Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea in Fisica

MEASUREMENT OF THE TOP QUARK MASS IN MULTIJET EVENTS WITH THE CMS EXPERIMENT AT LHC

Relatore:
Prof. Andrea Castro
Correlatori:
Prof. Daniele Bonacorsi
Dott. Giuseppe Codispoti

Presentata da:
Maria Giovanna Foti

Sessione autunnale, II appello
Anno Accademico 2013/2014

Desidero ringraziare in particolar modo il Professor Andrea Castro per la estrema cortesia, il garbo e la pazienza nei miei confronti, oltre che per la continua sollecitudine e costante presenza in ogni momento del percorso che ha portato a questa tesi.
Ringrazio molto anche il Professor Daniele Bonacorsi per il prezioso supporto, anche organizzativo, e per la attenzione che ha sempre avuto nei miei riguardi.
Un grande grazie va anche al Dottor Giuseppe Codispoti, per tutti i consigli che ha voluto regalarmi e la sua grande disponibilità ad ascoltare ogni mio dubbio o perplessità.
A chi mi è stato vicino lungo il cammino che ha portato a questo primo traguardo, va tutta la riconoscenza di cui sono capace.

Contents

Sommario

La massa del quark top è qui misurata per mezzo dei dati raccolti dall'esperimento CMS in collisioni protone-protone ad LHC, con energia nel centro di massa pari ad 8 TeV. Il campione di dati raccolto corrisponde ad una luminosità integrata pari a 18.2 fb^{-1}. La misura è effettuata su eventi con un numero di jet almeno pari a 6, di cui almeno due b-taggati (ovvero identificati come prodotto dell'adronizzazione di due quark bottom). Il valore di massa trovato è di $(173.95 \pm 0.43\ (\text{stat}))\ \text{GeV}/c^2$, in accordo con la media mondiale.

Abstract

The top quark mass is here measured by using the data that have been collected with the CMS experiment in proton-proton collisions at the LHC, at a center-of-mass energy of 8 TeV. The dataset which was used, corresponds to an integrated luminosiy of 18.2 fb^{-1}. The mass measurement is carried out by using events characterized by six or more jets, two of which identified as being originated by the hadronization of bottom quarks.
The result of the measurement of the top quark mass performed here is: (173.95 $\pm$ 0.43 (stat)) GeV/c^2, in accordance with the recently published world average.

Chapter 1

Introduction

The mass of the top quark is a fundamental parameter of the Standard Model (SM) and its large value makes the top quark contribution dominant in loop corrections to many observables, like the W boson mass. Precise measurements of the W boson and the top quark masses allow scientists to set indirect constraints on the mass of the Higgs boson. In this document a measurement of the t-quark mass is presented. This measurement is carried out by using proton-proton collision events at a center-of-mass energy of 8 TeV.

Top quarks are produced at the largest rate in pairs ($t\bar{t}$), with each top quark decaying immediately into a W boson and a b-quark nearly 100% of the time. In this analysis events where both the W's decay to a quark-antiquark pair are considered. This "all-hadronic" final state has the largest branching ratio among the possible decay channels (46%), but it is overwhelmed by the QCD multijet background processes, which surpass $t\bar{t}$ production by three orders of magnitude even after a dedicated trigger requirement. Nevertheless, it has been shown at CDF [1, 2] how this difficult background can be successfully controlled and significantly suppressed with a properly optimized event selection.

At the Large Hadron Collider (LHC) one can take advantage of the larger production cross section (approximately 40 times higher than at the Tevatron) [3, 4] to perform an accurate measurement of the top quark mass [5]. The large dataset which was used corresponds to an integrated luminosity of 18.2 fb^{-1} of proton-proton collisions collected between November 2012 and January 2013 by the Compact Muon Solenoid (CMS) detector at the LHC. Events used in this measurement are selected by a multijet trigger and retained only if they happen to have at least six "good" ($E_T \geq 30$ GeV and $\mid \eta \mid \leq 2.5$) jets, two of which identified as being b-jets. For each selected event, the six jets are assumed to come from the quarks of a $t\bar{t}$ all-hadronic

final state. Each of the different combinations where the jets are arranged in two doublets (the W bosons) and two triplets (the top quarks) is considered. To reduce the number of permutations, b-tagged jets are assumed to come from b-quarks only. For each permutation a top quark mass is obtained through a constrained fit based on the minimization of a χ^2-like function. A further request helps to improve the signal-to-background ratio (S/B).

After selecting the data, pseudo-experiments are performed assuming specific values for the top quark mass and these pseudo-data are therefore extracted from the corresponding signal (provided with Monte Carlo simulations, generated with different top quark masses) and background templates. The results of these pseudo-experiments have been compared to the input values and used to obtain, on average, a more reliable estimate of the true values and uncertainties.

Finally, a likelihood fit is applied to data and a top quark mass measurement has been obtained:

$$(173.95 \pm 0.43(stat))\ \mathrm{GeV}/c^2$$

where the uncertainty is statistical only.

As a final step, one of the most significant source of systematic uncertainty associated to the jet energy scale has been evaluated by performing pseudo-experiments using templates built with signal samples where effects due to systematic uncertainties have been included. The differences in the average values of the top quark mass with respect to the pseudo-experiments performed with default templates are then considered to estimate the associated systematic uncertainty.

Chapter 2

High-energy physics at LHC

This chapter describes the LHC, the CMS experiment and the main foucus of CERN's research.

2.1 The Large Hadron Collider

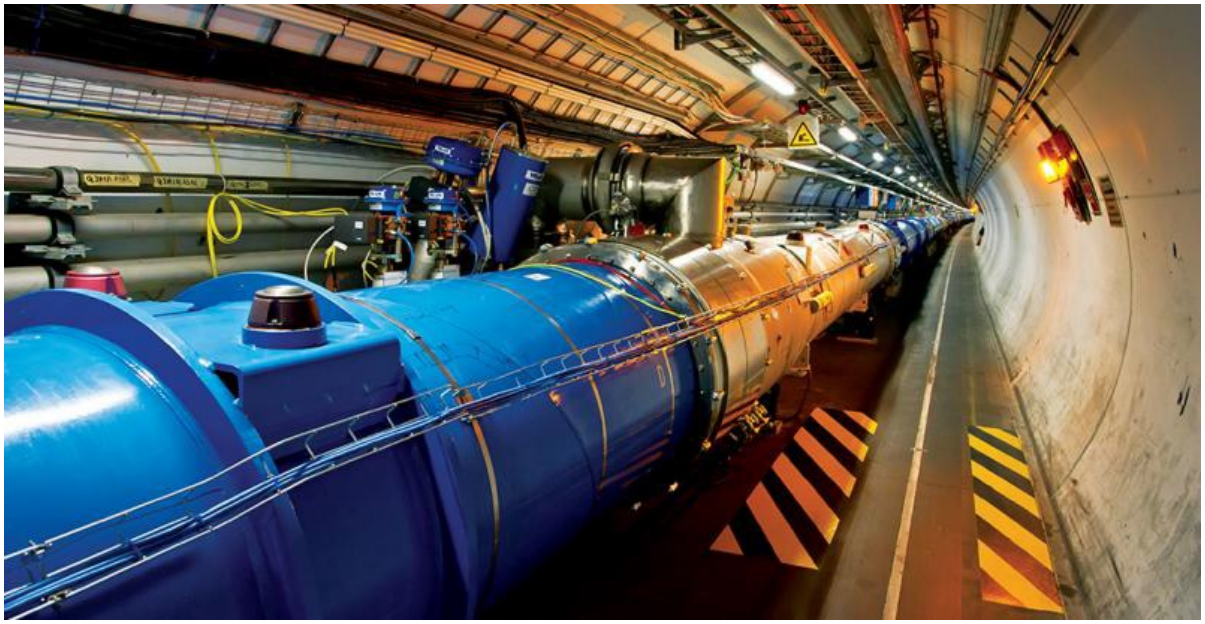


Figure 2.1: The LHC tunnel.

The LHC [6, 7, 8] is the world's largest and most powerful particle accelerator. LHC began operations on the 10 September 2008 and remains the latest addition to CERN's accelerator complex. The LHC consists of a

27-kilometre ring of superconducting magnets (see Fig. 2.1) with a number of accelerating structures to boost the energy of the particles along the way.

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. 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 $-271.3\ ^{\circ}C$ - 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, as well as to other supply services.

The history Back in the early 1980s, while the Large Electron-Positron (LEP) collider was being designed and built, groups at CERN were already busy looking at the long-term future. Scientists intended to use the existing LEP tunnel to install a higher energy machine - the LHC. When, on the 21st of October 1993, the US government voted to cancel the Superconducting Super Collider project, due to concerns linked to rising costs, the LHC became the sole candidate for a new high-energy hadron collider. After many years of work on the technical aspects and physics requirements of such a machine, these dreams came to fruition on the 16th of December 1994 when CERN's governing body, the CERN Council, voted to approve the construction of the LHC. The green light for the project was given under the condition that the new accelerator be built within a constant budget and on the understanding that any non-Member State contributions would be used to speed up and improve the project. Initially, the budgetary constraints implied that the LHC was to be conceived as a 2-stage project. However, following contributions from non-Member State such as Japan, the USA, India, Canada and Russia, the Council allowed the project to proceed in a single phase. On the 14th of April 1994, the first prototype (see Fig. 2.2) of the 1232 bending-magnets for the LHC reached a field of 8.73 Tesla (which is higher than the 8.4 Tesla field at which the LHC operated in 2012). It was produced by the Italian Instute of Nuclear Physics (INFN).

Between 1996 and 1998, four experiments - ALICE, ATLAS, CMS and LHCb - received official approval and construction work commenced on the four sites. In July 1998, as construction workers were preparing the work site for the CMS detector cavern, they unearthed 4th century Gallo-Roman ruins which turned out to be from an ancient villa with surroundings fields. The

Figure 2.2: The first prototype of bending-magnet for the LHC.

finding delayed work for 6 months while archaeologists excavated the site. On the 2nd of November 2000 the LEP collider was shut down for the last time and with the tunnel now available for work, teams began excavating the caverns to house the four big detectors on the LHC. In June 2003 the ATLAS detector cavern was complete and after six and a half years of work, on the 1st February 2005, CERN leaders and dignitaries celebrated the completion of a second detector cavern the CMS one: 53 metres long, 27 wide and 24 high. Finally, more than three years later, at 10.28 AM on the 10 September 2008 a beam of protons was successfully steered around LHC for the first time.

The machine The LHC is not a perfect circle. It is made of eight arcs and eight 'insertions', as shown in Fig. 2.3. The arcs contain the dipole 'bending' magnets (154 dipoles per arch), whilst an insertion consists of a long straight section plus two transition regions at its two ends. 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. There are three main ingredients in a particle accelerator:

Vacuum The LHC has the particularity of having three vacuum systems: the insulation vacuum for cryomagnets, the insulation one for the helium distribution line and the beam vacuum. The beam vacuum pres-

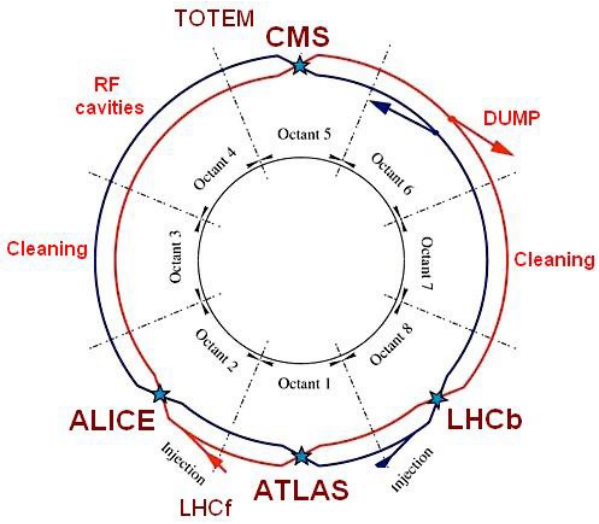


Figure 2.3: The structure of LHC.

sure is 10^{-13} atm (ultrahigh vacuum), because we need to avoid collisions with gas molecules.

Magnets There is a large variety of magnets in the LHC (dipoles, quadrupoles, sextupoles, octupoles, decapoles) giving a total of about 9600 magnets. Each type of magnet contributes to optimizing a particle's trajectory: dipoles have the function to maintain the beams in their circular orbit, whilst insertion quadrupoles are used to focus the beam down to the smallest possible size at the collision points, thereby maximizing the chance of two protons smashing head-on into each other.

The dipoles of the LHC represented the most important technological challenge for the LHC design. 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 which are able to provide the very high field of 8.3 T over their length. No practical solution could have been designed using 'warm' magnets instead of superconducting ones. The LHC dipoles use niobium-titanium (NbTi) cables, which become superconducting below a temperature of 10 K, that is, they conduct electricity without resistance (the LHC in fact operates at 1.9 K, even lower than the temperature of outer space, 2.7 K). The refrigeration process happens in three phases: firstly the accelerator is cooled down to 4.5 K, then the magnets are filled with liquid helium and then the final cool down to

1.9 K occurs.

Cavities The main role of the LHC cavities is to keep the 2808 proton bunches tightly bunched to ensure high luminosity at the collision points and hence, maximize the number of collisions. They also deliver radiofrequency (RF) power to the beam during acceleration to the top energy. Again superconducting cavities with small energy losses and large stored energy are the best solution.

LHC is not the only accelerator used at CERN, but a succession of machines (see Fig. 2.4) serves the purpose of accelerating the protons beams to increasingly higher energies. Each machine boosts the energy of a beam of particles, before injecting the beam into the next machine in the sequence, up to the record energy of 4 TeV per beam.

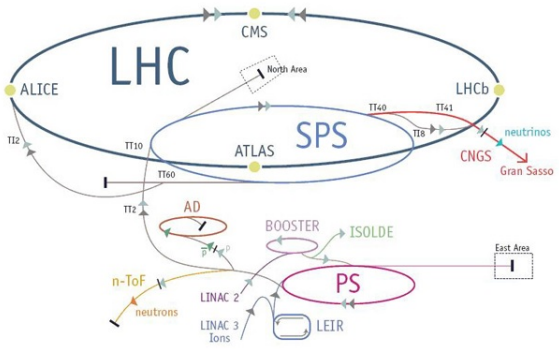


Figure 2.4: There is more to CERN than the LHC: a series of accelerators work together to push particles to nearly the speed of light.

The proton source is a simple bottle of hydrogen gas. An electric field is used to strip hydrogen atoms of their electrons to yield protons. Linac 2, the first accelerator in the chain, accelerates the protons to the energy of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV.

The protons are finally transferred to the two beam pipes of the LHC. The beam in one pipe circulates clockwise while the beam in the other pipe circulates anticlockwise. It takes 4 minutes and 20 seconds to fill each LHC ring and 20 minutes for the protons to reach their maximum energy of 4 TeV. Beams circulate for many hours inside the LHC beam pipes under normal

operating conditions. The two beams are brought into collision inside the four detectors - ALICE, ATLAS, CMS and LHCb - where the total energy at the collision point is equal to 8 TeV.

The accelerator complex includes the Antiproton Decelerator and the Online Isotope Mass Separator (ISOLDE) facility and feeds the CERN Neutrinos to Gran Sasso (CNGS) project and the Compact Linear Collider test area, as well as the neutron time-of-flight facility (nTOF).

Protons are not the only particles accelerated in the LHC. Lead ions for the LHC start from a source of vaporized lead and enter Linac 3 before being collected and accelerated in the Low Energy Ion Ring (LEIR). They then follow the same route to maximum energy as the protons.

2.2 The CMS experiment

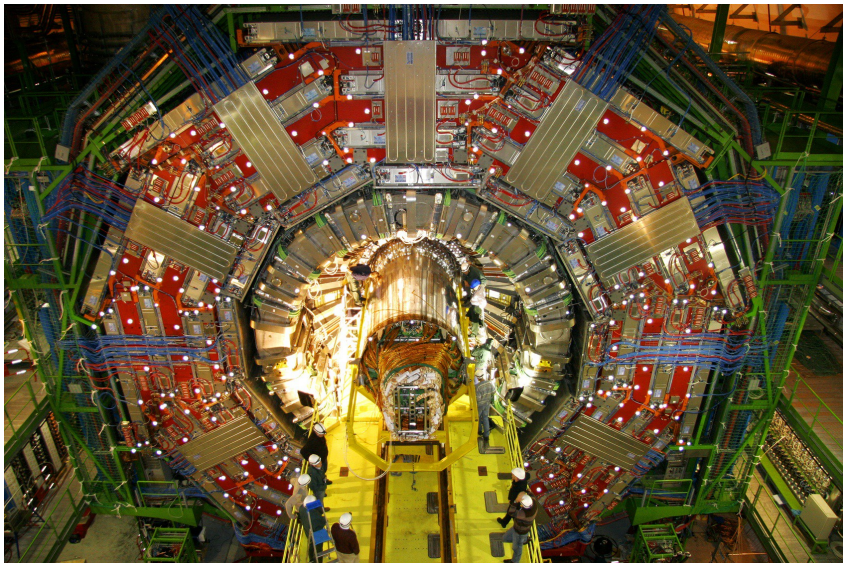


Figure 2.5: The CMS detector uses a huge solenoidal magnet to bend the paths of particles from collisions in the LHC.

The CMS [9, 10] is a general-purpose detector at the LHC. It is designed to investigate a wide range of physics, including the search for the Higgs boson, extra dimensions and particles that could make up dark matter. CMS is also designed to measure the properties of previously discovered particles with unprecedented precision and be on the lookout for completely new, unpredicted phenomena. Although it has the same scientific goals as

the ATLAS experiment, it uses different technical solutions and a different magnet-system design.

The CMS detector (see Figs. 2.5 and 2.6) is built around a huge solenoidal magnet. This takes the form of a cylindrical coil of superconducting cable that generates a field of 4 T, which is confined by a steel "yoke" that forms the bulk of the detector's 12,500-tonne weight.

The CMS experiment is one of the largest international scientific collaborations in history, involving 4300 particle physicists, engineers, technicians, students and support staff from 182 institutes in 42 countries (February 2014).

The CMS detector CMS is a particle detector that is designed to detect a wide range of particles and phenomena produced in high-energy collisions at the LHC. It is built around a huge superconducting solenoid and different layers of detectors measure the different particles and use this key data to build up a picture of events at the heart of the collision.

The detector is like a giant filter, where each layer is designed to stop, track or measure a different type of particle emerging from proton-proton and heavy ion collisions. Finding the energy and momentum of a particle gives clues to its identity and particular patterns of particles or "signatures" are indications of new and exciting physics.

The detector consists of layers of material that exploit the different properties of particles to catch and measure the energy and momentum of each one. In order to work correctly, CMS needs:

- a high quality central tracking system to give accurate momentum measurements;
- a high resolution method to detect and measure electrons and photons (an electromagnetic calorimeter);
- a "hermetic" hadron calorimeter, designed to entirely surround the collision and prevent particles from escaping;
- a high performance system to detect and measure muons.

With these priorities in mind, the first essential item is a very strong magnet. The higher a charged particle's momentum, the less its trajectory is curved in the magnetic field, therefore by knowing the details of this path, the particle's momentum can be easily measured. Thus, a strong magnet is needed to allow scientists to accurately measure even the very high momentum particles, such as muons.

Particles emerging from collisions first meet a tracker, made entirely of silicon, that charts their positions as they move through the detector, allowing scientists to measure their momentum. Outside the tracker are calorimeters that measure the energy of particles. In measuring the momentum, the tracker should interact with the particles as little as possible, whereas the calorimeters are specifically designed to stop the particles in their tracks.

As the name indicates, CMS is also designed to measure muons. The muon tracks are measured by four layers of muon detectors, while the neutrinos escape from CMS undetected, although their presence can be indirectly inferred from the "missing transverse energy" in the event.

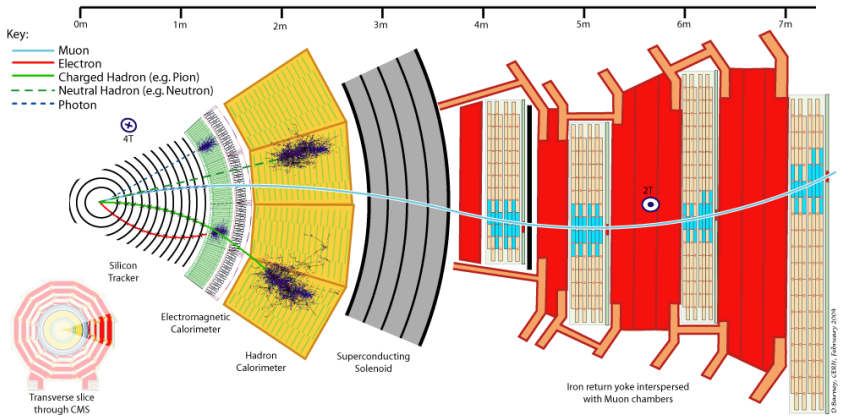


Figure 2.6: Different particles correspond to different trajectories troughout the layers of the CMS detector.

The magnet The CMS magnet is a solenoid, that is a coil of superconducting wire, and creates a magnetic field when electricity flows through it. In CMS the solenoid has an overall length of 13 m and a diameter of 7 m and a magnetic field of about 4 T. It is the largest magnet of its type ever constructed and allows the tracker and calorimeter detectors to be placed inside the coil, resulting in a detector that is, overall, "compact", compared to detectors of similar weight.

The tracker The momentum of particles is crucial in helping us build up a picture of events at the heart of the collision. One method to evaluate the momentum of a particle is to track its path through a magnetic field and the CMS tracker records the paths taken by charged particles by finding their positions at a number of key points. The tracker can reconstruct the paths of high-energy muons, electrons and hadrons as well as see tracks coming from the decay of very short-lived particles

such as b-quarks. Moreover, it needs to record particle paths accurately yet be lightweight so as to disturb the particle as little as possible. It does this by taking position measurements so accurate that tracks can be reliably reconstructed using just a few measurement points. Each measurement is accurate to 10 μm. It is also the inner most layer of the detector and so receives the highest volume of particles: the construction materials were therefore carefully chosen to resist radiation.

The final design consists of a tracker made entirely of silicon: the **pixels**, at the very core of the detector and dealing with the highest intensity of particles, and the silicon **microstrip detectors** that surround it. As particles travel through the tracker the pixels and microstrips produce tiny electric signals that are amplified and detected.

The **pixel detector**, about the size of a shoebox, contains 65 million pixels, allowing it to track the paths of particles emerging from the collision with extreme accuracy. It is also the closest detector to the beam pipe, with cylindrical layers at 4, 7 and 11 cm and disks at either end, and so it is crucial in reconstructing the tracks of very short-lived particles. However, being so close to the collision means that the number of particles passing through is huge: the rate at 8 cm from the beam line amounts to about 10 million particles per square centimetre per second. Despite this huge number of particles passing through, the pixel detector is able to disentangle and reconstruct all the tracks they leave behind. When a charged particle passes through, it gives enough energy for electrons to be ejected from the silicon atoms, creating electron-hole pairs. Each pixel uses an electric current to collect these charges on the surface as a small electric signal which is then amplified.

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 15,200 highly sensitive modules with a total of 10 million detector strips read by 80,000 microelectronic chips. Each module consists of three elements: a set of sensors, its mechanical support structure and readout electronics.

The ECAL The Electromagnetic Calorimeter (ECAL) measures the energy of photons and electrons. But to find them with the necessary precision in the very strict conditions of the LHC - a high magnetic field, high levels of radiation and only 25 nanoseconds between collisions - requires very particular detector materials. The lead tungstate crystal is made primarily of metal and is heavier than stainless steel, but with

a touch of oxygen in this crystalline form it is highly transparent and "scintillates" when electrons and photons pass through it. This means the cystal produces light in proportion to the particle's energy. Photodetectors are glued onto the back of each of the crystals to detect the scintillation light and convert it to an electrical signal that is amplified and processed. The ECAL, made up of a barrel section and two "endcaps", forms a layer between the tracker and the HCAL. The cylindrical "barrel" consists of 61,200 crystals formed into 36 "supermodules", each weighing around three tonnes and containing 1700 crystals. The flat ECAL endcaps seal off the barrel at either end and are made up of almost 15,000 further crystals.

The HCAL Following a particle emerging from the ECAL, the next layer is constituted by the Hadron Calorimeter. It measures the energy of hadrons. Additionally it provides indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos. It is made by 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. Measuring hadrons is important as they can tell us if new particles such as the Higgs boson or supersymmetric particles (much heavier versions of the standard particles we know) have been formed. As these particles decay, they may produce new particles that do not leave record of their presence in any part of the CMS detector. To spot these the HCAL must be "hermetic", so as to capture to the extent possible every particle emerging from the collisions. In this way if we see particles shoot out on one side of the detector, but not on the other, with an imbalance in the momentum and energy (measured in the sideways "transverse" direction relative to the beam line), we can deduce that "invisible" particles have been produced.

The muon detectors As the name "Compact Muon Solenoid" suggests, detecting muons is one of CMS's most important tasks. Because muons can penetrate several metres of iron without interacting, unlike most particles they are not stopped by any of CMS's calorimeters. Therefore, chambers to detect muons are placed at the very edge of the detector where they are the only particles likely to produce a signal. A particle is measured by fitting a curve to hits among the four muon stations. In total there are 1400 muon chambers: 250 *drift tubes* (DTs) and 540 *cathode strip chambers* (CSCs) track the particles' positions and

provide a trigger, while 610 *resistive plate chambers* (RPCs) form a redundant trigger system, which quickly decides to keep the acquired muon data or not. Because of the many layers of detector and different specialities of each type, the system is naturally robust and able to filter out background noise. DTs and RPCs are arranged in concentric cylinders around the beam line ("the barrel region") whilst CSCs and RPCs, make up the "endcaps" disks that cover the ends of the barrel.

Within the LHC, bunches of particles collide up to 40 million times per second, so a "trigger" system that saves only potentially interesting events is essential. This reduces the number of events recorded from one billion to around 100 per second.

When CMS is performing at its peak, about one billion proton-proton interactions take place every second inside the detector. There is no way that data from all these events could be read out and even if they could, most of the events would be less likely to reveal new phenomena; they might be low-energy glancing collisions for instance, rather than energetic, head-on interactions. We therefore need a "trigger" that can select the potentially interesting events, such as those which will produce the Higgs particle, and reduce the rate to just a few hundred "events" per second, which can be read out and stored on computer disks for subsequent analysis. However, with groups of protons colliding 40 million times per second there are only 25 nanoseconds before the next lot arrive. New waves of particles are being generated before those from the last event have even left the detector! The solution is to store the data in pipelines that can retain and process information from many interactions at the same time. To not confuse particles from two different events, the detectors must have very good time resolution and the signals from the millions of electronic channels must be synchronised so that they can all be identified as being from the same event.

2.3 Physics at the LHC

CERN's main focus is particle physics - i.e. the study of the fundamental constituents of matter - but the physics programme at the laboratory is much broader, ranging from nuclear to high-energy physics, from studies of antimatter to the possible effects of cosmic rays on clouds.

Since the 1970s, particle physicists have described the fundamental structure of matter using an elegant series of equations called the SM. The model describes how everything that they observe in the Universe is made from a few basic blocks called fundamental particles, governed by four forces.

Physicists at CERN use the world's most powerful particle accelerators and detectors to test the predictions and limits of the SM. Over the years the SM has explained many experimental results and precisely predicted a range of phenomena, such that today it is considered a well-tested physics theory.

However, our current understanding of the Universe through the SM is incomplete. The SM has been tested by various experiments and it has proven sucessful in anticipating the existence of previously undiscovered particles. Yet the model describes only 4% of the matter of the known Universe and leaves many unsolved questions. Will we see a unification of forces at the high energies of the LHC? Why is gravity so weak? Why is there more matter than antimatter in the Universe? Is there more exotic physics waiting to be discovered at higher energies? Will we discover evidence for a theory called supersymmetry at the LHC? Or understand the Higgs boson that gives particles mass? Hopefully, LHC will help to answer these important questions.

The SM does not explain the origin of mass, nor why some particles are very heavy while others have no mass at all. The answer may be the so-called Higgs mechanism. According to the theory of the Higgs mechanism, the whole space is filled with a 'Higgs field' and, by interacting with this field, particles acquire their masses. Particles that interact intensely with the Higgs field are heavy, while those that have feeble interactions are light. The Higgs field has at least one new particle associated with it, the Higgs boson, and experiments at the LHC have been able to detect it. The SM does not offer a unified description of all the fundamental forces, as it remains difficult to construct a theory of gravity similar to those for the other forces. Supersymmetry - a theory that hypothesises the existence of more massive partners of the standard particles we know - could facilitate the unification of fundamental forces. If supersymmetry is right, then the lightest supersymmetric particles should be found at the LHC.

Cosmological and astrophysical observations have shown that all of the visible matter accounts for only 4% of the Universe. The search is open for particles or phenomena responsible for dark matter (23%) and dark energy (73%). A very popular idea is that dark matter is made of neutral - but still undiscovered - supersymmetric particles. The first hint of the existence of dark matter came in 1933, when astronomical observations and calculations of gravitational effects revealed that there must be more 'stuff' present in the Universe than we could account for by sight. Researchers now believe that the gravitational effect of dark matter makes galaxies spin faster than expected and that its gravitational field deviates the light of objects behind it. Measurements of these effects show the existence of dark matter and can be used to estimate its density even though we cannot directly observe

it. Dark energy, on the other hand, is a form of energy that appears to be associated with the vacuum in space and makes up approximately 70% of the Universe. Dark energy is homogenously distributed throughout the Universe and in time. In other words, its effect is not diluted as the Universe expands. The even distribution means that dark energy does not have any local gravitational effects, but rather a global effect on the Universe as a whole. This leads to a repulsive force, which tends to accelerate the expansion of the Universe. The rate of expansion and its acceleration can be measured by experiments using the Hubble law. These measurements, together with other scientific data, have confirmed the existence of dark energy and have been used to estimate its quantity.

The LHC will also help us investigate the mystery of antimatter. Matter and antimatter must have been produced in the same amounts at the time of the Big Bang, but from what we have observed so far, our Universe is made only of matter. Why? The LHC could help to provide an answer. It was once thought that antimatter was a perfect 'reflection' of matter - that if you replaced matter with antimatter and looked at the result as if in a mirror, you would not be able to tell the difference. We now know that the reflection is imperfect and this could have led to the matter-antimatter imbalance in our Universe. The strongest limits on the amount of antimatter in the Universe come from the analysis of the "diffuse cosmic gamma-rays" and the inhomogeneities of the cosmic microwave background (CMB). Assuming that after the Big Bang, the Universe separated somehow into different domains where either matter or antimatter was dominant, it is evident that at the boundaries there should be annihilations, producing cosmic (gamma) rays. Taking into account annihilation cross-sections, distance and cosmic redshifts, this leads to a prediction of the amount of diffuse gamma radiation that should arrive on Earth. The free parameter in the model is the size of the domains. Comparing with the observed gamma-ray flux, this leads to an exclusion of any domain size below 3.7 giga light years, which is not so far away from the entire Universe. Another limit comes from analyzing the inhomogeneities in the CMB: antimatter domains (at any size) would cause heating of domain boundaries and show up in the CMB as density fluctuations. The observed value of 10^{-5} sets strong boundaries to the amount of antimatter in the early Universe.

In addition to the studies of proton-proton collisions, heavy-ion collisions at the LHC will provide a window onto the state of matter that would have existed in the early Universe, called "quark-gluon plasma". When heavy ions collide at high energies they form for an instant a "fireball" of hot, dense matter that can be studied by the experiments.

Chapter 3

The top quark

In this chapter information about the discovery, the properties and the phenomenology of the top quark will be provided.

3.1 Discovery of the top quark

In 1964 Murray Gell-Mann and George Zweig proposed the quark hypothesis to account for the explosion of subatomic particles discovered in accelerator and cosmic ray experiments during the 1950s and early 1960s. Over a hundred new particles, most of them strongly interacting and very short-lived, had been observed. These particles, called hadrons, are not elementary: they possess a definite size and internal structure and most of them can be transformed from one state into another. The quark hypothesis suggested that different combinations of three quarks - the up (u), down (d) and strange (s) quarks - and their antiparticles - could account for all of the hadrons then known.

Most physicists were initially reluctant to believe that quarks were anything more than convenient abstractions helping particle classification. The fractional electric charges seemed bizarre and experiments repeatedly failed to turn up any individual free quarks. Two major developments established the reality of quarks during the 1970s. Fixed-target experiments directing high-energy leptons at protons and neutrons showed that these hadrons contain point-like internal constituents whose charges and spins are just what the quark model had predicted. And in 1974 experiments at Brookhaven National Laboratory in New York and Stanford Linear Accelerator Center (SLAC) in California discovered a striking new hadron at the then very large mass of 3.1 GeV/c^2, over three times that of the proton. This hadron was found to be a bound state of a new kind of quark, called charm or c, with

its antiquark. With two quarks of each possible charge, a symmetry could be established between the quarks and the leptons. Two pairs of each were then known: (u,d) and (c,s) for quarks and (e, ν_e) and (μ, ν_μ) for leptons, satisfying theoretical constraints. But this symmetry was quickly broken by unexpected discoveries. In 1976 experiments at SLAC turned up a third charged lepton, the tau lepton or τ. A year later at the Fermi National Accelerator Laboratory in Illinois a new hadron was discovered, called the upsilon, at the huge mass of about 10 GeV/c^2 and it was soon found to be the bound state of another new quark: the *bottom* or b-quark and its antiparticle. Experiments at DESY in Germany and Cornell in New York measured its fundamental properties.

With these discoveries and through the development of the SM, physicists then understood that matter comes in two parallel but distinct classes: quarks and leptons. They occur in "generations" of two related pairs with differing electric charge, but the third-generation quark doublet seemed to be missing its charge $+2/3$ 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.

Using the ratios of the observed quark masses, some physicists naively suggested that the t might be about three times as heavy as the b and thus expected that the top quark would appear as a heavy new hadron containing a $t\bar{t}$ pair, at a mass around 30 GeV/c^2. The electron-positron colliders then under construction (PEP at SLAC and PETRA at DESY) raced to capture the prize, but they found no hint of the top quark.

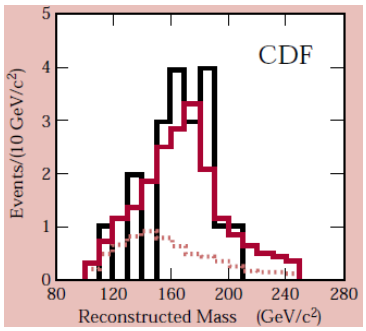
In the early 1980s a new class of accelerators came into operation at CERN in Switzerland, in which counter-rotating beams of protons and antiprotons collided with an energy of about 600 GeV. The protons and antiprotons brought their constituent quarks and antiquarks into collision with typical energies 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, producing *jets*, that is collimated sprays of particles.

In 1992 the DØ detector joined CDF as a long Fermilab Tevatron run began. Further searches would have to rely on the production of separate top and antitop quarks from annihilation of incoming quarks and antiquarks in the proton and antiproton, with subsequent decays into observable particles. The two experiments, while searching for the same basic decay sequence, had rather complementary approaches. First, DØ published a new lower limit of 131 GeV/c^2 on the possible top quark mass from the absence of events with

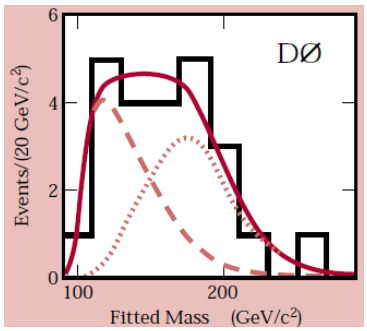
the characteristic dilepton or single lepton signatures.

After years spent on the long search, on the 24th February 1995, the observation of the top quark has been announced by the two experiments at the Tevatron proton-antiproton collider at Fermilab. In its paper [11], the CDF Colaboration reported finding six dilepton events plus 43 single-lepton events; they concluded that the odds were only one in a million that background fluctuations could account for these events. The DØ Collaboration, in its paper [12], observed three dilepton events plus 14 single-lepton events and concluded that the odds were two in a million that these could have been caused by backgrounds. The top quark masses reported by the two experiments were (176 ± 13) GeV/c^2 for CDF and (199 ± 30) GeV/c^2 for DØ (see Figs. 3.1(a) and 3.1(b)).

The top quark appears to be a point-like particle; it has no internal structure that we can discern. As one of the six fundamental constituents of matter, it has properties very similar to the up and charm quarks, with the exception of its remarkable massiveness and its very short lifetime. The top quark is about 200 times more massive than the proton, about 40 times heavier than the second heaviest quark (the b-quark) and roughly as heavy as the entire gold nucleus.



(a) Reconstructed mass (CDF).



(b) Fitted mass (DØ).

Figure 3.1: Since the top quark has a unique mass, the data (indicated by the black histograms) should show a "peak" in the reconstructed distribution. The non-top background (the red dashed curve for DØ and the red dotted curve for CDF) has very different shapes. The red dotted curve for DØ shows the expected contribution from the top quark for the best fit value of the top quark mass. The solid red curve shows what a simulated top quark mass distribution would look like when added to the background. These curves should be compared to the actual data.

3.2 Properties of the top quark

The top quark, also known as the t-quark, is one of the six existing quarks. According to the SM, it belongs to the third family, along with its "paired quark": the bottom or b-quark. Like all quarks, the top quark, as well as its antiparticle, is a fermion with spin $\frac{1}{2}\,\hbar$ and experiences all four fundamental interactions: gravitation, electromagnetism, weak interactions and strong interactions. It has an electric charge of $+\frac{2}{3}e$ and is the most massive of all observed elementary particles. Its large mass is the main reason why it was only discovered in 1995 at Fermilab collider with 1.8 TeV center of mass energy. Since its discovery the measurement of the top quark mass has become more and more accurate [13, 14].

3.3 Top quark production and decay

Top quark production At the LHC, top quarks are mostly produced in pairs via strong interaction ($\sigma_{t\bar{t}} \approx 250$ pb for $\sqrt{s} = 8$ TeV); however, there are a significant number of top quarks which are produced singly, via the weak interaction ($\sigma_t \approx 115$ pb for $\sqrt{s} = 8$ TeV).

In hadron collisions, the top quark is primarily produced via the strong interaction, therefore in couples of top and antitop quarks. At leading order (LO), there are only a few processes which describe the production of $t\bar{t}$ quark production: the dominant production mechanism is from gluon-fusion ($\approx$ 85%), while $q\bar{q}$ annihilation accounts for about 15%. The Feynman diagrams for the mechanism are shown in Fig. 3.2.

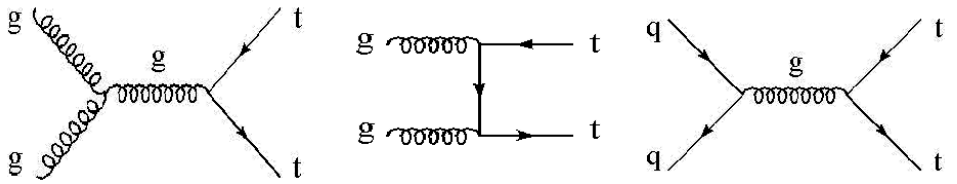


Figure 3.2: Feynman diagrams for $t\bar{t}$ production at LHC.

Single top production proceeds through three separate sub-processes at LHC: t-channel (the dominant process involves the exchange of a space-like W boson. This process is also called W-gluon fusion, because the b-quark ultimately arises from a gluon splitting to a bottom and antibottom quark), s-channel (involves the production of a time-like W boson, which then decays to a top and a bottom quark) and tW-channel, involving the production of a real W boson.

Top quark decay Top quarks are produced mainly by strong interactions and then decay through the weak force. The top quark decays almost exclusively to a W boson and a bottom quark, hence $t\bar{t}$ final states are classified by the decay products of the W boson which can decay to leptons and quarks. Three final states are distinguished:

- The dilepton channel, in which both W boson decay into a couple lepton-neutrino:

$$t\bar{t} \longrightarrow W^+b\ W^-\overline{b} \longrightarrow \bar{l}\nu_l b\ l'\overline{\nu_{l'}}\overline{b}$$

  with a branching ratio (BR) of 5%.

- The single-lepton channel, in which only one W decays into a couple lepton-neutrino:

$$t\bar{t} \longrightarrow W^+b\ W^-\overline{b} \longrightarrow \bar{l}\nu_l b\ qq'\overline{b}$$

  or:

$$t\bar{t} \longrightarrow W^+b\ W^-\overline{b} \longrightarrow qq'b\ l\overline{\nu_l}\overline{b}$$

  with a branching ratio (BR) of 30%.

- The all-hadronic channel, in which both W bosons decay hadronically, as represented in Fig. 3.3:

$$t\bar{t} \longrightarrow W^+b\ W^-\overline{b} \longrightarrow qq'b\ qq'\overline{b} \longrightarrow j_1j_2j_3\ j_4j_5j_6$$

  with a branching ratio (BR) of 46%.

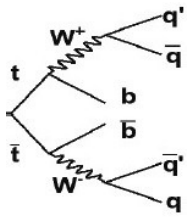


Figure 3.3: Feynman diagrams for $t\bar{t}$ all-hadronic decay.

Fig. 3.4 summarizes the possible final states and their branching ratios.

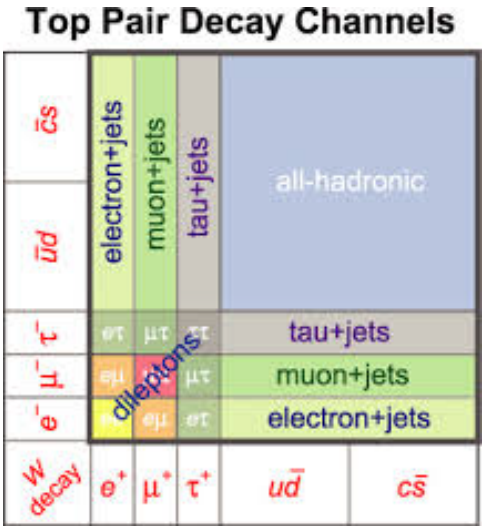


Figure 3.4: Possible final states for a $t\bar{t}$ system.The areas for each channel are proportional to the corresponding branching ratios.

Chapter 4

Data analysis

In this chapter the experimental procedure that has led to the mass measurement will be described.

4.1 Monte Carlo simulations

To perform a precision measurement, it is important to have the most complete understanding of the physics under study as well as the best possible description of the detector response formalized through its computer simulation. 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 p-p collision data, using the same reconstruction software and particle identification algorithms. Monte Carlo simulations have been used as a term of comparison of the data, since they have allowed to separate the signal events from the background ones, thus improving the S/B ratio. Since the purpose of this work is to provide a measurement of the top quark mass, seven MC samples have been used, each generated assuming a different value of the top quark mass: 169.5, 171.5, 172.5, 173.5, 175.5, 177.0, 178.5 GeV/c^2.

4.2 Selection of data

The final state of all-hadronic $t\bar{t}$ events is characterized by the presence of at least six jets from the decay of the two top quarks (we will call "quarks" both the particle and the anti-particle), where additional jets might come from initial or final state radiation. Events having such a topology are collected using a multijet trigger which relies on calorimeter information. The trigger

we have used here is the HLT_QuadJet50: a multijet trigger requiring four jets with transverse momentum $P_T \geq 50$ GeV/c. After applying this trigger, we expect to have a signal-background ratio $S/B \approx 1/76$. After a preliminary selection of multijet events, an additional selection based on a χ^2-function is used to further improve the purity of the sample.

Given a data sample, the number of expected signal events (with "signal events" we mean events from $t\bar{t}$ production) is:

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

Where BR is the branching ratio of the selected channel (in our case the all-hadronic one), $\sigma_{t\bar{t}}$ is the cross section of the $t\bar{t}$ process (we have used the value of 250 pb) and L the integrated luminosity of the used data sample (18.2 fb^{-1}). The first parameter ϵ is called "efficiency" and is defined as:

$$\epsilon = \frac{N_{selected}^{MC}}{N_{generated}^{MC}}$$

4.2.1 Preselection: trigger on the events, requests on the number of jets, kinematic cuts

Events satisfying the multijet trigger requirements are reconstructed in terms of their final state observables (tracks, vertices, charged leptons and jets). In our preselection (operated by the first part of the ROOT [15] macro `countEvents.C`), we retain only those events that present some properties:

- They are characterized by at least six jets, two of which identified as originating from a b-quark (b-tag);
- Each of the jets must have $P_T \geq 30$ GeV/c and $\mid \eta \mid \leq 2.5$;
- Each of the two b-tagged jets is identified using the CSV tagger on its "tight" working point.

After these kinematic cuts we expect to have $S/B \approx 1/6.5$ (see Tab. 4.1), where we define S/B as:

$$\frac{S}{B} = \frac{N_{evt}^{MC} \times L_{data}/L_{MC}}{N_{evt}^{data}}$$

where N_{evt}^{MC} is the number of the events that passes the required cuts for the default Monte Carlo $t\bar{t}$ sample (i.e. the one for 172.5 GeV/c^2), N_{evt}^{data} is the number of the events that passes the required cuts for the data sample and the ratio L_{data}/L_{MC} is a normalization factor that permits to scale the MC number of events in respect to the data luminosity (data and MC luminosity do not always have the same value).

4.2.2 The b-tagging

Jets that arise from bottom-quark hadronization (b-jets) are present in many physics processes, such as the decay of top quarks, the Higgs boson and various new particles predicted by supersymmetric models. Therefore, the ability to accurately identify b-jets is crucial in reducing the otherwise overwhelming background to these channels from processes involving jets from gluons (g) and light-flavour quarks (u, d, s), and from c-quark fragmentation. The properties of the bottom hadrons can be used to identify the hadronic jets into which the b-quarks fragment. These hadrons have relatively large masses, long lifetimes and daughter particles with hard momentum spectra. Their semileptonic decays can be exploited as well. The CMS detector, with its precise charged-particle tracking and robust lepton identification systems, is well matched to the task of b-jet identification (b-jet tagging) [16].

We will focus on two aspects of the b-jet tagging at CMS: the algorithms and the measurements of their performance.

B-tagging algorithms A variety of reconstructed objects - tracks, vertices and identified leptons - can be used to build observables that discriminate between b and light-parton jets. Several simple and robust algorithms use just a single observable, while others combine several of these objects to achieve a higher discrimination power. Each of these CMS algorithms yields a single discriminator value for each jet. The minimum thresholds on these discriminators define loose ("L"), medium ("M") and tight ("T") operating points with a misidentification probability for light-parton jets of close to 10%, 1% and 0.1%, respectively, at an average jet P_T of about 80 GeV/c. In this work the CSV b-tagging algorithm has been used and particularly we have chosen the tight operating point, in order to minimize the misidentification probability.

Performance of the b-tagging CSV algorithm Jets with $P_T \geq 60$ GeV/c in a sample of simulated multijet events are used to obtain the efficiencies and misidentification probabilities. For loose selections with 10% misidentification probability for light-parton jets, a b-jet tagging efficiency of $\approx$ 80-85% is achieved. For medium and tight selections the CSV algorithm shows the best performance in terms of S/B, where the efficiencies are 1% and 0.1%, respectively.

4.3 Mass reconstruction with kinematic fit

For each event we determine a reconstructed top quark mass, from the four-momenta of the selected jets. The total number of different permutations giving two doublets of jets corresponding to the W bosons and two triplets of jets corresponding to the top quarks is 90. Since we require the presence of b-tags, assigning the tagged jets only to b-quarks reduces this number to 30 for 1-tag events and 6 in case of two or more b-tags.

Once the preselection of data has been done, the `countEvents.C` ROOT macro reconstructs the top quark mass by using a kinematic fit carried out on the preselected events. Here are the steps that lead to the mass reconstruction:

1. The ROOT macro receives a parameter that indicates the samples that need to be processed (data, background or MC).

2. `countEvents.C` operates the pre-selection and if the considered event has not the requested features, continues by examining the following event

3. After the preselection the kinematic fit starts. The four-momenta $P^\mu = (E, \mathbf{p})$ of the six jets are reconstructed and, going backwards, the W bosons and b-jets four-momenta and masses are then reconstructed. Finally the four-momenta of the couple $t\bar{t}$ is obtained.

$$P^\mu_{W^+} = P^\mu_{j_1} + P^\mu_{j_2}$$

$$P^\mu_{W^-} = P^\mu_{j_3} + P^\mu_{j_4}$$

$$P^\mu_t = P^\mu_{W^+} + P^\mu_b$$

$$P^\mu_{\bar{t}} = P^\mu_{W^-} + P^\mu_{\bar{b}}$$

The top quark mass is then calculated by using a χ^2 function defined as:

$$\chi^2 = \frac{(m^{(1)}_{jj} + m_W)^2}{\sigma^2_W} + \frac{(m^{(2)}_{jj} + m_W)^2}{\sigma^2_W} + \frac{(m^{(1)}_{jjb} + m_t)^2}{\sigma^2_t} + \frac{(m^{(2)}_{jjb} + m_t)^2}{\sigma^2_t}$$

where $m^{(1,2)}_{jj}$ are the masses of the two double-jet systems coming from the W bosons, $m^{(1,2)}_{jjb}$ are the masses of the two three-jet systems coming from the two top quarks; σ_W and σ_t the uncertainties associated to the distributions respectively of W and t in MC $t\bar{t}$ events ($\sigma_W \approx$

Cuts	S	N	B=N-S	S/B
$N_{jets} \geq 6 + N_{tags} \geq 2$	64426	485867	421441	$\approx 1/6.5$
$N_{jets} \geq 6 + N_{tags} \geq 2 + \chi^2 \leq 5$	26838	80129	53290	$\approx 1/2$

Table 4.1: Improvements of the signal-background ratio after different cuts on the number of jets (N_{jets}) and tight tags (N_{tags}). The $t\bar{t}$ signal refers to the default MC sample.

15%m_{jj} and $\sigma_t \approx$11%m_{jjb}). The reader should know that the χ^2 function expresses the deviation of the observed data from the fit, weighted inversely by the uncertainties in the individual points. Therefore, the χ^2 function can be either used to test how well a particular model describes the data or, if the prediction is a function of some parameters, then the optimal values of the parameters can be found by minimizing χ^2. In the previous equation all the values are known, except for the m_t, which plays the role of a free parameter. By minimizing the χ^2-function thanks to the MINUIT algorithm, the best value for the top quark mass is then found.

Of course, it is impossibile to know a priori what the correct combination of jjb and j'j'b' is and therefore which light jets and b-jets come from the top quark and which from the antitop quark. This is the reason why it is necessary to repeat the minimization process for *every single permutation of the jets*. At this point, for each event, the two parameters (m_t and χ^2) of the permutation corresponding to the minimum (i.e. best) value of χ^2 are saved.

4. The macro `countEvents.C` produces an output ROOT file (`MassPlot.root`) containing plots of the m_t and χ^2 for the best permutation of every event. Moreover, there is also the plot of the "Cut-Masses", which are the "best" masses (in the sense that we have just explained) of each event that satisfies $\chi^2 \leq \chi^2_{cut}$, where the value chosen for χ^2_{cut} is 5.

By following this procedure we have significantly improved the S/B ratio, reaching $S/B \approx 1/2$. In Tab. 4.1 there is the evolution of the S/B ratio values for the MC default $t\bar{t}$ sample.

4.4 Experimental checks and calibrations

In this section the background issue will be analyzed as well as the MC P_T reweighting.

4.4.1 The Background

The data samples can be modeled as the sum of two different contribution: the signal (which is provided by the Monte Carlo simulations) and the background. Therefore, it is extremely important to evaluate the background mass distribution. In order to obtain such plots, it is necessary to create a sample of background-only events. This goal is achieved with the *event mixing* technique, which consists in mixing events that are characterized by: six or more jets, two b-tags and satisfying the kinematic cuts. In this way a new event is obtained from the completely arbitrary sum of two events. This mixed event has nothing to do with $t\bar{t}$ events and therefore behaves as pure background.

This background modelization could appear arbitrary and not fully satisfying, therefore a background validation will be provided. The purpose of this background validation is to show that there is no relevant discrepancy between the data sample and the background one when we consider events that are unlikely to contain signal, which means, in terms of the χ^2-function, events that satisfy:

$$\chi^2 \geq 10.$$

As shown in Fig. 4.1, the two plots are compatible, when normalized to the same area.

4.4.2 Monte Carlo reweighting

As previously stated, Monte Carlo samples are *simulations* and although the MC generators are extremely sofisticated, there may be slight imperfections in the models they generate. In this case, in order to better reproduce the actual data samples, it has been necessary to proceed with some reweighting operations in the jets P_T. These correction occur at three different stages:

1. P_T **reweighting**: the Monte Carlo top quarks are produced with a P_T that results slightly underestimated with respect to the real one;

2. **Pileup reweighting**: it is necessary to operate a calibration so as to take into consideration multiple interactions. In other words, it is possible that the p-p collisions do not involve only one proton per bunch

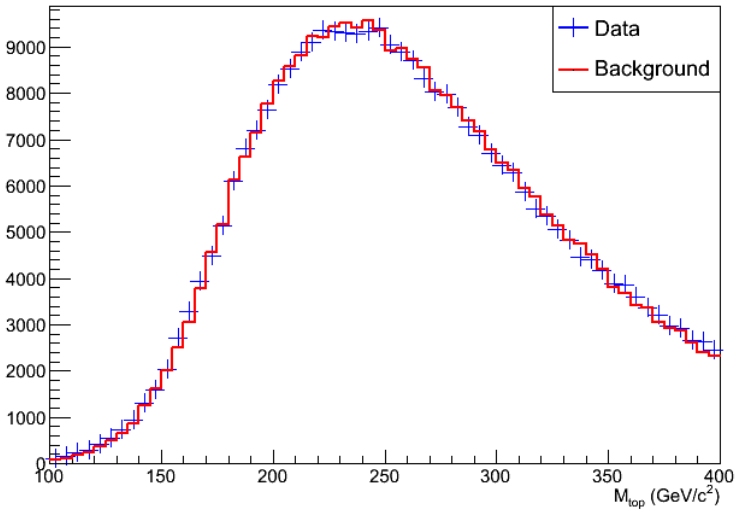


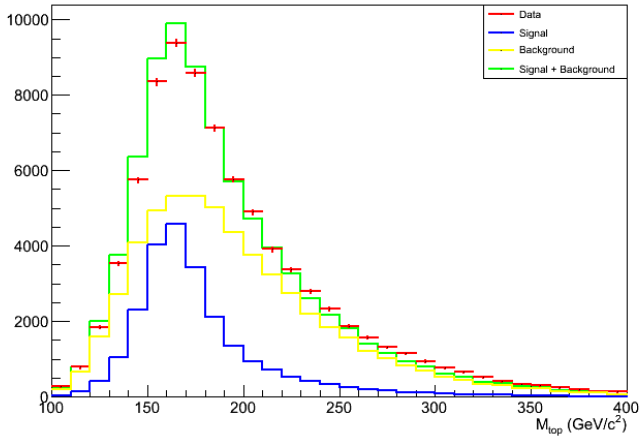
Figure 4.1: Reconstructed top quark mass for events with χ^2 $\geq$10. The blue markers refer to the data, while the red histogram is relative to the background from mixed events. The plots are normalized to the same area.

and the frequency of these events depends on the instantaneous luminosity. Not always do the Monte Carlo simulations consider these cases, therefore it is necessary to introduce a "Pileup weight" to calibrate the sample.

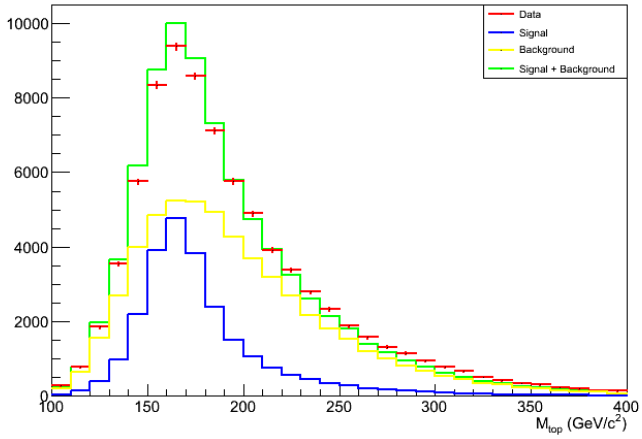
3. **Trigger reweighting**: finally, it is necessary to introduce a "trigger weight" so as to model the trigger efficiency.

Fig. 4.2 shows the plots we obtain with different MC samples. Data are drawn in red, while the expected plot (signal + background) is in green. Notice how the green peak slightly tends to move towards higher mass values, while going from the "171.5 GeV/c^2 MC" to the "175.5 GeV/c^2" one.

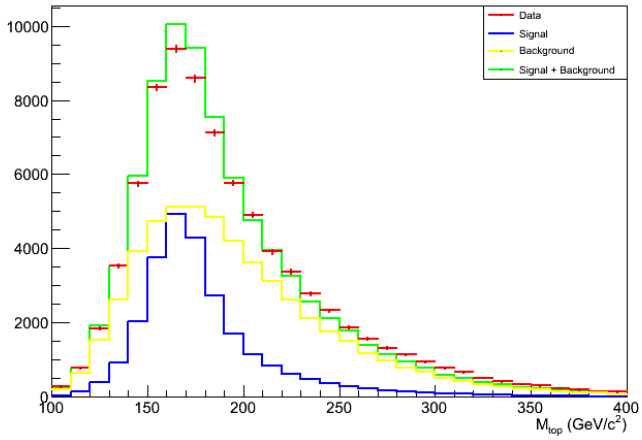
In Fig. 4.3 we have plotted the Monte Carlo top quark masses, that are the result of a Gaussian fit of the reconstructed-masses plots, vs the "true" masses (thus the ones that were assumed as top quark mass when generating the different MC samples). The Gaussian fit is not appropriate for the reconstructed-mass plots, therefore the value of the slope of the linear regression we get is not significant. What is significant, instead, is that the points on the plot lie on a straight line.



(a) Signal generated with $m_t = 171.5$ GeV/c^2



(b) Signal generated with $m_t = 173.5$ GeV/c^2



(c) Signal generated with $m_t = 175.5$ GeV/c^2

Figure 4.2: Reconstructed top quark mass for data, background and $t\bar{t}$ signal for different values of the input top quark masses.

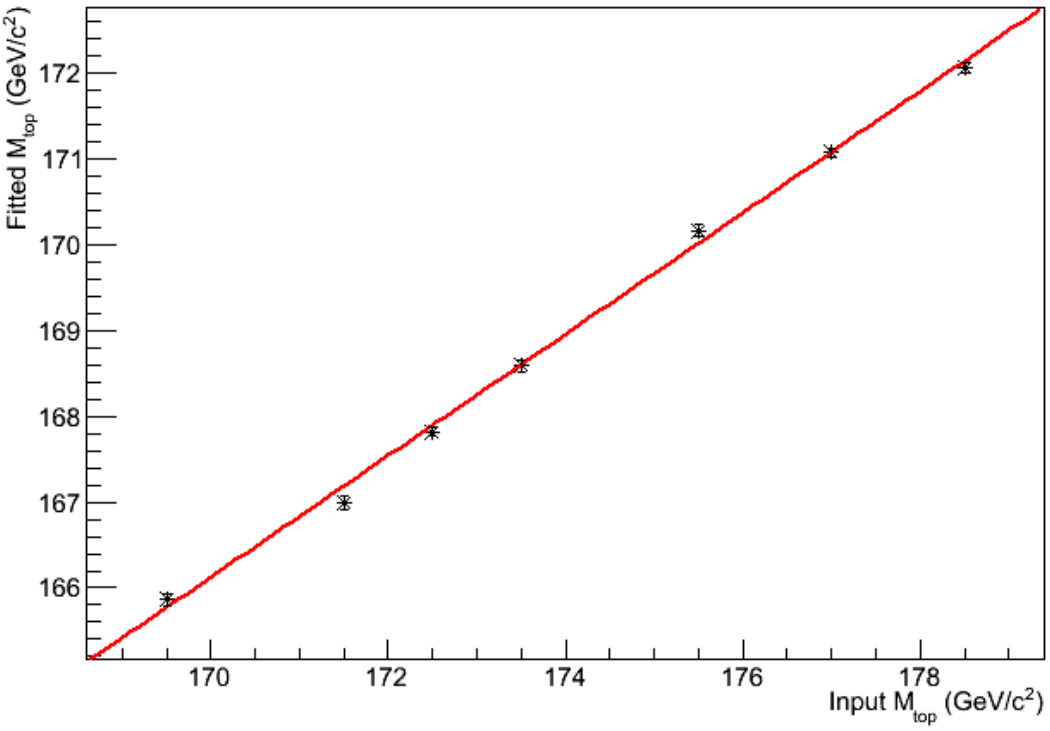


Figure 4.3: Mean of the reconstructed top quark mass versus the input mass for $t\bar{t}$ events. The mean is obtained assuming a Gaussian distribution. The red line shows the fitted linear dependence

4.5 Likelihood fit

In this section the procedure that has lead to the final measurement is described.

4.5.1 Pseudo-Data generation

After selecting and calibrating the data, pseudo-experiments are performed assuming specific values for the top quark mass and these pseudo-data are therefore extracted from the corresponding signal (provided with Monte Carlo simulations, generated with different top quark masses) and background templates. The results of these pseudo-experiments have been used to obtain, on average, a more reliable estimate of the true values and uncertainties. In our case the pseudo-experiments have been created with the ROOT macro `NewPseudoData.C`, which works as follows:

1. `NewPseudoData.C` receives as an input variable a flag that indicates which of the `MassPlotMC.root` files should be opened. After the required Monte Carlo file, the macro opens also the data file and the background one (respectively `MassPlotDATA.root` and `MassPlotBackground.root`).

2. The reconstructed-mass histogram is opened for every input file. We will refer to these histograms as *hData*, *hMC* and *hBkg*.

3. We have decided to create 100 pseudo-experiments for each Monte Carlo sample. For every pseudo-experiment, for every bin, the value of occurrence of the bin mass value is taken as mean value and its square root as the value of the standard deviation of a Gauss distribution. Then a random number from the standard Normal (Gaussian) distribution with the given mean and standard deviation is generated and it constitutes the value of occurrence of that certain bin in the considered pseudo-experiment. These actions are repeated both for the signal and the background, therefore at the end of the whole procedure there will be 100 pseudo-signal and 100 pseudo-background plots.

   Before creating the following pseudo-experiment (which means pseudo-signal and pseudo-background), each "pseudo-histogram" is normalized in order to obtain signal and background in the right proportion. The sum of the entries of the pseudo-signal and the pseudo-background histogram needs to equal the number of entries of the *hData* histogram. We then define the two parameters N_s and N_b:

$$N_s = \frac{S \times L \times \sigma_{\mathrm{t\bar{t}}}}{N^{MC}_{generated}} \qquad N_b = N - N_s$$

where N, S, B are the number of entries of the *hData*, *hMC* and *hBkg* histograms respectively, L the integrated luminosity of the data, $\sigma_{t\bar{t}}$ the $t\bar{t}$ cross section and, again, $N^{MC}_{generated}$ is the number of generated events for each MC sample.

4.5.2 Likelihood fit

After creating the pseudo-experiments, it is possible to actually measure the top quark mass. A likelihood-based method is adopted. The reason why we cannot use a χ^2-function, but we need a likelihood fit, is that the variables in this case are not purely Gaussian, because the data points come from Poisson-distributed numbers of events which are not well approximated by Gaussian distributions. Using a standard χ^2 approach in this case leads to biased estimates of both the mass and its uncertainties. A likelihood function simply expresses how likely the observed distribution is, given some model.

The ROOT macro that creates the likelihood function and then the fit is `makeLikelihood.C`. Its input files are the pseudo-experiments and the n events in the data file `MassPlotDATA.root`, therefore it can use the probability density function of signal (P_s) and background (P_b), with n_s and n_b entries respectively. The likelihood function is composed by two factors: the normalization factor (Poissonian):

$$\mathcal{L}_{norm} = e^{-\mu}\frac{\mu^n}{n!} \qquad \mu = number\,of\,expected\,events = n_s + n_b$$

and the second factor, tightly bound to the shape of the distributions:

$$\mathcal{L}_{shape} = \prod_{i=1}^{n}\Big(\frac{n_s \times P_s(i) \;+\; n_b \times P_b(i)}{n_s + n_b}\Big)$$

therefore:

$$\mathcal{L} = \mathcal{L}_{norm} \times \mathcal{L}_{shape} = e^{-\mu}\frac{\mu^n}{n!} \times \prod_{i=1}^{n}\Big(\frac{n_s \times P_s(i) \;+\; n_b \times P_b(i)}{n_s + n_b}\Big)$$

Looking at the last equation it is clear that the likelihood fit is characterized by two parameters: n_s and n_b. For computational reasons, finding the local maxima is rather tougher than finding the minima, therefore the MINUIT algorithm is applied to:

$$-\ln(\mathcal{L}) = \mu - n\ln\mu + ln(n!) - \sum_{i=1}^{n}\Big(\frac{n_s \times s(i) \;+\; n_b \times b(i)}{n_s + n_b}\Big)$$

This quantity is calculated for every Monte Carlo mass value and the uncertainties are given by the fluctuation provided by the pseudo-experiments.

As a final step, the obtained plot is fitted with a third degree polynomial, as shown in Fig. 4.4. Its minimum provides the final top quark mass value:

$$M_{top} = (173.95 \pm 0.43(stat)) \text{ GeV}/c^2$$

where the statistical uncertainty is evaluated from the mass value which corresponds to a variation of 0.5 units in likelihood.

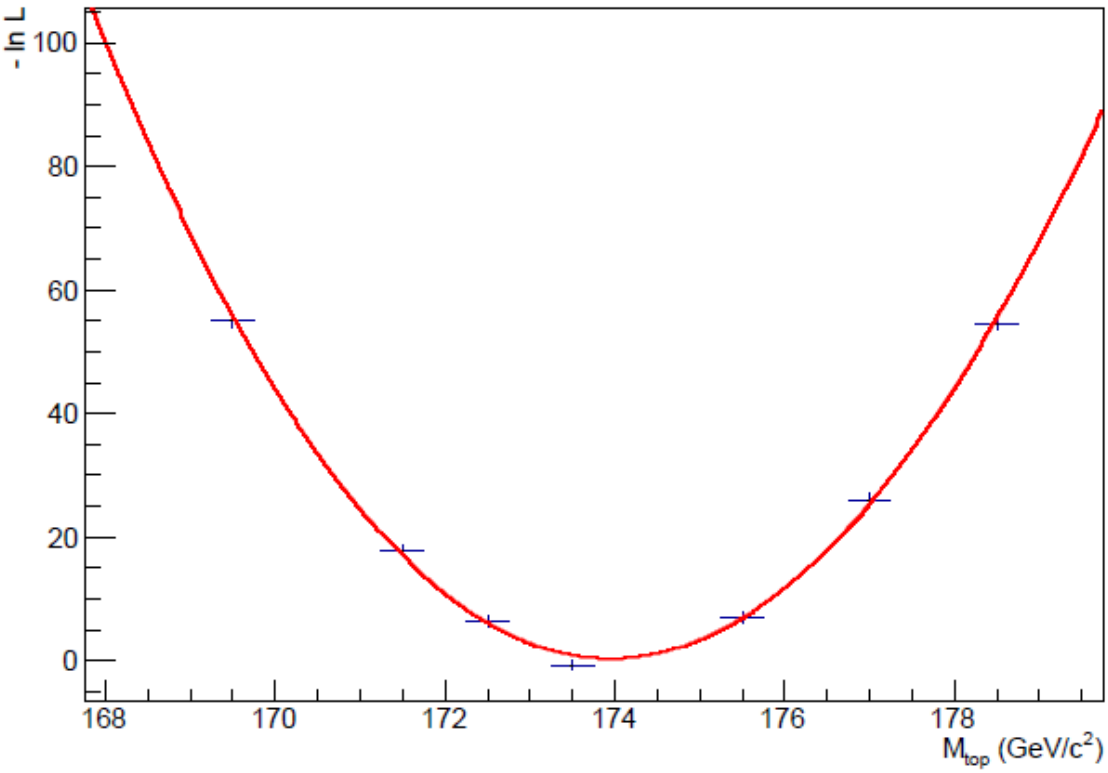


Figure 4.4: The Likelihood fit of $-\ln \mathcal{L}$ shows a minimum at a mass of 173.95 GeV/c^2

Chapter 5

Results

In this chapter the experimental results and the systematic uncertainties will be analyzed.

5.1 Measurement of the top quark mass

The result of the top quark mass measurement performed here is:

$$\boxed{M_{top} = (173.95 \pm 0.43(stat))\ \mathrm{GeV}/c^2}$$

in accordance with the recently published world average [17]:

$$M_{top} = (173.34 \pm 0.27(stat) \pm 0.71(syst))\ \mathrm{GeV}/c^2$$

This compatibility validates the methods that have been used to analyze the data and in particular confirms that the data selection and the Monte Carlo reweighting have been conducted properly.

5.2 Systematic uncertainties

Dealing with systematic uncertainties in an experiment such as CMS is extremely complex, since the sources of systematic uncertainties include those from biases in detector performance, precision of background estimates, accuracy of the signal reconstruction, the uncertainty on integrated luminosity and the uncertainty on the overall jet energy calibration (JEC_UncertaintyTot). This last term is one of the dominant sources of uncertainty on M_{top} and its value ranges from ≈1% to ≈3% of the jet energies. In order to show how systematic uncertainties are handled, at least in one case, we have evaluated the amount of the corresponding systematic uncertainty with the following procedure:

1. In `countEvents.C` the jets' P_x, P_y and P_z and energy E have been calibrated:
$$P_x^{\pm} = P_x(1 \pm JEC\_UncertaintyTot)$$
$$P_y^{\pm} = P_y(1 \pm JEC\_UncertaintyTot)$$
$$P_z^{\pm} = P_z(1 \pm JEC\_UncertaintyTot)$$
$$E = (P_x^{\pm})^2 + (P_y^{\pm})^2 + (P_z^{\pm})^2$$
(Notice that in the ROOT macro natural units are used.)

2. `countEvents.C` produces the output files `MassPlot`$\pm$`.root`.

3. The new "$\pm$"-pseudo-experiments are produced.

4. Finally the likelihood fit is applied to both the "plus" and "minus" sets of pseudo-experiments.

The final top quark mass "$\pm$"-values are:

$$M_{top}^{-} = (176.76 \pm 0.46(stat))\ \mathrm{GeV}/c^2 \qquad M_{top}^{+} = (171.62 \pm 0.41(stat))\ \mathrm{GeV}/c^2$$

Notice that: $M_{top}^{+} \leq M_{top} \leq M_{top}^{-}$. Although this may appear strange, it is instead perfectly consistent: when the JEC_UncertaintyTot is added to the x, y and z components of the impulse P, the reconstructed-mass plots "move" towards higher values of the top quark mass. This implies that the Monte Carlo simulation that fits best the data plot is a "lower mass"-MC. This exact same thing (but in the opposite direction) happens to the "minus set" of pseudo-experiments. Therefore the systematic relative uncertainty is:

$$\frac{M^{-} - M^{+}}{M^{-} + M^{+}} = 0.0148$$

Chapter 6

Conclusions

By using data that have been collected between November 2012 and January 2013 by the CMS detector at LHC, we have developed a procedure that has led to a measurement of the top quark mass. As a first step, candidate events from the data samples have been selected in order to improve the signal-background ratio and therefore separate $t\bar{t}$ (signal) events from other QCD (background) events. This step has been optimized by recurring to MC $t\bar{t}$ simulated events, chosen in the *all-hadronic* channel. Then data have been preselected by a multijet trigger and retained only if they happen to have at least six "good" jets, two of which identified as being b-jets. The triplets masses have been afterwards reconstructed through a constrained fit based on the minimization of a χ^2-like function. After selecting the data, pseudo-experiments are generated assuming certain specific values for the top quark mass, by adding background events to the signal MC ones in the right proportion. At this point, the likelihood fit is applied to the pseudo-data and the final top quark mass measurement has been obtained, yielding a value in agreement with the current world average. Finally, the effect of one the most significant sources of systematic uncertainty has been evaluated.

List of Figures

Bibliography

[1] CDF Collaboration, *Measurement of the top quark mass and* $p\bar{p} \to t\bar{t}$ *cross section in the all-hadronic mode with the CDF II detector*, "Phys. Rev. Lett. D" 81 (2010) 052010.

[2] CDF Collaboration, *Measurement of the top quark mass in the all-hadronic mode at CDF*, "Phys. Lett. B" 714 (2012) 24.

[3] CMS Collaboration, *Measurement of the* $t\bar{t}$ *production cross section in the dilepton channel in pp collisions at* $\sqrt{s} = 7\ TeV$, JHEP (2012) 067.

[4] CMS Collaboration, *Measurement of the* $t\bar{t}$ *production cross section in the all-jet final state in pp collisions at* $\sqrt{s} = 7\ TeV$, JHEP 05 (2013) 065.

[5] CMS Collaboration, *Measurement of the top-quark mass in* $t\bar{t}$ *events with dilepton final state in pp collisions at* $\sqrt{s} = 7\ TeV$, EPJC 72 (2012) 2202.

[6] *LHC the guide*, `http://cds.cern.ch/record/1092437/files/CERN-Brochure-2008-001-Eng.pdf`.

[7] `http://home.web.cern.ch/topics/large-hadron-collider`.

[8] L. Evans, P. Bryant, *LHC Machine*, JINST 3 (2008) S08001.

[9] `http://cms.web.cern.ch/`.

[10] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST 3 (2008) S08004.

[11] CDF Collaboration, *Observation of top quark production in* $p\bar{p}$ *collisions with the Collider Detector at Fermilab*, "Phys. Rev. Lett. 74, (1995) 2626".

[12] DØ Collaboration, *Observation of the top quark*, Pysical Review Letters, volume 74, number 14, 2632, 1995.

[13] A. Barbaro Galtieri, F. Margaroli, I.Volobouev, *Precision measurements of the top quark mass from the Tevatron in the pre-LHC era*, Rept. Prog. Phys. 75 (2012) 056201.

[14] B. Carithers, P. Grannis, *Discovery of the top quark*, SLAC Beam Line 25 (1995) 4.

[15] `http://root.cern.ch/root/html/`.

[16] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, JINST 8 (2013) P04013.

[17] ATLAS, CDF, CMS and DØ Collaborations, *First combination of Tevatron and LHC measurements of the top-quark mass*, arXiv:1403.4427 (2014).