ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea Magistrale in Fisica

Measurement of the inclusive and differential $t\bar{t}$ cross sections at 13 TeV in the all-jets boosted regime with CMS

Relatore:
Prof. Andrea Castro

Presentata da:
Fabio Iemmi

Anno Accademico 2015/2016

«If it disagrees with experiments, it's wrong.
In this simple statement is the key to Science.»
Richard P. Feynman, *Lectures at the Cornell University*

Sommario

La sezione d'urto di produzione di coppie $t\bar{t}$ viene misurata con l'esperimento CMS nel canale all-jets in collisioni pp ad un'energia nel centro di massa pari a 13 TeV. L'analisi si basa sullo studio di eventi $t\bar{t}$ nella topologia *boosted*, cioè eventi in cui i prodotti di decadimento del quark top hanno un grande boost di Lorentz e vengono perciò ricostruiti nel rivelatore come un unico, ampio jet. Il campione di dati utilizzato corrisponde a una luminosità integrata di 2.53 fb^{-1}. La sezione d'urto inclusiva misurata vale $\sigma_{t\bar{t}} = 727 \pm 46\ (\mathrm{stat.})^{+115}_{-112}\ (\mathrm{sys.}) \pm 20\ (\mathrm{lumi.})$ pb, un valore consistente con le previsioni teoriche. La sezione d'urto differenziale *detector-level* viene misurata in funzione della quantità di moto trasversa del jet più energetico e anch'essa confrontata con le previsioni teoriche della QCD. Infine viene riportata la sezione d'urto differenziale *parton-level*, misurata in funzione della quantità di moto trasversa del partone più energetico, estrapolata all'intero spazio delle fasi e confrontata con le previsioni della QCD.

Abstract

The $t\bar{t}$ production cross section is measured with the CMS detector in the all-jets channel in pp collisions at the centre-of-mass energy of 13 TeV. The analysis is based on the study of $t\bar{t}$ events in the boosted topology, namely events in which decay products of the quark top have a high Lorentz boost and are thus reconstructed in the detector as a single, wide jet. The data sample used in this analysis corresponds to an integrated luminosity of 2.53 fb^{-1}. The inclusive cross section is found to be $\sigma_{t\bar{t}} = 727 \pm 46$ (stat.)$^{+115}_{-112}$ (sys.) $\pm$ 20 (lumi.) pb, a value which is consistent with the theoretical predictions. The differential, detector-level cross section is measured as a function of the transverse momentum of the leading jet and compared to the QCD theoretical predictions. Finally, the differential, parton-level cross section is reported, measured as a function of the transverse momentum of the leading parton, extrapolated to the full phase space and compared to the QCD predictions.

I wish to express how lucky I feel to have been able to work with Professor Andrea Castro; I have to thank him for his support, his kindness and his constant care of me. Many thanks also to Konstantinos Kousouris that — even by long distance — constantly helped me and from whom I learned a lot. I wish my sincere gratitude to everyone who shared with me the path — long but joyful — which led to this wonderful achievement.

Contents

1 Introduction **1**

2 High Energy Physics at CERN: LHC and CMS **7**
- 2.1 The Large Hadron Collider 7
- 2.2 The CMS detector 11

3 Data analysis **15**
- 3.1 Samples 15
  - 3.1.1 Data sample 15
  - 3.1.2 Monte Carlo samples 15
- 3.2 Jet reconstruction and event selection 16
  - 3.2.1 Jet reconstruction 16
  - 3.2.2 Lepton reconstruction 19
  - 3.2.3 Trigger 19
  - 3.2.4 Event selection 20
- 3.3 Data vs. Simulations 22
- 3.4 Background prediction 25
- 3.5 Inclusive cross section measurement 32
  - 3.5.1 Signal yield extraction 32
  - 3.5.2 Postfit comparisons 32
  - 3.5.3 Systematic uncertainties 34
  - 3.5.4 Inclusive cross section 37
- 3.6 Differential measurement 38
  - 3.6.1 Fiducial cross section 38
  - 3.6.2 Detector effects - Unfolding 40
  - 3.6.3 Acceptance correction 43
  - 3.6.4 Parton-level cross section 44

4 Conclusions **47**

Chapter 1

Introduction

The top quark (t) is a Standard Model (SM) fermion belonging, along with the b quark and the τ, ν_τ leptons, to the third generation of fermions. It is found to be the heaviest fermion (having a mass comparable to that of a tungsten atom) and it has an electric charge equal to 2/3 e. Its discovery dates back to 1995, when the two experiments CDF [1] and D0 [2] operating at the Tevatron announced independently the observation of a particle compatible with a new quark.

While the main production mechanism at the Tevatron was the $q\bar{q}$ annihilation, giving rise to a $t\bar{t}$ pair (see Figure 1.1, upper row), the dominant process at the Large Hadron Collider (LHC) is the gluon-gluon fusion, which leads again to the production of a $t\bar{t}$ pair, as reported in Figure 1.1, lower row. The single top quark production is also possible, but with a three times lower cross section, so that its contribution is found to be almost negligible.

The top quark plays a very important role in the SM due to its very large mass m_t, so that the precise knowledge of its properties is critical for the general understanding of the theory. For example, it turns out from SM calculations that the mass of the W boson can be computed as:

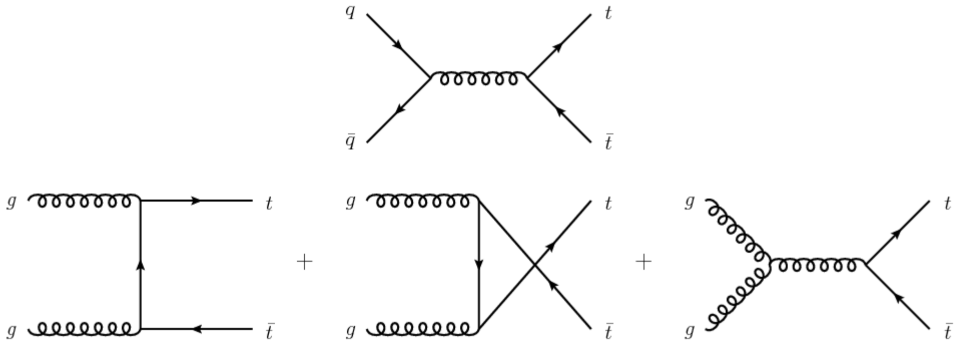

Figure 1.1: Dominant top quark production mechanisms at the Tevatron (upper row) and at the LHC (lower row).

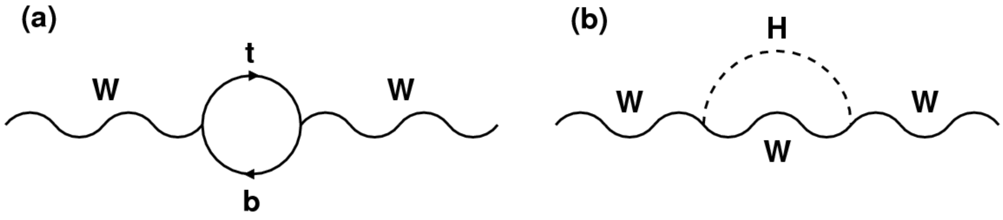

Figure 1.2: Radiative corrections to the W boson propagator. Top-bottom loop (left panel) and Higgs loop (right panel).

$$m_{\mathrm{W}} = \left(\frac{\pi\alpha}{G_{\mathrm{F}}\sqrt{2}}\right)^{\frac{1}{2}} \frac{1}{\sin\theta_{\mathrm{W}}\sqrt{1-\Delta r}},$$

where α is the fine-structure constant, G_{F} is the Fermi constant, θ_{W} is the Weinberg angle and Δr is a function related to the radiative corrections of the W propagator, see Figure 1.2, of the form $\Delta r \sim f(m_{\mathrm{t}}, \log m_{\mathrm{H}})$. Since the parameters α, G_{F} and θ_{W} are measured with very high precision, accurate measurements of m_{t} (and m_{W}) led to constrains on the value of the Higgs boson mass m_{H} and, now that the Higgs boson has been discovered, can be used to check the overall consistency of the theory. In addition, precise measurements of m_{t} and m_{H} are related to the stability of the vacuum [3].

Measurements of the top quark pair ($t\bar{t}$) production cross section are a good check of the quantum chromodynamics (QCD) predictions and have the potential to constrain the QCD parameters. In fact, the pair production cross section can be written as:

$$\sigma(pp \to t\bar{t}) = \sum_{i,j} \int \mathrm{d}x_i \; f_i(x_i, \mu^2) \int \mathrm{d}x_j \; f_j(x_j, \mu^2) \; \hat{\sigma}_{ij}(\hat{s}, \mu^2, m_{\mathrm{t}}),$$

where indices i, j indicates partons, f_k are the parton distribution functions (PDFs) of gluons and light quarks, $x_{i,j}$ are the momentum fractions of the partons, $\hat{s} = x_i x_j s$ is the centre-of-mass energy of the partons, $\hat{\sigma}_{i,j}$ is the cross section of the partonic process and μ is the factorization scale, related to the perturbative order of the calculations. Since the cross section is built up starting from the PDFs, a precise measurement can lead to constrains on them, along with constrains on the value of m_{t}.

Moreover, $t\bar{t}$ processes should be sensitive to physics beyond Standard Model (BSM): new physics should manifest itself as an excess in the production cross section with respect to the SM predictions. New particles, symmetries and unifications have been introduced; some of those new theories predict the existence of new, spin-1/2 particles called *vector-like quarks*. One of them, the T', is supposed to decay into a Z boson and a top quark. A good knowledge of the top quark properties thus

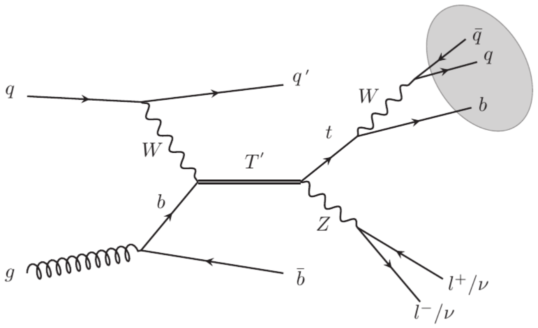

Figure 1.3: Feynman diagram for one of the possible BSM production processes and decays of a T' particle.

enhance the chance of detecting these new, exotic particles. Figure 1.3 shows the Feynman diagram for one of the possible processes which lead to the creation and decay of a T' quark. Finally, the $t\bar{t}$ process is a dominant SM background to search for some new physics phenomena, and therefore its precise knowledge is essential for new discoveries.

A t ($\bar{t}$) quark, once produced, decays in a time comparable to 10^{-24} s into a W^+b ($W^-\bar{b}$) pair. Different decay channels are then open for the W bosons originating from the decay of a $t\bar{t}$ pair:

1. Both W bosons decay into a lepton-neutrino pair. This is the so-called *dilepton channel*:

   $$t\bar{t} \longrightarrow W^+bW^-\bar{b} \longrightarrow \bar{\ell}\nu_\ell \; b \; \ell^{'}\bar{\nu}_{\ell'} \; \bar{b} \tag{1.1}$$

   where $\ell \equiv e, \mu$. The case in which $\ell = \tau$ is usually treated separately. The fraction of $t\bar{t}$ events ending in this state is given by the *branching ratio* (BR) of the channel, which for the dilepton channel is about 5%.

2. Only one W boson decays into a lepton-neutrino pair. In this case the channel is called *single-lepton channel*:

   $$t\bar{t} \longrightarrow W^+bW^-\bar{b} \longrightarrow \bar{\ell}\nu_\ell \; b \; q\bar{q}'\bar{b} \tag{1.2}$$

   or:

   $$t\bar{t} \longrightarrow W^+bW^-\bar{b} \longrightarrow q\bar{q}'b \; \ell\bar{\nu}_\ell \; \bar{b} \tag{1.3}$$

   The branching ratio of this channel is about 30%.

3. Both W bosons arising from the $t\bar{t}$ pair decay into a pair of light quarks. As a result, the final state consists of at least six partons (more are possible due

to initial- and final-state radiation), two of which are b quarks; the partons will then hadronize to form at least six jets in the final state. This is the so called *all-hadronic* or *all-jets channel*:

$$t\bar{t} \longrightarrow W^{+}bW^{-}\bar{b} \longrightarrow q\bar{q}'b\ q\bar{q}'\bar{b} \longrightarrow j_1 j_2 j_3 j_4 j_5 j_6 \tag{1.4}$$

with a branching ratio of about 46%. A peculiar feature of this channel is that it makes possible to fully reconstruct the $t\bar{t}$ decay products, in contrast to the leptonic channels, where the presence of one or two neutrinos makes the full event interpretation ambiguous. Since multiple jets are present in the final state, the measurement of the $t\bar{t}$ cross section in the all-jets channel involves larger uncertainties (from the jet energy scale) than in the leptonic channels, but it is complementary and unique in its possibility to fully reconstruct the $t\bar{t}$ system.

The large $t\bar{t}$ yield expected in proton-proton (pp) collisions at the LHC allows to perform more precise measurements of the $t\bar{t}$ properties in extended parts of the phase space and, more importantly, allows to make differential measurements, as a function of the $t\bar{t}$ kinematic properties, such as the transverse momentum p_{T} of the top quark. The measurements presented here have been performed with the Compact Muon Solenoid (CMS) detector during 2015. The inclusive cross section, measured in the all-jets decay channel, is reported, as well as the fiducial differential cross section at detector level. Finally, an unfolding procedure is set up in order to obtain the differential cross section at parton level.

In the latest runs of LHC at $\sqrt{s} = 13$ TeV very energetic top quarks can be produced. Considering a top quark with a Lorentz boost $\gamma = E/m$ [1], the distance between its W boson and b quark daughters will be approximately $\Delta R = 2/\gamma$. In the case in which the W boson decays hadronically, the resulting final jets (two from the decay of the W and one from the hadronization of the b quark) will often be contained into an angular region smaller than $2/\gamma$, so that a jet clustering algorithm (see section 3.2) with a distance parameter larger than $2/\gamma$ will typically collect all the jets into a single, "big" jet. The usual jet clustering algorithms are based on a parameter R which corresponds to the minimum radius that would have to be used in order to collect the three quarks of the chain $t \to Wb \to q\bar{q}'b$. Figure 1.4 shows the minimal size of a jet that contains the decay quarks in the decay $t \to Wb \to q\bar{q}'b$ as a function of the top quark p_{T} for one of such algorithms – the so-called Cambridge–Aachen algorithm – which is not the one used in the following analysis, but which is representative of the typical jet behaviour. Such figure indicates the need for a large R value when the top quark p_{T} increases. Such wide, large object are often referred to as *top jets*, and events in which the final state contains top jets are referred to as events in the *boosted topology* or *boosted regime*. It is now clear that the application of the usual analysis approach to the

---

[1] In natural units.

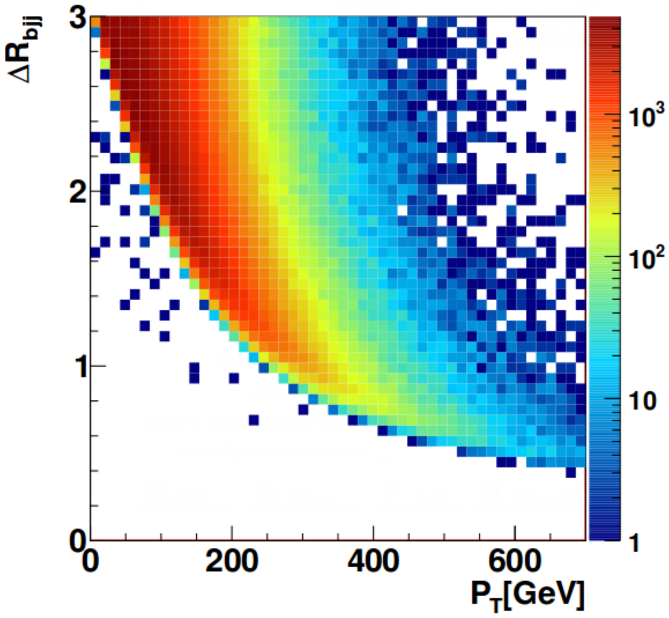

Figure 1.4: Minimal size of a Cambridge-Aachen jet that contains the decay quarks in the chain $t \to Wb \to q\bar{q}'b$ as a function of the top quark $p_{\rm T}$.

decay of $t\bar{t}$ pairs, based on events in the so called *resolved topology*, where the full kinematical event reconstruction starting from the six final jets is performed, is no longer possible. The analysis presented here thus follows a new *boosted* approach, based on the study of the properties of the top jets, which applies to high-energy top quarks. A pictorial representation of a resolved and a boosted $\bar{t}$ decay is shown in Figure 1.5. The combination of the boosted analysis with the resolved one (still valid for less energetic top quarks) extends the $p_{\rm T}$ reach of the differential measurements, almost doubling it, allowing us to reach the TeV region for the first time.

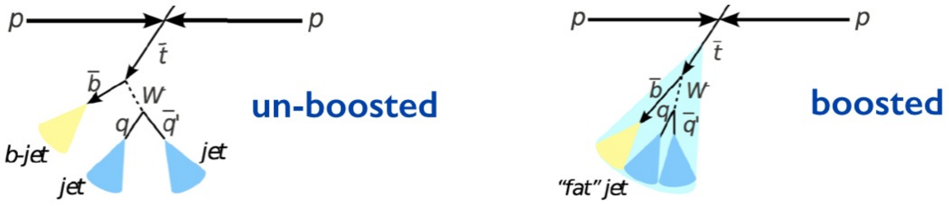

Figure 1.5: Decay of a $\bar{t}$ in the resolved (left) and boosted (right) topologies.

Chapter 2

High Energy Physics at CERN: LHC and CMS

In this chapter we shall describe the basic elements of the experimental apparatus which provided the data used in this analysis, namely the LHC machine and the CMS detector.

2.1 The Large Hadron Collider

The LHC [4] is a proton-proton accelerator and collider placed at CERN, Geneva, Switzerland. It is housed in the pre-existing 27 km tunnel that was used for the LEP machine. The approval of the LHC project was given by the CERN Council in December 1994. In that first stages, the plan was to build the accelerator in two steps, starting with a machine capable of a centre-of-mass energy of 10 TeV, to be upgraded to a final value of 14 TeV. However, during the years 1995-1996, negotiations with non-member states (such as United States, Japan, India, Russia and Canada) secured substantial contributions to the project, and in December 1996 the CERN Council approved the construction of a 14 TeV machine in a single stage.

LHC is not a perfect circle, since it is composed by eight arcs (see Figure 2.1) where the protons, bent by bipolar magnets, are free to circulate, and eight straight sections, each of them being approximately 530 meters long, which are used for various tasks, such as beam injection/ejection or beam collisions. Being a particle-particle collider, the beam pipe is composed by two rings with counter-rotating beams, unlike particle-antiparticle colliders that can exploit the charge dependence of the Lorentz force in order to have both beams occupying a single ring.

One of the main goal of the LHC is to detect rare events, possibly coming from processes beyond the SM. To achieve this, a high integrated luminosity is needed in order have a sufficiently high signal yield $N = \sigma \times L$, where σ is the cross section of the process of interest. The instantaneous luminosity depends exclusively on the machine design and can be written as

$$\mathcal{L} = \frac{n_1 n_2 N_b \gamma \nu}{4\pi \epsilon_n \beta^*} F,$$

where n_i stands for the particle content in the i-th bunch; N_b is the number of bunches per beam, γ is the relativistic factor, ν is the revolution frequency, ϵ_n is the transverse beam emittance, β^* is the beta function at the collision point and F is the the so-called geometric luminosity reduction factor, which accounts for the crossing angle at the interaction point. The LHC instantaneous luminosity is not constant in time, but decreases due to degradation effects, such as the degradation of intensities and emittances of the beams. It is found that the instantaneous luminosity decreases with a decay factor

$$\tau_{\text{decay}} = \frac{N_0}{\mathcal{L}_0 \sigma_{\text{tot}} k},$$

where N_0 is the initial beam intensity, $\mathcal{L}_0$ the initial luminosity, σ_{tot} is the total cross section ($\sigma_{\text{tot}} \approx 10^{25}\text{cm}^2$ at 14 TeV) and k is the number of interaction points, following a decay law of the form

$$\mathcal{L}(t) = \frac{\mathcal{L}_0}{(1 + t/\tau_{\text{decay}})^2}.$$

The integrated luminosity is obtained by integrating in time the instantaneous luminosity:

$$L = \int \mathcal{L}(t)\ \text{dt}.$$

The LHC has two high-luminosity, multi-purpose experiment: ATLAS (A Toroidal LHC ApparatuS) and CMS, both aiming at a peak, nominal, instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ (reached for the first time on June 26, 2016); there is then a third, low-luminosity experiment, LHCb, designed for the physics of the b quark (aiming at $\mathcal{L} = 2 \times 10^{29}\text{cm}^{-2}\text{s}^{-1}$) and finally a fourth experiment, ALICE (A Large Ion Collider Experiment), studying the heavy-ions collisions (aiming at $\mathcal{L} = 10^{27}\text{cm}^{-2}\text{s}^{-1}$).

One of the most outstanding features of the LHC is the magnet system, which is at the edge of the present technology. It is made by superconducting NbTi magnets, cooled by superfluid helium to the temperature of about 2 K, and operating at magnetic fields above 8 T. Space limitations in the LEP tunnel led to the choice of the so-called "two-in-one" or "twin-bore" magnet design: the windings for the two beam channels are housed in the same cryostat, making the magnet structure quite complicate and challenging. The LHC ring is covered by 1232 *main cryodipoles*, having all the same basic design. A schematic cross section of a cryodipole is shown in Figure 2.2. The core of a cryodipole is the dipole cold mass (the part inside the shrinking cylinder/He II vessel), which contains all the components cooled by superfluid helium. It has an overall length of about 16.5 m, a diameter of 570 mm

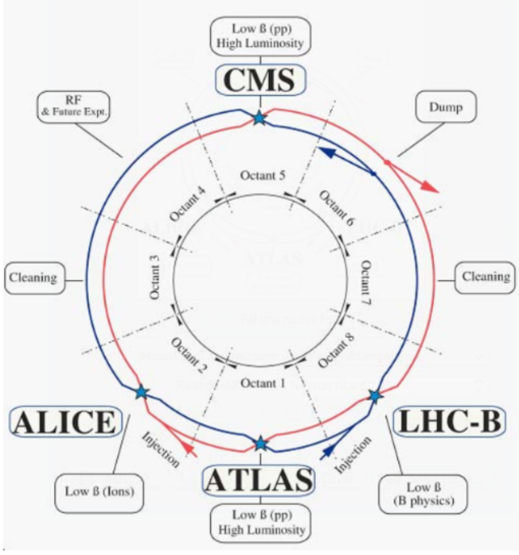
Low ß (pp)
High Luminosity
CMS
RF
& Future Expt.
Dump
Octant 5
Octant 4
Octant 6
Octant 3
Octant 7
Octant 2
Octant 8
Octant 1
Cleaning
Cleaning
ALICE
LHC-B
Injection
Injection
ATLAS
Low ß (Ions)
Low ß (pp)
High Luminosity
Low ß
(B physics)

Figure 2.1: Schematic layout of the LHC showing the four main experiments and the beam injection and dumping zones (beam 1 going clockwise, beam 2 going anticlockwise).

and a mass of about 27.5 t. The mass is also curved in the horizontal plane in order to correctly match the particle trajectories. In correspondence of the eight intersections along the LHC ring (four in the experiment areas, two for the beam cleaning, one for the radiofrequency cavities and one for the beam dumping, see Figure 2.1) the *main quadrupoles* are placed. These magnets accomplish many tasks, such as collimating the beams in proximity of the experiments in order to maximize the probability of pp collisions. A set of multipole corrector magnets are coupled with the quadrupoles, giving rise to 10 different combinations of magnets used for a fine tuning of the magnetic fields.

LHC exploits the properties of ultra-high vacuum in many ways. It has three separate vacuum systems, one for the beam pipe, one for insulating the cryogenically cooled magnets and one for insulating the helium distribution line. In order to ensure a high beam lifetime and low background in the experiments it is necessary to avoid collisions between protons and gas molecules. It is found that a pressure of 10^{-10} – 10^{-11} mbar is needed for a proper behaviour of the machine. The beam vacuum system is actually split in two independent chambers, one for each beam pipe, which merge into a single one in the collision regions. In the other two systems, vacuum is employed in order to reduce the amount of heat which seeps from the external room-temperature environment into the cryogenic parts.

The acceleration of particles at the LHC is obtained using radiofrequency (RF) cavities. A RF cavity is a metallic chamber containing a magnetic field supplied by a dedicated power generator. The design of the cavities creates resonant

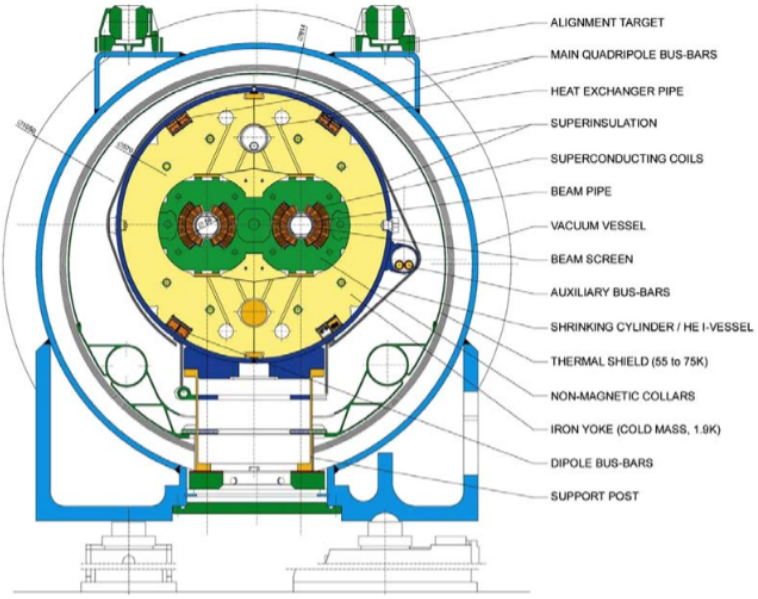
ALIGNMENT TARGET
MAIN QUADRIPOLE BUS-BARS
HEAT EXCHANGER PIPE
SUPERINSULATION
SUPERCONDUCTING COILS
BEAM PIPE
VACUUM VESSEL
BEAM SCREEN
AUXILIARY BUS-BARS
SHRINKING CYLINDER / HE I-VESSEL
THERMAL SHIELD (55 to 75K)
NON-MAGNETIC COLLARS
IRON YOKE (COLD MASS, 1.9K)
DIPOLE BUS-BARS
SUPPORT POST

Figure 2.2: Cross section of a cryodipole.

electromagnetic waves, so that particles passing through feel an overall force and are accelerated. The field in the RF cavities is made to oscillate at a frequency of 400 MHz. A perfectly timed proton, with precisely the requested energy, will see zero accelerating force. Protons with slightly different energies, arriving earlier or later, will be accelerated or decelerated so that they stay close to the energy of the ideal particle. In this way, the particle beam is sorted into bunches. The LHC is instrumented with 16 RF cavities, housed in four cylindrical refrigerators (two per beam) which keep the RF working in a superconducting regime. During the accelerating process the protons in the bunches see an overall acceleration on each passage through the cavities, picking up the energy needed to keep up with the increasing field in the LHC powerful magnets. The nominal energy can be reached in around 15 minutes, the bunches having passed the cavities around 1 million times.

The LHC is the last stage of a complex acceleration system (see Figure 2.3), and a long process of pre-acceleration is needed before the beams are injected in the machine. First of all, protons are obtained by ionizing hydrogen atoms; then the Linear accelerator 2 (Linac2) transfers to the particles the first amount of energy. In the following step, protons are accelerated up to 1.4 GeV in the Proton Synchrotron Booster, reaching then the Proton Synchrotron and the Super Proton Synchrotron. Finally, they are injected in the LHC with an energy of about 450 GeV and accelerated for the last time to the nominal energy of 6.5 TeV. A scheme of this acceleration system is presented in Figure 2.3.

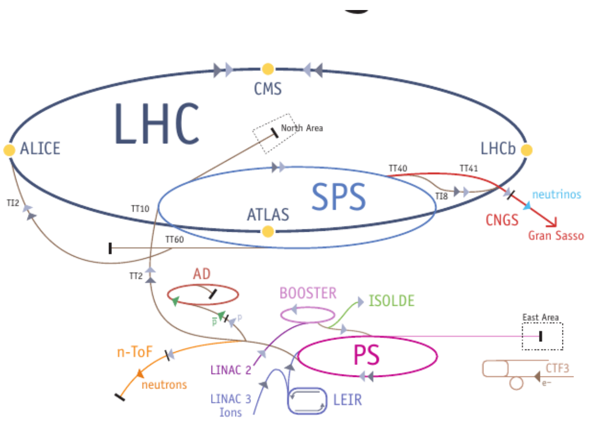

Figure 2.3: The acceleration system at CERN.

The LHC goals are several and different. First of all, the machine has been designed for the search of the Higgs boson, a search that ended successfully on July 4, 2012 when the ATLAS [5] and CMS [6] collaborations announced independently the discovery of a new particle compatible with the Higgs boson. Several questions are still open though: the LHC could eventually determine whether supersymmetric particles exist, or define the properties of the so-called dark matter, which accounts for the 27% of the mass of the universe. Moreover, it may clarify if the electroweak and strong forces correspond to different manifestations of the same force, a possibility theorized in the Grand Unification Theories; through the study of the physics of the b quark it could explain the asymmetry between matter and antimatter in the universe (related to the CP violation) and through the study of the heavy ion collisions it could clarify the properties of the quark-gluon plasma.

2.2 The CMS detector

CMS [7] is a multi-purpose detector operating at the CERN LHC. It is 21.6 m long and has a diameter of 14.6 m, it has a total weight of 12500 t and it is installed about 100 m underground near the village of Cessy (France). Working coupled to the LHC, the most powerful accelerator ever built, CMS has to accomplish many challenging tasks. For example, when operating at design, instantaneous luminosity (10^{34} cm^{-2}s^{-1}) and design centre-of-mass energy (14 TeV) the experiment will observe a rate of about 10^9 events/s. Very efficient triggers are needed in order to reduce the huge events rate and made possible the storage of data; ultra-fast electronics is needed in order to reduce as much as possible the pileup, namely the superimposition of several inelastic collision during a single acquisition; a radiation-hard environment is needed in order to guarantee the correct behaviour of the experimental apparatus for a sufficiently long time.

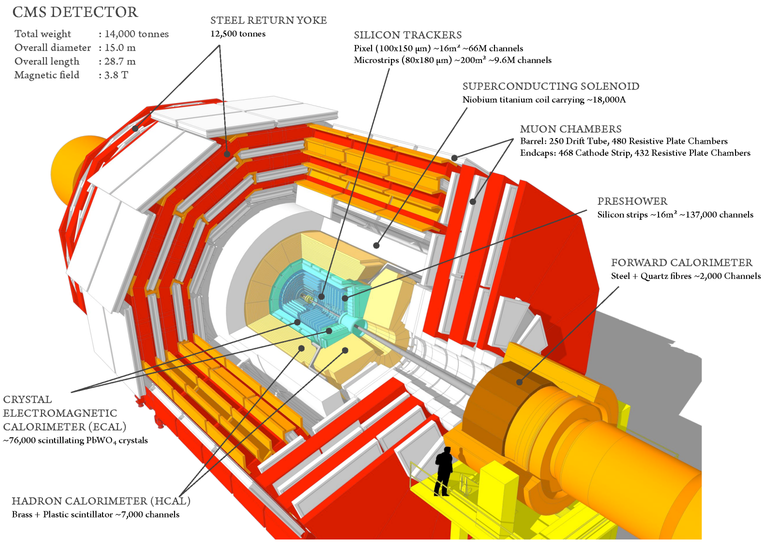
CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 2.4: Overall layout of CMS.

The layout of CMS is showed in Figure 2.4. It is structured in concentric layers, each of them designed for a different task. Starting from the core of the detector, we find:

- a tracking system;
- an electromagnetic calorimeter;
- a hadron calorimeter;
- a superconducting magnet;
- a muon system.

The coordinate system used inside CMS is shown in Figure 2.5 and is centred at the nominal collision point, with the y-axis pointing vertically upward, the x-axis pointing radially inward to the center of the LHC and the z-axis pointing along the beam direction. The azimuthal angle ϕ is measured from the x-axis in the x-y plane and the radial coordinate in this plane is denoted by r. The polar angle θ is measured from the z-axis.

The tracking system is used to measure precisely and efficiently the trajectories of charged particles emerging from the interaction points, along with secondary vertices of interaction. When the LHC works at design parameters, particles from more than 20 pp interactions traverse the tracker for each bunch crossing (which means every 25 ns). For this reason, a technology featuring high granularity and fast response is needed. These requirements led to the choice of a silicon detector technology. The tracker is composed of a silicon pixel detector disposed in three

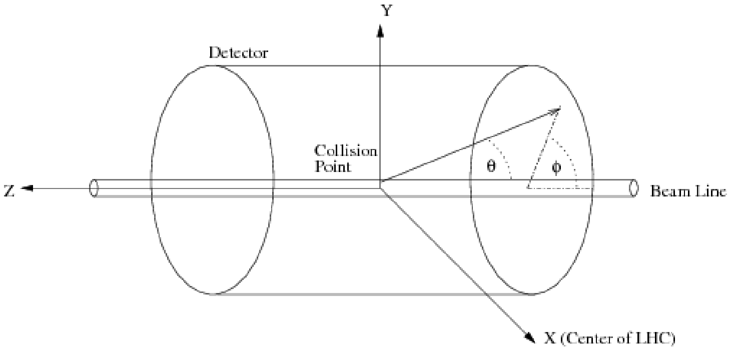

Figure 2.5: CMS coordinate system.

barrel layers at radii between 4.4 cm and 10.2 cm, plus a silicon strip detector with 10 barrel layers extending to a radius of 1.1 m. Each system is completed by endcaps which ensure an optimal geometrical acceptance. The CMS tracker is the largest silicon tracker ever built.

After the tracker, the Electromagnetic Calorimeter (ECAL) is placed between radii 1.30 m and 1.70 m approximately. It is a hermetic, homogeneous calorimeter made of 61200 lead tungstate ($PbWO_4$) scintillating crystals mounted in the central barrel part, covering a pseudorapidity [1] range of $|\eta| < 1.48$, closed by 7324 crystals in each of the two endcaps covering the range $1.48 < |\eta| < 3.0$. The high density, short radiation length and small Molière radius result in a compact, high-granularity calorimeter. Avalanche photodiodes (APDs) are used in the barrel as photodetectors, while vacuum phototriodes (VPTs) are used in the endcaps.

The following layer houses the Hadron Calorimeter (HCAL). It is a sampling calorimeter split in two half barrels and radially restricted between radii 1.77 m and 2.95 m, covering a pseudorapidity range of $|\eta| < 1.3$. It is composed by 36 azimuthal wedges, each of them divided in 4 sectors. The absorber is made of alternating layers of stainless steel and brass, while the active material is a plastic scintillator (Kuraray SCSN81) which exhibits long term stability and a good radiation hardness. The scintillator is used in the tile plus wavelength-shifting fibre design. The CMS HCAL is made of about 70000 tiles. The light collected by the fibres is transferred to photodiodes which enhance the signal and make possible its usage. The HCAL is completed by two endcaps which cover the pseudorapidity range of $1.3 < |\eta| < 3.0$. Beyond $|\eta| = 3.0$, a forward hadron calorimeter placed about 11 m after the interaction point extends the pseudorapidity coverage to $|\eta| = 5.2$, making the CMS HCAL an almost hermetic calorimeter.

After the Hadron Calorimeter we find possibly the most important component of the CMS detector: the superconducting magnet. The use of a strong magnetic field is critical for the correct behaviour of the experiment, since it is used to bend the particle trajectories and thus contribute to make particle discrimination. The magnet has been designed to reach a magnetic field of 4 T and consists in a solenoid, whose windings are made of NbTi and which is composed by 4 layers instead of the usual 1 (or 2, in detectors such as ZEUS or BaBar). As in the case of the LHC magnets, the CMS solenoid must be kept at very low temperature in order to exploit its superconducting properties.

---

[1] The pseudorapidity is defined as $\eta = -\log[\tan(\theta/2)]$

The final layer of the detector houses the muon system. As implied in the detector name, muon detection is crucial to CMS: for example, the decay of a Higgs boson into a ZZ pair, which then decays in 4 leptons, has been called "gold plated" channel in the case in which all the leptons are muons. The muon system has 3 functions: muon identification, momentum measurement, and triggering. As in the previous cases, the muon system is composed by a barrel plus two endcaps. The barrel region, where the muon rate is relatively low and the magnetic field uniform, is instrumented with drift tubes (DTs). In the endcap regions, where the muon rate is high and the magnetic field is non-uniform, the muon system uses cathode strip chambers (CSC).

Chapter 3

Data analysis

3.1 Samples

In this section we give a brief description of the samples used to perform the analysis. We discuss separately data and Monte Carlo samples.

3.1.1 Data sample

The data used for this analysis were collected during the 2015 Run from *pp* collisions at 13 TeV, with a bunch spacing of 25 ns, corresponding to an integrated luminosity of 2534 pb^{-1}. They were stored in ROOT files as trees that contain all the relevant informations on the major physics objects reconstructed in the detector (jets, leptons, photons, tracks).

3.1.2 Monte Carlo samples

The Monte Carlo (MC) program POWHEG (v2) [8, 9] has been used to simulate $t\bar{t}$ events. It includes a next-to-leading order (NLO) QCD matrix element calculation, which is matched to a parton shower simulation with PYTHIA [10, 11]. Events contained in the samples were generated in all the decay channels with a top quark mass of 172.5 GeV as a default. This is the standard value commonly used to generate MC $t\bar{t}$ events and, even though it is not constantly updated, it is fairly close to the current world average of $m_\text{t} = (173.21 \pm 0.51 \pm 0.71)$ GeV [12].

The dominant background process is the QCD multijet production, which has been simulated with MADGRAPH [13]. Speaking of subdominant backgrounds, the tW single top quark production has been simulated using POWHEG (v1), while W plus jets and Z plus jets events have been simulated with MADGRAPH. In all the cases, the hard process is interfaced with the PYTHIA parton shower.

The detector effects are simulated using GEANT4 [14]. We summarize in Table 3.1 the MC samples used during the analysis, reporting them in the same order

Table 3.1: Monte Carlo samples used in the analysis. QCD simulated events are divided in slices of H_{T}, which stands for the scalar sum of jet transverse momenta.

Sample	σ(pb)	Events
$t\bar{t}$ `pair production`	832	91644400
`QCD multijet production` ($300 < H_{\mathrm{T}} < 500$)	3.67e+5	16909000
`QCD multijet production` ($500 < H_{\mathrm{T}} < 700$)	2.94e+4	19289000
`QCD multijet production` ($700 < H_{\mathrm{T}} < 1000$)	6.524e+3	15387600
`QCD multijet production` ($1000 < H_{\mathrm{T}} < 1500$)	1.064e+3	5049270
`QCD multijet production` ($1500 < H_{\mathrm{T}} < 2000$)	121.5	3771420
`QCD multijet production` ($H_{\mathrm{T}} > 2000$)	2.542e+1	1921300
`Single` t `production` (tW)	35.6	1000000
`Single` $\bar{t}$ `production` ($\bar{t}W$)	35.6	999400
W`+jets production`	3539	22689600
Z`+jets production`	1460	12051600

as presented above: $t\bar{t}$ events, QCD background and subdominant backgrounds. The cross sections used for the computation and the number of event in each sample, previous to any selection, are showed.

3.2 Jet reconstruction and event selection

3.2.1 Jet reconstruction

Jet reconstruction algorithms are among the most useful tools used in particle physics in order to analyse data from hadron collisions. Theoretical QCD calculations provide results in terms of quarks and gluons in the final state. Once produced in the detectors, these particles undergo hadronization and give rise to the characteristic signature of QCD jets. A quantitative mapping between hadrons in the form of jets and final state quarks and gluons is then required. This mapping is provided by jet reconstruction algorithms. A wide variety of algorithms is available, differing from one another by the internal parameters used to reconstruct a jet. A good algorithm should show some fundamental properties: it should be easy to implement during data analysis and theoretical calculations, and it must be well defined at any order of perturbation theory. This last property, known as infrared safety, was not achieved until 2007: usage of infrared unsafe algorithms strongly degraded the quality of jet measurements because it was impossible to compare them with many theoretical calculations. Fortunately, jet reconstruction algorithm used at the present time at LHC are all infrared safe, and they belong to the class of so called *sequential recombination jet algorithms*. They are based on some kind of definition of how likely two partons arise from a QCD splitting, and they proceed sequentially to construct a jet evaluating which partons are close in this measure. This kind of algorithms can be described as follows: first, the following two quantities are defined

$$\begin{aligned} d_{ij} &= \min(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p})\frac{\Delta R_{ij}^2}{R^2} \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \\ d_{iB} &= p_{\mathrm{T},i}^{2p} \end{aligned} \tag{3.1}$$

where d_{ij} is the "distance" between particle i and particle j, d_{iB} is the distance between particle i and the beam, R is the so called distance parameter, y_i and ϕ_i are the rapidity [1] and azimuthal angle of particle i and p is an integer which can take values $-1, 0, 1$. Then the procedure starts:

1. Evaluate all the distances d_{ij} and d_{iB} from the list of all the final-state particles;
2. Find the minimum distance;
3. If it is a d_{ij}, recombine together particle i and j, then come back to step 1;
4. Otherwise declare particle i to be a jet, remove it from the final state and come back to step 1;
5. Algorithm stops when no particles remain.

Jet reconstruction is implemented here using the anti-k_T algorithm [15], which corresponds to $p = -1$, with distance parameter $R = 0.8$ (AK8 algorithm). The particle which are used as inputs to the the jet reconstruction are identified using the *particle-flow* algorithm (PF) [16, 17]. This algorithm is able to reconstruct a very wide set of particles, such as electrons, muons, photons, charged and neutral hadrons, exploiting a combination of information coming from all of the CMS subdetectors (calorimeters, trackers etc.) in order to obtain an optimal determination of directions, energies, types, etc.. The use of PF allows to enhance the energy resolution of jets, if compared to methods which exploit information from calorimeters only; resolutions of 15% at 10 GeV, 8% at 100 GeV and 4% at 1 TeV can be reached. Finally, using simulations, corrections to the jet energies are applied, defined as functions of variables such as η ad p_T of the jets. In order to reject "false jets", that is single particles eventually reconstructed as a whole jet, the energy fraction coming from photons, electrons and neutral hadrons is requested to be less than 99%, while the energy fraction coming from charged hadrons must be much greater than zero.

As we stated before, if a top quark has enough boost (typically $p_\mathrm{T} > 350$ GeV) its decay products are strongly collimated and are reconstructed as a single top jet with large distance parameter (hence the usage of the AK8 jets instead of the AK4

---

[1] The rapidity is defined as $y = 0.5 \times \log[(E + p_\mathrm{z})/(E - p_\mathrm{z})]$

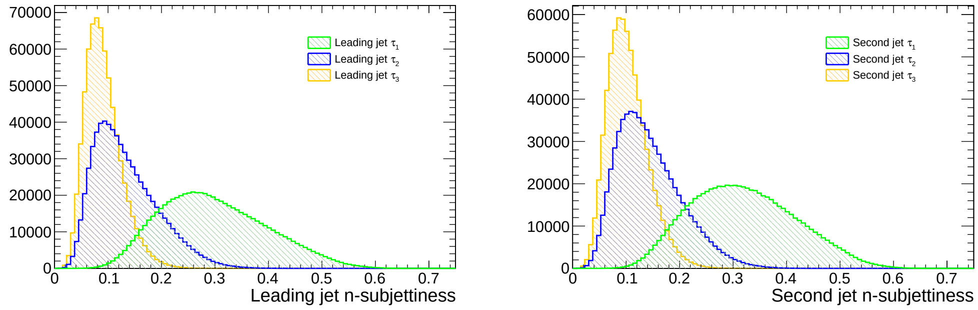

Figure 3.1: n-subjettiness distributions for the leading jet and the second jet for $t\bar{t}$ simulated events.

jets (with $R = 0.4$), usually used in the resolved analysis). An useful experimental feature of these jets is their number of subjets. Jets coming from hadronic decays of top quarks are expected to show three regions of space where most of the energy is deposited, corresponding to the three partons resulting from the decay. On the other side, jets arising from the hadronization of light quarks are expected to have one or two such regions. In order to reveal the substructure of such jets, we look at the n-subjettiness [18] variable τ_i, defined as follows:

$$\tau_i = \frac{1}{\sum_k p_{\mathrm{T},k} R} \sum_k p_{\mathrm{T},k} \mathrm{min}\,(\Delta R_{1k}, \Delta R_{2k}, ...\Delta R_{ik}), \tag{3.2}$$

where the index k enumerates the constituents of the input jet, $p_{\mathrm{T},k}$ is the p_{T} of the k-th constituent, R is the distance parameter of the original jet, and ΔR_{ik} is the angular distance in the $\eta - \phi$ space between the i-th subjet and the k-th constituent. The subjettiness variable τ_i measures the compatibility of a jet with the hypothesis that it is composed of i subjets. In the case of exactly i subjets, the value of τ_i tends to zero. Since for top quark decays a topology with three subjets is expected, the ratios τ_3/τ_2 and τ_3/τ_1 have been proven to be robust variables for discriminating between boosted top quarks (three subjets) and ordinary QCD jets that tend to have fewer subjets. Figure 3.1 shows the n-subjettiness distributions for the leading jet and the second jet in the simulations. The distributions are obtained after the preselection (see Section 3.2.4) and follow the expected behaviour: since in the simulations we are dealing with $t\bar{t}$ events only, jets are less likely to be constituted by one or two subjets, so that τ_1 and τ_2 are distributed quite far from the value zero, while τ_3 has a peak closer to this region.

For the calculation of the jet invariant mass, the *softdrop method* [19] is used, which removes from the jet clustering soft and/or collinear particles thus suppressing the contamination by underlying event, pileup etc.. In this method, the clustering of the jet j with distance parameter R is reverted step by step, breaking j into j_1 and j_2 at each iteration. Then, the softdrop condition is examined:

$$\frac{\min\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)}{p_{\mathrm{T},1}+p_{\mathrm{T},2}} > z_{cut} \cdot \left(\frac{\Delta R_{12}}{R}\right)^{\beta}. \tag{3.3}$$

If the condition holds then j is considered the final jet and the procedure stops. If this is not the case, the leading subjet is relabelled as j and the softer one is discarded. The two parameters of the algorithm, z_{cut} and β, control the strength of the fractional p_{T} selection and the suppression of the collinear radiation, respectively. In this analysis the default CMS reconstruction is used, in which these parameters take the values $z_{cut} = 0.1$ and $\beta = 0$.

The identification of jets that likely originate from the hadronization of b quarks is done with the *combined secondary vertex* version 2 (CSVv2) b-tagger [20]. b-flavoured hadrons have, on average, a quite long lifetime, so they can leave the primary vertex of interaction and form secondary vertices. The CSVv2 algorithm looks for these secondary vertices and provides a useful discriminator to separate top quark decays from background events.

3.2.2 Lepton reconstruction

Both muons and electrons are defined in the region $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.4$, using the default particle flow present in the CMS Software (CMSSW) [21]. The chosen $t\bar{t}$ final state contains no high-p_{T} isolated leptons from W boson decay. The muon isolation variable is defined as the sum of the p_{T} of all tracks, except for the muon track, originating from the $t\bar{t}$ interaction vertex within a cone of $\Delta R = 0.3$. It is required to be less than 5% of the muon p_{T}. For electrons the isolation variable is the sum of the p_{T} of neutral hadrons, charged hadrons, and photon PF candidates in a cone of $\Delta R = 0.3$ around the electron. Contributions of the electron to the isolation variable are suppressed excluding a small region around the electron. This isolation variable is required to be smaller than 7% of the electron p_{T}. To distinguish the possible $t\bar{t}$ final states we count the number of such leptons, N_{leptons}.

3.2.3 Trigger

When the LHC works at nominal parameters, the proton bunches cross at a rate of 40 MHz. At CMS this rate is reduced by a factor 1000 using a Level-1 hardware trigger and subsequently by another factor 1000 using a software-implemented High Level Trigger (HLT) [22]. The trigger used for this analysis is `HLT_AK8PFJet360_TrimMass30`. At HLT level the decision is made in two steps, one based on calorimetric jets (CaloJet) and one based on particle-flow jets: first a CaloJet filter is applied and then a PFJet filter makes the final decision. The CaloJet filter asks for the presence of at least one AK8 CaloJet with $p_{\mathrm{T}} > 260$ GeV. If events pass the filter, the PF algorithm is run and the final filter requires at least one AK8 PFJet with $p_{\mathrm{T}} > 360$ GeV and trimmed mass [23] greater than

Table 3.2: Summary of the selection for the signal sample. The indices in parentheses refer to the leading and subleading jets/subjets.

Boosted selection
`HLT_AK8PFJet360_TrimMass30` AK8 jets $p_{\mathrm{T}} > 200$ GeV, $\lvert\eta\rvert < 2.4$, $m_{\mathrm{SD}} > 50$ GeV $N_{\mathrm{jets}} \geq 2$ $p_{\mathrm{T}}^{(1)} > 450$ GeV $N_{\mathrm{leptons}} = 0$ $CSVv2^{(1)}_{\mathrm{subjet}} > 0.8$, $CSVv2^{(2)}_{\mathrm{subjet}} > 0.8$ $\mathcal{F} > 0$ $150 < m_{\mathrm{SD}}^{(1)} < 200$ GeV

30 GeV. The online trimming algorithm is used to remove the contribution of soft PF jets that are probably associated with pileup and not to the prompt parton decay. These trigger criteria are aimed to suppress low mass QCD jets.

3.2.4 Event selection

Data samples coming from experiments like CMS are composed by a wide variety of events, only a little fraction of them being $t\bar{t}$ events. It is thus impossible to perform a data analysis using this "raw" samples, since the signal events are "hidden" by the big amount of background events. One of the crucial tasks to be accomplished in a data analysis is then to apply a proper event selection in order to strongly suppress the background yield in data samples, while preserving as much signal events as possible. The study of the ideal selection criteria is based on the MC simulation, for which we can define the quantity called efficiency

$$\epsilon = \left[\frac{N_{\mathrm{sel}}}{N_{\mathrm{gen}}}\right]_{\mathrm{MC}}, \tag{3.4}$$

where N_{gen} is the total number of $t\bar{t}$ events generated in the MC sample, while N_{sel} is the number of MC events passing the event selection. The efficiency of the selection is a fundamental quantity since it is an essential ingredient for the calculation of the inclusive cross section.

The selection of $t\bar{t}$ events in the boosted topology starts with the request of the trigger `HLT_AK8PFJet360_TrimMass30`. This is followed by a preselection asking for at least two AK8 jets with $p_{\mathrm{T}} > 200$ GeV, $\lvert\eta\rvert < 2.4$, and softdrop mass $m_{\mathrm{SD}} > 50$ GeV. Figure 3.2 shows the trigger efficiency as measured in the data sample compared to the one measured in the simulated sample. The MC sample corresponds to $t\bar{t}$ and QCD events mixed with the expected cross sections. Since

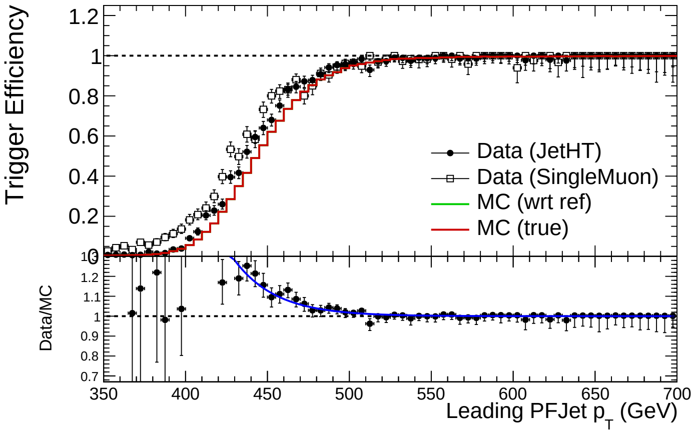

Figure 3.2: Trigger efficiency, measured in data, as a function of the leading jet p_T, along with the corresponding efficiency in MC (upper panel); data/MC scale factor (bottom panel).

the two efficiencies are slightly different, a data/MC scale factor (bottom panel) is needed in order to account for this effect and correct the selection efficiency, which is evaluated using the simulated sample.

We then ask the leading jet p_T to be greater than 450 GeV and apply the lepton veto $N_\mathrm{leptons} = 0$ in order to select all-jets decays. Furthermore we request the two AK8 jets to be b-tagged, that is to have at least one CSVv2 subjet. As shown in Figure 3.3, this request rejects the cases where the AK8 jets are the result of the merging of only the jets from the W decay: the peak at the W mass disappears. In addition, a Fisher discriminant $\mathcal{F}$ is constructed from the n-subjettiness ratios τ_3/τ_2 and τ_3/τ_1 of the two leading jets. The samples used for the training are a simulated $t\bar{t}$ sample for signal and a simulated QCD sample for the background. Both signal and background events are requested to satisfy the previous selection, except for the b-tagging (this is necessary in order to have a sufficiently large background yield and have a more generic training). Figure 3.4 shows the output of the multivariate training. After the training, it turns out that an appropriate selection requires the condition $\mathcal{F} > 0$. Finally, for the differential measurement only, the softdrop mass of the leading jet is required to lie in the range $150 - 200$ GeV. A sample of events fulfilling this selection will be referred to as a *signal sample*. The full set of selection requirements is summarized in Table 3.2.

In order to evaluate the quality of the selection, some quantities have been studied, using the MC $t\bar{t}$ sample. Figure 3.5 shows how the $t\bar{t}$ events are distributed after the selection, with respect to the decay channel. It appears that events in the all-jets channel are strongly dominating, as expected. We then study how the reconstructed jets are matched to the two $t\bar{t}$ partons. Since there could be sources of jets other than top quarks (e.g. underlying event, pileup), we find the

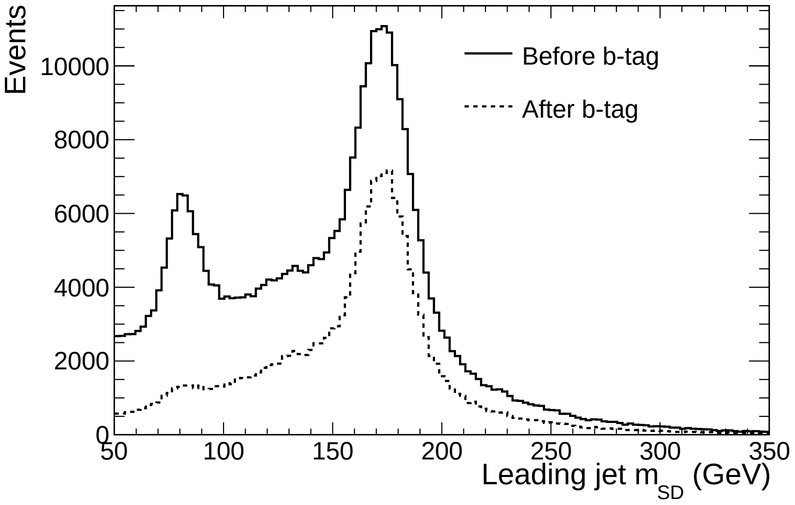

Figure 3.3: Softdrop mass m_{SD} for the leading jet before (solid line) and after (dashed line) the request of at least one CSVv2 subjet, in simulated $t\bar{t}$ events.

probability that the leading jet "comes" from a top quark. For each event, the ΔR between the leading jet and each of the two top quarks is calculated. If the lower ΔR obtained fulfils the request $\Delta\mathrm{R} < 0.4$ we then call this a jet matching to a top quark. Figure 3.6 shows the quantity known as sample purity, that is the probability, expressed in function of the leading jet p_{T}, of having the leading jet matching to one of the two top quarks. The sample purity appears to be flat and close to 100%, the little discrepancy being due to the arbitrary choice of the value $\Delta R = 0.4$ as a discrimination between matching and non matching jets: looser cuts on this variable lead to values closer and closer to 100%.

3.3 Data vs. Simulations

In this section we present comparisons between data and simulations for a wide range of variables of interest. The $t\bar{t}$, single top quark, W + jets and Z + jets contributions are normalized to the integrated luminosity, using the respective theoretical cross sections reported in Table 3.1. The QCD multijet background is normalized a first time to the integrated luminosity, using the leading-order cross section of each H_{T} slice reported in Table 3.1; then a global k-factor is applied in such a way that the total simulated yield is equal to the number of events in data. It should be pointed out that the goal of these comparisons is to demonstrate the overall consistency of the data, but they are not used in the actual measurement.

We start showing the softdrop mass of the two leading jets, see Figure 3.7. Note that in both the distributions we observe the top quark resonance, the peak being

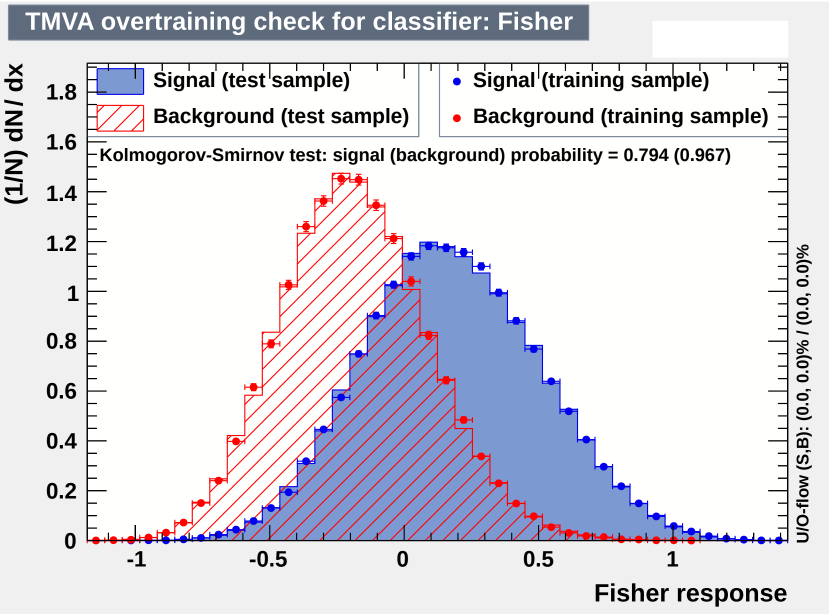

Figure 3.4: Output of the multivariate training.

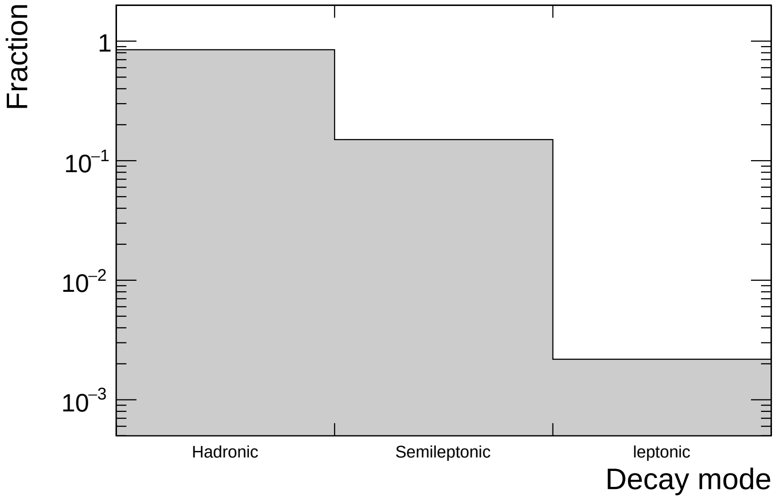

Figure 3.5: Decay mode of simulated $t\bar{t}$ after the selection.

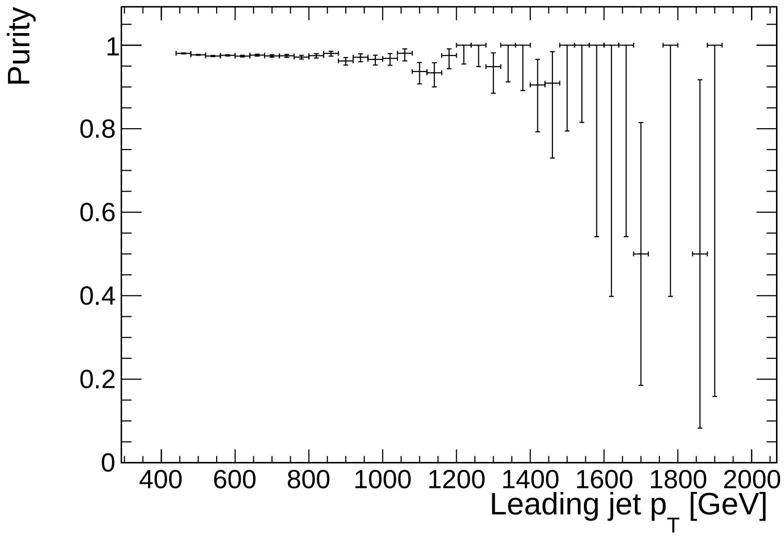

Figure 3.6: Sample purity after the selection.

more pronounced in the leading jet distribution due to the higher p_{T} cut required in the selection. We then focus on the properties of the two AK8 jets and on some global variables. Figure 3.8 shows the p_{T} and η distributions of the most energetic jets, showing good agreement between data and simulations. The agreement is also present for event quantities like the jet and b-jet multiplicities, as shown in Figure 3.9. Figure 3.10 shows some other event properties such as the sum H_{T} of the E_{T} of all jets, and the Fischer discriminant output, where it is clearly visible the cut $\mathcal{F} > 0$ which helps in discriminating the signal from the background. We studied also variables concerning the two leading jets, as shown in Figure 3.11. Finally, Figure 3.12 shows the substructure properties of the two leading jets: the mass distribution of the leading subjet within the leading jet clearly shows a peak around the W mass, which can be interpreted as a highly-boosted jet originated from the merging of the decay products of the W boson. In the case of the mass distribution of the second subjet we still see a W peak, though less pronounced. We then see in Figure 3.13 a reasonable agreement between data and simulations in the substructure properties τ_3/τ_2 and τ_3/τ_1, with the exception of the τ_3/τ_2 ratio of the leading jet. Causes of this discrepancy could be a mismodelling of the boosted top jets in the $t\bar{t}$ sample, or even a mismodelling of the QCD jets that mimic the properties of top jets. Since the biggest discrepancy is observed in the QCD-dominated region, the second hypothesis is probably the correct one. In order to investigate this, we select in the data a *control sample*, QCD-enriched, by asking the usual selection combined with a veto on the presence of b-tagged subjets (for a detailed description of the control sample, see Section 3.4). Substructure

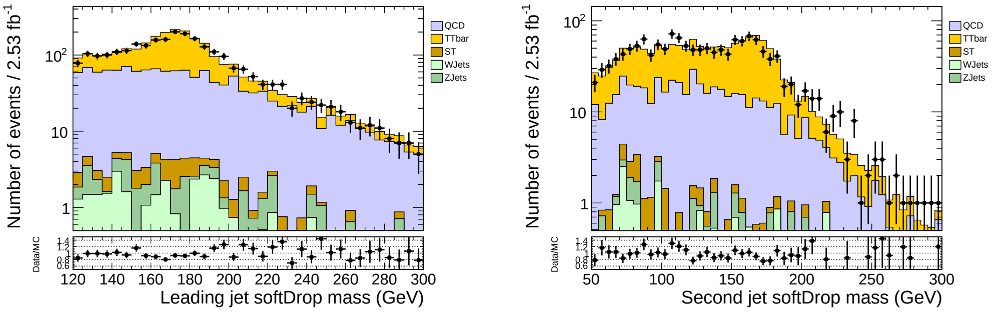

Figure 3.7: Leading jet and second jet softdrop mass, after the selection.

properties in the control sample are shown in Figure 3.14, where we still see a partial disagreement between data and simulation in the variable τ_3/τ_2. This should be a quite strong hint of the fact that the properties of the QCD jets are not very well modelled. On the other hand, we see a good agreement when we consider the Fischer discriminant, as shown in Figure 3.15. We must point out, however, that the QCD simulation is not directly used in the measurement, since the background prediction will be made starting from a control sample in data (see Section 3.4).

3.4 Background prediction

Due to its large cross section in pp collisions, the QCD multijet production turns out to be the dominant background for the $t\bar{t}$ events in the all-jets final state. The subdominant backgrounds, such as single top quark, W + jets, Z + jets production are found to contribute up to a maximum of 4% to the total background. We saw in Section 3.3 the first hint of the fact that the QCD multijet production cannot be safely reproduced by simulations. It turned out that data are not described accurately enough; moreover, despite the large number of simulated events, the efficiency after the selection is very small, resulting in a relatively small number of available events. We therefore define a *control sample*, QCD-enriched, which can be used for a data-based estimate of the shape of the QCD process for each observable of interest.

Since one of the critical properties exploited in the signal selection is the presence of at least two b-tagged subjets, we revert this condition asking for exactly no b-tagged subjets. Acting this way we should obtain an almost pure QCD sample showing a behaviour similar to that of QCD multijet events in the signal region. A summary of the selection for the control sample is shown in Table 3.3. However, it must be pointed out that this sample is obtained in the control region, so that we must now find a way to extrapolate its behaviour to the signal region, where the request of two b-tagged subjets is made. This is performed through a *transition*

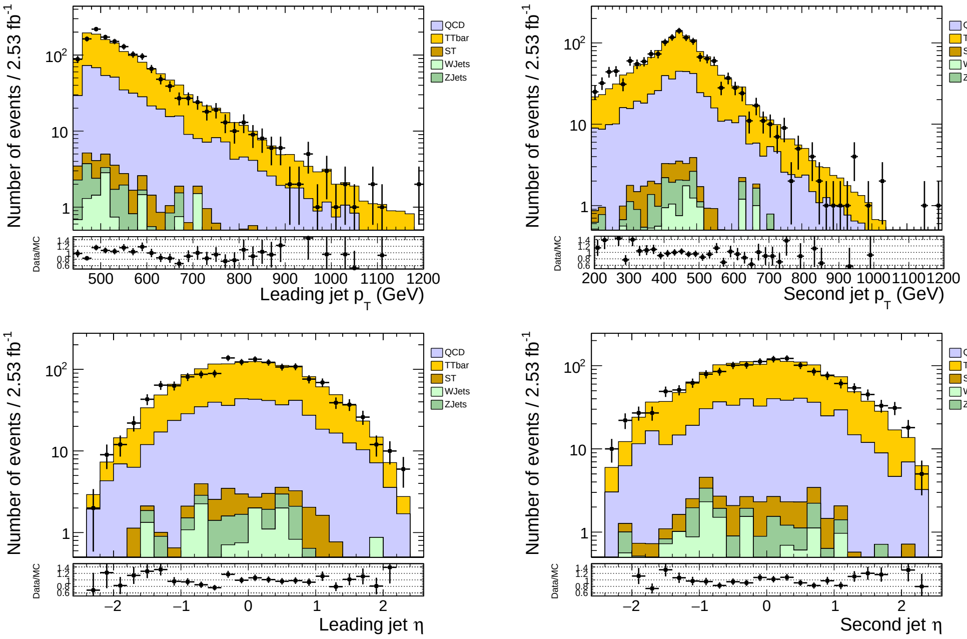

Figure 3.8: Distributions of some relevant jet properties: the p_{T} (first row) and η (second row) of the two leading jets are shown, after the selection.

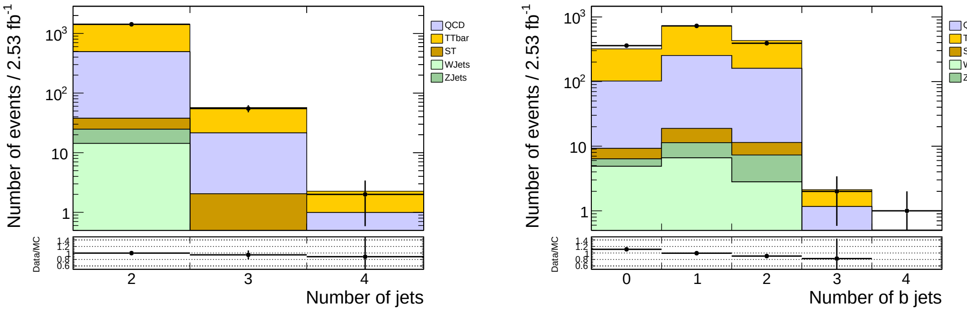

Figure 3.9: Distributions of the jet and b-jet multiplicities, after the selection.

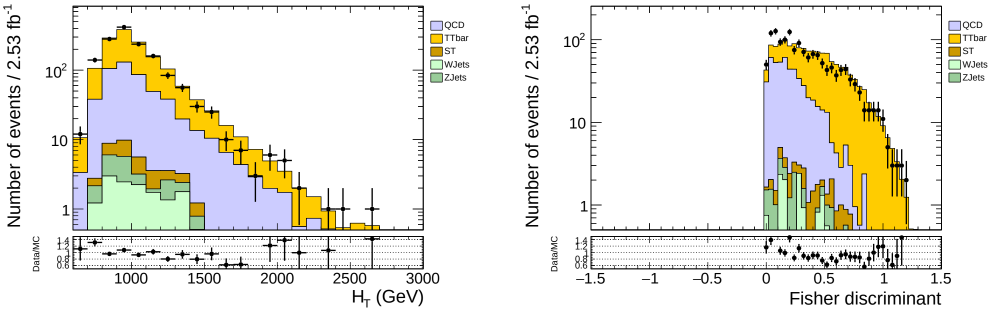

Figure 3.10: Distributions of H_T and of the Fisher discriminant output, after the selection.

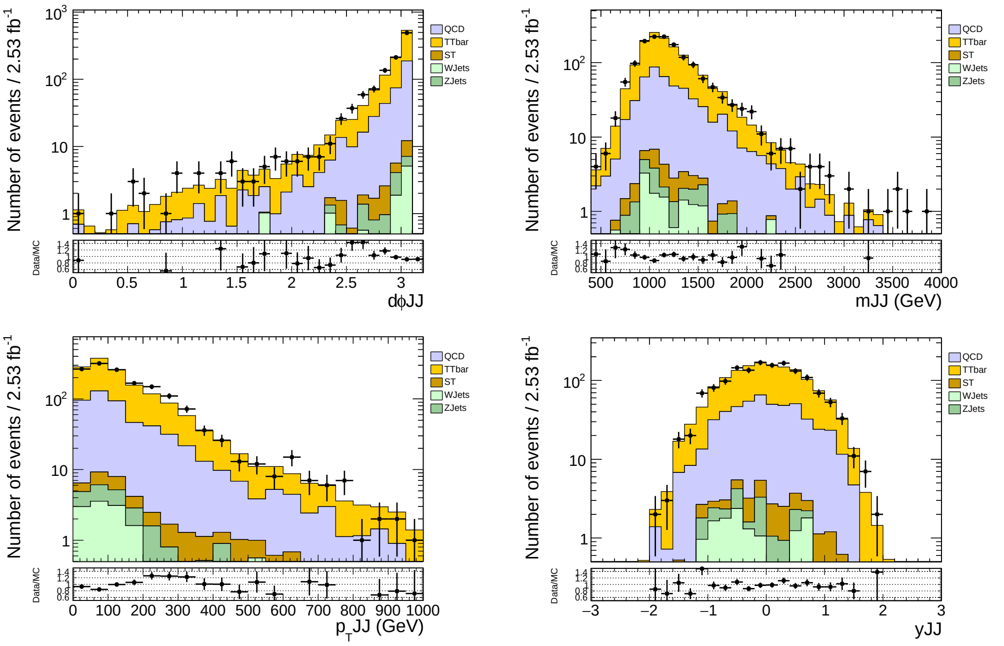

Figure 3.11: Distributions of some relevant event properties after the selection: some dijet kinematic distributions such as $\Delta\phi_{\rm jj}$ and $m_{\rm jj}$ (upper row), $p_{\rm T,jj}$ and $y_{\rm jj}$ (bottom row) are shown.

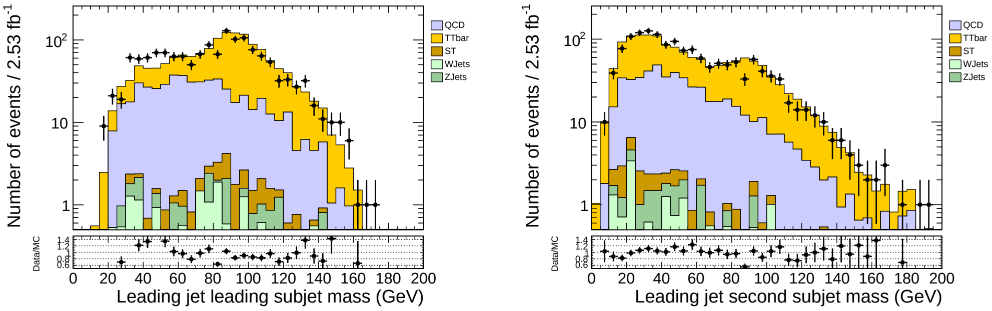

Figure 3.12: Softdrop mass distributions of the two subjets inside of the leading jet, after the selection.

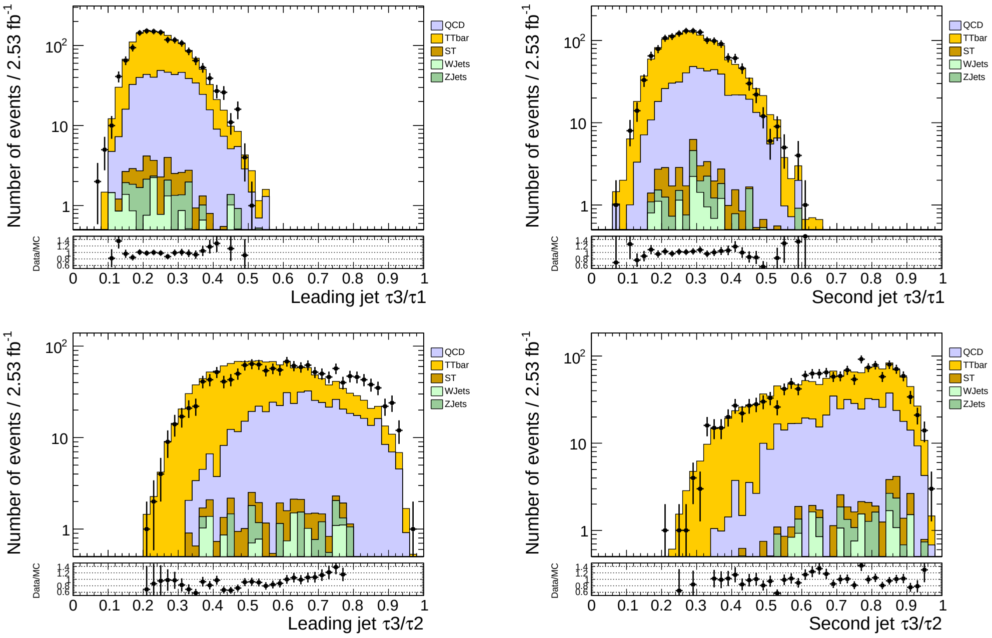

Figure 3.13: Substructure properties of the two leading jets in the signal sample, after the selection: the n-subjettiness ratios τ_3/τ_1 (upper row) and τ_3/τ_2 (lower row) of the two leading jets are shown.

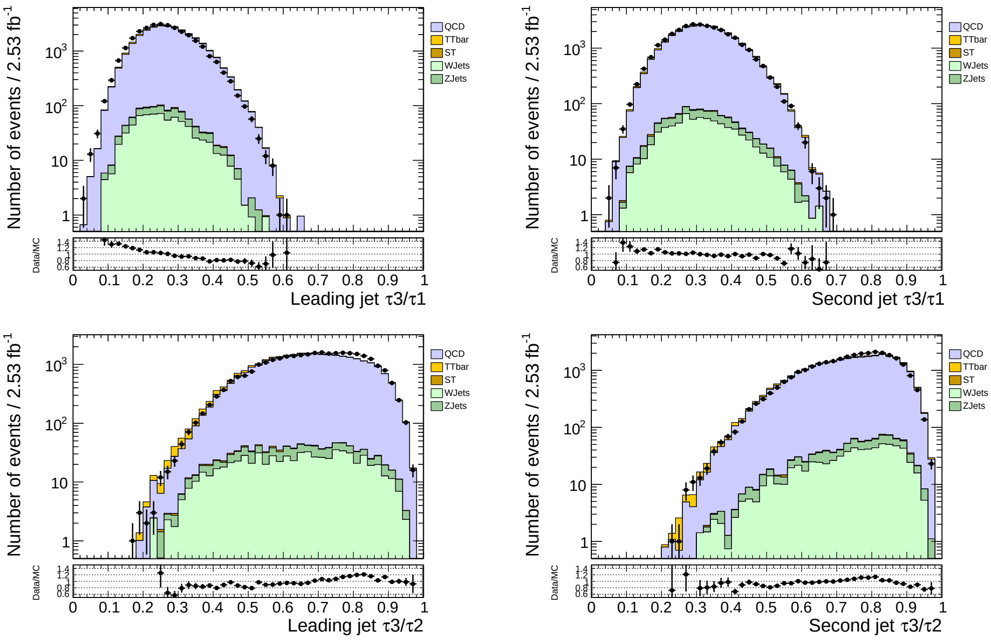

Figure 3.14: Substructure properties of the two leading jets in the control sample (passing the selection but with no b-tagged subjets). The *n*-subjettiness ratios τ_3/τ_1 (upper row) and τ_3/τ_2 (lower row) of the two leading jets are shown.

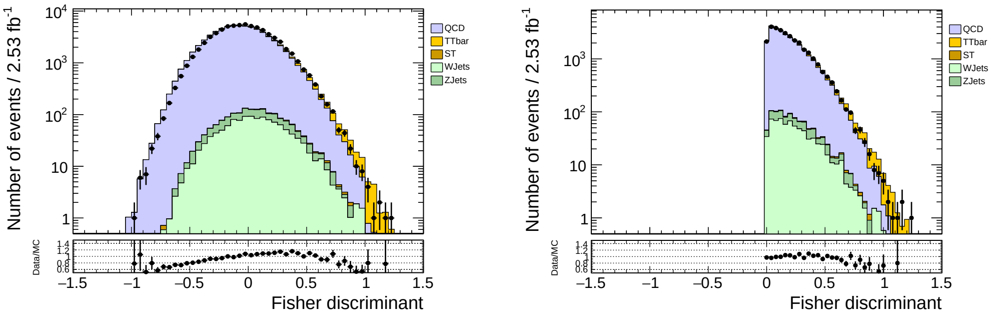

Figure 3.15: Fisher discriminant in the control sample before and after the cut on it.

Table 3.3: Summary of the selection for the control sample. The indices in parentheses refer to the leading and subleading jets/subjets.

Control sample selection
`HLT_AK8PFJet360_TrimMass30` AK8 jets $p_{\mathrm{T}} > 200$ GeV, $\lvert\eta\rvert < 2.4$, $m_{\mathrm{SD}} > 50$ GeV $N_{\mathrm{jets}} \geq 2$ $p_{\mathrm{T}}^{(1)} > 450$ GeV $N_{\mathrm{leptons}} = 0$ b-tag veto $\mathcal{F} > 0$ $150 < m_{\mathrm{SD}}^{(1)} < 200$ GeV

factor between the control and signal regions, based on QCD simulations. The distributions in the signal region are extrapolated from those in the control region using the following formula:

$$B^{\mathrm{data}}_{\mathrm{signal}}(x) = \frac{B^{\mathrm{MC}}_{\mathrm{signal}}(x)}{B^{\mathrm{MC}}_{\mathrm{control}}(x)} \cdot B^{\mathrm{data}}_{\mathrm{control}}(x) \ , \tag{3.5}$$

where x denotes the variable of interest (for example m_{SD}), and “control”, “signal” refer to the selections with zero b-tagged and two b-tagged jets, respectively. The ratio $B^{\mathrm{MC}}_{\mathrm{signal}}(x)/B^{\mathrm{MC}}_{\mathrm{control}}(x)$ is the transition factor.

Figure 3.16 shows a first “closure test” for the background prediction, which indicates how well the procedure works. The comparison between distributions of m_{SD} for the leading jet in the signal and control samples is reported for the simulated QCD events. The left panel also shows the ratio $B^{\mathrm{MC}}_{\mathrm{signal}}(m_{\mathrm{SD}})/B^{\mathrm{MC}}_{\mathrm{control}}(m_{\mathrm{SD}})$ along with a smooth interpolating curve which is taken as a transition function between the two samples. Since the ratio is almost flat, the fitting curve is chosen to be a straight line. The right panel shows the control distribution corrected with the application of the transition function.

Figure 3.17 shows a second closure test for the background prediction. The comparison between p_{T} distributions for the leading jet in the signal and control samples is reported for the simulated QCD events. The left panel also shows the ratio $B^{\mathrm{MC}}_{\mathrm{signal}}(p_{\mathrm{T}})/B^{\mathrm{MC}}_{\mathrm{control}}(p_{\mathrm{T}})$ along with the interpolating curve which is taken as a transition function. Since also in this case the ratio is essentially flat, the fitting curve is again chosen to be a straight line. The right panel shows the control distribution corrected with the application of the transition function. In this second case the fitting curve appears to be a straight line with slope $m \simeq 0$ and y-intercept $q \simeq 1$, so that almost no difference is found between the corrected and original control distributions.

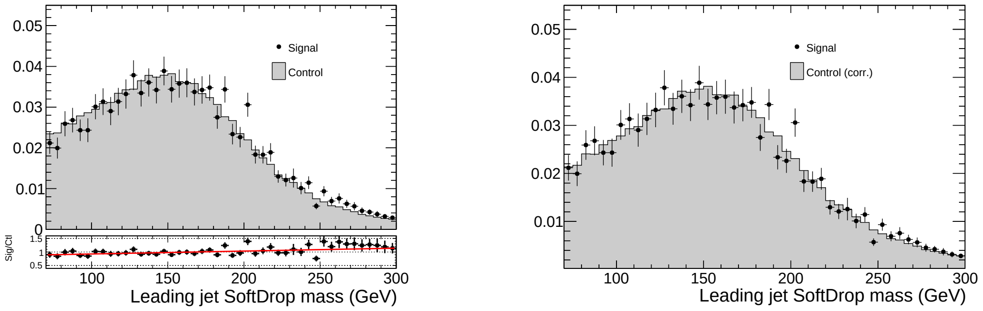

Figure 3.16: Closure of the background prediction method for the leading jet m_{SD}. The signal and control distributions in simulated QCD events, along with their ratio and the fitting transition function (right panel) and the corrected control distribution (left panel) are shown.

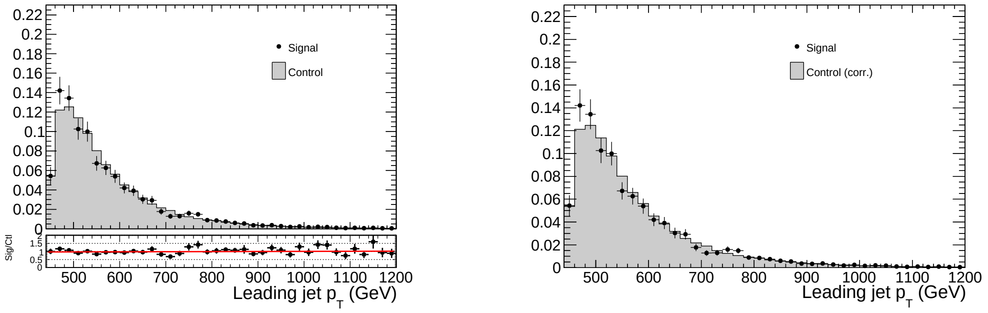

Figure 3.17: Closure of the background prediction method for the leading jet p_{T}. The signal and control distributions in simulated QCD events, along with their ratio and the fitting transition function (right panel) and the corrected control distribution (left panel) are shown.

3.5 Inclusive cross section measurement

The inclusive cross section measurement is performed following some standard steps: first of all, the $t\bar{t}$ signal and QCD background normalized distributions (templates), taken from simulations and the control sample in data respectively, are fitted with parametrized functions. Then, keeping these parameters fixed, a second extended maximum-likelihood fit to the data is performed with the signal and background yields being the free parameters. All the fits are implemented with the package RooFit [24]. Once the number of signal events is available, we explicitly calculate the inclusive cross section.

3.5.1 Signal yield extraction

The variable used to extract the signal yield is the softdrop mass m_{SD} of the leading jet. Figure 3.18 shows the signal template, taken from the MC simulation. The fitting function $S(m_{\mathrm{SD}})$ (blue, solid line) is the sum of two components: a Crystal Ball function (red, dashed line) describing the core and a Bernstein Polynomial of degree 4 (green, dashed line) describing the combinatorial background. Figure 3.19 shows the background template, taken from the control sample in data and corrected with the transition function reported in Figure 3.16. The fitting function $B(m_{\mathrm{SD}})$ (blue, solid line) is again the sum of two components: a Crystal Ball function (red, dashed line) and a Bernstein Polynomial of degree 4 (green, dashed line). As shown in the figures, the chosen functions reproduce quite well the mass distributions for both signal and background.

The functions $S(m_{\mathrm{SD}})$ and $B(m_{\mathrm{SD}})$ resulting from the fit of the $t\bar{t}$ and QCD templates are then used to perform an extended maximum-likelihood fit of the data, starting from a function of the form

$$D(m_{\mathrm{SD}}) = N_{\mathrm{sig}} \cdot S(m_{\mathrm{SD}}) + N_{\mathrm{bkg}} \cdot B(m_{\mathrm{SD}}), \tag{3.6}$$

where N_{sig} and N_{bkg} are the signal and background yields which are treated as free parameters. The fit of the data is shown is Figure 3.20. The final output of the extended maximum-likelihood fit gives the following yields:

$$\begin{aligned} N_{\mathrm{sig}} &= 1350 \pm 86; \\ N_{\mathrm{bkg}} &= 2000 \pm 89. \end{aligned} \tag{3.7}$$

3.5.2 Postfit comparisons

To support the goodness of the results obtained by the fit, we present some postfit comparisons, where the background distributions are taken from the control sample. Figure 3.21 shows the postfit m_{SD} distribution of the leading jet and of the second jet. The distributions are normalized to the fitted yields (Equation 3.7). All the remaining distributions are normalized to the signal and background yields in the

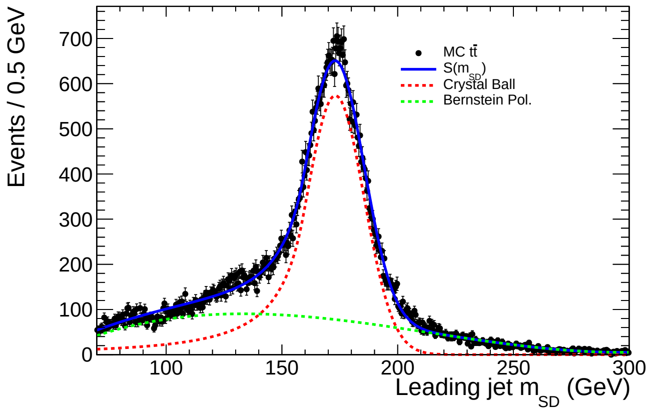

Figure 3.18: $m_{\rm SD}$ signal template for $t\bar{t}$ events and the fitted function $S(m_{\rm SD})$ (blue line), along with the individual components (red and green dashed lines).

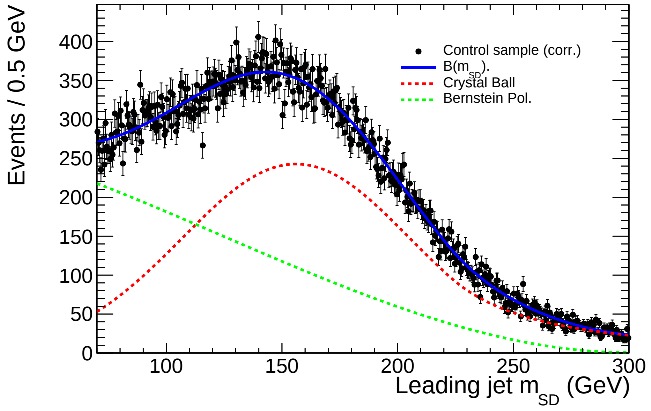

Figure 3.19: $m_{\rm SD}$ background template from the control sample and the fitted function $B(m_{\rm SD})$ (blue line), along with the individual components (red and green dashed lines).

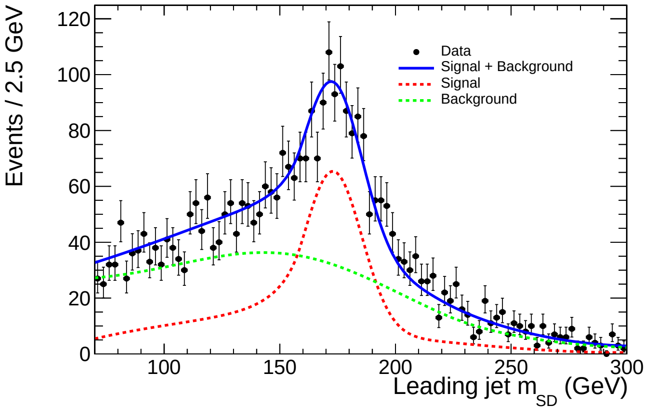

Figure 3.20: Fitted m_{SD} distribution (blue line), along with the signal (red, dashed line) and the background (green, dashed line) components.

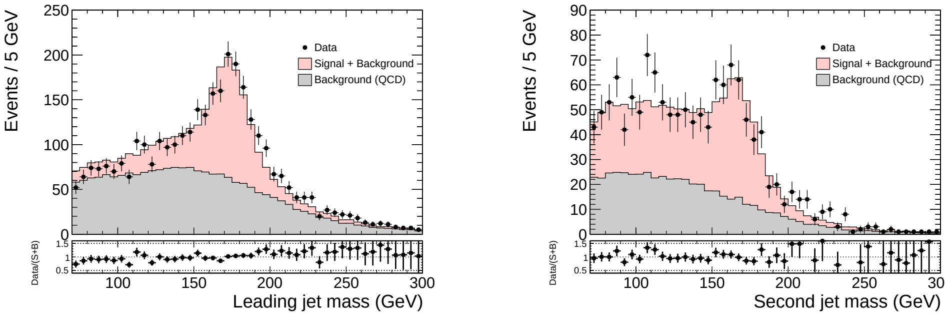

Figure 3.21: Postfit distributions for the softdrop mass of the two leading jets. Signal and background distributions are normalized to the fitted yields.

window $150 < m_{\mathrm{SD}} < 200$ GeV. Figures 3.22 and 3.23 show some other variables of interest such as jet p_{T} and substructure variables. Figure 3.24 shows some additional relevant quantities, including the Fischer discriminant. Finally, Figure 3.25 shows the mass of the two subjets inside of the leading jet. In all cases, the distributions obtained from the sum of N_{sig} signal events and N_{bkg} background events match well those observed in data candidates. A comparison between Figure 3.21 and Figure 3.25 shows that indeed the request $150 < m_{\mathrm{SD}} < 200$ GeV enhances the signal purity of the sample.

3.5.3 Systematic uncertainties

The inclusive measurement is affected by several sources of systematic uncertainties, some of which are estimated here, while others are taken from official results of

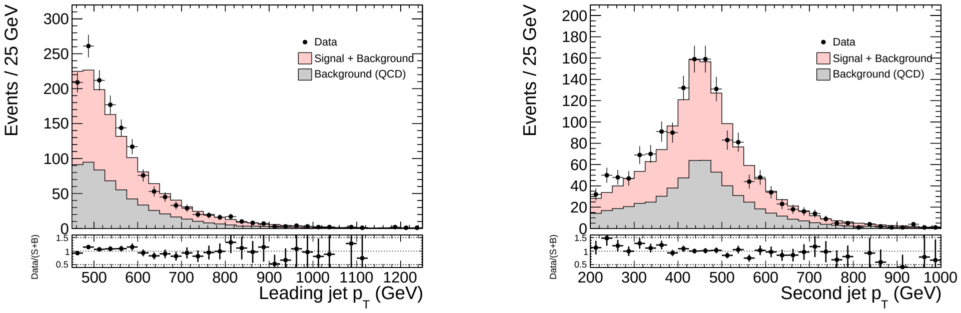

Figure 3.22: Postfit distributions of the leading jet and second jet p_{T}. Signal and background distributions are normalized to the fitted yields in the mass window $150 < m_{\mathrm{SD}} < 200$ GeV.

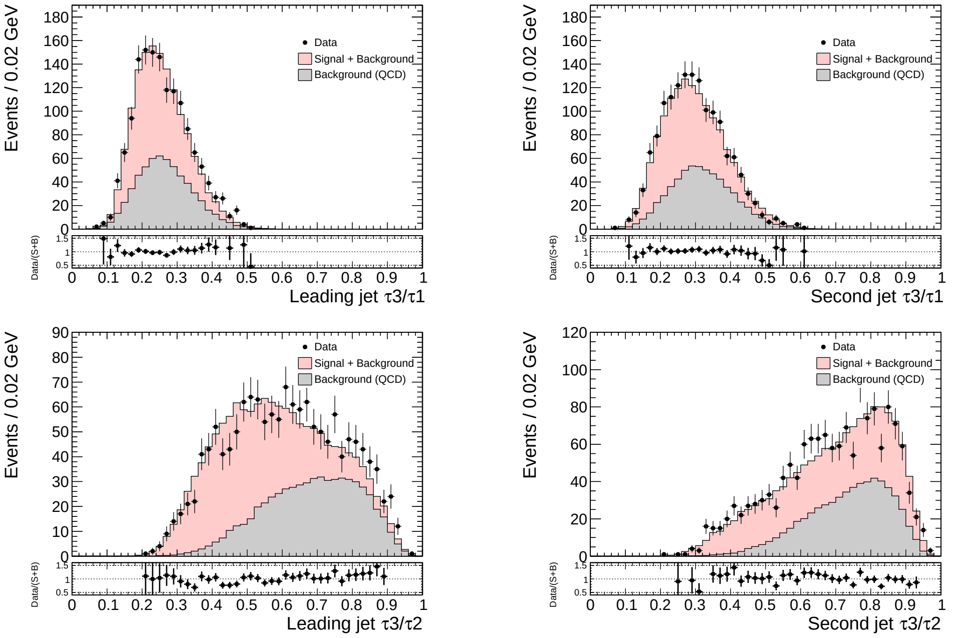

Figure 3.23: Postfit jet substructure distributions. Signal and background distributions are normalized to the fitted yields in the mass window $150 < m_{\mathrm{SD}} < 200$ GeV.

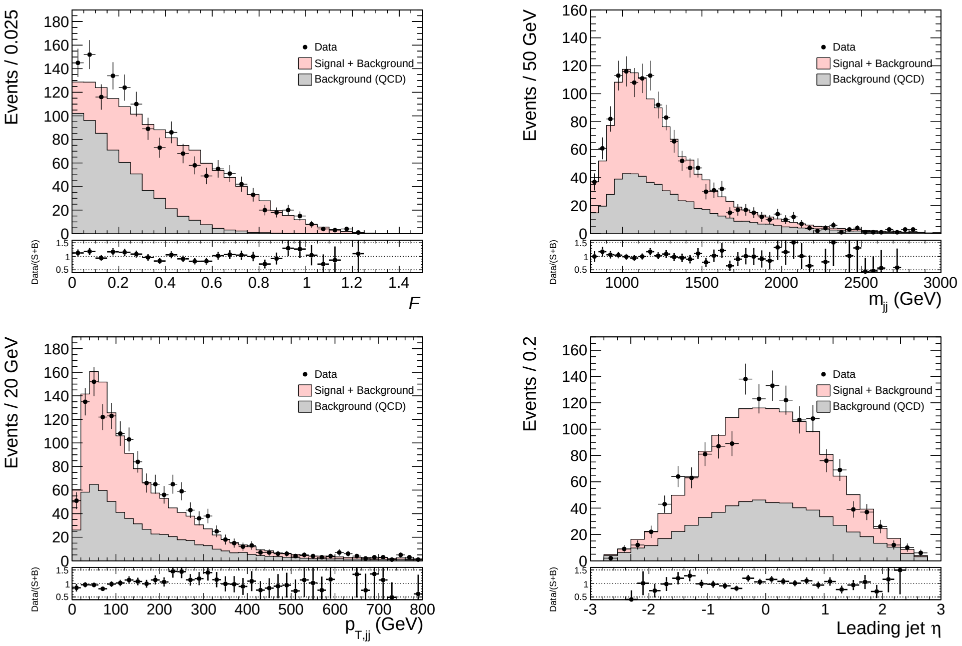

Figure 3.24: Postfit distributions for the Fischer discriminant $\mathcal{F}$, m_{jj}, $p_{\mathrm{T,jj}}$ and leading jet η. Signal and background distributions are normalized to the fitted yields in the mass window $150 < m_{\mathrm{SD}} < 200$ GeV.

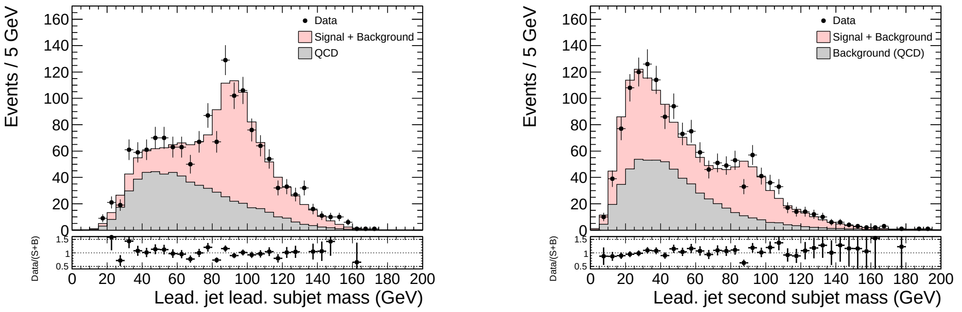

Figure 3.25: Postfit distributions of the mass of the two subjets of the leading jet. Signal and background distributions are normalized to the fitted yields in the mass window $150 < m_{\mathrm{SD}} < 200$ GeV.

the CMS top quark physics analysis group. The process of estimation is quite the same for each source: systematic-affected quantities are used to compute again the cross section (see Subsection 3.5.4); the fractional change in the cross section from the nominal value is then reported as relative uncertainty.

- QCD background modelling: QCD background distributions are obtained from a control sample in data and then corrected with a control-to-signal transition function, which is found to be a straight line, in order to extrapolate them to the signal region. The uncertainty on this prediction is conservatively estimated by shifting up and down the slope of the transition function by a half of its value, and it is found to affect the measurement by 3% at most.
- Subdominant background modelling: the background contributes from single-t/single-$\bar{t}$ production, $W +$ jets production and $Z +$ jets production are found to contribute up to 4% of the signal. This amount is taken as systematic uncertainty.
- Jet energy scale (JES) and jet energy resolution (JER): in the MC samples, jets are shifted and smeared up and down according to the p_{T}- and η-dependent JES and JER uncertainties. The effect of the JER is found to be negligible, while the JES is found to affect the measurement by 2% at most.

Other systematic uncertainties which should be taken into account are the b-tagging efficiency (which is found to be the most relevant source of uncertainty) and some “theoretical” uncertainties related to the choice of PDFs, the modelling of parton showers, details on the MC generator and so on. The impact of each systematic uncertainty is reported in Table 3.4. The global systematic uncertainty is computed as the quadrature sum of the individual contributions.

3.5.4 Inclusive cross section

The inclusive cross section is extracted starting from the fitted signal yield using the standard formula

$$\sigma_{t\bar{t}} = \frac{N_{\mathrm{sig}}}{\epsilon \times L} \tag{3.8}$$

where N_{sig} is the signal yield, ϵ is the efficiency of the selection computed starting from the $t\bar{t}$ simulation using Equation 3.4 and L is the integrated luminosity. In principle, the selection efficiency should be corrected for the trigger efficiency and b-tagging efficiency data-vs-MC scale factors, but it is happens that the two contributions quite balance each other, combining to give a correction factor very close to 1, which is the reason why we do not need to consider this effect. Putting together the values $N_{\mathrm{sig}} = 1350 \pm 86$, $\epsilon = 7.3 \times 10^{-4}$ and $L = 2534\ \mathrm{pb}^{-1}$ we obtain:

$$\sigma_{t\bar{t}} = 727 \pm 46\ (\mathrm{stat.})^{+115}_{-112}\ (\mathrm{sys.}) \pm 20\ (\mathrm{lumi.})\ \mathrm{pb}, \tag{3.9}$$

Table 3.4: Fractional uncertainties on the inclusive $t\bar{t}$ production cross section.

Source	(%)
QCD background modelling	$-2.7, +2.4$
Subdominant backgrounds	± 4.0
Jet energy scale	$-1.8, +1.6$
Jet energy resolution	$\ll 1$
b-jet tagging	$-10.5, +12.9$
Trigger efficiency	$-1.1, +0.9$
Scale (μ_F and μ_R)	$-1.5, +0.0$
PDF	± 1.0
Parton shower	$-7.0, +3.0$
NLO generator	± 7.0
Total systematic unc.	$-15.4, +15.8$
Statistical unc.	± 6.4
Integrated luminosity	± 2.7

a value which is consistent with the theoretical value

$$\sigma_{t\bar{t}}^{\text{th}} = 832^{+20}_{-29}\ (\text{scale}) \pm 35\ (\text{PDF} + \alpha_s)\ \text{pb}$$

as calculated with the Top++ [25] program at next-to-next-to-leading-order (NNLO) in perturbative QCD assuming $\mathrm{m_t} = 172.5$ GeV. This cross section measured in the all-jets channel is also consistent with the other CMS measurements in different channels (dilepton channel [26] and single-lepton channel [27]), and in the same channel but in the resolved topology [28], as shown in Figures 3.26 and 3.27.

3.6 Differential measurement

In this section we compute two different kinds of differential cross section for $t\bar{t}$ production: first, the fiducial cross section or detector-level cross section, which is affected by detector effects; second, we account for those effects in order to obtain a parton-level cross section which can be compared to the theoretical predictions.

3.6.1 Fiducial cross section

The differential $t\bar{t}$ cross section at detector level is computed, for each bin of the spectrum, starting from the following formula:

$$\frac{d\sigma_{t\bar{t}}}{dx} = \frac{S(x)}{\Delta x \cdot L} , \tag{3.10}$$

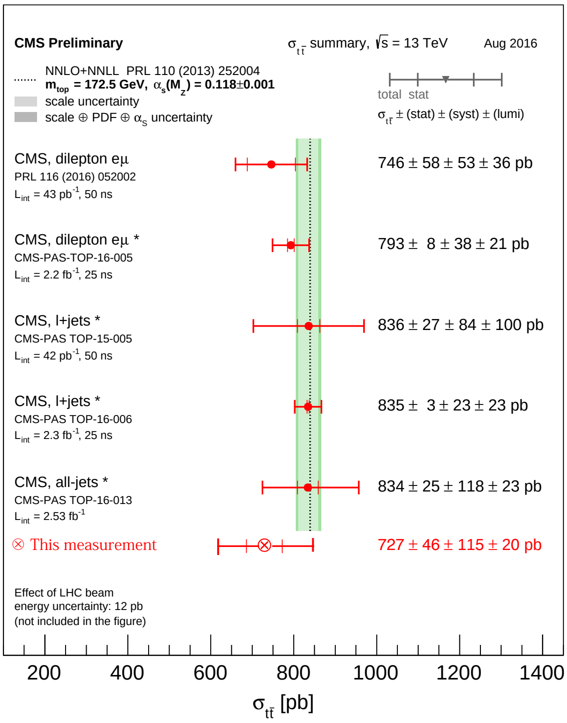

Figure 3.26: Summary of the CMS inclusive $t\bar{t}$ cross section measurements at 13 TeV, along with the measurement performed in this analysis, which is found to be consistent with the other values.

where x is the variable of interest (in our case the $p_{\rm T}$ of the leading jet), Δx is the bin width, L is the integrated luminosity, and $S(x)$ is the signal yield, background subtracted, in each bin. In our case the $S(x)$ is computed as follows:

$$S(p_{\rm T}) = D(p_{\rm T}) - f_{\rm bkg}^{m_{\rm SD}} \cdot N_{\rm bkg} \cdot {\rm TF}_{\rm MC}(p_{\rm T}) \cdot B(p_{\rm T}) \ , \tag{3.11}$$

where $D(p_{\rm T})$ is the $p_{\rm T}$ distribution in data, $N_{\rm bkg}$ is the fitted background yield determined in the inclusive measurement, $B(p_{\rm T})$ is the background distribution from the control sample in data (normalized to unit area), ${\rm TF}_{\rm MC}(p_{\rm T})$ is the transition function from the control to the signal sample (see Figure 3.17), and $f_{\rm bkg}^{m_{\rm SD}}$ is the background fraction in the $m_{\rm SD}$ range $(150-200)$ GeV, defined as:

$$f_{\rm bkg}^{m_{\rm SD}} = \frac{\int_{150}^{200} {\rm TF}_{\rm MC}(m_{\rm SD}) \cdot B_{\rm control}^{\rm data}(m_{\rm SD})\, {\rm d}m_{\rm SD}}{\int_{m_{\rm SD,min}}^{m_{\rm SD,max}} {\rm TF}_{\rm MC}(m_{\rm SD}) \cdot B_{\rm control}^{\rm data}(m_{\rm SD})\, {\rm d}m_{\rm SD}} . \tag{3.12}$$

Basically, what we do here is the following: we take the background distribution in the control sample $B(p_{\rm T})$, normalized to unit area, and correct it with the proper transition function ${\rm TF}_{\rm MC}(p_{\rm T})$ in order to obtain the same distribution in the signal

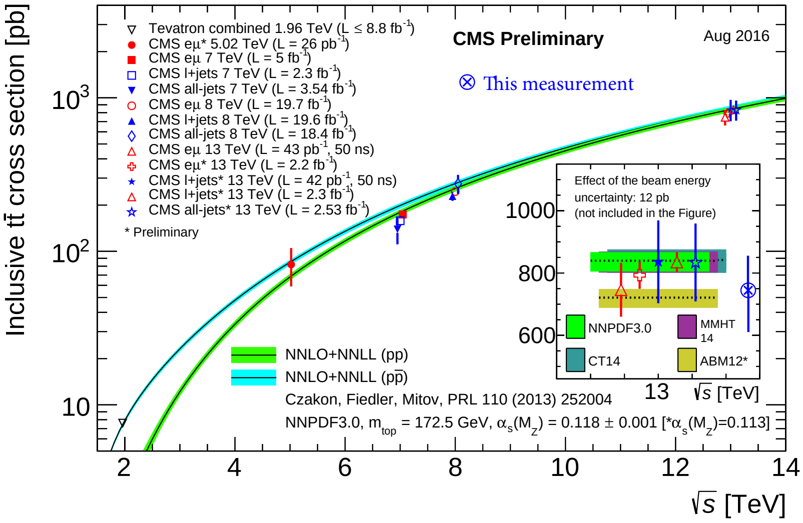

Figure 3.27: Evolution of the inclusive $t\bar{t}$ cross section as a function of the centre-of-mass energy, along with previous measurements at lower energies and current measurements at 13 TeV.

sample. We then normalize the resulting distribution to the fitted background yield in the m_{SD} window (150 – 200 GeV) and subtract this final distribution from the data distribution $D(p_{\mathrm{T}})$ to obtain $S(p_{\mathrm{T}})$.

The background fraction is found to be $f_{\mathrm{bkg}}^{m_{\mathrm{SD}}} = 0.31$. The resulting differential $t\bar{t}$ cross section at detector level is reported in Figure 3.28, where a comparison with the POWHEG prediction is shown. A good agreement is found, but the measured spectrum appears to be softer than the simulated one, that is it shows a distribution falling more steeply. However, the measurement is statistically-limited above ~ 600 GeV and should improve with new data.

3.6.2 Detector effects - Unfolding

The differential result we just obtained is expressed in terms of reconstructed quantities, such as jet p_{T}. This means that it is somehow influenced by the design of the detector, by its geometry, its resolution and so on. Thus, it is important to find a way to remove these detector effects and make possible the comparison between results coming from different experiments, along with the comparison with theoretical calculations.

For the measured observables there is a bin-to-bin migration effect due to the finite experimental resolution. Events where a top quark has been produced/generated in a given p_{T} bin might be reconstructed/measured in a different bin. These bin-to-bin migrations are characterized by two quantities: the stability

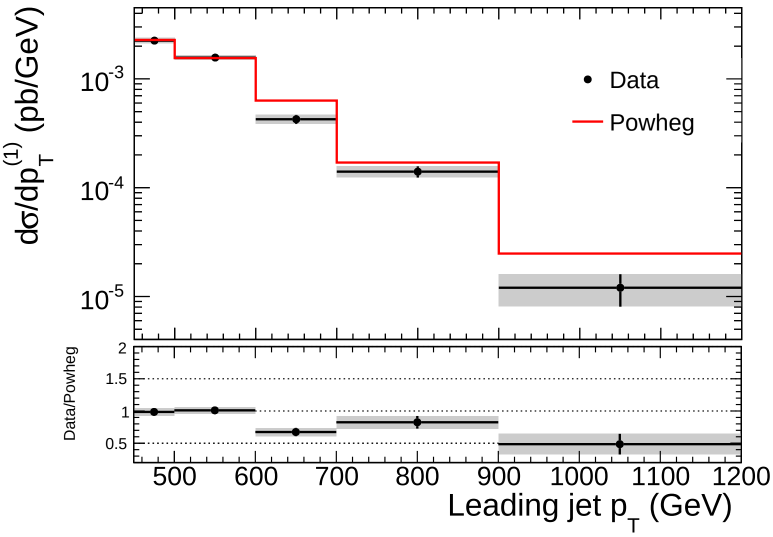

Figure 3.28: Differential $t\bar{t}$ cross section at detector level, along with the prediction of the POWHEG simulation. The bottom panel shows the ratio between the measurement and the expected value The shaded regions represent the statistical uncertainties.

s^i and purity p^i, defined as:

$$s^i = \frac{N^i_{\text{gen\&rec}}}{N^i_{\text{gen}}} \ ,$$

and

$$p^i = \frac{N^i_{\text{gen\&rec}}}{N^i_{\text{rec}}} \ ,$$

namely, the stability s^i denotes the ratio between the number of events generated and correctly reconstructed in a given bin i and the events generated in that bin but reconstructed anywhere, while the purity p^i is given by the ratio between the number of events generated and correctly reconstructed in a given bin i and the events reconstructed in that bin but generated anywhere. These two quantities are determined by the simulation of the $t\bar{t}$ signal. In the absence of bin-to-bin migrations, both purity and stability would be 1, but resolution effects lead to migrations outside the bin and to a reduction in the values of purity and stability.

To account and correct for bin-to-bin migration, a regularized unfolding procedure is applied to the differential measured distributions. The unfolding is implemented via RooUnfold [29], the ROOT unfolding framework, using the iterative or Bayesian algorithm proposed by D'Agostini [30]. The known effects of detection efficiencies, measurements resolutions and systematic biases are taken into account using a response matrix, derived from signal simulation, which maps

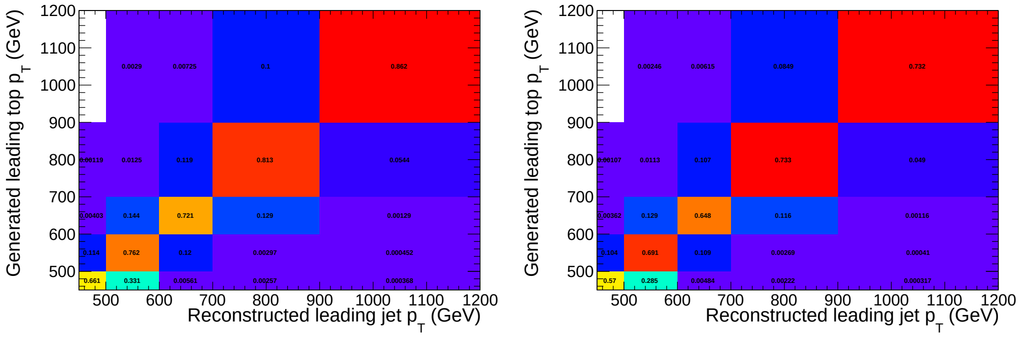

Figure 3.29: Response matrices for simulated $t\bar{t}$ events, having stabilities (left) and purities (right) in the diagonal bins.

the reconstructed differential distribution into the generated differential distribution, each element of this matrix representing the number of events generated and reconstructed in a specific p_{T} range. Such a matrix is non-diagonal and is used by the unfolding algorithm to reconstruct the generated distribution starting from the measured one, removing detector effects. The response matrix is shown in Figure 3.29. Each bin is normalized by row (by column), in such a way that diagonal bins contain stabilities (purities). The binning is chosen in order to obtain stabilities and purities greater than about 60%. We do not take very much care of the stability and purity of the first bin ($450 < p_{\mathrm{T}} < 500$) since this bin is used as a reference point only and the unfolded spectrum will be extrapolated for a parton p_{T} greater than 500 GeV.

In order to validate the unfolding procedure, a closure test is implemented. In Figure 3.30 we show the ratio between the unfolded and the generated p_{T} spectra. Since the ratio is equal to 1 for each bin, the closure test validates the Bayesian unfolding. Implementation of the Bayesian unfolding algorithm requires the use of a regularization parameter which must be optimized for the number of bins in the distributions and the size of the sample. The regularization parameter of the Bayesian algorithm has been optimized evaluating, for each bin in the spectrum, the ratio between the relative error in the unfolded spectrum and the relative error in the reconstructed spectrum. It should be noticed that the unfolded spectrum is built up by the unfolding procedure starting from the reconstructed spectrum, considering the contribution of bin-to-bin migrations. We thus expect the errors in the bins of the unfolded spectrum to be greater than the errors in the corresponding bins of the measured spectrum, since we have here additional sources of uncertainty coming from the migrations from neighbouring bins. The best value of the regularization parameter thus depends on the number of bins in the spectra and is the first integer for which all the ratios are greater than 1. These ratios are shown in Figure 3.30 for the best value of the regularization parameter,

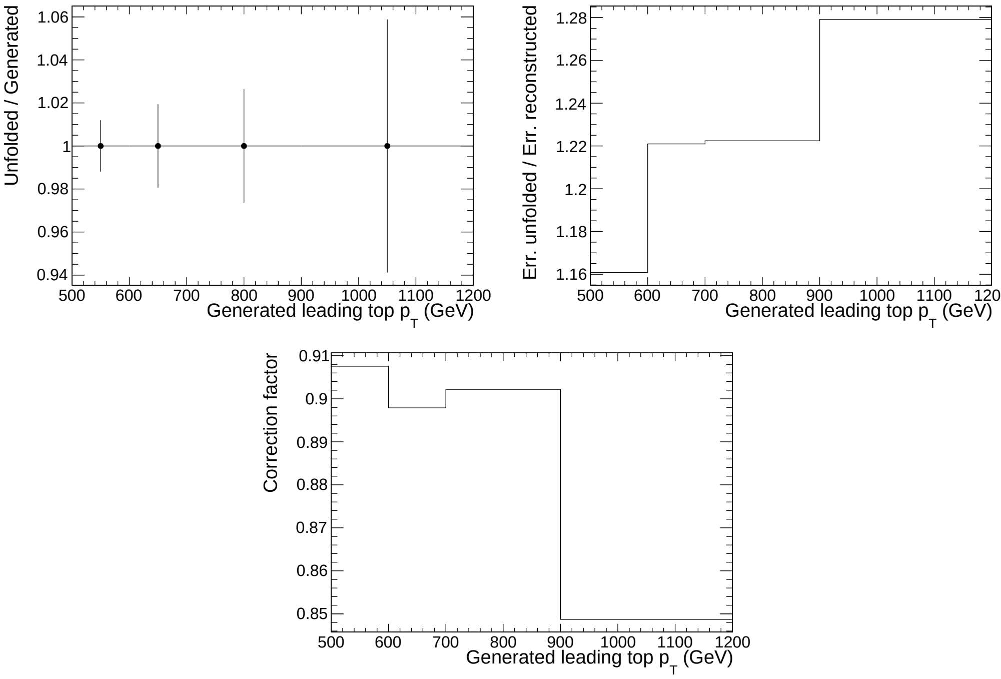

Figure 3.30: Ratio of the unfolded and generated p_{T} spectra (upper left), error ratio (upper right) and bin-to-bin correction factor (bottom) for simulated $t\bar{t}$ events.

which is found to be 3.

Finally, Figure 3.30 also shows the so called bin-to-bin correction factor, namely the ratio between the unfolded and reconstructed p_{T} spectra. Notice that in the boosted regime, since we are dealing with a falling spectrum, for each two close-lying bins the left bin will always contain a number of events greater than the right one. Therefore, because of bin-to-bin migrations, the number of events which will enter the right bin from the left will be greater than the number of events that leave it. This explains why the bin-to-bin correction factor is always less than one.

Once the unfolding procedure is validated, the unfolding is performed on the distribution $S(p_{\mathrm{T}})$ in Equation 3.11 to obtain the unfolded signal yield.

3.6.3 Acceptance correction

As we stated above, we are now dealing with quantities which do not depend upon the detector design. For this reason, it must be pointed out that we also have to discard effects from the event selection (which of course involves measured quantities). An acceptance correction is then introduced in order to extrapolate the leading top quark p_{T} at parton level from the fiducial space of the measurement

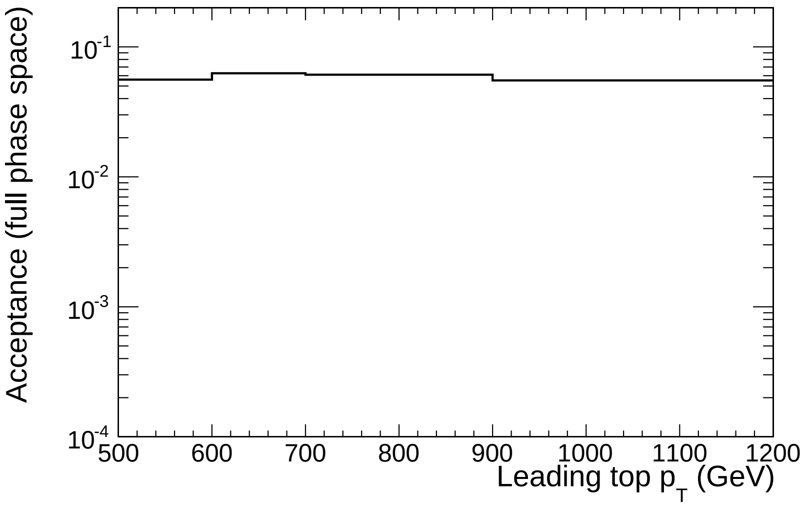

Figure 3.31: Acceptance correction factor.

to the full phase space. This correction is defined, starting from the $t\bar{t}$ simulation, as the ratio between the leading top quark p_{T} after the selection and the same quantity prior to any selection. Figure 3.31 shows this acceptance correction factor for each bin of the unfolded spectrum.

3.6.4 Parton-level cross section

The unfolded differential cross section at parton level, extrapolated to the full phase space is computed in the following way:

$$\frac{d\sigma}{dx_{\text{parton}}} = \frac{S_{\text{unfolded}}(x_{\text{parton}})}{L \cdot \mathcal{A}(x_{\text{parton}}) \cdot \Delta x_{\text{parton}}}, \tag{3.13}$$

where $x_{\text{parton}} = p_{\mathrm{T,top}}^{(1)}$ is the p_{T} of the leading top quark, $S_{\text{unfolded}}(x_{\text{parton}})$ is the unfolded signal yield, $\mathcal{A}(x_{\text{parton}})$ is the acceptance correction, and Δx_{parton} is the bin width. Figure 3.32 shows the parton level cross section, extrapolated to the full phase space, as a function of the leading parton p_{T}, and compared to the prediction of MadGraph5_aMC@NLO [31]. As in the case of the fiducial cross section, the parton-level spectrum is found to be softer than the prediction. In any case the boosted-topology approach allows to reach and pass the TeV region for the characterization of the top quark p_{T} spectrum, extending the range explored with the resolved analysis, as shown in Figure 3.33 [28].

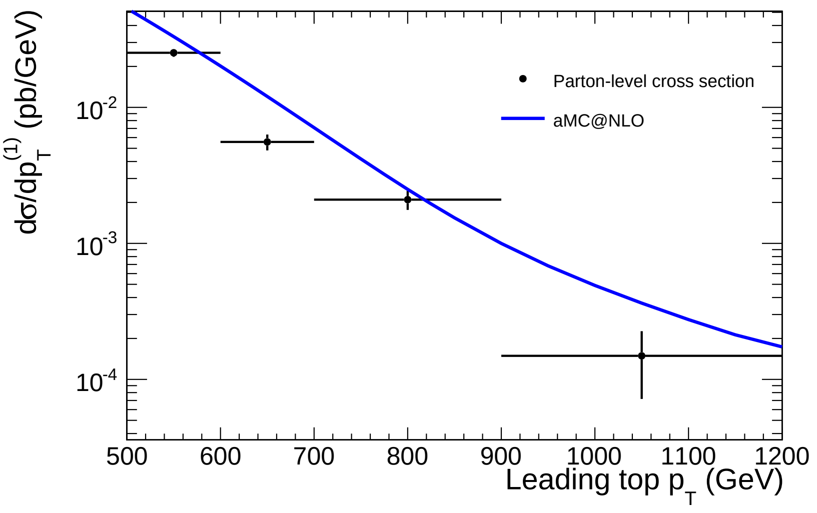

Figure 3.32: Differential $t\bar{t}$ cross section at parton level, extrapolated to the full phase space, along with the prediction of MadGraph5_aMC@NLO.

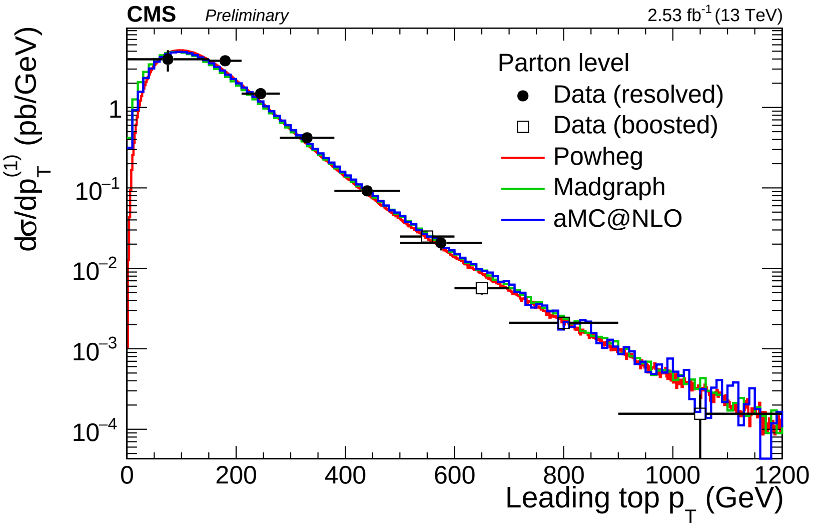

Figure 3.33: Unfolded differential cross section, extrapolated to the full phase space, as a function of the leading quark p_{T} (resolved and boosted analyses).

Chapter 4

Conclusions

The $t\bar{t}$ production cross section has been measured both inclusively and differentially starting from a data sample corresponding to an integrated luminosity of 2.53 fb^{-1} collected in the year 2015 by the CMS experiment.

First of all, an event selection has been applied to the data sample in order to reject background events and enhance the purity of the sample. This selection, based on the study of MC simulated events passing the trigger `HLT_AK8PFJet360_TrimMass30`, required the presence of at least two AK8 jets with $p_{\mathrm{T}} > 200$ GeV, $|\eta| < 2.4$, and a softdrop mass $m_{\mathrm{SD}} > 50$ GeV, with the leading jet p_{T} greater than 450 GeV; moreover, we asked for a lepton veto, thus selecting specifically the all-jets final state. We then requested the two AK8 jets to be b-tagged, in order to select event containing top quark decay products, and calculated a Fisher discriminant $\mathcal{F}$ starting from the n-subjettiness ratios τ_3/τ_2 and τ_3/τ_1 of the two leading jets; the ideal cut on this variable was found to be $\mathcal{F} > 0$. Finally, for the differential measurement only, we asked the softdrop mass of the leading jet to lie in the range $150 - 200$ GeV.

After a comparison between data and simulations, it turned out from the τ_3/τ_2 distribution of the leading jet that the QCD modelling was not very accurate, so that a new background modelling method was needed. Since the presence of b-tagged subjets is a strong characteristic of top quark decays, we chose to define a control sample, QCD-enriched, by asking the usual selection combined with the request of exactly no b-tagged subjets. The transition between control (0 b-tag) and signal (2 b-tag) samples was made using transition functions obtained from the MC simulations.

Once the signal and background templates were available, maximum-likelihood fits were performed in order to obtain the signal and background yields, the former being used to compute the value of the inclusive cross section, which turned out to be $\sigma_{t\bar{t}} = 727 \pm 46$ (stat.)$^{+115}_{-112}$ (sys.) $\pm$ 20 (lumi.) pb, a value consistent with the theoretical predictions, with the values obtained in the other decay channels and with the values obtained in the same channel, but in the resolved topology. After that, we computed the differential, detector-level cross section and compared it with the prediction of the simulations, finding an overall good agreement, but with the measured spectrum being softer than the simulated one. Finally, an unfolding procedure was implemented in order to account for detector effects (such as detection efficiencies, measurements resolutions and systematic biases) and obtain a differential, parton-level cross section, extrapolated to the full phase space with the use of a proper acceptance correction. Again, the measured spectrum is compared to theoretical calculations and it is found to be softer than the predicted one.

This work and the details of the measurements have been included into an official CMS analysis note [32] and have been made public as a CMS Physics Analysis Summary [28]. My major contributions to this complex data analysis concerned the characterization of the events in the boosted topology, focusing on the calculation of quantities such as the top quark reconstruction efficiency of the CMS detector and the purity of the sample, a study based of the simulated $t\bar{t}$ events. I also set up the implementation of all the steps of the unfolding procedure which led to the parton-level differential cross section, starting from the calculation of response matrices up to the computation of unfolded spectra.

The techniques applied in this measurement, and the results here obtained, will serve as the ground from which to develop the future measurements based on the data samples to be collected in 2016, which are going to contain about 10 times more events and are going to be free from statistical limitation that sometimes affected this measurement.

Bibliography

[1] CDF Collaboration, *Observation of top quark production in $p\bar{p}$ collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (1995) 2626, `doi:10.1103/PhysRevLett.74.2626`, `arXiv:hep-ex/9503002`.

[2] D0 Collaboration, *Observation of the Top Quark*, Phys. Rev. Lett. **74** (1995) 2632, `doi:10.1103/PhysRevLett.74.2632`, `arXiv:hep-ex/9503003`.

[3] G. Degrassi et al., *Higgs mass and vacuum stability in the Standard Model at NNLO*, JHEP **1208** (2012) 098, `doi:10.1007/JHEP08(2012)098`, `arXiv:1205.6497`.

[4] L. Evans and P. Bryant, *LHC machine*, JINST **03** (2008) S08001, `doi:10.1088/1748-0221/3/08/S08001`.

[5] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, `doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[6] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, `doi:10.1016/j.physletb.2012.08.021`, `arXiv:1207.7235`.

[7] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004 `doi:10.1088/1748/0221/3/08/S08004`.

[8] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, `doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092`.

[9] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, `doi:10.1007/JHEP06(2010)043, arXiv:1002.2581`.

[10] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, `doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175`.

[11] T. Sjöstrand, S. Mrenna and P. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `doi:10.1016/j.cpc.2008.01.036`, `arXiv:0710.3820`.

[12] K. A. Olive et al., Particle Data Group, Chin. Phys. C, **38**, 090001 (2014) and 2015 update.

[13] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going Beyond*, JHEP **06** (2011) 128 `doi:10.1007/JHEP06(2011)128`, `arXiv:1106.0522v1`.

[14] J. Allison et al., *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270, `doi:10.1109/TNS.2006.869826`.

[15] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_T clustering algorithm*, JHEP **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[16] CMS Collaboration, *Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}*, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[17] 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.

[18] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, JHEP **1103** (2011) 015, `doi:10.1007/JHEP03(2011)015`, `arXiv:1011.2268`.

[19] A. Larkoski, S. Marzani, G. Soyez and T. Jesse, *Soft Drop*, JHEP **05** (2014) 146, `doi:10.1007/JHEP05(2014)146`, `arXiv:1402.2657`.

[20] CMS Collaboration, *Identification of b quark jets at the CMS experiment in the LHC Run2 Startup*, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.

[21] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework`.

[22] The CMS Trigger and Data Acquisition Group, *The CMS High Level Trigger*, Eur. Phys. J. **C46**, (2006) 605 `doi:10.1140/epjc/s2006-02495-8`, `arXiv:hep-ex/0512077`.

[23] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, JHEP **1002** (2010) 084 `doi:10.1007/JHEP02(2010)084`, `arXiv:0912.1342`.

[24] Toolkit for Data Modeling with ROOT (RooFit), available at `root.cern.ch/roofit`.

[25] M. Czakon and A. Mitov, *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. 185 (2014) 2930, `doi:10.1016/j.cpc.2014.06.021`, `arXiv:1112.5675`.

[26] CMS Collaboration, *Measurement of the top quark pair production cross section using $e\mu$ events in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS detector*, CMS Physics Analysis Summary, CMS-PAS-TOP-16-005, 2016.

[27] CMS Collaboration, *Measurement of the inclusive and differential $t\bar{t}$ production cross section in lepton + jets final states at 13 TeV*, CMS Physics Analysis Summary, CMS-PAS-TOP-16-008, 2016.

[28] CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section at 13 TeV in the all-jets final state*, CMS Physics Analysis Summary, CMS-PAS-TOP-16-013, 2016.

[29] T. Adye, *Unfolding algorithms and tests using RooUnfold*, `doi:10.5170/CERN-2011-006.313`, `arXiv:1105.1160`.

[30] G. D'Agostini, *A multidimensional unfolding method based on Bayes theorem*, Nucl. Instrum. Meth. A **362** (1995) 487, `doi:0.1016/0168-9002(95)00274-X`.

[31] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079, `doi:10.1007/JHEP07(2014)079` `arXiv:1405.0301`.

[32] K. Kousouris, A. Castro and F. Iemmi, *Measurement of the $t\bar{t}$ cross section at 13 TeV in the all-jets final state*, CMS Analysis note, CMS-AN-2016-106.

List of Tables

3.1 Monte Carlo samples. 16
3.2 Summary of the selection for the signal sample. 20
3.3 Summary of the selection for the control sample. 30
3.4 Relative uncertainties on the inclusive $t\bar{t}$ cross section. 38

List of Figures

1.1 Dominant top quark production mechanisms at the Tevatron and at the LHC. 1
1.2 Radiative corrections to the W boson propagator. 2
1.3 Production and decay of a T' particle. 3
1.4 Jet size as a function of the top quark p_{T}. 5
1.5 Resolved and boosted topologies. 5

2.1 Schematic layout of the LHC. 9
2.2 Cross section of a cryodipole. 10
2.3 The acceleration system at CERN. 11
2.4 Overall layout of CMS. 12
2.5 CMS coordinate system. 13

3.1 n-subjettiness for the leading jet and the second jet in simulations. 18
3.2 Trigger efficiency and data/MC scale factor. 21
3.3 Softdrop mass m_{SD} for the leading jet before and after the request on subjets. 22
3.4 Output of the multivariate training. 23
3.5 Decay mode of simulated $t\bar{t}$ after the selection. 23
3.6 Sample purity after the selection. 24
3.7 Leading jet and second jet softdrop mass. 25
3.8 Jet properties. 26
3.9 Jet and b-jet multiplicities. 26
3.10 H_T and Fisher discriminant output. 27
3.11 Event properties. 27
3.12 Mass of the two subjets of the leading jet. 28
3.13 Substructure properties of the two leading jets in the signal sample. 28
3.14 Substructure properties of the two leading jets in the control sample. 29
3.15 Fischer discriminant in the control sample. 29
3.16 Closure of the background prediction for the leading jet m_{SD}. . . . 31
3.17 Closure of the background prediction for the leading jet p_{T}. . . . 31
3.18 m_{SD} signal template. 33
3.19 m_{SD} control template. 33

3.20 Fitted m_{SD} distribution. 34
3.21 Postfit distributions of the leading jet and second jet mass. 34
3.22 Postfit distributions of the leading jet and second jet p_{T}. 35
3.23 Postfit jet substructure distributions. 35
3.24 Postfit distributions of some variables of interest. 36
3.25 Postfit mass of the two subjets of the leading jet. 36
3.26 CMS inclusive cross section measurements. 39
3.27 Evolution of the inclusive cross section as a function of the centre-of-mass energy. 40
3.28 Detector-level differential cross section. 41
3.29 Response matrices. 42
3.30 Closure test for the Bayesian unfolding, error ratio and bin-to-bin correction factor. 43
3.31 Acceptance correction. 44
3.32 Parton-level differential cross section. 45
3.33 Unfolded differential cross section. 45