e an exponential functions describing each of the tails.

The determination of the detector resolution is important for correctly selecting the bin widths of the distributions that are measured. If the bin widths are smaller than the estimated resolutions, the measured distributions could be highly affected by migration of events from one bin to another one. This statement directly implies that, for example, an event contributing to a certain bin at stable particle level, can appear in a neighboring bin at the detector level. Normally, such migrations effects are corrected by the unfolding procedure. But, if the distribution is strongly affected by those migration effects, difficulties might appear during the unfolding procedure. Hence, the bin width should be larger than the estimated resolution. In the analysis presented in this thesis, they are selected as at least two times larger than the estimated resolution.

The resolution response are estimated using $t\bar{t}$ simulated events with the POWHEG+PYTHIA8 event stable particle. They are defined for each observable, either because they correspond to the final measured distributions (i.e p_T, $\Delta\phi$), or because they affect the selection procedure (i.e jet mass, η).

In order to estimate the resolution response, selection criteria at both levels need to be applied. The selection criteria at both levels correspond to the baseline selection of the analysis: the boosted selection. Additionally at stable particle level, a B-hadron is required inside the jet, to ensure that it corresponds to a top jet. The events, selected at both levels, are then considered for the determination of the resolution response. The stable particle and detector level jets are then matched in order to ensure that the jet at detector level truly corresponds to the specific jet at stable particle level. The matching criteria are defined using different observables from those whose response is being determined.

The resolution response for a specific observable X is expressed as ΔR_X. The definition of each resolution response and their respective matching criteria are:

- ✓ ϕ-resolution:

$$\Delta R_\phi = \phi^{gen} - \phi^{reco} \tag{9.1}$$

  The matching criteria for the jets at both levels is: $\Delta R_{p_T} < 0.2$ and $\Delta\eta < 0.3$, where $\Delta R_{p_T} = (p_T^{reco} - p_T^{gen})/p_T^{gen}$ and $\Delta\eta = |\eta^{gen} - \eta^{det}|$.

- ✓ $\Delta\phi_{t\bar{t}}$-resolution:

$$R_{\Delta\phi_{t\bar{t}}} = \Delta\phi_{t\bar{t}}^{gen} - \Delta\phi_{t\bar{t}}^{reco} \tag{9.2}$$

  where $\Delta\phi_{t\bar{t}}$ is the azimuthal separation between the top jet candidates, with the same matching criteria as for the ϕ observable, since the $\Delta\phi_{t\bar{t}}$ involves the leading and subleading jet simultaneously, the condition is that each of the two leading jets at stable particle level should be matched to one of the two leading jets at detector level.

- ✓ p_t-resolution :

$$\Delta R_{p_t} = (p_t^{reco} - p_t^{gen})/p_t^{gen} \tag{9.3}$$

  with the matching criteria $\Delta R_{\eta\phi} < 0.3$, $\Delta R_{\eta\phi} = \sqrt{(\phi^{gen} - \phi^{reco})^2 + (\eta^{gen} - \eta^{reco})^2}$

- ✓ η-resolution:

$$\Delta R_\eta = \eta^{gen} - \eta^{reco} \tag{9.4}$$

  with matching criteria: $|\Delta R_{p_t}| < 0.2$ and $|\Delta R_\phi| < 0.3$

- ✓ jet mass resolution

$$\Delta R_{m_j} = (m_j^{gen} - m_j^{det})/m_j^{gen} \tag{9.5}$$

  where m_j refers to the soft drop jet mass, with the considered matching criteria $\Delta R_{\eta\phi} < 0.3$.

In order to estimate the specific resolution values (σ), a fit of a function given by the equations 9.1-9.5 is performed. The function which properly describes the shapes is a combination of a Gaussian function with two Exponential functions describing the tails of the response distribution. This function known as *ExpGaussExp* [169] is defined as :

$$f(x;\bar{x},\sigma,k_L,k_R) = \begin{cases} e^{\frac{k_L^2}{2}+k_L(\frac{x-\bar{x}}{\sigma})} & \text{for } \frac{x-\bar{x}}{\sigma} \le -k_L \\ e^{\frac{1}{2}(\frac{x-\bar{x}}{\sigma})^2} & \text{for } -k_L < \frac{x-\bar{x}}{\sigma} \le k_H \\ e^{\frac{k_R^2}{2}-k_R(\frac{x-\bar{x}}{\sigma})} & \text{for } k_R < \frac{x-\bar{x}}{\sigma} \end{cases} \tag{9.6}$$

where $\bar{x}$, σ, k_L, k_R are the four free parameters to be determined in the fit.

The parameters k_L and k_H are always required to be larger than 1. They represent the boundaries where a Gaussian function is fitted, hence the core of the function (the response region $\pm 1.\sigma$), is always described by the Gaussian component.

Figure 9.2 shows the resolution response related to the two main observables, p_T and $\Delta\phi$. These resolution response are an average over the whole phase space (p_T or $\Delta\phi$). The fitted values of the free parameters of the function given by the formula 9.6 are shown in the figures.

The relative resolution ΔR_{p_t} takes values of $\sim$ 5%. This implies, for example, that for a jet with $p_T \sim 450$ GeV, the detector resolution is $\sim$ 22.5 GeV. The centered value given by the parameter μ in the fit ($\mu \sim 0.4\%$, representing for a jet of 450 GeV approximately 1.1 GeV) reflects a good transverse momentum jet (p_T) calibration of the jets.

The relative resolution related to the p_T observable might depend on the p_T region. Usually, this relative value is expected to decrease, when the p_T increases. This is a direct consequence of the fact that jets with higher transverse momentum are better reconstructed by the calorimeter, which leads to better p_T resolution. Nevertheless, since the selected jets already have a high transverse momentum, no major changes are expected. The resolution values are estimated in different intervals of p_T, to corroborate the last statement. Table 9.1 shows the specific values in five intervals of p_T. The changes of the relative resolution ΔR_{p_t} from the lower p_T region to the higher p_T region is of the order of $\sim$ 0.53%. The selected bin widths, in the measured distributions differential in p_T, are larger than two times the absolute resolutions estimated for each specific bin.

In the case of $\Delta R_{\Delta\phi_{t\bar{t}}}$ (the response resolution related to the $\Delta\phi$ observable) the estimated σ value is 0.015 rad (0.86°). The smaller bin widths used in the differential measurements with respect to $\Delta\phi$ correspond to the measurements in the most back-to-back region. The selected bin widths for this region are equal to 1.6°, which is almost twice the estimated resolution. The small value of the μ parameter ($\mu = 0.0015$ rad), indicates a good detector calibration with respect to the azimuthal angle. The obtained resolution values related to the azimuthal angle of the individual jets (given by equation 9.1) are similar to the one of $\Delta\phi$ ($\Delta R_{\phi} \sim 0.6°$).

Table 9.1: Detector resolutions values related to the p_T observable, in different p_T intervals. The resolution are given by the equation 9.3.

p_T Range (GeV)	ΔR_{p_t}(%)
400-500	5.00
500-560	4.95
560-650	4.91
650-750	4.86
> 750	4.47

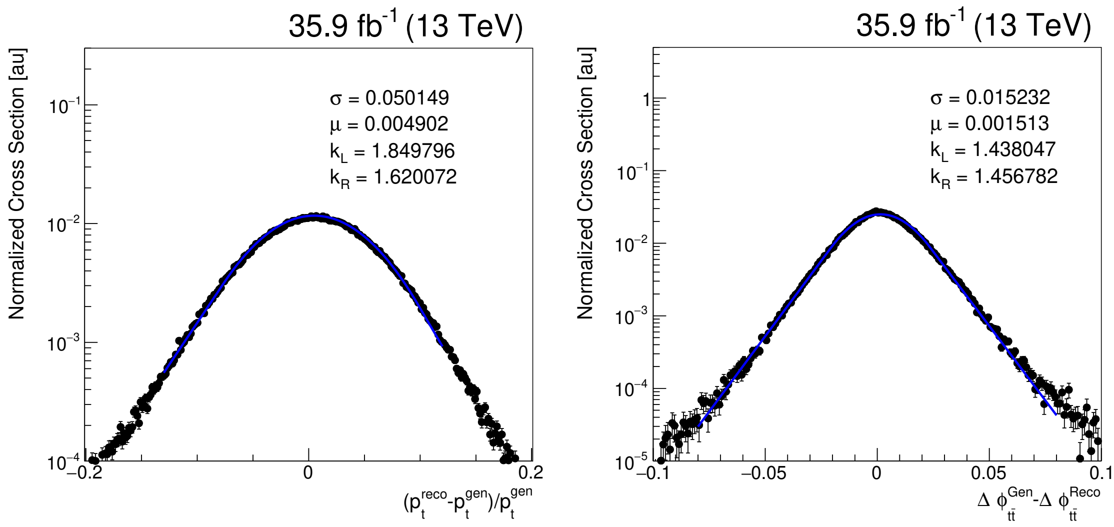


Figure 9.2: Resolution response with respect to: (left) p_T observable (given by equation 9.3), (right) $\Delta\phi$ observable (given by equation 9.2). The resolution response are fitted to a *ExpGaussianExp* (9.6) distribution with four free parameters, whose values after the fits are illustrated in each plot.

The resolution values related to $\Delta\phi$ observable are estimated, in different $\Delta\phi$ regions. Table 9.2 shows the results in six different $\Delta\phi$ intervals. A better resolution for $\Delta\phi$ closer to π is observed. The resolution changes from 1.02° in the region $\Delta\phi \sim \pi/2$ to values of 0.785° in the region $\Delta\phi \sim \pi$.

Table 9.2: Detector resolutions values related to the $\Delta\phi$ observable, in different $\Delta\phi$ intervals. The resolution are given by the equation 9.2.

$\Delta\phi(^\circ)$ Range	$\Delta R_{\Delta\phi_{t\bar{t}}}(^\circ)$
90 – 157.5	1.0233
157.5 – 163.8	0.9547
163.8 – 170	0.9540
170 – 174.3	0.943
174.3 – 177.1	0.9410
177.1 – 180	0.785

Additionally, the resolutions related to the soft drop jet mass (ΔR_{m_j}, equation 9.5) and to the η observable (ΔR_η, equation 9.4) were measured. The former one takes values of $\Delta R_{m_j} \sim 5.9\%$, while the latter one $\Delta R_\eta \sim 0.009$.

9.1.2 Purity, stability, background and acceptance.

In this subsection, the effects on the measured distributions caused by the finite detector resolution are studied in more detail. Thus effects are studied combining the Monte Carlo simulation at stable particle and detector level.

Several scenarios can be distinguished, affecting the measurements in different ways :

- ✓ *"fake events"*: wrong events can be selected at detector level. Those scenarios can be identified if a selected event at detector level doesn't have a corresponding event at stable particle level.
- ✓ *"miss events"*: true events are missed due to detector inefficiency. These scenarios can be identified if they are selected at stable particle level, but not at detector level.
- ✓ *"selected events"*: events selected at both levels might correspond to different bins at both levels. Those are known as migration bin-by-bin effects.

The detector effects can be classified in two categories, depending on their direct impact on the measurements. For each of those categories, different magnitudes are then defined in order to quantify them. The two categories are:

1. *Migrations occurring inside the fiducial phase space*: events are selected at both levels, but they fill different bins at stable particle and detector levels. This case corresponds to the third scenario previously described. The quantities, which can be defined for quantifying these effects are:

   - ✓ Purity (p_i), represents the fraction of the events at detector level in specific bin i that at stable particle level filled the same bin. This quantity is estimated for each bin of the distribution of a certain observable x, and can be written as:

   $$p_i = \frac{N(\mathit{both\ selected}, x_{det} \in \text{bin i}, x_{gen} \in \text{bin i})}{N(\mathit{both\ selected}, x_{det} \in \text{bin i})} \tag{9.7}$$

   - ✓ Stability (s_i), represents the fraction of the events at stable particle level in specific bin i, that at detector level filled the same bin. This quantity can be written as:

   $$s_i = \frac{N(\mathit{both\ selected}, x_{gen} \in \text{bin i}, x_{det} \in \text{bin i})}{N(\mathit{both\ selected}, x_{gen} \in \text{bin i})} \tag{9.8}$$

   Note: for both quantities the event need to be selected at both levels.

2. *Migrations into or outside of the fiducial phase space*: the event is selected at exclusively one of the two levels. Similarly as before, two magnitudes can be defined to quantify the effects on the measurements:

   - ✓ Background (b_i), represents the fraction of the events selected at detector level in specific bin i, that doesn't have a corresponding event in the same bin at stable particle level, also known as *"fake"* rate, since they are events probably wrongly selected. This quantity is estimated for each bin of the distribution of a certain observable x, and can be written as:

   $$b_i = 1 - \frac{N(\mathit{both\ selected}, x_{det} \in \text{bin i})}{N(\mathit{det\ selected}, x_{det} \in \text{bin i})} \tag{9.9}$$

✓ Acceptance (a_i), represents the fraction of the events selected at stable particle level in a specific bin i, that are also selected at detector level, corresponding to the same bin i. Also known as *"efficiency"* (or *"acceptance"*) rate, since represents the percentage of true events that after detector effects are still selected. This quantity can be written as:

$$a_i = \frac{N(both\ selected, x_{gen} \in \text{bin i})}{N(gen\ selected, x_{gen} \in \text{bin i})} \tag{9.10}$$

The observable (m) associated to this value, given by the formula $m = (1 - a)$ is known as inefficiency rate.

In all the equations previously defined, N refers to the total number of events, *"both selected"* means that the considered event was selected at both levels, *"det (gen) selected"* indicates that the event was selected at least at detector (stable particle) level.

Detector and *stable particle* selection criteria need to be applied in order to estimate the quantities given by the equations 9.7-9.10. Table 9.3 summarizes the selection criteria at both levels. The selection at detector level corresponds to the fiducial region (signal region). At stable particle level, the selection criteria are inspired by the top jet definition (see Chapter 3), considering events where the leading and subleading jets satisfy the top-jets requirement.

Detector level selection	stable particle level selection
Two jets with: $p_T > 400$ GeV, $\lvert\eta\rvert < 2.4$, $m_{sd} > 50GeV$	Two jets with $p_T > 400$ GeV, $\lvert\eta\rvert < 2.4$, $m_{sd} > 50GeV$.
$m^{lead}_{sub0} > 55$ GeV & $m^{sublead}_{sub0} > 59$ GeV	$m^{lead}_{sub0} > 40$ GeV & $m^{sublead}_{sub0} > 40$ GeV
$d_{MVA} > 0.45$	$\Delta R_{B-jet} < 0.4$
2-btagged jets	

Table 9.3: Selection criteria at stable particle and detector level. The detector selection criteria correspond to the signal region, while the stable particle selection criteria are the hadronic top jet definition. m^{lead}_{sub0} and $m^{sublead}_{sub0}$ refer to the masses of the subjet with higher p_T in the leading and subleading jets respectively. ΔR_{B-jet} is the distance in the $\phi - \eta$ plane between the B-hadron and the top jet candidate. The d_{MVA} is referred to the output discriminant of the multivariate analysis.

Due to migration effects, scenarios where the leading jet at detector corresponds to the subleading jet at stable particle level, can occur. Since for the quantities defined by equations 9.7−9.10, is important to avoid wrong assignments, a matching criterion is considered. Two jets are considered as matched if the ΔR (space distance in the $\eta - \phi$ plane) is taking values smaller than 0.3 (less than the half of the cone size).

Figure 9.3 shows the quantities defined by equations 9.7−9.10, applying the selection criteria listed in Table 9.3. The simulation of $t\bar{t}$ events is performed using the POWHEG+PYTHIA8 event stable particle and the full CMS detector simulation. The results are shown for different distributions: (upper plots) the p_T spectra of the leading and subleading jets, and (lower plots) for the $\Delta\phi$ observable, in the complete $\Delta\phi$ phase space region, and when a fine binning is considered in the most back-to-back region. The distributions correspond to the measurements in the inclusive fiducial phase space.

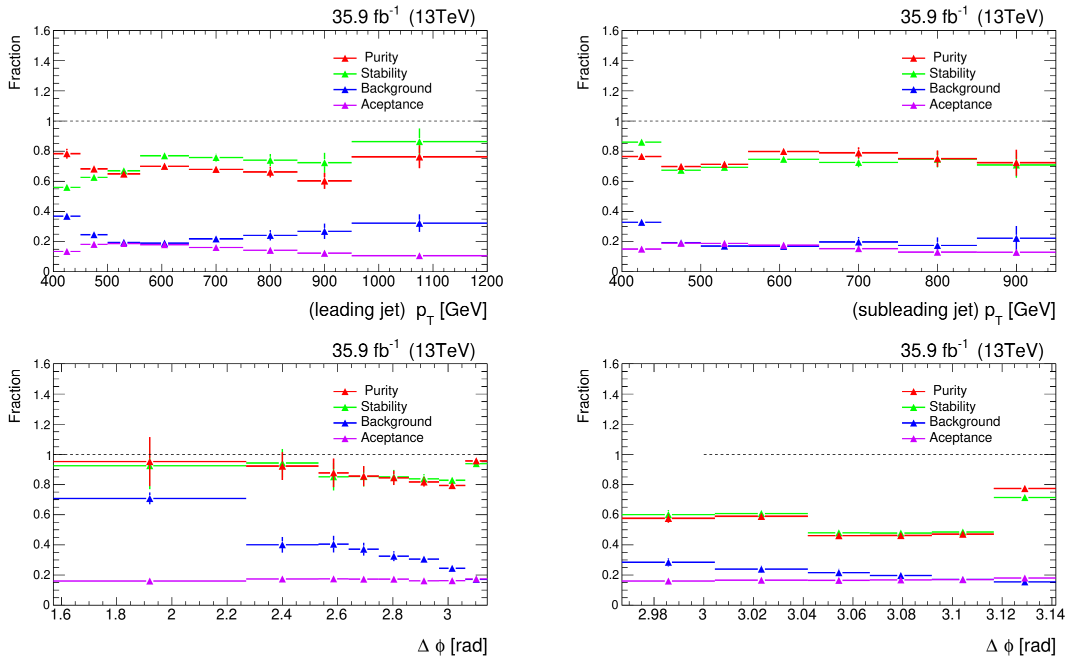


Figure 9.3: Purity, stability, background and acceptance in the fiducial - particle phase space, for the following observables: (upper left) p_T leading, (upper right) p_T subleading, (lower) $\Delta\phi$.

Purity and stability values, for all the observables, over the whole phase space, take relatively high values. In the case of $\Delta\phi$ observable, they are between 80%-95%, while for the p_T observable between 60%-90%. When smaller bin widths are considered for the $\Delta\phi$ observable, in the region $\Delta\phi \sim \pi$ (illustrated in the last plot in Figure 9.3), the purity and stability values are lower: 45%-80%, as a direct consequence of the selection of smaller bin widths. The bin widths are chosen carefully in order to have values of both, purity and stability, above 45%. By selecting the bin widths with this criterion, considerable migrations effects are avoided, and a better control on the unfolding procedure is expected.

Approximately 20% of the events selected at true level are effectively selected at detector level (acceptance values). The low values of the acceptance is due to the tight selection applied at detector level, in order to discriminate the QCD background. The acceptance values have an approximately flat behavior in the whole phase space (with respect to p_T and $\Delta\phi$).

The background values indicate that, in the case of p_T, around $20\% - 40\%$ of the events, selected at detector level don't have a correspondent event at particle level (fake events). In the case of $\Delta\phi$, the background values vary from 20%, in the region near to π, up to 70% in the $\Delta\phi \sim \pi/2$. Both magnitudes, background and acceptance are taken into account in the unfolding procedure as miss and fake contribution, respectively.

9.1.3 Response Matrices

Migrations effects can be also studied by defining response matrices. The response matrices provide the information about the correlation between stable particle and detector level. They are $2D$ histograms, filled by the stable particle and detector values of a certain observable, for events selected at both levels.

The response matrices are inputs needed for the unfolding procedure, therefore, in order to not introduce any bias, no matching criteria for the jets at both levels is applied. In other words, if an event is selected at both levels, the leading jet at stable particle level is combined with the information of the leading jet at detector level, and the same criteria for the subleading jet. A completely diagonal response matrix would mean that the stable particle and detector level information for the corresponding observable is fully correlated, and, that no migration between the bins, due to detector effects, are expected. Nevertheless, diagonal matrices are never foreseen, and the unfolding procedure corrects the non diagonal contributions. If the non diagonal terms have relative small contributions, a better convergence of the unfolding method can be anticipated.

The detector level information is usually given after applying the full selection (same criteria that is applied to the data). The stable particle level information which is provided depends to which level the data are unfolded. It can be given at parton level (i.e applying cut at the partonic level, for example on the top quark), or can be given at particle level (i.e applying certain cuts in terms of stable objects, for example, jets). In the analysis presented here, the response matrices correspond to the particle level. The selection criteria applied at both levels were given in Table 9.3.

Figure 9.4 shows the response matrices corresponding to $\Delta\phi$ observable. The left plot represents the response in the whole considered $\Delta\phi$ phase space ($\pi/2 < \Delta\phi < \pi$), while the next plot, represents the back-to-back region when fine bin widths ($\Delta\phi_{bin} \sim 0.017$ rad) are considered. Each of the rows are normalized to the total number of events at stable particle level belonging to the specific stable particle bin. In the first case, a diagonal behavior with almost no migration effects is obtained (over 83% of the events in each stable particle bin is reconstructed in the corresponding detector bin). In the case of the $\Delta\phi$ in the back-to-back configuration larger migration effects between bins can be observed. Nevertheless, over 48% of the events in all the bins at stable particle level is well reconstructed at the corresponding detector level bin.

Figure 9.5 shows the response matrices for the jet p_T of the leading and subleading jets. Each of the rows is normalized to the total number of events in the corresponding bin at stable particle level. In the case of the leading (subleading jet), between $62\% - 85\%$ ($69\% - 84\%$) of the events is properly reconstructed at the same bin as is generated.

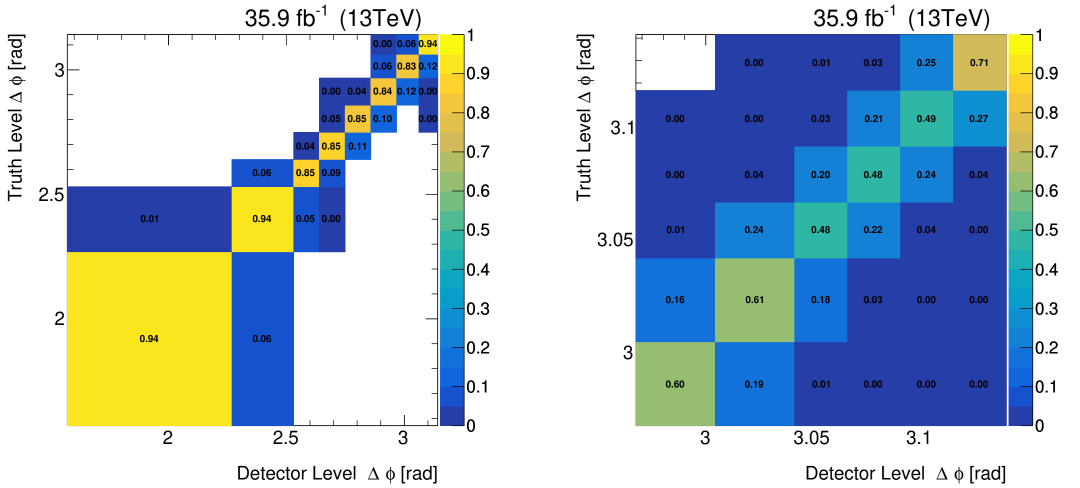


Figure 9.4: Response matrices for the $\Delta\phi$ observable in the inclusive fiducial phase space ($p_T > 400$ GeV, $|\eta| < 2.4$) taking into account different regions in $\Delta\phi$: (left) $\pi/2 < \Delta\phi < \pi$ (right) $2.96 < \Delta\phi < \pi$ (back to back region). The two last bins in the left plot correspond to the bins shown in the right plot. Each row has been normalized to the number of events in each corresponding stable particle bin.

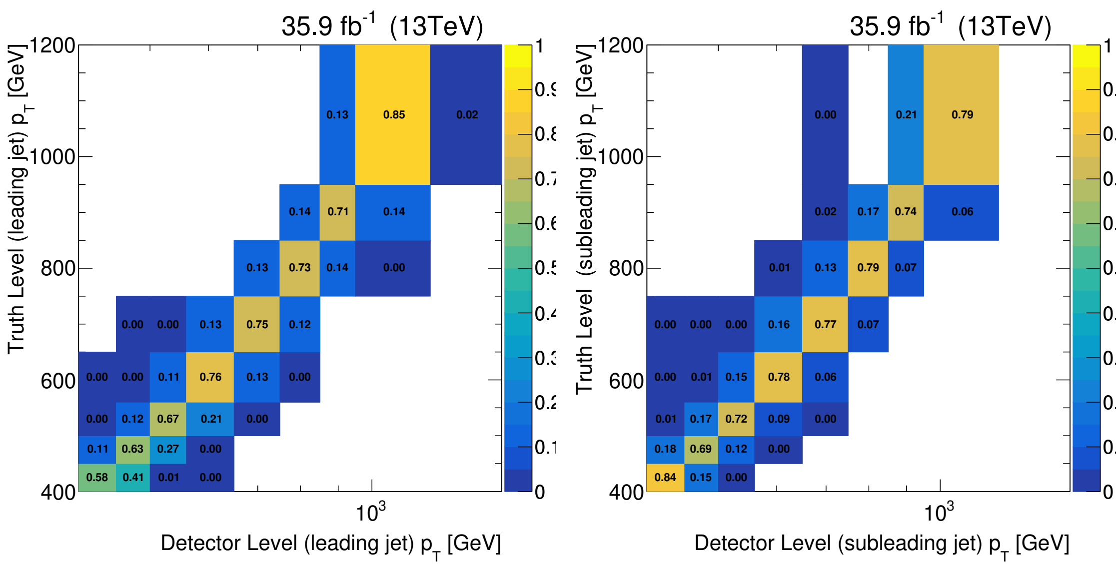


Figure 9.5: Response matrices for (left) p_T of the leading jet (right) p_T of the subleading jet. Each row has been normalized to the number of events in each corresponding stable particle bin.

9.2 Unfolding

In order to be able to compare the measured distributions to theory predictions, detector effects need to be taken into account. Detector effects are not only related to finite detector resolutions but also other stochastic processes occurring when particles travel through the detector material. One could think about incorporating a very precise detector simulation to the Monte Carlo predictions (given by the theory of the collisions), and then, comparing the outcome distributions to the measured ones. The results would be detector-dependent, meaning that it would be impossible to compare measurements between different experiments. The most convenient way of comparing measured distributions to theory predictions is to correct the estimated detector effects. This procedure is known as unfolding, and are explained in detail in this section.

Figure 9.6 shows a sketch of the unfolding procedure. The main idea is to provide a measurement equivalent to the stable particle level. The input information is provided through the response matrices previously described (see Section 9.1) given by $t\bar{t}$ simulated events.

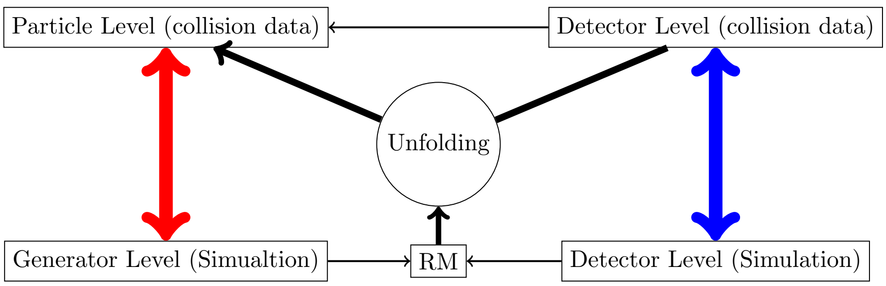


Figure 9.6: Sketch showing the unfolding procedure: the response matrix (RM) is obtained from the simulation of both levels (particle level and detector level distributions). Based on this input information, the unfolding method corrects the collision data at detector level to the data distribution at the corresponding stable particle level.

This section is dedicated to discuss different unfolding methods. These procedures are then applied to the measured distributions (transverse momentum spectra and the spectra of the azimuthal separation between top jets), after the QCD background has been subtracted. This section is organized as follows: in the subsection 9.2.1 the description of some of the available unfolding methods is provided. In subsection 9.2.2 the application of one of the described method (D´Agostini method [168]) to the measured distributions is presented. Other unfolding methods are used in order to provide closure tests to the default method, and to study possible biases introduced by the implemented method. The closure test studies are provided in the Appendix G.

9.2.1 Different Unfolding methods.

In this section several unfolding methods are briefly explained. The unfolding problem can be mathematically expressed as function of the *truth* distribution of the data (the one which is the outcome of the procedure) (x_i), and the actual measured distribution (y_i) as follows:

$$y_j^{data} = \sum_{i=1}^{m} A_{ji} x_i + b_j, \tag{9.11}$$

where A_{ji} terms represent the response matrix, given by the exact probability of migrations between bins. All the methods further described are aimed to find the solutions for this problem, given by the components x_i.

Correction Factor Method.

The Correction Factor Method (also known as *Bin by Bin Correction Method*), is the simplest of the unfolding methods, but only valid in limited cases, i.e when migrations effects are small. Basically, by using this method, the corrected quantity (x_i) is obtained by modifying the measured quantity by a multiplicative factor estimated in the simulation as follows:

$$x_i = y_i^{data} \frac{N_i^{gen}}{N_i^{det}}, \tag{9.12}$$

where y_i^{data} is the measured quantity in the data, N_i^{gen} corresponds to the respective magnitude estimated in the simulation, at stable particle level, while N_i^{det} corresponds to detector level in the simulation.

This method assumes that all the events are reconstructed at the same bin that were originally generated in the simulation. In other words, the purity, defined by equation 9.7, is equal to unity. A more realistic approximation to the cases where migrations effects occur, within the fiducial phase space, would be to considered the unfolded magnitude as follows [170]:

$$x_i = N_i^{gen} \frac{y_i^{data} - (N_i^{rec} - N_i^{rec\&gen})}{N_i^{rec\&gen}} = N_i^{gen} \frac{y_i^{data} - N_i^{det}(1 - P_i)}{N_i^{det} P_i}, \tag{9.13}$$

where P_i is the magnitude known as purity ($P_i = N_i^{rec\&gen}/N_i^{rec}$) corresponding to a specific bin. The last equation is derived by subtracting those events which were reconstructed, but do not have an equivalent generated event in the same bin from the events initially contained in a specific bin ($y_i^{data} - (N_i^{rec} - N_i^{rec\&gen})$), and by considering in equation 9.12 the denominator factor (N_i^{det}), only those events that have an equivalent event at stable particle level ($N_i^{rec\&gen}$). In the last expression, however, migrations to inside or to outside of the fiducial phase space have not been considered.

D´Agostini method

The D´Agostini unfolding method [171] is motivated in Bayesian theory. The original method strongly depends on the Monte Carlo simulation used to provide the response matrix. This model dependence can be reduced by iterating the method having as input the results of the previous step. The iterative solution to this problem [168] is the unfolding method used in this thesis.

The unfolded magnitude in a specific bin i can be estimated as:

$$x_i = N_i^{gen} \sum_j \frac{A_{ji}}{\epsilon_i} \frac{y_j^{data}}{N_i^{rec}}, \tag{9.14}$$

where A_{ji} is the response matrix which represents the probability that a given event generated in the bin i, will be reconstructed at a specific bin j, and ϵ_i represents the efficiency (acceptance), i.e. the probability that an event selected at stable particle level, will have an equivalent event at detector level. Both magnitudes can be written as following:

$$A_{ji} = \frac{N_{ji}^{rec\&gen}}{N_j^{gen}}, \quad \epsilon_i = \sum_j A_{ji} \quad , \tag{9.15}$$

where the index j refers to the reconstructed level and i to the stable particle level.

The acceptance and the response matrix are estimated from the Monte Carlo simulations (already discussed in sections 9.1.2, 9.1.3 respectively).

The iterative method is then considered as follows:

$$x_i^{iter+1} = x_i^{iter} \sum_j \frac{A_{ji}}{\epsilon_i} \frac{y_j^{data}}{\sum_k A_{kj} x_k^{iter}} \quad , \tag{9.16}$$

When a sufficient number of iterations are performed, this method can be considered as a good solution of the unfolding problem. One of the difficulties relies on the propagation of the statistical uncertainty through the iterations. Dedicated studies on the performance of the method with respect to the number of iterations, for the specific distributions studied in this thesis, are provided in section 9.2.2.

Singular Value Decomposition (SVD) method

The Singular Value Decomposition (SVD) unfolding method [172] is a matrix inversion unfolding method with a regularization parameter to prevent statistical fluctuations.

The idea of this method is first of all, to decompose the response matrix A in a diagonal and two orthogonal matrices in the following way:

$$A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T, \tag{9.17}$$

where Σ is a diagonal matrix with ordered singular values $\sigma_1.....\sigma_n$ ($\sigma_1 \leq \sigma_2....\sigma_n$), and U,V^T are orthogonal components. The direct problem (folding distribution) can be expressed in terms of the response matrix (A), and the vectors $\mathbf{y}$ and $\mathbf{x}$ corresponding to linear vectors associated to each bin content for the measured distribution:

$$\mathbf{y} = A\mathbf{x}\,. \tag{9.18}$$

The outcome of the unfolding procedure corresponds to the unfolded data at stable particle level. With the decomposition of the response matrix, the last expression can be written in the following way:

$$\mathbf{y} = \sum_{i=1}^{n} \sigma_i (v_i^T \mathbf{x}) u_i\,, \tag{9.19}$$

The last expression means that the true vector $\mathbf{x}$ is decomposed in components $v_i^T\mathbf{x}$. Hence, the measured vector $\mathbf{y}$ is a superposition of the vectors u_i, weighted with the values given by the components σ_i.

The inverse problem (unfolding) is to estimate the vector $\mathbf{x}$. This step can be expressed using the previous equations as follows:

$$\mathbf{x} = A^T\mathbf{y} = V\Sigma^{-1}(U^T\mathbf{y}) = \sum_j^n \frac{1}{\sigma_j}(u_j^T y)v_j \tag{9.20}$$

By this definition, which relies on the matrix inversion problem, it is clear that certain convergence issues are expected. First of all, when estimating the singular values σ_j of the response matrix A, small values could appear, causing an unstable behavior of the estimated vector $\mathbf{x}$. In addition, the statistical errors of each of the components of the vector $\mathbf{y}$ are usually not equal, which could lead as well to an unstable solution to the unfolding method.

The SVD unfolding method faces this problem on the basis of a regularization approximation. The problem is then reduced to the following minimization problem:

$$(A\mathbf{x} - \mathbf{y})^T(A\mathbf{x} - \mathbf{y}) + \tau(C\mathbf{x})^T C\mathbf{x} = min \tag{9.21}$$

where C now is a priory defined matrix affecting the solution of the problem, and τ is the regularization parameter. Often, the matrix C is selected as $C_{ij} = \delta_{ij}$. An optimal regularization parameter need to be determined being specific for each distribution under study. A too small τ value could lead to oscillations of the solution, while too large τ values could lead to a bias solution towards the Monte Carlo simulation.

TUnfold Algorithm.

The TUnfold algorithm [173] is based on a least square minimization problem using the Tikhonov regularization [174]. The problem can be written as follows:

$$\chi^2(x) = (A\mathbf{x} - \mathbf{y})^T V_{yy}^{-1}(A\mathbf{x} - \mathbf{y}) + \tau^2(\mathbf{x} - \mathbf{x}_b)^T(L^T L)(\mathbf{x} - \mathbf{x}_b) \tag{9.22}$$

where, $\mathbf{x}$, $\mathbf{y}$ and A are the same quantities already defined in this section, while V_{yy} is the covariance matrix of the measured distributions $\mathbf{y}$. The second term represents the regularization term, where τ is the regularization strength, L is a matrix depending on the regularization conditions and $\mathbf{x}_b$ counts for the regularization bias. The matrix L is taken as the unity matrix, and x_b is taken as zero at the beginning.

However, the regularization term is not considered in this thesis (when this method is used for comparing results of different unfolding algorithms). Without this term, stable unfolded results are obtained, meaning that introducing it, would rather increase the chances of considering certain bias through the unfolding implementation.

9.2.2 Unfolding the fiducial measurements from detector level to stable particle level

The iterative D´Agostini method [168] is applied to the measured distributions: p_T spectra and $\Delta\phi$ spectra. The stable particle level to which the distributions are unfolded, is given in Table 9.3.

In order to study the effect of the number of iterations on the unfolded distributions, the procedure is performed by varying this parameter. The unfolded distributions are expected *ideally* to become closer to the "*truth*" (particle) level distribution, when more iterations are considered. Nevertheless, when the number of iterations increases, the unfolded distribution might start to be biased by the method. The optimal value of iterations needs to be estimated, based on the stability of the outcome distributions.

In order to estimate the optimal number of iterations, and to understand the effect of each iterative step on the unfolded distributions, the following aspects are taken into account:

- ✓ for each additional iteration the unfolded distribution and the stable particle level distribution (from the Monte Carlo simulation) are compared. The criterion for the comparison is the χ^2 estimator given by the following equation:

$$\frac{\chi^2}{N_{bins}} = \frac{1}{N_{bins}} \sum_{i=1}^{Nbins} \left(\frac{\sigma_i^{gen} - \sigma_i^{unf}}{\sqrt{(\epsilon_i^{gen})^2 + (\epsilon_i^{unf})^2}} \right)^2 \quad , \tag{9.23}$$

  where in the sum each of the bins of the specific distribution ($Nbins$) is considered, σ_i is the cross section predicted for each bin, either in the distribution at stable particle level (gen), or at unfolded level (unf). The parameter ϵ_i corresponds to the error in the specific bin.
  An essential point to notice is that decreasing χ^2 value, when more iterations are considered, could be obtained when the error of the unfolded distribution increases. The optimal number of iterations would correspond to the lower χ^2, but avoiding a decrease of χ^2 driven by the increase of the statistical uncertainty.

- ✓ additionally, for each iteration, the ratio between the relative statistical uncertainty (in percentage) between the unfolded distribution, and the data distribution before unfolding are estimated. This ratio is calculated for each specific bin i as follows:

$$r_i = \frac{\epsilon_i^{data} / \sigma_i^{data}}{\epsilon_i^{unf} / \sigma_i^{unf}} \quad . \tag{9.24}$$

  where ϵ_i^{data} is the uncertainty in the data distribution before unfolding. As a requirement, the relative uncertainty in the unfolded distribution need to be larger than the errors in the data distribution. If with a certain number of iterations, this isn't satisfied, then more number of iterations are needed.

The outcome of this procedure depends on the observable and on the migration effects reflected in the response matrices. Therefore they are estimated for each measured distribution.

In the Appendix G, closure tests to the unfolding procedure are provided. The tests are performed testing different unfolding methods, previously described in this Chapter, and additionally testing different Monte Carlo predictions to provide the input information to the unfolding.

Unfolding the p_T spectra

Figure 9.7 shows the reduced χ^2 and the statistical ratio (r) quantities (defined by equations 9.23 and 9.24), for the p_T spectra of the leading and subleading jets. The distributions of the p_T spectra in the inclusive fiducial phase space, as well as the exclusive regions defined by the η observable, have been considered. The r quantity is presented in the first p_T bin (400 GeV$< p_T <$ 450 GeV). In all cases, the decreasing behavior of χ^2, with respect to the number of iterations is observed. Similar behavior is observed for the r observable, indicating that one of the reasons for the continuously decreasing χ^2 might be given by the increase of the statistical uncertainty of the unfolded distribution.

For example, in the limit when 600 iterations are considered, in most of the represented cases, the statistical uncertainty in the unfolded distribution are approximately 10 times larger than in the data ($r \sim 0.1$). Stables results are observed after 10 iterations.

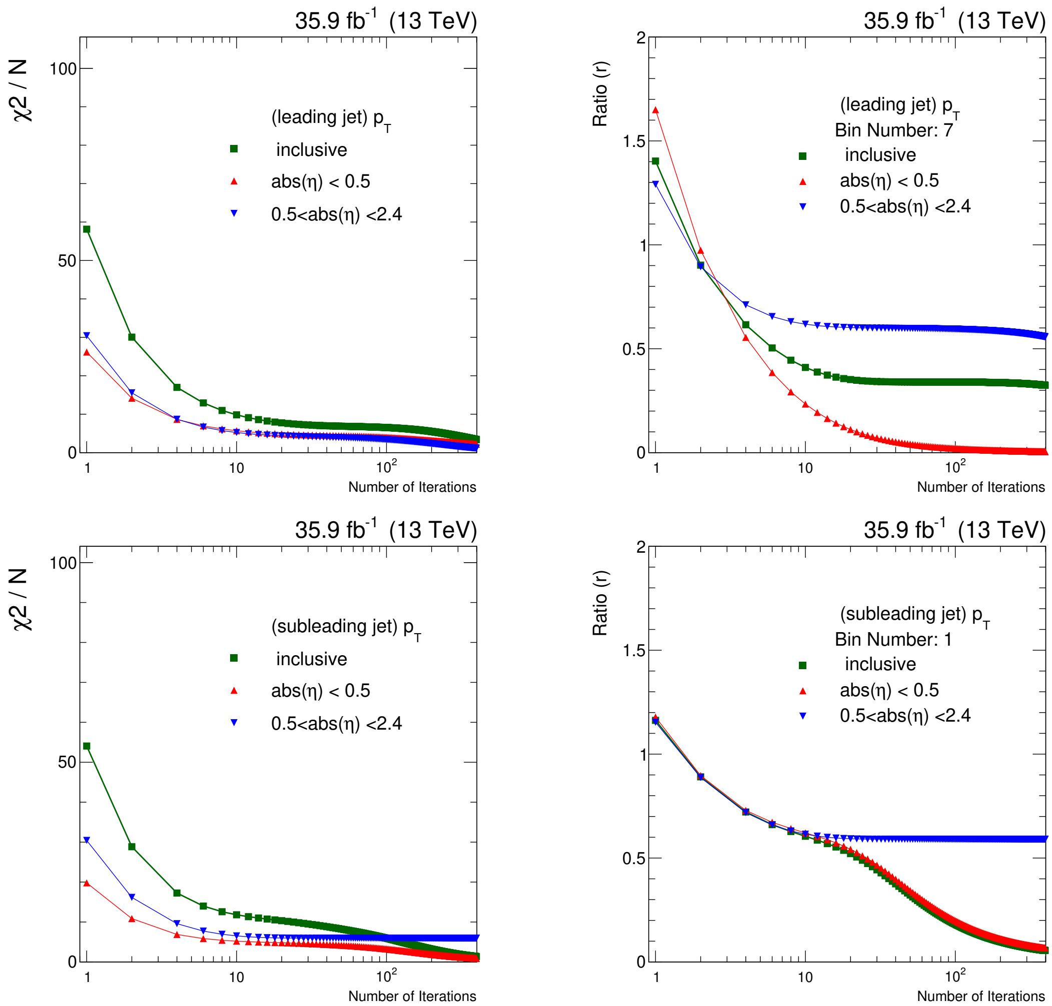


Figure 9.7: (left) χ^2 (equation 9.23) distribution and (right) r (equation 9.24) distributions as function of the number of iterations of the unfolding procedure, performed for the p_T spectra of the leading jet (upper plots), subleading jet (lower plots). The inclusive fiducial phase space, and the exclusive regions defined with $|\eta|$ observable are shown in each plot.

Figure 9.8 shows the comparison between unfolded p_T spectra, considering different number of iterations of the D´Agostini method [168], for the leading (upper plots) and subleading jets (lower plots), for different regions in η ($|\eta| < 2.4$ and $0.5|\eta| < 2.4$). In each plot, the distributions at detector and stable particle level in the Monte Carlo simulations are shown. Additionally the measured distribution in data, and the unfolded distributions (considering different number of iterations) are illustrated. The considered number of iterations are: 1, 10 and 20. In the lower part of each plot, the following ratios are shown:

- the *Data-MC* ratio: the ratio between the data and Monte Carlo predictions at detector level; and the ratio between the unfolded distribution (considering 10 iterations) and Monte Carlo predictions at particle level.
- the *Ratio Unfolded*: the ratio between each of the considered unfolded distribution (changing the number of iterations) and the distribution at stable particle level given from the simulation.

From the results shown in Figure 9.8 it can be noticed that the ratio between Data and Monte Carlo predictions at both levels (detector and stable particle level) are similar, meaning that no large migrations effects are present. Additionally, it is checked that the result given by the unfolding procedure is stable.

Unfolding the $\Delta\phi$ spectra

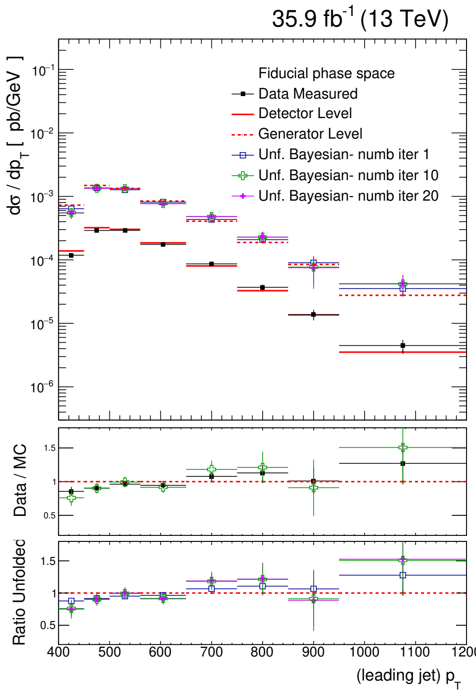
In the following, the unfolding method is applied to the differential distributions in the azimuthal separation between the two leading jets ($\Delta\phi$). The migrations effects are expected to be smaller compared to the p_T spectra, because the response matrices are more diagonal.

Figure 9.9 shows the reduced χ^2 and the ratio r for the $\Delta\phi$ distributions. The plots correspond to the whole $\Delta\phi$ region (upper plots) and the most back-to-back region (lower plots), when the fine binning is considered. In the former case in addition to the inclusive fiducial phase space region, exclusive regions are considered: 400 GeV $< p_T^{lead} <$ 600 GeV, $p_T^{lead} >$ 600 GeV. The quantity r is estimated for the last bin, corresponding to the region $\Delta\phi \sim \pi$, where most of the events are.

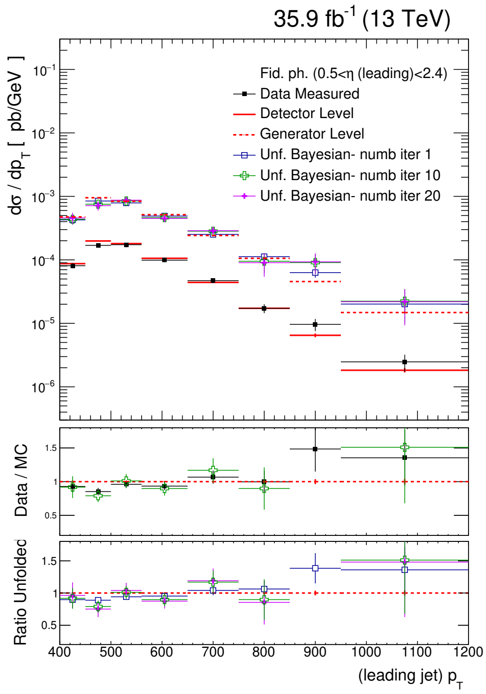
After a certain number of iterations, the studied values χ^2 and r are basically constant. After $\sim$ 10 iterations, no significantly change is observed neither for the central values of the unfolded distributions, nor for the statistical uncertainty.

Figure 9.10 shows the comparison between simulation and data at stable particle level and detector levels for the $\Delta\phi$ observable.

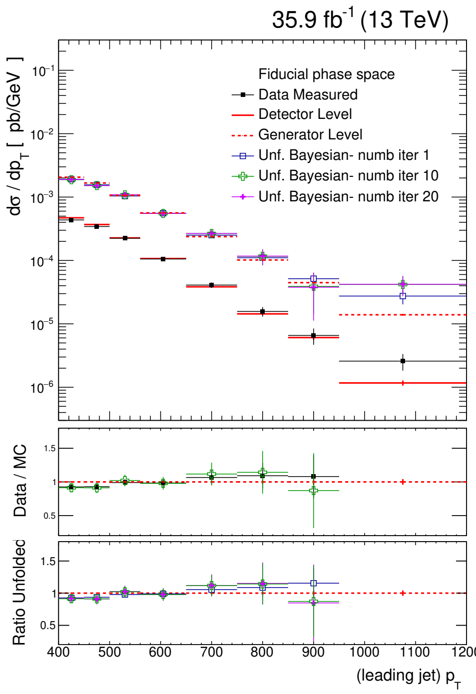
At least two iterations in the unfolded method are required in order to fulfill the condition: $r < 1$. Stable results are reached after 10 iterations, which is further considered to perform the unfolding for these distributions.



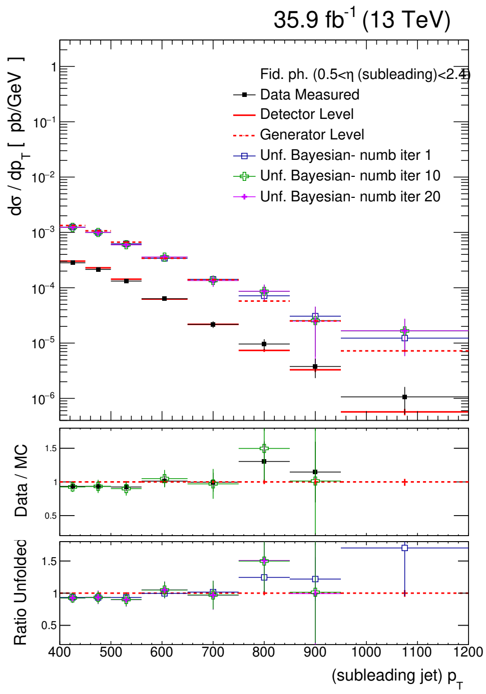
(a)



(b)



(c)



(d)

Figure 9.8: Comparison of the distributions given by the Monte Carlo simulation and measurements at both levels: stable particle level and detector levels. The data at stable particle level is obtained by unfolding the data with the D´Agostini method, and changing the number of iterations: 1, 6, 10, 500. (a) p_T of the leading jet in the inclusive phase (b)p_T of the leading jet in the exclusive region $0.5 < |\eta^{lead}| < 2.4$ (c) p_T of the subleading jet in the inclusive phase (d) p_T of the subleading jet in the exclusive region $0.5 < |\eta^{sublead}| < 2.4$. The $Data/MC$ ratio refers to the ratio between the data distributions and the Monte Carlo simulations at either stable particle level or detector levels. The $Ratio\ Unfolded$, refers to the ratio between the unfolded distribution and the distribution at stable particle level. The Monte Carlo distributions are scaled by a factor of 0.70.

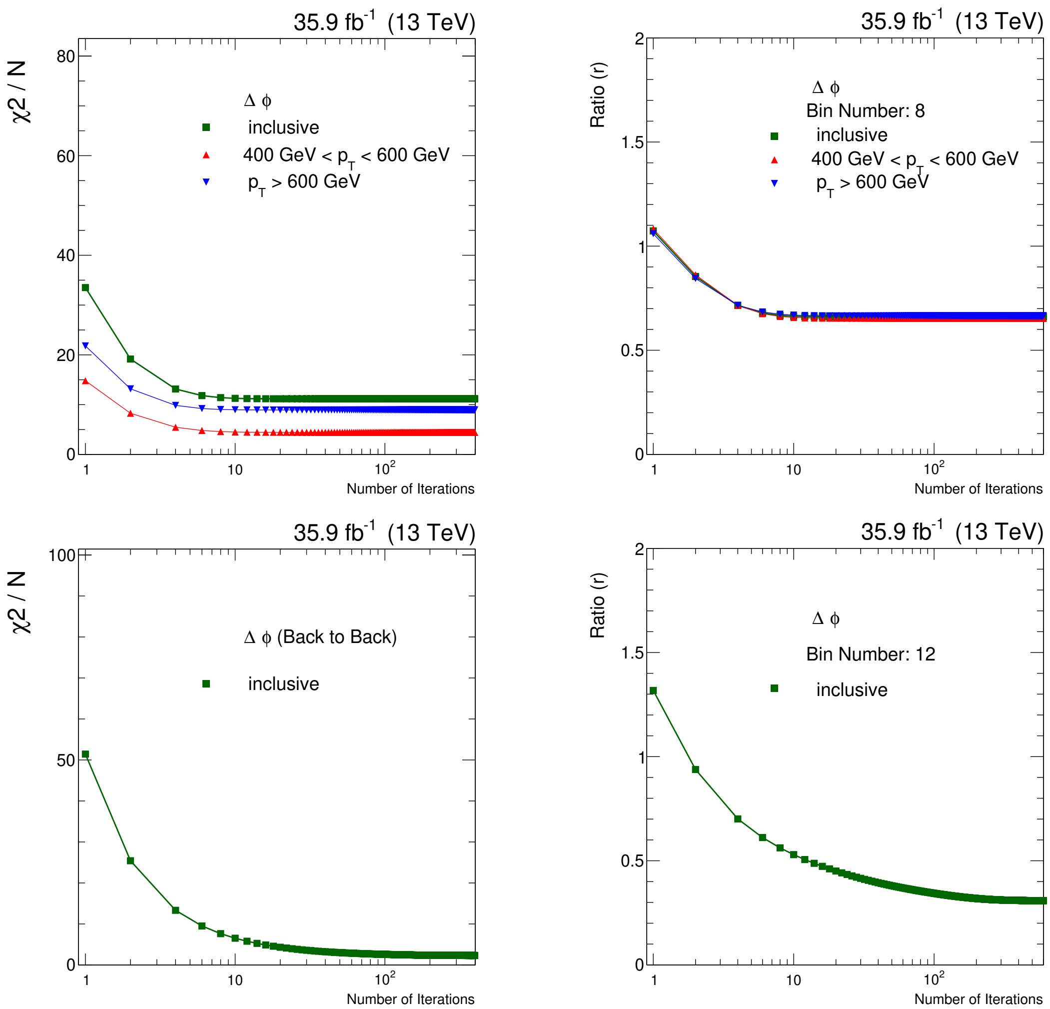
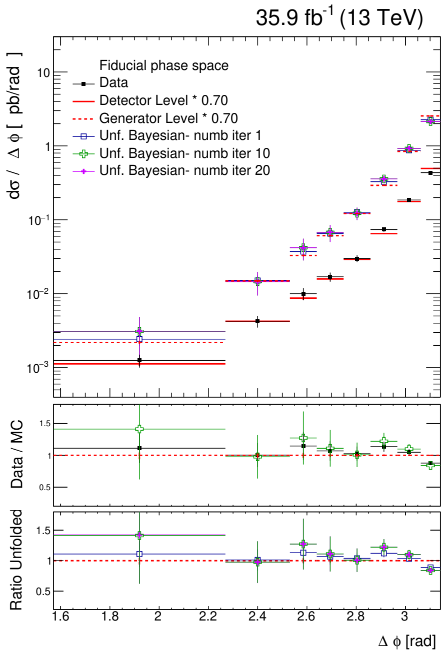


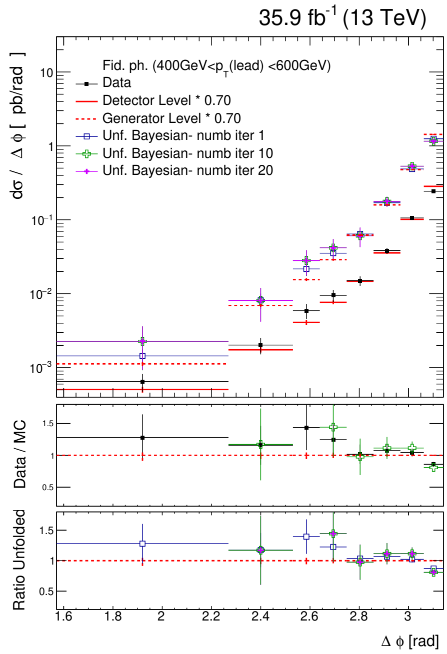
Figure 9.9: (left) χ^2 (9.23) distribution and (right) r (9.24) distributions as function of the number of iterations of the unfolding procedure for $\Delta\phi$ spectra. The whole $\Delta\phi$ phase space (upper plots), and a fine binning in the most back-to- back region (lower plots) are illustrated. For the plots, the inclusive fiducial phase space and exclusive regions defined by the p_T^{lead} are considered.

Unfolding the measurements including the QCD multijet contribution as part of the signal.

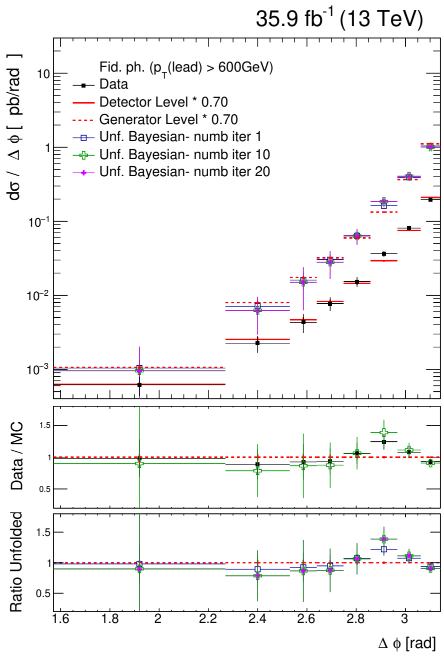
Until now, the unfolding procedure was applied to the $t\bar{t}$ measurements. Therefore, all the Monte Carlo samples are obtained from $t\bar{t}$ simulated events. In the case of the measurements in which the QCD multijet contribution is included as part of the signal, the input information to the unfolded procedure is provided by the sum of the $t\bar{t}$ and QCD contribution simulated with the POWHEG+PYTHIA8 particle and MADGRAPH+PYTHIA8 particle, respectively. The results obtained are similar, since the migrations effects are independent to the source of events. However, it is needed to consider both contributions in order to consider the correct stable particle level at which the data is unfolded.



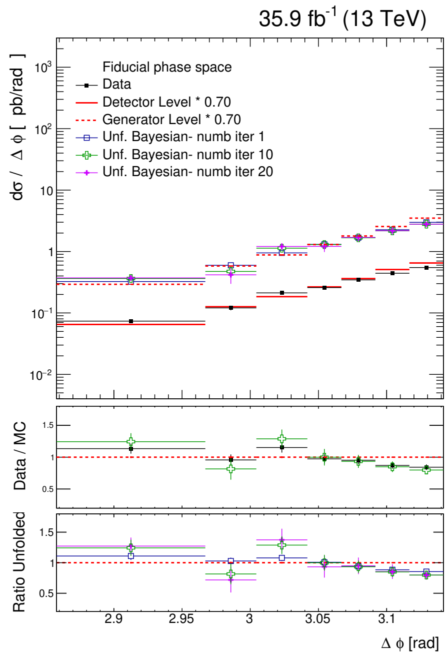
(a)



(b)



(c)



(d)

Figure 9.10: Comparison of the $\Delta\phi$ distributions given by the Monte Carlo simulation and measurements at both levels: stable particle level and detector levels. The stable particle level distributions are obtained by unfolding the data with the D´Agostini method, changing the number of iterations: 1, 1, 4, 10, 80. (a) $\Delta\phi$ in the inclusive phase (b) $\Delta\phi$ in the exclusive phase space given by the region 400GeV $< p_T^{lead} <$ 600GeV, (c) $\Delta\phi$ in the exclusive phase space given by the region $p_T^{lead} >$ 600GeV (d) $\Delta\phi$ in the inclusive phase , when the refined binning in the most back to back region is considered. The *Data/MC* ratio refers to the ratio between the data distributions and the Monte Carlo simulations at either stable particle level or detector levels. The *Ratio Unfolded*, refers to the ratio between the unfolded distribution and the distribution at stable particle level. The Monte Carlo distributions are scaled by a factor of 0.70.

Chapter 10

Systematic Uncertainties

The measured distributions can be affected by uncertainties from systematic effects. Those uncertainties are known as systematic errors reflecting the inaccuracy of the measurement. They can be grouped in two categories: experimental and modeling uncertainties. The former ones are those related to detector resolutions and to performance of some algorithms, which could have a different behavior in Monte Carlo simulations and data. The modeling uncertainties are related directly to the simulations used for estimating the migration effects due to detector resolutions. They are usually handled by changing parameters in the simulations at detector level.

The chapter is divided into three main sections. The first section (10.1) is focused on presenting the experimental uncertainties, while on the second one (10.2) the modeling uncertainties are studied. In the last section (10.3), a summary of the combined uncertainties is provided.

The different sources of uncertainties are:

- ✓ Experimental uncertainties: the jet energy scale uncertainties (`JES Unc`), the jet energy resolution uncertainties (`JER Unc`), the uncertainties related to the b-tagging algorithm (`Btag Unc`), the uncertainties related to the background estimation and subtraction (`Backg Unc`) [1], uncertainties related to the measured luminosity.
- ✓ Modeling theory uncertainties: uncertainties related to the modeling of all the Underlying Events (`UE Unc`), uncertainties related to the modeling of the Hard Scattering process (`HS Unc`), and uncertainties related to the top-quark mass used in the simulation for $t\bar{t}$ events.

Both mentioned groups of uncertainties are estimated in different ways. The modeling uncertainties are taken into account through the unfolding procedure, where the input information (acceptance, background and response matrices) is provided to determine migration effects. For each considered parameter variation, a new unfolding is performed using the new input information, and the difference with the default unfolded distribution is taken as uncertainty. In the case of the

[1] The uncertainties introduced through the background subtraction procedure are considered as experimental uncertainties since data driven methods are applied.

experimental uncertainties, they are taken into account in the whole procedure, meaning that, for each source of uncertainties, a new background contribution is estimated. After a new detector level distribution has been estimated after the background subtraction, the unfolding procedure is then repeated, taking into account also the corresponding variations in the simulation [1].

Since two main measurements are performed, in the following, the treatment of the systematic uncertainties for each of them is briefly explained:

- $t\bar{t}$ **measurements**: Figure 10.1 shows a sketch of how the uncertainties are considered. The way that they are considered depends on if the measurement is at detector level or at the stable particle level.
  - ✓ *Fiducial detector level measurements*: the systematic uncertainties are taken into consideration through the background subtraction procedure. The Monte Carlo simulations used to estimate and to subtract the background are affected by: `JES Unc`, `JER Unc`, `BTag Unc`. The procedure of subtracting the background has also uncertainties (for instance in the fitted yield). The uncertainties related to the measured luminosity is also considered in the measured distributions.
  - ✓ *Stable particle level measurements*: for each of the following sources: `JEC Unc`, `JER Unc`, `BTag Unc`, new unfolded measurements are performed [2]. In the case of the uncertainties related to the background estimation, and the luminosity uncertainties, the new detector-level measurements are unfolded taking into account the nominal response matrices. In order to estimate the modeling uncertainties, the nominal detector-level measurements are unfolded taking into account the new response matrices obtained for each source of theory uncertainties (`UE Unc`, `HS Unc`, t`-quark mass Unc`).
- **Inclusive top-jet measurements**: analogously to the previous case, in Figure 10.2, a sketch illustrates the systematic uncertainties in those measurements. The two measurements at detector level and stable particle level are considered as well.
  - ✓ *Fiducial detector level*: in this case, since no background subtraction is performed (no simulation needed for this measurement), only two main sources are affecting the data distributions: `JEC Unc.`, and the luminosity uncertainties.
  - ✓ *Particle level measurements*: in this case, first, the new distributions obtained by considering the `JES Unc.` are unfolded, considering the respective response matrices. The nominal fiducial measurements are unfolded considering new response matrices estimated by applying the `BTag Unc.`, the `JER Unc`, and all the modeling uncertainties: `UE Unc`, `HS Unc`, t`-quark mass Unc`. Finally the systematic uncertainties at the detector level measurements obtained by considering the luminosity uncertainties are propagated to the particle level measurement performing the unfolding, considering the nominal response matrices.

[1] for those uncertainties affecting the Monte Carlo simulated events through the response matrices.

[2] the detector level distributions obtained by the consideration of the respective uncertainties in the background subtraction procedure, are the distributions that will be unfolded. Additionally, the response matrices also consider the respective variations.

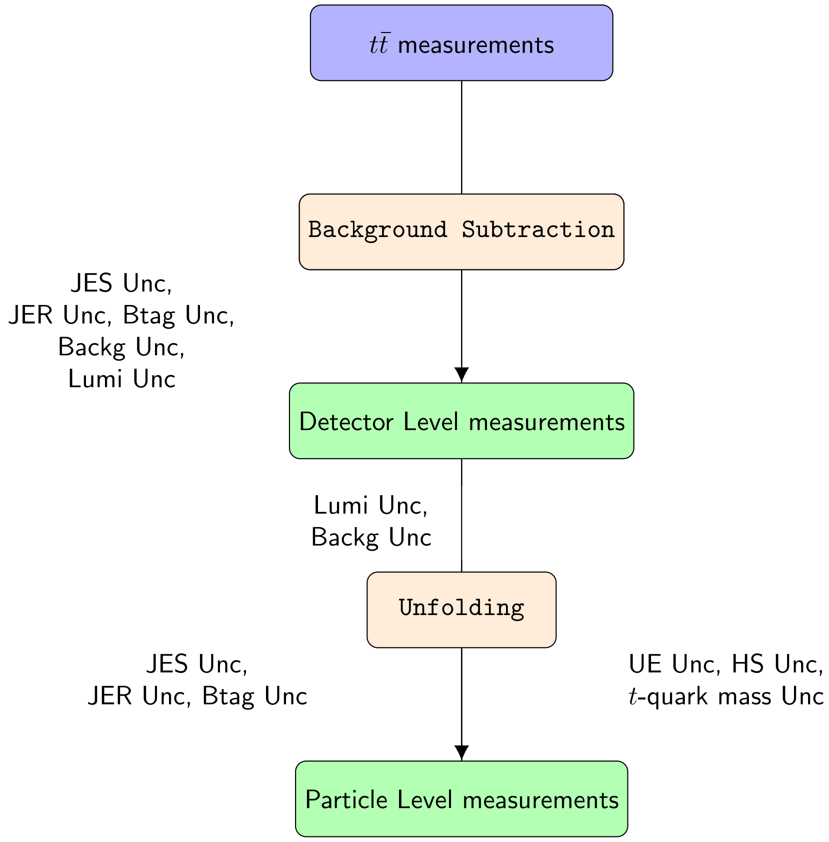


Figure 10.1: Sketch representing the systematic uncertainties in the measurement when the QCD background contribution has been subtracted.

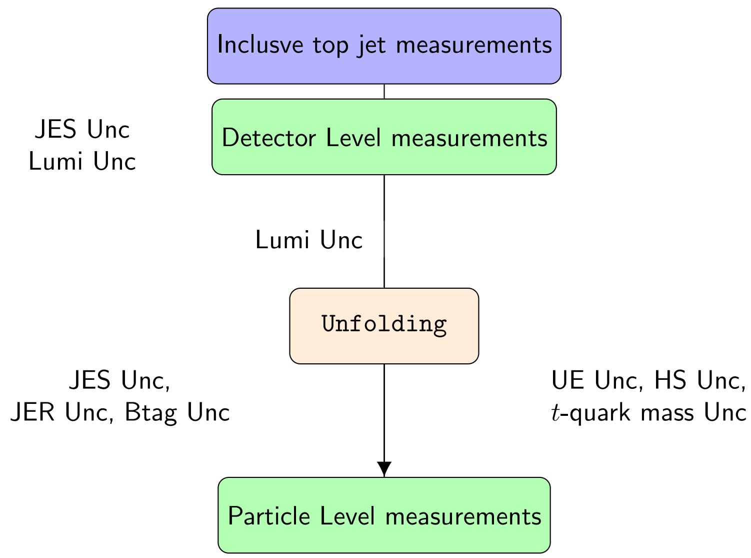


Figure 10.2: Sketch representing the systematic uncertainties in the measurement when the QCD multijet events are considered as part of the signal.

10.1 Experimental Uncertainties

In subsection 10.1.1 the way that the experimental uncertainties are treated in the thesis is explained. Other subdominant sources of uncertainties, like for example the ones related to the pileup modeling and to the modeling of the trigger efficiencies were also considered and found to be well below 1%. In the subsection 10.1.2, the results of the main experimental systematic uncertainties are presented. In the case of the uncertainties related to the background estimation and subtraction procedure only the estimated quantities are given in this section, since they are already discussed in Chapter 8,

10.1.1 Handling of the experimental uncertainties.

Uncertainties related to the jet energy corrections (`JES Unc.`)
The jets are calibrated by applying the jet energy scale factors (JES) (already explained in section 5.3). Those corrections are applied on a jet-by-jet basis, in Monte Carlo simulations and in data. The corrections are centrally provided by the experiment, as well as their uncertainties. The effect of the uncertainties related to the corrections on the specific measurements, need to be studied.

In section 5.3, the uncertainties of the corrections, as a function of p_T and η of the jet, have been presented. The uncertainties are provided in 24 independent sources. The total uncertainties have been considered (the quadratic sum of those 24 sources). In the phase space where this analysis is performed ($\eta < 2.4$, $p_T > 400$ GeV) the JES uncertainties are of the order of $1\% - 2\%$.

The uncertainties are implemented on a jet-by-jet basis, smearing the transverse momentum p_T of each jet (the two leading jets):

$$p_T \rightarrow (1 \pm \sigma_{JES}) p_T \, , \tag{10.1}$$

where σ_{JES} is the provided uncertainty for the specific jet. The $\pm$ sign, refers to either smearing to higher values (up), or to lower values (down).

The smearing can be either applied to the data or to the Monte Carlo simulation. In this specific case, it is applied to the data, since the background subtraction procedure is insensitive to the smearing applied to the Monte Carlo simulations (the background subtraction is a data driven method).

By smearing the p_T of the jets, not only the shape of the measured distribution might be changed, but also the total number of selected events (some events that were selected might be rejected and vice-versa). Those variations also give the possibility that the top jet candidates are not anymore the leading and subleading jets: for instance the subleading jet could be after the smearing the third p_T jet, and is not considered as top jet candidate anymore. However, in the fiducial phase space (after all the selection criteria have been applied), such scenarios are negligible. Therefore, the top jets candidates are still considered as the leading and subleading jets.

It is known that the differential inclusive cross section decreases as function of p_T::

$$\sigma(p_T) \propto \frac{1}{p_T^5} . \tag{10.2}$$

Therefore, uncertainties on the p_T of the order of 2% could be translated to a 10% uncertainty on the measured cross section [1].

The uncertainty related to this source is estimated in a bin-by-bin basis of each specific distribution, taking the maximum deviation given by the *up* and *down* variations with respect to the nominal distribution. The new measurements, corresponding to each of the considered variations, are then unfolded to particle level using the response matrices provided by the Monte Carlo simulation by considering the p_T smearing given by the equation 10.1.

JER Uncertainties
The observable mostly affected by detector resolution effects is the p_T. In section 6.5, the resolution effects have been discussed. The way how to evaluate resolutions effects consists of smearing the p_T of each jet, at detector level by the following formula:

$$p_T^{det} = p_T^{det}(1 + SF_{res}\Delta R_{p_T}^{MC}) \;, \tag{10.3}$$

where $\Delta R_{p_T}^{MC} = (p_T^{reco} - p_T^{gen})/p_T^{gen}$, is the relative difference estimated in Monte Carlo, of the transverse momentum at both levels. The SF_{res} are factors provided by the CMS collaboration. Those factors are estimated for each data taking period (i.e year 2016), and they are in the range between 1.06 and 1.17. The uncertainties for these parameters are additionally provided.

Therefore, the uncertainties on the final measured distribution are estimated by applying the p_T smearing of the jets considering the variations up and *down* of the SF_{res} parameter. The maximum deviation of those variations with respect to the nominal distribution is considered as systematic uncertainties.

Uncertainties related to the b-tagging performance (`Btag Unc`)
The factors to be applied to the Monte Carlo simulations, in order to correct the performance of the CSVv2 tagger with respect to the data, are also affected by uncertainties. Following the recommendations given by a specific group in the CMS Collaboration dedicated to provide b-tagging scale factors, [160], the effect of the uncertainties on the measured cross section is determined. The *up* and *down* variations of the factors by 1σ are then considered in the event-by-event reweighting. Those variations are applied for the Monte Carlo simulations used through the background subtraction procedure, as well as for estimating the response matrices. The systematic uncertainties are estimated as the maximum deviation of those two variations with respect to the nominal distribution.

Uncertainties related to the measured luminosity (`Lumi Unc`)
The measured luminosity has a relative uncertainty of 2.5% [175] and it is directly considered in the measured cross sections at detector level. At particle level the uncertainty is propagated by unfolding the measurements with the respective up and down variations, using the nominal response matrices.

[1] It is difficult to predict the exact value, and larger values could be expected, since that 10% was for inclusive jets, while for top-jets, the tendency is not necessarily to the power of 5.

10.1.2 Estimated experimental uncertainties

Summarizing, the uncertainties related to the experimental sources are estimated at detector and particle level. At detector level they are estimated through repeating the background subtraction procedure, by taking into account the correspondent variations. After obtaining the variations of the measurement at detector level this new distribution is unfolded to particle level. In order to perform the unfolding procedure, in the case of the uncertainties related to the jet energy scale correction (`JES Unc`), to the jet energy resolution (`JER Unc.`) and to the b-tagging estimated efficiencies (`Btag. Unc`), the unfolding procedure considers a new set of response matrices provided with the correspondent variations. In the specific case of the propagation of the uncertainties from detector level to particle level related to the measured luminosity and to the background estimation procedure, the nominal response matrices are then considered.

Figures 10.3 and 10.4 show the individual sources of experimental uncertainties, estimated in the cross section measurements differential with respect to p_T of the two leading jets and with respect to $\Delta\phi$, respectively, at both levels: detector and particle level. In addition, the statistical uncertainties and the total uncertainties are illustrated. The total uncertainties are estimated as the quadratic sum of all the individual experimental uncertainties and the statistical uncertainties.

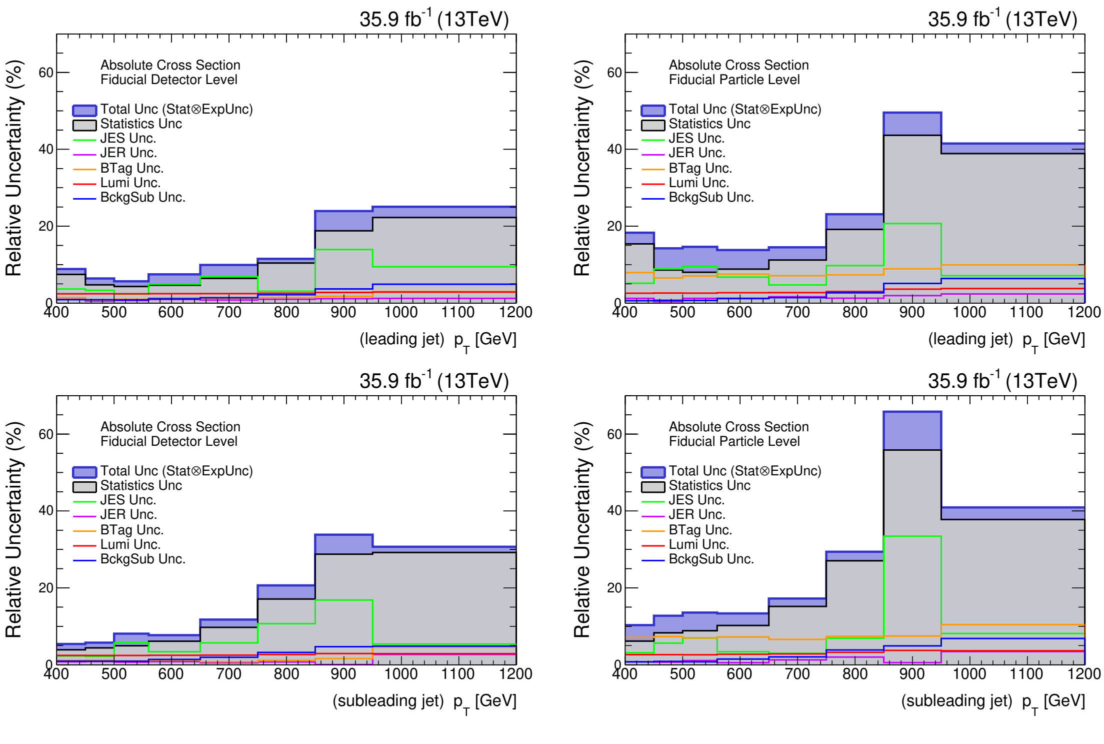


Figure 10.3: Breakdown of the experimental systematic uncertainties of the cross section measurements differential in p_T of the leading jet (upper plots) and subleading jet (lower plots) at: detector level (left plots) and particle level (right plots). The plots correspond to the measurements of the $t\bar{t}$ cross section.

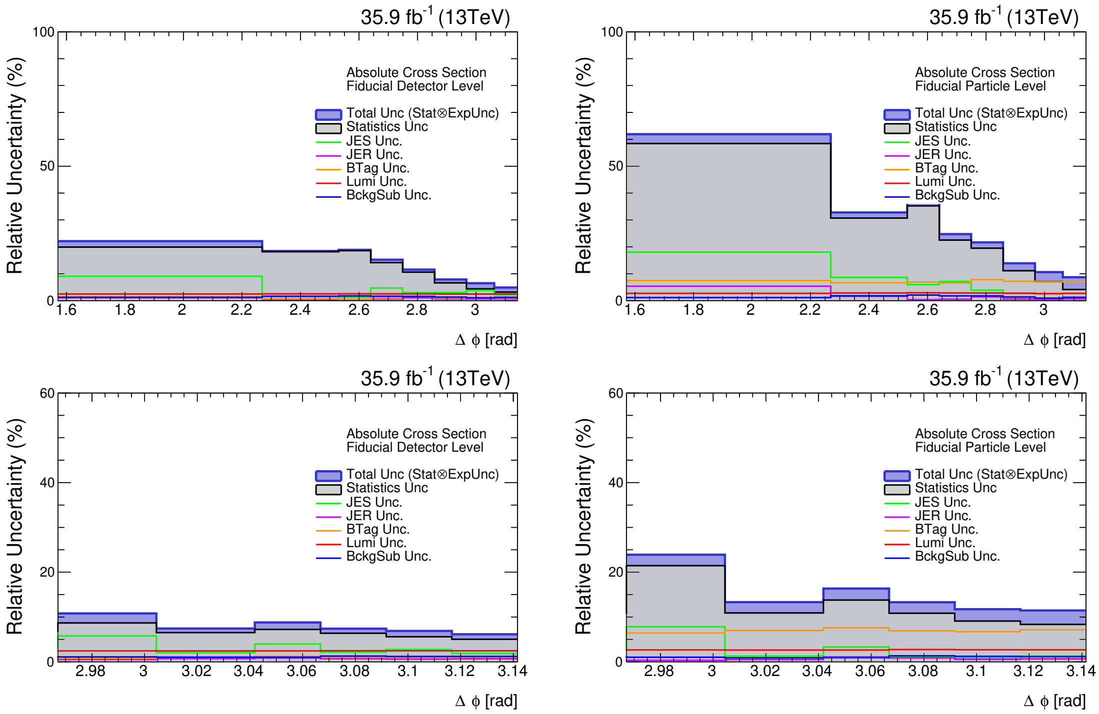


Figure 10.4: Breakdown of the experimental systematic uncertainties of the cross section measurements differential in $\Delta\phi$ in the whole phase space (upper plots) and in $\Delta\phi$ in the most back to back region (lower plots) at: detector level (left plots) and particle level (right plots). The plots correspond to the measurements of the $t\bar{t}$ cross section.

From the presented results, the following conclusions can be drawn:

- ✓ by propagating the uncertainties from detector level to particle level, the latter ones are larger than the former ones. The major change is observed in the uncertainties related to the b-tagging performance.
- ✓ in the case of the uncertainties predicted in the distributions differential in p_T (Figure 10.3), the two major contributions are given by the uncertainties related to the JES, and the b-tagging performance. The former ones change from 7% to 30% in the whole p_T phase space, while the second one behaves moreless flat with 7%. The uncertainties given by the background subtraction procedure and by JER are negligible in the lower p_T region, while on the higher p_T region they reach values of $\sim 7\%$
- ✓ in the case of the uncertainties predicted in the distributions differential in $\Delta\phi$ (Figure 10.4), in the most back-to-back region the major contributions are given by the b-tagging performance ($\sim 8\%$), while the total uncertainties is $\sim 10\%$.
- ✓ the b-tagging related uncertainties in the unfolded measurements have a rather flat behavior. This source of uncertainty, due to the constant behavior, cancels when normalized distributions are considered.

Tables H.1 and H.2 (in the appendix H) shows the obtained experimental uncertainties for all the measured distributions.

10.2 Monte Carlo signal modeling uncertainties

The input information needed for the unfolding procedure is given by simulated $t\bar{t}$ events with the POWHEG+PYTHIA8 event generator. The unfolded distributions could be biased through the specific event simulation. The variation of any parameters in the simulation of $t\bar{t}$ events, could change the final results, given the different possible migration effects together with fluctuations. In this section, those effects are studied in detail. The differences with respect to the nominal distributions are assigned as systematic uncertainties of the measurements associated to modeling dependence.

The uncertainties can be grouped in three main categories. Those categories, and each of their independent sources are listed in Table 10.1. The systematic uncertainties are assessed by considering specific variations of parameters in the simulation. The first group is related to the simulation of the hard scattering process, which is performed with POWHEG. The variations of the parameters are applied through event weights provided in the LHE files [176]. The second group of uncertainties is related to the modeling of all the underlying events (done within PYTHIA8). The systematic uncertainties are estimated by using alternative Monte Carlo samples, where variations of the specific PYTHIA8 parameters are applied. One exception is the uncertainties related to the matching of the Matrix Elements and the Parton Shower (ME-PS Unc.), which is estimated by changing the hdamp parameter in the POWHEG generator. The uncertainties related to the Monte Carlo t-quark mass used to simulate $t\bar{t}$ events are considered through variations on this parameter in the POWHEG+PYTHIA8 simulation. The way in which all the modeling uncertainties are handled, is explained in 10.2.1. The estimated quantities are provided in 10.2.2.

Table 10.1: Modeling uncertainties affecting the measurements through the efficiencies and response matrices (unfolding method).

Hard Scattering Unc.	Underlying Event Unc.	t-quark mass Unc.
✓PDF Unc.	✓ISR α_s Unc.	✓ $m_{top} = 172.5 \pm 1$GeV
✓Scale Unc.	✓FSR α_s Unc.	
	✓MPI Unc.	
	✓ME-PS Unc.	
	✓Color Reconnection Unc.	

The multivariate discriminant used in the selection procedure was obtained by training the multivariate technique with signal events ($t\bar{t}$ events) provided with of the POWHEG+PYTHIA8 event generator. In principle, small changes in the output discriminant d_{MVA}, when variations of parameters in the simulation are considered, could lead also to small changes in the acceptance, and therefore in the unfolded results. The systematic modeling uncertainty related to this selection criterion is considered within the systematic uncertainties by using different Monte Carlo samples to obtain the signal events.

Some observables used as input information of the multivariate technique (i.e τ_1, τ_2, τ_3) might be strongly sensitive to some of the considered variations, so that the acceptance (efficiency) considerably differs with respect to ones obtained with the nominal simulation. An example of the mentioned cases is when the variations of α_s in the Final State radiation (FSR) modeling (PYTHIA8) are considered. This can be understood by the influence of the FSR on the substructure of the jets, where a global scaling factor is applied at detector level, to consider the acceptance similar to the nominal case.

In the case of the inclusive top-jets measurements, the systematic uncertainties have been estimated only by considering variations in the $t\bar{t}$ simulated events. In those measurements, $\sim$ 80% of the events are originating from $t\bar{t}$ pair production, therefore the QCD multijet events has a minor relevance in the variations.

10.2.1 Handling modeling uncertainties.

In the following, the systematic uncertainties coming from unfolding are discussed.

PDF uncertainties (`PDF Unc.`)
The uncertainties related to the Parton Distribution Function (PDF) and to the strong coupling (α_s) used in the Monte Carlo simulation might play a role in the unfolding procedure. The PDF set used in this analysis is the `NNPDF3.0` [157] at NLO. The Monte Carlo predictions are provided with 100 replicas of the PDF considering $\alpha_s = 0.18$. Additional replicas are provided to account for the α_s variations in the PDF ($\alpha_s = 0.117$, $\alpha_s = 0.119$). For each of the PDF replicas the unfolding procedure is repeated. The uncertainty is estimated considering the standard deviation of all variations with respect to the nominal value.

Renormalization-factorization scales uncertainties (`Scale Unc.`)
In order to estimate the uncertainties related to the choice of the renormalization and factorization scales (μ_R, μ_F), the variations of those parameters by factor of 1/2 and 2 are considered. In total six independent variations are considered (up-down variations for μ_R, up-down variations for μ_F, up-down variations of both simultaneously). Similarly to the `PDF Unc`, the variations are assessed by events weights provided in the LHE files. For each of the variations, the unfolded procedure is performed, and the systematic uncertainties is estimated as the maximum deviation with respect to the nominal distribution.

Initial (Final) State radiation (α_s) uncertainties (`ISR (FSR)` α_s `Unc.`).
The uncertainties related to the α_s value used in initial and final state radiation modeling by the PYTHIA8 event generator are evaluated by variations of the α_s value to $\alpha_s = 0.117$ and $\alpha_s = 0.119$. Each of the uncertainties (ISR Unc, FSR Unc) are evaluated independently. For each variation, dedicated Monte Carlo samples with increased (decreased) α_s value are considered. The uncertainty is the envelope of the deviation after the unfolding procedure.

MPI Uncertainties (`UE Unc.`).
The MPI uncertainties are related to the specific tune used within the PYTHIA8 event generator. The uncertainties are estimated using different Monte Carlo samples where variations of the parameters used in the `CUETP8M2` tune are considered. Each tune is provided with variations of its parameters [177]. The response matrices considering the up-down variations of the parameters are provided as input information for the unfolding procedure. The estimated uncertainty is the largest difference between the nominal result and the up-down variations.

Matching the Matrix Element calculations and Parton Shower uncertainties (`ME-PS Unc.`).
In the matching scheme between the Matrix Elements (POWHEG) and the Parton Shower (PYTHIA8), the `hdamp` parameter plays a fundamental role. This parameter controls the high p_T radiation calculated in NLO processes. The uncertainties related to this parameter are estimated by considering the variation of this parameter. The nominal value used in $t\bar{t}$ simulated events is $h_{damp} = 1.581 \cdot m_{top}$, while the considered variations are $2.239 \cdot m_{top}$ and $0.9959 \cdot m_{top}$ [178]. The unfolding procedure is repeated with the new response matrices and the associated systematic uncertainties are taken as the largest variation with respect to the nominal distribution.

Color Reconnection modeling Uncertainties (`CR Unc.`)
The modeling of color reconnection between the initial and final state particles has an important impact in the simulation of the $t\bar{t}$ events [179]. Some alternative models for the colour reconnection scheme can be found in [180]. In order to access to the uncertainties related to this model dependence, an alternative Monte Carlo sample, also provided with POWHEG has been considered. This sample considers the QCD based Colour Reconnection model and the gluon-move approach [180]. As systematic uncertainty estimated from this source, the difference to the nominal distribution is taken.

Uncertainties related to $m_{top-quark}$ (`Mass Unc.`)
The uncertainties related to different t-quark mass values used in the Monte Carlo simulation of $t\bar{t}$ events are estimated by considering Monte Carlo samples with variation of this parameter: $m_{top} = 172.5$ GeV± 1 GeV. The uncertainty is then estimated as the largest deviation from the nominal unfolded distribution.

10.2.2 Estimated modeling uncertainties.

Figure 10.5 shows the systematic uncertainties for the $t\bar{t}$ cross section measurement differentially in p_T of the leading and subleading jets. The first row illustrates the uncertainties related to the modeling of the Underlying Events. In the second row, the uncertainties related to the simulation of the Hard Scattering process are shown. In the third row, the total uncertainties are given, combining the statistical with the modeling uncertainties quadratically.

As shown in Figure 10.5, the uncertainties related to the modeling of the Underlying Events vary in the range $\sim 1\% - 10\%$ ($\sim 1\% - 20\%$) for the leading (subleading) jet. The contribution related to the variations of the α_s value on the FSR modeling has in general a dominant effect. This can be understood since by those variations the jet substructure is affected. The modeling uncertainties of the hard scattering process predominantly is given by the variations of μ_F and μ_R scales, taking values of $1\% - 8\%$ in the whole p_T phase space region. The systematic uncertainty related to the value of m_{top} in the simulation of the $t\bar{t}$ events is taking values up to 2%. The statistical uncertainty changes in the range $\sim 8\% - 55\%$, being the dominant contribution compared to the theory modeling uncertainties.

Analogously to the previous plots, Figure 10.6 shows the modeling systematic and the statistical uncertainties, for the measurements differential in $\Delta\phi$. Similarly, the larger source of uncertainty is coming from the statistical, while the theory uncertainties, for instance, in the most back-to-back region are taking values of less than $\sim 2\%$. Figure 10.7 shows the corresponding plots for when the fine binning for the most back-to-back region is considered. The theory modeling uncertainties in this case is less than 5%, while the statistical uncertainties is $\sim 10\%$.

In the Appendix H in Tables H.3 and H.4 the estimated modeling systematic uncertainties for all measured distributions are given. The uncertainties are reported for the $t\bar{t}$ and inclusive top-jet measurements. For each measurement, the uncertainties are presented for the absolute distributions (cross sections), as well as, for the normalized distributions. The dependence on p_T or $\Delta\phi$ for all distributions is similar to plots shown in Figures 10.5, 10.6, 10.7.

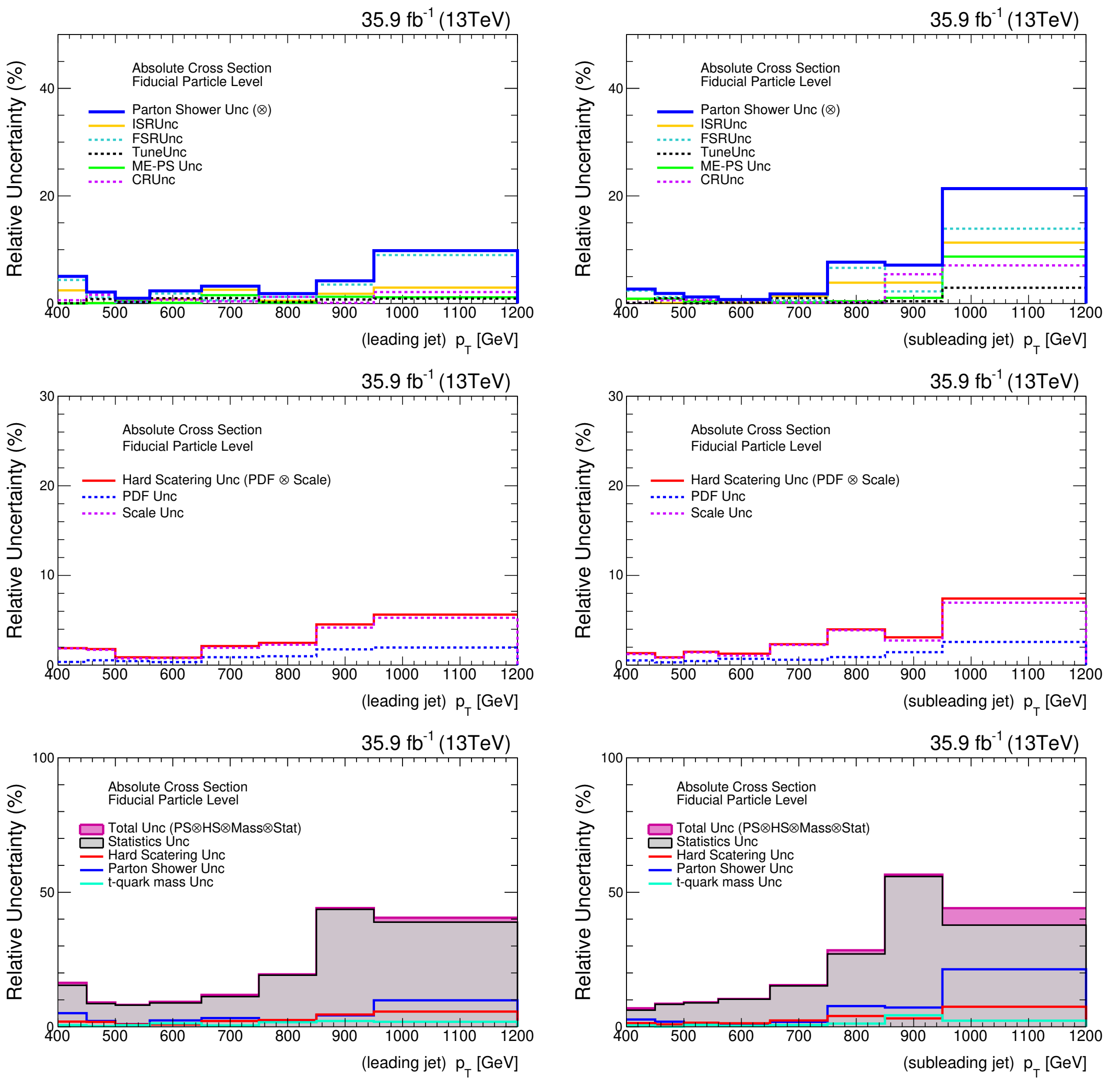


Figure 10.5: Breakdown of the relative systematic uncertainties of the cross section measurements differential with respect to p_T of the leading jet (left), subleading jet (right). (Upper plots): relative uncertainties related to the Parton Shower modeling. (Middle plots): relative uncertainties related to the Hard Scattering. (Lower plots): represent the total uncertainties considering all sources and the statistical uncertainties.

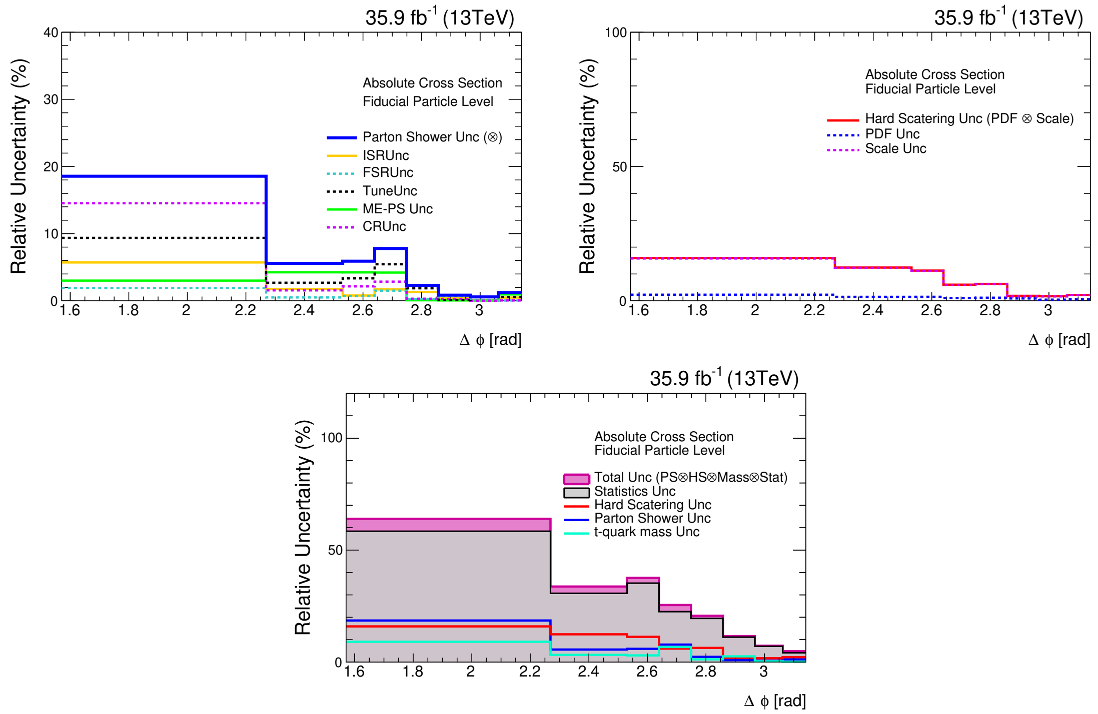


Figure 10.6: Breakdown of the relative systematic uncertainties of the cross section measurements differential in $\Delta\phi$: (upper-left plot) the Parton Shower modeling uncertainties, (upper-right) the Hard Scattering modeling uncertainties, (lower plot) the total uncertainties counting the modeling and the statistical uncertainties

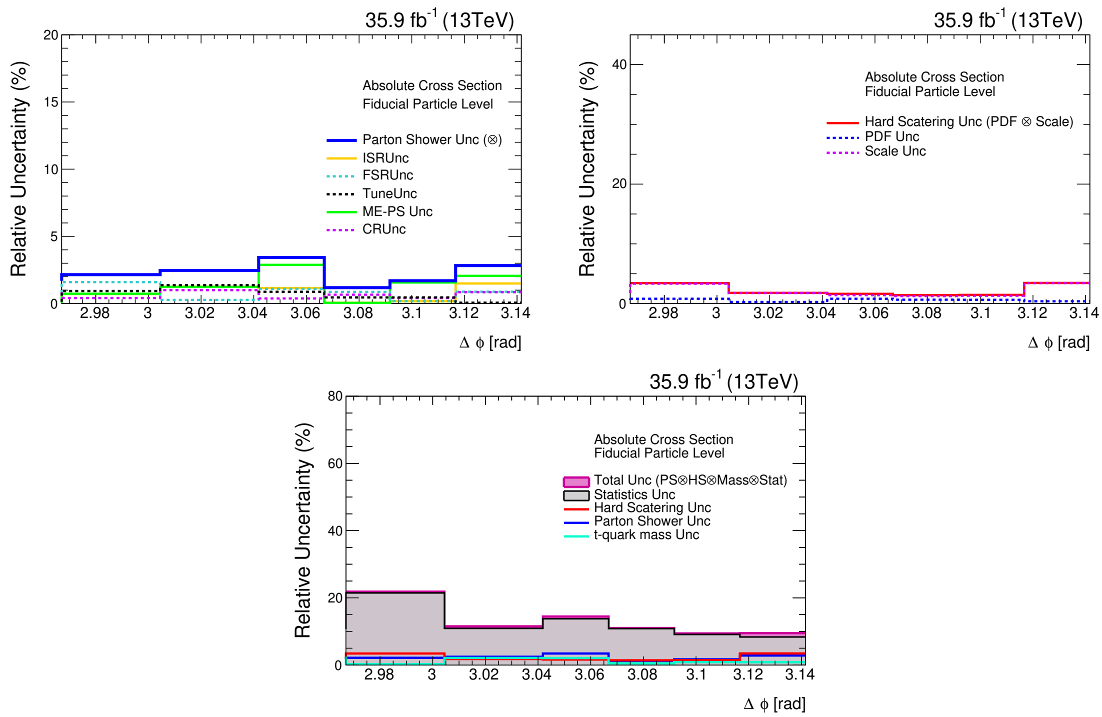


Figure 10.7: Breakdown of the relative systematic uncertainties of the cross section measurements differential in $\Delta\phi$: (upper-left plot) the Parton Shower modeling uncertainties, (upper-right) the Hard Scattering modeling uncertainties, (lower plot) the total uncertainties counting the modeling and the statistical uncertainties

10.3 Summary of uncertainties.

In this section, a summary of the different sources of uncertainties is given. The systematic uncertainties are later combined with the statistical uncertainties forming the total uncertainties of the measurements.

Figure 10.8 shows the systematic and statistical uncertainties for typical examples. The uncertainties for both, the absolute and normalized distributions are shown, with the individual contributions grouped in the following main sources:

- ✓ experimental: JES⊗JER⊗Lumi⊗t-quark mass,
- ✓ Hard Scattering,
- ✓ Underlying event,
- ✓ b-tagging.

The plots correspond to measurements differentially in p_T of the leading and subleading jets and $\Delta\phi$. The last row corresponds to the measurements in the most back-to-back region. In the Appendix H, analogous plots are provided (Figures H.2, H.2, H.1), for measurements in the exclusive regions of the fiducial particle level. The total uncertainties are then estimated as the quadratic sum of all the uncertainties.

Tables 10.2, 10.3, 10.4, and 10.5, give a summary of the uncertainties intervals (minimum and maximum values) for p_T of the leading jet, p_T of the subleading jet, $\Delta\phi$ in the extended phase space, and $\Delta\phi$ in the back-to-back region, respectively. The values correspond to the cross section measurements provided in Figure 10.8, and the reported intervals cover the whole phase space, although, for instance, for the p_T observables, the minimum value is the typical values, while the maximum is reached in the higher p_T region, where statistical fluctuations become important.

In the measurements differential in p_T, the dominant sources of uncertainties are related to the jet energy scales, b-tagging efficiency, and luminosity. All the other sources are small (less than 1%) in the low p_T region, and increase with the p_T. For the normalized distributions, the b-tagging and underlying event uncertainties are very small, at least in the region $p_T < 800$ GeV, and the systematic uncertainties are generally reduced. The total systematic uncertainty, in the lower p_T region are comparable with the statistical uncertainties, while in the higher p_T region, the statistical uncertainties dominate.

The statistical uncertainties in the distributions differentially in $\Delta\phi$ are larger in the lower $\Delta\phi$ region. In the region nearer to $\Delta\phi \sim \pi$, the dominant systematic uncertainties are given by the b-tagging efficiency ($\sim$ 6.5%) and the luminosity $\sim$ 2.6%. Therefore, in this region, the systematic uncertainties dominate over the statistical uncertainties (the systematic are of the order $\sim$ 8%, while the statistical uncertainties of the order of $\sim$ 4%).

Tables H.5 and H.6, in the Appendix H provide the summary of the systematic uncertainties for all the measured distributions.

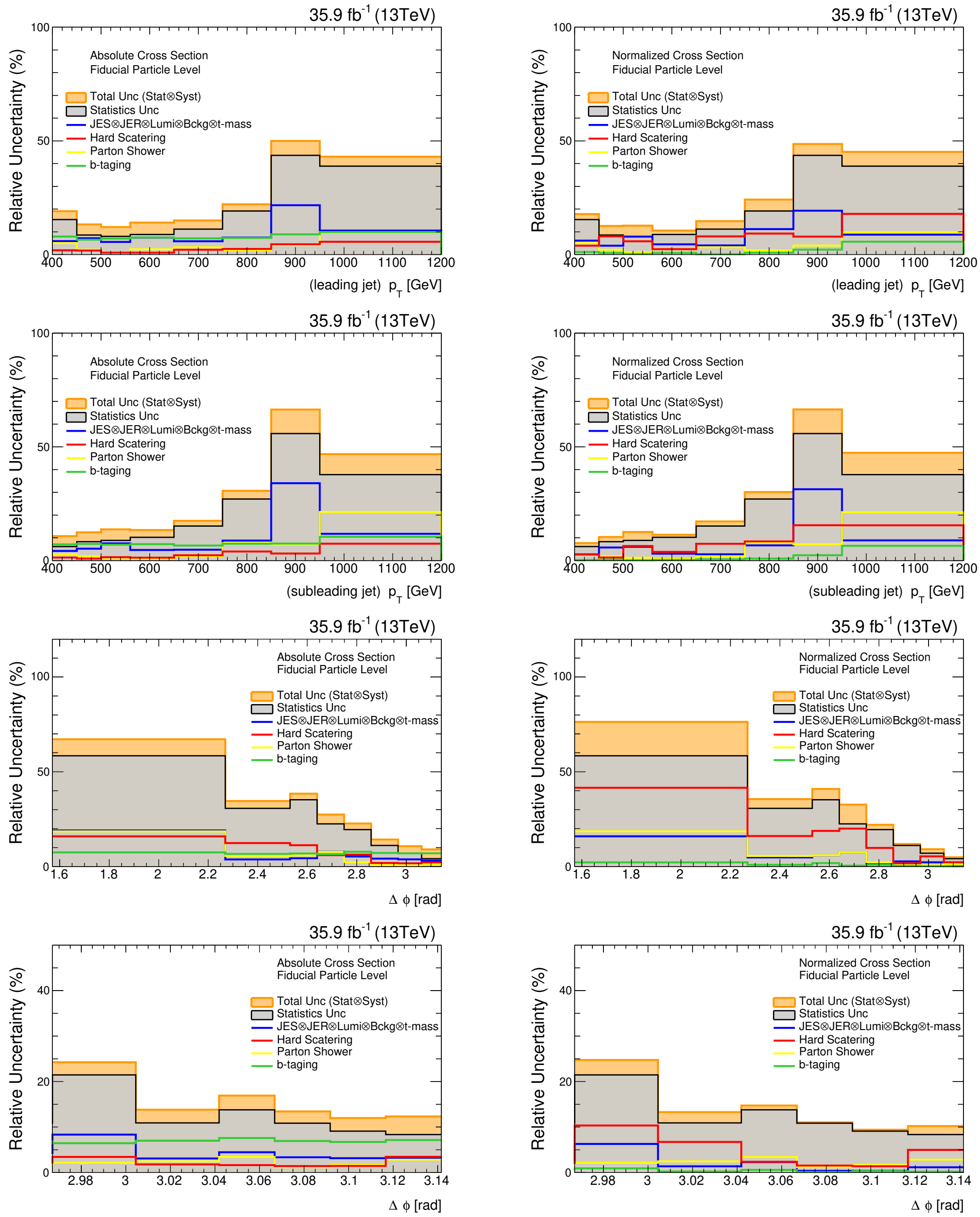


Figure 10.8: Systematic and statistical uncertainties corresponding to the distributions differential in p_T of the leading jet (two first plots), p_T of the subleading jets (second row), $\Delta\phi$ (third row), and $\Delta\phi$ in the back to back region (lower plots), in the fiducial particle level. The systematic uncertainties corresponding to the absolute and normalized distributions are illustrated.

Table 10.2: Summary of the minimum-maximum interval of the systematic and statistical relative uncertainties for the fiducial particle level cross section measurement differential in p_T of the leading jet. The values correspond to the $t\bar{t}$ measurements. The minimum reported value corresponds to the typical uncertainties, while the maximum is reach in the higher p_T region.

Source of Uncertainty	Percentage (min,max)
Hard Scattering	$0.84\% - 5.63\%$
Parton Shower	$0.97\% - 9.82\%$
t-quark mass	$0.23\% - 2.12\%$
JES	$4.71\% - 20.69\%$
JER	$0.39\% - 2.38\%$
Luminosity	$2.57\% - 3.79\%$
Background estimation	$0.63\% - 6.40\%$
b-tagging efficiency	$6.54\% - 9.92\%$
Total Systematic	$9.10\% - 24.36\%$
statistical Unc	$8.01\% - 43.64\%$
Total	$12.12\% - 49.98\%$

Table 10.3: Summary of the minimum-maximum interval of the systematic and statistical relative uncertainties for the fiducial particle level cross section measurement differential in p_T of the subleading jet. The values correspond to the $t\bar{t}$ measurements. The minimum reported value corresponds to the typical uncertainties, while the maximum is reach in the higher p_T region.

Source of Uncertainty	Percentage (min,max)
Hard Scattering	$0.87\% - 7.43\%$
Parton Shower	$0.74\% - 21.36\%$
t-quark mass	$0.03\% - 4.26\%$
JES	$2.98\% - 33.45\%$
JER	$0.55\% - 3.43\%$
Luminosity	$2.60\% - 3.74\%$
Background estimation	$0.77\% - 6.85\%$
b-tagging efficiency	$6.62\% - 10.45\%$
Total Systematic	$8.69\% - 35.94\%$
Statistical	$6.16\% - 55.85\%$
Total	$10.76\% - 66.42\%$

Table 10.4: Summary of the minimum-maximum interval of the systematic and statistical relative uncertainties for the fiducial particle level cross section measurement differential in $\Delta\phi$ (in the whole $\Delta\phi$ studied phase space).The values correspond to the $t\bar{t}$ measurements. The minimum reported value corresponds to the typical uncertainties in the most back-to-back region, while the maximum is reach in the higher lower $\Delta\phi$ region.

Source of Uncertainty	Percentage (min,max)
Hard Scattering	1.71% − 15.93%
Parton Shower	0.59% − 18.56%
t-quark mass	0.15% − 9.05%
JES	0.37% − 18.03%
JER	0.34% − 5.36%
Luminosity	2.62% − 2.91%
Background estimation	0.92% − 2.07%
b-tagging efficiency	6.53% − 7.81%
Total Systematic	8.06% − 33.14%
Statistical	4.14% − 58.44%
Total	9.06% − 67.18%

Table 10.5: Summary of the minimum-maximum interval of the systematic and Statistical relative uncertainties for the fiducial particle level cross section measurement differential in $\Delta\phi$ when the fine binning in the most back-to-back region is considered. The values correspond to the $t\bar{t}$ measurements

Source of Uncertainty	Percentage (min,max)
Hard Scattering	1.41% − 3.45%
Parton Shower	1.18% − 3.43%
t-quark mass	0.27% − 2.10%
JES	0.94% − 7.83%
JER	0.32% − 1.09%
Luminosity	2.62% − 2.73%
Background estimation	0.89% − 1.33%
b-tagging efficiency	6.44% − 7.58%
Total Systematic	7.80% − 11.29%
Statistical	8.35% − 21.46%
Total	11.98% − 24.25%

Chapter 11

Results

The fiducial measurements at detector level are unfolded to stable particle level. The measurements were already presented in Chapter 8, and the unfolding procedure was covered in Chapter 9. Additionally, in Chapter 10, the systematic uncertainties on each of the measured distributions were estimated. This chapter focuses on the results at particle level and their respective comparison to theory predictions. Two main measurements are distinguished:

- ✓ $t\bar{t}$ measurements: considering only $t\bar{t}$ events, where the QCD background has been subtracted
- ✓ inclusive top-jet measurements: the particle level is defined with two top-jet candidates, considering the combination of QCD multijet and $t\bar{t}$ events as the signal (the QCD contribution is not subtracted).

The particle level, to which the measured distributions are unfolded, was defined in Chapter 9 by by requiring at least two jets with:

- ✓ $p_T > 400$ GeV, $|\eta| < 2.4$, $m_j^{sd} > 50$ GeV, [1],
- ✓ $m_{sub0} > 40$ GeV [2],
- ✓ a B-hadron inside the jet : $\Delta R_{B-jet} < 0.4$ [3].

The cross section measurements differential in an observable x [4] (considering the finite interval Δx_i defined by the bin selection), at particle level, can be defined as follows:

$$\frac{d\sigma_i}{dx} = \frac{1}{\mathcal{L}\Delta x_i}\frac{1}{e_i}\sum_j M_{ij}^{-1} a_j (N_j^{data} - N_j^{bg}), \tag{11.1}$$

[1] m_j^{sd} is the soft drop jet mass.
[2] m_{sub0} is the mass of the first soft drop subjet.
[3] ΔR is the spacial distance in the $\phi - \eta$ plane.
[4] p_T of the leading and subleading jets, and $\Delta\phi$ between the two top jets.

where $\mathcal{L}$ is the measured integrated luminosity, e_i is the efficiency in a given bin i defined as the fraction of events at particle level which has an equivalent event at detector level, in a specific bin i; a_j is the acceptance [5]; M_{ij}^{-1} counts for the inverse of the response matrix considered by the unfolding; N^{data} and N^{bg} counts for the total events in the data and the background events that were falsely selected at reconstructed level, respectively.

The measurements are also presented as normalised distributions which can be written as follows:

$$\frac{1}{\sigma}\frac{d\sigma_i}{dx} = \frac{1}{\sum_i^{bins}(N_i^{data} - N_i^{bg})}\frac{1}{e_i}\sum_j M_{ij}^{-1} a_j (N_j^{data} - N_j^{bg}) \tag{11.2}$$

For the comparison to theory predictions, the following Monte Carlo simulations are considered:

- ✓ **POWHEG+PYTHIA8**: $t\bar{t}$ events are simulated at NLO in the perturbative QCD matrix element calculations using the POWHEG V2 box [149, 150] with m_t = 172.5 GeV. The Parton Shower is simulated with PYTHIA8 [34] using the CUETP8M2T4 Tune [181] [37].
- ✓ **POWHEG+HERWIGPP**: $t\bar{t}$ events are simulated at NLO in the perturbative QCD expansion with the POWHEG V2 box, as before, but considering the parton shower, hadronization and MPI from HERWIG++ [43] with the EE5C [37] Tune.
- ✓ **aMC@NLO+PYTHIA8**: $t\bar{t}$ events are simulated at NLO in the perturbative QCD expansion with aMC@NLO [153] matched to PYTHIA8 using the CUETP8M2T4 Tune.
- ✓ QCD multijet events are simulated at LO with MADGRAPH5 [153] considering the MLM [156] matching algorithm between Matrix Element and Parton Shower. The latter are simulated with PYTHIA8 using the CUETP8M1 Tune [37].

In the case of $t\bar{t}$ events, the total cross section is normalized to the cross section predicted at NNLO accuracy [154, 155], which is the most precise available calculation for the signal predictions. The QCD multijet predictions are combined with the $t\bar{t}$ predictions for comparing to the top-jet inclusive measurements. A K-factor ($K = 0.65$) is considered for the latter, since the LO accuracy does not properly predict this contribution.

11.1 Cross section measurements of the $t\bar{t}$ pair production

In this section, the results of the cross section measurements for $t\bar{t}$ pair production are presented and compared to theory predictions. The measurements differential in p_T and $\Delta\phi$ are presented in 11.1.1 and 11.1.2 respectively. In the subsection 11.1.3, the integrated cross section is estimated.

11.1.1 Measurements differential in p_T

Figures 11.1 and 11.2 show the measured distributions (absolute and normalized) differentially in p_T of the leading and subleading jets respectively. The estimated total uncertainties in the measurements are shown while for the theory predictions only the statistical uncertainty is considered. The measured p_T phase space is in the interval from 400 GeV to 1.2 TeV.

[5]fraction of the events at detector level which has an equivalent event at particle level.

Interesting features are observed from the comparison to theory predictions:

- ✓ the cross section decreases approximately two orders of magnitude in the p_T range from 400 GeV to 1.2 TeV.
- ✓ in the comparison, disagreements of about $\sim$ 40% can be observed, being the measured cross section smaller than the predicted ones.
- ✓ the distributions predicted from POWHEG+PYTHIA8 and POWHEG+HERWIGpp are similar, meaning that the p_T spectra is not sensitive to the modeling of the parton shower.
- ✓ the predictions for the leading jet corresponding to aMC@NLO+PYTHIA8 show a slightly harder spectrum compared to the other predictions. The differences are observed in the region $p_T > 700$ GeV, meaning that the p_T spectra in the most boosted region are sensitive to the Matrix Element calculations.
- ✓ the shape of the distributions predicted from all Monte Carlo simulations is in good agreement with the measured ones. The prediction better describing the shape of the measured distribution is aMC@NLO+PYTHIA8 [1].

Figures 11.3 and 11.4 show the measured distributions as a function of the p_T of the leading and subleading jets, when exclusive regions of the fiducial particle level are considered: $|\eta| < 0.5$ and $0.5 < |\eta| < 2.4$, respectively. The comparison to theory predictions show a similar behavior as for the inclusive fiducial phase space.

In Table 11.1 the estimated χ^2 values, between the data distributions and the theory predictions are provided. They are estimated with the normalized distributions as:

$$\chi^2 = \sum_{i}^{N_{bins}} \frac{(\Sigma_i^{data} - \Sigma_i^{mc})^2}{(\epsilon_i^{data})^2 + (\epsilon_i^{mc})^2} \tag{11.3}$$

where N_{bins} is the number of considered bins, Σ^{data} and Σ^{mc} are the values of the normalized distributions in the considered bin for data and Monte Carlo, while ϵ_i, refers to the uncertainties in both distributions [2]. The values of the number of degree of freedom (ndf) is $N_{bins} - 1$.

Table 11.1: Comparison of the measurements and theory predictions in the normalized distributions with the χ^2 estimator and the number of degree of freedom (ndf).

Observable	POWHEG+PYTHIA8 χ^2/ndf	POWHEG+HERWIGpp χ^2/ndf	AMC@NLO+PYTHIA8 χ^2/ndf
p_T leading jet (fiducial particle level)	9.39/7	16.09/7	6.64/7
p_T subleading jet (fiducial particle level)	4.01/7	8.90/7	2.52/7
p_T leading jet ($\|\eta\| < 0.5$)	11.03/7	17.94/7	9.69/7
p_T subleading jet ($\|\eta\| < 0.5$)	6.90/7	11.6/7	4.65/7
p_T leading jet ($0.5 < \|\eta\| < 2.4$)	4.07/7	6.20/7	4.51/7
p_T subleading jet ($0.5 < \|\eta\| < 2.4$)	2.14/7	3.32/7	3.38/7

[1] this can be better noticed from results shown in Table 11.1 .

[2] No correlation between the uncertainties have been considered.

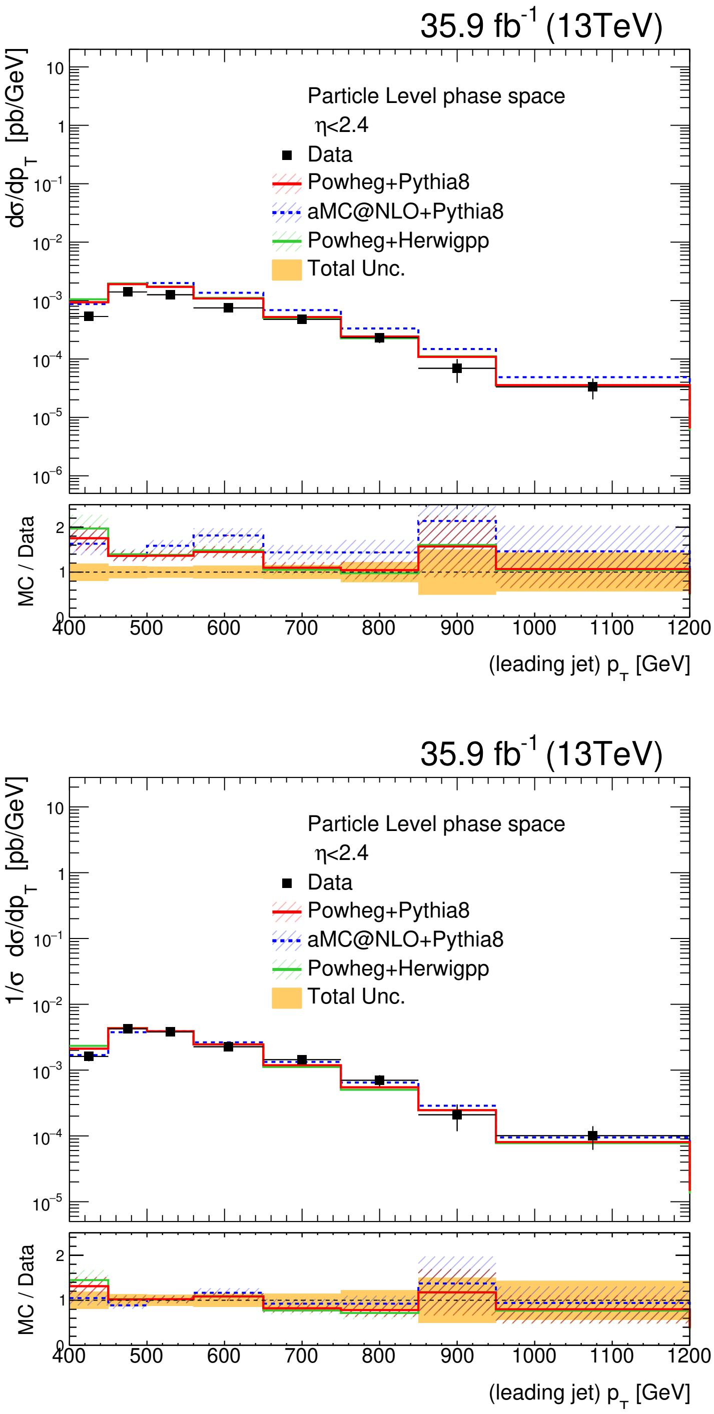
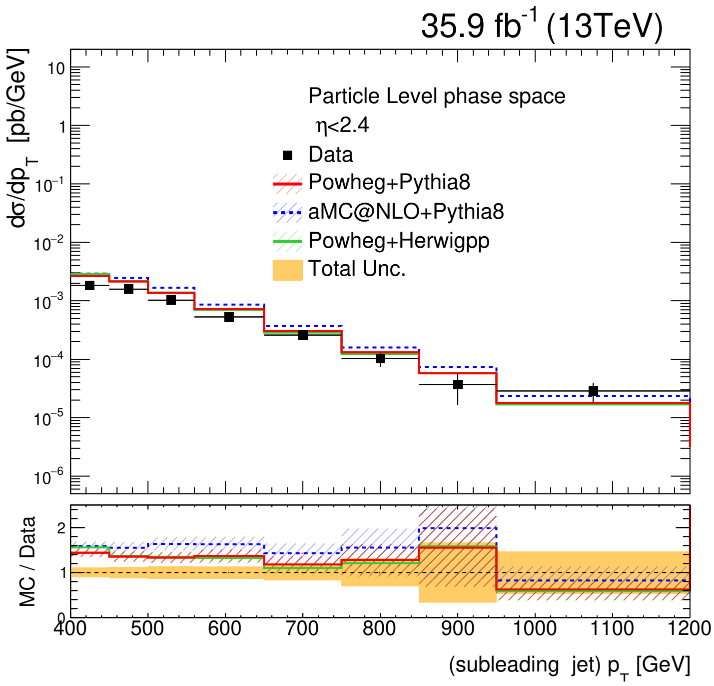


Figure 11.1: Cross section (upper plot) and normalized distributions (lower plot) differential in p_T of the leading top-jet, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The $t\bar{t}$ measurements are performed in the inclusive fiducial particle level ($|\eta| < 2.4$).



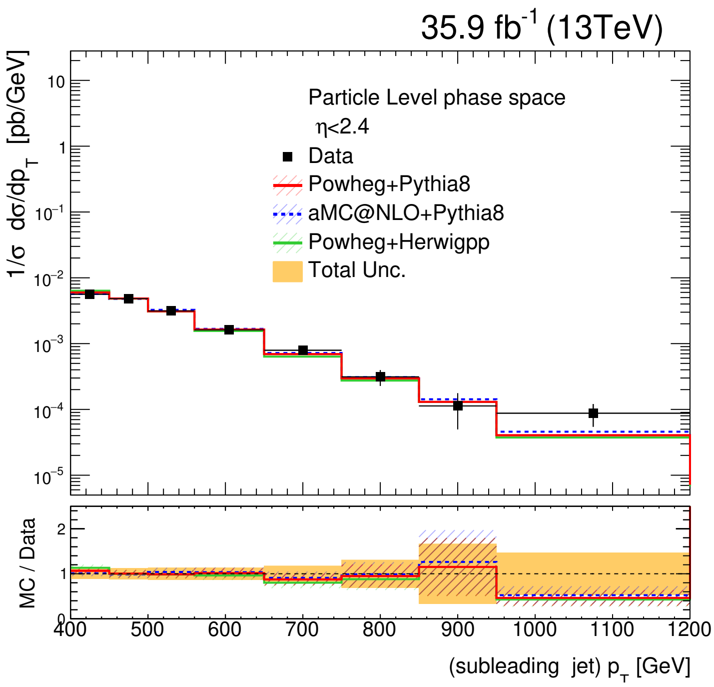


Figure 11.2: Cross section (upper plot) and normalized distributions (lower plot) differential as function of the p_T of the subleading top-jet, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The $t\bar{t}$ measurements are performed in the inclusive fiducial particle level ($|\eta| < 2.4$).

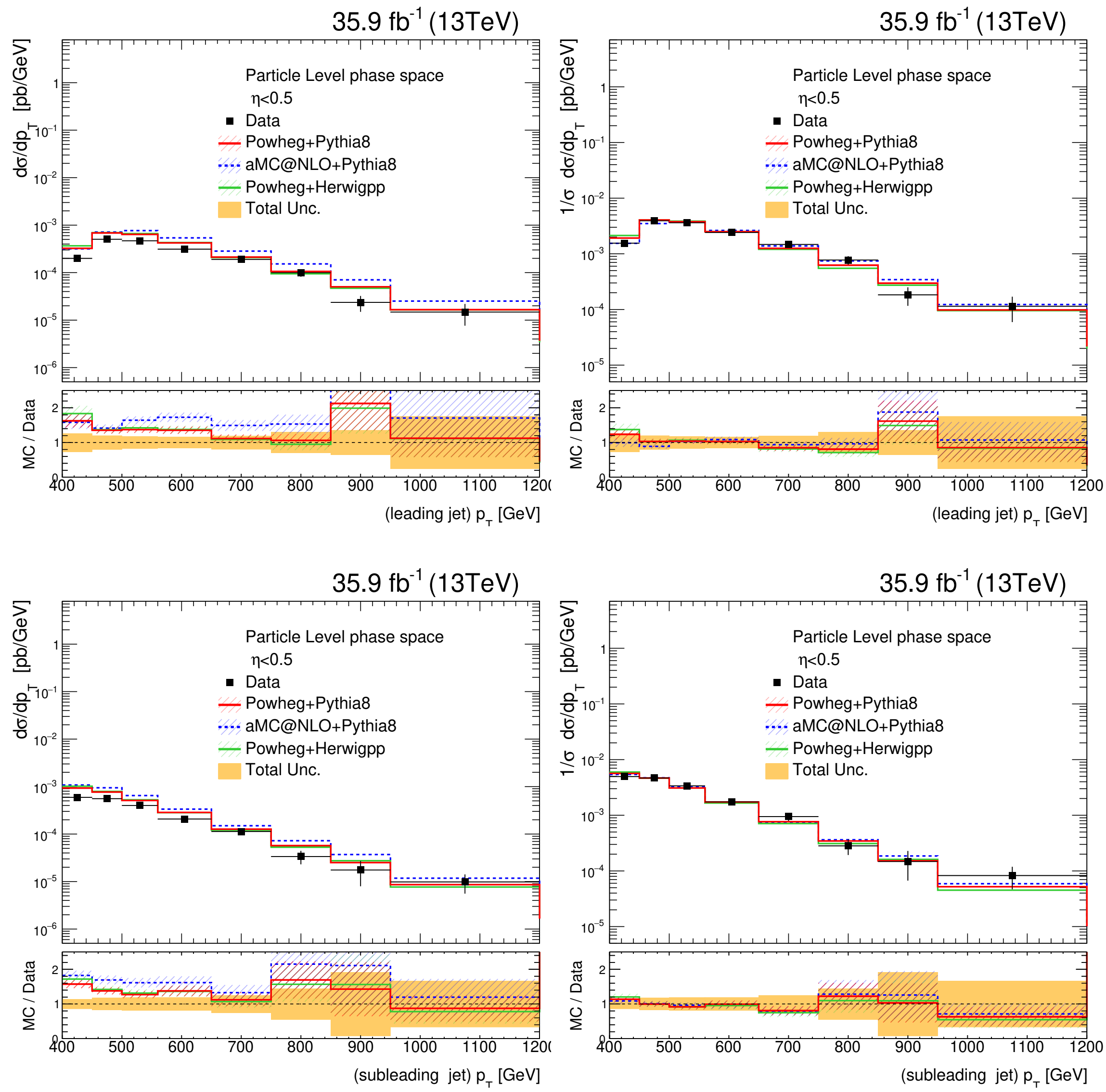


Figure 11.3: Cross section (left plots) and normalized distributions (right plots) differential as function of the p_T of the leading jet (upper plots), subleading jet (lower plots), compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The $t\bar{t}$ measurements are performed in $|\eta| < 0.5$.

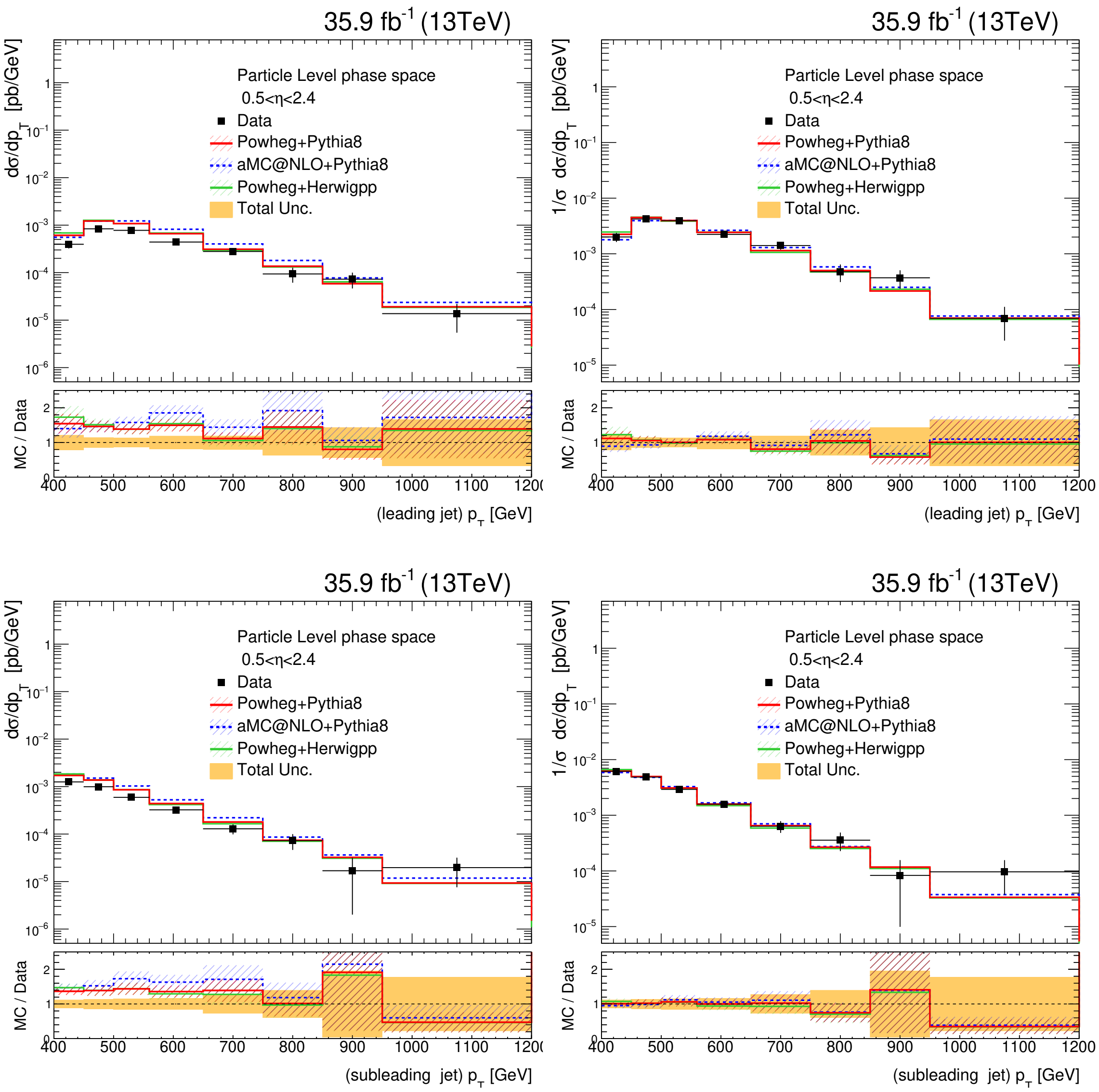


Figure 11.4: Cross section (left plots) and normalized distributions (right plots) differential as function of the p_T of the leading jet (upper plots), subleading jet (lower plots)compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The $t\bar{t}$ measurements are performed in $0.5 < |\eta| < 2.4$.

11.1.2 Measurements differentially in $\Delta\phi$

Figure 11.5 and 11.6 show the measured cross section and normalized distributions differential in $\Delta\phi$ between the two top-jets compared to theory predictions. In Figure 11.6, the fine binning in the most back-to-back region is considered. Measurements in exclusive regions defined by the p_T of the leading jet (400 GeV$< p_T^{lead} <$ 600 GeV, and $p_T^{lead} >$ 600 GeV) are shown in Figure 11.7.

Table 11.2 provides the χ^2 values computed for comparing the normalized distributions.

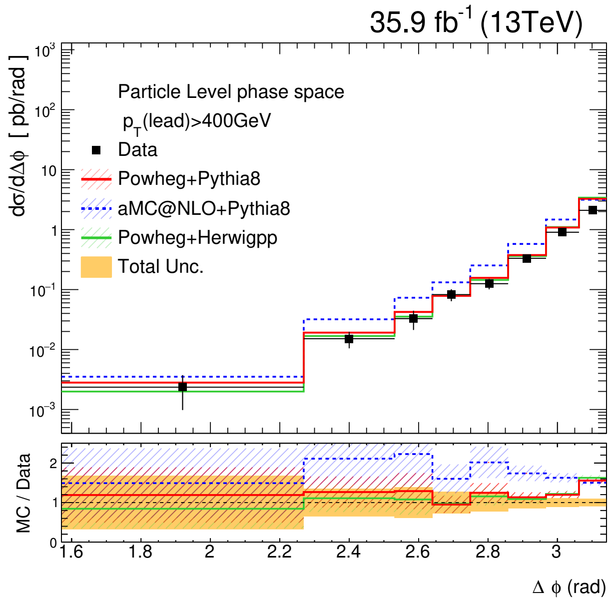
Table 11.2: Comparison by using the χ^2 estimator of the measured normalized distributions with Standard Model theory predictions.

Observable	POWHEG+PYTHIA8 χ^2/ndf	POWHEG+HERWIGpp χ^2/ndf	AMC@NLO+PYTHIA8 χ^2/ndf
$\Delta\phi$ inclusive particle level	18.39/7	25.61/7	6.77524/7
$\Delta\phi$,back-to-back fine binning [2]	22.36/5	28.51/5	20.13/5
$\Delta\phi$, 400 GeV$< p_T^{lead} <$ 600GeV	16.2/7	19.06/7	10.60/7
$\Delta\phi$, $p_T^{lead} >$ 600 GeV	9.36/7	13.82/7	5.51/7

The following features are observed comparing the measured distributions to the theory predictions:

- ✓ for the cross section measurements:
  - ✓ the POWHEG+PYTHIA8 and POWHEG+HERWIGpp predictions differ up to $\sim$ 60% from the measured cross section. The differences are enhanced towards increasing $\Delta\phi$. This is better noticed in the measurements in the most back-to-back region, where the difference increases in the region $3.05 < \Delta\phi < \pi$,
  - ✓ the aMC@NLO+PYTHIA8 predictions shows a disagreement of $\sim$ 60%, but as an overall factor in the whole $\Delta\phi$ phase space.
- ✓ for the normalized distributions:
  - ✓ the shape of the distributions predicted from POWHEG+PYTHIA8 and POWHEG+HERWIGpp show similar behavior, meaning that there is not noticeable influence of the parton shower modeling on the predicted distributions.
  - ✓ different behavior is observed in the aMC@NLO+PYTHIA8 predictions with respect to the POWHEG+PYTHIA8 and POWHEG+HERWIGpp, having the former one a broader opening angle [1]
  - ✓ The measured distribution in the most back-to-back region (Figure 11.6) is broader than predicted from POWHEG+PYTHIA8, but narrower than the one predicted from aMC@NLO+PYTHIA8. The best prediction describing the shape is given by aMC@NLO+PYTHIA8.
- ✓ Similar behavior is observed when considering exclusive regions in the particle level phase space, i.e 400 GeV $< p_T^{lead} <$ 600 GeV and $p_T^{lead} >$ 600 GeV.

[1] the decreasing of the differential distribution when decreasing $\Delta\phi$ is softer



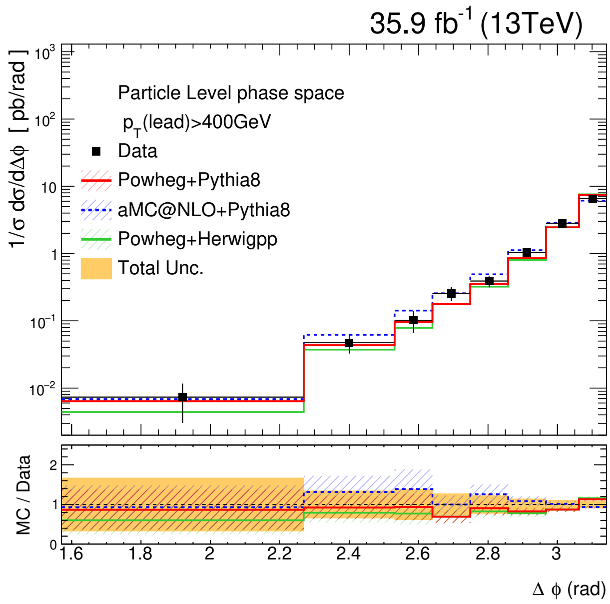


Figure 11.5: Cross section (upper plot) and normalized distributions (lower plot) differential as function of the $\Delta\phi$ between the two top-jets, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The $t\bar{t}$ measurements are performed in the inclusive fiducial phase space region ($|\eta| < 2.4$, $p_T > 400$ GeV).

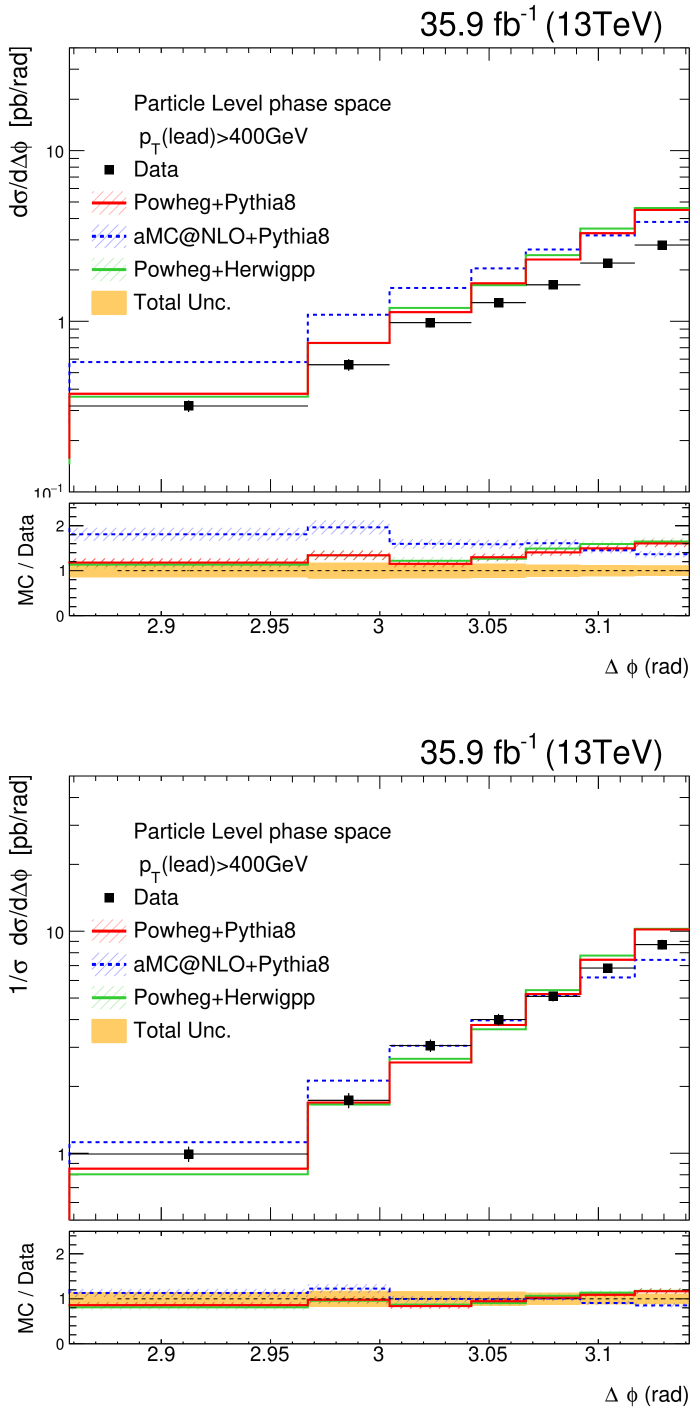


Figure 11.6: Cross section (upper plot) and normalized distributions (lower plot) differential as function of the $\Delta\phi$ between the two top-jets compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The $t\bar{t}$ measurements are performed in the inclusive fiducial phase space region ($|\eta| < 2.4$, $p_T > 400$ GeV). The fine binning in the most back-to-back region is considered.

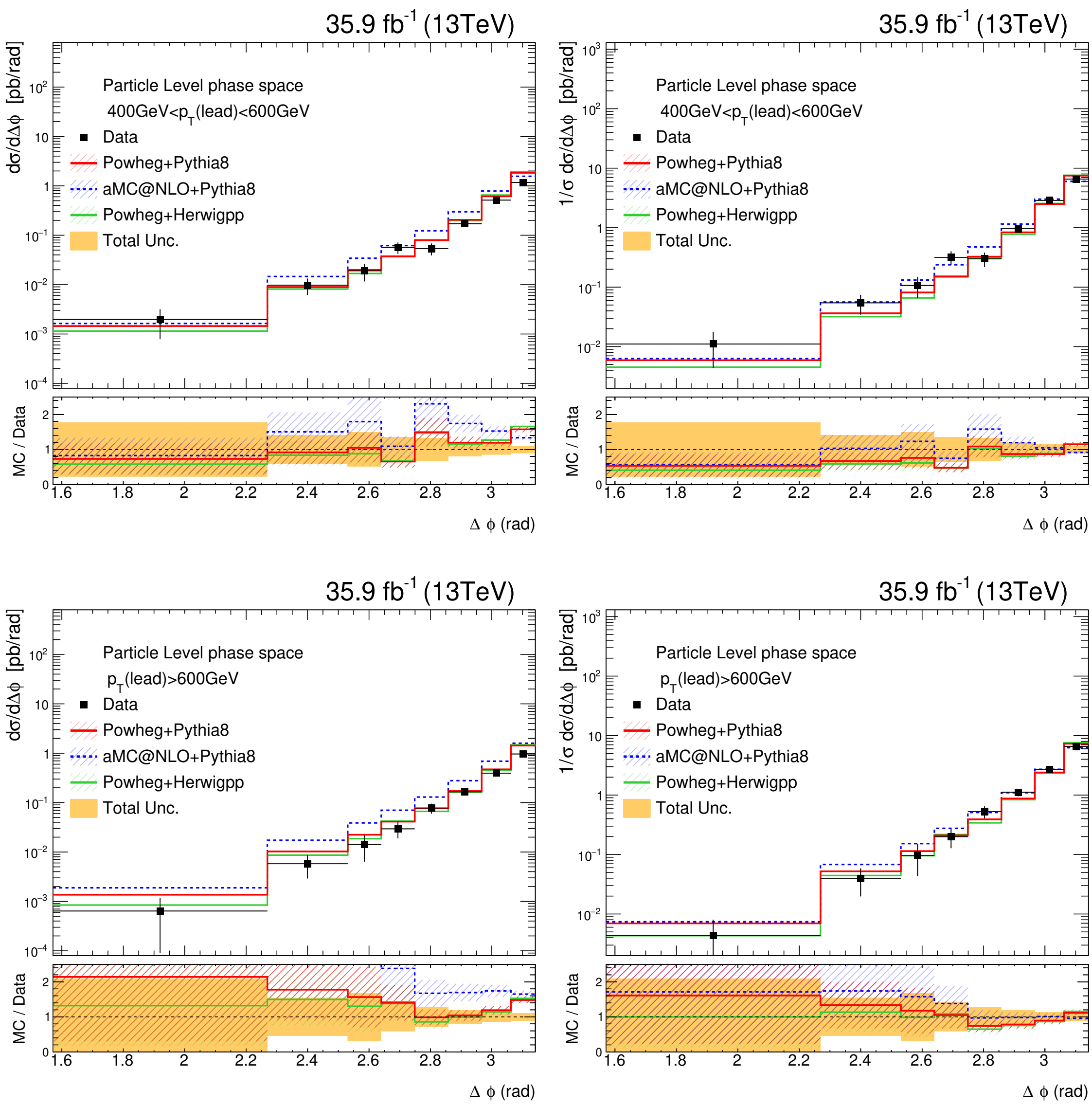


Figure 11.7: Cross section (left plots) and normalized distributions (right plots) differential as function of the $\Delta\phi$ between the two top-jets, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The $t\bar{t}$ measurements are performed in: (upper plots400 GeV$< p_T^{lead} <$ 600 GeV), (lower plots) $p_T^{lead} >$ 600 GeV .

11.1.3 Inclusive Cross section and comparison with theory predictions.

Figure 11.8 shows the integrated cross section, which is estimated by integrating the differential distribution in $\Delta\phi$ in the whole phase space.

The measured value and uncertainties (represented by the pink band) are:

$$\sigma = 321.1\text{fb} \pm 26.7\text{fb (stat. unc.)} \pm 28.4\text{fb (syst. unc.)} \tag{11.4}$$

The predicted inclusive cross section from different theory predictions are also illustrated.

The predictions corresponding to POWHEG+HERWIGpp and aMC@NLO+PYTHIA8 are reported with their statistical uncertainties, while in the case of the prediction with POWHEG+PYTHIA8, the systematic uncertainties are added in quadrature with the statistical uncertainties. The predicted value for the latter one is:

$$\sigma = 429.11\text{fb} \pm 4.15\text{fb (stat. unc.)} \pm 47.02\text{fb (syst. unc.)} \tag{11.5}$$

Two sources of theoretical uncertainties have been considered for estimating the systematic uncertainties on the POWHEG+PYTHIA8: the PDF uncertainty and the renormalization and factorization scales (μ_R, μ_F) uncertainties [1].

The estimated integrated cross section shows a $\sim$ 34% difference from the POWHEG+PYTHIA8 prediction.

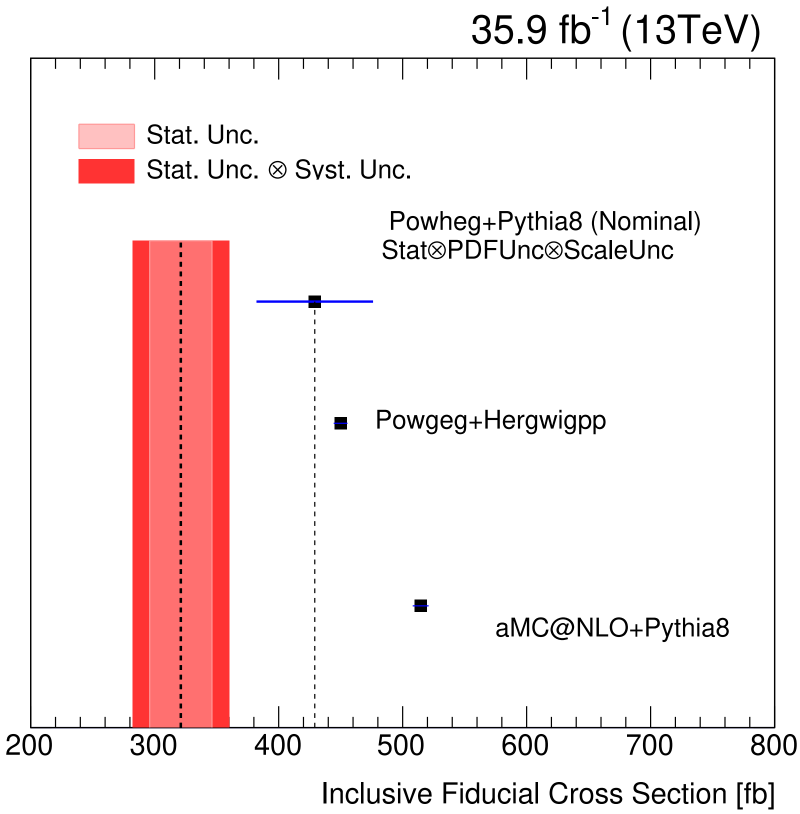


Figure 11.8: Inclusive $t\bar{t}$ pair cross section. The measured value is represented by the dashed black line with by the pink colored areas showing the statistical and total uncertainties. Additionally, the predicted values from theory predictions are shown: POWHEG+PYTHIA8, POWHEG+HERWIGpp, aMC@NLO+PYTHIA8. The former one is shown with the total uncertainties (systematic and statistic), while for the others only the statistical uncertainties are included.

[1]The PDF uncertainties are estimated by using the events weights for each $NNPDF$3.0 MC replicas, while the scale uncertainties are estimated by the variations of μ_R, μ_F parameters

11.2 Inclusive top-jet measurements

In this section, the inclusive top-jet measurements are presented. These measurements benefit from two points: first, the events are selected and the QCD contribution is not subtracted, and secondly, the statistical uncertainties are lower due to more events selected. Since the dominant contribution is given from the $t\bar{t}$ events, similar behavior observed for the $t\bar{t}$ measurements (presented in section 11.1) is expected.

Figures 11.9 and 11.10 show the measured distributions as a function of the p_T of the leading and subleading jets, and as function of $\Delta\phi$, respectively. The measurements correspond to the inclusive phase space ($|\eta| < 2.4$ and $p_T^{lead} > 400$ GeV). The measurements in exclusive regions are provided in Figures I.1-I.4 in the Appendix I.

The absolute cross section as a function of p_T (given in Figure 11.9) shows a difference to predictions of $\sim 40\%$, while the shape is well described. For the measurements as a function of $\Delta\phi$ (Figure 11.10), two cases are shown, one considering the extended $\Delta\phi$ region (upper plots), and one focusing on the most back-to-back region (lower plots). The measured cross sections show a difference to predictions gradually increasing towards larger $\Delta\phi$ (nearer to π) up to $\sim 60\%$ ($\Delta\phi \sim \pi$). The shape is rather well described by the predictions when the $t\bar{t}$ contribution is simulate with aMC@NLO+PYTHIA8, while for the other predictions a difference of $\sim 20\%$ is observed.

Table 11.3 presents the χ^2 values (given by equation 11.3) comparing the normalized distribution in the measurements and the theory predictions predictions.

The results from the inclusive top-jet measurement are similar to the $t\bar{t}$ measurement indicating that there have been not introduced any bias through the background subtraction procedure.

Table 11.3: Comparison by using the χ^2 estimator of the measured normalized distributions with Standard Model theory predictions. The reported values are the chi_2 values and the number of degree of freedom $((N_{bins} - 1)$, where N_{bins} is the number of bins of the considered phase space for each distribution).

Observable	POWHEG+PYTHIA8 χ^2/ndf	POWHEG+HERWIGPP χ^2/ndf	aMC@NLO+PYTHIA8 χ^2/ndf
p_T leading jet (fiducial particle level)	4.84/7	6.8/7	4.45/7
p_T subleading jet (fiducial particle level)	3.7/7	3.1/7	3.90/7
p_T leading jet ($\|\eta\| < 0.5$)	11.58/7	13.0/7	8.09/7
p_T subleading jet ($\|\eta\| < 0.5$)	5.84/7	7.84/7	4.5/7
p_T leading jet ($0.5 < \|\eta\| < 2.4$)	3.6/7	4.29/7	5.13/7
p_T subleading jet ($0.5 < \|\eta\| < 2.4$)	5.98/7	4.76/7	8.21/7
$\Delta\phi$ (fiducial particle level)	41.57/7	49.38/7	11.0/7
$\Delta\phi$ (.. with fine binning in the back to back)	42.16/5	47.15/5	5.97/5
$\Delta\phi$ 400 GeV$< p_T^{lead} <$ 600GeV	26.3574	30.8921	8.9163/7
$\Delta\phi$, $p_T^{lead} > 600$ GeV	23.32/7	27.5/7	9.79/7

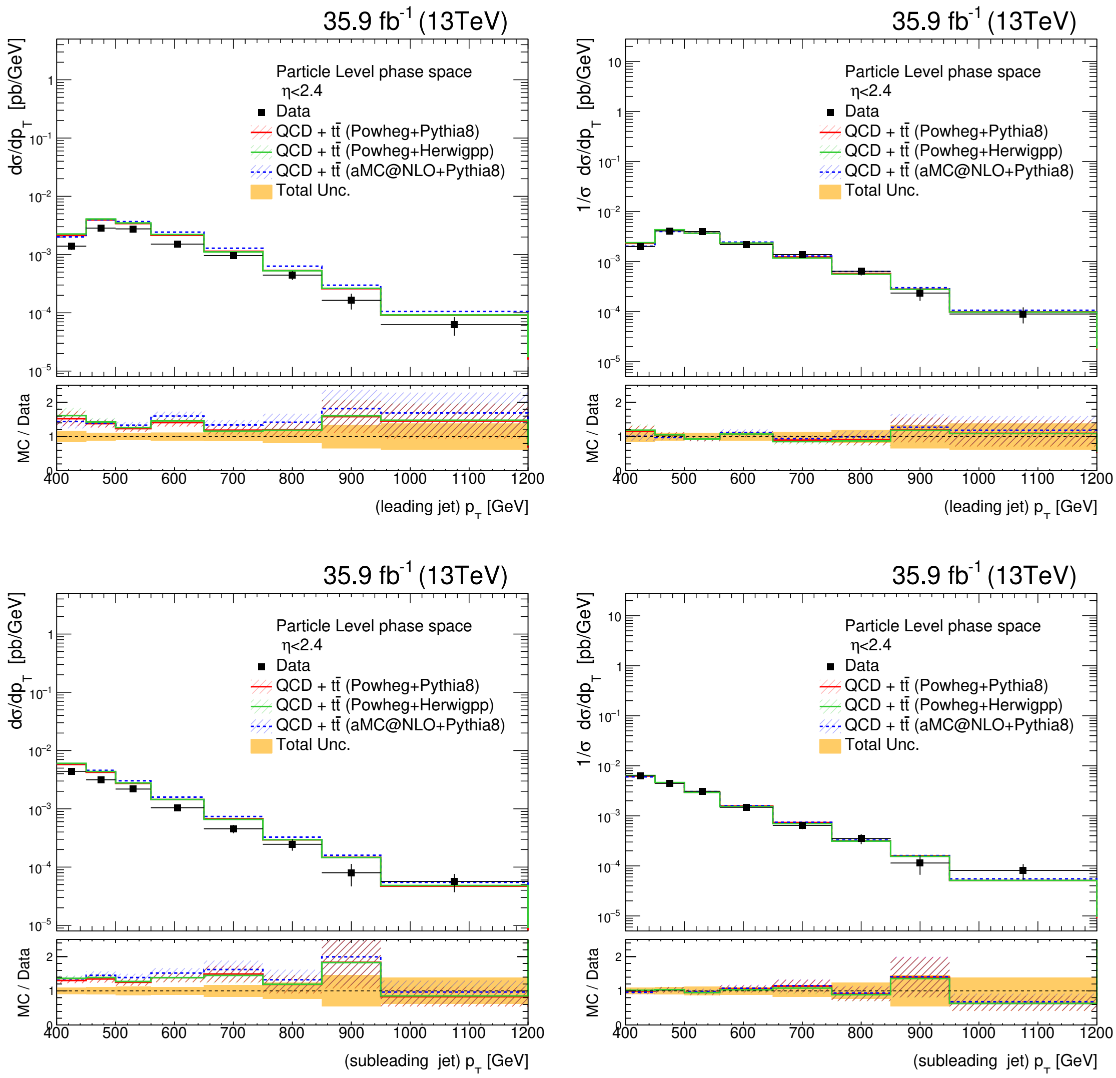


Figure 11.9: Cross section (left plots) and normalized distributions (right plots) differentially in (upper plots) the p_T of the leading jet, (lower plots), the p_T of the subleading jets, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The distributions correspond to the inclusive top-jet cross section.

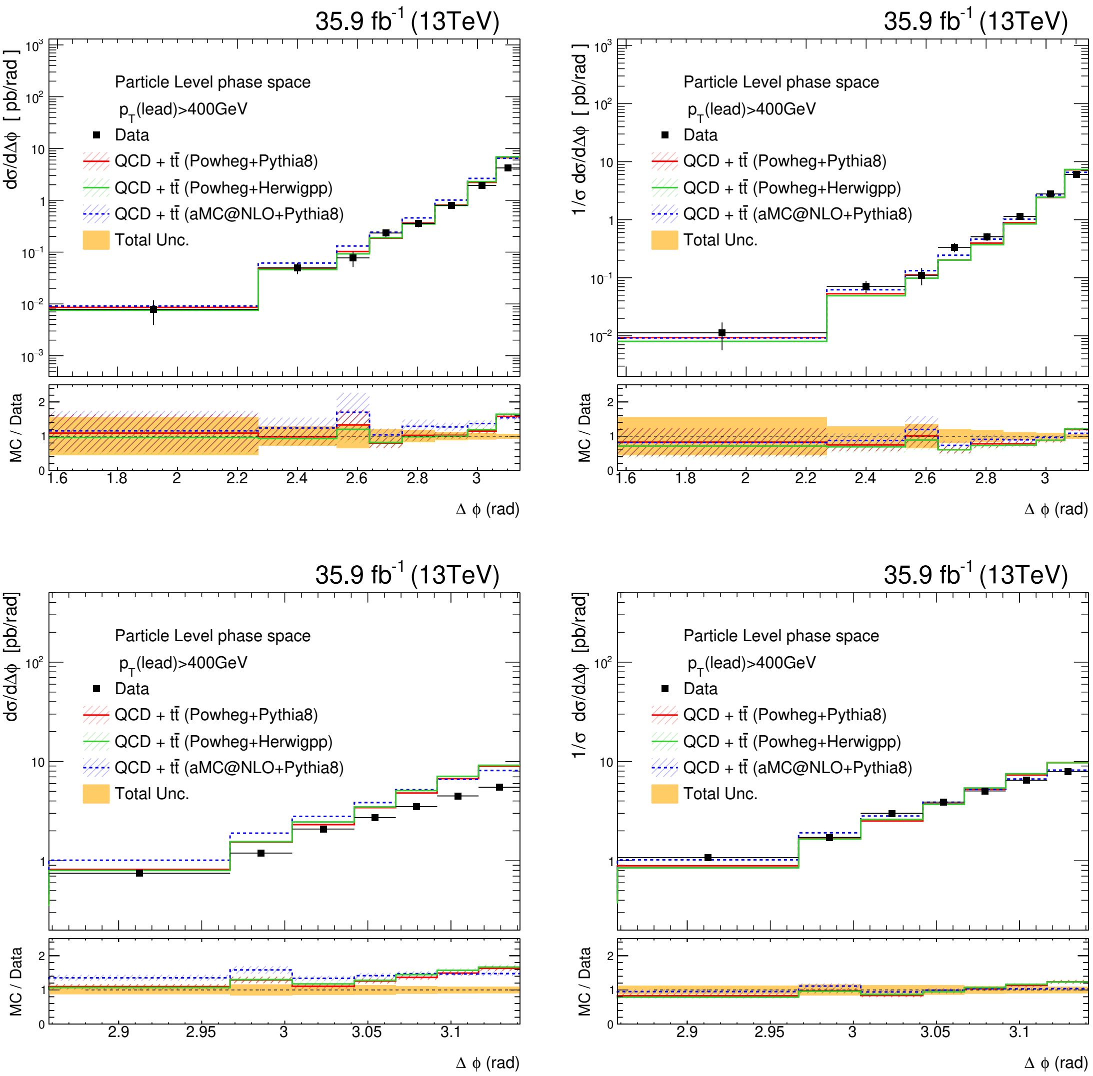


Figure 11.10: Cross section (left plots) and normalized distributions (right plots) differential in $\Delta\phi$ between the two leading jets, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The distributions correspond to the inclusive top-jet cross section.

11.3 Summary and interpretation of the results

In this chapter, the $t\bar{t}$ and inclusive top-jet measurements were compared to theory prediction, in the highly boosted regime. The predictions overestimate the data by $\sim$ 34%. The cross section differential in p_T show a difference up to $\sim$ 40%, while in the case of the distributions differential in $\Delta\phi$ differences are up to $\sim$ 60% increasing towards the $\Delta\phi$. The shape of the distributions is in good agreement within the uncertainties in the measurements, although for $\Delta\phi$, it is described slightly better with the aMC@NLO+PYTHIA8 prediction.

Other studies for $t\bar{t}$ measurements have reported softer spectra in comparison to the theory predictions at very high boosted topologies [182]. For instance, the ATLAS collaboration recently published the measurements of the $t\bar{t}$ pair production in the boosted regime at $\sqrt{s} = 13$ TeV [183], showing similar results to the ones obtained in this thesis.

Differences up to 40% are observed for the measured p_T spectra in comparison to theory predictions, while up to 60% for the spectra differential in $\Delta\phi$. The observed differences in the former one is an overall factor, meaning that if one rescale down the predicted cross section, a good agreement is obtained. In contradiction, in the case of $\Delta\phi$ the differences are observed increasing towards higher $\Delta\phi$ values. If one applies the same scale factor to bring the p_T distribution from Monte Carlo prediction in agreement to the data, to the $\Delta\phi$ predictions, differences are still observed in the most back-to-back region. However those discrepancies are within the systematics uncertainties obtained in the measurements.

The observed differences might come from Monte Carlo simulations, not taking soft gluon resummation and factorization breaking phenomena fully into account. In the predictions, soft gluon resummations are treated with the Parton Shower and colour reconnection models are also implemented, however, it is not clear, whether these simulations take into account all effects needed to properly describe the boosted $t\bar{t}$ scenarios, when both top quarks are separated in $\Delta\phi \sim \pi$. One would need to incorporate to Monte Carlo simulations higher order calculations (i.e NNLO, with soft gluon resummations at NNLL) to probably improve the description of the measurements.

Chapter 12

Summary and Outlook

The measured $t\bar{t}$ cross section differential in p_T of the leading and subleading jets, and in the azimuthal separation between the top-jets ($\Delta\phi$) have been presented, for very high p_T top-quarks, in proton proton collisions at a center of mass energy of $\sqrt{s} = 13$ TeV. The all-hadronic decay channel of the $t\bar{t}$ pair is considered. The studied phase space corresponds to boosted topologies, where the top quark decay products are collimated enough and clustered within a large cone-size jet. In signal events, the two leading jets have been considered as top-jets candidates. Top-jets with p_T from 400 GeV up to 1.2 TeV are measured. Additionally, the integrated cross section in the fiducial particle level phase space is compared to Standard Model Predictions, as well as, the differential distributions in their absolute and normalized forms.

The first achievement of the studies presented in this thesis is the particle level top-jet definition. Phenomenological studies are presented in order to identify in signal events, those ones where the two leading jets are properly reconstructing the top-quarks. The soft drop jet mass of the first subjet, obtained after applying the Mass Drop Tagger (MDT) turned to be a good observable for defining top-jets, since this subject is generally reconstructing the hadronic decays of the W-boson.

The all-hadronic $t\bar{t}$ decay channel is embedded in an overwhelming QCD background, hence, the suppression of this background contribution is one of the major challenge faced in the selection strategy. The procedure is built by our own top-tagging, trying to provide selection criteria at detector level similar as much as possible to the top-jet particle level definition. The selection strategy is based in four set of criteria, where two of them play an essential role for the background suppression: a multivariate selection criterion, derived from a Multilayer perceptron (MLP) technique; and the b-tagging criteria, requiring in both leading jets that at least one of the subleading jets are b-tagged by the CSVv2 algorithm. After the complete selection, the number of signal events is approximately four times larger that the amount of background events. The selection efficiency of signal events varies from 4% up to 8% with respect to the number of signal events passing the boosted dijet selection criteria. Those low efficiency values are consequence of the tight selection criteria needed for suppressing the background contribution. In total, 3094 events are selected in data.

After having applied the event selection, the background contribution are subtracted using a data driven method. The method consists of defining a control region, where the dominant contribution comes from QCD multidijet events, and hence, the background templates are estimated from the data.

After the background subtraction measurements are unfolded to particle level defined with our top-jet particle definition. The measured cross section in the fiducial particle level are determined to be:

$$\sigma = 321.1\text{fb} \pm 26.7\text{fb (stat. unc.)} \pm 28.4\text{fb (syst. unc.)} \tag{12.1}$$

In the Monte Carlo prediction provided with the POWHEG+PYTHI8 sample, the cross section takes value of:

$$\sigma = 429.11\text{fb} \pm 4.15\text{fb (stat. unc.)} \pm 47.02\text{fb (syst. unc.)}, \tag{12.2}$$

Hence, a difference form the Monte Carlo predictions of about 34% are observed. However, the Standard Model predictions are able to describe the shape of the distributions differential in p_T, although some discrepancies are observed in the case of the distributions differential in $\Delta\phi$.

The presented measurements give the possibility for studying phenomena of soft gluon resummation. I presented preliminary results of further investigations in the Workshop on Resummation, Evolution, Factorization (REF) in November 2018 [184]. Boosted topologies are studied using different Transverse Momentum Dependent (TMD) parton distribution functions, obtained from the Parton Branching method [185, 186]. Similar effects have been obtained in case of the p_T spectra o the Z-boson, when studying the lower p_T phase space [187], which is also sensitive to soft gluon resummation effects. Therefore, the measurements presented in this thesis open the possibility of testing interesting phenomena predicted in the Standard Model appearing for boosted $t\bar{t}$ topologies.

The measurements presented in this thesis can be further improved, for instance, considering the whole data set corresponding to the RunII period of the LHC. Larger statistics would benefit specially the smaller measured $\Delta\phi$ region. Another way to improve the measurements is understanding better the estimated systematic uncertainty.

Appendix A

High-p_T jet measurements at the HL-LHC

CMS Physics Analysis Summary

In this Chapter, studies related to high p_T jets in the High Luminosity LHC scenarios (HL-LHC) are presented. Standard Model predictions are provided for high p_T jets coming from different processes: inclusive jets, b-jets, boosted W-boson, and boosted $t\bar{t}$. The particle level cross section predictions differential in p_T and $\Delta\phi$ are compared between different processes. I was involved in these studies providing predictions for boosted top-jets.

These studies were included in the Ref. [3] *"Standard Model Physics at the HL-LHC and HE-LHC"*, included in the Yellow CERN report.

Available on the CERN CDS information server CMS PAS FTR-18-032

CMS Physics Analysis Summary

Contact: cms-phys-conveners-ftr@cern.ch 2018/12/14

High-p_{T} jet measurements at the HL-LHC

The CMS Collaboration

Abstract

Processes containing jets with high transverse momenta are studied for the upgraded CMS Phase-2 detector design at the High-Luminosity LHC assuming a center-of-mass energy of 14 TeV and an integrated luminosity of 3 ab^{-1}. The high luminosity allows to fully exploit high transverse momentum jets (boosted jets) and to differentiate between various jet types. Inclusive jet production, the production of jets originating from b or t quarks, as well as from W bosons are studied, with emphasis on the transverse momentum spectrum of the jets and angular correlations between the two jets with highest transverse momenta.

1 Introduction

The theory of quantum chromodynamics (QCD) is the underlying theory to describe interactions among quarks and gluons, i.e., partons. Inclusive jet production is a QCD process that allows to probe perturbative QCD calculations and the proton structure at the highest accessible scales. With the expected integrated luminosity of 3 ab^{-1} at the High Luminosity LHC (HL-LHC) [1] the accessible range in transverse momentum p_{T} can reach a few TeV, the highest p_{T} ever reached in a collider. A wide collection of inclusive jet measurements was carried out at the LHC by the ATLAS and CMS collaborations at center-of-mass energies $\sqrt{s} = 2.76$ [2, 3], 7 TeV [4–8], 8 TeV [9, 10] and 13 TeV [11, 12], and at lower $\sqrt{s}$ by experiments at other hadron colliders [13–17]. Measurements of inclusive jet cross sections are generally in agreement with theoretical calculations at next-to-leading order (NLO), or at next-to-next-to-leading order (NNLO) or NLO including resummation of next-to-leading logarithmic soft gluon terms. The jet cross sections play a crucial role in the determination of parton density functions and the strong coupling α_S, especially at the highest scales.

The improved tracking and b tagging performance at the HL-LHC [18, 19] and jet substructure analysis techniques will allow to discriminate jets of different origin. In this document, we study kinematic distributions of jets in inclusive jet production, as well as in final states containing bottom quark (b), top quark (t) jets, and W boson jets. In addition to the cross section as a function of the transverse momentum p_{T}, angular correlations between the jets with highest p_{T} are investigated. Higher order QCD radiation affects the distribution of the angular correlation, and especially the region where the jets are back-to-back in the transverse plane is sensitive to multiple "soft" gluon contributions, treated by all-order resummation and parton showers. This region is of particular interest since soft-gluon interference effects between the initial and final state can be significant [20, 21]. The azimuthal correlations in $\mathrm{t\bar{t}}$ production is of particular interest because of color interference effects [22, 23].

In inclusive jet production at 13 TeV [11] jet transverse momenta of up to about 2 TeV were reached. The main uncertainties in the high-p_{T} ($p_{\mathrm{T}} > 800$ GeV) region come from the jet energy calibration and statistical accuracy. Measurements of jets originating from b quarks are important to investigate the heavy-flavor contribution to the total jet cross section and to study the agreement of the measurement with available theoretical predictions. In particular, inclusive b production is very sensitive to higher-order corrections and to parton showers. By exploiting the long lifetime of the B hadrons produced by b quarks, one can identify b jets. Since the b tagging algorithm strongly relies on the tracking information, only jets within the tracker acceptance can be considered. Measurements of inclusive b jet cross sections were already performed at the Tevatron [24, 25] and at HERA [26, 27]. They exhibited a large disagreement between data and theory and helped to improve our understanding of the b quark production and fragmentation. Measurements performed at $\sqrt{s} = 7$ TeV by the ATLAS [28, 29] and CMS [30, 31] collaborations show a reasonable agreement with theoretical calculations.

In top quark production processes, t jets can be defined when the top quark decays hadronically and all decay products can be clustered into a single jet. The production of W bosons is studied in the high-p_{T} region, where the W boson decays hadronically and are reconstructed as jets. We apply jet substructure techniques [32] to discriminate the jets originating from top quarks and W bosons from the QCD background. Measurements of t-jet cross sections were performed at $\sqrt{s} = 8$ TeV in Ref. [33] and at $\sqrt{s} = 13$ TeV in Refs. [34, 35] where jets with p_{T} up to 1 TeV were observed.

Angular correlations between the two leading p_{T} jets and their dependency on the production process are also investigated. The analysis technique is inspired by previous analyses on azimuthal correlations in high-p_T dijet production [36, 37] .

With the luminosity expected at HL-LHC, measurements of cross sections of jet production can reach transverse momenta of a few TeV with reasonable precision. The program of jet physics will substantially profit from the HL-LHC data since higher scales can be reached and the region of very low partonic momentum fractions x can be accessed, where the parton density becomes large.

2 Analysis strategy

All results discussed in this note are based on PYTHIA 8 [38] with tune CUETP8M1 [39] supplemented with the Delphes simulation [40] of the CMS Phase-2 detector, except the study of boosted W bosons, where particle level distributions are presented. In inclusive jet and b jet production, the size of the higher order corrections are estimated using the POWHEG generator [41] and were found to be of the order of 20%. For $\mathrm{t\bar{t}}$ jet production the size of the higher order corrections can be even larger. For example, a 20% difference in the cross section will lead to a difference of up to 10% in the predicted statistical uncertainty.

The higher luminosity at the HL-LHC will allow to extract jet energy corrections and b tagging scale factors at high p_T with much higher precision, leading to smaller systematic uncertainties. The extended tracker coverage up to a pseudorapidity of $|\eta| = 4$, better tracking performance and expected progress in machine learning (ML) techniques will especially improve the jet flavor tagging based on jet substructure. With the extended tracker coverage, b jets can be measured in the forward region, which is currently inaccessible.

However, even the jet energy calibration can benefit from these methods, for example in the extraction of the flavor-dependent jet energy corrections. The analysis of inclusive jet production can also benefit from the extended tracker coverage since jets reconstructed from particle-flow [42–44] objects incorporating tracks are typically much more precise than jets reconstructed from only calorimeter objects. In Run 2 this was visible in both the size of jet energy resolution in the central and forward direction and in the uncertainties of the jet energy scale and jet energy resolution corrections. Since the jet energy corrections are extracted from in-situ measurements, such as dijet or γ-jet final states, their precision is expected to improve with increasing luminosity.

3 Systematic uncertainties

3.1 The b tagging at Phase-2 and related systematic uncertainties

Most of the presented studies rely on b tagging. The cross section of b jet production is about 3–4% of inclusive jet production cross section. In order to achieve sufficiently high purity of the measured b-tagged jets, the light-flavor (udsg) tagging efficiency (referred to as mistagging efficiency) must be as low as possible. For analyses presented in this note, the DeepCSV b tagging algorithm [45] trained for the HL-LHC is used.

The b tagging efficiencies predicted by the simulation are slightly different compared to that measured in data. To correct for this difference, so-called scale factors (SF) are introduced, which are defined as the ratio between the b tagging efficiency in data and simulation. These scale factors are obtained from measurements of b jet enhanced processes [45]. The efficiencies of b tagging, c tagging and light-flavor tagging are corrected by the corresponding scale factors. In this note, we assume that the b tagging scale factors are equal to one, but with uncertainties according to the ones obtained in Ref. [46].

In our studies we use a tight working point, defined by a light-flavor (udsg) mistag rate of 0.1% (for a medium working point, the mistag rate is 1 %, leading to a much higher background contribution). The expected uncertainty of the b tagging scale factor is 15% [46] as shown in Fig. 1. The b tagging uncertainty grows towards higher p_T, since it is more difficult to reconstruct a secondary vertex as the tracks become nearly collinear. An overview of the systematic uncertainties in b tagging is given in Table 1, more details are given in Ref. [46].

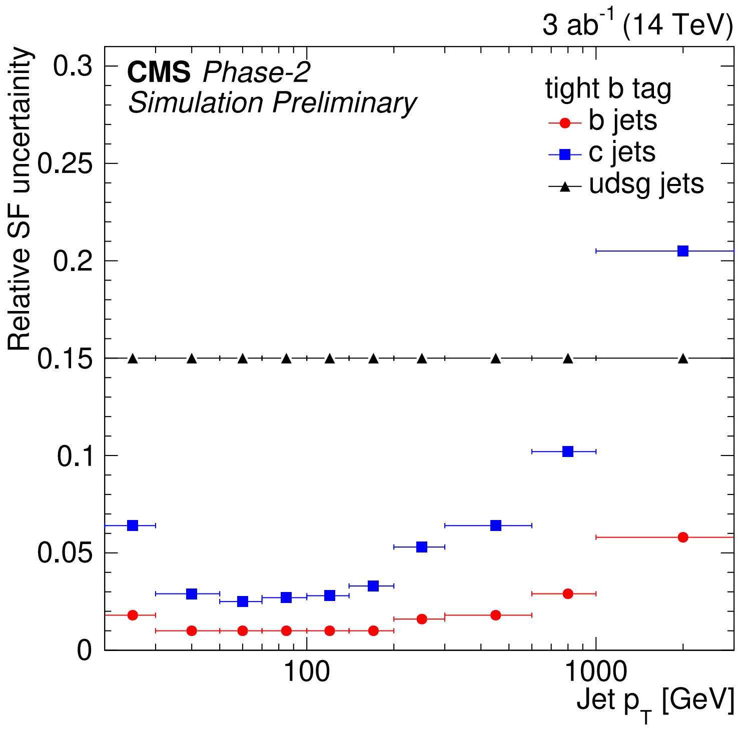


Figure 1: Expected b- tagging scale factor uncertainties as a function of jet p_T for the tight working point [46].

Table 1: The b tagging scale factor (SF) uncertainties for several p_T values [46]. The scale factor uncertainties for jets with $R = 0.4$ and $R = 0.8$ are assumed to be identical.

p_T [GeV]	100	500	2000
b tagging SF unc.	1%	2%	6%
c tagging SF unc.	3%	7%	20%
light-flavor tagging SF unc.	15%	15%	15%

The tagging efficiencies, as obtained from the Delphes simulation, and the related flavor composition of the b-tagged sample of inclusive jets are shown in Fig. 2.

Figure 2 (left) shows the tagging efficiencies as a function of the jet p_T. The b tagging efficiency decreases from $\sim$ 70% at $p_T = 100$ GeV to about 20% at $p_T = 1$ TeV, which leads to a larger light-flavor contamination of the b-tagged sample as shown in Fig. 2 (right). Jets containing charm hadrons have similar properties as jets with a B hadron, e.g., the presence of a displaced vertex, and there is a non-negligible probability to misidentify a c jet as a b jet. This probability is rather constant as a function of p_T, as shown in Fig. 2 (right).

To evaluate the expected systematic uncertainties from b tagging, we assume, for simplicity,

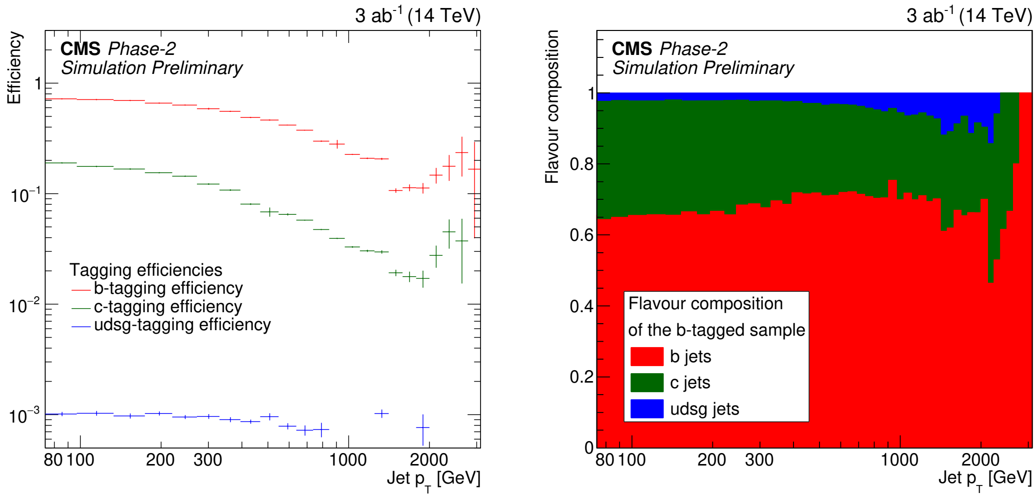


Figure 2: Predicted b-tagging efficiencies with the tight working point for jets with $R = 0.4$ (left). Predicted flavor composition of the b-tagged sample (right).

that only the b-tagged events are used to obtain the cross section:

$$\sigma_b^{\text{data}} = \frac{\sigma_{\text{b tag}}^{\text{data}}}{\sigma_{\text{b tag}}^{\text{MC}}} \sigma_b^{\text{MC}}, \tag{1}$$

where $\sigma_b^{\text{data,MC}}$ is the b jet cross section in the data and Monte Carlo simulation (MC), respectively. The cross section of b-tagged jet production in the MC simulation $\sigma_{\text{b tag}}^{\text{MC}}$ can be calculated as:

$$\sigma_{\text{b tag}}^{\text{MC}} = \sigma_b^{\text{MC}} \epsilon_b + \sigma_c^{\text{MC}} \epsilon_c + \sigma_l^{\text{MC}} \epsilon_l \,, \tag{2}$$

where $\epsilon_{b,c,l}$ are the probabilities that b jet, c jet or light-flavor jet is b tagged and $\sigma_{b,c,l}^{\text{MC}}$ are the b jets, c jets and light-flavor jet cross sections in the Monte Carlo simulation.

In Eq. (1), the background from wrongly tagged b jets is implicitly subtracted. This background fraction increases the resulting statistical uncertainty of the true level cross section:

$$\frac{\Delta\sigma}{\sigma} = \frac{\sqrt{N_b + N_{bg}}}{N_b} \tag{3}$$

where N_b is the number of events with tagged b jets, i.e., the signal, and N_{bg} is the number of events in which other flavors were mistagged, i.e., the background.

In the calculation of the resulting systematic uncertainty of the predicted cross section, the b tagging and c tagging SF uncertainties are assumed to be correlated (as treated in Run 2), whereas light flavor tagging is taken as uncorrelated with the other two.

The expected uncertainty of the inclusive b jet cross section as a function of p_{T} shown in Fig. 3. The uncertainty coming from the uncertainty of the light-flavor and heavy-flavor SF varies between 2% at low p_{T} and 10% at large p_{T}. The b tagging systematic uncertainty is dominated by the b+c SF uncertainty in the high-p_{T} region.

The b tagging performance is also crucial for top quark tagging, since a b-tagged subjet is required (Section 4.4). The b tagging performance of jets with larger cone size is comparable to one of the jets with $R = 0.4$. It is important to mention that (in case of dijet production)

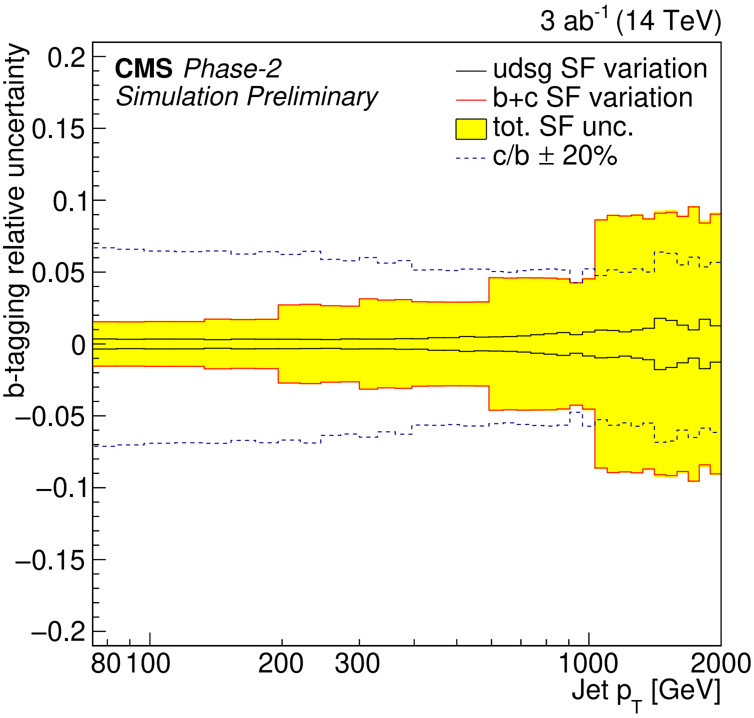


Figure 3: Expected b-tagging systematic uncertainty of the inclusive b-jet cross section.

requiring one jet to be b tagged increases the probability that the other jet is also tagged and the background contamination is lower compared to inclusive jet production.

The model uncertainty of the b tagging is related to differences in jet flavor composition in MC and data. This can affect the predicted amount of background from c and light flavors and, consequently, the measured particle-level cross section. To evaluate this model dependence, the flavor composition in PYTHIA 8 and in HERWIG ++ was compared in Ref. [45]. The flavor fractions b/c were found to differ maximally by 20% and this value is considered as a model uncertainty (as indicated in Fig. 3). The amount of light-flavor jets is well constrained by the inclusive jet cross section and, therefore, no model dependence is considered.

3.2 Other sources of systematic uncertainties

In addition to the uncertainties from b tagging, the uncertainties related to the jet energy calibration can significantly contribute. Based on previous experience [47], they can be about 1–2% within the tracker acceptance, where the 2% value is expected at lower p_T mainly due to the uncertainty introduced by the subtraction of effects from additional proton-proton collisions (pileup). In the high-p_T region the dominant component in the jet energy scale uncertainty (JES) is due to the jet flavor dependence of the detector response, which is slightly different for quark- and gluon-induced jets. A 1% shift in the energy calibration leads to about 5% change of the cross section $d\sigma/dp_T$ if the cross section falls as $\propto p_T^{-5}$.

The uncertainty in the measured integrated luminosity is assumed to be 1%.

4 Results

4.1 Inclusive jet production

The inclusive jet cross section at particle level, without any flavor requirement, is shown as a function of p_T for a rapidity range of $|y| < 0.5$ in Fig. 4 (left). The statistical uncertainty, visible in the ratio, corresponds to an integrated luminosity of 3 ab^{-1}. The systematic uncertainty (shown as the grey band) is dominated by the jet energy scale uncertainty (JEC). Also shown is the expected inclusive jet cross section at $\sqrt{s} = 13$ TeV with uncertainties corresponding to an integrated luminosity of 150 fb^{-1}.

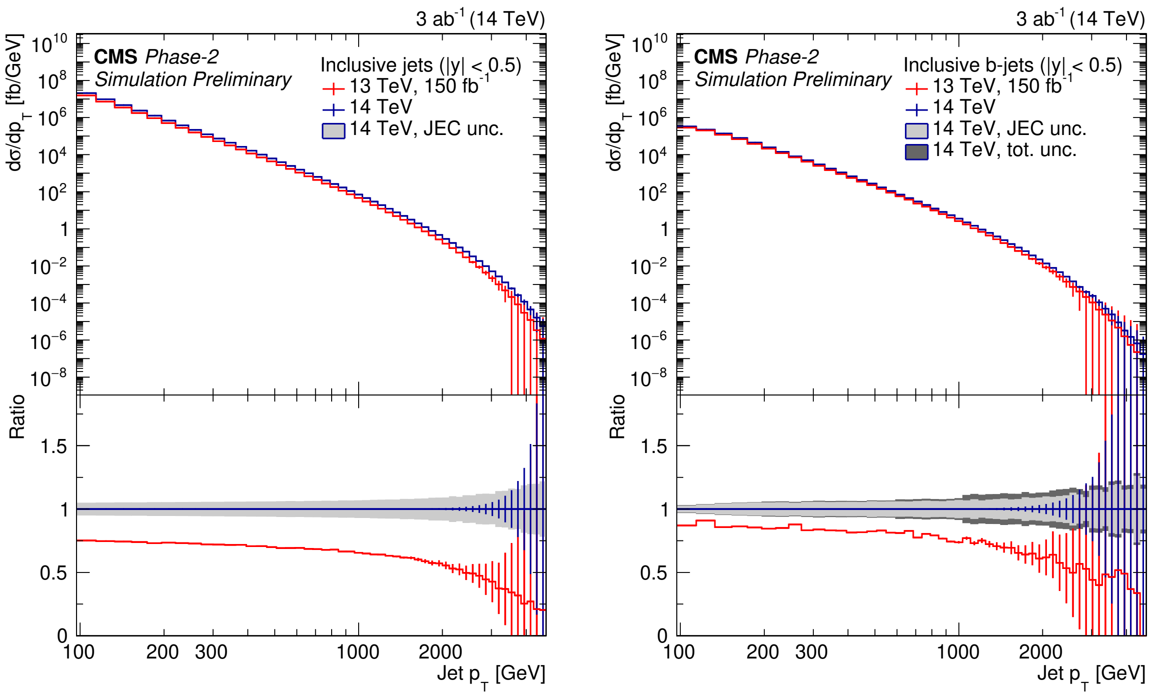


Figure 4: Comparison of the 13 TeV and 14 TeV cross sections for inclusive jet (left) and inclusive b jet (right) production at particle level as a function of p_T in $|y| < 0.5$. The lower panel shows the ratio to the jet cross section at 14 TeV. The uncertainties in the ratio represent the expected statistical uncertainty assuming 150 fb^{-1} and 3 ab^{-1}, respectively. The systematic uncertainty is shown for 14 TeV and is dominated by the jet energy scale uncertainty for inclusive jet production, and by the jet energy scale uncertainty and by the uncertainties from b tagging for the inclusive b jets.

Compared to Run 2 measurements at $\sqrt{s} = 13$ TeV the increase of the center-of-mass energy leads to about twice larger cross section at highest p_T. Taking into account the much higher luminosity and the higher cross section, the statistical uncertainty is expected to be around six times smaller, compared to the analysis of the Run 2 data. A measurement of the inclusive jet cross section up to $p_T \sim 4$ TeV can be performed with about 10 events above this threshold.

4.2 Inclusive b jet production

In Fig. 4 (right), the inclusive b jet cross section at particle level as a function of p_T for $|y| < 0.5$ is shown. The statistical uncertainty corresponds to an integrated luminosity of 3 ab^{-1}, where the b tagging efficiency, as described in Section 3.1, is included. The systematic uncertainty of around 5% in the low-p_T region rising to around 10% at high-p_T includes uncertainties from jet energy scale calibration as well as uncertainties from b tagging. For comparison, also the expected cross section at 13 TeV with uncertainties corresponding to an integrated luminosity of 150 fb^{-1} is shown. Compared to Run 2 measurements at $\sqrt{s} = 13\,\text{TeV}$, the increase of the

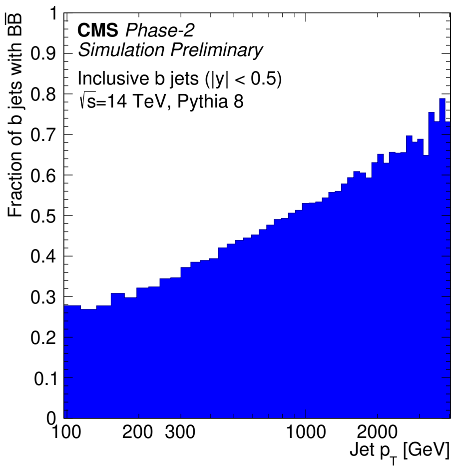


Figure 5: Fraction of b jets containing both a B and a $\overline{\text{B}}$ hadron as a function of the jet p_T.

center-of-mass energy leads to about twice larger cross section at largest p_T. A measurement of the inclusive b jet cross section can reach transverse momenta of $p_T \sim 3\,\text{TeV}$ with about 30 events above this threshold, where the details depend crucially on the b tagging performance at highest p_T (the b tagging SF uncertainties were derived only up to 2 TeV [46] and the uncertainty is expected to increase with p_T).

In the high-p_T region, the mass of the b quark becomes negligible with respect to the jet momentum. This leads to a high probability that the b quark is not only produced in the hard subprocess, but also during further QCD radiation, simulated with a parton shower. In such cases, a pair of B hadrons inside the b jet can be observed, where one consists of a b quark, and the second of a $\overline{\text{b}}$ quark. The fraction of such jets as a function of p_T, as predicted by PYTHIA 8, is shown in Fig. 5.

4.3 High-p_T b$\overline{\text{b}}$ jets

The angular correlations $\Delta\phi = |\phi_2 - \phi_1|$ and $|\Delta y| = |y_2 - y_1|$ between the two leading p_T jets are studied. The flavor dependence of the angular correlations are investigated by selecting dijet events with at least one or two b-jets. The leading jet p_T must satisfy $400 < p_T < 800\,\text{GeV}$ or $p_T > 1600\,\text{GeV}$ while the subleading jet is required to be above 200 GeV. The event selection follows closely the Run 1 and Run 2 measurements [36, 48, 49].

The angular resolution is found to be 0.07 rad for $|\Delta\phi|$, obtained from the Delphes simulation (and consistent with the resolution found in Run 2 [36]). The resolution in $|y|$ has a similar size. The systematic uncertainties are treated as in the previous section and are dominated by the jet energy scale and b tagging scale factors uncertainties.

In Fig. 6, the particle-level cross section as a function of $\Delta\phi$ is shown. The statistical uncertainty corresponds to an integrated luminosity of 3 ab^{-1} including b tagging as described in Section 3.1. The systematic uncertainty includes uncertainties from jet energy scale calibration as well as uncertainties from b tagging. It is around 5% in the low-p_T region and rises to 10% at high p_T.

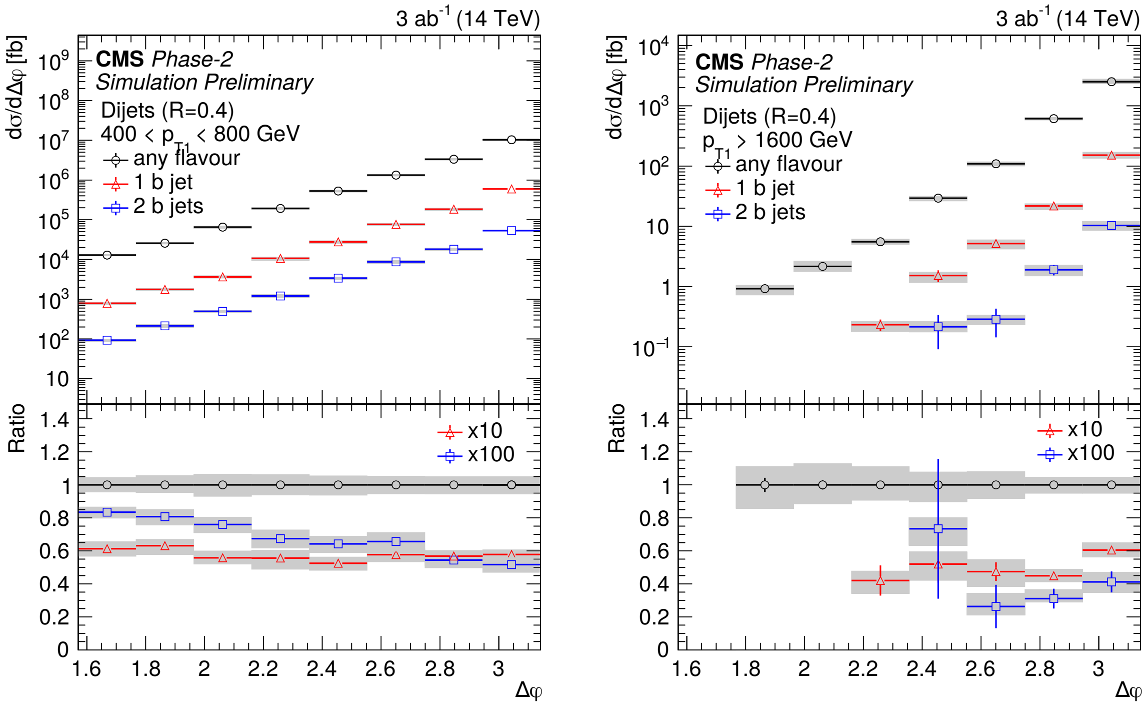


Figure 6: Distribution of the azimuthal correlation $\Delta\phi$ between two leading jets at the particle level for leading jet p_T between 400 GeV and 800 GeV (left) and above 1600 GeV (right). The uncertainties represent the expected statistical uncertainty assuming 3 ab^{-1}. The systematic uncertainty includes the jet energy scale uncertainty (JEC) and uncertainties from b tagging.

The shape of the $\Delta\phi$ distribution of inclusive dijet production differs from the one of $\mathrm{b\bar{b}}$ jet production. When both leading jets are required to be b jets, the dominant production channel is $gg \to \mathrm{b\bar{b}}$. Since the gluons in the initial state radiate more than quarks, the p_T of the $\mathrm{b\bar{b}}$ system is expected to be higher and, consequently, the jets are more decorrelated in $\Delta\phi$. At larger p_T ($p_\mathrm{T} > 1600\,\mathrm{GeV}$) this effect becomes less visible, also because of the restricted range in $\Delta\phi$ due to statistics. There is no apparent difference between single b jet production and the inclusive cross section. The figures in this section include the ratio with respect to the jet+jet differential cross section (the relative uncertainties shown in the lower panel correspond the statistical and systematic uncertainties of the production cross sections) to visualize the size of the uncertainties and the difference in shape.

In Fig. 7 the particle-level cross section as a function of $|\Delta y|$ is shown, with statistical uncertainties corresponding to an integrated luminosity of 3 ab^{-1} including b tagging as described in Section 3.1. Larger differences between the cross sections of different flavors can be seen, where the b jets are preferably produced in the central region. The main reason for this observation is the suppression of the b quark density in the proton with respect to the light flavors at high x. In Run 2 similar distributions were studied for inclusive dijet production [37].

In conclusion, different regions in rapidity and $\Delta\phi$ are sensitive to the different parton-level processes and thus can provide constraints on the parton densities, especially when the jet flavor is measured.

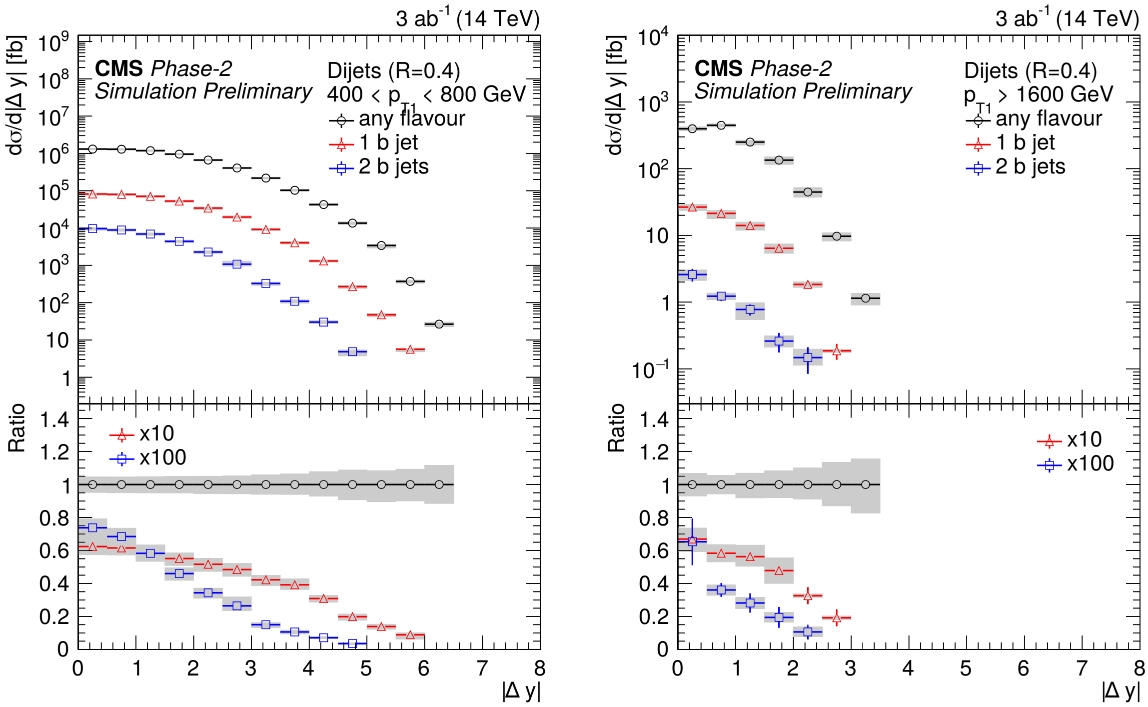


Figure 7: Distribution of the rapidity difference $|\Delta y|$ between two leading jets at the particle level for leading jet p_T between 400 GeV and 800 GeV (left) and above 1600 GeV (right). The uncertainties represent the expected statistical uncertainty assuming 3 ab^{-1}. The systematic uncertainty includes the jet energy scale uncertainty (JEC) and uncertainties from b tagging.

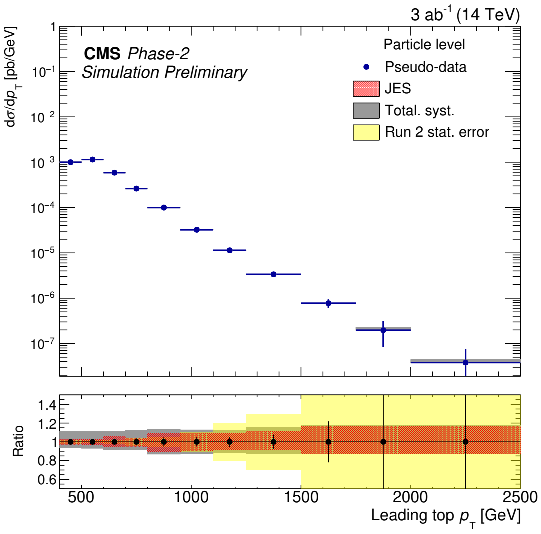
4.4 High-p_{T} $\mathrm{t\bar{t}}$-jets

Jets originating from t quarks provide further information on the flavor dependence of QCD cross sections. The t jets are defined in the fully hadronic decay mode, where the t quark decays into a W boson and a b quark with the W boson decaying hadronically. The measurement can be efficiently performed in the boosted region, with jet $p_{\mathrm{T}} > 400$ GeV. In contrast to the inclusive and b jet measurements, a jet radius of $R = 0.8$ is used to ensure all decay products are clustered into one jet. We use a particle level definition for the t jet, i.e., the jet must contain a B hadron as well as two subjets, where the subjet with largest p_{T} must have a mass of $50 < m_{\mathrm{subjet}} < 150\,\mathrm{GeV}$ and can be identified as a W boson candidate. The subjets are found by applying the soft-drop algorithm [50] which also suppresses the contribution from soft partons, as well as from underlying event and (at detector level) pileup.

Of particular interest are the azimuthal correlations between $\mathrm{t\bar{t}}$ jets in the back-to-back region in the transverse plane, as they might be subject to significant corrections due to color correlations between initial- and final-state soft gluons [22, 23].

Top quark jets can be distinguished from the dominant background of QCD multijets through substructure techniques at the detector level: the soft-drop algorithm (with $z_{\mathrm{cut}} = 0.1$ and $\beta = 0$) is applied to remove the contribution from soft partons [50]. The soft-drop mass is required to be around the top quark mass and the N-subjettiness variables τ_{N} are used to suppress the QCD background [51]. Since the b quark should be present in the jet, the b tagging technique can be used to further suppress QCD background. Only leading and subleading jets with $p_{\mathrm{T}} > 400$ GeV and $|\eta| < 2.5$, $m_{\mathrm{SD}} > 105$ GeV, and $\tau_3/\tau_2 < 0.68$ together with a b tag (with tight working point) are kept as $t\bar{t}$ jets candidates at the detector level. These selection criteria are based on the experience from Run 2 analyses [33], giving confidence on good signal selection and significant background rejection.

In Fig. 8 (left), the particle level cross section for $t\bar{t}$ jets is shown as a function of the leading jet transverse momentum. The statistical uncertainties correspond to an integrated luminosity of $3\ \mathrm{ab}^{-1}$ including efficiencies for selecting t jets at the detector level. The efficiency for selecting $t\bar{t}$ jets ranges from 25 % at $p_{\mathrm{T}} \sim 500$ GeV to about 5 % at $p_{\mathrm{T}} > 1.5$ TeV, as obtained from the Delphes simulation. Systematic uncertainties originate from b tagging, jet energy scale, and the uncertainty related to the jet substructure, i.e., to the jet mass scale and the jet mass resolution. Both of them affect the shape of the m_{SD} distribution. Based on the analyses from Run 2, the jet mass scale uncertainty in the barrel region is around 1% and the jet mass resolution uncertainty is around 10%.



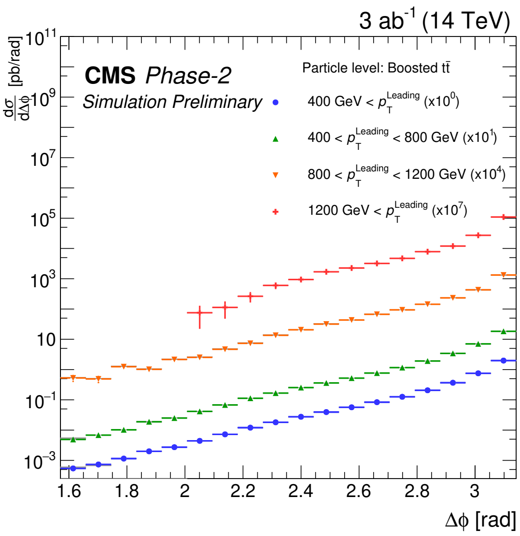


Figure 8: The cross section at particle level as a function of the leading-jet p_{T} in $t\bar{t}$ events (left), and as a function of $\Delta\phi$ between the two leading $t\bar{t}$ jets (right). The statistical uncertainties correspond to an integrated luminosity of $3\ \mathrm{ab}^{-1}$, including efficiencies from the selection of t jets at detector level. The systematic uncertainties are described in the main text.

In Fig. 8 (right), the azimuthal correlation for $t\bar{t}$ jets is shown for various ranges of the leading jet p_{T}. The uncertainties are obtained in the same way as for Fig. 8 (left). The efficiency for selecting $t\bar{t}$ jets ranges from 10% at small $\Delta\phi$ to about 20% at $\Delta\phi \sim \pi$, as obtained from the Delphes simulation.

4.5 W boson production at large p_{T}

Jets originating from hadronic decays of W and Z bosons form also a contribution to inclusive jet cross sections. For simplicity, we consider here only W boson production which has a hadronic branching fraction of $\sim$70%. As in the case of the t jet, jets with a radius of $R = 0.8$ have to be considered to ensure that all decay products of the W boson are included in the jet. Of particular interest are again the azimuthal correlations between a highly boosted, high-p_{T} W boson decaying hadronically and the recoiling jet. The kinematic situation is very similar as in the case of $t\bar{t}$ jets, with the difference that the jet from the hadronically decaying vector boson has no color connection to the initial-state partons, and thus the azimuthal correlation does not suffer from color correlations between initial and final-state partons.

The W boson jets are identified by anti-k_T jets with $R = 0.8$, where the hadronic decay products of the W boson are fully contained inside the jet. The major background is coming from the QCD multijets. To suppress this background, the soft-drop mass of the jet is required to be close to the W mass, namely $65 < m_{SD} < 105$ GeV. The particle-level cross section as a function of the p_T of the W boson candidates of W+jet events where the W boson jet is required to have a $p_T > 400$ GeV and $|\eta| < 2.5$ is shown in Fig. 9 (left). In Fig. 9 (right) the azimuthal correlation between the jet originating from the W boson and the recoil jet is shown for several intervals of the W boson transverse momentum. The statistical uncertainties do not include any correction from efficiencies, since the background from QCD processes is large and would need further studies.

One of the interesting features of this process is the absence of color connection between the W boson jet and the initial and/or final state, in contrast to dijets or $t\bar{t}$ jets.

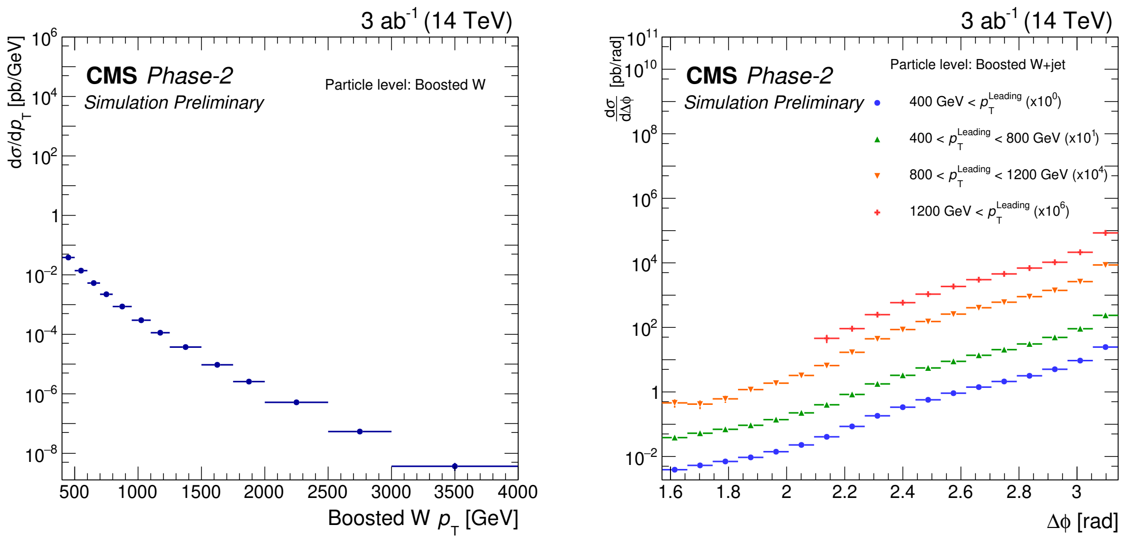


Figure 9: The cross section as a function of p_T for hadronically decaying W bosons (left), and as a function of $\Delta\phi$ between the jet originating from the W boson and the recoil jet (right). The statistical uncertainties do not include selection efficiencies.

4.6 Overview of the jet measurements

In Fig. 10 we show a comparison of the jet cross sections as a function of p_T and as a function of $\Delta\phi$ for the different processes discussed above. For comparison, here all use $R = 0.8$. In Fig. 10 (left) the inclusive b jet cross section is shown (for comparison with the inclusive jet cross section), while in Fig. 10 (right) the two-b-jet cross section is shown. Except for the cross section for W production, the statistical uncertainties shown correspond to an integrated luminosity of 3 ab^{-1} including efficiencies due to b tagging and selection at the detector level, estimated from the Delphes simulation.

It can be seen that the shapes of the p_T spectra are comparable but in the normalization the $t\bar{t}$ cross section is about ten thousand times smaller than the inclusive jet cross section. The ratio to the inclusive dijet cross section as a function of $\Delta\phi$ illustrates the differences in shape of the $\Delta\phi$ distribution of the different processes (all processes are normalized at $\Delta\phi = \pi$), which depend on the partonic configuration of the initial state.

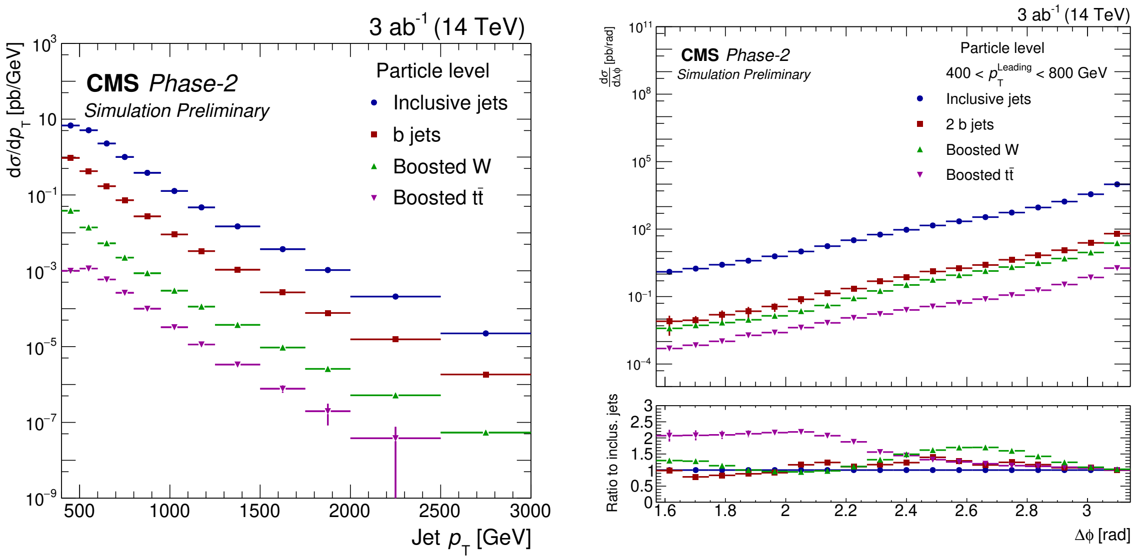


Figure 10: The overview of the particle-level differential jet cross sections (with $R = 0.8$) as a function of p_T (left) and $\Delta\phi$ (right) for various processes. In the left plot the inclusive b jet cross section is shown (for comparison with the inclusive jet cross section), while for $\Delta\phi$ the two-b-jet cross section is shown. For the ratio the normalization is fixed arbitrarily at $\Delta\phi = \pi$. The cross section of W production does not include statistical uncertainties corrected for efficiencies and background subtraction.

5 Conclusion

We have determined the expected reach in p_T for inclusive jets and b jets at the HL-LHC. The HL-LHC data will allow to probe the proton structure and perturbative QCD in general at the highest ever achieved scales. The inclusive b jet production is a process, which can be identified with high accuracy. We show that at high p_T, the b jets are more and more affected by gluon splitting.

The angular correlation between the two leading p_T jets is evaluated as a function of the $\Delta\phi$ and $|\Delta y|$ variables. It is demonstrated that these variables together with the possible b-jet requirement enhance the sensitivity to the different partonic content of the proton. The studies are complemented with a particle-level study of boosted W+jet events. The angular correlation variables are sensitive to perturbative soft-gluon radiation and are important for calculations involving soft gluon resummation.

The boosted $t\bar{t}$ cross section in the high p_T region is studied, where even the top quark mass becomes negligible. Consequently, the top quark pair is produced at a rate comparable to that of light quarks. However, the prominent process at high p_T is the quark-quark scattering which makes the top quark pair production still suppressed, as the probability to produce top quarks within the QCD evolution (in the shower) is low. This is in contrast to the case of b quarks, which at high p_T typically are produced within the QCD evolution, i.e., in the initial-state shower.

With an integrated luminosity of 3 ab^{-1}, inclusive jet cross section measurements can reach a $p_T \sim 4$ TeV, inclusive b jet measurements can reach a $p_T \sim 3$ TeV, jets originating from hadronic top quarks can reach a $p_T \sim 2$ TeV, and boosted hadronically decaying W bosons can access the region of $p_T \sim 2.5$ TeV.

References

[1] G. Apollinari et al., “High-Luminosity Large Hadron Collider (HL-LHC) : preliminary design report”, doi:10.5170/CERN-2015-005.

[2] ATLAS Collaboration, “Measurement of the inclusive jet cross-section in pp collisions at $\sqrt{s} = 2.76$ TeV and comparison to the inclusive jet cross-section at $\sqrt{s} = 7$ TeV using the atlas detector”, *Eur. Phys. J. C* **73** (2013) 2509, doi:10.1140/epjc/s10052-013-2509-4, arXiv:1304.4739.

[3] CMS Collaboration, “Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76$ TeV”, *Eur. Phys. J. C* **76** (2016), no. 5, 265, doi:10.1140/epjc/s10052-016-4083-z, arXiv:1512.06212.

[4] ATLAS Collaboration, “Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector”, *Eur. Phys. J. C* **71** (2011) 1512, doi:10.1140/epjc/s10052-010-1512-2, arXiv:1009.5908.

[5] CMS Collaboration, “Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. Lett.* **107** (2011) 132001, doi:10.1103/PhysRevLett.107.132001, arXiv:1106.0208.

[6] ATLAS Collaboration, “Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector”, *Phys. Rev. D* **86** (2012) 014022, doi:10.1103/PhysRevD.86.014022, arXiv:1112.6297.

[7] CMS Collaboration, “Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector”, *Phys. Rev. D* **87** (2013) 112002, doi:10.1103/PhysRevD.87.112002, arXiv:1212.6660.

[8] ATLAS Collaboration, “Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using 4.5 fb^{-1} of data with the ATLAS detector”, *JHEP* **02** (2015) 153, doi:10.1007/JHEP02(2015)153, arXiv:1410.8857. [Erratum: doi:10.1007/JHEP09(2015)141].

[9] CMS Collaboration, “Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV”, *JHEP* **03** (2017) 156, doi:10.1007/JHEP03(2017)156, arXiv:1609.05331.

[10] ATLAS Collaboration, “Measurement of the inclusive jet cross-sections in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *JHEP* **09** (2017) 020, doi:10.1007/JHEP09(2017)020, arXiv:1706.03192.

[11] CMS Collaboration, “Measurement of the double-differential inclusive jet cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **76** (2016), no. 8, 451, doi:10.1140/epjc/s10052-016-4286-3, arXiv:1605.04436.

[12] ATLAS Collaboration, “Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *JHEP* **05** (2018) 195, doi:10.1007/JHEP05(2018)195, arXiv:1711.02692.

[13] UA2 Collaboration, “Observation of very large transverse momentum jets at the CERN $p\bar{p}$ collider”, *Phys. Lett. B* **118** (1982) 203, doi:10.1016/0370-2693(82)90629-3.

[14] UA1 Collaboration, "Hadronic jet production at the CERN proton-antiproton collider", *Phys. Lett. B* **132** (1983) 214, doi:10.1016/0370-2693(83)90254-X.

[15] CDF Collaboration, "Measurement of the Inclusive Jet Cross Section using the k_T algorithm in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV with the CDF II Detector", *Phys. Rev. D* **75** (2007) 092006, doi:10.1103/PhysRevD.75.119901,10.1103/PhysRevD.75.092006, arXiv:hep-ex/0701051. [Erratum: Phys. Rev.D75,119901(2007)].

[16] D0 Collaboration, "Measurement of the inclusive jet cross section in p$\bar{\text{p}}$ collisions at $\sqrt{s} = 1.96$ tev", *Phys. Rev. Lett.* **101** (2008) 062001, doi:10.1103/PhysRevLett.101.062001, arXiv:0802.2400.

[17] CDF Collaboration, "Measurement of the inclusive jet cross section at the fermilab tevatron p$\bar{\text{p}}$ collider using a cone-based jet algorithm", *Phys. Rev. D* **78** (2008) 052006, doi:10.1103/PhysRevD.78.052006, arXiv:0807.2204. [Erratum: doi:10.1103/PhysRevD.79.119902].

[18] CMS Collaboration, "Technical proposal for the Phase-II upgrade of the CMS detector", Technical Report CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02, 2015.

[19] CMS Collaboration, "The Phase 2 upgrade of the CMS tracker", Technical Report CERN-LHCC-2017-009, CMS-TDR-014, 2017.

[20] P. Sun, C.-P. Yuan, and F. Yuan, "Transverse momentum resummation for dijet correlation in hadronic collisions", arXiv:1506.06170.

[21] J. Collins and J.-W. Qiu, "k_T factorization is violated in production of high-transverse-momentum particles in hadron-hadron collisions", *Phys. Rev. D* **75** (2007) 114014, doi:10.1103/PhysRevD.75.114014, arXiv:0705.2141.

[22] S. Catani, M. Grazzini, and H. Sargsyan, "Transverse-momentum resummation for top-quark pair production at the LHC", *JHEP* **11** (2018) 061, doi:10.1007/JHEP11(2018)061, arXiv:1806.01601.

[23] S. Catani, M. Grazzini, and A. Torre, "Transverse-momentum resummation for heavy-quark hadroproduction", *Nucl.Phys. B* **890** (2015) 518–538, doi:10.1016/j.nuclphysb.2014.11.019, arXiv:1408.4564.

[24] CDF Collaboration, "Measurement of the b-Hadron Production Cross Section using decays to $\mu^-\mathrm{D}^0$X final states in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96-TeV", *Phys. Rev. D* **79** (2009) 092003, doi:10.1103/PhysRevD.79.092003, arXiv:0903.2403.

[25] D0 Collaboration, "The $b\bar{b}$ production cross section and angular correlations in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Lett. B* **487** (2000) 264–272, doi:10.1016/S0370-2693(00)00844-3, arXiv:hep-ex/9905024.

[26] ZEUS Collaboration, "Measurement of beauty photoproduction using decays into muons in dijet events at HERA", *JHEP* **04** (2009) 133, doi:10.1088/1126-6708/2009/04/133, arXiv:0901.2226.

[27] H1 Collaboration, "Measurement of beauty production at HERA using events with muons and jets", *Eur. Phys. J. C* **41** (2005) 453–467, doi:10.1140/epjc/s2005-02267-0, arXiv:hep-ex/0502010.

[28] ATLAS Collaboration, "Measurement of the inclusive and dijet cross-sections of b jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Eur. Phys. J. C* **71** (2011) 1846, doi:10.1140/epjc/s10052-011-1846-4, arXiv:1109.6833.

[29] ATLAS Collaboration, "Measurement of the $b\bar{b}$ dijet cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Eur. Phys. J. C* **76** (2016), no. 12, 670, doi:10.1140/epjc/s10052-016-4521-y, arXiv:1607.08430.

[30] CMS Collaboration, "Studies of inclusive four-jet production with two b-tagged jets in proton-proton collisions at 7 TeV", *Phys. Rev. D* **94** (2016), no. 11, 112005, doi:10.1103/PhysRevD.94.112005, arXiv:1609.03489.

[31] CMS Collaboration, "Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **04** (2012) 084, doi:10.1007/JHEP04(2012)084, arXiv:1202.4617.

[32] ATLAS Collaboration, "Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC", arXiv:1808.07858.

[33] CMS Collaboration, "Measurement of the differential $t\bar{t}$ cross section with high-p_t top-quark jets in the all-hadronic channel at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-018, 2016.

[34] ATLAS Collaboration, "Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector", *Phys. Rev. D* **98** (2018), no. 1, 012003, doi:10.1103/PhysRevD.98.012003, arXiv:1801.02052.

[35] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ cross section at 13 TeV in the all jets final state", CMS Physics Analysis Summary CMS-PAS-TOP-16-013, 2018.

[36] CMS Collaboration, "Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **78** (2018), no. 7, 566, doi:10.1140/epjc/s10052-018-6033-4, arXiv:1712.05471.

[37] CMS Collaboration, "Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV and constraints on parton distribution functions", *Eur. Phys. J. C* **77** (2017), no. 11, 746, doi:10.1140/epjc/s10052-017-5286-7, arXiv:1705.02628.

[38] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159.

[39] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[40] DELPHES 3 Collaboration, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **02** (2014) 057, doi:10.1007/JHEP02(2014)057, arXiv:1307.6346.

[41] S. Alioli et al., "Jet pair production in POWHEG", *JHEP* **04** (2011) 081.

[42] CMS Collaboration, "Particle–flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[43] CMS Collaboration, "Commissioning of the particle–flow event reconstruction with the first LHC collisions recorded in the CMS detector", CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.

[44] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[45] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[46] CMS Collaboration, "Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC", CMS Physics Analysis Summary CMS-PAS-FTR-18-012, 2018.

[47] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) 02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[48] CMS Collaboration, "Measurement of dijet azimuthal decorrelation in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur. Phys. J. C* **76** (2016), no. 10, 536, doi:10.1140/epjc/s10052-016-4346-8, arXiv:1602.04384.

[49] CMS Collaboration, "Azimuthal angular correlations in high transverse momentum dijet events", CMS Physics Analysis Summary CMS-PAS-SMP-17-009, 2017.

[50] 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.

[51] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* **02** (2012) 093, doi:10.1007/JHEP02(2012)093, arXiv:1108.2701.

Appendix B

Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV

Eur. Phys. J. C (2018) 78:566

In this chapter, studies published in the European Physical Journal C (Eur. Phys. J. C), Ref. [80] are presented. Measurements of the azimuthal correlation between jets in 2,3,4 inclusive jets events topologies are presented and compared to theory predictions provided from different Monte Carlo at LO and NLO accuracy. The measurements are performed in proton-proton collisions, at a center of mass energy of $\sqrt{s} = 13$ TeV using the data collected by the CMS experiment during the 2016 data taking period. I contributed to this analysis with studies related to unfolding and jet energy resolutions.

Eur. Phys. J. C (2018) 78:566
https://doi.org/10.1140/epjc/s10052-018-6033-4

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV

CMS Collaboration*

CERN, 1211 Geneva 23, Switzerland

Received: 14 December 2017 / Accepted: 28 June 2018 / Published online: 10 July 2018

Abstract Azimuthal correlations between the two jets with the largest transverse momenta p_T in inclusive 2-, 3-, and 4-jet events are presented for several regions of the leading jet p_T up to 4 TeV. For 3- and 4-jet scenarios, measurements of the minimum azimuthal angles between any two of the three or four leading p_T jets are also presented. The analysis is based on data from proton–proton collisions collected by the CMS Collaboration at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 35.9 fb^{-1}. Calculations based on leading-order matrix elements supplemented with parton showering and hadronization do not fully describe the data, so next-to-leading-order calculations matched with parton shower and hadronization models are needed to better describe the measured distributions. Furthermore, we show that azimuthal jet correlations are sensitive to details of the parton showering, hadronization, and multiparton interactions. A next-to-leading-order calculation matched with parton showers in the MC@NLO method, as implemented in HERWIG 7, gives a better overall description of the measurements than the POWHEG method.

1 Introduction

Particle jets with large transverse momenta p_T are abundantly produced in proton–proton collisions at the CERN LHC through the strong interactions of quantum chromodynamics (QCD) between the incoming partons. When the momentum transfer is large, the dynamics can be predicted using perturbative techniques (pQCD). The two final-state partons at leading order (LO) in pQCD are produced back-to-back in the transverse plane, and thus the azimuthal angular separation between the two highest-p_T jets, $\Delta\phi_{1,2} = |\phi_{\text{jet1}} - \phi_{\text{jet2}}|$, equals π. The production of additional high-p_T jets leads to a deviation of the azimuthal angle from π. The measurement of azimuthal angular correlations (or decorrelation from π) in inclusive 2-jet topologies is a useful tool to test theoretical predictions of multijet production processes. Previous measurements of azimuthal correlation in inclusive 2-jet events were reported by the D0 Collaboration in $\mathrm{p\bar{p}}$ collisions at $\sqrt{s} = 1.96$ TeV at the Fermilab Tevatron [1,2], and by the ATLAS Collaboration in pp collisions at $\sqrt{s} = 7$ TeV [3] and the CMS Collaboration in pp collisions at $\sqrt{s} = 7$ and 8 TeV [4,5] at the LHC. Multijet correlations have been measured by the ATLAS Collaboration at $\sqrt{s} = 8$ TeV [6,7].

This paper reports measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections as a function of the azimuthal angular separation between the two highest p_T (leading) jets, $\Delta\phi_{1,2}$,

$$\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\Delta\phi_{1,2}},$$

for several regions of the leading jet p_T, $p_T^{\max}$, for the rapidity region $|y| < 2.5$. The measurements cover the region $\pi/2 < \Delta\phi_{1,2} \leq \pi$; the region $\Delta\phi_{1,2} \leq \pi/2$ includes large backgrounds due to $\mathrm{t\bar{t}}$ and Z/W+jet(s) events. Experimental and theoretical uncertainties are reduced by normalizing the $\Delta\phi_{1,2}$ distribution to the total dijet cross section within each region of $p_T^{\max}$.

For 3- and 4-jet topologies, measurements of the normalized inclusive 3- and 4-jet cross sections are also presented as a function of the minimum azimuthal angular separation between any two of the three or four highest p_T jets, $\Delta\phi_{2\mathrm{j}}^{\min}$,

$$\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\Delta\phi_{2\mathrm{j}}^{\min}},$$

for several regions of $p_T^{\max}$, for $|y| < 2.5$. This observable, which is infrared safe (independent of additional soft radiation), is especially suited for studying correlations amongst the jets in multijet events: the maximum value of $\Delta\phi_{2\mathrm{j}}^{\min}$ is $2\pi/3$ for 3-jet events (the "Mercedes star" configuration), while it is $\pi/2$ in the 4-jet case (corresponding to the "cross" configuration). The cross section for small angular separations is suppressed because of the finite jet sizes for a particular jet algorithm. The observable $\Delta\phi_{2\mathrm{j}}^{\min}$ is sensitive to the

* e-mail: cms-publication-committee-chair@cern.ch

contributions of jets with lower p_T than the leading jet, i.e. the subleading jets, and one can distinguish nearby (nearly collinear) jets (at large $\Delta\phi_{2j}^{min}$) from other additional high p_T jets (small $\Delta\phi_{2j}^{min}$), yielding information additional to that of the $\Delta\phi_{1,2}$ observable. The 4-jet cross section differential in $\Delta\phi_{2j}^{min}$ has also been measured by the ATLAS Collaboration [7].

The measurements are performed using data collected during 2016 with the CMS experiment at the LHC, and the event sample corresponds to an integrated luminosity of 35.9 fb^{-1} of proton–proton collisions at $\sqrt{s} = 13$ TeV.

2 The CMS detector

The central feature of the CMS detector is a superconducting solenoid, 13 m in length and 6 m in inner diameter, providing an axial magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Charged-particle trajectories are measured by the tracker with full azimuthal coverage within pseudorapidities $|\eta| < 2.5$. The ECAL, which is equipped with a preshower detector in the endcaps, and the HCAL cover the region $|\eta| < 3.0$. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors to the region $3.0 < |\eta| < 5.2$. Finally, muons are measured up to $|\eta| < 2.4$ by gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. A detailed description of the CMS detector together with a definition of the coordinate system used and the relevant kinematic variables can be found in Ref. [8].

3 Theoretical predictions

Predictions from five different Monte Carlo (MC) event generators are compared with data. The PYTHIA 8 [9] and HERWIG++ [10] event generators are used, both based on LO $2 \rightarrow 2$ matrix element calculations. The PYTHIA 8 event generator simulates parton showers ordered in p_T and uses the Lund string model [11] for hadronization, while HERWIG++ generates parton showers through angular-ordered emissions and uses a cluster fragmentation model [12] for hadronization. The contribution of multiparton interactions (MPI) is simulated in both PYTHIA 8 and HERWIG++, but the number of generated MPI varies between PYTHIA 8 and HERWIG++ MPI simulations. The MPI parameters of both generators are tuned to measurements in proton–proton collisions at the LHC and proton–antiproton collisions at the Tevatron [13], while the hadronization parameters are determined from fits to LEP data. For PYTHIA 8 the CUETP8M1 [13] tune, which is based on the NNPDF2.3LO PDF set [14,15], is employed, while for HERWIG++ the CUETHppS1 tune [13], based on the CTEQ6L1 PDF set [16], is used.

The MADGRAPH [17,18] event generator provides LO matrix element calculations with up to four outgoing partons, i.e. $2 \rightarrow 2$, $2 \rightarrow 3$, and $2 \rightarrow 4$ diagrams. It is interfaced to PYTHIA 8 with tune CUETP8M1 for the implementation of parton showers, hadronization, and MPI. In order to match with PYTHIA 8 the k_T-MLM matching procedure [19] with a matching scale of 14 GeV is used to avoid any double counting of the parton configurations generated within the matrix element calculation and the ones simulated by the parton shower. The NNPDF2.3LO PDF set is used for the hard-process calculation.

Predictions based on next-to-leading-order (NLO) pQCD are obtained with the POWHEGBOX library [20–22] and the HERWIG 7 [23] event generator. The events simulated with POWHEG are matched to PYTHIA 8 or to HERWIG++ parton showers and MPI, while HERWIG 7 uses similar parton shower and MPI models as HERWIG++, and the MC@NLO [24,25] method is applied to combine the parton shower with the NLO calculation. The POWHEG generator is used in the NLO dijet mode [26], referred to as PH- 2J, as well as in the NLO three-jet mode [27], referred to as PH- 3J, both using the NNPDF3.0NLO PDF set [28]. The POWHEG generator, referred to as PH- 2J- LHE, is also used in the NLO dijet mode without parton showers and MPI. A minimum p_T for real parton emission of 10 GeV is required for the PH- 2J predictions, and similarly for the PH- 3J predictions a minimum p_T for the three final-state partons of 10 GeV is imposed. To simulate the contributions due to parton showers, hadronization, and MPIs, the PH- 2J is matched to PYTHIA 8 with tune CUETP8M1 and HERWIG++ with tune CUETHppS1, while the PH- 3J is matched only to PYTHIA 8 with tune CUETP8M1. The matching between the POWHEG matrix element calculations and the PYTHIA 8 underlying event (UE) simulation is performed using the shower-veto procedure, which rejects showers if their transverse momentum is greater than the minimal p_T of all final-state partons simulated in the matrix element (parameter PTHARD = 2 [26]). Predictions from the HERWIG 7 event generator are based on the MMHT2014 PDF set [29] and the default tune H7-UE-MMHT [23] for the UE simulation. A summary of the details of the MC event generators used for comparisons with the experimental data is shown in Table 1.

Uncertainties in the theoretical predictions of the parton shower simulation are illustrated using the PYTHIA 8 event generator. Choices of scale for the parton shower are expected to have the largest impact on the azimuthal distributions. The parton shower uncertainty is calculated by independently varying the renormalization scales (μ_r) for initial- and final-state radiation by a factor 2 in units of the p_T of the emitted

Table 1 Monte Carlo event generators used for comparison in this analysis. Version of the generators, PDF set, underlying event tune, and corresponding references are listed

Matrix element generator	Simulated diagrams	PDF set	Tune
PYTHIA 8.219 [9]	$2 \to 2$ (LO)	NNPDF2.3LO [14, 15]	CUETP8M1 [13]
HERWIG++ 2.7.1 [10]	$2 \to 2$ (LO)	CTEQ6L1 [16]	CUETHppS1 [13]
MADGRAPH5_AMC@NLO 2.3.3 [17, 18] + PYTHIA 8.219 [9]	$2 \to 2$, $2 \to 3$, $2 \to 4$ (LO)	NNPDF2.3LO [14, 15]	CUETP8M1 [13]
PH- 2J V2_Sep2016 [20–22] + PYTHIA 8.219 [9]	$2 \to 2$ (NLO), $2 \to 3$ (LO)	NNPDF3.0NLO [28]	CUETP8M1 [13]
PH- 2J- LHE V2_Sep2016 [20–22]	$2 \to 2$ (NLO), $2 \to 3$ (LO)	NNPDF3.0NLO [28]	
PH- 3J V2_Sep2016 [20–22] + PYTHIA 8.219 [9]	$2 \to 3$ (NLO), $2 \to 4$ (LO)	NNPDF3.0NLO [28]	CUETP8M1 [13]
PH- 2J V2_Sep2016 [20–22] + HERWIG++ 2.7.1 [10]	$2 \to 2$ (NLO), $2 \to 3$ (LO)	NNPDF3.0NLO [28]	CUETHppS1 [13]
HERWIG 7.0.4 [23]	$2 \to 2$ (NLO), $2 \to 3$ (LO)	MMHT2014 [29]	H7-UE-MMHT [23]

Table 2 The integrated luminosity for each trigger sample considered in this analysis

HLT p_T threshold (GeV)	140	200	320	400	450
p_T^{max} region (GeV)	200–300	300–400	400–500	500–600	> 600
$\mathcal{L}$ (fb^{-1})	0.024	0.11	1.77	5.2	36

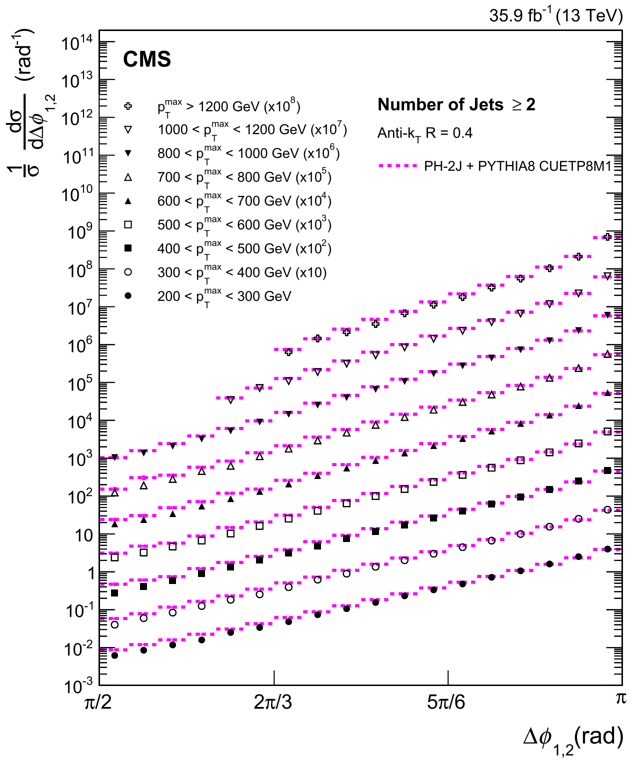


Fig. 1 Normalized inclusive 2-jet cross section differential in $\Delta\phi_{1,2}$ for nine p_T^{max} regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH- 2J + PYTHIA 8 event generator

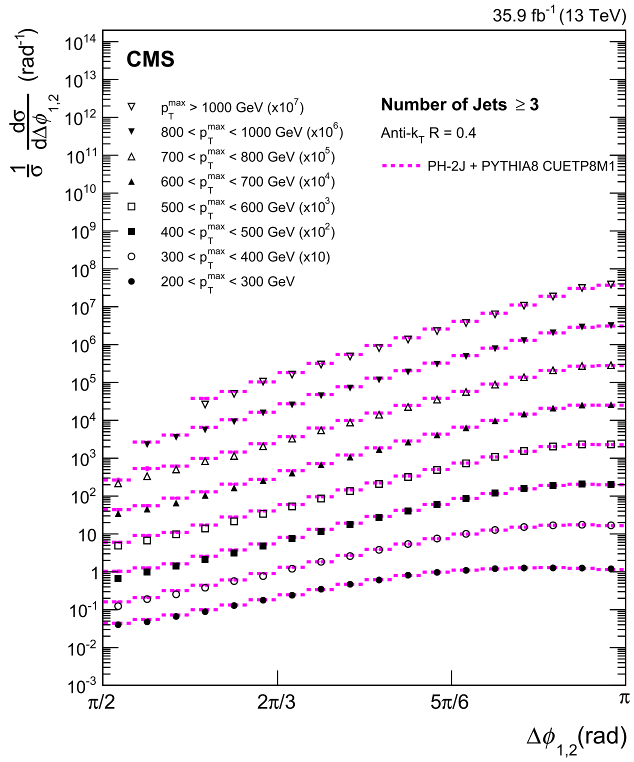


Fig. 2 Normalized inclusive 3-jet cross section differential in $\Delta\phi_{1,2}$ for eight p_T^{max} regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH- 2J + PYTHIA 8 event generator

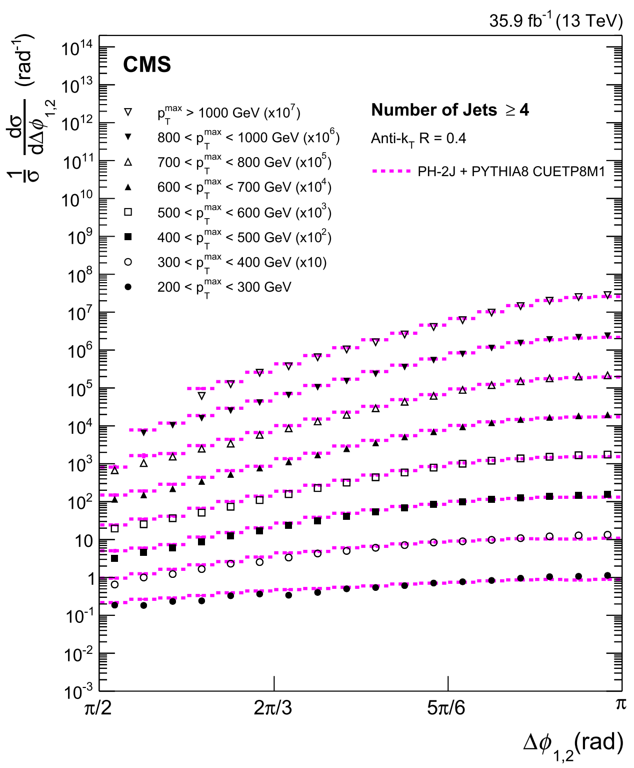


Fig. 3 Normalized inclusive 4-jet cross section differential in $\Delta\phi_{1,2}$ for eight p_T^{max} regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH- 2J + PYTHIA 8 event generator

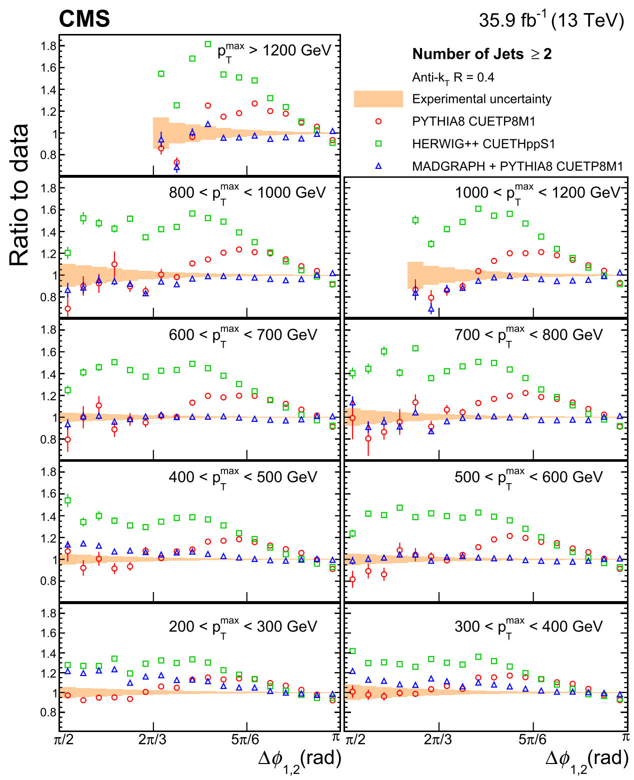


Fig. 4 Ratios of PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to the normalized inclusive 2-jet cross section differential in $\Delta\phi_{1,2}$, for all p_T^{max} regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

partons of the hard scattering. The maximum deviation found is considered a theoretical uncertainty in the event generator predictions.

4 Jet reconstruction and event selection

The measurements are based on data samples collected with single-jet high-level triggers (HLT) [30,31]. Five such triggers are considered that require at least one jet in an event with $p_T > 140$, 200, 320, 400, or 450 GeV in the full rapidity coverage of the CMS detector. All triggers are prescaled except the one with the highest threshold. Table 2 shows the integrated luminosity $\mathcal{L}$ for the five trigger samples. The relative efficiency of each trigger is estimated using triggers with lower p_T thresholds. Using these five jet energy thresholds, a 100% trigger efficiency is achieved in the region of $p_T^{max} > 200$ GeV.

Particles are reconstructed and identified using a particle-flow (PF) algorithm [32], which uses an optimized combination of information from the various elements of the CMS detector. Jets are reconstructed by clustering the Lorentz vectors of the PF candidates with the infrared- and collinear-safe anti-k_T clustering algorithm [33] with a distance parameter $R = 0.4$. The clustering is performed with the FASTJET package [34]. The technique of charged-hadron subtraction [35] is used to remove tracks identified as originating from additional pp interactions within the same or neighbouring bunch crossings (pileup). The average number of pileup interactions observed in the data is about 27.

The reconstructed jets require energy corrections to account for residual nonuniformities and nonlinearities in the detector response. These jet energy scale (JES) corrections [35] are derived using simulated events that are generated with PYTHIA 8.219 [9] using tune CUETP8M1 [13] and processed through the CMS detector simulation based on GEANT4 [36]; they are confirmed with in situ measurements with dijet, multijet, photon+jet, and leptonic Z+jet events. An offset correction is required to account for the extra energy clustered into jets due to pileup. The JES corrections, which depend on the η and p_T of the jet, are applied as multiplicative factors to the jet four-momentum vectors. The typical overall correction is about 10% for central jets having $p_T = 100$ GeV and decreases with increasing p_T.

Resolution studies on the measurements of $\Delta\phi_{1,2}$ and $\Delta\phi_{2j}^{min}$ are performed using PYTHIA 8.219 with tune

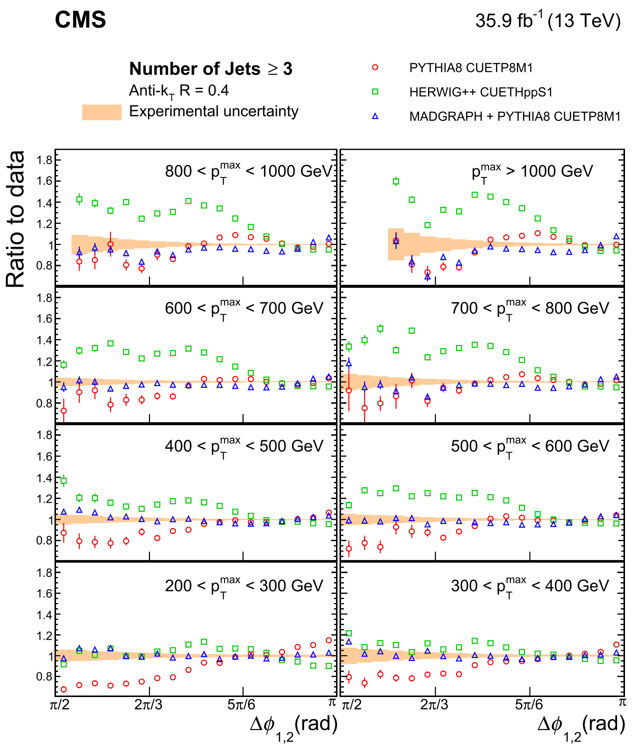


Fig. 5 Ratios of PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to the normalized inclusive 3-jet cross section differential in $\Delta\phi_{1,2}$, for all $p_{\mathrm{T}}^{\max}$ regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

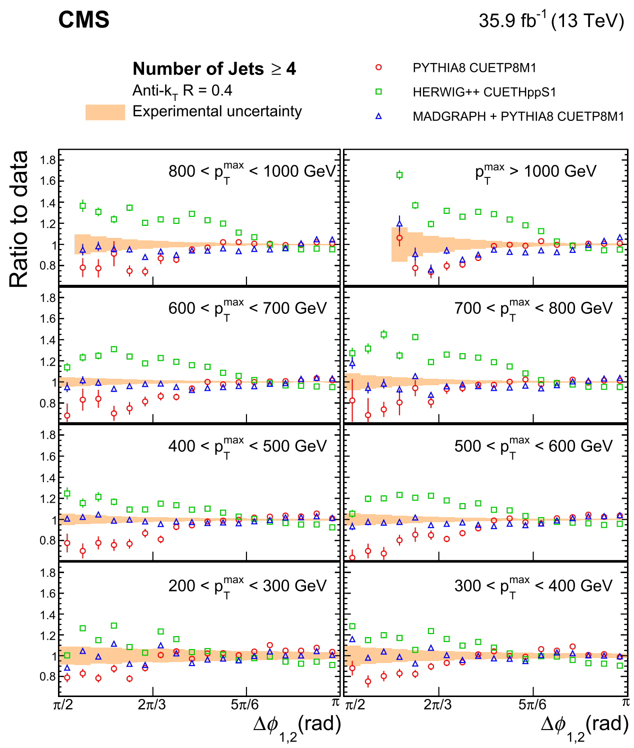


Fig. 6 Ratios of PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to the normalized inclusive 4-jet cross section differential in $\Delta\phi_{1,2}$, for all $p_{\mathrm{T}}^{\max}$ regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

CUETP8M1 processed through the CMS detector simulation. The azimuthal angular separation is determined with an accuracy from 1° to 0.5° (0.017 to 0.0087 in radians) for $p_{\mathrm{T}}^{\max} = 200\,\mathrm{GeV}$ to 1 TeV, respectively.

Events are required to have at least one primary vertex candidate [37] reconstructed offline from at least five charged-particle tracks and lies along the beam line within 24 cm of the nominal interaction point. The reconstructed vertex with the largest value of summed physics-object p_{T}^2 is taken to be the primary pp interaction vertex. The physics objects are the objects determined by a jet finding algorithm [33,34] applied to all charged tracks associated with the vertex plus the corresponding associated missing transverse momentum. Additional selection criteria are applied to each event to remove spurious jet-like signatures originating from isolated noise patterns in certain HCAL regions. Stringent criteria [38] are applied to suppress these nonphysical signatures; each jet should contain at least two particles, one of which is a charged hadron, and the jet energy fraction carried by neutral hadrons and photons should be less than 90%. These criteria have a jet selection efficiency greater than 99% for genuine jets.

For the measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections as a function of $\Delta\phi_{1,2}$ or $\Delta\phi_{2\mathrm{j}}^{\min}$ all jets in the event with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ and a rapidity $|y| < 5$ are considered and ordered in p_{T}. Events are selected where the two highest-p_{T} jets have $|y| < 2.5$, (i.e. events are not counted where one of the leading jets has $|y| > 2.5$). Also, events are only selected in which the highest-p_{T} jet has $|y| < 2.5$ and exceeds 200 GeV. The inclusive 2-jet event sample includes events where the two leading jets lie within the tracker coverage of $|y| < 2.5$. Similarly the 3-jet (4-jet) event sample includes those events where the three (four) leading jets lie within $|y| < 2.5$, respectively. In this paper results are presented in bins of $p_{\mathrm{T}}^{\max}$, corresponding to the p_{T} of the leading jet, which is always within $|y| < 2.5$.

5 Measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections in $\Delta\phi_{1,2}$ and $\Delta\phi_{2\mathrm{j}}^{\min}$

The normalized inclusive 2-, 3-, and 4-jet cross sections differential in $\Delta\phi_{1,2}$ and $\Delta\phi_{2\mathrm{j}}^{\min}$ are corrected for the finite detector resolution to better approximate the final-state particles, a procedure called "unfolding". In this way, a direct comparison of this measurement to results from other experiments and to QCD predictions is possible. Particles are considered stable if their mean decay length is $c\tau > 1\,\mathrm{cm}$.

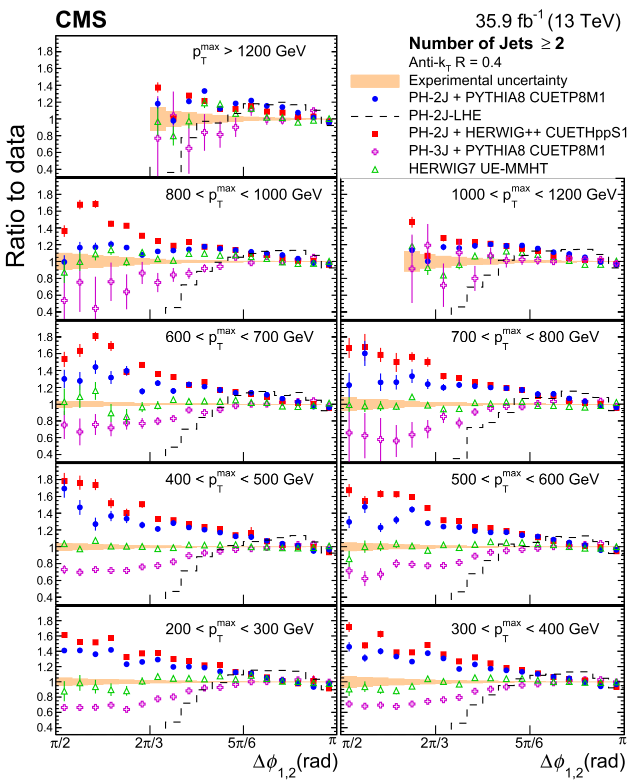


Fig. 7 Ratios of PH-2J + PYTHIA 8, PH-2J-LHE, PH-2J + HERWIG++, PH-3J + PYTHIA 8, and HERWIG 7 predictions to the normalized inclusive 2-jet cross section differential in $\Delta\phi_{1,2}$, for all $p_{\mathrm{T}}^{\max}$ regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

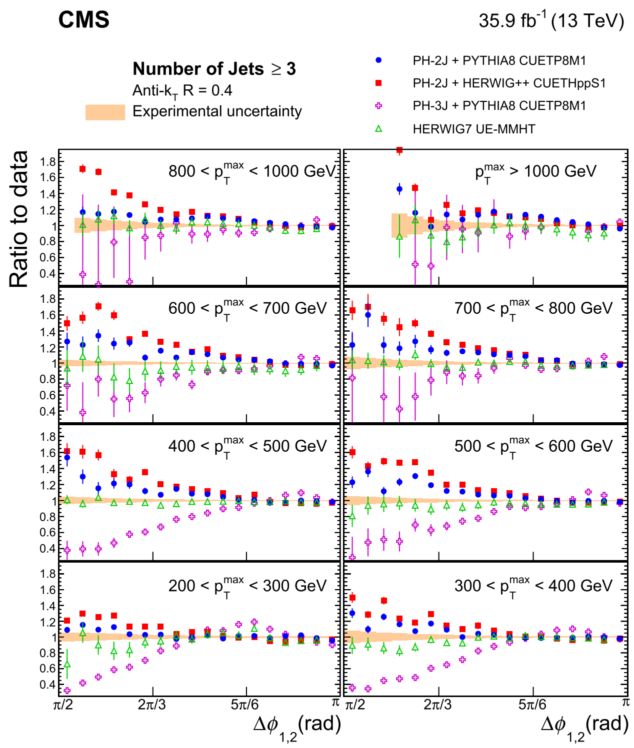


Fig. 8 Ratios of PH-2J + PYTHIA 8, PH-2J + HERWIG++, PH-3J + PYTHIA 8, and HERWIG 7 predictions to the normalized inclusive 3-jet cross section differential in $\Delta\phi_{1,2}$, for all $p_{\mathrm{T}}^{\max}$ regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

The bin width used in the measurements of $\Delta\phi_{1,2}$ and $\Delta\phi_{2j}^{\min}$ is set to $\pi/36 = 0.087$ rads (5°), which is five to ten times larger than the azimuthal angular separation resolution. The corrections due to the unfolding are approximately a few per cent.

The unfolding procedure is based on the matrix inversion algorithm implemented in the software package RooUnfold [39] using a 2-dimensional response matrix that correlates the modeled distribution with the reconstructed one. The response matrix is created by the convolution of the $\Delta\phi$ resolution with the generator-level inclusive 2-, 3-, and 4- cross section distributions from PYTHIA 8 with tune CUETP8M1. The unfolded distributions differ from the distributions at detector level by 1–4%. As a cross-check, the above procedure was repeated by creating the response matrix with event samples obtained with the full Geant4 detector simulation, and no significant difference was observed.

We consider three main sources of systematic uncertainties that arise from the estimation of the JES calibration, the jet energy resolution (JER), and the unfolding correction. The relative JES uncertainty is estimated to be 1–2% for PF jets using charged-hadron subtraction [35]. The resulting uncertainties in the normalized 2-, 3-, and 4-jet cross sections differential in $\Delta\phi_{1,2}$ range from 3% at $\pi/2$ to 0.1% at π. For the normalized 3- and 4-jet cross sections differential in $\Delta\phi_{2j}^{\min}$ the resulting uncertainties range from 0.1 to 1%, and 0.1–2%, respectively.

The JER [35] is responsible for migration of events among the $p_{\mathrm{T}}^{\max}$ regions, and its parametrization is determined from a full detector simulation using events generated by PYTHIA 8 with tune CUETP8M1. The effect of the JER uncertainty is estimated by varying its parameters within their uncertainties [35] and comparing the normalized inclusive 2-, 3-, and 4-jet cross sections before and after the changes. The JER-induced uncertainty ranges from 1% at $\pi/2$ to 0.1% at π for the normalized 2-, 3-, and 4-jet cross sections differential in $\Delta\phi_{1,2}$ and is less than 0.5% for the normalized 3- and 4-jet cross sections differential in $\Delta\phi_{2j}^{\min}$.

The above systematic uncertainties in the JES calibration and the JER cover the effects from migrations due to the p_{T} thresholds, i.e. migrations between the 2-, 3-, and 4-jet samples and migrations between the various $p_{\mathrm{T}}^{\max}$ regions of the measurements.

The unfolding procedure is affected by uncertainties in the parametrization of the $\Delta\phi$ resolution. Alternative response matrices, generated by varying the $\Delta\phi$ resolution by ±10%, are used to unfold the measured spectra. This variation is

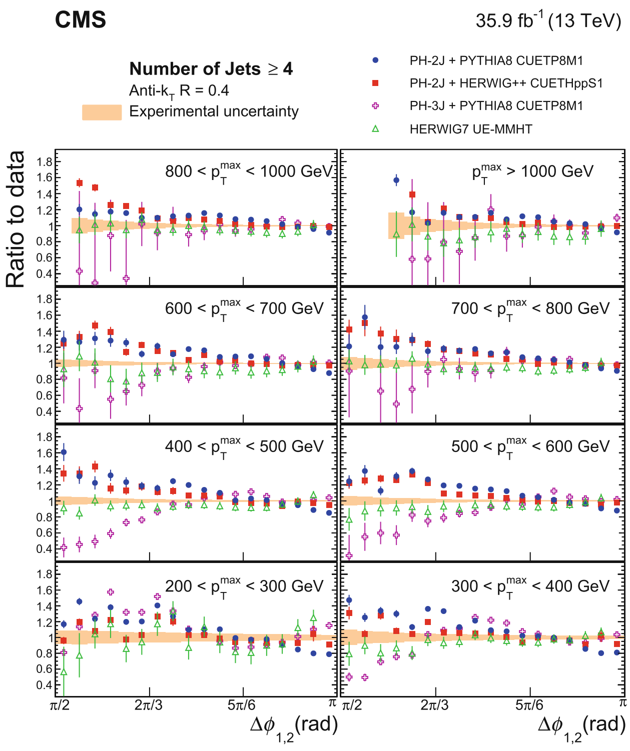


Fig. 9 Ratios of PH- 2J + PYTHIA 8, PH- 2J + HERWIG++, PH- 3J + PYTHIA 8, and HERWIG 7 predictions to the normalized inclusive 4-jet cross section differential in $\Delta\phi_{1,2}$, for all $p_{\mathrm{T}}^{\max}$ regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

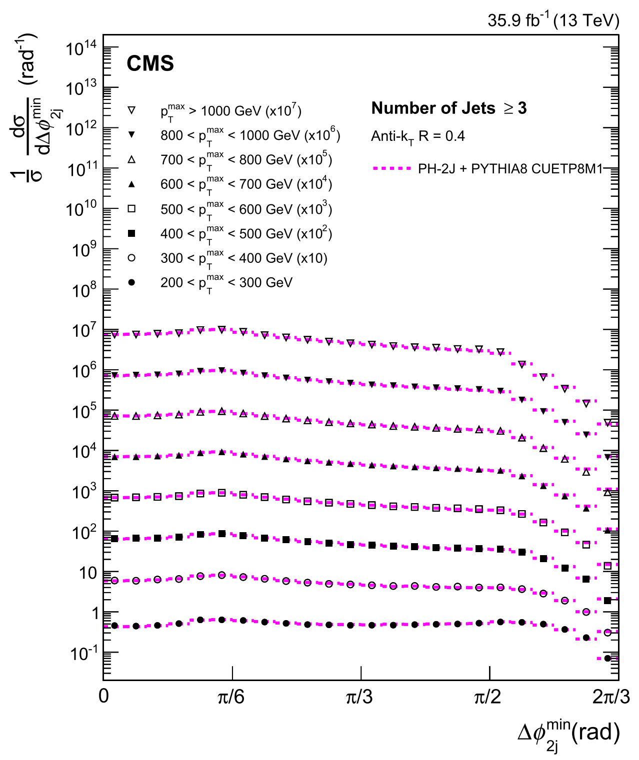


Fig. 10 Normalized inclusive 3-jet cross section differential in $\Delta\phi_{2\mathrm{j}}^{\min}$ for eight $p_{\mathrm{T}}^{\max}$ regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH- 2J + PYTHIA 8 event generator

motivated by studies on the $\Delta\phi$ resolution for simulated dijet events [32]. The uncertainty in the unfolding correction factors is estimated to be about 0.2%. An additional systematic uncertainty is obtained by examining the dependence of the response matrix on the choice of the MC generator. Alternative response matrices are constructed using the HERWIG++ event generator [10] with tune EE5C [40]; the effect is <0.1%. A total systematic unfolding uncertainty of 0.2% is considered, which accounts for all these various uncertainty sources.

6 Comparison with theoretical predictions

6.1 The $\Delta\phi_{1,2}$ measurements

The unfolded, normalized, inclusive 2-, 3-, and 4-jet cross sections differential in $\Delta\phi_{1,2}$ are shown in Figs. 1, 2, 3 for the various $p_{\mathrm{T}}^{\max}$ regions considered in this analysis. In the 2-jet case the $\Delta\phi_{1,2}$ distributions are strongly peaked at π and become steeper with increasing $p_{\mathrm{T}}^{\max}$. In the 3-jet case, the $\Delta\phi_{1,2}$ distributions become flatter at π, since by definition dijet events do not contribute, and in the 4-jet case they become even flatter. The data points are overlaid with the predictions from the PH- 2J + PYTHIA 8 event generator.

The ratios of the PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 event generator predictions to the normalized inclusive 2-, 3-, and 4-jet cross section differential in $\Delta\phi_{1,2}$ are shown in Figs. 4, 5, and 6, respectively, for all $p_{\mathrm{T}}^{\max}$ regions. The solid band around unity represents the total experimental uncertainty and the error bars on the points represent the statistical uncertainties in the simulated data. Among the LO dijet event generators, HERWIG++ exhibits the largest deviations from the experimental measurements, whereas PYTHIA 8 behaves much better than HERWIG++, although with deviations of up to 30–40%, in particular around $\Delta\phi_{1,2} = 5\pi/6$ in the 2-jet case and around $\Delta\phi_{1,2} < 2\pi/3$ in the 3- and 4-jet case. Predictions from HERWIG++ tend to overestimate the measurements as a function of $\Delta\phi_{1,2}$ in the 2-, 3-, and 4-jet cases, especially at $\Delta\phi_{1,2} < 5\pi/6$ for $p_{\mathrm{T}}^{\max} > 400$ GeV. However, it is remarkable that predictions based on the $2 \to 2$ matrix element calculations supplemented with parton showers, MPI, and hadronization describe the $\Delta\phi_{1,2}$ distributions rather well, even in regions that are sensitive to hard jets not included in the matrix element calculations. The MADGRAPH + PYTHIA 8 calculation using up to 4 partons in the

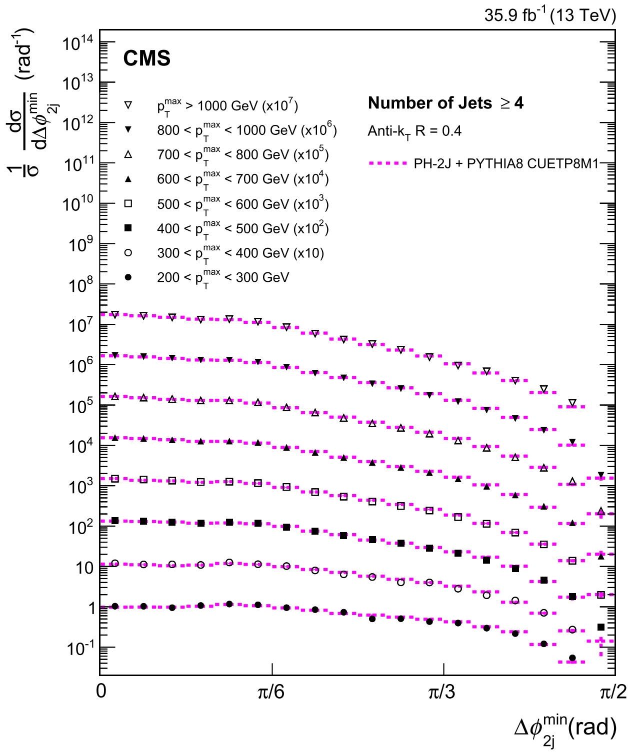


Fig. 11 Normalized inclusive 4-jet cross section differential in $\Delta\phi_{2j}^{min}$ for eight p_T^{max} regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH- 2J + PYTHIA 8 event generator

matrix element calculations provides the best description of the measurements.

Figures 7, 8 and 9 show the ratios of the PH- 2J matched to PYTHIA 8 and HERWIG++, PH- 3J + PYTHIA 8, and HERWIG 7 event generators predictions to the normalized inclusive 2-, 3-, and 4-jet cross section differential in $\Delta\phi_{1,2}$, for all p_T^{max} regions. The solid band around unity represents the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data. The predictions of PH- 2J and PH- 3J exhibit deviations from the measurement, increasing towards small $\Delta\phi_{1,2}$. While PH- 2J is above the data, PH- 3J predicts too few events at small $\Delta\phi_{1,2}$. These deviations were investigated in a dedicated study with parton showers and MPI switched off. Because of the kinematic restriction of a 3-parton state, PH- 2J without parton showers cannot fill the region $\Delta\phi_{1,2} < 2\pi/3$, shown as PH- 2J- LHE with the dashed line in Fig. 7, whereas for PH- 3J the parton showers have little impact. Thus, the events at low $\Delta\phi_{1,2}$ observed for PH- 2J originate from leading-log parton showers, and there are too many of these. In contrast, the PH- 3J prediction, which provides $2 \rightarrow 3$ jet calculations at NLO QCD, is below the measurement. The NLO PH- 2J calculation and the LO POWHEG three-jet calculation are equivalent when initial- and final-state radiation are not allowed to occur.

The predictions from PH- 2J matched to PYTHIA 8 describe the normalized cross sections better than those where PH- 2J is matched to HERWIG++. Since the hard process calculation is the same, the difference between the two predictions might be due to the treatment of parton showers in PYTHIA 8 and HERWIG++ and to the matching to the matrix element calculation. The PYTHIA 8 and HERWIG++ parton shower calculations use different α_S values for initial- and final-state emissions, in addition to a different upper scale for the parton shower simulation, which is higher in PYTHIA 8 than in HERWIG++. The dijet NLO calculation of HERWIG 7 provides the best description of the measurements, indicating that the MC@NLO method of combining parton showers with the NLO parton level calculations has advantages compared to the POWHEG method in this context.

For $\Delta\phi_{1,2}$ generator-level predictions in the 2-jet case, parton shower uncertainties have a very small impact ($< 5\%$) at values close to π and go up to 40–60% for increasing p_T^{max} at $\Delta\phi_{1,2} \sim \pi/2$. For the 3- and 4-jet scenarios, parton shower uncertainties are less relevant, not exceeding ∼20% for $\Delta\phi_{1,2}$.

6.2 The $\Delta\phi_{2j}^{min}$ measurements

The unfolded, normalized, inclusive 3- and 4-jet cross sections differential in $\Delta\phi_{2j}^{min}$ are shown in Figs. 10 and 11, respectively, for eight p_T^{max} regions. The measured distributions decrease towards the kinematic limit of $\Delta\phi_{2j}^{min} \rightarrow 2\pi/3(\pi/2)$ for the 3-jet and 4-jet case, respectively. The data points are overlaid with the predictions from the PH- 2J + PYTHIA 8 event generator. The size of the data symbol includes both statistical and systematic uncertainties.

Figures 12 and 13 show, respectively, the ratios of the PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 event generators predictions to the normalized inclusive 3- and 4-jet cross sections differential in $\Delta\phi_{2j}^{min}$, for all p_T^{max} regions. The PYTHIA 8 event generator shows larger deviations from the measured $\Delta\phi_{2j}^{min}$ distributions in comparison to HERWIG++, which provides a reasonable description of the measurement. The MADGRAPH generator matched to PYTHIA 8 provides a reasonable description of the measurements in the 3-jet case, but shows deviations in the 4-jet case.

The predictions from MADGRAPH + PYTHIA 8 and PYTHIA 8 are very similar for the normalized cross sections as a function of $\Delta\phi_{2j}^{min}$ in the four-jet case. It has been checked that predictions obtained with the MADGRAPH matrix element with up to 4 partons included in the calculation without contribution of the parton shower are able to reproduce the data very well. Parton shower effects increase the number of events with low values of $\Delta\phi_{2j}^{min}$.

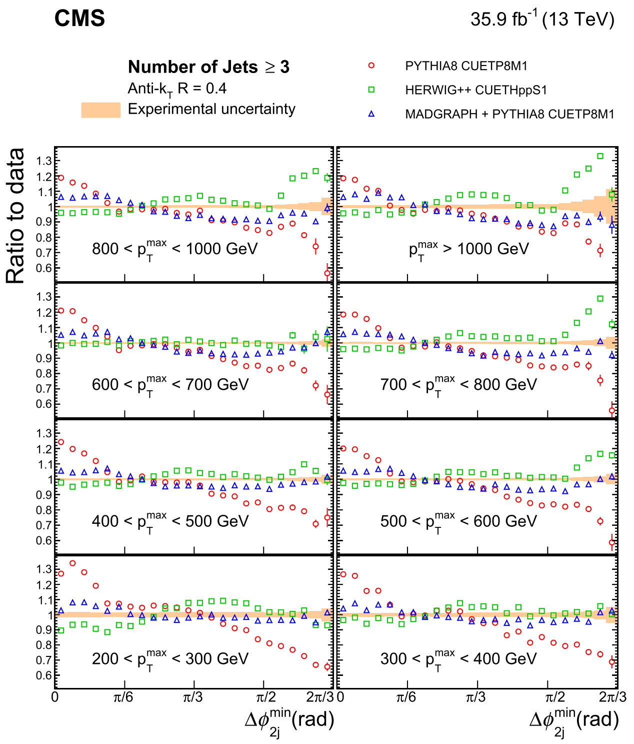


Fig. 12 Ratios of PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to the normalized inclusive 3-jet cross section differential in $\Delta\phi_{2j}^{min}$, for all p_T^{max} regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

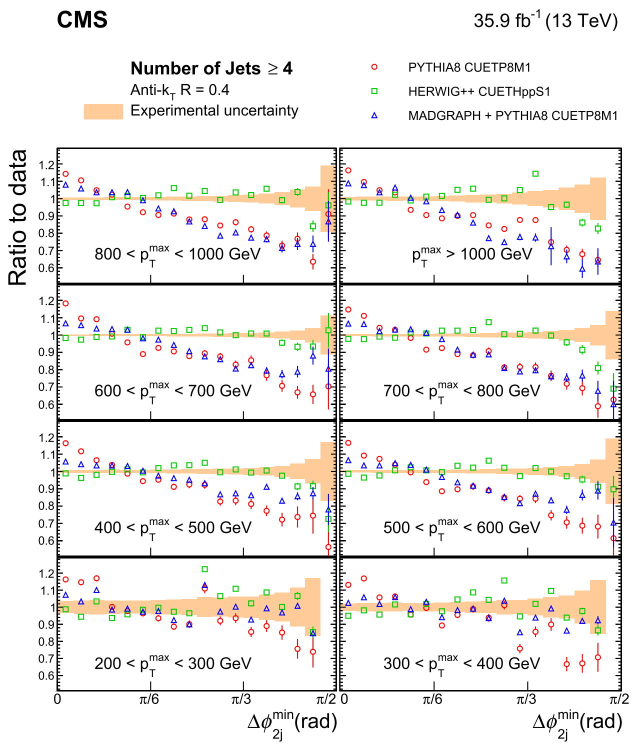


Fig. 13 Ratios of PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to the normalized inclusive 4-jet cross section differential in $\Delta\phi_{2j}^{min}$, for all p_T^{max} regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data

Figures 14 and 15 illustrate the ratios of predictions from PH- 2J matched to PYTHIA 8 and HERWIG++, PH- 3J + PYTHIA 8, and HERWIG 7 to the normalized inclusive 3- and 4-jet cross sections differential in $\Delta\phi_{2j}^{min}$, for all p_T^{max} regions. Due to an unphysical behavior of the HERWIG 7 prediction (which has been confirmed by the HERWIG 7 authors), the first $\Delta\phi_{2j}^{min}$ and last $\Delta\phi_{1,2}$ bins are not shown in Figs. 8, 9, 14 and 15. An additional uncertainty is introduced to the prediction of HERWIG 7, that is evaluated as the difference between this prediction and the prediction when the first bin is replaced with the result from HERWIG++. The additional uncertainty ranges from 2 to 10%. Among the three NLO dijet calculations PH- 2J matched to PYTHIA 8 or to HERWIG++ provides the best description of the measurements.

For the two lowest p_T^{max} regions in Figs. 13 and 15, which correspond to the 4-jet case, the measurements become statistically limited because the data used for these two regions were collected with highly prescaled triggers with p_T thresholds of 140 and 200 GeV (c.f. Table 2).

The PH- 3J predictions suffer from low statistical accuracy, especially in the highest interval of p_T^{max}, because the same p_T threshold is applied to all 3 jets resulting in low efficiency at large p_T. Nevertheless, the performance of the PH- 3J simulation on multijet observables can already be inferred by the presented predictions, especially in the low p_T region.

The effect of parton shower uncertainties in the event generator predictions of $\Delta\phi_{2j}^{min}$ is estimated to be less than 10% over the entire phase space.

7 Summary

Measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections differential in the azimuthal angular separation $\Delta\phi_{1,2}$ and of the normalized inclusive 3- and 4-jet cross sections differential in the minimum azimuthal angular separation between any two jets $\Delta\phi_{2j}^{min}$ are presented for several regions of the leading-jet transverse momentum p_T^{max}. The measurements are performed using data collected during 2016 with the CMS detector at the CERN LHC corresponding to an integrated luminosity of 35.9 fb^{-1} of proton–proton collisions at $\sqrt{s} = 13$ TeV.

The measured distributions in $\Delta\phi_{1,2}$ and $\Delta\phi_{2j}^{min}$ are compared with predictions from PYTHIA 8, HERWIG++, MADGRAPH + PYTHIA 8, PH- 2J matched to PYTHIA 8 and HERWIG++, PH- 3J + PYTHIA 8, and HERWIG 7 event generators.

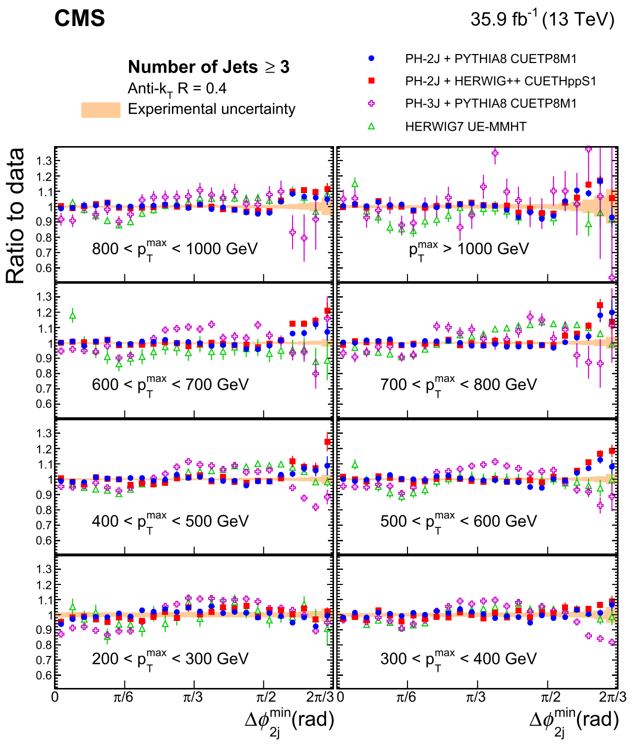


Fig. 14 Ratios of PH- 2J + PYTHIA 8, PH- 2J + HERWIG++, PH- 3J + PYTHIA 8, and HERWIG 7 predictions to the normalized inclusive 3-jet cross section differential in $\Delta\phi_{2j}^{min}$, for all p_T^{max} regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties of the simulated data

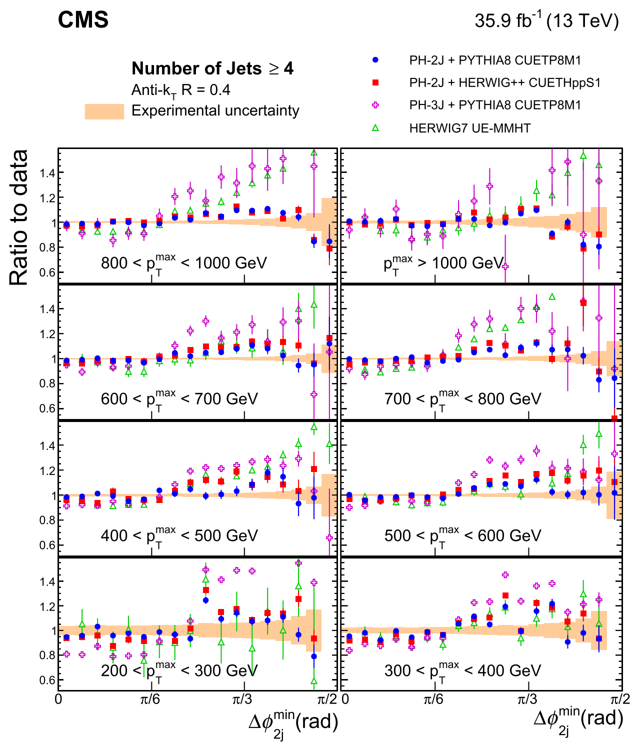


Fig. 15 Ratios of PH- 2J + PYTHIA 8, PH- 2J + HERWIG++, PH- 3J + PYTHIA 8, and HERWIG 7 predictions to the normalized inclusive 4-jet cross section differential in $\Delta\phi_{2j}^{min}$, for all p_T^{max} regions. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties of the simulated data

The leading order (LO) PYTHIA 8 dijet event generator exhibits small deviations from the $\Delta\phi_{1,2}$ measurements but shows significant deviations at low-p_T in the $\Delta\phi_{2j}^{min}$ distributions. The HERWIG++ event generator exhibits the largest deviations of any of the generators for the $\Delta\phi_{1,2}$ measurements, but provides a reasonable description of the $\Delta\phi_{2j}^{min}$ distributions. The tree-level multijet event generator MADGRAPH in combination with PYTHIA 8 for showering, hadronization, and multiparton interactions provides a good overall description of the measurements, except for the $\Delta\phi_{2j}^{min}$ distributions in the 4-jet case, where the generator deviates from the measurement mainly at high p_T^{max}.

The dijet next-to-leading order (NLO) PH- 2J event generator deviates from the $\Delta\phi_{1,2}$ measurements, but provides a good description of the $\Delta\phi_{2j}^{min}$ observable. The predictions from the three-jet NLO PH- 3J event generator exhibit large deviations from the measurements and describe the considered multijet observables in a less accurate way than the predictions from PH- 2J. Parton shower contributions are responsible for the different behaviour of the PH- 2J and PH- 3J predictions. Finally, predictions from the dijet NLO HERWIG 7 event generator matched to parton shower contributions with the MC@NLO method provide a very good description of the $\Delta\phi_{1,2}$ measurements, showing improvement in comparison to HERWIG++.

All these observations emphasize the need to improve predictions for multijet production. Similar observations, for the inclusive 2-jet cross sections differential in $\Delta\phi_{1,2}$, were reported previously by CMS [5] at a different centre-of-mass energy of 8 TeV. The extension of $\Delta\phi_{1,2}$ correlations, and the measurement of the $\Delta\phi_{2j}^{min}$ distributions in inclusive 3- and 4-jet topologies are novel measurements of the present analysis.

Acknowledgements We thank Simon Plätzer and Simone Alioli for discussion and great help on setting up, respectively, the HERWIG 7 and the PH- 3J simulation. We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centres and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWFW and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); SENESCYT (Ecuador); MoER, ERC IUT, and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIP and NRF (Republic of Korea); LAS (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MBIE (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS, RFBR and RAEP (Russia); MESTD (Serbia); SEIDI, CPAN, PCTI and FEDER (Spain); Swiss Funding Agencies (Switzerland); MST (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU and SFFR (Ukraine); STFC (UK); DOE and NSF (USA). Individuals have received support from the Marie-Curie programme and the European Research Council and Horizon 2020 Grant, contract No. 675440 (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Council of Science and Industrial Research, India; the HOMING PLUS programme of the Foundation for Polish Science, cofinanced from European Union, Regional Development Fund, the Mobility Plus programme of the Ministry of Science and Higher Education, the National Science Center (Poland), contracts Harmonia 2014/14/M/ST2/00428, Opus 2014/13/B/ST2/02543, 2014/15/B/ST2/03998, and 2015/19/B/ST2/02861, Sonata-bis 2012/07/E/ST2/01406; the National Priorities Research Program by Qatar National Research Fund; the Programa Severo Ochoa del Principado de Asturias; the Thalis and Aristeia programmes cofinanced by EU-ESF and the Greek NSRF; the Rachadapisek Sompot Fund for Postdoctoral Fellowship, Chulalongkorn University and the Chulalongkorn Academic into Its 2nd Century Project Advancement Project (Thailand); the Welch Foundation, contract C-1845; and the Weston Havens Foundation (USA).

Funded by SCOAP[3].

References

1. D0 Collaboration, Measurement of dijet azimuthal decorrelations at central rapidities in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Phys. Rev. Lett. **94**, 221801 (2005). https://doi.org/10.1103/PhysRevLett.94.221801. arXiv:hep-ex/0409040
2. D0 Collaboration, Measurement of the combined rapidity and p_T dependence of dijet azimuthal decorrelations in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Phys. Lett. B **721**, 212 (2013). https://doi.org/10.1016/j.physletb.2013.03.029. arXiv:1212.1842
3. ATLAS Collaboration, Measurement of dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Rev. Lett. **106**, 172002 (2011). https://doi.org/10.1103/PhysRevLett.106.172002. arXiv:1102.2696
4. CMS Collaboration, Dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Rev. Lett. **106**, 122003 (2011). https://doi.org/10.1103/PhysRevLett.106.122003. arXiv:1101.5029
5. CMS Collaboration, Measurement of dijet azimuthal decorrelation in pp collisions at $\sqrt{s} = 8$ TeV. Eur. Phys. J. C **76**, 536 (2016). https://doi.org/10.1140/epjc/s10052-016-4346-8. arXiv:1602.04384
6. ATLAS Collaboration, Measurement of three-jet production cross-sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector. Eur. Phys. J. C **75**, 228 (2015). https://doi.org/10.1140/epjc/s10052-015-3363-3. arXiv:1411.1855
7. ATLAS Collaboration, Measurement of four-jet differential cross sections in $\sqrt{s} = 8$ TeV proton–proton collisions using the ATLAS detector. JHEP **12**, 105 (2015). https://doi.org/10.1007/JHEP12(2015)105. arXiv:1509.07335
8. CMS Collaboration, The CMS experiment at the CERN LHC. JINST **3**, S08004 (2008). https://doi.org/10.1088/1748-0221/3/08/S08004
9. T. Sjöstrand et al., An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159 (2015). https://doi.org/10.1016/j.cpc.2015.01.024. arXiv:1410.3012
10. M. Bähr et al., Herwig++ physics and manual. Eur. Phys. J. C **58**, 639 (2008). https://doi.org/10.1140/epjc/s10052-008-0798-9. arXiv:0803.0883
11. B. Andersson, The lund model. Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **7**, 1 (1997). https://doi.org/10.1016/0375-9474(87)90510-0
12. B.R. Webber, A QCD model for jet fragmentation including soft gluon interference. Nucl. Phys. B **238**, 492 (1984). https://doi.org/10.1016/0550-3213(84)90333-X
13. CMS Collaboration, Event generator tunes obtained from underlying event and multiparton scattering measurements. Eur. Phys. J. C **76** 155, (2016). https://doi.org/10.1140/epjc/s10052-016-3988-x. arXiv:1512.00815
14. NNPDF Collaboration, Parton distributions with QED corrections. Nucl. Phys. B **877**, 290 (2013). https://doi.org/10.1016/j.nuclphysb.2013.10.010. arXiv:1308.0598
15. NNPDF Collaboration, Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. Nucl. Phys. B **855**, 153 (2012). https://doi.org/10.1016/j.nuclphysb.2011.09.024. arXiv:1107.2652
16. J. Pumplin et al., New generation of parton distributions with uncertainties from global QCD analysis. JHEP **07**, 012 (2002). https://doi.org/10.1088/1126-6708/2002/07/012. arXiv:hep-ph/0201195
17. 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**, 079 (2014). https://doi.org/10.1007/JHEP07(2014)079. arXiv:1405.0301
18. J. Alwall et al., Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. Eur. Phys. J. C **53**, 473 (2008). https://doi.org/10.1140/epjc/s10052-007-0490-5. arXiv:0706.2569
19. S. Mrenna, P. Richardson, Matching matrix elements and parton showers with HERWIG and PYTHIA. JHEP **05**, 040 (2004). https://doi.org/10.1088/1126-6708/2004/05/040. arXiv:hep-ph/0312274
20. S. Frixione, P. Nason, C. Oleari, Matching NLO QCD computations with parton shower simulations: the POWHEG method. JHEP **11**, 070 (2007). https://doi.org/10.1088/1126-6708/2007/11/070. arXiv:0709.2092
21. S. Alioli, P. Nason, C. Oleari, E. Re, A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. JHEP **06**, 043 (2010). https://doi.org/10.1007/JHEP06(2010)043. arXiv:1002.2581
22. P. Nason, A new method for combining NLO QCD with shower Monte Carlo algorithms. JHEP **11**, 040 (2004). https://doi.org/10.1088/1126-6708/2004/11/040. arXiv:hep-ph/0409146
23. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. Eur. Phys. J. C **76**, 196 (2016). https://doi.org/10.1140/epjc/s10052-016-4018-8. arXiv:1512.01178
24. S. Frixione, B.R. Webber, Matching NLO QCD computations and parton shower simulations. JHEP **06**, 029 (2002). https://doi.org/10.1088/1126-6708/2002/06/029. arXiv:hep-ph/0204244

25. R. Frederix, S. Frixione, Merging meets matching in MC@NLO. JHEP **12**, 061 (2012). https://doi.org/10.1007/JHEP12(2012)061. arXiv:1209.6215
26. S. Alioli et al., Jet pair production in POWHEG. JHEP **11**, 081 (2011). https://doi.org/10.1007/JHEP04(2011)081. arXiv:1012.3380
27. A. Kardos, P. Nason, C. Oleari, Three-jet production in POWHEG. JHEP **04**, 043 (2014). https://doi.org/10.1007/JHEP04(2014)043. arXiv:1402.4001
28. NNPDF Collaboration, Parton distributions for the LHC Run II. JHEP **04**, 040 (2015). https://doi.org/10.1007/JHEP04(2015)040. arXiv:1410.8849
29. L.A. Harland-Lang, A.D. Martin, P. Motylinski, R.S. Thorne, Parton distributions in the LHC era: MMHT 2014 PDFs. Eur. Phys. J. C **75**, 204 (2015). https://doi.org/10.1140/epjc/s10052-015-3397-6. arXiv:1412.3989
30. CMS Collaboration, The CMS high level trigger. Eur. Phys. J. C **46**, 605 (2006). https://doi.org/10.1140/epjc/s2006-02495-8. arXiv:hep-ex/0512077
31. CMS Collaboration, The CMS trigger system. JINST **12**, P01020 (2017). https://doi.org/10.1088/1748-0221/12/01/P01020. arXiv:1609.02366
32. CMS Collaboration, Particle-flow reconstruction and global event description with the CMS detector. JINST **12**, P10003 (2017). https://doi.org/10.1088/1748-0221/12/10/P10003. arXiv:1706.04965
33. M. Cacciari, G.P. Salam, G. Soyez, The anti-k_t jet clustering algorithm. JHEP **04**, 063 (2008). https://doi.org/10.1088/1126-6708/2008/04/063. arXiv:0802.1189
34. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). https://doi.org/10.1140/epjc/s10052-012-1896-2. arXiv:1111.6097
35. CMS Collaboration, Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. JINST **12**, 02014 (2017). https://doi.org/10.1088/1748-0221/12/02/P02014. arXiv:1607.03663
36. GEANT4 Collaboration, GEANT4—a simulation toolkit. Nucl. Instrum. Methods A **506**, 250 (2003). https://doi.org/10.1016/S0168-9002(03)01368-8
37. CMS Collaboration, Description and performance of track and primary-vertex reconstruction with the CMS tracker. JINST **9**, P10009 (2014). https://doi.org/10.1088/1748-0221/9/10/P10009. arXiv:1405.6569
38. CMS Collaboration, Jet algorithms performance in 13 TeV data. CMS Physics Analysis Summary CMS-PAS-JME-16-003 (2017)
39. T. Adye, Unfolding algorithms and tests using RooUnfold. In PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, H. Prosper, L. Lyons, eds., p. 313. Geneva, Switzerland, (2011). https://doi.org/10.5170/CERN-2011-006.313. arXiv:1105.1160
40. M.H. Seymour, A. Siodmok, Constraining MPI models using $\sigma_$eff and recent Tevatron and LHC underlying event data. JHEP **10**, 113 (2013). https://doi.org/10.1007/JHEP10(2013)113. arXiv:1307.5015

Appendix C

Particle level top-jet definition

In this chapter the studies related to the particle level top-jet definition are presented. These studies are important for understanding the results presented in Chapter 3.

In simulated $t\bar{t}$ events, the information about the t-quarks, b-quarks and W-bosons can be assessed. For instance, one can directly know whether a t-quark is associated to a certain jet by estimating the spatial separation in the $\eta - \phi$ plane between the jet and the direction of the t-quark. However, this information cannot be assessed in actual data, therefore a particle level top-jet definition is further required. First, the parton level information is used to study boosted top jets (section C.1), and later in section C.2, the particle level top-jet definition is derived. $t\bar{t}$ events are simulated with POWHEG+PYTHIA8 Monte Carlo samples.

C.1 Studies of top jets at parton level.

The condition that a jet might be associated to a t quark, using the parton level information can be considered by the following expression:

$$\Delta R_{tj} = \sqrt{(\phi_t - \phi_{jet})^2 + (\eta_t - \eta_{jet})^2} < R_{par} \tag{C.1}$$

where ΔR_{tj} is the distance between the jet-axis and the t-quark direction in the $\eta - \phi$ plane, and R_{par} is a distance parameter to be defined. The smaller the value of the R_{par} parameter is, the better the jet under consideration is reconstructing the t-quark.

A second parton level level infromation useful to study the top-jets is the distance between the jet axis and the W-boson. This parameter is considered as:

$$\Delta R_{W-jet} = \sqrt{(\phi_W - \phi_{jet})^2 + (\eta_W - \eta_{jet})^2} \tag{C.2}$$

Scenarios where the b-quark is outside the jet associated to the t-quark have a peculiar feature concerning the distance parameter ΔR_{W-jet}. Jets reconstructing the W-boson (i.e the Soft Drop jet mass is peaked around the W-boson mass window) satisfy always the condition: $\Delta R_{W-jet} < \Delta R_{tj}$. Therefore those unwanted scenarios are easy to identify using the parton level information.

However, the opposite requirement: $\Delta R_{tj} < \Delta R_{W-jet}$, does not guarantee a good t-quark reconstruction. In the phase space $\Delta R_{tj} < \Delta R_{W-jet}$, the R_{par} can be used to identify top-jets.

The following jet categories, using the observables presented in Equations C.1 and C.2, can be defined:

- ✓ *signal jets (top-jets):* $\Delta R_{tj} < R_{par}$ and $\Delta R_{tj} < \Delta R_{Wj}$
- ✓ *non-top jets:* $\Delta R_{tj} > R_{par}$ and $\Delta R_{Wj} > \Delta R_{tj}$
- ✓ *W-jets:* $\Delta R_{tj} < 0.8$ and $\Delta R_{Wj} < \Delta R_{tj}$
- ✓ *QCD-like jets:* $\Delta R_{j-jet} > 0.8$ or $\Delta R_{W-jet} > 0.8$.

Figure C.1 (left) shows a schematic picture of the different jet definitions considering the $\Delta R_{tj} - \Delta R_{Wj}$ plane. The signal jets (top-jets) and the non-top jets (contaminated jets) categories depend on the exact value of the R_{par} parameter. The W-jets are those reconstructing the W-boson, while the QCD-like jets are those jets which cannot be associated to any of the t-quarks (probably containing recoiling QCD radiation of the $t\bar{t}$ system). Figure C.1 (right) shows the Soft Drop jet mass distribution considering the leading jet in simulated $t\bar{t}$ events. The different jet categories are illustrated, and the R_{par} is considered as $R_{par} = 0.02$.

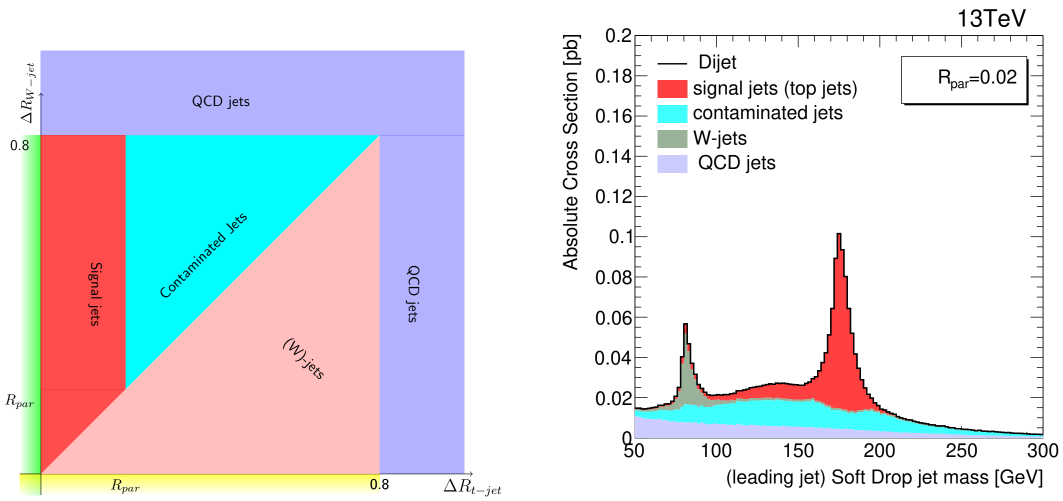


Figure C.1: (left) Sketch picture representing the phase space region defined by ΔR_{W-jet} and ΔR_{t-jet} observables. Different jet categories defined in this section are illustrated. (right) Soft Drop jet mass distribution for the leading jet in $t\bar{t}$ events. Jets are classified following the jet categories in the sketch.

From Figure C.1 the following can be conclude:

- ✓ the jets classified as QCD-like jets have a Soft Drop mass distribution similar to the case of simulated QCD multijet events, showing that in those cases, the leading jet is only containing QCD radiation.
- ✓ the jets classified as W-jets have a Soft Drop jet mass distribution peaked at the W-boson mass window.
- ✓ the jets classified as non-top jets have a Soft Drop jet mass distribution spread around the whole phase space.

Following the results presented here, a top-jet can be defined using the parton level information as follows:

- ✓ a t-quark can be associated to the jet: $\Delta R_{t-jet} < 0.02$
- ✓ the jet-axis is closer to the t-quark than to the W-boson: $\Delta R_{tj} < \Delta R_{Wj}$

This definition is further considered as a base-line in order to derive the top-jet hadron definition.

C.2 Top-jet definition at hadron level

Figure C.2 shows the Soft Drop mass distribution for leading and subleading jets, in simulated $t\bar{t}$ events. Jets have been classified as top-jets using the parton level definition presented in the Section C.1. In the following jets not satisfying the definition are considered under the category of non-top jets.

The hadron level top-jet definition is derived with observables defined at stable particle level (hadron level). The optimal selection criteria are derived, considering a balance of signal efficiency (fraction of top jets which are actually considering after applying a selection criterion) and misidentification probability (fraction of non-top jets considered as top jets). Four variables are

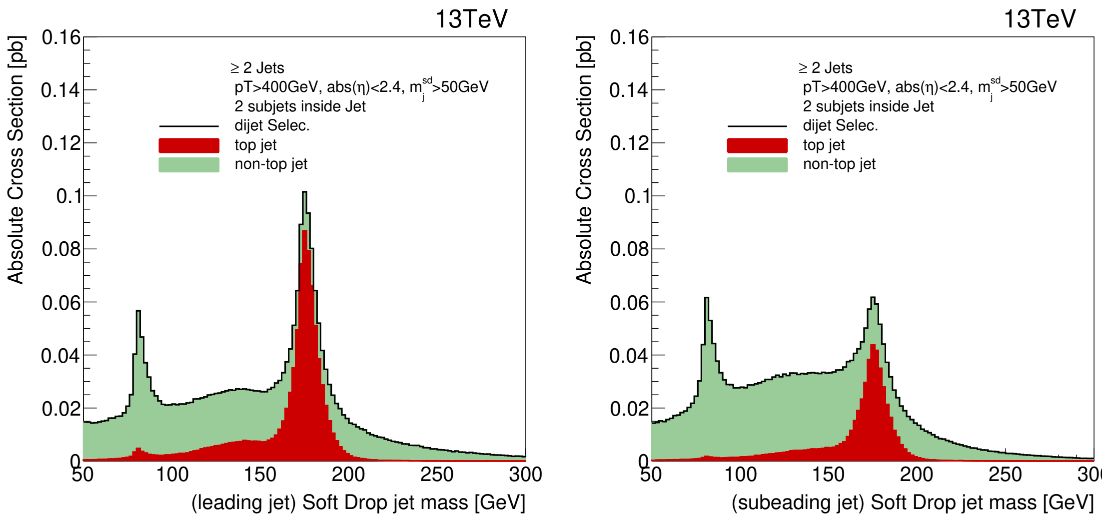


Figure C.2: Soft Drop jet mass distribution for (left) leading jet, (right) subleading jet, in $t\bar{t}$ simulated events. Evens were selected with the boosted dijet criteria. Two jet categories are shown: top jets, and non-top jets, using the parton level infromation.

considered as possible discriminating observables: related to the subjettiness observables (τ_3/τ_2, τ_3/τ_1), and the Soft Drop mass of two subjets (m_j^{sub0},m_j^{sub1}). Figure C.3 shows the respective distributions of those observables, considering the leading jet in $t\bar{t}$ simulated events and following the top-jet parton definition presented in Section C.1.

Figure C.4 represents the correlation between misidentification probability and signal efficiency by considering a selection cut on the respective observables. The efficiencies curves are illustrated for the leading and subleading jets separately.

The two observables performing the best (better discriminating non-top jets from top jets) are τ_3/τ_2 ratio, and the m_j^{sub0} (the mass of the subjet with higher p_T). The specific selection cut for those observables further consider correspond to a selection efficiency of 80% ($\tau_3/\tau_2 < 0.65$ and $m_j^{sub0} > 40$ GeV). Table C.1 shows the two possible top-jet hadron definitions considering these selection criteria.

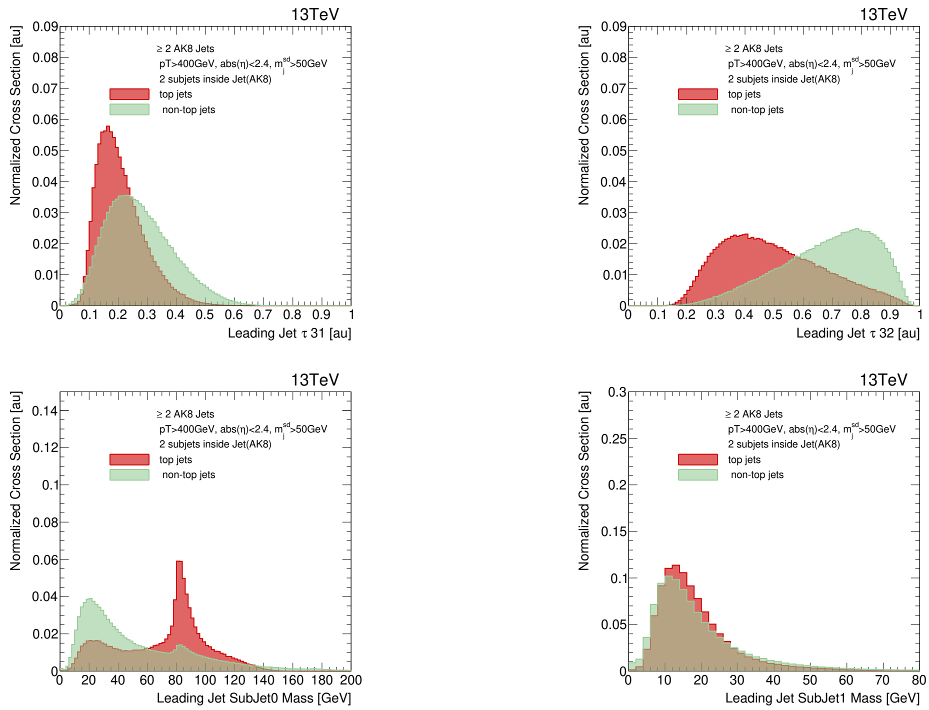


Figure C.3: Normalized distributions considering top jets and non-top jets for the leading jet in simulated $t\bar{t}$ events. The classification for top-jets and non-top jets corresponds to the parton level definition presented in section C.1. The following observables are considered: (upper, left) τ_3/τ_1; (upper, right) τ_3/τ_2; (down, left) m_j^{sub0}; (down, right) m_j^{sub1}.

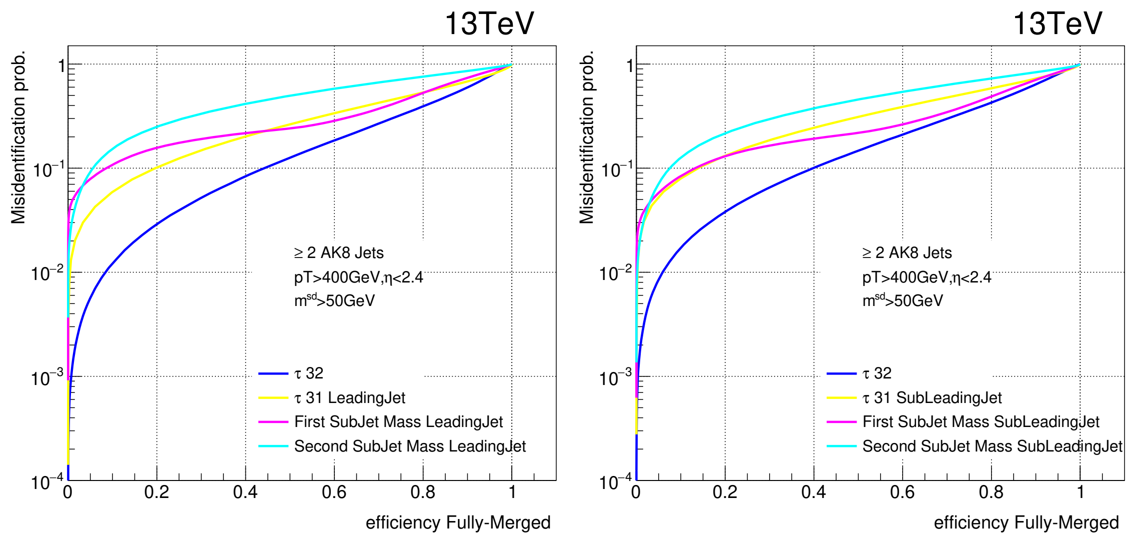


Figure C.4: Signal efficiency for top-jet selection as function of the misidentinfication probability of classifying as signal jet a non-top jet. Different observables are considered as discriminating variables. The leading jet (left plot) and the subleading jet (right plot) are considered independently.

Table C.1: Top-jet hadronic definition. Two variations are provided considering the selection criterion $m_{sub0} > 40$ GeV and $\tau_3/\tau_2 < 0.65$

Definition (A)	Definition (B)
$p_T > 400$ GeV, $m_j^{sd} > 50$ GeV	$p_T > 400$ GeV, $m_j^{sd} > 50$ GeV
contains a B-hadorn	contains a B-hadorn
$m_j^{sub0} > 40$ GeV	$\tau_3/\tau_2 < 0.65$

Figure C.5 shows the Soft Drop jet mass distribution for the leading jet in simulated $t\bar{t}$ events. The two respective plots correspond to the definitions presented in Table C.1 (the left plot to the definition A, the right plot to the definition B).

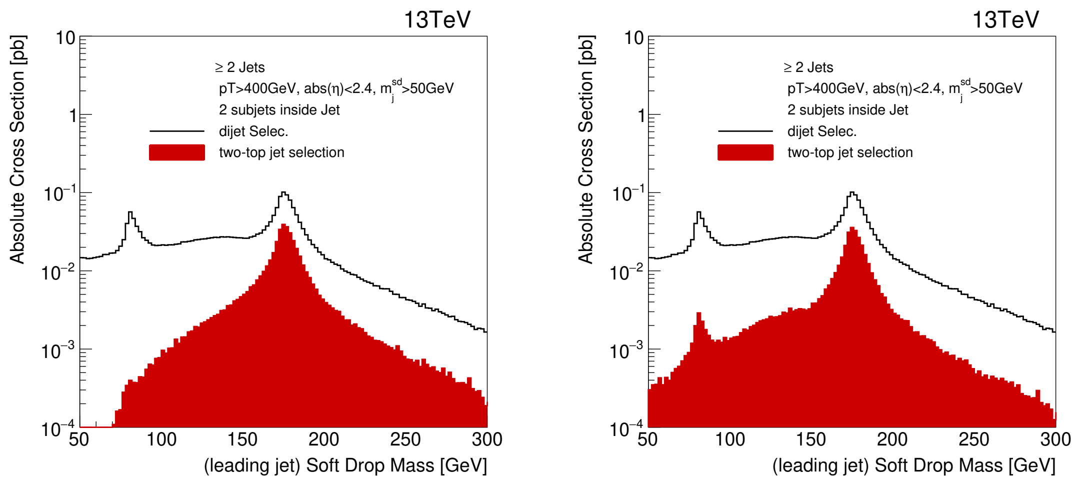


Figure C.5: Soft Drop jet mass distribution corresponding to the leading jet in simulated $t\bar{t}$ events. First, the events are selected with the boosted dijet selection criteria, and secondly by requiring two top jets using: the definition A (left plot), the definition B (right plot) provided in Table C.1

From the results presented in C.5, it can be notices a clear advantage of the definition A with respect to the definition B: considering this selection criteria, the contribution given in the W-mass window is completely removed. Therefore, this is the top-jet hadron definition further considered in the studies presented in this Thesis.

Appendix D

Top-jet definition using detector level information

This chapter presents similar studies that the one presented in Chapter C, but using detector level information. The top-jet definition deduced in this appendix is applied in the selection strategy discussed in Chapter 6.

In order to find the optimal variable and its respective cuts, two different categories for jets are defined:

- ✓ fully merged (top jet): the jet is defined by the Particle Flow as b-flavour jet and additionally it is matched with a W-boson ($\Delta R < 0.4$), considering $\tau_{32} < 0.70$,
- ✓ non-merged (non-top jet): all the jets not satisfying the previous condition.

Figure D.1 shows the soft drop mass distribution, where the contributions from these two categories in the both leading jets are illustrated.

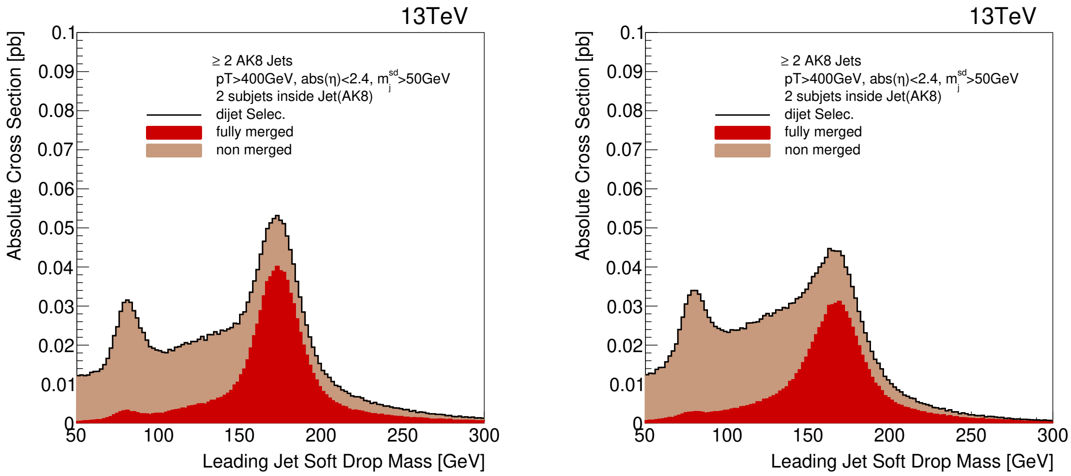


Figure D.1: Soft Drop jet mass distributions in $t\bar{t}$ events considering two categories: fully merged and non-merged. Fully merged refers to the condition that the jet is classified as $b\text{-}flavour$ jet with the additional requirement that $\tau_{32} < 0.7$, while the non-merged sample are the remaining jets. The (left) leading jet, and (right) subleading jet are considered separately.

In order to study the discriminating power of different observables considering the two categories, six variables are taken into account. Figure D.2 illustrates the distribution of the considered variables for fully-merged and non-merged jets. The distributions are obtained at detector level of simulated $t\bar{t}$ events, considering in this case the leading jet. The variables are: τ_1, τ_2, τ_3, τ_{31}, the soft drop mass of the first subjet and the soft drop mass of the second subjet.

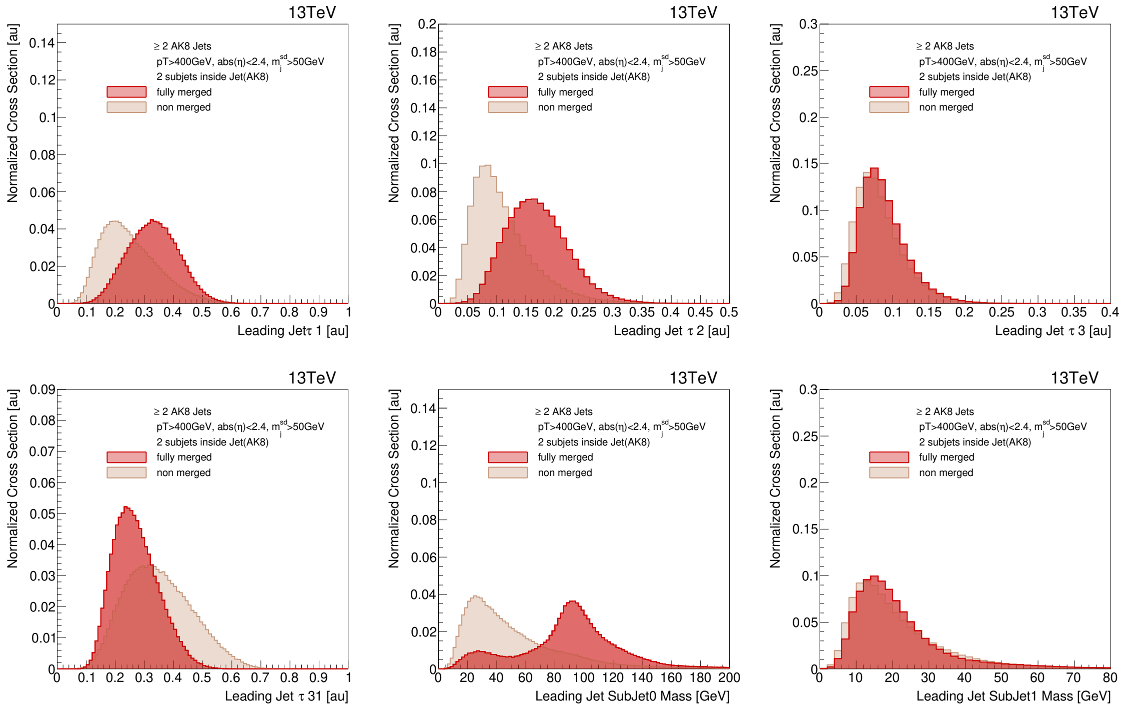


Figure D.2: Discriminating variables considering the leading jet in order to select fully and non merged jets: (upper,left) τ_1,(upper,middle) τ_2, (upper,right) τ_3, (down, left) τ_{31}, (down middle) first subjet soft drop mass, (down right) second subjet soft drop mass.

Figure D.3 shows the efficiencies of selecting fully merged top jets, with respect to the misidentifaction probability for non-merged jets, considering the leading and subleading jets separately. The curves are reported for each of the mentioned observables.

The selected variable to perform the cut was the mass of the first subjet, since in the region of higher efficiency, it is the variable with better performance. The selected cut values correspond to the point where the efficiency of selecting fully merged jets takes values of $\sim$ 80%. The cuts were defined as $m_{subjet0} > 59$ GeV for the leading jet, and $m_{subjet0} > 55$ GeV for the subleading jet. The misidentification rates for those selected cuts were $\sim$ 30%. The efficiency is an average value, which does not give any information about the performance with respect to p_T. Nevertheless, it was found that the performance is approximately constant as a function of p_T.

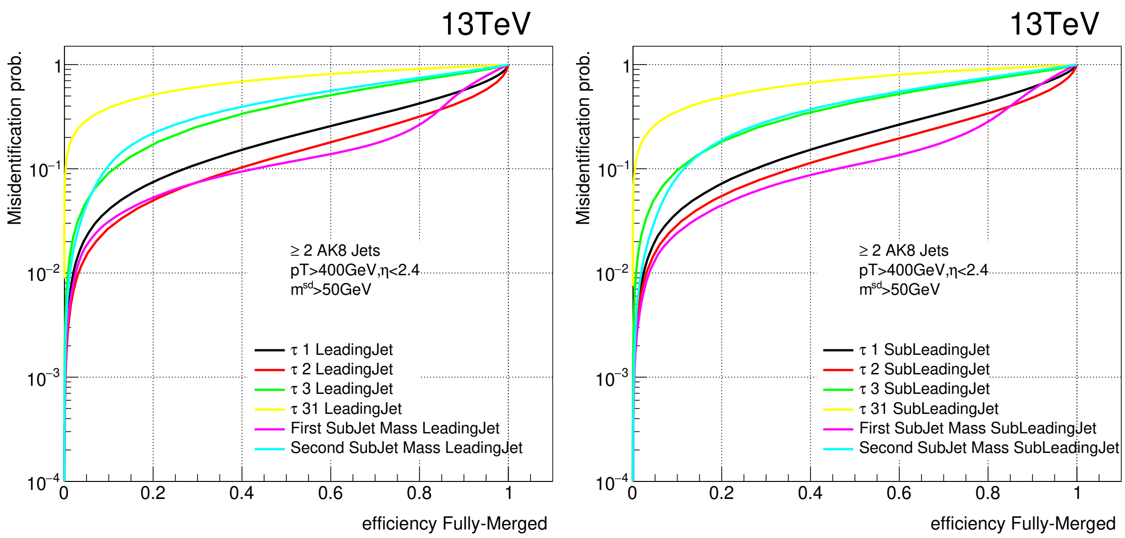


Figure D.3: Efficiency of correctly identifying fully merged top jets with respect to the misidentification probability of classifying non-merged jets as fully merged: (left) leading jet, (right) suleading jet.

Appendix E

b-tagging efficiency studies in boosted top topologies

In this Chapter, the studies on the b-tagging efficiency at the very high p_T region are presented. Simulated $t\bar{t}$ and QCD multijet events at detector level are considered. The former ones are simulated with POWHEG+PYTHIA8, while the latter with MADGRAPH+PYTHIA8.

Figure E.1 shows the efficiency selection, as function of the jet p_T. The CSVv2 tagger is applied in two ways: to the subjets in the fat jet, or to the fat jet itself. The plot at the left is obtained for *tbart* events, wile the plot at the right for QCD events. For the estimation of this efficiency, the jet flavour definition by the Monte Carlo CMS tool [163] is used. Hence, the efficiency can be defined as function of p_T of the jet by the following expression:

$$\varepsilon(p_T) = \frac{N(\text{btagged, bflavour jet})}{N(\text{bflavour jet})} \tag{E.1}$$

where $N(\text{bflavour jet})$ is the number of jets classified as b-flavour jets, and $N(\text{btagged, bflavour jet})$ is the number of jets classified as b-flavour jets which are actually b-tagged by the CSVv2 tagger using the MWP. The efficiency is calculated having as baseline those selected events by the boosted dijet selection criteria. In both cases ($t\bar{t}$ and QCD events) the performance of the b-tagging applied to the subjets is better in the whole p_T range, although for the higher p_T phase space region the performances become similar.

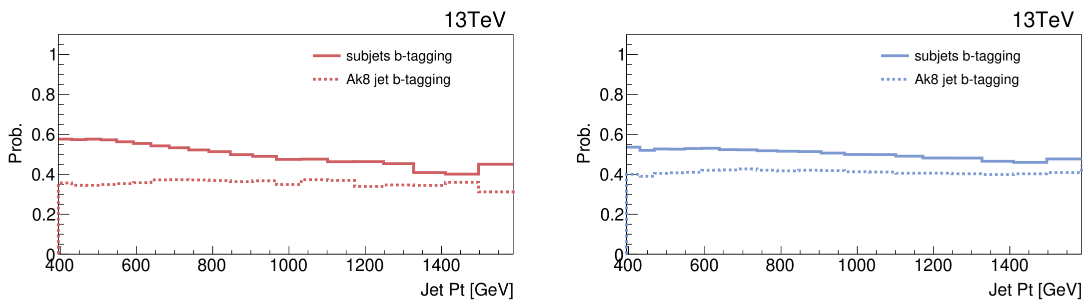


Figure E.1: B-tagging efficiency as a function of p_t: the probability is estimated with (left) $t\bar{t}$ and (right) QCD simulated events with at least two jets with $p_t > 400\text{GeV}$ and $|\eta| < 2.4$. The b-tagging requirement is applied to either the subjets, or the fat jet.

Figure E.2 shows the correlation of the b-tagging efficiency with the mistagging identification probability, applying the CSVv2 tagger to the subjets within the fat jet. Simulated $t\bar{t}$ and QCD events are considered, in the boosted dijet phase space region. The dependence between the two quantities (efficiency-mistagging probabilities) is reported for QCD and $t\bar{t}$ events, considering light jets and c-jets independently.

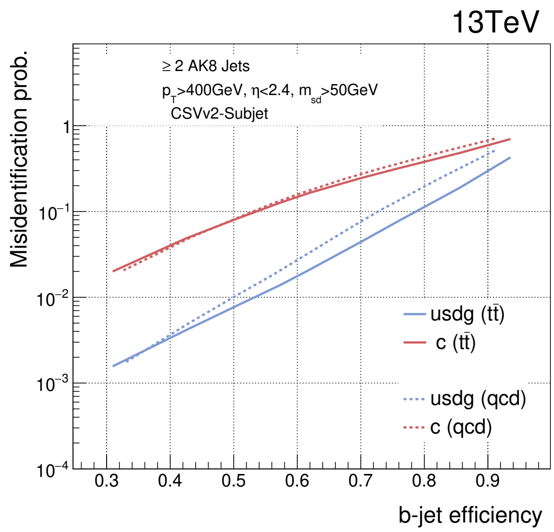


Figure E.2: b-tagging efficiency performance for the CSVv2 algorithm applied to the subjets obtained by the Soft Drop Mass mechanism in simulated $t\bar{t}$ and QCD events with two $Ak8$ jets with $p_t > 400$GeV, $|\eta| < 2.4$ and $m_j^{sd} > 50$ GeV.

Table E.1 shows the exact values of the b-tagging efficiency and misidentification probability for QCD and $t\bar{t}$ samples. These values were deduced from the results presented in Figure E.2. The efficiencies for the defined working point are provided: light working point (LWP), medium working point (MWP), and tight working point (TWP). The performance of the MWP (the one that is further used in the analysis) has $\sim$ 1.4% of misidentification probability for light jets (the expected value for this working point is 1.%) and a b-tagging efficiency of $\sim$ 84% (the expected value is $\sim$ 80%).

$t\bar{t}$ events

Working Point	c-Jets prob. %	light-Jets prob %	b-Jets prob %
Light WP	48	18.7	85.5
Medium WP	12.6	1.38	57.3
Tight WP	4.8	0.41	42.45

QCD events

Working Point	c-Jets prob. %	light-Jets prob %	b-Jets prob %
Light WP	47.6	23.30	81.9
Medium WP	12.5	1.84	56.26
Tight WP	4.90	0.50	43.1

Table E.1: Efficiency of b-tagging truth b-jets, or mistagging light-jets and c-jets for different working points of the CSVv2 tagger: (upper table) $t\bar{t}$ events, and (down table) QCD events.

In order to evaluate the performance of a tagger, not only the efficiency and the mistagging ratios are of interest. Among the tagged jets, there might be non truth b-jets (*contamination* of the sample). This contamination depends strongly on the nature of the events since the fraction of each quark flavour is different for QCD multijet events and $t\bar{t}$ events.

Figure E.3 shows the flavour composition of the b-tagged sample as a function of p_T of subjets jet for $t\bar{t}$ and QCD simulated events.

As expected, the contamination of the QCD sample is larger and increasing with the p_T of the jet. The purity of the sample is by definition the fraction of b-truth jets in the tagged sample, and is also shown in the Figure E.3, where for $t\bar{t}$ events is behaving constantly, while for QCD events, is strongly decreasing as a function of p_T. In the case of QCD events, the contamination of c-quarks is affecting the purity.

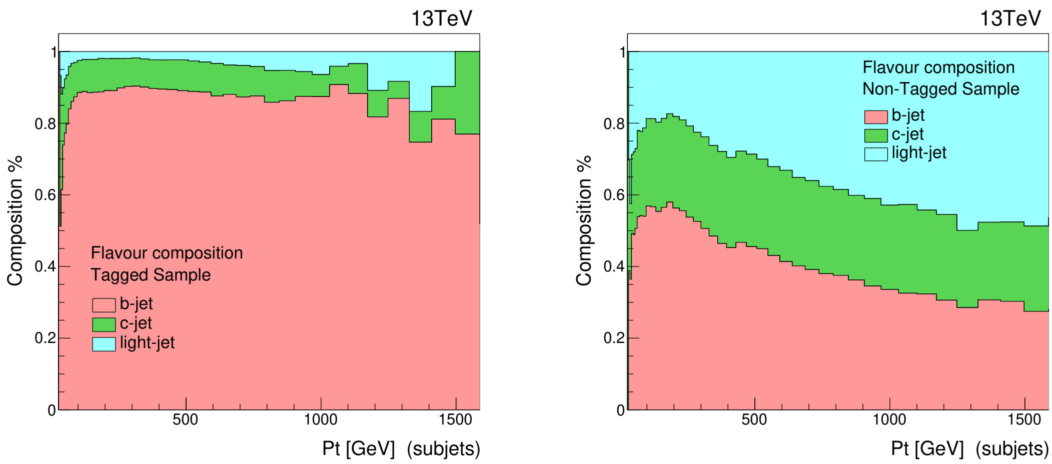


Figure E.3: Flavour composition for the b-tagged sample: (left) $t\bar{t}$ events, (right) QCD events.

On the other hand, other composition that is of interest corresponds to the non-btagged sample. There, the most important message is how much of the sample categorized by "non b-jets" was actually truth *b-jets*. Figure E.4 shows the composition of the non b-tagged sample, for $t\bar{t}$ and QCD events.

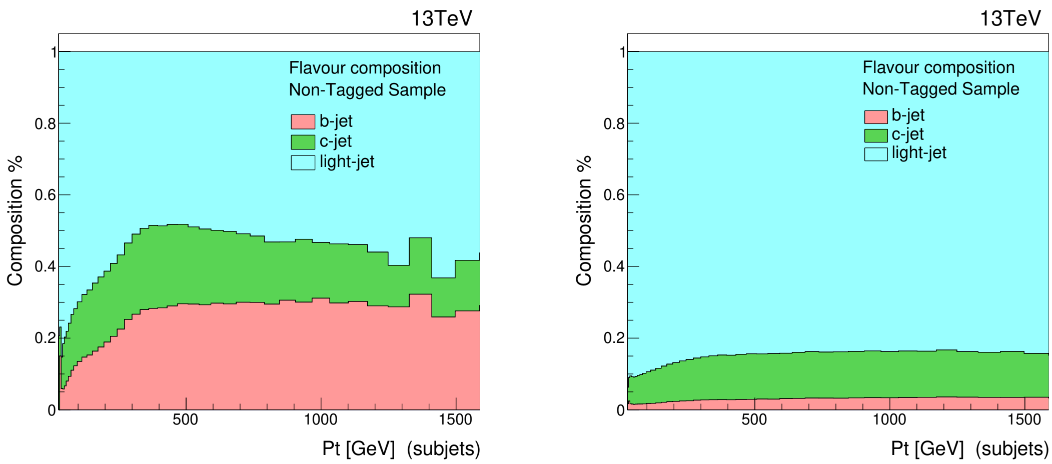


Figure E.4: Flavour composition for the non-b tagged sample: (left) $t\bar{t}$ events, (right) QCD events.

Appendix F

Background Estimation. Additional material

In this Chapter, additional material to the Chapter 8, related to the background subtraction procedure are presented.

The following list of plots are provided:

- ✓ d_{MVA} distributions for the leading jet in different regions
- ✓ soft drop jet mass of the first subjet (Subjet0) distributions for the leading jet ,
- ✓ soft drop jet mass of the second subjet (Subjet1) distributions for the leading jet,

the regions correspond to the ones illustrated in Figure 8.5.

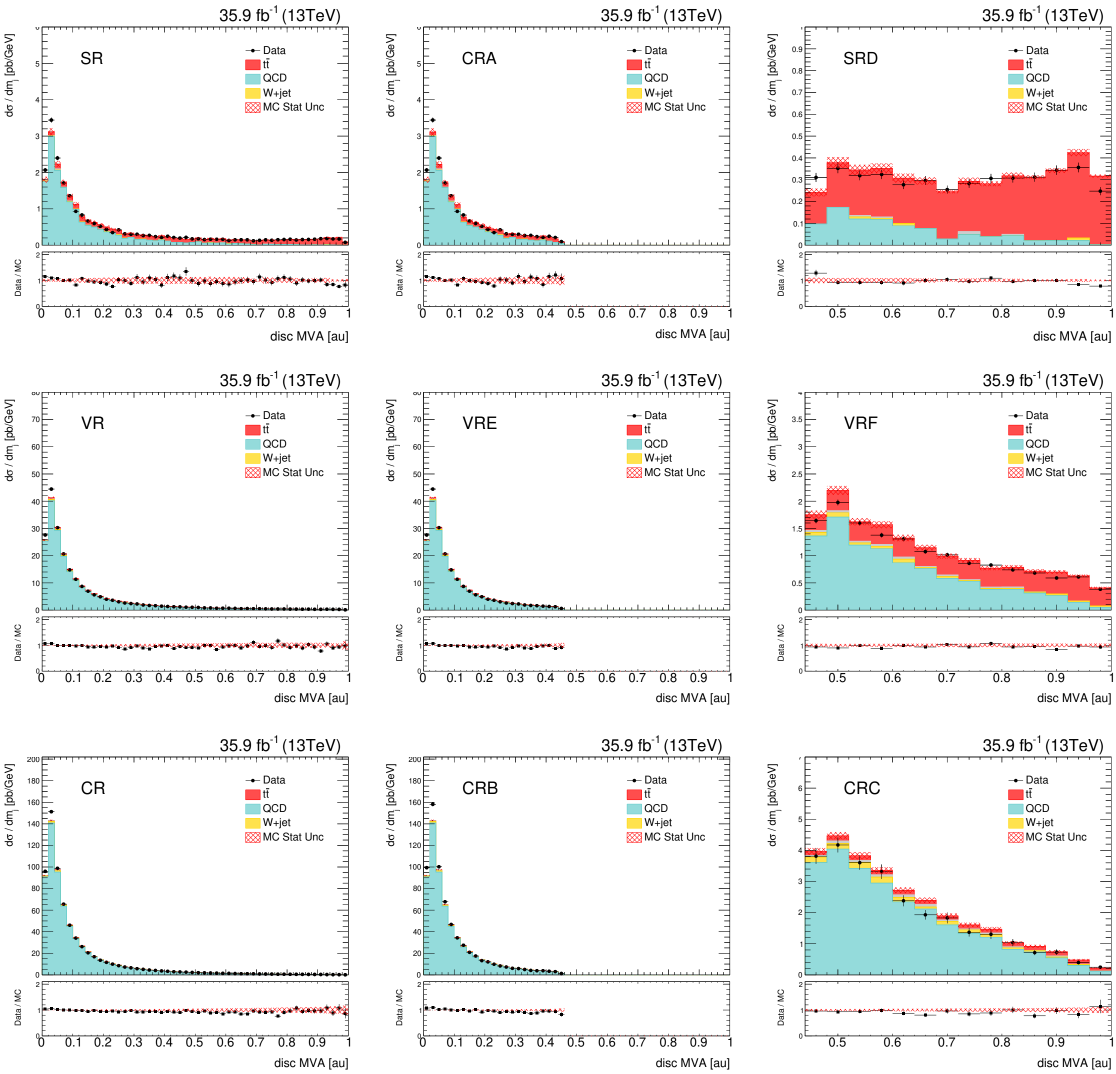


Figure F.1: d_{MVA} distributions for the leading jet in different regions defined in Figure 8.5. The normalization factor for the $t\bar{t}$ contribution has been taken as $N_{t\bar{t}} = 0.654$, while for the QCD multijet contribution: $N_{QCD} = 0.74$ in the first row (2-btagged jets), $N_{QCD} = 0.67$ in the second row (exclusively 1-btagged jet), $N_{QCD} = 0.6$ in the third row (zero btagged jet).

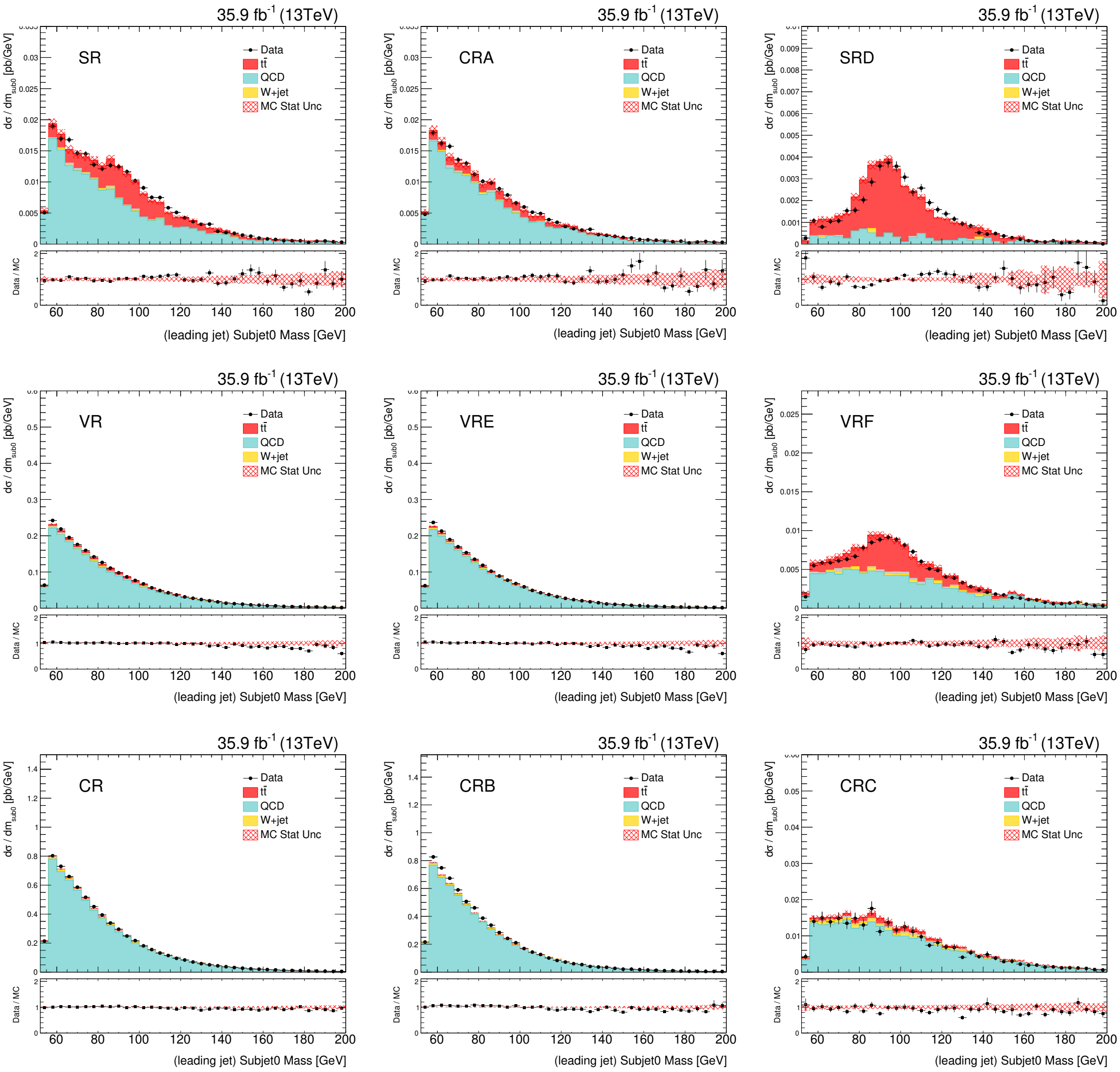


Figure F.2: First subjet soft drop mass distributions for the leading jet in different regions defined in Figure 8.5. The normalization factor for the $t\bar{t}$ contribution has been taken as $N_{t\bar{t}} = 0.654$, while for the QCD multijet contribution: $N_{QCD} = 0.74$ in the first row (2-btagged jets), $N_{QCD} = 0.67$ in the second row (exclusively 1-btagged jet), $N_{QCD} = 0.6$ in the third row (zero btagged jet).

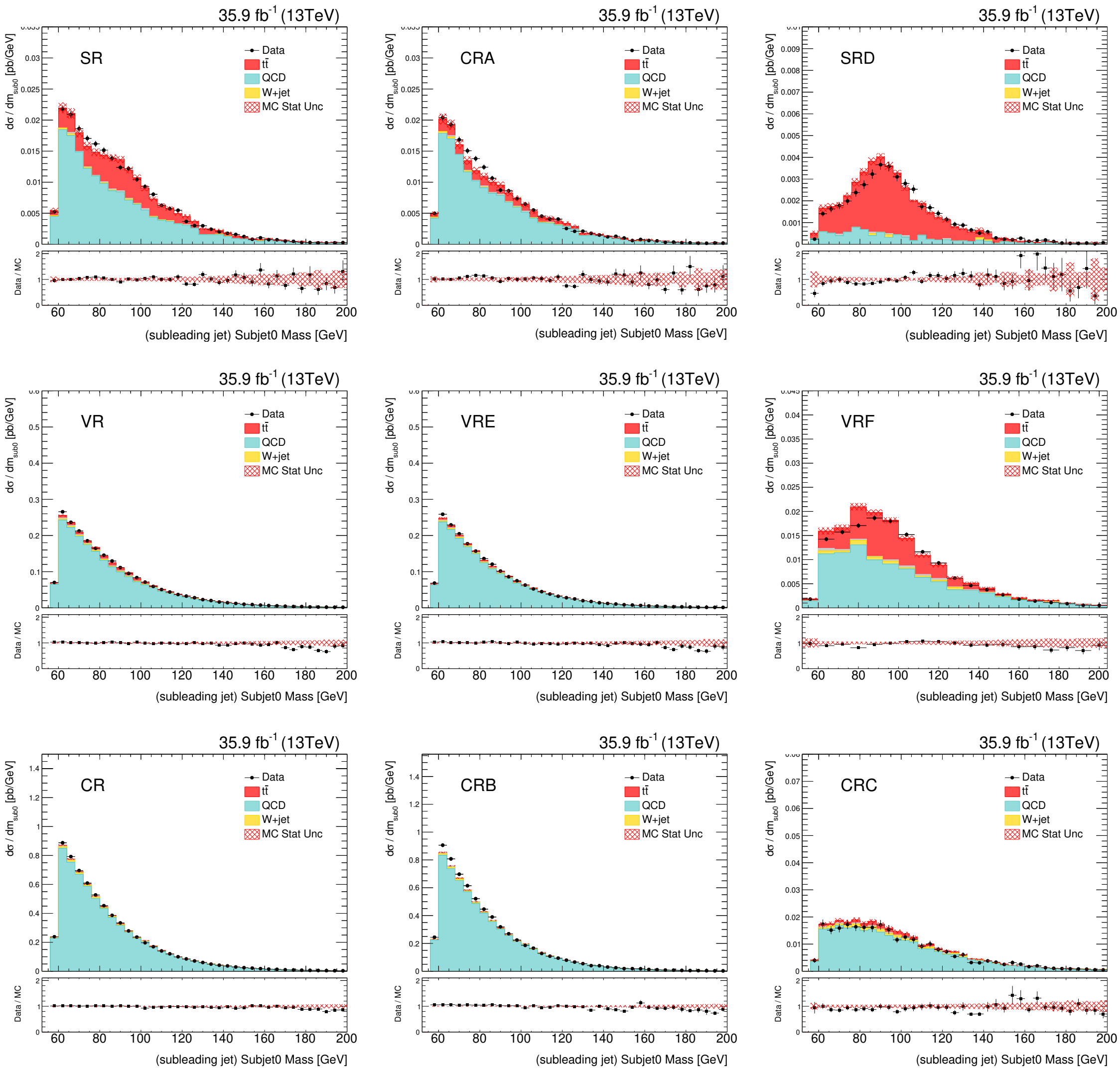


Figure F.3: Second subjet soft drop mass distributions for the leading jet in different regions defined in Figure 8.5. The normalization factor for the $t\bar{t}$ contribution has been taken as $N_{t\bar{t}} = 0.654$, while for the QCD multijet contribution: $N_{QCD} = 0.74$ in the first row (2-btagged jets), $N_{QCD} = 0.67$ in the second row (exclusively 1-btagged jet), $N_{QCD} = 0.6$ in the third row (zero btagged jet).

Appendix G

Detector Effects and Unfolding. Additional material

This Chapter presents closure test to the unfolding procedure performed in Chapter 9.

The final unfolded results should be independent on the unfolding procedure. However, possible bias introduced through the unfolding method need to be checked. There are several ways to perform tests for the unfolding procedure, and some of them are further discussed in this section.

The closure tests to the unfolded procedure are then organized in two main categories. The first category is related to different unfolding algorithms. Those unfolding methods, already discussed in the subsection 9.2.1, are then applied to the measured distributions in this thesis. The results are presented in the subsection G.1. The second category consists of using different Monte Carlo predictions to provide the input information to the unfolding procedure. The results are then given in the subsection G.2.

G.1 Comparison with other Unfolding Methods.

The comparison between the unfolded measured distributions obtained by using different methods are shown in Figure G.1. The plots correspond to the following observables: (upper plots) $\Delta\phi$ in the extended phase space, as well as, in the most back to back region, (lower plots) the p_T of the leading and subleading jets. The provided input input information to the unfolded procedure has been produced from the sample corresponding to the Powheg and Pythia8 (with the CUETP8M1 tune) event generators. The distributions have double size bins in order to compare with the output of the TUnfond unfolding procedure. The two level distributions (generator and detector) in the Data and Monte Carlo are illustrated. The data distributions at generator level are the result of unfolding the measurements with different methods. The Monte Carlo distributions at both levels are scaled down by an overall factor ($s = 0.70$). The ratios between the unfolded distributions, with respect to the default result in this thesis (given by the output of the D´Agostini method with 10 iterations), are as well provided. The statics uncertainties obtained by the default method are illustrated by the yellow band in the ratio plots.

The SVD unfolding method is implemented by using as regularization parameter the number of bins of the original data distribution, having considerable entries. However, several regularization parameters for this method were implemented and no major changes were found, when the method converge. The "*bin by bin*" method considers, in a simple way, the corrections factors obtained by dividing the generator and detector Monte Carlo distributions. Finally, the distributions obtained by the TUnfold method are estimated without considering the regularization parameter ($\tau = 0$), given in the equation 9.22.

In the case of the $\Delta\phi$ observable (the upper plots), similar results are obtained independently which method is implemented. Both, the central bins values and the estimated statistics uncertainties of the unfolded distributions are totally compatible for all the methods.

In the case of the p_T observable (the lower plots), slightly different results are obtained by using the TUnfold method in the higher p_T region. The TUnfold method however relies on a scaling factor applied to the fake contribution (given by events reconstructed at detector level but not having an equivalent event at generator level). The scaling factor is considered as $b = 0.70$, bringing rather good agreement in the lower p_T region. Nevertheless, all the distributions agree within the statistic uncertainties.

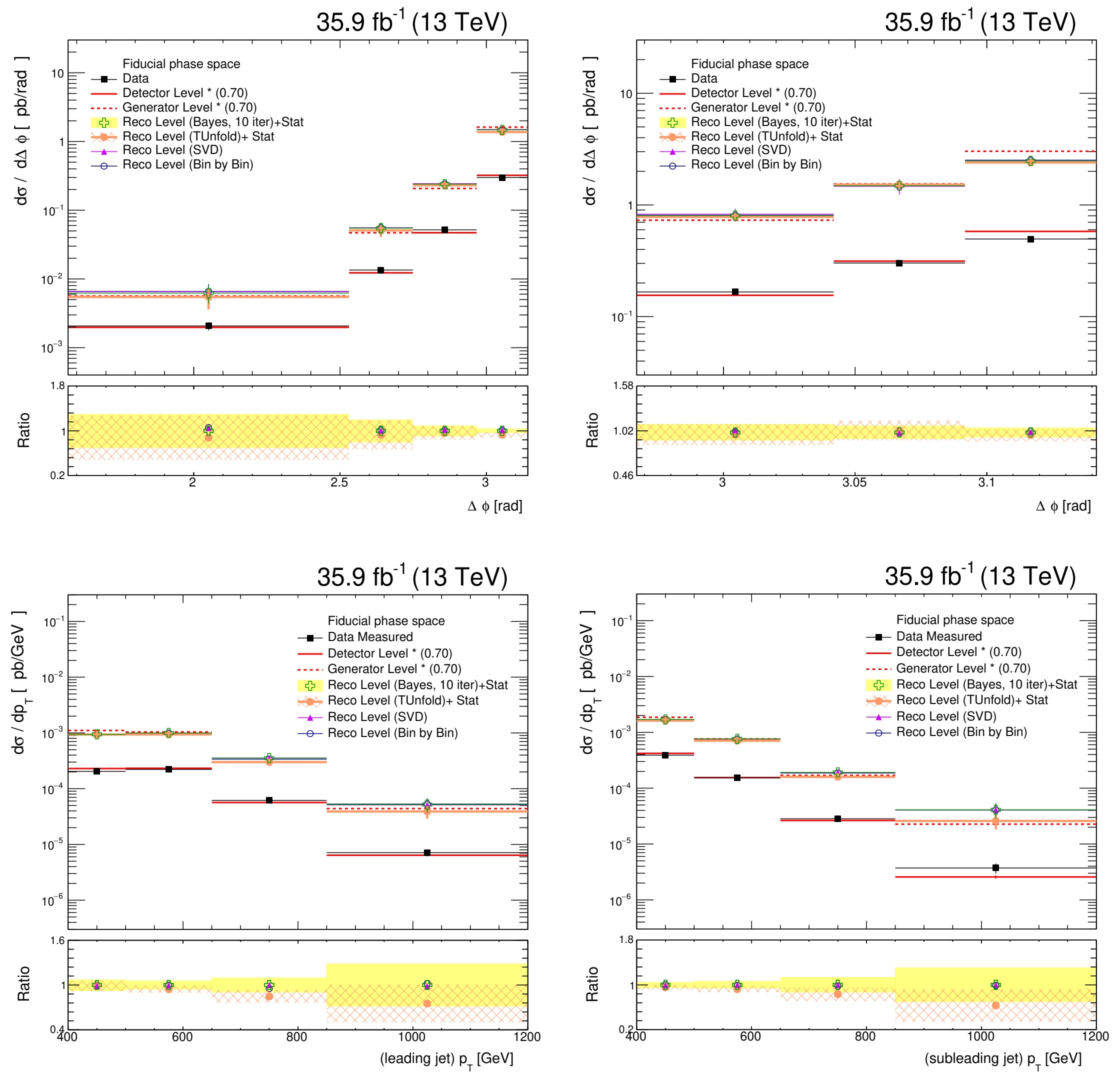


Figure G.1: Comparison between stable particle level and detector levels. The stable particle level distributions are obtained by unfolding the data distribution at detector level using different unfolding methods. The plots correspond to the following distributions: (upper left) $\Delta\phi$ in the inclusive phase space (upper rigth) $\Delta\phi$ when the refined binning in the most back-to-back region is considered, (lower left) the p_T spectra of the leading jet, (lower right) the p_T spectra of the subleading jet. The lower panel shows the ratio of the unfolded distributions using different methods, with respect the unfolded distribution using the D´Agostini method with 10 iterations. The Monte Carlo distributions have been scaled down by factor of $s = 0.70$.

G.2 Unfolding using different Monte Carlos samples.

The unfolded results could be also biased through the input information given by the Monte Carlo simulation. The default Monte Carlo sample, which is used to provide the response matrices as input information to the unfolding procedure corresponds to the Powheg and Pythia8 event generators, with the CUETP8M1 [37] tune (for simulating $t\bar{t}$ events).

Figure G.2 shows the test consisting of unfolding the detector level distribution (considered as pseudo-data) provided by the default Monte Carlo sample. The input information for the unfolding algorithm (the D´Agostini method with 10 iterations) is provided alternatively using two different Monte Carlo samples: one using the CUTEP8M2 tune, and the other using the Color-Reconnection tune. The plots show the stable level distributions, comparing the unfolded results with the theory predictions. Four different observable are illustrated: the $\Delta\phi$ observable for the two considered binning, and the p_T observables for the leading and subleading jets. The ratio between the predictions and the unfolded pseudo-data are also provided. A good agreement is observed for all the observebles, therefore the results are considered independent on the Monte Carlo simulations.

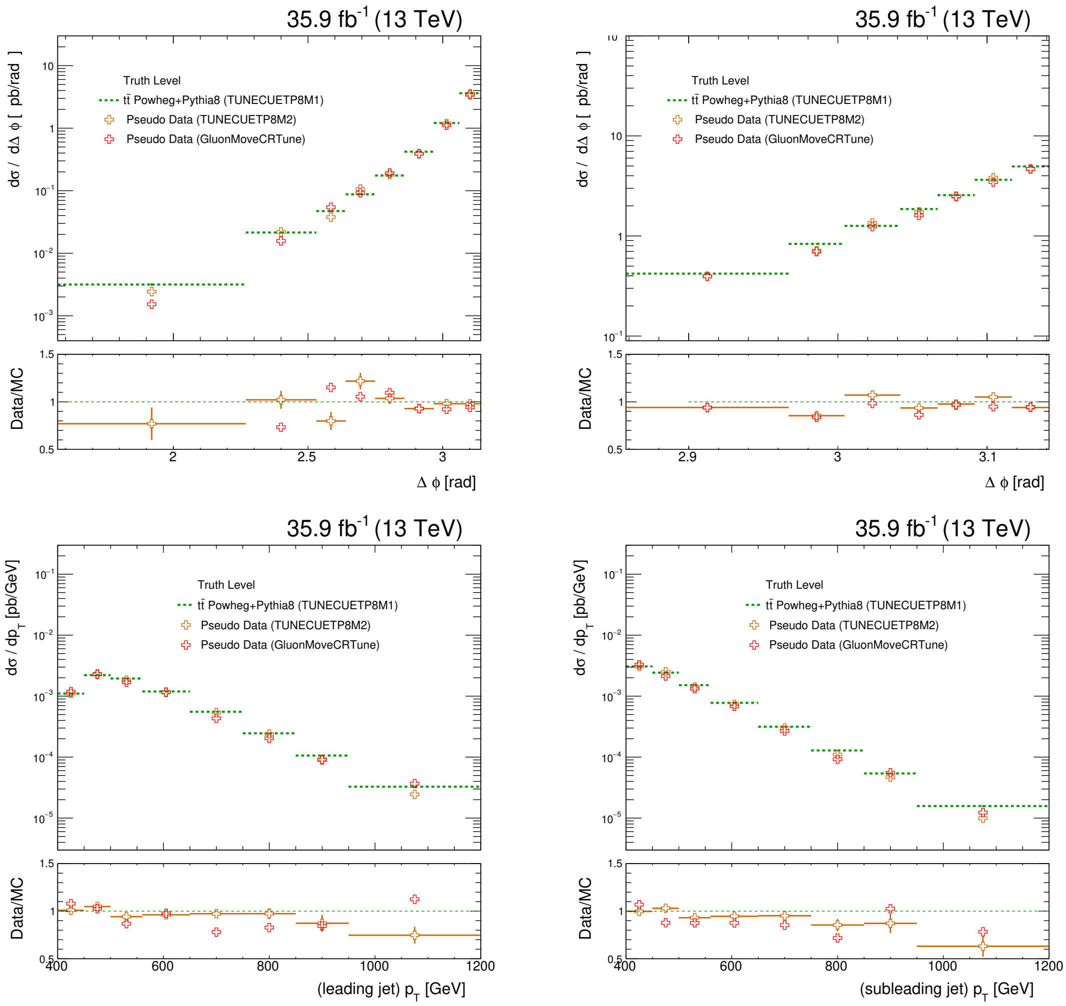


Figure G.2: Comparison of the Monte Carlo predictions and the unfolded pseudo-data at stable level (truth level). The unfolded procedure is implemented using different Monte Carlo samples for the input information. The following observables are illustrated: (upper left) the $\Delta\phi$ observable in the whole azimuthal phase space, (upper right) the $\Delta\phi$ observable in the most back-to-back region, (down left) the p_T of the leading jet, (down right) the p_T of the subleading jet.

Appendix H

Systematic Uncertainties. Additional material

In this Chapter, additional material related to the estimated systematic uncertainty (presented in Chapter 10) are provided.

Tables H.1 and H.2 give the estimated experimental uncertainty in all the measured distributions for the both measurements considered in this thesis: $t\bar{t}$ measurements, and those ones where the QCD multijet contribution is considered as part of the signal. Tables H.3 and H.4 report the results of the modeling systematic uncertainty also for the mentioned measurements respectively. All the presented results correspond to the particle level measurements.

The breakdown of the total systematic uncertainty at particle level (unfolded level) are illustrated in Figures H.1, H.2 and H.3. The plots correspond to the $t\bar{t}$ measurements, and for each observable, the uncertainty in the absolute (left plots) and in the normalized distributions (right plots) are illustrated.

Figure H.1 illustrates the systematic uncertainty as function of the $\Delta\phi$ observable, when the two exclusive regions with respect to the p_T^{lead} (transverse momentum of the leading jet) are considered: 400GeV$< p_T <$ 600GeV (upper plots) and $p_T >$ 600GeV (lower plots).

Figure H.2 (H.3) illustrate the systematic uncertainty as function of the p_T of the leading (subleading) jet, when the two exclusive regions with respect to the η are considered: $|\eta| < 0.5$ (upper plots) and $0.5 < |\eta| < 2.4$ (lower plots).

The Table H.5 shows a summary of all sort of estimated uncertainties for the $t\bar{t}$ measurements (when the QCD contribution has been subtracted), in the fiducial particle level. The uncertainty estimated in the absolute and in the normalized distributions are provided. The systematic uncertainty has been breakdown in modeling and experimental uncertainty (the two first columns), and later compared to the statistical uncertainty. Finally the total uncertainty on the measurements are provided. Analogously to this results, in Table H.6, the uncertainty on the measurements in which the QCD background has been considered as part of the signal are provided.

Table H.1: Breakdown of the experimental uncertainty for the $t\bar{t}$ measurements

Observable	JES Unc.	JER Unc	BTag Unc.	Bckg Unc
Cross Section Measurements				
p_T leading jet	4.71% − 20.69%	0.39% − 2.38%	6.54% − 9.92%	0.63% − 6.40%
p_T leading jet $\eta < 0.5$	2.36% − 56.95%	0.75% − 8.49%	6.51% − 16.26%	0.62% − 18.98%
p_T leading jet $0.5 < \eta < 2.4$	1.33% − 18.57%	0.29% − 1.90%	5.54% − 7.94%	2.44% − 2.57%
p_T subleading jet	2.98% − 33.45%	0.55% − 2.01%	6.62% − 7.40%	2.60% − 3.25%
p_T subleading jet $\eta < 0.5$	3.60% − 38.69%	0.51% − 7.61%	6.75% − 14.26%	0.77% − 7.83%
p_T subleading jet $0.5 < \eta < 2.4$	1.19% − 31.41%	0.37% − 14.24%	6.56% − 21.28%	1.08% − 10.58%
$\Delta\phi$	0.62% − 18.03%	0.34% − 5.36%	6.53% − 7.81%	0.92% − 2.07%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	4.38% − 26.59%	0.18% − 1.97%	5.31% − 8.69%	0.29% − 0.88%
$\Delta\phi$, $p_T >$ 600GeV	3.12% − 48.41%	0.14% − 11.06%	6.37% − 11.58%	1.73% − 3.95%
$\Delta\phi$ (back to back)	0.94% − 7.83%	0.32% − 1.09%	6.44% − 7.58%	0.89% − 1.33%
Normalized Measurements				
p_T leading jet	3.92% − 18.87%	0.52% − 3.07%	0.16% − 5.67%	0.05% − 5.10%
p_T leading jet $\eta < 0.5$	3.19% − 55.70%	0.32% − 2.18%	0.30% − 0.89%	0.27% − 17.04%
p_T leading jet $0.5 < \eta < 2.4$	0.87% − 18.19%	0.48% − 2.35%	0.50% − 1.82%	0.08% − 6.64%
p_T subleading jet	2.57%31.13%	0.03% − 4.08%	0.10% − 6.47%	0.21% − 5.61%
p_T subleading jet $\eta < 0.5$	3.27% − 38.48%	0.29% − 7.70%	0.29% − 9.31%	0.29% − 6.41%
p_T subleading jet $0.5 < \eta < 2.4$	0.29% − 34.84%	0.47% − 14.11%	0.31% − 16.78%	0.19% − 9.01%
$\Delta\phi$	0.51% − 18.08%	0.07% − 4.71%	0.11% − 2.21%	0.01% − 0.87%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	3.20% − 25.16%	0.07% − 1.94%	0.21% − 2.30%	$<$ 0.41%
$\Delta\phi$, $p_T >$ 600GeV	2.40% − 47.30%	0.12% − 10.83%	0.17% − 6.81%	0.06% − 1.56%
$\Delta\phi$ (back to back)	0.28% − 7.22%	0.06% − 0.69%	0.11% − 0.92%	0.01% − 0.32%

Table H.2: Breakdown of the experimental uncertainty for the measurements in which the QCD mulitjet events has been considered as part of the signal.

Observable	JES Unc.	JER Unc	BTag Unc.
Cross Section Measurements			
p_T leading jet	2.45% − 14.79%	0.08% − 1.45%	2.04% − 6.47%
p_T leading jet $\eta < 0.5$	2.42% − 26.67%	0.20% − 2.28%	2.60% − 6.17%
p_T leading jet $0.5 < \eta < 2.4$	1.18% − 15.92%	0.06% − 1.05%	1.74% − 7.28%
p_T subleading jet	2.27% − 17.39%	0.04% − 0.98%	1.40% − 5.82%
p_T subleading jet $\eta < 0.5$	1.46% − 21.05%	0.12% − 2.17%	1.47% − 6.14%
p_T subleading jet $0.5 < \eta < 2.4$	1.70% − 33.33%	0.06% − 0.62%	1.61% − 6.03%
$\Delta\phi$	$<$ 15.38%	$<$ 3.09%	4.43% − 5.52%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	$<$ 21.05%	0.03% − 2.06%	3.33% − 6.82%
$\Delta\phi$, $p_T >$ 600GeV	$<$ 20.00%	0.03% − 5.57%	1.48% − 6.31%
$\Delta\phi$ (back to back)	0.59% − 7.63%	0.02% − 0.28%	4.27% − 5.15%
Normalized Measurements			
p_T leading jet	1.41% − 11.95%	0.07% − 1.46%	0.18% − 7.14%
p_T leading jet $\eta < 0.5$	1.43% − 22.80%	0.09% − 2.35%	0.21% − 6.95%
p_T leading jet $0.5 < \eta < 2.4$	0.88% − 13.94%	0.02% − 1.01%	0.11% − 7.10%
p_T subleading jet	1.95% − 14.13%	0.03% − 0.97%	0.38% − 6.72%
p_T subleading jet $\eta < 0.5$	0.95% − 19.01%	0.10% − 2.19%	1.23% − 5.40%
p_T subleading jet $0.5 < \eta < 2.4$	1.59% − 27.67%	0.06% − 0.61%	0.28% − 7.44%
$\Delta\phi$	0.62% − 12.07%	0.04% − 3.13%	0.17% − 0.83%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	1.50% − 20.71%	0.03% − 2.08%	0.14% − 2.83%
$\Delta\phi$, $p_T >$ 600GeV	1.81% − 12.79%	0.02% − 5.61%	0.08% − 3.56%
$\Delta\phi$ (back to back)	0.43% − 5.94%	0.03% − 0.24%	0.11% − 0.52%

Table H.3: Model Uncertainties in the $t\bar{t}$ measurements.

Observable	Parton Shower Unc.	Hard Scattering Unc.	t-quark mass Unc.
Cross Section Measurements			
p_T leading jet	0.97% − 9.82%	0.84% − 5.63%	0.23% − 2.12%
p_T leading jet $\eta < 0.5$	1.66% − 11.80%	1.26% − 6.67%	0.18% − 4.09%
p_T leading jet $0.5 < \eta < 2.4$	1.23% − 11.22%	0.83% − 7.33%	< 2.35%
p_T subleading jet	0.74% − 21.36%	0.87% − 7.43%	0.03% − 4.26%
p_T subleading jet $\eta < 0.5$	0.87% − 21.26%	1.14% − 8.05%	0.24% − 7.44%
p_T subleading jet $0.5 < \eta < 2.4$	1.51% − 17.84%	0.49% − 12.57%	0.19% − 1.80%
$\Delta\phi$	0.59% − 18.56%	1.71% − 15.93%	0.15% − 9.05%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	0.57% − 28.61%	1.09% − 18.66%	0.01% − 10.82%
$\Delta\phi$, $p_T >$ 600GeV	1.15% − 20.09%	1.76% − 20.55%	0.04% − 2.25%
$\Delta\phi$ (back to back)	1.18% − 3.43%	1.41% − 3.45%	0.27% − 2.10%
Normalized Measurements			
p_T leading jet	0.97% − 9.81%	2.39% − 17.89%	0.23% − 2.12%
p_T leading jet $\eta < 0.5$	1.66% − 11.80%	1.61% − 30.48%	0.18% − 4.09%
p_T leading jet $0.5 < \eta < 2.4$	1.23% − 10.99%	3.63% − 17.86%	< 2.35%
p_T subleading jet	0.73% − 21.27%	1.36% − 15.54%	0.03% − 4.26%
p_T subleading jet $\eta < 0.5$	0.87% − 20.88%	1.81% − 25.15%	0.24% − 7.44%
p_T subleading jet $0.5 < \eta < 2.4$	1.51% − 17.82%	0.90% − 30.88%	0.19% − 1.80%
$\Delta\phi$	0.59% − 18.56%	1.76% − 41.56%	0.15% − 9.05%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	0.57% − 28.55%	1.01% − 74.23%	0.01% − 10.82%
$\Delta\phi$, $p_T >$ 600GeV	1.15% − 20.09%	3.50% − 121.97%	0.04% − 2.25%
$\Delta\phi$ (back to back)	1.18% − 3.42%	1.38% − 10.34%	0.27% − 2.10%

Table H.4: Model Uncertainties in the measurements where the QCD multijet events are considered as part of the signal.

Observable	Parton Shower Unc.	Hard Scattering Unc.	t-mass Unc.
Cross Section Measurements			
p_T leading jet	0.35% − 3.65%	1.05% − 5.46%	0.10% − 0.86%
p_T leading jet $\eta < 0.5$	0.67% − 4.75%	1.43% − 5.30%	< 1.52%
p_T leading jet $0.5 < \eta < 2.4$	0.56%4.22%−	1.17% − 7.05%	< 1%
p_T subleading jet	0.59% − 7.69%	1.05% − 7.24%	0.01% − 1.67%
p_T subleading jet $\eta < 0.5$	0.51% − 8.52%	1.23% − 8.07%	0.11% − 1.80%
p_T subleading jet $0.5 < \eta < 2.4$	0.61% − 4.98%	0.83% − 11.44%	0.09% − 0.78%
$\Delta\phi$	0.42% − 6.05%	2.17% − 16.02%	0.07% − 2.95%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	0.36% − 15.15%	1.53% − 18.67%	0.00% − 4.65%
$\Delta\phi$, $p_T >$ 600GeV	0.43% − 5.40%	1.80% − 19.93%	0.02% − 0.95%
$\Delta\phi$ (back to back)	0.68% − 1.47%	1.36% − 4.01%	0.13% − 0.96%
Normalized Measurements			
p_T leading jet	1.08% − 4.31%	2.24% − 10.97%	0.00% − 2.96%
p_T leading jet $\eta < 0.5$	0.93% − 4.72%	3.50% − 10.83%	0.28% − 2.16%
p_T leading jet $0.5 < \eta < 2.4$	0.75% − 6.97%	2.60% − 17.78%	0.06% − 2.36%
p_T subleading jet	1.29% − 6.49%	1.85% − 23.82%	0.25% − 1.66%
p_T subleading jet $\eta < 0.5$	0.65% − 5.57%	2.18% − 29.00%	0.05% − 2.55%
p_T subleading jet $0.5 < \eta < 2.4$	1.95% − 7.47%	0.94% − 23.70%	0.11% − 1.12%
$\Delta\phi$	0.41% − 12.70%	2.37% − 45.01%	0.03% − 1.67%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	0.36% − 22.40%	1.62% − 70.33%	0.09% − 5.17%
$\Delta\phi$, $p_T >$ 600GeV	1.18% − 4.91%	4.42% − 110.42%	0.01% − 4.65%
$\Delta\phi$ (back to back)	0.84% − 1.57%	1.77% − 8.61%	0.03% − 0.95%

Table H.5: Breakdown of the total uncertainty for the $t\bar{t}$ measurements

Observable	Modeling Unc [a]	Exp. Unc. [b]	Syst Unc.	Stat Unc	Total Unc
CROSS SECTION MEASUREMENTS					
p_T `leading jet`	1.43% – 11.47%	8.99% – 23.46%	9.10% – 24.36%	8.01% – 43.64%	12.12% – 49.98%
p_T `leading jet` $\eta < 0.5$	2.50% – 13.75%	8.32% – 63.00%	9.36% – 63.63%	12.73% – 145.60%	16.97% – 158.90%
p_T `leading jet` $0.5 < \eta < 2.4$	2.40% – 12.19%	8.12% – 27.83%	8.46% – 28.86%	10.38% – 60.17%	13.40% – 66.73%
p_T `subleading jet`	1.50% – 22.72%	8.16% – 34.83%	8.69% – 35.94%	6.16% – 55.85%	10.76% – 66.42%
p_T `subleading jet` $\eta < 0.5$	2.24% – 21.58%	8.55% – 42.78%	8.92% – 44.28%	10.83% – 84.94%	14.49% – 95.79%
p_T `subleading jet` $0.5 < \eta < 2.4$	2.23% – 21.83%	8.35% – 42.12%	8.83% – 47.44%	7.62% – 88.04%	11.67% – 95.74%
$\Delta\phi$	1.84% – 26.08%	7.63% – 20.45%	8.06% – 33.14%	4.14% – 58.44%	9.06% – 67.18%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	1.23% – 35.09%	8.10% – 24.18%	8.81% – 42.18%	5.55% – 61.89%	10.41% – 74.89%
$\Delta\phi$, $p_T >$ 600GeV	2.69% – 28.74%	7.88% – 30.17%	8.36% – 41.67%	6.02% – 110.64%	10.70% – 118.23%
$\Delta\phi$ `(back to back)`	1.91% – 4.54%	7.42% – 10.54%	7.80% – 11.29%	8.35% – 21.46%	11.98% – 24.25%
NORMALIZED MEASUREMENTS					
p_T `leading jet`	3.58% – 20.49%	3.98% – 19.43%	5.81% – 22.99%	8.01% – 43.64%	10.57% – 48.63%
p_T `leading jet` $\eta < 0.5$	2.85% – 32.94%	3.36% – 59.52%	5.65% – 60.78%	12.73% – 145.60%	14.75% – 157.78%
p_T `leading jet` $0.5 < \eta < 2.4$	5.68% – 18.08%	2.83% – 22.68%	8.10% – 24.87%	10.38% – 60.17%	14.54% – 65.08%
p_T `subleading jet`	2.31% – 26.42%	2.77% – 31.45%	4.60% – 36.05%	6.16% – 55.85%	7.69% – 66.48%
p_T `subleading jet` $\eta < 0.5$	2.66% – 27.05%	3.82% – 40.85%	4.65% – 48.71%	10.83% – 84.94%	12.46% – 97.92%
p_T `subleading jet` $0.5 < \eta < 2.4$	3.19% – 32.60%	2.97% – 37.87%	4.36% – 49.89%	7.62% – 88.04%	8.95% – 96.93%
$\Delta\phi$	2.58% – 46.41%	0.53% – 16.10%	2.63% – 49.12%	4.14% – 58.44%	4.90% – 76.34%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	1.16% – 79.94%	3.21% – 23.61%	3.41% – 83.35%	5.55% – 61.89%	6.52% – 103.82%
$\Delta\phi$, $p_T >$ 600GeV	3.70% – 123.61%	2.20% – 22.14%	4.41% – 125.58%	6.02% – 110.64%	7.47% – 167.37%
$\Delta\phi$ `(back to back)`	2.00% – 10.57%	0.38% – 6.35%	2.04% – 12.33%	8.35% – 21.46%	9.41% – 24.75%

[a] PS. Unc $\otimes$ HS. Unc $\otimes$ t-quark mass. Unc

[b] JEC. Unc $\otimes$ JER. Unc $\otimes$ Bckg. Unc $\otimes$ BTag Unc.$\otimes$ Lumi Unc

Table H.6: Breakdown of the total uncertainty for the measurements in which the QCD mulitjet events has been considered as part of the signal.

Observable	Modeling Unc [a]	Exp. Unc.[b]	Syst Unc.	Stat Unc	Total Unc
CROSS SECTION MEASUREMENTS					
p_T leading jet	1.21%6.60%	6.28% – 15.13%	6.92% – 15.92%	7.12% – 35.12%	9.94% – 37.96%
p_T leading jet $\eta < 0.5$	2.25% – 7.27%	5.42% – 26.91%	6.47% – 27.37%	10.86% – 65.45%	13.54% – 70.94%
p_T leading jet $0.5 < \eta < 2.4$	1.70% – 7.13%	5.12% – 16.23%	6.66% – 17.38%	9.39% – 50.36%	11.52% – 51.50%
p_T subleading jet	1.38% – 10.58%	5.44% – 17.62%	6.00% – 18.17%	5.81% – 41.63%	9.21% – 45.42%
p_T subleading jet $\eta < 0.5$	1.88% – 9.20%	5.89% – 21.25%	6.22% – 22.83%	10.05% – 56.99%	12.07% – 61.39%
p_T subleading jet $0.5 < \eta < 2.4$	1.75% – 12.48%	3.87% – 33.57%	6.76% – 35.81%	7.09% – 59.43%	10.22% – 61.78%
$\Delta\phi$	2.30% – 17.37%	5.13% – 16.48%	5.63% – 23.95%	3.70% – 49.86%	6.73% – 55.32%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	1.58% – 24.24%	5.89% – 21.46%	6.18% – 32.37%	5.13% – 67.24%	8.03% – 74.63%
$\Delta\phi$, $p_T >$ 600GeV	2.10% – 20.64%	5.73% – 21.84%	6.10% – 30.05%	5.36% – 74.13%	8.38% – 79.99%
$\Delta\phi$ (back to back)	1.54% – 4.14%	5.00% – 9.54%	5.26% – 10.39%	7.69% – 18.80%	9.42% – 21.48%
NORMALIZED MEASUREMENTS					
p_T leading jet	2.54% – 12.15%	1.47% – 12.26%	4.76% – 16.33%	7.12% – 35.12%	9.30% – 38.73%
p_T leading jet $\eta < 0.5$	4.71% – 10.98%	1.59% – 22.90%	6.41% – 25.40%	10.86% – 65.45%	13.48% – 70.20%
p_T leading jet $0.5 < \eta < 2.4$	4.42% – 19.25%	1.01% – 14.25%	6.14% – 21.35%	9.39% – 50.36%	12.36% – 54.70%
p_T subleading jet	2.47% – 24.69%	2.08% – 14.55%	3.22% – 28.66%	5.81% – 41.63%	9.88% – 50.54%
p_T subleading jet $\eta < 0.5$	3.06% – 29.49%	2.21% – 19.26%	3.78% – 33.92%	10.05% – 56.99%	12.29% – 66.32%
p_T subleading jet $0.5 < \eta < 2.4$	2.56% – 23.91%	2.02% – 28.65%	5.85% – 31.21%	7.09% – 59.43%	11.16% – 64.98%
$\Delta\phi$	2.91% – 46.78%	0.70% – 12.50%	3.46% – 48.42%	3.70% – 49.86%	5.07% – 69.51%
$\Delta\phi$, 400GeV $< p_T <$ 600GeV	1.66% – 73.99%	2.60% – 20.91%	3.11% – 76.89%	5.13% – 67.24%	5.99% – 102.14%
$\Delta\phi$, $p_T >$ 600GeV	4.61% – 110.63%	1.82% – 14.09%	5.56% – 111.52%	5.36% – 74.13%	7.79% – 133.91%
$\Delta\phi$ (back to back)	1.99% – 8.67%	0.60% – 5.96%	2.08% – 10.52%	7.69% – 18.80%	8.10% – 21.54%

[a] PS. Unc $\otimes$ HS. Unc $\otimes$ t-quark mass. Unc

[b] JEC. Unc $\otimes$ JER. Unc $\otimes$ BTag Unc.$\otimes$ Lumi Unc

The following Figures illustrate the total systematic uncertainty when exclusive phase space regions are considered

- ✓ H.1 Systematical and statistical uncertainty corresponding to the distributions differential in $\Delta\phi$ in two p_T^{lead} regions: 400GeV$< p_T <$ 600GeV , $p_T >$ 600GeV.
- ✓ H.2 Systematic and statistical uncertainty corresponding to the distributions differential in p_T of the leading jet in two different η regions: $|\eta| < 0.5$ and $0.5 < |\eta| < 2.4$.
- ✓ H.3 Systematic and statistical uncertainty corresponding to the distributions differential in p_T of the subleading jet in two different η regions: $|\eta| < 0.5$ and $0.5 < |\eta| < 2.4$.

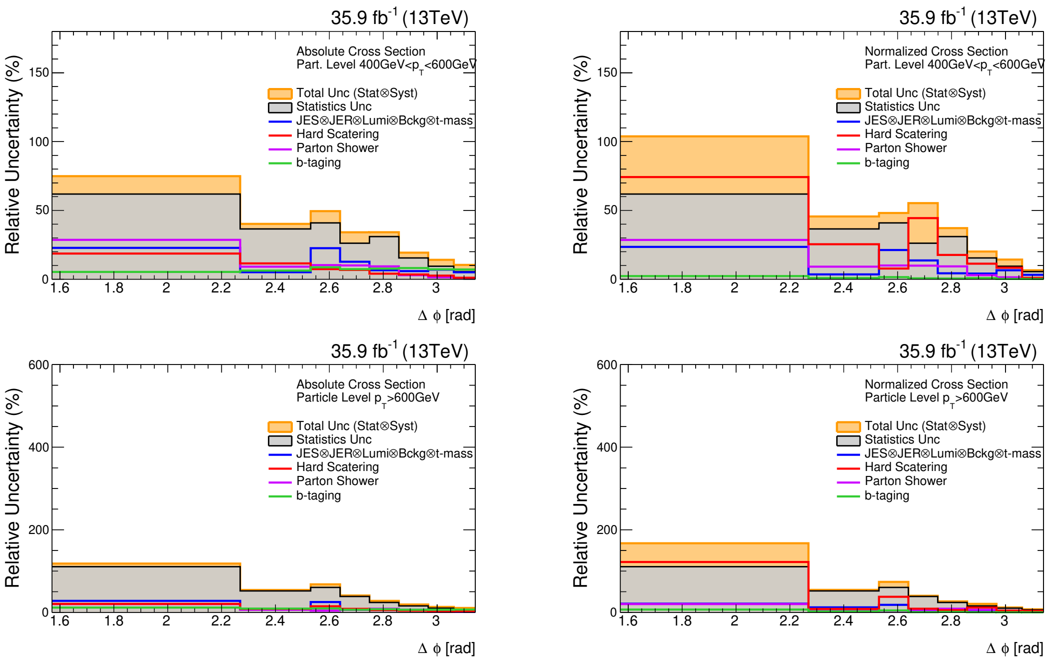


Figure H.1: Systematic and statistical uncertainty corresponding to the distributions differential in $\Delta\phi$ in two p_T^{lead} regions: (upper plots) 400GeV$< p_T <$ 600GeV , (down plots) $p_T >$ 600GeV. The systematic uncertainty corresponding to the absolute and normalized distributions are illustrated.

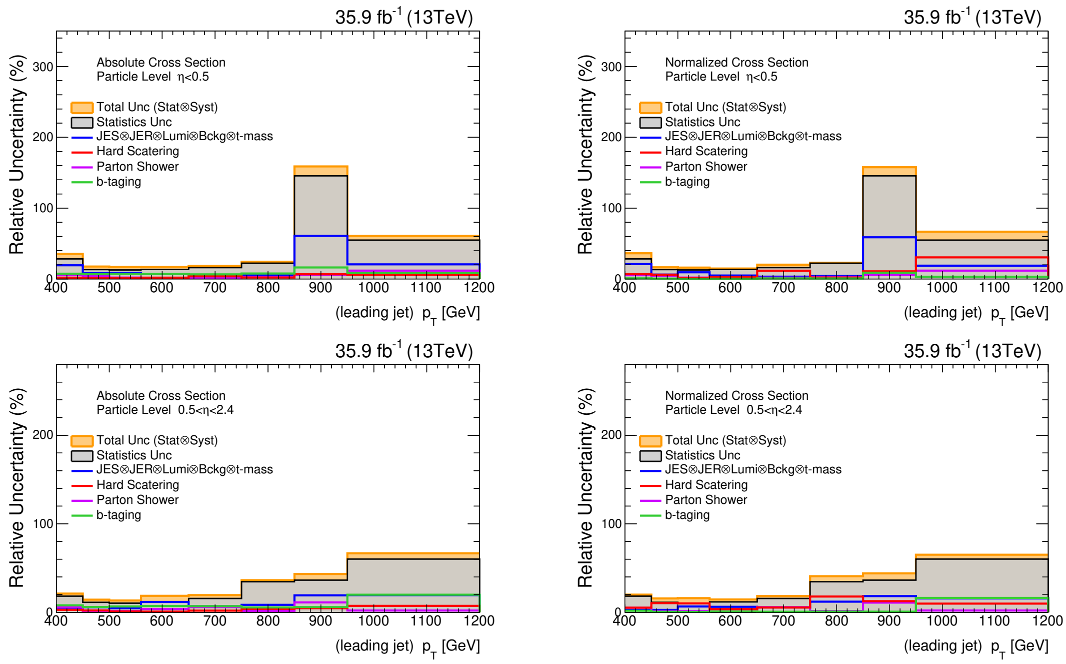


Figure H.2: Systematic and statistical uncertainty corresponding to the distributions differential in p_T of the leading jet in two different η regions: $|\eta| < 0.5$ (upper plots) and $0.5 < |\eta| < 2.4$ (lower plots). The systematic uncertainty corresponding to the absolute (left) and normalized (right) distributions are illustrated.

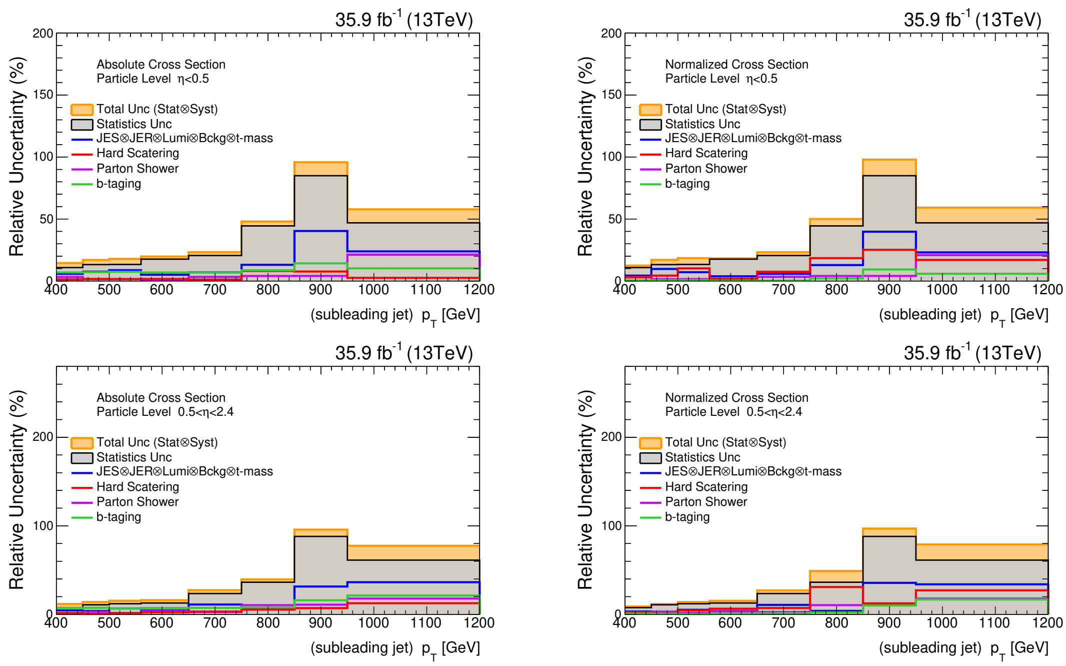


Figure H.3: Systematic and statistical uncertainty corresponding to the distributions differential in p_T of the subleading jet in two different η regions: $|\eta| < 0.5$ (upper plots) and $0.5 < |\eta| < 2.4$ (lower plots). The systematic uncertainty corresponding to the absolute (left) and normalized (right) distributions are illustrated.

Appendix I

Results. Additional material

In this Chapter, additional material to Chapter 11 are provided. The plots included in this additional material correspond to the inclusive top-jet measurements differential in p_T of the two leading jets and their azimuthal angle separation $\Delta\phi$. In some of the provided plots, exclusive phase space regions of the fiducial phase space are considered.

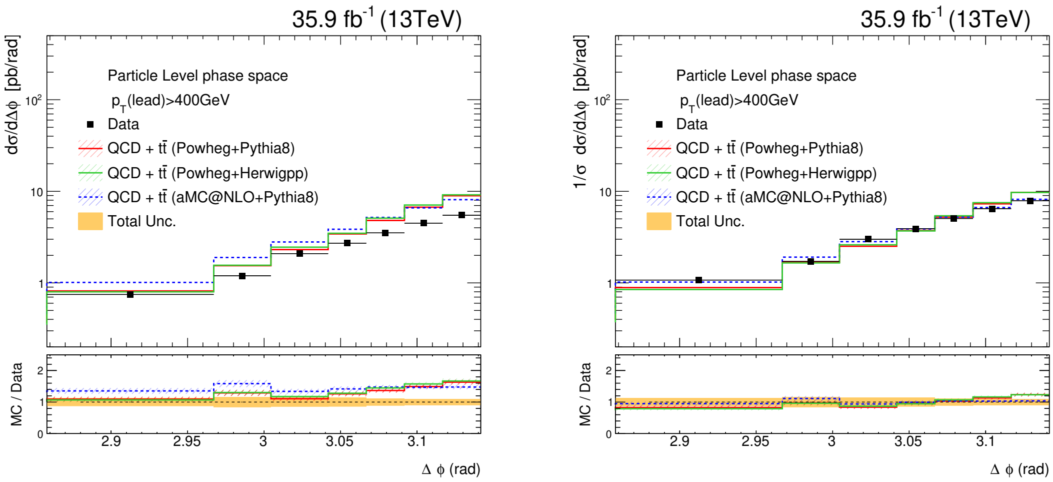


Figure I.1: Cross section (left plots) and normalized distributions (right plots) differentially in $\Delta\phi$, when the fine binnig has been considered in the most back-to-back region. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The distributions correspond to the inclusive top-jet cross section, and considering the fiducial phase space

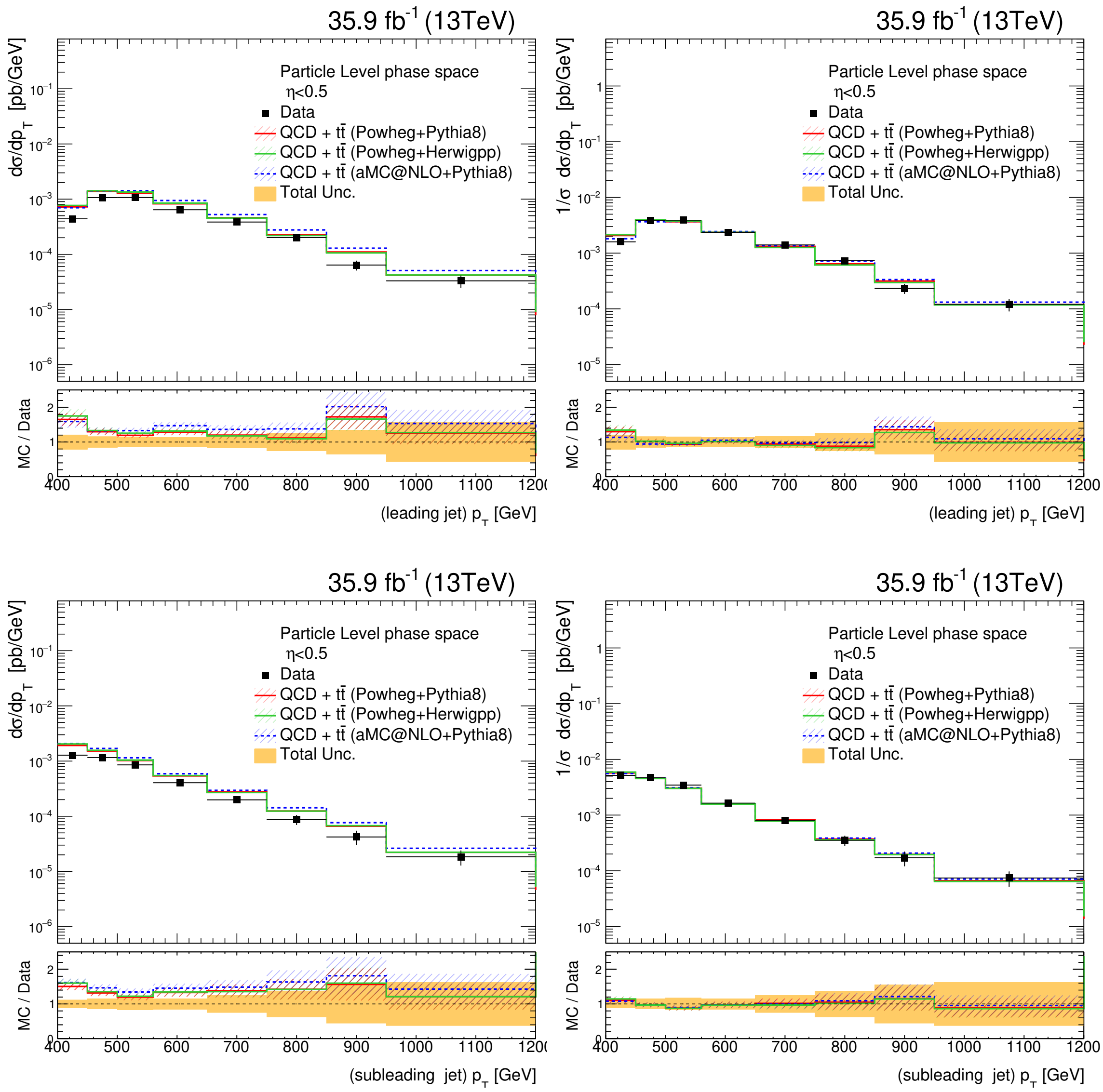


Figure I.2: Cross section (left plots) and normalized distributions (right plots) differentially in (upper plots) the p_T of the leading jet, (lower plots) the p_T of the subleading jet, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The distributions correspond to the inclusive top-jet cross section, when the exclusive region of the phase space: $|\eta| < 0.5$ is considered.

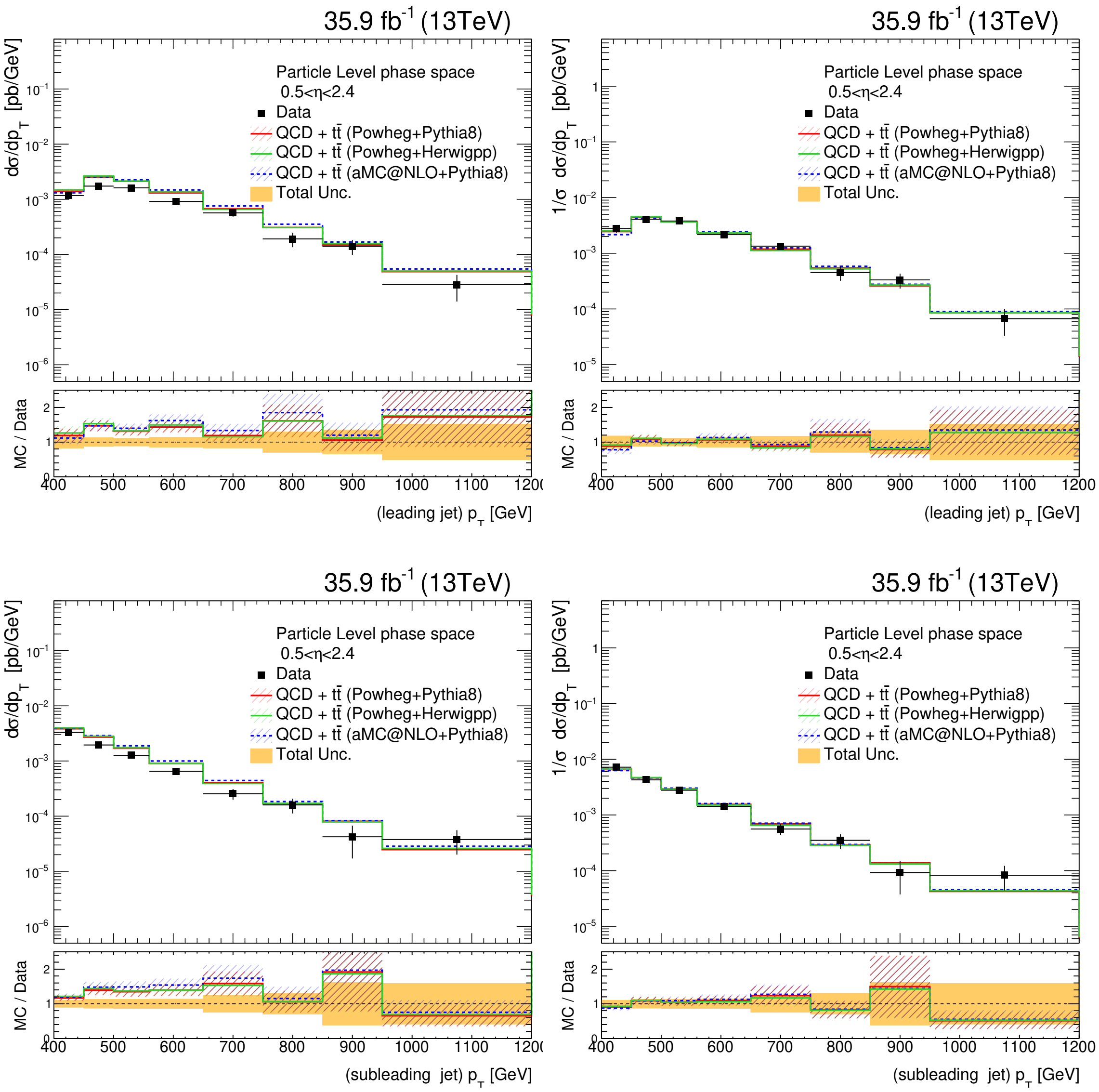


Figure I.3: Cross section (left plots) and normalized distributions (right plots) differentially in (upper plots) the p_T of the leading jet, (lower plots) the p_T of the subleading jet, compared to theory predictions. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The distributions correspond to the inclusive top-jet cross section, when the exclusive region of the phase space: $0.5 < |\eta| < 2.4$ is considered.

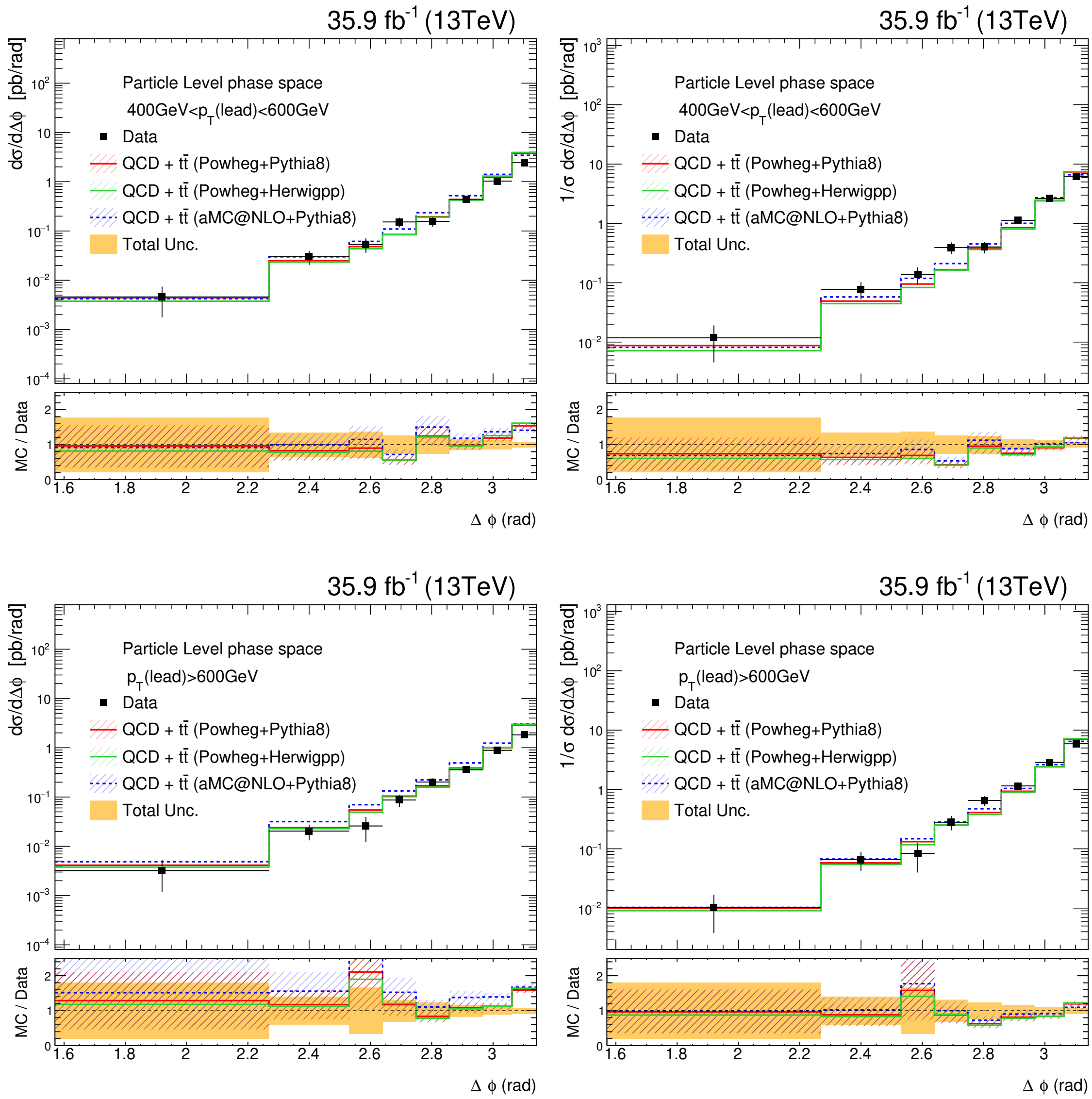


Figure I.4: Cross section (left plots) and normalized distributions (right plots) differentially in $\Delta\phi$. The total uncertainty are shown for each measurements in the lower panels, while in the upper panels only the statistical uncertainties are shown. The shown uncertainty in the theory predictions are the statistical uncertainties. The distributions correspond to the inclusive top-jet cross section, when the exclusive region of the phase space are considered: (upper plots) 400 GeV $p_T^{lead} < 600$ GeV, (down plots) $p_T^{lead} > 600$ GeV.

Bibliography

[1] Michael H. Seymour. "Searches for new particles using cone and cluster jet algorithms: a comparative study". In: *Zeitschrift für Physik C Particles and Fields* 62.1 (1994), pp. 127–138. ISSN: 1431-5858. DOI: 10.1007/BF01559532. URL: https://doi.org/10.1007/BF01559532.

[2] CMS Collaboration. "High-p_T jet measurements at the HL-LHC". In: (2018).

[3] P. Azzi et al. "Standard Model Physics at the HL-LHC and HE-LHC". In: (2019). arXiv: 1902.04070 [hep-ph].

[4] Chen-Ning Yang and Robert L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1954). [,150(1954)], pp. 191–195. DOI: 10.1103/PhysRev.96.191.

[5] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964). [,160(1964)], pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

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

[7] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[8] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[9] *CMS Expreriment, LHC*. URL: https://cms.cern/.

[10] *The Atlas Collaboration*. URL: https://atlas.web.cern.ch/Atlas/.

[11] Stefan Weinzierl. "Introduction to Monte Carlo methods". In: *Springer, New York, NY* (2000). DOI: 10.1007/978-0-387-87837-9_1. arXiv: hep-ph/0006269 [hep-ph].

[12] *Latex Templates for Physics Models*. URL: http://www.texample.net/tikz/examples/model-physics/.

[13] Paul A. M. Dirac. "On the Theory of quantum mechanics". In: *Proc. Roy. Soc. Lond.* A112 (1926), pp. 661–677. DOI: 10.1098/rspa.1926.0133.

[14] W. Pauli. "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren". In: *Zeitschrift für Physik* 31.1 (1925), pp. 765–783. ISSN: 0044-3328. DOI: 10.1007/BF02980631. URL: https://doi.org/10.1007/BF02980631.

[15] B. Pontecorvo. "Neutrino Experiments and the Problem of Conservation of Leptonic Charge". In: *Sov. Phys. JETP* 26 (1968). [Zh. Eksp. Teor. Fiz.53,1717(1967)], pp. 984–988.

[16] Alan R. White. "Electroweak high-energy scattering and the chiral anomaly". In: *Phys. Rev.* D69 (2004), p. 096002. DOI: 10.1103/PhysRevD.69.096002. arXiv: hep-ph/0308287 [hep-ph].

[17] Bose. "Plancks Gesetz und Lichtquantenhypothese". In: *Zeitschrift für Physik* 26.1 (1924), pp. 178–181. ISSN: 0044-3328. DOI: 10.1007/BF01327326. URL: https://doi.org/10.1007/BF01327326.

[18] Vardan Khachatryan et al. "Measurement of the inclusive 3-jet production differential cross section in protonâĂŞproton collisions at 7 TeV and determination of the strong coupling constant in the TeV range". In: *Eur. Phys. J.* C75.5 (2015), p. 186. DOI: 10.1140/epjc/s10052-015-3376-y. arXiv: 1412.1633 [hep-ex].

[19] T. N. Pham. “CKM Matrix Elements”. In: (2011). arXiv: 1110.6050 [hep-ph].

[20] A. Ceccucci, Z. Ligeti, and Y. Sakai. “The CKM quark-mixing matrix”. In: *Phys.Lett.* B667:1 (2008), pp. 145–152.

[21] John C. Collins, Davison E. Soper, and George F. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313 [hep-ph].

[22] Guido Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. In: *Nucl. Phys.* B126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[23] Jun Gao, Lucian Harland-Lang, and Juan Rojo. “The Structure of the Proton in the LHC Precision Era”. In: *Phys. Rept.* 742 (2018), pp. 1–121. DOI: 10.1016/j.physrep.2018.03.002. arXiv: 1709.04922 [hep-ph].

[24] Vardan Khachatryan et al. “Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV”. In: *JHEP* 03 (2017), p. 156. DOI: 10.1007/JHEP03(2017)156. arXiv: 1609.05331 [hep-ex].

[25] J. G. Morfín and W. K. Tung. “Parton distributions from a global QCD analysis of deep inelastic scattering and lepton-pair production”. In: *Zeitschrift für Physik C Particles and Fields* 52.1 (1991), pp. 13–29. ISSN: 1431-5858. DOI: 10.1007/BF01412323. URL: https://doi.org/10.1007/BF01412323.

[26] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Phys. Rev.* D93.3 (2016), p. 033006. DOI: 10.1103/PhysRevD.93.033006. arXiv: 1506.07443 [hep-ph].

[27] F. D. Aaron et al. “Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA”. In: *JHEP* 01 (2010), p. 109. DOI: 10.1007/JHEP01(2010)109. arXiv: 0911.0884 [hep-ex].

[28] H. Abramowicz et al. “Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data”. In: *Eur. Phys. J.* C75.12 (2015), p. 580. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[29] Stefano Forte et al. “Neural network parametrization of deep inelastic structure functions”. In: *JHEP* 05 (2002), p. 062. DOI: 10.1088/1126-6708/2002/05/062. arXiv: hep-ph/0204232 [hep-ph].

[30] I. Abt et al. “The H1 detector at HERA”. In: *Nucl. Instrum. Meth.* A386 (1997), pp. 310–347. DOI: 10.1016/S0168-9002(96)00893-5.

[31] *The ZEUS detector: Status report 1993*. Tech. rep. ZEUS-STATUS-REPT-1993. 1993, p. 597. DOI: 10.3204/PUBDB-2017-12635. URL: http://bib-pubdb1.desy.de/record/396236.

[32] M. J. G. Veltman. “Diagrammatica: The Path to Feynman rules”. In: *Cambridge Lect. Notes Phys.* 4 (1994), pp. 1–284.

[33] *Monte Carlo Simulations*. URL: https://theory.slac.stanford.edu/our-research/simulations.

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

[35] Serguei Chatrchyan et al. “Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 7$ TeV”. In: *Phys. Lett.* B722 (2013), pp. 5–27. DOI: 10.1016/j.physletb.2013.03.024. arXiv: 1210.6718 [hep-ex].

[36] C. Marquet and Robert B. Peschanski. “Testing saturation at hadron colliders”. In: *Phys. Lett.* B587 (2004), pp. 201–210. DOI: 10.1016/j.physletb.2004.03.005. arXiv: hep-ph/0312261 [hep-ph].

[37] Vardan Khachatryan et al. “Event generator tunes obtained from underlying event and multiparton scattering measurements”. In: *Eur. Phys. J.* C76.3 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

[38] Peter Skands, Stefano Carrazza, and Juan Rojo. "Tuning PYTHIA 8.1: the Monash 2013 Tune". In: *Eur. Phys. J.* C74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].

[39] A. M. Sirunyan et al. "Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 06 (2018), p. 002. DOI: 10.1007/JHEP06(2018)002. arXiv: 1803.03991 [hep-ex].

[40] CMS Collaboration. "Measurement of the top quark mass with lepton+jets final states in pp collisions at $\sqrt{s} = 13$ TeV". In: (2017).

[41] Bo Andersson et al. "Parton Fragmentation and String Dynamics". In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[42] S. Gieseke et al. "Herwig++ 2.5 Release Note". In: (2011). arXiv: 1102.1672 [hep-ph].

[43] M. Bahr et al. "Herwig++ Physics and Manual". In: *Eur. Phys. J.* C58 (2008), pp. 639–707. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[44] John C. Collins. "Sudakov form-factors". In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 573–614. DOI: 10.1142/9789814503266_0006. arXiv: hep-ph/0312336 [hep-ph].

[45] Stefan Hoeche et al. "Matching parton showers and matrix elements". In: *HERA and the LHC: A Workshop on the implications of HERA for LHC physics*. 2005, pp. 288–289. DOI: 10.5170/CERN-2005-014.288. arXiv: hep-ph/0602031 [hep-ph].

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

[47] Paolo Torrielli and Stefano Frixione. "Matching NLO QCD computations with PYTHIA using MC@NLO". In: *JHEP* 04 (2010), p. 110. DOI: 10.1007/JHEP04(2010)110. arXiv: 1002.4293 [hep-ph].

[48] F. Abe et al. "Observation of top quark production in $\bar{p}p$ collisions". In: *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002 [hep-ex].

[49] S. Abachi et al. "Observation of the top quark". In: *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. arXiv: hep-ex/9503003 [hep-ex].

[50] *CDF Collaboration, Fermilab*. URL: https://www-cdf.fnal.gov/.

[51] *D0 collaboration, Tevatron, Fermilab*. URL: https://www-d0.fnal.gov/.

[52] T. Aaltonen et al. "Combination of the top-quark mass measurements from the Tevatron collider". In: *Phys. Rev.* D86 (2012), p. 092003. DOI: 10.1103/PhysRevD.86.092003. arXiv: 1207.1069 [hep-ex].

[53] Michal Czakon, Paul Fiedler, and Alexander Mitov. "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\frac{4}{S})$". In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254 [hep-ph].

[54] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys.* G43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

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

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

[57] A. M. Sirunyan et al. "Measurement of the inclusive $\mathrm{t\bar{t}}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton". In: *JHEP* 03 (2018), p. 115. DOI: 10.1007/JHEP03(2018)115. arXiv: 1711.03143 [hep-ex].

[58] Morad Aaboud et al. "Measurements of top-quark pair differential cross-sections in the $e\mu$ channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector". In: *Eur. Phys. J.* C77.5 (2017), p. 292. DOI: 10.1140/epjc/s10052-017-4821-x. arXiv: 1612.05220 [hep-ex].

[59] Vardan Khachatryan et al. "Measurement of the $t\bar{t}$ production cross section using events in the eμ final state in pp collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J.* C77 (2017), p. 172. DOI: 10.1140/epjc/s10052-017-4718-8. arXiv: 1611.04040 [hep-ex].

[60] The ATLAS collaboration. "Measurement of the inclusive $t\bar{t}$ cross-section in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: (2017).

[61] FrÃl'dÃl'ric Derue. "Latest ATLAS measurements of top quark properties using tt̄ events". In: *10th International Workshop on Top Quark Physics (TOP2017) Braga, Portugal, September 17-22, 2017*. 2017. arXiv: 1711.08778 [hep-ex].

[62] Morad Aaboud et al. "Combination of inclusive and differential tt̄ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 04 (2018), p. 033. DOI: 10.1007/JHEP04(2018)033. arXiv: 1709.05327 [hep-ex].

[63] Michaa L. Czakon et al. "Top quark pair production at NNLO+NNLL′ in QCD combined with electroweak corrections". In: *11th International Workshop on Top Quark Physics (TOP2018) Bad Neuenahr, Germany, September 16-21, 2018*. 2019. arXiv: 1901.08281 [hep-ph].

[64] Andrea Giammanco. "TOP2017: Experimental Summary". In: *10th International Workshop on Top Quark Physics (TOP2017) Braga, Portugal, September 17-22, 2017*. 2017. arXiv: 1712.02177 [hep-ex].

[65] P. Kant et al. "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions". In: *Comput. Phys. Commun.* 191 (2015), pp. 74–89. DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403 [hep-ph].

[66] "Top Feyman diagrams". In: (). URL: https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html.

[67] Andrea Giammanco and Reinhard Schwienhorst. "Single top-quark production at the Tevatron and the LHC". In: *Rev. Mod. Phys.* 90.3 (2018), p. 035001. DOI: 10.1103/RevModPhys.90.035001. arXiv: 1710.10699 [hep-ex].

[68] *Physics Results TOP Summary Figures, CMS TWiki*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryFigures.

[69] Stefano Catani, Massimiliano Grazzini, and Hayk Sargsyan. "Transverse-momentum resummation for top-quark pair production at the LHC". In: *JHEP* 11 (2018), p. 061. DOI: 10.1007/JHEP11(2018)061. arXiv: 1806.01601 [hep-ph].

[70] Stefano Catani, Daniel de Florian, and Massimiliano Grazzini. "Universality of nonleading logarithmic contributions in transverse momentum distributions". In: *Nucl. Phys.* B596 (2001), pp. 299–312. DOI: 10.1016/S0550-3213(00)00617-9. arXiv: hep-ph/0008184 [hep-ph].

[71] Michal Czakon et al. "Top-pair production at the LHC through NNLO QCD and NLO EW". In: *JHEP* 10 (2017), p. 186. DOI: 10.1007/JHEP10(2017)186. arXiv: 1705.04105 [hep-ph].

[72] Aneesh V. Manohar and Michael Trott. "Electroweak Sudakov Corrections and the Top Quark Forward-Backward Asymmetry". In: *Phys. Lett.* B711 (2012), pp. 313–316. DOI: 10.1016/j.physletb.2012.04.013. arXiv: 1201.3926 [hep-ph].

[73] Werner Bernreuther and Zong-Guo Si. "Top quark and leptonic charge asymmetries for the Tevatron and LHC". In: *Phys. Rev.* D86 (2012), p. 034026. DOI: 10.1103/PhysRevD.86.034026. arXiv: 1205.6580 [hep-ph].

[74] Michal Czakon, David Heymes, and Alexander Mitov. "High-precision differential predictions for top-quark pairs at the LHC". In: *Phys. Rev. Lett.* 116.8 (2016), p. 082003. DOI: 10.1103/PhysRevLett.116.082003. arXiv: 1511.00549 [hep-ph].

[75] R. Keith Ellis, W. James Stirling, and B. R. Webber. "QCD and collider physics". In: *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 8 (1996), pp. 1–435.

[76] Vardan Khachatryan et al. "Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76\,\text{TeV}$". In: *Eur. Phys. J.* C76.5 (2016), p. 265. DOI: 10.1140/epjc/s10052-016-4083-z. arXiv: 1512.06212 [hep-ex].

[77] Serguei Chatrchyan et al. "Measurement of the ratio of inclusive jet cross sections using the anti-k_T algorithm with radius parameters R=0.5 and 0.7 in pp collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev.* D90.7 (2014), p. 072006. DOI: 10.1103/PhysRevD.90.072006. arXiv: 1406.0324 [hep-ex].

[78] CMS Collaboration. *Measurement of the double-differential inclusive jet cross section at sqrt(s) = 8 TeV*. Tech. rep. CMS-PAS-SMP-14-001. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2060744.

[79] Vardan Khachatryan et al. "Measurement of the double-differential inclusive jet cross section in protonâĂŞproton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J.* C76.8 (2016), p. 451. DOI: 10.1140/epjc/s10052-016-4286-3. arXiv: 1605.04436 [hep-ex].

[80] Albert M Sirunyan et al. "Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J.* C78.7 (2018), p. 566. DOI: 10.1140/epjc/s10052-018-6033-4. arXiv: 1712.05471 [hep-ex].

[81] CMS Collaboration. "Performance of quark/gluon discrimination in 13 TeV data". In: (2016). URL: https://cds.cern.ch/record/2234117.

[82] Jessie Shelton. "Jet Substructure". In: *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012): Boulder, Colorado, June 4-29, 2012*. 2013, pp. 303–340. DOI: 10.1142/9789814525220_0007. arXiv: 1302.0260 [hep-ph].

[83] Mrinal Dasgupta et al. "Towards an understanding of jet substructure". In: *JHEP* 09 (2013), p. 029. DOI: 10.1007/JHEP09(2013)029. arXiv: 1307.0007 [hep-ph].

[84] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "PYTHIA 6.4 Physics and Manual". In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

[85] Andrew J. Larkoski et al. "Soft Drop". In: *JHEP* 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[86] CMS Collaboration. *Measurements of inclusive 2-jet, 3-jet and 4-jet azimuthal correlations in pp collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-SMP-16-014. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257685.

[87] Gavin P. Salam and Gregory Soyez. "A Practical Seedless Infrared-Safe Cone jet algorithm". In: *JHEP* 05 (2007), p. 086. DOI: 10.1088/1126-6708/2007/05/086. arXiv: 0704.0292 [hep-ph].

[88] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev.* D48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[89] Yuri L. Dokshitzer et al. "Better jet clustering algorithms". In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

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

[91] S. Catani et al. "New clustering algorithm for multi - jet cross-sections in e+ e- annihilation". In: *Phys. Lett.* B269 (1991), pp. 432–438. DOI: 10.1016/0370-2693(91)90196-W.

[92] Albert M Sirunyan et al. "Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair". In: *Phys. Rev. Lett.* 120.7 (2018), p. 071802. DOI: 10.1103/PhysRevLett.120.071802. arXiv: 1709.05543 [hep-ex].

[93] Vardan Khachatryan et al. "Identification techniques for highly boosted W bosons that decay into hadrons". In: *JHEP* 12 (2014), p. 017. DOI: 10.1007/JHEP12(2014)017. arXiv: 1410.4227 [hep-ex].

[94] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jet Trimming". In: *JHEP* 02 (2010), p. 084. DOI: 10.1007/JHEP02(2010)084. arXiv: 0912.1342 [hep-ph].
[95] Georges Aad et al. "Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector". In: *JHEP* 09 (2013), p. 076. DOI: 10.1007/JHEP09(2013)076. arXiv: 1306.4945 [hep-ex].
[96] Jonathan M. Butterworth et al. "Jet substructure as a new Higgs search channel at the LHC". In: *AIP Conf. Proc.* 1078.1 (2009), pp. 189–191. DOI: 10.1063/1.3051904. arXiv: 0809.2530 [hep-ph].
[97] Mrinal Dasgupta et al. "Towards an understanding of jet substructure". In: *JHEP* 09 (2013), p. 029. DOI: 10.1007/JHEP09(2013)029. arXiv: 1307.0007 [hep-ph].
[98] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: *JHEP* 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].
[99] Rolf-Dieter Heuer. "The future of the CERN accelerator complex". In: *Proceedings, Symposium on From the PS to the LHC: 50 Years of Nobel Memories in High-Energy Physics: Geneva, Switzerland, December 3-4, 2009*. 2009, pp. 119–127. DOI: 10.1007/978-3-642-30844-4_8.
[100] Thomas Sven Pettersson and P Lefevre. *The Large Hadron Collider: conceptual design*. Tech. rep. CERN-AC-95-05-LHC. 1995. URL: https://cds.cern.ch/record/291782.
[101] S. Myers and E. Picasso. "The Design, construction and commissioning of the CERN Large Electron Positron collider". In: *Contemp. Phys.* 31 (1990), pp. 387–403.
[102] G. L. Bayatian et al. "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". In: (2006). URL: http://cds.cern.ch/record/922757.
[103] G. L. Bayatian et al. "CMS technical design report, volume II: Physics performance". In: *J. Phys.* G34.6 (2007), pp. 995–1579. DOI: 10.1088/0954-3899/34/6/S01.
[104] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.
[105] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.
[106] G. Anelli et al. "The TOTEM experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08007. DOI: 10.1088/1748-0221/3/08/S08007.
[107] O. Adriani et al. "The LHCf detector at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08006. DOI: 10.1088/1748-0221/3/08/S08006.
[108] James Pinfold et al. "Technical Design Report of the MoEDAL Experiment". In: (2009).
[109] Christiane LefÃÍvre. "The CERN accelerator complex. Complexe des accÃl'lÃl'rateurs du CERN". 2008. URL: https://cds.cern.ch/record/1260465.
[110] CMS Collaboration. *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. CMS-PAS-LUM-17-001. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257069.
[111] *CMS Luminosity Public results, CMS TWiki.* URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.
[112] G. Acquistapace et al. "CMS, the magnet project: Technical design report". In: Technical Design Report CMS (1997). URL: https://cds.cern.ch/record/331056.
[113] *Sketch CMS Experiment, CMS TWiki*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS.
[114] Wikipedia. *Pseudorapidity*. URL: https://en.wikipedia.org/wiki/Pseudorapidity.
[115] V Karimaki et al. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/368412.
[116] S Chatrchyan et al. "Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays". In: *JINST* 5 (2010), T03009. DOI: 10.1088/1748-0221/5/03/T03009. arXiv: 0910.2505 [physics.ins-det].
[117] "CMS: The electromagnetic calorimeter. Technical design report". In: Technical Design Report CMS (1997). URL: https://cds.cern.ch/record/349375.

[118] "CMS: The hadron calorimeter technical design report". In: Technical Design Report CMS (1997). URL: http://cds.cern.ch/record/357153.

[119] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[120] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[121] L. Tuura et al. "CMS data quality monitoring: Systems and experiences". In: *J. Phys. Conf. Ser.* 219 (2010), p. 072020. DOI: 10.1088/1742-6596/219/7/072020.

[122] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

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

[124] D. Buskulic et al. "Performance of the ALEPH detector at LEP". In: *Nucl. Instrum. Meth.* A360 (1995), pp. 481–506. DOI: 10.1016/0168-9002(95)00138-7.

[125] Uwe Holm. "The ZEUS detector: Status report 1993". In: (1993), p. 597. DOI: 10.3204/PUBDB-2017-12635.

[126] F. Abe et al. "The CDF Detector: An Overview". In: *Nucl. Instrum. Meth.* A271 (1988), pp. 387–403. DOI: 10.1016/0168-9002(88)90298-7.

[127] T. Atkinson. "Reconstruction of electrons with the Gaussian-Sum filter in the ATLAS inner detector". In: *Proceedings, 2005 IEEE Nuclear Science Symposium and Medical Imaging Conference: El Conquistador Resort, Puerto Rico, October 23-29, 2005*. 1. 2005, pp. 148–152. DOI: 10.1109/NSSMIC.2005.1596226.

[128] W. Adam et al. "Track reconstruction in the CMS tracker". In: CERN-CMS-NOTE-2006-041, CMS-NOTE-2006-041 (2005). URL: https://cds.cern.ch/record/934067.

[129] Serguei Chatrchyan et al. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].

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

[131] S. Abdullin et al. "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c". In: *Eur. Phys. J.* C60 (2009). [Erratum: Eur. Phys. J.C61,353(2009)], pp. 359–373. DOI: 10.1140/epjc/s10052-009-0959-5,10.1140/epjc/s10052-009-1024-0.

[132] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[133] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches". In: *Phys. Rev.* D81 (2010), p. 094023. DOI: 10.1103/PhysRevD.81.094023. arXiv: 0912.0033 [hep-ph].

[134] Daniele Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[135] CMS Collaboration. "Top Tagging with New Approaches". In: CMS-PAS-JME-15-002 (2016). URL: http://cds.cern.ch/record/2126325.

[136] CMS Collaboration. "Jet algorithms performance in 13 TeV data". In: CMS-PAS-JME-16-003 (2017). URL: https://cds.cern.ch/record/2256875.

[137] CMS Collaboration. "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016". In: (2018). URL: http://cds.cern.ch/record/2622157.

[138] CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. rep. CMS-PAS-PFT-09-001. Geneva: CERN, 2009. URL: https://cds.cern.ch/record/1194487.

[139] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[140] CMS Collaboration. "Jet energy scale and resolution performances with 13TeV data". In: (2016). URL: https://cds.cern.ch/record/2160347.

[141] Serguei Chatrchyan et al. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[142] CMS Collaboration. "Identification of b quark jets at the CMS Experiment in the LHC Run 2". In: (2016). URL: https://cds.cern.ch/record/2138504.

[143] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[144] CMS Collaboration. "CMS Phase 1 heavy flavour identification performance and developments". In: (2017). URL: http://cds.cern.ch/record/2263802.

[145] Vardan Khachatryan et al. "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV". In: *JHEP* 03 (2011), p. 136. DOI: 10.1007/JHEP03(2011)136. arXiv: 1102.3194 [hep-ex].

[146] *Wikipedia B-tagging*. URL: https://en.wikipedia.org/wiki/B-tagging.

[147] *B-Tagging POG, CMS TWiki*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BtagPOG.

[148] John M. Campbell et al. "Top-Pair Production and Decay at NLO Matched with Parton Showers". In: *JHEP* 04 (2015), p. 114. DOI: 10.1007/JHEP04(2015)114. arXiv: 1412.1828 [hep-ph].

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

[150] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction". In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

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

[152] Simone Alioli, Sven-Olaf Moch, and Peter Uwer. "Hadronic top-quark pair-production with one jet and parton showering". In: *JHEP* 01 (2012), p. 137. DOI: 10.1007/JHEP01(2012)137. arXiv: 1110.5251 [hep-ph].

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

[154] Michal Czakon and Alexander Mitov. "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders". In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].

[155] Nikolaos Kidonakis. "Top Quark Production". In: *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013*. 2014, pp. 139–168. DOI: 10.3204/DESY-PROC-2013-03/Kidonakis. arXiv: 1311.0283 [hep-ph].

[156] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *Eur. Phys. J.* C53 (2008), pp. 473–500. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].

[157] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[158] Eren Engin. “Determination of the Parton Distribution Functions and Strong Coupling by Using Measurement of Jet Cross Sections at $\sqrt{s} = 13$ TeV with the CMS detector with a combined QCD Analysis”. Dissertation, Universität Hamburg, 2018. Dissertation. Hamburg: Universität Hamburg, 2018, p. 158. DOI: 10.3204/PUBDB-2018-03915. URL: https://bib-pubdb1.desy.de/record/414381.

[159] Sourav Dey. “Measurement of the Double-Differential Inclusive Jet Cross Section in Proton-Proton Collisions at 13 TeV Centre-of-Mass Energy with the Compact Muon Solenoid Detector”. In: *Springer Proc. Phys.* 203 (2018), pp. 923–925. DOI: 10.1007/978-3-319-73171-1_226.

[160] *BTagging BTV, CMS TWiki*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco.

[161] CMS Collaboration. “Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events”. In: (2013). URL: https://cds.cern.ch/record/1581306.

[162] *Pileup, CMS TWiki*. URL: twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities.

[163] *Jet Flavour definition MC truth; CMS TWiki*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagMCTools.

[164] *Top pt Reweighting, CMS TWiki*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting2017.

[165] Andreas Hocker et al. “TMVA - Toolkit for Multivariate Data Analysis”. In: (2007). arXiv: physics/0703039 [physics.data-an].

[166] Kim Albertsson et al. “Machine Learning in High Energy Physics Community White Paper”. In: *J. Phys. Conf. Ser.* 1085.2 (2018), p. 022008. DOI: 10.1088/1742-6596/1085/2/022008. arXiv: 1807.02876 [physics.comp-ph].

[167] Wouter Verkerke and David P. Kirkby. “The RooFit toolkit for data modeling”. In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].

[168] G. D’Agostini. “Improved iterative Bayesian unfolding”. In: *Alliance Workshop on Unfolding and Data Correction Hamburg, Germany, May 27-28, 2010*. 2010. arXiv: 1010.0632.

[169] Souvik Das. “A simple alternative to the Crystal Ball function”. In: (2016). arXiv: 1603.08591 [hep-ex].

[170] *Presentation, Stefan Schmit. Statisctic School 2014 DESY*. URL: http://www.desy.de/~sschmitt/talks/UnfoldStatSchool2014.pdf.

[171] G. D’Agostini. “A Multidimensional unfolding method based on Bayes’ theorem”. In: *Nucl. Instrum. Meth.* A362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)00274-X.

[172] Andreas Hocker and Vakhtang Kartvelishvili. “SVD approach to data unfolding”. In: *Nucl. Instrum. Meth.* A372 (1996), pp. 469–481. DOI: 10.1016/0168-9002(95)01478-0. arXiv: hep-ph/9509307 [hep-ph].

[173] S Schmitt. “TUnfold, an algorithm for correcting migration effects in high energy physics”. In: *Journal of Instrumentation* 7.10 (2012), T10003–T10003. DOI: 10.1088/1748-0221/7/10/t10003. URL: https://doi.org/10.1088%2F1748-0221%2F7%2F10%2Ft10003.

[174] A. N. Tikhonov. “Solution of incorrectly formulated problems and the regularization method”. In: *Soviet Math. Dokl.* 4 (1963), pp. 1035–1038.

[175] *Luminosity, CMS TWiki*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM.

[176] Johan Alwall et al. “A Standard format for Les Houches event files”. In: *Comput. Phys. Commun.* 176 (2007), pp. 300–304. DOI: 10.1016/j.cpc.2006.11.010. arXiv: hep-ph/0609017 [hep-ph].

[177] CMS Collaboration. “Underlying Event Tunes and Double Parton Scattering”. In: (2014). URL: http://cds.cern.ch/record/1697700.

[178] G Anton et al. *AMADEUS: a new type of large area scintillation detector with position-, energy- and time-of-flight determination*. Tech. rep. BONN-ME-91-02. Bonn: Bonn Univ., 1991. URL: https://cds.cern.ch/record/223519.

[179] Spyros Argyropoulos and TorbjÃűrn SjÃűstrand. “Effects of color reconnection on $t\bar{t}$ final states at the LHC”. In: *JHEP* 11 (2014), p. 043. DOI: 10.1007/JHEP11(2014)043. arXiv: 1407.6653 [hep-ph].

[180] Jesper R. Christiansen. “Colour Reconnection - Models and Tests”. In: *PoS* EPS-HEP2015 (2015), p. 443. DOI: 10.22323/1.234.0443. arXiv: 1510.04859 [hep-ph].

[181] CMS Collaboration. “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV”. In: CMS-PAS-TOP-16-021 (2016). URL: http://cds.cern.ch/record/2235192.

[182] Vardan Khachatryan et al. “Measurement of the $\mathrm{t\bar{t}}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Eur. Phys. J.* C76.3 (2016), p. 128. DOI: 10.1140/epjc/s10052-016-3956-5. arXiv: 1509.06076 [hep-ex].

[183] Vardan Khachatryan et al. “Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV”. In: *Phys. Rev.* D95.9 (2017), p. 092001. DOI: 10.1103/PhysRevD.95.092001. arXiv: 1610.04191 [hep-ex].

[184] Daniela Dominguez Damiani. “Talk given in the Workshop on Resummation, Evolution, Factorization (REF), 2018. Tittle: Top-pair production at very high p_T: study of resummation effects using TMD approach.” In: (). URL: https://indico.cern.ch/event/696311/contributions/3204294/attachments/1756261/2847515/DanielaREF2018.pdf.

[185] F. Hautmann et al. “Soft-gluon resolution scale in QCD evolution equations”. In: *Phys. Lett.* B772 (2017), pp. 446–451. DOI: 10.1016/j.physletb.2017.07.005. arXiv: 1704.01757 [hep-ph].

[186] F. Hautmann et al. “Collinear and TMD Quark and Gluon Densities from Parton Branching Solution of QCD Evolution Equations”. In: *JHEP* 01 (2018), p. 070. DOI: 10.1007/JHEP01(2018)070. arXiv: 1708.03279 [hep-ph].

[187] A. Bermudez Martinez et al. “Production of Z-bosons in the parton branching method”. In: *Phys. Rev.* D100.7 (2019), p. 074027. DOI: 10.1103/PhysRevD.100.074027. arXiv: 1906.00919 [hep-ph].