

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Corso di Laurea in Fisica

Characteristics of multijet $t\bar{t}$ events
reconstructed by CMS in pp collisions
at 13 TeV

Relatore:
Prof. Andrea Castro

Presentata da:
Federico Celli

Correlatore:
Dott. Giuseppe Codispoti

Sessione II
Anno Accademico 2014/2015

Desidero ringraziare il Professor Andrea Castro per avermi dato la possibilità di partecipare a questo progetto e per avermi seguito con grande disponibilità e pazienza in ogni fase del lavoro di tesi.

Un grande ringraziamento va, inoltre, al Dottor Giuseppe Codispoti per l'aiuto e gli utili consigli che mi ha fornito, anche dal punto di vista tecnico.

Grazie alla mia famiglia, ai miei amici e ai compagni di corso che mi hanno consigliato e sostenuto in questi anni e a Federica per essermi accanto in ogni momento.

Contents

1	Introduction	1
2	Physics at the LHC	2
2.1	The Large Hadron Collider	2
2.1.1	Generalities	2
2.1.2	The CERN complex	3
2.1.3	The LHC machine	4
2.2	The Compact Muon Solenoid	6
2.2.1	Generalities	6
2.2.2	Structure	6
2.3	High- P_T physics	9
3	The top quark	12
3.1	Top quark discovery	12
3.2	Main properties of the top quark	14
3.3	Why is top quark mass so important?	14
3.4	Production and decay	15
3.4.1	Production	15
3.4.2	Decay	18
4	$t\bar{t}$ signal description	20
4.1	Monte Carlo simulations	20
4.2	Data analysis	20
4.3	Mass reconstruction	23
5	Comparison between multijet events at 8 and 13 TeV	25
5.1	Efficiency of the standard selection	28
5.2	Mass distributions	29
5.3	Possible improvements	32
6	Conclusions	39

Sommario

Presentiamo qui una caratterizzazione dei campioni Monte Carlo usati a CMS nel run attuale di LHC (Run 2, $\sqrt{s} = 13$ TeV) e li confrontiamo con quelli utilizzati nel run precedente (Run 1, $\sqrt{s} = 8$ TeV). Usiamo poi questi campioni per ricostruire la massa del quark top dai prodotti di decadimento totalmente adronico e confrontiamo le efficienze del metodo di ricostruzione standard quando applicato ai due differenti campioni. Infine troviamo un modo di migliorare l'efficienza della ricostruzione sui campioni a 13 TeV usando jets ricostruiti con un differente algoritmo, l'algoritmo Cambridge-Aachen.

Abstract

We present here a characterization of the Monte Carlo samples used at CMS in the current LHC run (Run 2, $\sqrt{s} = 13$ TeV) and we compare them to the ones used in the previous run (Run 1, $\sqrt{s} = 8$ TeV). We then use these samples to reconstruct the top quark mass from the all-hadronic decay products and we compare the efficiencies of the standard reconstruction method when applied to the two different samples. We finally find a way to improve the efficiency for 13 TeV samples by using jets reconstructed with a different algorithm, the Cambridge-Aachen algorithm.

Chapter 1

Introduction

The top quark is the latest discovered quark and nowadays it is one of the fundamental particles of the Standard Model. The peculiar properties deriving from its huge mass, make the top quark a very interesting subject of studies. For instance, the large value of its mass makes the top quark contribution dominant in loop corrections to many observables, like the W boson mass. Also, precise measurements of the W boson and the top quark masses allow to set indirect constraints on the mass of the Higgs boson.

Top quarks are produced at the *Large Hadron Collider* (LHC), mainly in pairs ($t\bar{t}$), from proton-proton high energy collisions. Thanks to their large mass, top quarks decay before hadronizing in a bottom quark and a W boson. In this work we will study the so called *all-hadronic* decay, namely the case in which both W bosons coming from a $t\bar{t}$ state, decay in a couple of quarks each (which is the most likely decay process for a $t\bar{t}$ state). This all-hadronic final state is therefore characterized by the production of at least six jets, with at least two of them coming from the bottom quark hadronization.

In order to evaluate the consistency of the theory, top quark events need to be compared to Standard Model predictions. The most powerful way to generate these simulated samples are Monte Carlo (MC) methods, computational algorithms which rely on repeated random sampling to obtain numerical results.

The aim of this document is to characterize the MC samples used in the current LHC run (Run 2, $\sqrt{s} = 13$ TeV), studying their efficiency for the top quark mass reconstruction with respect to the samples used in the previous run (Run 1, $\sqrt{s} = 8$ TeV).

In the final part of this work, we present a different way to compute the top quark mass, improving the efficiency of the standard mass reconstruction method. This method is based on the use of jets reconstructed with a different clustering algorithm, the Cambridge-Aachen algorithm (CA_jets).

Chapter 2

Physics at the LHC

2.1 The Large Hadron Collider

2.1.1 Generalities

The LHC is a superconducting accelerator [1, 2] and collider installed in a 27 km long circular tunnel buried about 100 m underground in the border of France and Switzerland, near the city of Geneva. It first started up on 10 September 2008, and remains the latest addition to the CERN accelerator complex, the European Organization for Nuclear Research.

Nowadays LHC is the world largest and most powerful particle accelerator producing collisions between protons with a record center-of-mass energy of $\sqrt{s} = 13$ TeV, or lead ions.

Inside the accelerator, two high-energy particle beams travel at close to the speed of light before they are made to collide. The beams travel in opposite directions in separate beam pipes, two tubes kept at ultrahigh vacuum, and they are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets. The electromagnets are built from coils of special electric cable that operates in a superconducting state, efficiently conducting electricity without resistance or loss of energy. This requires chilling the magnets to -271.3 °C, a temperature colder than outer space. For this reason, much of the accelerator is connected to a distribution system of liquid helium, which cools the magnets.

All the controls for the accelerator, its services and technical infrastructure are housed under one roof at the CERN Control Center. From there, the beams inside the LHC are made to collide at four locations around the accelerator ring, corresponding to the positions of four particle detectors: ATLAS, CMS, ALICE and LHCb.[2]

2.1.2 The CERN complex

LHC is not the only accelerator at CERN: the CERN complex is composed of various machines and accelerating rings which have different power. Each machine injects the particle beam into the next one, which takes over to bring the beam to a higher energy. The particles at the end of this process enter into LHC where they are accelerated to the maximum energy.

Most of the accelerators in the CERN complex have their own experimental halls, where the beams are used for experiments at different energies.[1]

A simple schematic of the CERN complex is shown in Fig. 2.1.

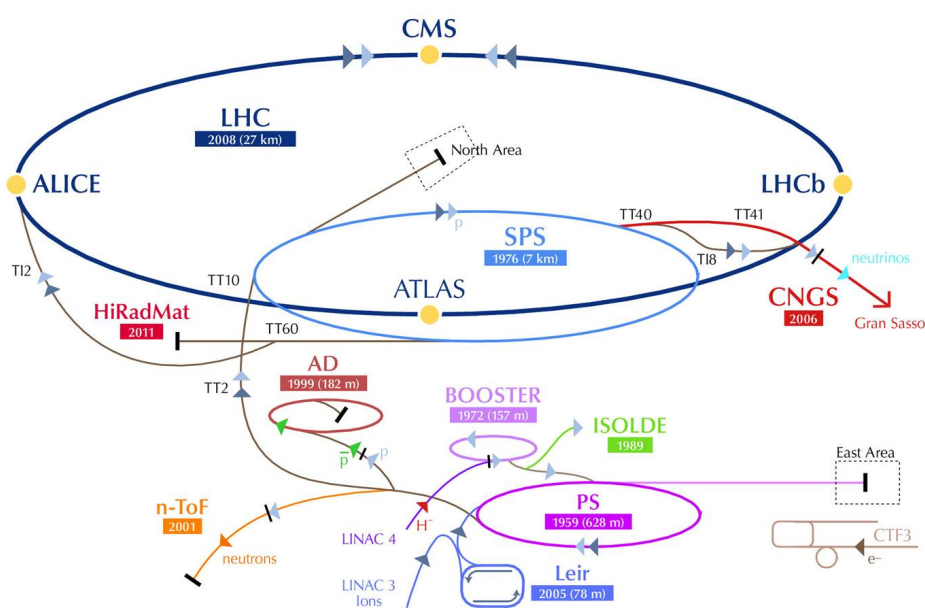


Figure 2.1: CERN accelerator complex.

The process of accelerating protons at CERN starts from a bottle of hydrogen. Protons are extracted from hydrogen atoms by stripping orbiting electrons thanks to a strong electric field. Protons are then injected into the PS Booster (PSB) at an energy of 50 MeV from Linac2. The booster accelerates them to 1.4 GeV. The beam is then fed to the Proton Synchrotron (PS) where it is accelerated to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. They are finally transferred to the LHC (both in a clockwise and an anticlockwise direction) where they reach the maximum energy.

The complex can accelerate not exclusively protons but also lead ions, which are produced from a highly purified lead sample heated to a temperature of about 500 °C. The

lead vapour is ionized by an electron current. Many different charge states are produced with a maximum around Pb^{29+} . These ions are selected and accelerated to 4.2 MeV/u (energy per nucleon) before passing through a carbon foil, which strips most of them to Pb^{54+} .

The Pb^{54+} beam is accumulated, then accelerated to 72 MeV/u in the Low Energy Ion Ring (LEIR), which transfers it to the PS. The PS accelerates the beam to 5.9 GeV/u and sends it to the SPS after first passing it through a second foil where it is fully stripped to Pb^{82+} . The SPS accelerates it to 177 GeV/u and finally sends it to the LHC.[1]

2.1.3 The LHC machine

Structure The LHC ring is made of eight arcs and eight “insertions”. The arcs contain the dipole bending magnets, with 154 magnets in each arc. Their aim is to bend the beams using a strong magnetic field so that the particles can fly in the almost circular orbit of the LHC ring.

An insertion consists of a long straight section plus two transition regions (one at each end), the so-called “dispersion suppressors”. The exact layout of the straight section depends on the specific use of the insertion: physics (beam collisions within an experiment), injection, beam dumping, beam cleaning.

A sector is defined as the part of the machine between two insertion points. The eight sectors are the working units of the LHC: the magnet installation happens sector by sector, the hardware is commissioned sector by sector and all the dipoles of a sector are connected in series and are in the same continuous cryostat. Powering of each sector is essentially independent.

An octant starts from the middle of an arc and ends in the middle of the following arc and thus spans a full insertion. Therefore, this description is more practical when we look at the use of the magnets to guide the beams into collisions or through the injection, dumping, and cleaning sections.[1]

A simple schematic of the LHC structure is shown in Fig. 2.2.

Vacuum LHC has three vacuum systems made up to handle three different tasks: insulation vacuum for cryomagnets, insulation vacuum for the helium distribution line, beam vacuum.

Since their only aim is insulation, the first two systems do not provide a vacuum as high as the last system. The beam vacuum instead has to be very high, 10^{-13} atm (ultrahigh vacuum), because we want to avoid collisions between the beam particles and the gas in the beam pipes.

Magnets There is a large variety of magnets in the LHC, including dipoles, quadrupoles, sextupoles, octupoles, decapoles, etc. giving a total of about 9600 magnets. Each type of magnet contributes to optimizing the beam trajectory: the dipoles bend the beam in

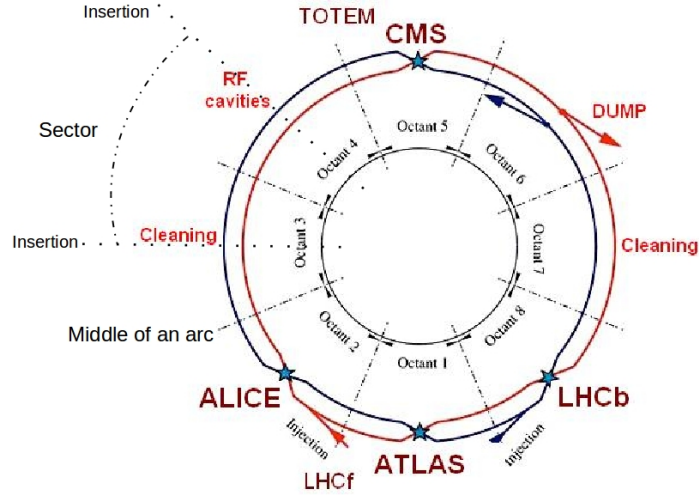


Figure 2.2: LHC layout.

the correct direction along the LHC ring, the other multipoles focus the beam reducing its transverse section to increase the interaction probability during the collisions.

The dipoles of the LHC represented the most important technological challenge for the LHC design. Each dipole is 15 m long and weighs around 35 t. In a proton accelerator like the LHC, the maximum energy that can be achieved is directly proportional to the strength of the dipole field, given a specific acceleration circumference. At the LHC the dipole magnets are superconducting electromagnets and able to provide the very high field of 8.33 T over their length. No practical solution could have been designed using “warm” magnets instead of superconducting ones.

The LHC dipoles use coils made of niobium-titanium (NbTi) cables, which become superconducting below a temperature of 10 K (-263.2 °C), that is, they conduct electricity without resistance. In fact, the LHC will operate at 1.9 K (-271.3 °C), which is even lower than the temperature of outer space (2.7 K or -270.5 °C). A current of 11850 A flows in the dipoles, to create the high magnetic field of 8.33 T required to bend the beam.

The temperature of 1.9 K (-271.3 °C) is reached by pumping superfluid helium into the magnet systems.

Cavities The main role of the LHC cavities is to keep the proton bunches, which constitute the beam, tightly bunched to ensure high luminosity at the collision points and hence, maximize the number of collisions. Each bunch contains about 10^{11} protons and measures a few centimetres in length and a millimetre in width when far from the

collision points. However, as they approach the collision points, they are squeezed to about $16 \mu\text{m}$ in width to allow for a greater chance of proton-proton collisions. Increasing the number of bunches is one of the ways to increase the luminosity \mathcal{L} in a machine, namely the number which, multiplied by the total cross section, gives the total number of collisions per unit time. At full luminosity the LHC uses a bunch spacing of 25 ns (or about 7 m) which corresponds to a frequency of 40 MHz. However, for practical reasons there are several bigger gaps in the pattern of bunches which leads to a frequency of 31.6 MHz.

The cavities also deliver radiofrequency (RF) power to the beam during acceleration to the top energy. Protons can only be accelerated when the RF field has the correct orientation when particles pass through an accelerating cavity, which happens at well specified moments during an RF cycle. The LHC will use eight cavities per beam, each delivering 2 MV (an accelerating field of 5 MV/m) at 400 MHz. The cavities will operate at 4.5 K (-268.7 °C).

2.2 The Compact Muon Solenoid

2.2.1 Generalities

The Compact Muon Solenoid (CMS) [3] is one of the six detectors installed at the LHC. The other five detectors are: A Large Ion Collider Experiment (ALICE), A Toroidal LHC ApparatuS (ATLAS), the Large Hadron Collider beauty (LHCb), the Large Hadron Collider forward (LHCf) and the TOTal Elastic and diffractive cross section Measurement (TOTEM). ALICE, ATLAS, CMS and LHCb are installed in four huge underground caverns built around the four collision points of the LHC beams.

These detectors have different research purposes in nuclear physics.

CMS is a general-purpose detector built around a huge superconducting solenoid which takes the form of a cylindrical coil of superconducting cable. It can generate a magnetic field of 4 T.

CMS was mainly designed to look for the Higgs boson, to measure its properties, and also to scrutinize various currently unproven models as supersymmetry, the existence extra dimensions, or the origin of the dark matter. CMS possesses the necessary versatility to uncover unexpected phenomena at LHC energies.

2.2.2 Structure

To work as a precise detector and achieve its goals, CMS must be able to reconstruct the collision events in the best possible way. For this reason it is composed of several sub-detectors of different type to reveal most of the particles produced in the collisions, measuring their energy and momentum.

Since a magnetic field can be used to measure the momentum of a particle through the

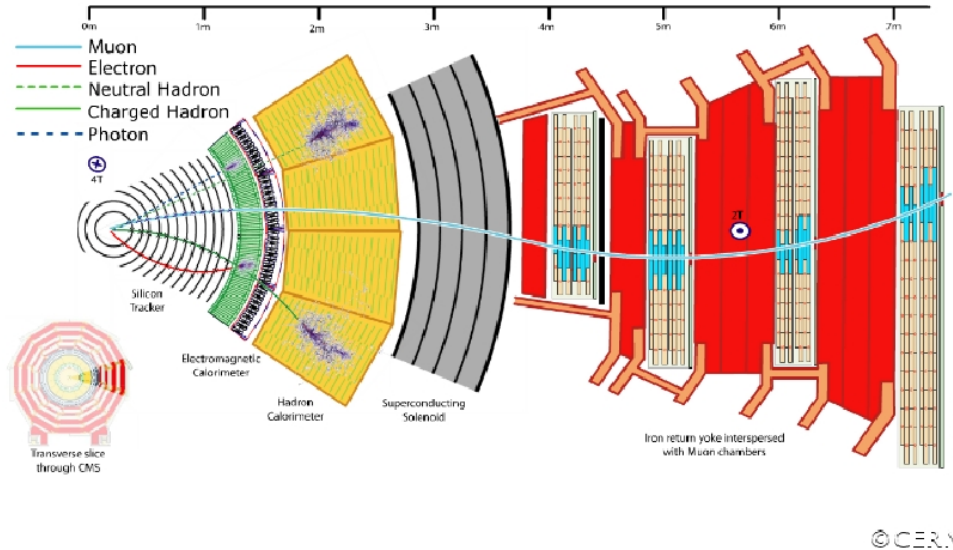


Figure 2.3: CMS transverse section.

bending of the track left in the detector, the central feature of the CMS apparatus is a superconducting solenoid, of 6 m internal diameter, which provides a magnetic field of 3.8 T. Such a strong magnetic field is needed to produce a large bending power to measure precisely the momentum of high-energy charged particles like muons. This forces a choice of superconducting technology for the magnets.

Within the field volume are the silicon tracker, the crystal electromagnetic calorimeter, and the brass/scintillator hadron calorimeter. Muons are measured in gas-ionization detectors located externally to the other elements of the detector.

A simple schematic of the CMS transverse section is shown in Fig. 2.3.

The tracker The inner tracking system of CMS is designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions (like electrons, protons, muons), as well as a precise reconstruction of secondary vertices. It surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m.

Being the nearest part of the detector to the collision point, the tracker will experience a huge particle flux. Therefore, a detector technology featuring high granularity and fast response is required, such that the trajectories can be identified reliably and attributed to the correct bunch crossing. However, these features imply a high power density of the on-detector electronics which in turn requires efficient cooling. This is in direct conflict with the aim of keeping to the minimum the amount of material in order to limit multiple scattering, bremsstrahlung, photon conversion and nuclear interactions. A compromise

had to be found in this respect. The intense particle flux will also cause severe radiation damage to the tracking system. The main challenge in the design of the tracking system was to develop detector components able to operate in this harsh environment for an expected lifetime of 10 years. These requirements on granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology.

The silicon tracker is composed of a pixel detector and strip detectors.

The pixel detector covers an area of about 1 m^2 and has 66 million pixels. When a charged particle passes through this detector it releases enough energy for electrons to be ejected from the silicon atoms of the pixels, creating electron-hole pairs. Each pixel uses an electric field to collect these charges on the surface as a small electric signal which is then amplified. In this way the pixel detector provides three high precision space points for each charged particle and thus gives the possibility to reconstruct the track.

After the pixels and on their way out of the tracker, particles pass through ten layers of silicon strip detectors, reaching out to a radius of 130 centimetres. This part of the tracker contains 15200 highly sensitive modules with a total of 10 million detector strips read by 80000 microelectronic chips. Each module consists of three elements: a set of sensors, its mechanical support structure and readout electronics.

The electromagnetic calorimeter (ECAL) The aim of the electromagnetic calorimeter is to measure the energy of photons and electrons. It is made of lead tungstate (PbWO_4) crystals whose usage guarantees high speed, fine granularity and radiation resistance, all important characteristics in the LHC environment.

These crystals have the important property to scintillate when electrons or photons pass through them, namely they produce light in proportion to the particles energy.

The ECAL has a cylindrical shape with two endcaps. The central part is called “the barrel”. A total of about 61200 crystals are located in the barrel and 7324 in the endcaps. Avalanche photodiodes are used as photodetectors in the barrel and vacuum phototriodes in the endcaps. Each photodetector produces an electric signal whose intensity is proportional to the energy of the photons coming from the crystal. As a result it is possible to measure the energy of the particles (electrons or photons) produced during the collisions.

The hadron calorimeter (HCAL) The CMS detector is designed to study a wide range of high-energy processes involving different signatures of final states. For this reason the hadron calorimeter (HCAL) [4] is particularly important for the measurement of hadron jets and neutrinos or exotic particles resulting in apparent missing transverse energy.

The hadron calorimeter is radially restricted between the outer extent of the electromagnetic calorimeter ($R = 1.77 \text{ m}$) and the inner extent of the magnet coil ($R = 2.95 \text{ m}$). As the electromagnetic calorimeter, the HCAL has a cylindrical shape composed of a barrel

and two endcaps.

The HCAL is a sampling calorimeter meaning that it finds a particle position, energy and arrival time using alternating layers of “absorber” and fluorescent “scintillator” materials that produce a rapid light pulse when the particle passes through. Special optic fibres collect up this light and feed it into readout boxes where photodetectors amplify the signal. When the amount of light in a given region is summed up over many layers of tiles in depth, called a “tower”, this total amount of light is a measure of a particle’s energy.

Measuring hadrons is important as they can tell us if new particles such as the Higgs boson or supersymmetric particles have been formed.

The muon detectors As the name suggest, one of the CMS main aims is to measure muons, fundamental particles similar to the electron but with a mass about 200 times heavier. A precise measure of these particles is important because we expect them to be produced in the decay of a number of potential new particles; for instance, one of the clearest “signatures” of the Higgs Boson is its decay into four muons.

Unlike most of the particles produced in the LHC collisions, muons can penetrate several layers of matter without interacting. For this reason, while most of the other particles are stopped in the internal calorimeters (ECAL, HCAL), muons are revealed in special detectors located externally to them.[5]

The track of a muon is measured by fitting a curve to hits among the four muon stations, which sit outside the magnet coil and are interleaved with iron plates.

By tracking their position through the multiple layers of each station, combined with tracker measurements, the detectors precisely trace the muons paths. Furthermore, thanks to the strong magnetic field, the muons momenta can be measured by observing the bending of their tracks.

2.3 High- P_T physics

As previously said, at LHC two beams of protons collide at high energy. The process of collision is very complex because every proton is composed of partons, namely quarks and gluons, which can interact strongly.

The description of proton-proton scatterings can be factorized into three parts: the momentum distribution of the partons inside the proton, the scattering matrix element for any two partons and the fragmentation function of the outgoing particles into measurable particles.

Each incoming parton is described by the variable x , the fraction of the proton momentum carried by the parton, i.e.

$$x = \frac{P_{parton}}{P_{proton}}.$$

In general, the two colliding partons have different values of x , so the center-of-mass frame for the scattering does not correspond to the laboratory (LAB) frame.

The scattering center-of-mass energy of the two partons p_1 and p_2 is

$$\sqrt{s'} = \sqrt{x_{p_1} x_{p_2} s},$$

where \sqrt{s} is the center-of-mass energy of the two protons.

The parton momentum distribution function inside the proton have been measured in neutrino scattering experiments at low momentum transfers. These have been evolved up to momentum transfers relevant for the colliders, using the prescription of QCD.

The scattering matrices have been calculated in QCD to leading order in α_s and in some cases up to α_s^3 .

The fragmentation of partons has been measured thoroughly at electron-positron colliders, and at $p\bar{p}$ colliders.

According to the characteristics of the final state of a collision, we can divide the scattering processes into three groups: elastic, inelastic and diffractive.

Elastic processes are characterized by the fact that both protons must be detected in the final state at small angles.

Inelastic processes are instead characterized by the presence of a reconstructed vertex and a large number of tracks.

Finally, in diffractive processes only one proton remains at small angle.

We define the variables P_T , η and ϕ to describe outgoing particles from the most general inelastic interaction.

Suppose to choose a cartesian frame with the positive z -axis lying along the counter-clockwise proton beam direction. The transverse momentum P_T is the modulus of the projection of the total momentum on the xy plane and it is defined as

$$P_T = \sqrt{P_x^2 + P_y^2},$$

or, equivalently, in terms of the polar angle θ and the modulus of the total momentum P

$$P_T = P \sin \theta.$$

Often in particle physics the quantity used to express directions is not the polar angle θ but the *pseudorapidity* η , which is defined as

$$\eta = -\ln \tan \frac{\theta}{2}.$$

As polar angles approach zero, pseudorapidities tend towards infinity, while pseudorapidities are zero for polar angles of $\pm 90^\circ$.

Finally, ϕ is the azimuthal angle of polar coordinates, defined as

$$\phi = \tan^{-1} \frac{P_x}{P_y}.$$

We finally define another useful quantity, the distance in the η - ϕ plane, defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}.$$

The inelastic processes that we want to describe are characterized by a large number of tracks often arranged in *jets*. A jet is a cluster of hadrons and other particles, produced by the hadronization of a quark or gluon, whose momenta are contained in a tight cone in the three-dimensional space.

The collisions that produce jets are characterised by a small distance scale, or high momentum transfer q^2 . These are the so called *hard scatterings*. [6]

Being LHC a proton-proton collider, a system of two partons interacting receives a boost in the LAB frame, which depends on the x momentum fractions carried by the partons. So, for the momentum conservation law, the final state of the collision will be characterized by jets with a small component of the total momentum along the z axis. On the other hand, being the energy of the collision very high, we expect to see jets with high momentum flying in almost opposite directions transversally to the z axis. These directions are the same of the scattered partons during the collision process, before hadronization occurred. We are interested in event like this, characterised by jets with high P_T . In the following we will adopt natural units ($c = \hbar = 1$) to simplify the notation.

Chapter 3

The top quark

Quarks and leptons constitute the basic building blocks of matter in the Standard Model. There are three generations of quarks and leptons in the model with different properties and masses. The top quark is the most massive of them, being approximately 200 times heavier than the proton and about 40 times heavier than the next-lightest quark, the bottom quark.

Despite the fact that a lot of properties have been measured since its discovery, there are still many things we do not know about the top quark.

The study of top quark physics tries to answer these questions.

3.1 Top quark discovery

The discovery of the top quark may be seen as one of the most evident confirmation of the Standard Model.

After the discovery of the charm quark (c) in 1974, physicists noted the existence of a symmetry between the particles known at the time. There were two quarks, the up (u) and the charm (c), which had an electric charge $+\frac{2}{3}e$, and two quarks, the down (d) and the strange (s), which otherwise had charge $-\frac{1}{3}e$. So it was possible to divide the quarks into two generations with two quarks of each possible charge: the first generation being composed of the u and d quarks, the second being composed of the c and s quarks. The first generation (u,d) contains the only stable quarks, which form all the ordinary matter and have lower masses, while the second one (c,s) contains unstable quarks with larger masses. Furthermore, this division into two generations could be made even with the leptons known at the time: the first composed of the electron (e) and the electron neutrino (ν_e), the second composed of the muon (μ) and the muon neutrino (ν_μ).

This symmetry however was quickly broken by unexpected discoveries. In 1976, experiments at SLAC (the *Stanford Linear Accelerator Center* in California) showed proofs of the existence of a third charged lepton, the tau lepton (τ). A year later at Fermi National Accelerator Laboratory (also known as *Fermilab* in Illinois) a new hadron, called

the upsilon (Y), was discovered at the large mass of about 10 GeV. The existence of that new particle could only be explained as a bounded state of a new quark, the bottom quark (b), and its antiparticle.

Experiments at DESY in Germany and Cornell in New York showed that the b quark has spin $\frac{1}{2}\hbar$ and a charge of $-\frac{1}{3}e$, just like the d and s quarks.

The discovery of the new lepton τ and the new quark b suggested to physicists that a third generation of particles could exist. But the third generation quark doublet seemed to be missing its charge $+\frac{2}{3}e$ member, whose existence was inferred from the existing pattern. In advance of its sighting, physicists named it the top (t) quark. Thus began a search that lasted almost twenty years.

In the early 1980s a new class of accelerators came into operation at CERN, in which beams of protons and antiprotons collided with an energy of about 600 GeV. The protons and antiprotons brought their constituent quarks and antiquarks into collision with typical energies of 50 to 100 GeV, so the top quark search could be extended considerably. Besides the important discovery of the W and Z bosons that act as carriers of the unified electroweak force, the CERN experiments demonstrated another aspect of quarks. Though quarks had continued to elude direct detection, they can be violently scattered in high energy collisions. The high energy quarks emerging from the collision region are subject to the strong interaction as they leave the scene of the collision, creating additional quark-antiquark pairs from the available collision energy (using $E = mc^2$). The quarks and antiquarks so created combine into ordinary hadrons that the experiment can detect. These hadrons tend to cluster along the direction of the original quark, and are thus recorded as a *jet* of rather collinear particles.

Such quark jets were clearly observed at CERN and became a key ingredient in the next round of top quark searches.

The final chapter in finding the top quark began at the Tevatron, a circular particle accelerator at Fermilab

The Tevatron involves a collection of seven separate accelerators with a complex web of connecting beam lines. In this accelerator beams of protons and antiprotons were accelerated initially to 900 GeV and finally to 980 GeV and were made to collide.

The two main Tevatron experiments which worked on top quark discovery were called $D\bar{O}$ and CDF (*Collider Detector at Fermilab*).

After several efforts, in 1995, the two experiments announced the top quark discovery to the scientific community: the CDF measurement for the top quark mass was 176 ± 13 GeV [7], while the $D\bar{O}$ measurement was 199 ± 30 GeV [8]. Therefore, the results were consistent and in agreement with the Standard Model predictions.

3.2 Main properties of the top quark

In this section let us summarize the most important properties of the top quark.

According to the Standard Model, the top quark is a fundamental particle belonging to the third generation. It is a fermion, which means it has a half-integer spin ($\frac{1}{2}\hbar$) and so it obeys the Pauli exclusion principle. Like the other quarks, it feels the electromagnetic interaction (having an electric charge, $+\frac{2}{3}e$), the strong nuclear interaction (having a color charge), the weak nuclear interaction (being part, together with the bottom quark, of a weak isospin doublet).

The top quark is the most massive quark: $M_t = 173.21 \pm 0.51 \pm 0.71$ GeV (the first uncertainty is statistical while the second is systematical). This is the latest *Particle Data Group* (PDG) evaluation of the top quark mass, the average of the measurements made by the most important experiments all over the world [9].

Thanks to its huge mass (bigger than the W boson), top quark is the only quark which can decay semi-weakly into a real W boson and a b quark, before hadronization can occur. In fact, differently from the other quarks, it has never been observed a bounded state involving this quark. Top quark lifetime is so short ($\sim 5 \cdot 10^{-25}$ s) that it decays weakly before it can form mesons or baryons with other quarks.

3.3 Why is top quark mass so important?

One of the principal aims of the contemporary research in subnuclear physics is to understand the limits of the Standard Model and to eventually provide a more general theory. In this field an accurate measurement of the top quark mass plays a key role for several reasons:

- As it was said previously, the top quark decays well before hadronizing. This allows us to measure the mass of this quark directly from the observation of its decay products. For this reason the top quark is the most accurately measured of all quarks.
- The top quark participates to quantum loop radiative corrections to the W mass constraining the Higgs boson mass. Now that the Higgs boson has been discovered, the measurement of these masses constitutes a verification of the self-consistency within of the Standard Model.
- In the Standard Model mass is generated by electroweak symmetry breaking. The large mass of the top quark, close to the electroweak scale, indicates that it couples strongly to the field that breaks this symmetry. The top quark could thus play an important role in understanding the generation of mass in the Standard Model.

- The top quark mass is related to the electroweak vacuum stability described by the Standard Model. According to the Standard Model a really accurate measurement of the top quark and Higgs boson masses will allow to understand whether the universe is stable, unstable or metastable leading to possible different endings. The values of the two masses measured so far seem to place the universe near the limit between the stable and the metastable behavior. The high experimental uncertainty related to the current measurements of the top quark and Higgs boson masses however does not allow us to make an accurate prediction.

If the universe were stable thus all the laws of physics we know would remain the same in the future. Otherwise if the universe were metastable, the electroweak vacuum would currently lie in a local energy minimum. In this case a global energy minimum would however exist and so the electroweak vacuum could evolve into this energetically convenient state. The transition of the universe to this state could have a great impact on the structure of the universe itself, altering the nature of the laws of physics we know.

3.4 Production and decay

3.4.1 Production

At a hadron collider, top quark production may occur thanks mainly to two different processes:

- A pair $t\bar{t}$ may be produced through the strong nuclear interaction (described by QCD: *Quantum Chromo-Dynamics*). This process is also known as *pair production*;
- A single top quark may be produced through the charged-current coupling (electroweak interaction).

Pair production: The first production channel (the one used for the top quark discovery at Tevatron) is the most frequent at LHC with a production cross section of $\sigma_{t\bar{t}} \approx 250$ (830) pb for $\sqrt{s} = 8$ (13) TeV, while the second is less likely, with $\sigma_t \approx 115$ (300) pb for $\sqrt{s} = 8$ (13) TeV.

At leading order, the most important Feynman diagrams for top quark pair production are shown in Figs. 3.1, 3.2 and 3.3.

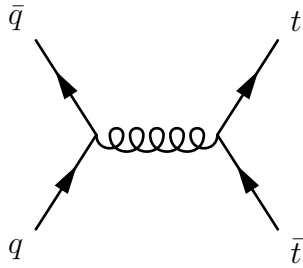


Figure 3.1: $t\bar{t}$ production through $q\bar{q}$ annihilation.

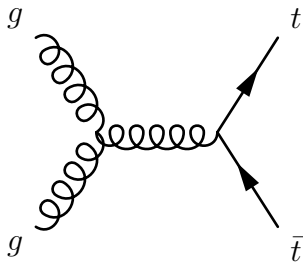


Figure 3.2: $t\bar{t}$ production through gluon fusion (s-channel).

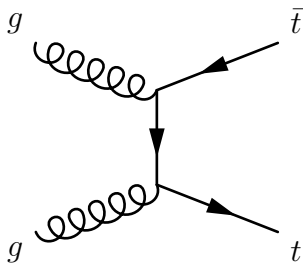


Figure 3.3: $t\bar{t}$ production through gluon fusion (t-channel).

So top quark production can occur by $q\bar{q}$ annihilation (first diagram) or by gluon fusion (second and third diagram).

With the energy rising, the cross section for the gluon fusion processes become increasingly dominant with respect to the $q\bar{q}$ annihilation process. For instance at LHC with $\sqrt{s} = 14$ TeV, the gluon fusion process is predicted to occur 90% of times, while $q\bar{q}$ only 10% of times.

Single top quark production: Single top quark production was first observed in 2009 by DØ and CDF at the Tevatron. The production cross section at the Tevatron was roughly half that of the pair production cross section, but the final state was much more difficult to distinguish from the background.

The dominant production is through s-channel and t-channel W-boson exchange.

Another process is the W-t channel where the top quark production is associated to the emission of a real W boson.

The Feynman diagrams of these three single top quark production channels are shown in Figs. 3.4, 3.5 and 3.6.

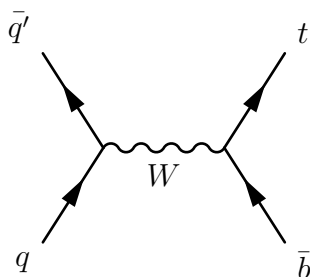


Figure 3.4: Single top quark production (s-channel).

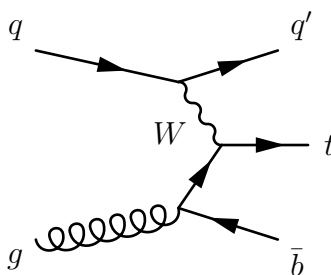


Figure 3.5: Single top quark production (t-channel).

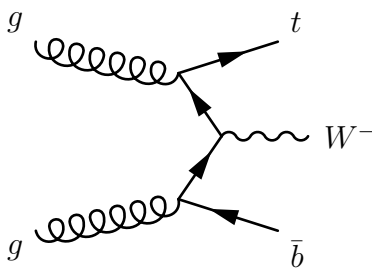


Figure 3.6: W-t production.

3.4.2 Decay

The phenomenology of the top quark decay is driven by its large mass. Being heavier than a W boson, it is the only quark that decays semi-weakly, namely into a real W boson and a b quark, before hadronization can occur.

Since top quarks decay almost exclusively in this way, the possible outcomes are classified according to the W boson decay.

Let us only analyze the possible outcomes for the $t\bar{t}$ state:

- Dilepton channel: in this case both the W bosons decay in a couple lepton-neutrino ($(l, \bar{\nu}_l)$ and $(\bar{l}', \nu_{l'})$), as described in Fig. 3.7.

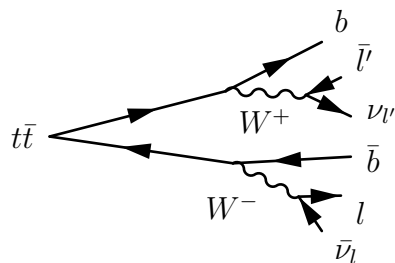


Figure 3.7: Dilepton channel.

This process has a branching ratio of 5% if we consider only electrons and muons, and not tau leptons.

- Single-lepton channel: in this case only a W decays into a couple lepton-neutrino, while the other decays hadronically into a couple of quarks, as shown in Figs. 3.8 or 3.9. This process has a branching ratio of 30% when considering only electrons and muons, but not tau leptons.

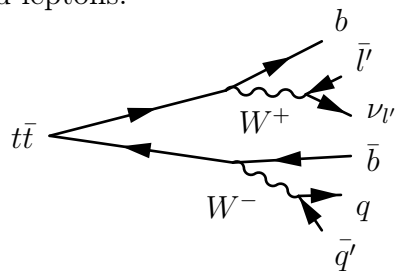


Figure 3.8: Single-lepton channel (first way).

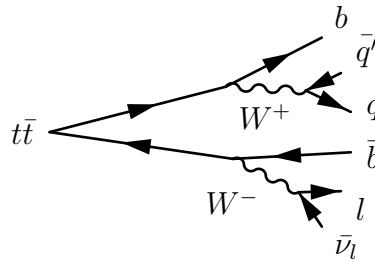


Figure 3.9: Single-lepton channel (second way).

- All-hadronic channel: in this case both the W bosons decay hadronically into a couple of quarks, and so in the final state there are six quarks which then hadronize and are detected as jets.

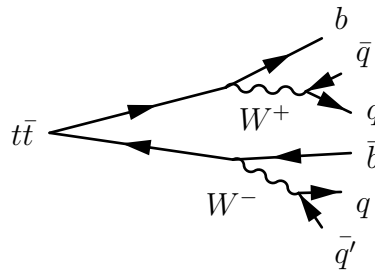


Figure 3.10: All hadronic channel.

This process, represented in Fig. 3.10, has a branching ratio of 46%.

Chapter 4

$t\bar{t}$ signal description

4.1 Monte Carlo simulations

The main goal of CERN experiments is to verify the Standard Model predictions on particle physics or, eventually, to find something new. For this reason scientists need to compare experimental data provided by the CERN detectors to theoretical predictions. Because of the complexity of the processes involved in LHC collisions, it is very difficult to find a way to generate the theoretical samples needed for this comparison. The best choice are computational methods.

The most used computational methods in particle physics are the MC methods. MC methods are a large class of computational algorithms which rely on repeated random sampling to obtain numerical results. For this reason, these results depend on the sequence of random numbers which is generated during the simulation. With a second, different sequence of random numbers the simulation will not give identical results but will yield values which agree with those obtained from the first sequence to within some statistical error.

The CMS collaboration uses MC generation programs to model a number of physics processes relevant to $t\bar{t}$ production and decay. Events are subsequently passed through a complete simulation of the detector response.

The resulting simulated samples are treated just like the recorded pp collision data, using the same reconstruction software and particle identification algorithms.

The principal aim of this study is to compare the simulated events of top quark decay generated by CMS at two different center-of-mass energies: $\sqrt{s} = 8$ TeV as in the 2012 run, and $\sqrt{s} = 13$ TeV as in the current run (Run2).

4.2 Data analysis

Between the various possible channels of $t\bar{t}$ decay presented in the previous chapter, we study here the all-hadronic one. As we have seen, we therefore expect six jets in the final

state: four jets coming from the light quarks hadronization and two jets coming from the b quarks hadronization. More than six jets may however appear, coming from the initial or final state radiation. On the other hand, jets might be lost if they are not very energetic or fall outside the acceptance.

Suppose to make a selection over the events of a given MC sample. The number of expected $t\bar{t}$ events is given by

$$N_{t\bar{t}}^{exp} = \epsilon \sigma_{t\bar{t}} L,$$

where ϵ is the *efficiency* of the selection we have chosen. The efficiency is defined as

$$\epsilon = \frac{N_{evt}^{cut}}{N_{evt}^{gen}},$$

where N_{evt}^{gen} is the total number of events in the MC sample (8 TeV or 13 TeV), while N_{evt}^{cut} is the number of event which passed the selection.

Furthermore, $\sigma_{t\bar{t}}$ is the $t\bar{t}$ pair production cross section at 8 or 13 TeV and L is the integrated luminosity, which is defined as the integral of the instantaneous luminosity,

$$L = \int \mathcal{L} dt.$$

The instantaneous luminosity \mathcal{L} is defined in terms of the collider parameters as

$$\mathcal{L} = \frac{fn_1n_2N_bG}{4\pi r^2},$$

where n_1 and n_2 are the numbers of particles in the counter-rotating bunches, N_b is the number of bunches, r is the average size of the bunches in the transverse plane, f is the revolution frequency and finally G takes into account the finite length of the bunches.

However, in this study we are working with simulated samples, not with data samples. For this reason the integrated luminosity of the MC sample can be expressed as

$$L_{MC} = \frac{N_{evt}^{gen}}{\sigma_{t\bar{t}}}$$

In Table 4.1 are listed the values of N_{evt}^{gen} , $\sigma_{t\bar{t}}$ and integrated luminosity for the 8 TeV and 13 TeV MC samples.

	8 TeV	13 TeV
N_{evt}^{gen}	61913917	42730273
$\sigma_{t\bar{t}}$ (fb)	253000	832000
L_{MC} (fb ⁻¹)	244	51

Table 4.1: N_{evt}^{gen} , $\sigma_{t\bar{t}}$ and integrated luminosity values for the 8 TeV and 13 TeV simulated samples.

Trigger Although the trigger requirement is the preliminary request of any selection, we do not apply it here because the MC samples at 8 and 13 TeV do not have a common multijet event trigger. To avoid the introduction of biases in our comparison we skipped the trigger request.

Jets The particle-flow algorithm [10] reconstructs and identifies each individual particle in an event with an optimised combination of information from the various elements of the CMS detector.

The energy of photons is directly obtained from the ECAL measurement. The energy of electrons is determined from a combination of the electron momentum as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy. To mitigate the effect of pileup, i.e. additional proton-proton collisions whose signals in the detector sum to the products of the primary interaction that triggered the event, charged particles associated to non-leading primary vertices are vetoed.

Jets are then reconstructed from the PF particle candidates using the anti- k_T clustering algorithm [11] with a distance parameter of 0.4.

Cuts on P_T and η , request on the number of jets Setting a threshold on some kinematical quantities may help excluding background events. First of all jets coming from a $t\bar{t}$ decay should have high transverse momentum and a not too small angle between their total momentum and the z -axis.

Thus, in order to discard low-energetic jets with small angles with respect to the z -axis, we have chosen to keep only the jets with $P_T > 30$ GeV and $|\eta| < 2.4$.

These thresholds on P_T and η have been chosen as a starting point which replaces the trigger selection and guarantees a reduced presence of multijet background events.

As previously said, we want to study the $t\bar{t}$ all-hadronic decay, characterized by the presence of at least six jets in the final state. For this reason we discard all the events containing less than six jets ($N_{jets} \geq 6$). Furthermore, events with 6 or more jets will allow a full kinematical reconstruction of the two decaying top quarks.

B-tagging algorithms In order to reconstruct the top quark mass we need to know which of the jets are b -jets, namely jets coming from bottom quark hadronization. To this end CMS provides several different algorithms which use a variety of reconstructed tracks, track vertices and identified leptons to build observables that discriminate between b -jets and other light-quark jets. Each of these CMS algorithms yields a single discriminator

value for each jet.

Furthermore, each b-tagging algorithm is associated to three thresholds: Loose (L), Medium (M) and Tight (T). We decide whether to consider or not a jet as a b-jet according to these thresholds. The higher the threshold, the higher the probability that the jets passing the cut are actually b-jets. In spite of this, the higher the threshold, the lower the number of jets which can pass the cut, which implies a decrease in the cut efficiency. The minimum thresholds on these discriminators define L, M and T operating points with a misidentification probability for light-quark jets of close to 10%, 1% and 0.1% respectively.

In this work we used the BtagCSV algorithm (*Combined Secondary Vertex*) [12] with a medium cut (M) for both the 8 TeV and the 13 TeV simulated samples. In Table 4.2 are listed the BtagCSV values of Loose, Medium and Tight thresholds for 8 TeV and 13 TeV simulated samples. The values of the threshold represent the current official working points established by the CMS group.

	L	M	T
8 TeV	0.244	0.679	0.898
13 TeV	0.605	0.890	0.970

Table 4.2: BtagCSV Loose, Medium and Tight thresholds for 8 TeV and 13 TeV simulated samples.

An all-hadronic $t\bar{t}$ decay is characterized by the presence of at least two b-jets in the final state. So, if an event has passed the $N_{jets} \geq 6$ cut, we decide to keep it only if $N_{b-jets} \geq 2$. We call this selection on the events the *standard selection*.

4.3 Mass reconstruction

A reconstructed top quark mass is determined fitting the kinematic variables of the jets in the event to a $t\bar{t}$ final state, according to the all-hadronic decay hypothesis ($t\bar{t} \rightarrow W^+bW^-b \rightarrow j_1j_2bj_3j_4\bar{b}$).

There are several possibilities for the matching of the jets to the final state quarks of a $t\bar{t}$ decay, but we limit ourselves to those regarding the 6 leading jets where two of these leading jets are tagged (Medium tag) and associated to the b quarks.

A ROOT macro (*CompareMC.C*) is used to reconstruct the top quark mass.

When the macro starts, a pre-selection decides whether to work on 8 TeV MC or on 13 TeV MC. Then the macro controls, event by event, if the current event passes the cuts described in the previous section. If so, the kinematic fit starts: first of all the four-momenta $P^\mu = (E, \mathbf{P})$ of the jets are reconstructed and used to compute the di-jet and trijet four-momenta (representing respectively the W bosons and the top quarks).

$$P_{W^+}^\mu = P_{j_1}^\mu + P_{j_2}^\mu = (m_{jj}^{(1)}, \mathbf{P}_{jj}^{(1)})$$

$$P_{W^-}^\mu = P_{j_3}^\mu + P_{j_4}^\mu = (m_{jj}^{(2)}, \mathbf{P}_{jj}^{(2)})$$

$$P_t^\mu = P_{W^+}^\mu + P_b^\mu = (m_{jjb}^{(1)}, \mathbf{P}_{jjb}^{(1)})$$

$$P_{\bar{t}}^\mu = P_{W^-}^\mu + P_{\bar{b}}^\mu = (m_{jjb}^{(2)}, \mathbf{P}_{jjb}^{(2)})$$

where j_1, j_2, j_3 and j_4 are the light jets, while b, \bar{b} are the two b-tagged jets.

So, the invariant masses of each dijet or trijet systems can be obtained directly from their four-momenta.

For each combination in the event the top quark reconstructed mass m_t^{rec} can be determined from a fit based on the χ^2 -like function

$$\chi^2 = \frac{(m_{jj}^{(1)} - M_W)^2}{\Gamma_W^2} + \frac{(m_{jj}^{(2)} - M_W)^2}{\Gamma_W^2} + \frac{(m_{jjb}^{(1)} - m_t^{rec})^2}{\Gamma_t^2} + \frac{(m_{jjb}^{(2)} - m_t^{rec})^2}{\Gamma_t^2} + \sum_{i=1}^6 \frac{(P_{T,i}^{fit} - P_{T,i}^{meas})^2}{\sigma_i^2},$$

where $m_{jj}^{(1,2)}$ are the invariant masses of the dijet systems associated to light flavor quarks, $m_{jjb}^{(1,2)}$ are the invariant masses of the trijet systems including one b quark (Medium tag), $M_W = 80.4$ GeV and $\Gamma_W = 2.1$ GeV are the mass and natural width of the W boson, $\Gamma_t = 1.5$ GeV is the assumed natural width of the top quark, σ_i is the resolution on the $P_{T,i}^{meas}$ measurement, $P_{T,i}^{fit}$ is the value returned by the fit.

Instead of this χ^2 -like function, we used a simpler one

$$\chi_{std}^2 = \frac{(m_{jj}^{(1)} - M_W)^2}{\sigma_W^2} + \frac{(m_{jj}^{(2)} - M_W)^2}{\sigma_W^2} + \frac{(m_{jjb}^{(1)} - m_t^{rec})^2}{\sigma_t^2} + \frac{(m_{jjb}^{(2)} - m_t^{rec})^2}{\sigma_t^2},$$

where $\sigma_W = 12$ GeV and $\sigma_t = 20$ GeV are the standard deviations associated to the dijet and trijet reconstructed masses, obtained by putting together the resolution on the P_T measurements and the values of natural widths Γ_W and Γ_t .

For each permutation of the 6 jet-to-quark assignments in the event, the χ_{std}^2 function is minimized with respect to the free parameter m_t^{rec} , using the MINUIT algorithm; the combination which gives the lowest χ_{std}^2 value, is selected and the corresponding mass m_t^{best} is chosen as estimator of the top quark mass. The m_t^{best} values obtained in this way can be plotted on a histogram for further analysis. The χ_{std}^2 value associated to each m_t^{rec} can also be used to generate histograms with a stricter request on χ_{std}^2 (i.e. $\chi_{std}^2 < 5$, $\chi_{std}^2 < 3$).

Finally, the macro saves all the histograms in an output ROOT file called *MC8TeV.root* if the input file is a 8 TeV MC, *MC13TeV.root* if the input file is a 13 TeV MC.

Chapter 5

Comparison between multijet events at 8 and 13 TeV

We begin the comparison between 8 and 13 TeV MC samples by analyzing some general characteristics of events and jets, without having applied any trigger on them. In the ROOT files only those jets which have $P_T > 30$ GeV and $\eta < 2.4$ have been saved.

P_T and η distributions Let us now study the jets P_T and η distributions. In Figs. 5.1 and 5.3 are respectively shown the normalized histograms for the P_T and η distributions.

Turning from 8 to 13 TeV MC samples we expect to see an increased jets production due to the higher energy in the center-of-mass of the collisions. Not all of these jets in general come from $t\bar{t}$ decay as they might be produced from initial or final state radiation. These jets are less energetic than the ones produced by $t\bar{t}$ decay and thus they even have lower P_T .

We can notice this effect in Fig. 5.1 as an increased peak in the 13 TeV P_T distribution, for small values of P_T . Furthermore, this effect lowers the mean value of the 13 TeV distribution with respect to the 8 TeV one.

On the other hand, the increased energy characterizing the 13 TeV collisions can produce highly boosted jets which may be merged into a single jet with a large P_T .

This effect can be clearly seen in Fig. 5.2, where we have plotted the P_T distributions for 8 and 13 TeV MC samples in a logarithmic scale. We can easily see as for 13 TeV samples there is an higher number of jets for large P_T values ($P_T > 400$ GeV).

Furthermore, as we can see in Fig. 5.3, jets coming from 13 TeV MC samples have higher values of $|\eta|$ than the 8 TeV ones. This effect can be explained in terms of relativistic boost. As explained in section 2.3, the boost between the center-of-mass of the interacting partons and the LAB frame, determines the magnitude of the total momentum z component of the final states. Because of the energy increasing from 8 to 13 TeV MC samples, we will have, in general, a higher boost. This implies an increasing of the

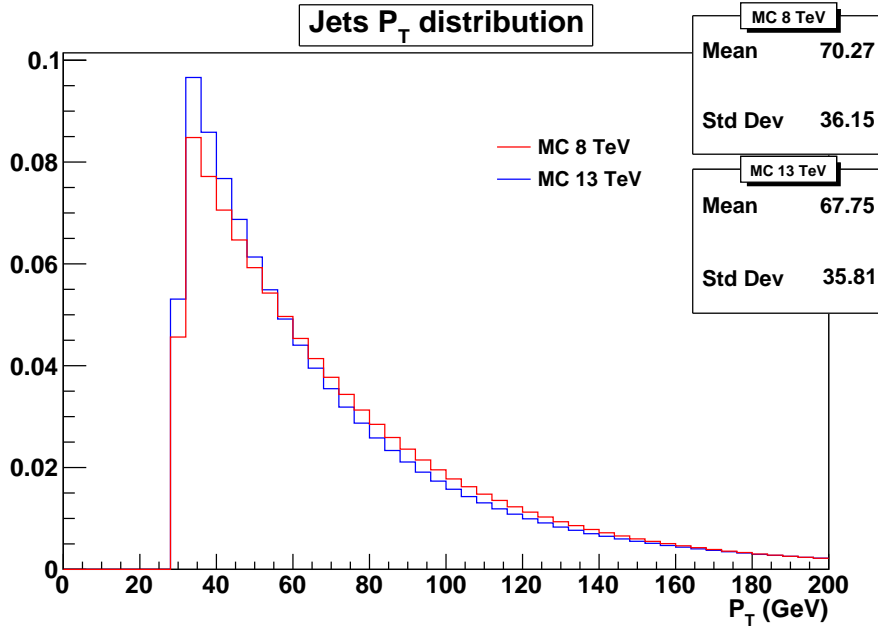


Figure 5.1: Jets P_T distributions for 8 and 13 TeV MC samples, normalized to equal area.

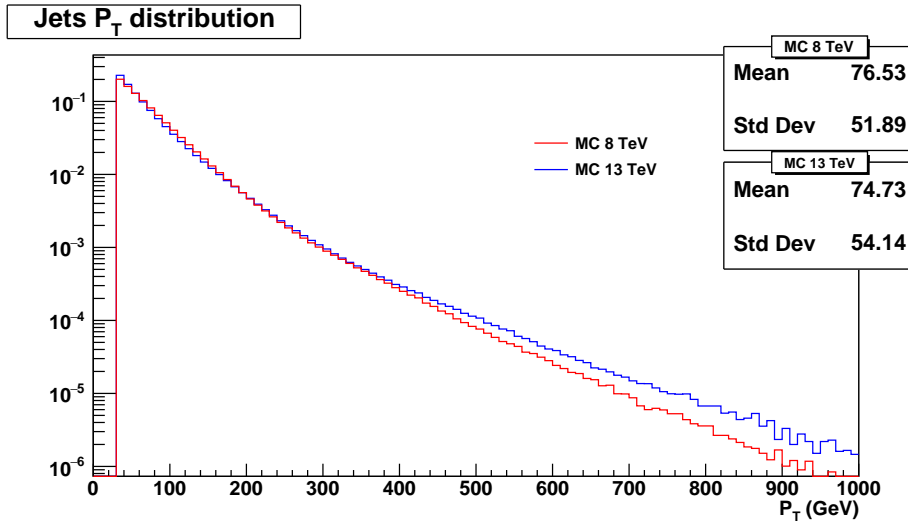


Figure 5.2: Jets P_T distributions for 8 and 13 TeV MC samples (logarithmic scale), normalized to equal area.

P_z component which brings jets to higher $|\eta|$ values.

In Fig. 5.3, the black line shows the η distribution for isotropic $t\bar{t}$ production, namely the extreme case in which two partons interact with a perfectly central collision with

equal momenta. In this case, t and \bar{t} have the same chance to be emitted in any spacial direction. This histogram is created using ten million randomly-generated θ values in the interval $[0, \pi]$ and applying the cut $|\eta| < 2.4$.

Because this condition is very rare, we will have in general a boost of the center-of-mass of the partons colliding, which increases for higher energies of the collider.

As a matter of facts, we can notice by inspecting Fig. 5.3 as 13 TeV jets have higher values of η with respect to the isotropic and the 8 TeV distribution. This means that there will be a larger amount of 13 TeV jets with smaller angles with respect to the z -axis. Furthermore, we shall notice how the $|\eta| < 2.4$ cut acts in different ways on the two samples: because 13 TeV jets have in general higher values of η , the cut has a lower efficiency on these samples because we loose a larger amount of jets with $|\eta| > 2.4$.

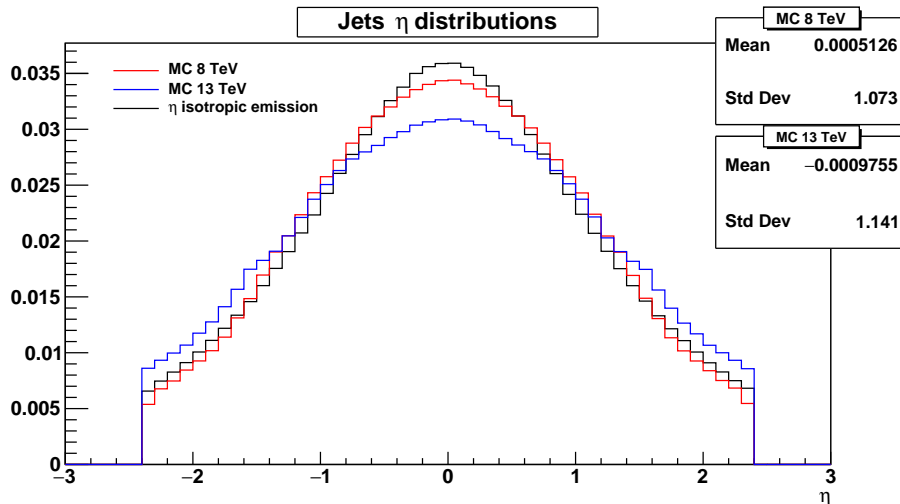


Figure 5.3: Jets pseudorapidity distributions for 8 and 13 TeV MC samples and pseudorapidity isotropic distribution, normalized to equal area.

Jet multiplicity We now turn to the jets multiplicity comparison, namely the number of jets (N_{jets}) per event. We remember that the samples we used for this comparison have at least six jets per event.

In Fig. 5.4 we can notice that 13 TeV MC samples are characterized by an increased amount of events with more than six jets. As explained before, this effect is due to the higher energy in the center-of-mass system of the partons interacting. For this reason, a larger number of jets (in general with small P_T) will be produced from the initial or final state radiation.

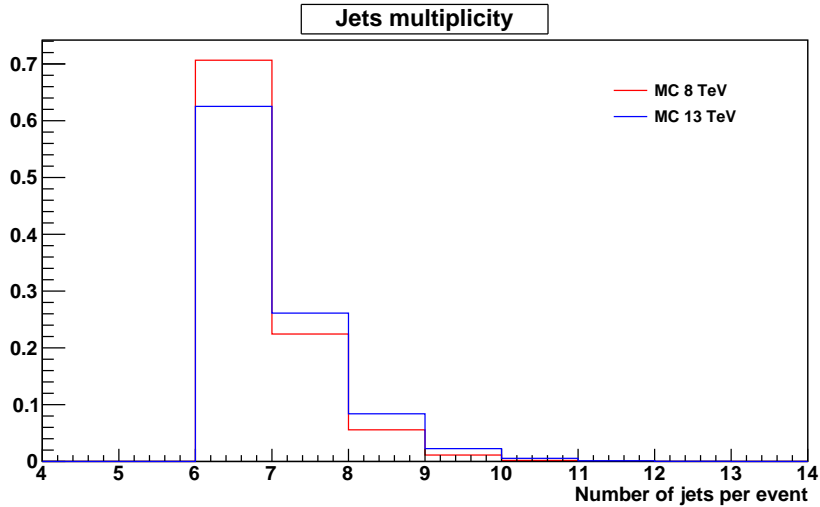


Figure 5.4: Jets multiplicity distributions for 8 and 13 TeV MC samples, normalized to equal area.

5.1 Efficiency of the standard selection

In order to compare correctly 8 and 13 TeV standard selection efficiencies, the same trigger should be applied on all samples. Unfortunately, the useful *HLT_QuadJet50* trigger defined for the 8 TeV samples, is not defined for the 13 TeV ones. As a matter of fact 13 TeV samples currently have multijet triggers which contain the b-tagging. The triggers will be used in the final analysis but cannot be used at this stage because they bias these studies.

To solve this problem we have decided not to apply any trigger on both the 8 and 13 TeV samples, but to select an appropriate threshold on the jets P_T . In fact we have noticed that the 8 TeV MC *HLT_QuadJet50* trigger has an efficiency of about 95.6% if we ask the presence of six jets with $P_T > 40$ GeV. The same request is thus temporarily applied even on 13 TeV MC samples, simulating in this way the effect of a trigger on both the samples.

Let us now study the efficiencies of the standard selection requiring $P_T > 40$ GeV. In Table 5.1 we show the standard selection efficiencies for 8 and 13 TeV MC samples.

Looking at Table 5.1 we can notice a decreasing in the efficiencies for the $N_{jets} \geq 6$, $N_{b-jets} \geq 2$ selection for 13 TeV samples with respect to the 8 TeV ones.

As a matter of fact, if we compute the relative efficiencies of the $N_{jets} \geq 6$, $N_{b-jets} \geq 2$ selection with respect to the $N_{jets} \geq 6$ one, we obtain a value about 43.9% for 8 TeV samples and of about 35.3% for the 13 TeV ones.

This higher loss in efficiency is caused by the choice of a too high value for the Medium

	8 TeV		13 TeV	
	N_{evt}	ϵ (%)	N_{evt}	ϵ (%)
N_{evt}^{gen}	61913917	-	42730273	-
$N_{jets} \geq 6$	2065107	3.33	1101716	2.58
$N_{jets} \geq 6, N_{b-jets} \geq 2$	907376	1.47	388301	0.91

Table 5.1: Standard selection efficiencies for 8 and 13 TeV MC samples.

b-tagging threshold for 13 TeV samples.

These official values for the thresholds are preliminary and will need to be tuned.

In the following analysis we will keep the official thresholds, even if there will be probably improvements in the early future.

Turning back to the standard selection, we can notice, as expected, a general loss in efficiency for the standard selection turning from 8 to 13 TeV MC samples. This effect is due to the larger fraction of events having ≤ 5 jets, due to the merging of jets because of the boosted decays of the top quark. Such an effect cannot be seen in the figures we presented so far because the samples used for this study have a prerequisite of having at least 6 jets.

5.2 Mass distributions

We now turn to the most important part of the analysis, the top quark mass reconstruction. We want to study how the mass distributions and the χ_{std}^2 cut efficiencies have changed turning from 8 to 13 TeV MC samples.

In Table 5.2 we compare the χ_{std}^2 cut efficiencies, having applied the tighter requirement $P_T > 40$ GeV, $|\eta| < 2.4$.

	8 TeV		13 TeV	
	N_{evt}	ϵ (%)	N_{evt}	ϵ (%)
$\chi_{std}^2 \leq 10$	347924	0.56	138365	0.32
$\chi_{std}^2 \leq 5$	236992	0.38	90904	0.21
$\chi_{std}^2 \leq 3$	166340	0.27	62402	0.15

Table 5.2: χ_{std}^2 cut efficiencies of the mass fit for 8 and 13 TeV MC samples.

In Figs. 5.5 and 5.6 we show the comparison between m_t^{best} and $m_t^{\chi_{std}^2 \leq 3}$ distributions respectively for 8 and 13 TeV MC samples. The former is the distribution of the top quark mass which gives the best combination in each event, while the latter is obtained from the fit by applying the cut $\chi_{std}^2 \leq 3$.

Finally in Fig. 5.7 we show the direct comparison between $m_t^{\chi^2_{std} \leq 3}$ distributions for 8 and 13 TeV MC samples. We remind that all the plots are normalized to equal area.

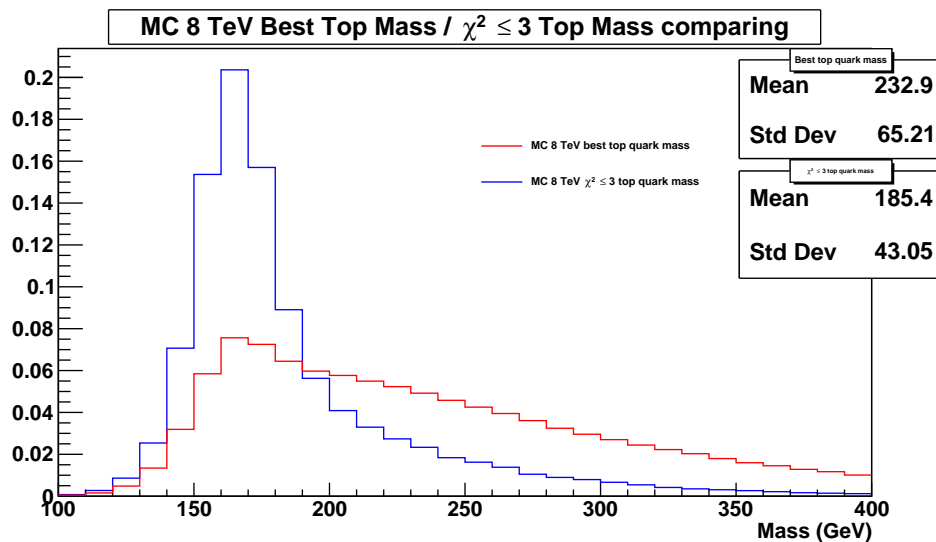


Figure 5.5: Comparison between jets m_t^{best} and $m_t^{\chi^2_{std} \leq 3}$ distributions for 8 TeV MC sample, normalized to equal area.

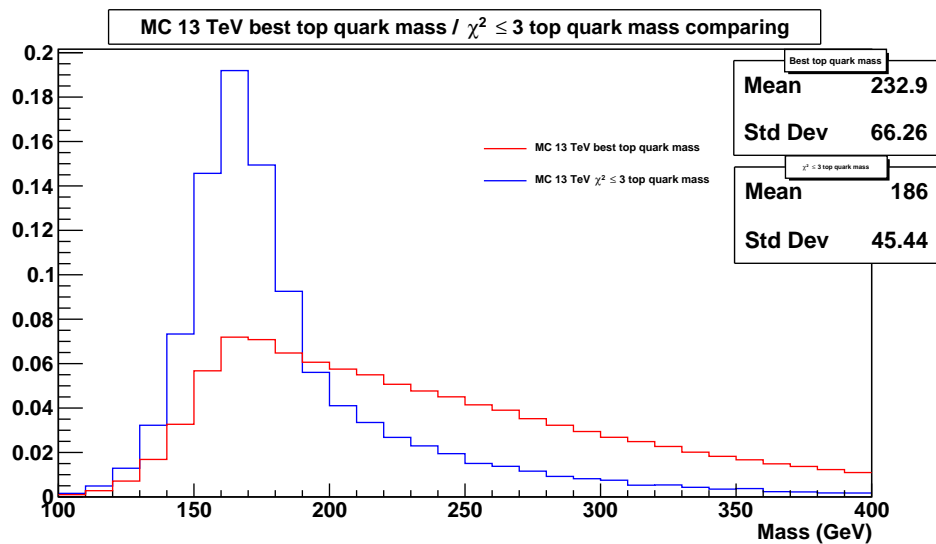


Figure 5.6: Comparison between jets m_t^{best} and $m_t^{\chi^2_{std} \leq 3}$ distributions for 13 TeV MC sample, normalized to equal area.

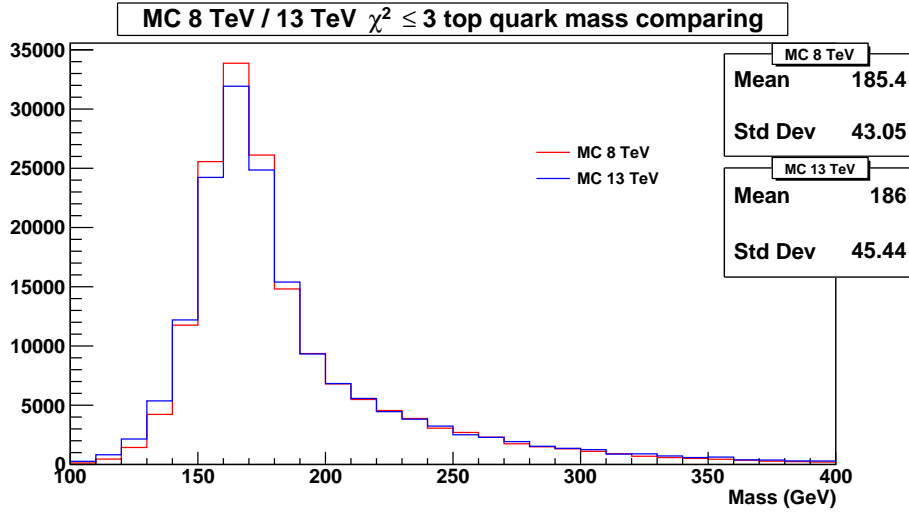


Figure 5.7: Comparison between $m_t^{\chi_{std}^2 \leq 3}$ distributions for 8 and 13 TeV MC samples, normalized to equal area.

Let us make some considerations about the plots: as we can see, there are not strong differences between 8 and 13 TeV $m_t^{\chi_{std}^2 \leq 3}$ distributions. This implies that we can keep the same mass reconstruction algorithm obtaining good results. However, looking at Table 5.2, we can notice a loss in efficiency of the 13 TeV cuts. In Fig. 5.8 we plot the χ_{std}^2 distributions for 8 and 13 TeV MC samples, namely the distributions of the χ_{std}^2 values computed for every possible jet combination.

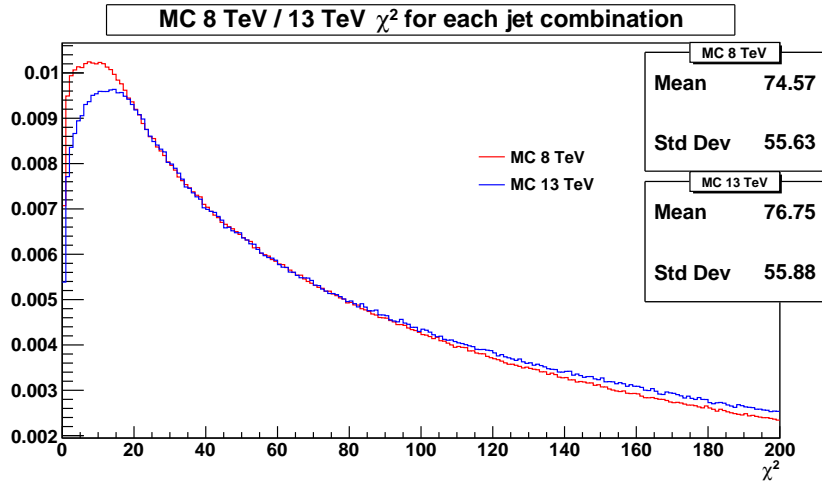


Figure 5.8: Comparison between χ_{std}^2 distributions for 8 and 13 TeV MC samples, normalized to equal area.

As we can see, 13 TeV samples have higher values of χ_{std}^2 . This effect is due to the merging of jets because of the higher boost at 13 TeV. For this reason, the use of these merged jets instead of the jets associated to the six quarks, yields higher χ_{std}^2 values. This study suggests to use a looser cut on χ_{std}^2 if we want to keep a high enough efficiency at 13 TeV. The choice of the χ_{std}^2 cut will be done in the future, studying the data and the possible values of S/B (the ratio between the signal and the background).

5.3 Possible improvements

Because of the high energy involved in 13 TeV proton-proton collisions, top quarks can be produced with a relevant boost with respect to the LAB frame. For this reason, their decay products will be boosted too and the standard jets clustering algorithm may fail, wrongly merging them into a single jet. This is an issue for the standard mass fit which is not able to correctly reconstruct the top quark mass from these boosted jets.

Because we are working with MC samples, we can get the information of the parton that generated each jet. So we can prove the presence of boosted top quarks in the samples by showing, in Fig. 5.9, the distance in the $\Delta\phi$ - $\Delta\eta$ plane between the directions of each top quark and its related W boson for an all-hadronic decay.

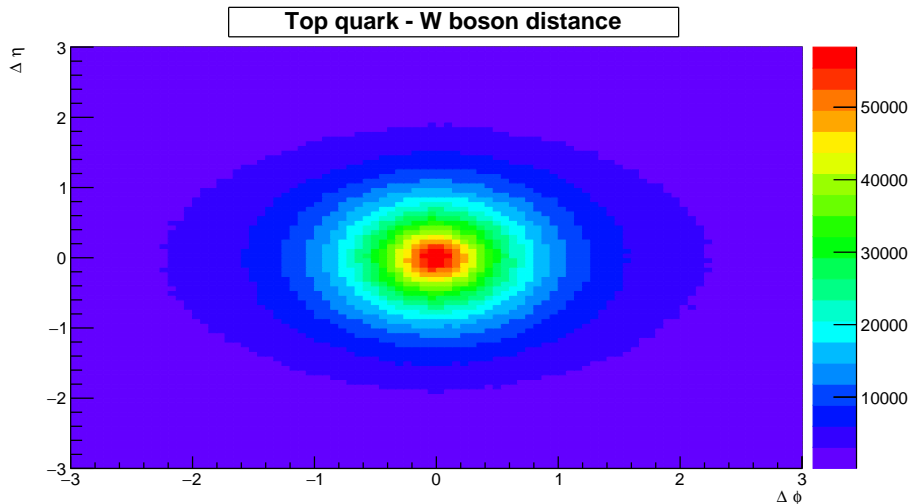


Figure 5.9: Distance in the $\Delta\phi$ - $\Delta\eta$ plane between the directions of each top quark and its related W boson.

As we can see, the high population of the region $|\Delta\phi| < 1, |\Delta\eta| < 1$ proves the relevance of boosted top quark decays.

In this section we try to find a way to improve the mass reconstruction and make use somehow of the events with boosted top quark decays.

The Cambridge-Aachen and Top-Tag algorithms As we said in section 3.3, the study of the top quark produced by LHC provides profound insights into the Standard Model and its possible extensions. In particular, almost every new physics scenario will include new heavy particles which decay to top quarks (like squarks in supersymmetry). If their masses are even a factor of a few above the top quark mass, the top quarks that they produce will be highly boosted, namely they will decay to collimated collections of particles that look like single jets. In this case, the standard top quark identification techniques described so far, may falter: b-tagging is difficult because the tracks are crowded and unresolvable and the W decay products (light-jets) are not always isolated from each other or from the b-jet. In the current LHC run (Run 2), there is also a larger probability to find highly boosted top quarks, because of the increased energy of the collider.

The Cambridge-Aachen algorithm [13] has been developed to solve these problems: it is a clustering algorithm useful to decompose highly boosted top quark jets into subjet components and examine the kinematic properties of these subjets. CA_jets are obtained by clustering particles into jets using the distance parameter $R = 0.8$. This is an iterative procedure which begins with all four-momenta in an event, as defined by the energy deposits in the calorimeter. It then finds the pair which is closest in $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, merges it into a single four-momentum, and then repeats. The procedure ends when no two four-momenta have $\Delta R < R$.

The information about the internal structure of the CA_jets may be useful to understand if these wider jets are produced by the boosted decay of top quarks. As a matter of fact, we can use the following Top-Tag algorithm [14]: at first each CA_jet in the event is declustered to look for subjets. In each step the jet is divided into two objects: if their energies are larger than a certain parameter and they are not too close, they are considered as subjets. Then the process is repeated on the subjets until all of them are irreducible. The algorithm then imposes kinematic cuts on the CA_jets and their subjets to understand if they come from a top quark decay: the total invariant mass should be near the top quark mass m_t , two subjets should reconstruct the W boson mass m_W , and the W boson helicity angle should be consistent with a top quark decay. The Top-Tag algorithm returns several quantities, like TopTagTopMass and TopTagTopMassMin. The first quantity is the invariant mass of the top-tagged CA_jets, while the second is the minimum invariant mass between all the possible couplings of the light subjets. Using these quantities we can make a selection on the top-tagged jets to increase the purity of the sample, excluding jets which do not come from top quark decay. Imposing $100 \text{ GeV} < \text{TopTagTopMass} < 250 \text{ GeV}$, $\text{TopTagMinMass} > 50 \text{ GeV}$ and the presence of at least three subjets, we obtain a non-top quark rejection of approximately 98-99% and a top quark efficiency of approximately 33%.

In this study, we use top-tagged CA_jets to find a different way to describe the all-hadronic decay and to reconstruct the top quark mass.

Boosted top quarks hypothesis validation Before explaining the mass reconstruction method based on CA_jets, we want to show the number of events containing boosted top quark decays with respect to the total amount of $t\bar{t}$ all-hadronic decays in the sample. Because we are working with MC samples, we can get the information of the parton that generated each jet. Thus we can select only those events which are characterized by one or two boosted top quark decays, asking that the decay products are in a cone of a tight enough opening. We say that a top quark decay is boosted if the ΔR between the bottom quark and the relative W boson is less than 0.8, namely when the decay products can be clustered in a single CA_jet.

In Table 5.3 we show the number of events N_{evt}^{boost} with only one, only two and at least one boosted top quarks for 13 TeV samples and their percentage with respect to the total number of $t\bar{t}$ all-hadronic decay events in the sample $N_{evt}^{all-had}$.

	N_{evt}	%
$N_{evt}^{all-had}$	7793065	-
$N_{evt}^{boost} = 1$	427400	5.5
$N_{evt}^{boost} = 2$	59207	0.8
$N_{evt}^{boost} \geq 1$	486607	6.2

Table 5.3: Number of boosted top quark decays in the 13 TeV MC sample.

We can thus see that at 13 TeV there is a relevant fraction of boosted top quarks. This means that we are allowed to look for a different mass reconstruction method which use the information about the boosted top quark decays clustered in CA_jets.

Mass reconstruction Let us describe the all-hadronic decay in terms of both jets and CA_jets.

Suppose to consider an event with at least a reconstructed CA_jet. As explained before, we will keep the jet only if it comes from a boosted top quark decay, namely if it is composed of at least three sub-jets, $100 \text{ GeV} < \text{TopTagTopMass} < 250 \text{ GeV}$, $\text{TopTagMinMass} > 50 \text{ GeV}$. [13]

If the CA_jet pass this selection, we search if at a distance of $\Delta R > 1$ in the η - ϕ space, there are at least three jets with at least one of them b-tagged (Medium tag).

Then, the following χ^2 -like function is minimized with respect to the free parameter m_t^{rec} for every compatible combinations of CA_jets and jets. For this so-called CA fit, We keep the best combination for each event.

$$\chi_{CA}^2 = \frac{(m_{jj} - M_W)^2}{\sigma_W^2} + \frac{(m_{jtb} - m_t^{rec})^2}{\sigma_t^2} + \frac{(m_{CA} - m_t^{rec})^2}{\sigma_t'^2},$$

where m_{jj} and m_{jjb} are respectively the invariant masses of the dijet and trijet systems, m_{CA} is the invariant mass of the subjets in the CA_jet and $\sigma_W = 12$ GeV, $\sigma_t = 20$ GeV, $\sigma'_t = 30$ GeV are the standard deviations associated to the dijet, trijet and CA_jets reconstructed masses. The value chosen for σ'_t accounts for the worse mass resolution observed using CA_jets.

Mass distributions We begin the mass analysis by showing, in Fig. 5.10, the action of the cuts $100 \text{ GeV} < \text{TopTagTopMass} < 250 \text{ GeV}$, $\text{TopTagMinMass} > 50 \text{ GeV}$ and $\text{TopTagNSubJets} \geq 3$ on the “raw” mass distribution of the top-tagged CA_jets.

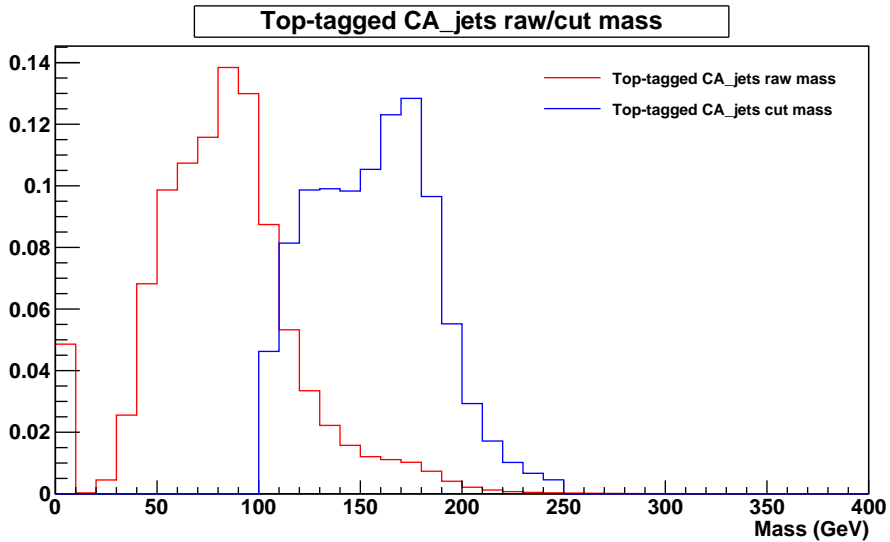


Figure 5.10: Comparison between the raw mass distribution from generic CA_jets and from top-tagged CA_jets, normalized to equal area.

As we can see, the application of the cuts produces a mass distribution which may be compatible with the top quark mass distribution, even if we can still notice the presence of background, especially in the interval 100-150 GeV, due to CA_jets which do not contain subjets that really come from a top quark decay. This background will be removed next with the CA fit, and a cut on χ^2_{CA} .

We want to prove now that we can rely on the new CA fit to reconstruct correctly the top quark mass. This happens when the mass distribution obtained with the new method has a peak value compatible with the one obtained with the standard method. Furthermore, the two distributions must have almost the same accuracy.

In Fig. 5.11 we plot together the mass distributions for the 13 TeV MC sample obtained using the standard and the CA fit, and asking for $\chi^2 \leq 3$ in both cases. Because we are testing the new method, we have applied both reconstruction algorithms on the same 13 TeV samples asking the standard selection and at least one CA_jet for each event.

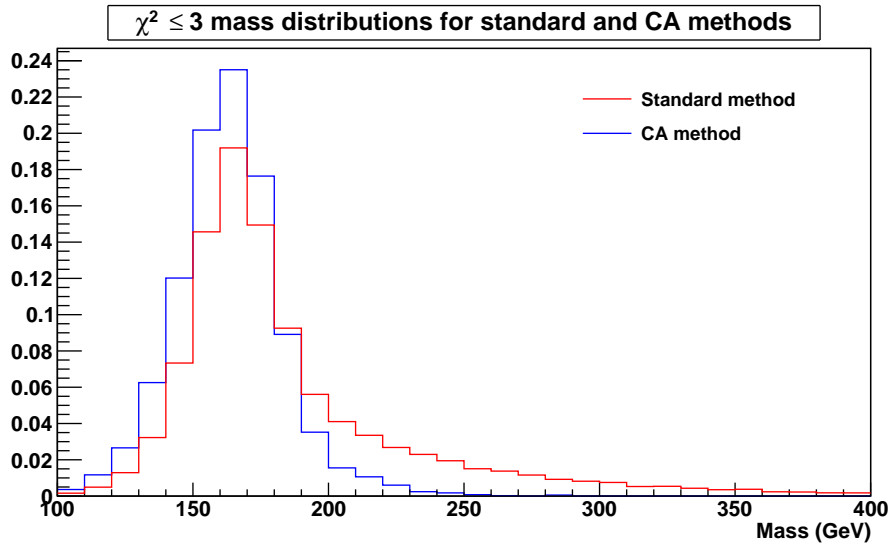


Figure 5.11: Comparison between the mass from the standard fit and the mass from the CA fit ($\chi^2 \leq 3$ for both cases), normalized to equal area.

As we can see, the peak values of the two distributions are compatible with each other. On the other hand, the distribution from the standard fit has a tail at high masses due to wrong jet-to-quark combinations. In fact the CA fit is inherently more precise because of the top-tag requirements.

Efficiency improvement In the previous paragraph we understood that we can rely on `CA_jets` to describe the $t\bar{t}$ all-hadronic decay in a different way and to reconstruct correctly the top quark mass. We want to use now the new reconstruction algorithm to improve the efficiency of the standard method when applied to 13 TeV MC samples.

Standard reconstruction (at either 8 and 13 TeV) works on MC samples with at least six jets for each event. This method may fail in presence of boosted top quark decays in the events. As a matter of fact it may happens that the decay products are too close in the η - ϕ space and they are then clustered in a single jet with a large momentum.

Using standard reconstruction two things may happen:

- if some jets merge, it may happens that the total number of jets in the event is less than six, thus the standard mass reconstruction discards the event and this results in a loss in efficiency;
- if the event has six or more jets including the merged jet, the standard mass reconstruction starts, but, because we are using the boosted top quark decay products as a single jet, the algorithm will yield high χ_{std}^2 values and so we loose part of the signal when a the cut on χ_{std}^2 is made.

We want to solve these two problems using the CA fit in order to increase the efficiency of the standard mass reconstruction, getting back the discarded events. We solve them separately, using different procedures.

Let us analyze the first problem, namely when some jets are clustered together and the event has less than six jets. In this case we have run the CA_jets reconstruction method on 13 TeV MC samples, asking for less than six jets, at least one b-tagged jet (Medium tag) and at least one CA_jet per event. We expect that some events lost using the standard reconstruction have small values of χ_{CA}^2 .

In Table 5.4 we show the number of events which pass different cuts on χ_{CA}^2 . The third column of the table indicates the relative increase in efficiency with respect to the standard selection $N_{jets} \geq 6$, $N_{b-jets} \geq 2$, $\chi_{std}^2 \leq 3$.

	N_{evt}	efficiency (%)	$\Delta\epsilon/\epsilon$ (%)
$\chi_{CA}^2 \leq 10$	2434	0.006	3.9
$\chi_{CA}^2 \leq 5$	1919	0.005	3.1
$\chi_{CA}^2 \leq 3$	1538	0.004	2.5

Table 5.4: Number of reconstructed events using the CA fit for different cuts on χ_{CA}^2 , asking for less than six jets per event. Shown are the absolute efficiency and the relative efficiency increase with respect to the standard fit ($\chi_{std}^2 \leq 3$).

Let us now analyze the second problem, namely when some jets are clustered together but the event still has at least six jets. We expect high χ_{std}^2 values for events with boosted top quark decays when the standard reconstruction is used. In turn, we expect for the same events, small χ_{CA}^2 values when the CA fit is applied. For this reason, if the standard mass reconstruction discard the event because it returns a χ_{std}^2 value higher than 3, we try to reconstruct the same event using the CA fit and compute the invariant mass. We then apply some cuts on χ_{CA}^2 to exclude undesired events.

The results are shown in Table 5.5.

	N_{evt}	efficiency (%)	$\Delta\epsilon/\epsilon$ (%)
$\chi_{CA}^2 \leq 10$	10060	0.024	16.1
$\chi_{CA}^2 \leq 5$	7796	0.018	12.5
$\chi_{CA}^2 \leq 3$	6063	0.014	9.8

Table 5.5: Number of reconstructed events using the CA fit for different cuts on χ_{CA}^2 , asking for at least six jets per event and a standard fit with $\chi_{std}^2 > 3$. Shown are the absolute efficiency and the relative efficiency increase with respect to the standard fit ($\chi_{std}^2 \leq 3$).

We summarize in Table 5.6 the effect of applying both procedures and using the CA fit for $t\bar{t}$ events failing the standard fit. A relative gain of as high as 20% can be obtained applying the CA fit with $\chi_{CA}^2 \leq 10$.

	N_{evt}	efficiency (%)	$\Delta\epsilon/\epsilon$ (%)
$\chi_{CA}^2 \leq 10$	12494	0.029	20.0
$\chi_{CA}^2 \leq 5$	9715	0.023	15.6
$\chi_{CA}^2 \leq 3$	7601	0.018	12.2

Table 5.6: Number of reconstructed events using the CA fit for $t\bar{t}$ events which failed the standard fit ($\chi_{std}^2 > 3$). Shown are the absolute efficiency and the relative efficiency increase with respect to the standard fit ($\chi_{std}^2 \leq 3$).

In Fig. 5.12 we show the mass distributions of the events recovered using the two procedures just described, comparing them to the mass distribution reconstructed with the standard fit. We asked $\chi_{std}^2 \leq 3$ for the standard distribution and $\chi_{CA}^2 \leq 10$ for the CA_jets distributions. The distributions have a similar width indicating that they will have a similar effectiveness when used for a future measurement of the top quark mass.

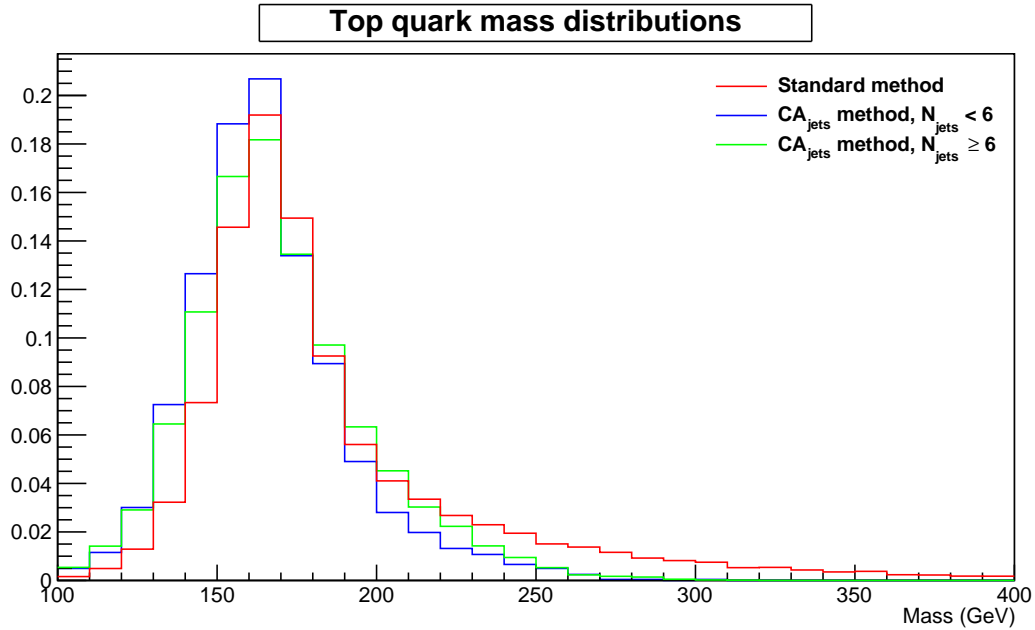


Figure 5.12: Top quark mass distributions reconstructed with the standard and CA fits (for $N_{jets} < 6$ and $N_{jets} \geq 6$), asking $\chi_{std}^2 \leq 3$ for the standard fit and $\chi_{CA}^2 \leq 10$ for the CA fit. The distributions are normalized to equal area.

Chapter 6

Conclusions

The new LHC run at 13 TeV has started this June, and a few fb^{-1} of pp data are expected to be collected by CMS before the end of the year.

The goal of this work has been the study of the kinematical characteristics of multijet $t\bar{t}$ events in order to reconstruct and eventually measure the top quark mass.

Given the higher center-of-mass energy with respect to the previous run, the standard mass reconstruction used so far might not be adequate, and in fact this study searched for improvements in the method. Indeed the use of boosted top quarks and a specific top-tagger based on them has been proven to be effective in improving the accuracy of a top quark mass measurement, while increasing the efficiency of the event selection. The final tuning of the procedure and the choice of the cuts on the mass fits requires a comparison with the data to assess the expected accuracy of a top quark mass measurement based on multijet events.

Bibliography

- [1] *LHC the guide*
<http://cds.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf>.
- [2] <http://home.web.cern.ch/topics/large-hadron-collider>.
- [3] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST 3 (2008) S08004.
- [4] <http://cms.web.cern.ch/news/hadron-calorimeter>.
- [5] <http://cms.web.cern.ch/news/muon-detectors>.
- [6] Melissa Franklin, *Physics at the proton anti-proton colliders*, Harvard University, 1990.
- [7] CDF Collaboration, *Observation of top quark production in pp collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. 74, (1995) 2626.
- [8] DØ Collaboration, *Observation of the top quark*, Phys. Rev. Lett. 74 (1995) 2632.
- [9] K. A. Olive *et al.* [Particle Data Group Collaboration], “Review of Particle Physics,” Chin. Phys. C **38** (2014) 090001.
- [10] CMS Collaboration, *Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E/T*, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.
- [11] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-kt jet clustering algorithm*, JHEP 04582 (2008) 063, arXiv:0802.1189.
- [12] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, JINST 8 (2012) P04013.
- [13] P. Maksimovic, M. Swartz, S. Rappoccio, *A Boosted Top-Jet Tagging Algorithm: CATopTag.*, CMS AN-2008/069.

- [14] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz and Brock Tweedie, *Top-tagging: A Method for Identifying Boosted Hadronic Tops*, Phys. Rev. Lett. 101 (2008) 142001.