Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Corso di Laurea Magistrale in Fisica

Measurement of $t\bar{t}$ production cross section at 13 TeV in the fully-hadronic channel with CMS

Relatore:
Prof. Andrea Castro

Presentata da:
Nico Zanetti

Sessione III
Anno Accademico 2014/2015

«Our lives and our choices, like quantum trajectories, are understood moment to moment. At each point of intersection, each encounter suggests a new potential direction.»

David Mitchell, *Cloud Atlas*

Abstract

Lo scopo di questa tesi è la misura di sezione d'urto di produzione di coppie *top-antitop* nel canale adronico. Per la misura sono stati utilizzati i dati raccolti dall'esperimento CMS in collisioni protone-protone ad LHC, con un'energia nel centro di massa pari a 13 TeV. Il campione di dati utilizzato corrisponde ad una luminosità integrata di 2.474 fb^{-1}. L'analisi dati inizia selezionando gli eventi che soddisfano determinate condizioni (e.g. trigger, tagli cinematici, sei o più jet, almeno 2 jet provenienti dall'adronizzazione di due quark *bottom*) con lo scopo di incrementare la purezza del segnale scartando il più possibile gli eventi di fondo. A seguire, viene ricostruita la massa del quark top usando un fit cinematico. Sulle distribuzioni di tale massa si basa la stima degli eventi di fondo e di segnale. Infine, attraverso un fit di verosimiglianza, si ottiene il valore della sezione d'urto: $\sigma_{t\bar{t}} = 893 \pm 57\ (stat)\ \pm 104\ (syst)\ pb$. Questo risultato è in buon accordo con il valore teorico di 832 pb e con altre misure di CMS effettuate in canali differenti.

Abstract

In this work the $t\bar{t}$ production cross section ($\sigma_{t\bar{t}}$) in the fully-hadronic channel is measured using the data collected by the CMS experiment in proton-proton collisions at the LHC, at a centre-of-mass energy of 13 TeV. The dataset which was used corresponds to an integrated luminosiy of 2.474 fb^{-1}. The analysis starts selecting events which satisfy certain conditions (e.g. trigger, kinematic cuts, six or more jets, two jets coming from the hadronization of bottom quarks) with the aim of rejecting the background events as much as possible thus increasing the signal purity. Then, the top quark mass is reconstructed through a kinematic fit. These mass distributions are used as a way to estimate the background and signal yields. Finally, a likelihood fit is performed in order to measure the cross section value. The obtained result is: $\sigma_{t\bar{t}} = 893 \pm 57\ (stat)\ \pm 104\ (syst)\ pb$. This result is in good agreement with the theoretical value of 832 pb and with other CMS measurements in different channels.

Contents

Chapter 1

Introduction

On April 5, 2015, the most powerful particle accelerator of the world, the *Large Hadron Collider* (LHC), restarted after a two-year break during which it was extensively upgraded to run at operating energy of 13 TeV (6.5 TeV per beam). On June 3, 2015 the LHC started delivering data good for physics. In the following months it was used for proton-proton collisions while in November the machine switched to collisions of lead ions. In December the usual winter shutdown started.

The study of top quark physics and the measurement of important quantities like the production cross section or the top quark mass, represent a relevant fraction of the campaign of measurements performed at CMS (*Compact Muon Solenoid*), one of the four LHC detectors.

In the following, I will describe the analysis I have performed for the measurement of the $t\bar{t}$ production cross section ($\sigma_{t\bar{t}}$) at $\sqrt{s} = 13\,TeV$, in the challenging fully-hadronic multijet topology.

Chapter 2

The Top Quark

2.1 Introduction to the Standard Model

The *Standard Model of particle physics* is a theory which classifies all the known subatomic particles and describes the electromagnetic, weak, and strong nuclear interactions between those particles. It was developed throughout the latter half of the 20th century, as a collaborative effort of scientists around the world. The current formulation was finalized in the mid-1970s upon experimental confirmation of the existence of quarks. Since then, discoveries of the top quark (1995), the tau neutrino (2000), and more recently the Higgs boson (2012), have given further credence to the Standard Model, although many issues are still unsolved, as explained later in section 3.5.

The Standard Model particles can be divided into two classes: fermions and bosons.

2.1.1 Fermions

Fermions are spin-1/2 particles so, according to the spin-statistics theorem, they respect the Pauli exclusion principle. In the Standard Model there are 12 fermions and each of them has a corresponding antiparticle. Fermions are classified according to how they interact: there are six **quarks** (up, down, charme, strange, top, bottom), and six **leptons** (electron, muon, tau lepton and related neutrinos). Pairs from each classification are grouped together to form a generation (see Fig. 2.1)

The defining property of the quarks is that they carry colour charge, and hence, interact via the strong interaction. Because of a phenomenon called colour confinement quarks form *colour-neutral* composite particles called *hadrons*. These can be composed either by a quark and an antiquark (mesons) or by three quarks (baryons). Quarks also carry electric charge and weak isospin. Hence they interact with other fermions both electromagnetically and via the weak interaction.

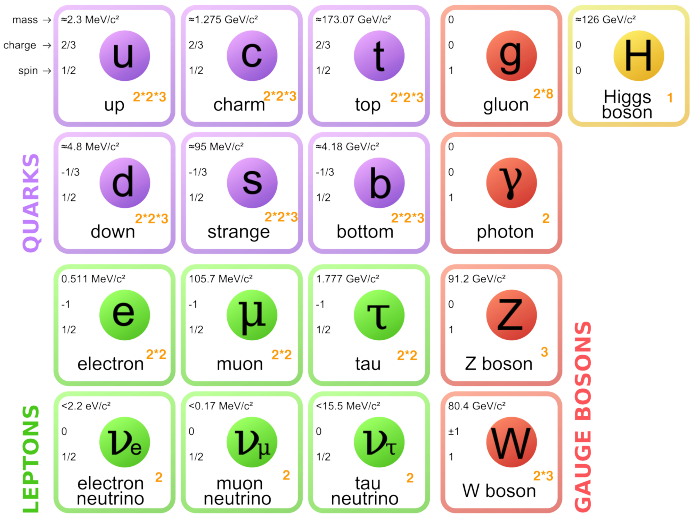

Figure 2.1: Standard Model of elementary particles.

Leptons, on the other hand, do not carry colour charge. The three neutrinos do not carry electric charge either, so their motion is directly influenced only by the weak nuclear force, which makes them notoriously difficult to detect. The electron, muon, and tau lepton, by virtue of carrying an electric charge, interact also electromagnetically.

Each member of a generation has greater mass than the corresponding particles of lower generations. The charged particles of the first generation do not decay; hence all ordinary (baryonic) matter is made of such particles. Specifically, all atoms consist of electrons orbiting around atomic nuclei, ultimately constituted of up and down quarks. Charged particles of the second and third generation, on the other hand, decay with very short half lives, and are observed only in very high-energy environments. Neutrinos of all generations do not decay, and pervade the Universe, but rarely interact with baryonic matter.

2.1.2 Bosons

In the Standard Model, **gauge bosons** are defined as force carriers that mediate the strong, weak, and electromagnetic fundamental interactions. They all have spin 1 and, as a result, they do not follow the Pauli exclusion principle that constrains fermions: thus bosons do not have a theoretical limit on their spatial density. The different types of gauge bosons are described below:

- Photons mediate the electromagnetic force between electrically charged particles. The photon is massless and is well-described by the theory of quantum electrodynamics;

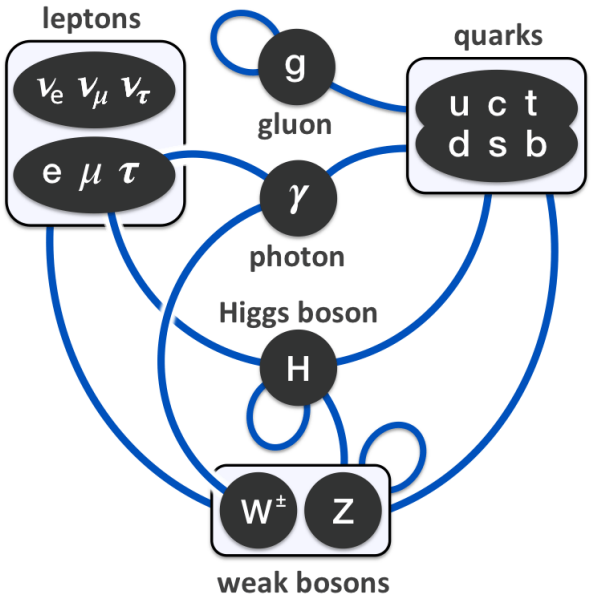

Figure 2.2: Interactions between the Standard Model particles.

- The W^+, W^- and Z gauge bosons mediate the weak interactions between particles of different flavours (all quarks and leptons). They are massive, with the Z being more massive than the $W^\pm$. The latter carries an electric charge of $+1$ or -1 and then can interact electromagnetically. These three gauge bosons along with the photons are grouped together, as collectively mediating the *electroweak* interactions;

- Eight gluons mediate the strong interactions between the quarks. Gluons are massless and have an effective colour charge so they can also interact among themselves. The gluons and their interactions are described by the theory of quantum chromodynamics.

There is another boson in the Standard Model: the **Higgs boson**. It is a massive scalar (spin-0) particle and it is associated with the *Higgs field* which, according to the theory, permeates the universe giving mass to elementary particles (except the photon and the gluon). As the Higgs boson is massive, it must interact with itself.

The interactions between all the particles described by the Standard Model are summarized in Fig. 2.2.

2.2 The Top quark

According to the Standard Model, the *top* quark (t) belongs to the third generation of quarks, like its electroweak partner, the *bottom* quark (b). It has an electric charge of $+2/3$ e and has also one of the three colour charges provided by the QCD. Its antiparticle, the top antiquark ($\bar{t}$) has the same features of the top quark but opposite internal quantum numers (e.g. the electric charge). Like all quarks it experiences all four fundamental interactions: gravitation, electromagnetism, weak interactions and strong interactions. The fundamental feature of the top quark is its huge mass, which makes it the most massive particle of the Standard Model. Experimentally the mass of the top quark is [1]:

$$M_t = 173.21 \pm 0.51\ (stat) \pm 0.71\ (syst)\ GeV^1 \tag{2.1}$$

Because top quarks are very massive, large amounts of energy are needed to create one. The only way to achieve such high energies is through high energy collisions. These occur naturally in the upper atmosphere of the earth as cosmic rays collide with particles in the air. Otherwise top quarks can be created at particle accelerators with very high energy. This is why the top quark was discovered only in 1995 after about twenty years of searches.

2.2.1 The quark model and the Top quark discovery

In 1964 *Murray Gell-Mann* and *George Zweig* proposed the quark hypothesis to account for the abundance of subatomic particles discovered in accelerator and cosmic ray experiments during the 1950s and early 1960s. Over a hundred new particles (the hadrons), most of them strongly interacting and very short-lived, had been observed. The quark hypothesis suggested that different combinations of three quarks (u, d, s) and their antiparticles could account for all of the hadrons that had been discovered [2].

So far, quarks appear to have no size or internal structure and thus represent the smallest known constituents of matter. To explain the observed spectrum of hadrons quarks had to have electric charges that are fractions of the electron charge. The u quark has charge $+2/3$ e while the d and s quarks have charges $-1/3$ e. The observed hadron spectrum agreed remarkably well with the expected states formed from combinations of three quarks or a quark-antiquark pair. Quarks also seemed to form a counterpart to the other class of elementary particles, the leptons, which then included the electron (e) and muon (μ), both with unit charge, and their companion chargeless neutrinos (ν_e and ν_μ). Also leptons seems to have no discernible size or internal structure.

[1] In this work natural units are used.

But most physicists were initially reluctant to believe that quarks were anything more than convenient abstractions aiding 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 contained point-like internal constituents whose charges and spins were just what the quark model had predicted.

The discovery of the Charm quark In 1974, experiments at the Brookhaven National Laboratory in New York and the Stanford Linear Accelerator Center (SLAC) in California discovered a striking new hadron at the mass of 3.1 GeV. This hadron (called J/Ψ) was found to be a bound state of a new kind of quark, called charm (c) with its antiquark ($\overline{c}$). The c quark has a much greater mass than the first three, and its charge is $+2/3$ e. With two quarks of each possible charge, a symmetry could be established between the quarks and the leptons, and theoretical constraints could be satisfied. But this symmetry was quickly broken by unexpected discoveries.

The discovery of the Bottom quark In 1976, experiments at SLAC turned up a third charged lepton, the tau lepton (τ). A year later at the Fermi National Accelerator Laboratory in Illinois a new hadron, called the upsilon (Υ), was discovered at the huge mass of about 10 GeV; it was soon found to be the bound state of yet another new quark: the bottom quark (b) and its antiparticle ($\overline{b}$). Experiments at DESY in Germany and Cornell in New York showed that the b quark had spin 1/2 and a charge of $-1/3$ e, just like the d and s quarks.

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

Searching for the Top quark Using the ratios of the observed quark masses, some physicists suggested that the top quark might be about three times as heavy as the bottom quark, and thus expected that the top would appear as a heavy new hadron containing a $t\overline{t}$ pair, at a mass around 30 GeV. 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 accelerator, the Super Proton Synchrotron (SPS), in which counter-rotating beams of protons and antiprotons collided with an energy of about 600 GeV, came into operation at CERN in Switzerland. 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 they had continued to elude direct detection, the quarks could be violently scattered in high-energy collisions, producing the so-called *jets*, that are collimated sprays of particles. The accelerator at CERN reached its limits without creating a single top quark, pushing the lower bound on its mass up to 77 GeV.

In 1988, with the advent of the more powerful collider at Fermilab (Tevatron), with a centre-of-mass energy of 1800 GeV, the search for the top quark turned to new avenues. At the large masses now accessible, the $t\bar{t}$ bound state was unlikely to form and isolated top quarks were expected. By 1990, the CDF (*Collider Detector at Fermilab*) experiment had extended the top mass limit to 91 GeV, thus eliminating the possibility for W decay to top.

In 1992, the DØ detector joined CDF as a long Tevatron run began. Further searches would have to rely on the production of separate t and $\bar{t}$ quarks from annihilation of incoming quarks and antiquarks in the proton and antiproton, with subsequent decays into observable particles. Meanwhile CDF had installed a new vertex detector of silicon microstrips near the beams intended to detect short-lived particles that survive long enough to travel few millimeters from the interaction point. This detector was particularly good at sensing the presence of the b quarks, characteristic of top decay.

The discovery of the Top quark After several years spent on the long search, on February 24, 1995, CDF and DØ finally announced the observation of the top quark [3] [4]. CDF reported finding six dilepton events plus 43 single-lepton events. It concluded that the probabilities that background fluctuations could account for these events were only one in a million. DØ observed three dilepton events plus 14 single-lepton events and concluded that the probabilities that these could have been caused by backgrounds were two in a million.

The top quark masses reported by the two experiments were 176 ± 13 GeV for CDF and 199 ± 30 GeV for DØ. The results of the two experiments are shown in figures 2.3 and 2.4.

In 2008 Kobayashi and Maskawa won the Nobel Prize in Physics for the prediction of the existence of top and bottom quarks, which together form the third generation of quarks.

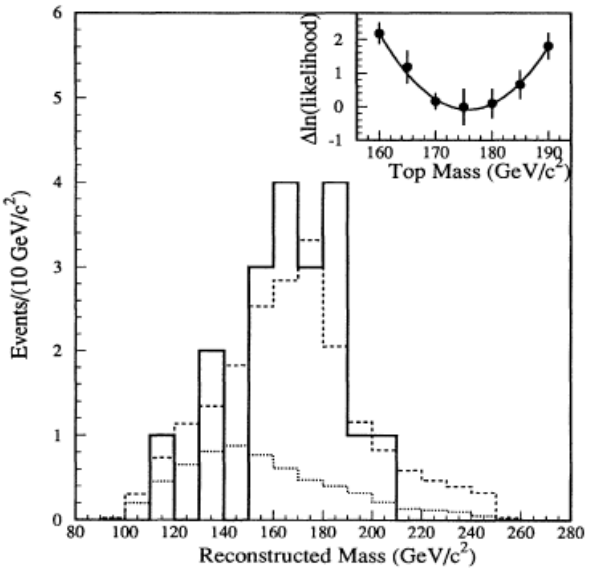

Figure 2.3: Reconstructed mass distribution for top quark candidates events from CDF top quark discovery paper. The dotted line represents the background shape while the dashed one is the sum of background plus $t\bar{t}$ Monte Carlo simulations for $M_t = 175$ GeV. The inset shows the likelihood fit used to determine the top mass.

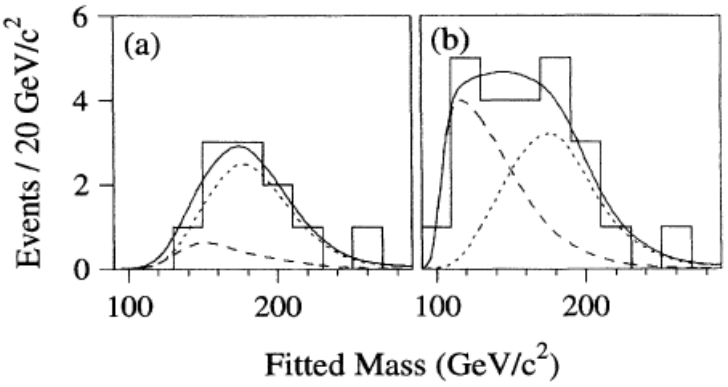

Figure 2.4: Mass distributions from DØ top quark discovery paper. Fitted mass distribution for candidate events (histogram) with the expected mass distribution for $M_t = 199$ GeV top quark events (dotted curve), background (dashed curve), and the sum of $t\bar{t}$ and background (solid curve) for (a) standard and (b) loose events selection.

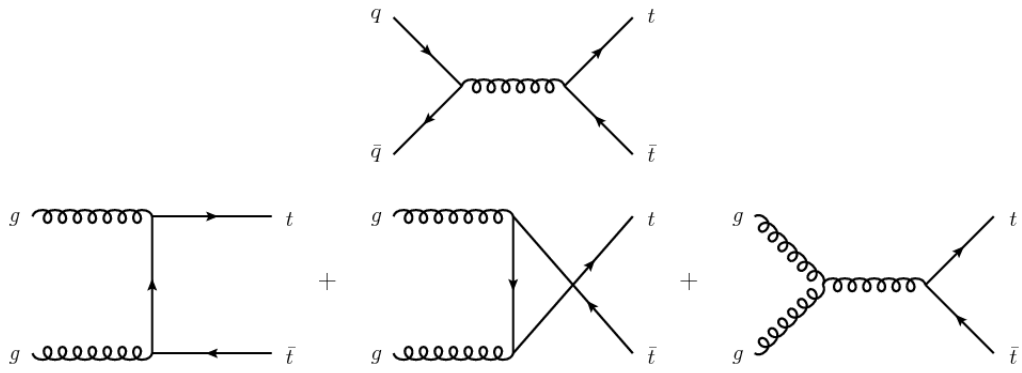

Figure 2.5: Lowest-order diagrams contributing to $t\bar{t}$ pair production at hadron colliders: quark-antiquark annihilation (upper diagram) and gluon-gluon fusion (lower diagrams).

2.2.2 Top quark production

There are multiple processes that can lead to the production of a top quark. The most common is the production of a $t\bar{t}$ pair via strong interaction. There are two main processes (see Fig. 2.5): quark-antiquark annihilation ($q\overline{q} \rightarrow t\bar{t}$) and gluon-gluon fusion ($gg \rightarrow t\bar{t}$).

At a proton-proton collider, like the LHC (see chapter 3), for $\sqrt{s} = 13$ TeV, about 90% of the $t\bar{t}$ production is from gluon-gluon fusion, while the remaining 10% is from quark-antiquark annihilation.

It is also possible to produce $t\bar{t}$ pairs through the decay of an intermediate photon or Z boson. However, these processes are predicted to be much rarer and in addition they have an identical experimental signature in hadron colliders like LHC.

A distinctly different process is the production of single top quarks via weak interaction (see Fig. 2.6). This can happen at a lower rate in three ways:

- a bottom quark transforms into a top quark by exchanging a W boson with an up or down quark (*t-channel*);
- associated production of a top quark and a W boson (tW);
- an intermediate W boson decays into a top and antibottom quark (*s-channel*).

These three channels have decreasing cross sections.

2.2.3 Top quark decays

With its correspondingly short lifetime of $\sim 0.5 \times 10^{-24} s$, the top quark is expected to decay before top-flavoured hadrons or $t\bar{t}$ bound states can

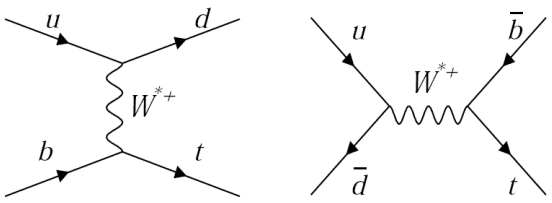

Figure 2.6: Feynman diagrams for electroweak single top quark production: a t-channel diagram (left) and a s-channel diagram (right).

form. Being heavier than a W boson, t is the only quark that, through weak interaction, decays semi-weakly into a real W boson and a down type quark (d, s, b). The coupling of the top quark to the different down type quarks is given by the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix. From existing experimental constraints on the CKM matrix and assuming unitarity and 3 generations of quarks, the CKM matrix element $|V_{tb}|$ is constrained to be very close to unity. Therefore the top quark decays into Wb in nearly 100% of the cases while the CKM-suppressed decays represent $\sim 0.1\%$ for $t \to Ws$ and $\sim 0.01\%$ for $t \to Wd$.

The final states of $t\bar{t}$ event are completely determined by the decay products of the two W bosons from the two top quarks decays, and contains in addition two jets of hadronic particles arising from the hadronization of the two b quarks. Three different final states can be observed:

- *Dilepton channel*: both W boson decay into a lepton-neutrino pair:

$$t\bar{t} \to W^+ b \;\; W^- \bar{b} \to l^+ \nu_l b \;\; l^- \overline{\nu}_l \bar{b}, \tag{2.2}$$

  with a branching ratio (BR) of about 9%. The signature of the dilepton decay is the presence of two charged leptons, missing energy in the transverse plane (associated to the two neutrinos) and two jets from the b quarks. Most of the analyses distinguish the e and μ channel from the τ channel, which is more difficult to reconstruct.

- *Single-lepton channel*: one W boson decays into a lepton-neutrino pair while the other decays hadronically:

$$t\bar{t} \to W^+ b \;\; W^- \bar{b} \to l^+ \nu_l b \;\; q q' \bar{b} \tag{2.3}$$

  or

$$t\bar{t} \to W^+ b \;\; W^- \bar{b} \to q q' b \;\; l^- \overline{\nu}_l \bar{b}, \tag{2.4}$$

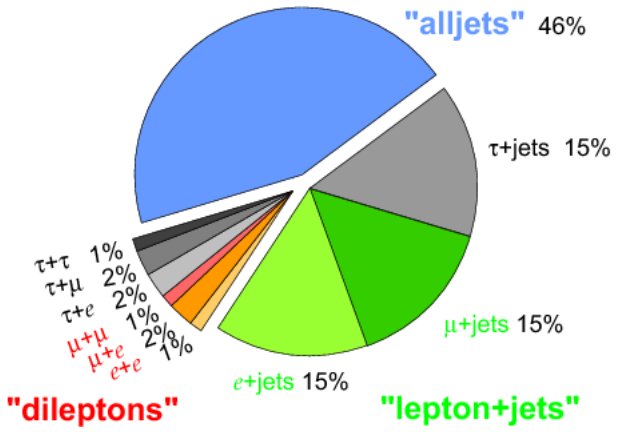

Figure 2.7: Branching ratios (BR) for top-quark decays.

with a branching ratio of 45% if including decays into τ leptons, or 30% if not. In the final state there will be one charged lepton, missing energy in the transverse plane (from the neutrino) and four jets, arising from the four quarks produced in the decay.

- *Fully-hadronic channel*: both W bosons decay hadronically:

$$t\bar{t} \rightarrow W^+b \;\; W^-\bar{b} \rightarrow qq'b \;\; qq'\bar{b} \rightarrow j_1j_2j_3 \;\; j_4j_5j_6 \tag{2.5}$$

with a branching ratio of 46%. In the final state there will be six jets arising from the six quarks produced in the decay, two of which coming from b quarks.

In addition to the quarks resulting from the top-quark decays, extra QCD radiation (quarks and gluons) from the coloured particles in the event can lead to extra jets.

In Fig. 2.7 are summarized all the possible final states and their branching ratios.

Chapter 3

High-Energy Physics at LHC

This chapter describes the LHC, the CMS experiment and the main goals of the LHC project.

3.1 The Large Hadron Collider

The LHC is the largest and most powerful particle accelerator of the world. It first started up on 10 September 2008, and it was the latest addition to the CERN accelerator complex. It is situated inside an underground tunnel 100 meters deep on average located on the border between France and Switzerland (see Fig. 3.1 and Fig. 3.2). The LHC [5] [6] consists of a 27-kilometre ring of superconducting magnets with a certain number of accelerating structures which increase the energy of the particles along the way.

Inside the accelerator, two high-energy particle beams, consisting of protons or heavy ions, travel with velocity close to the speed of light before they are made to collide (within one of four detectors). The beams travel in opposite directions in separate beam pipes, which are two tubes kept at ultrahigh vacuum, and they are guided around the accelerator ring by a strong

Figure 3.1: Aerial view of the LHC in Geneva (Switzerland).

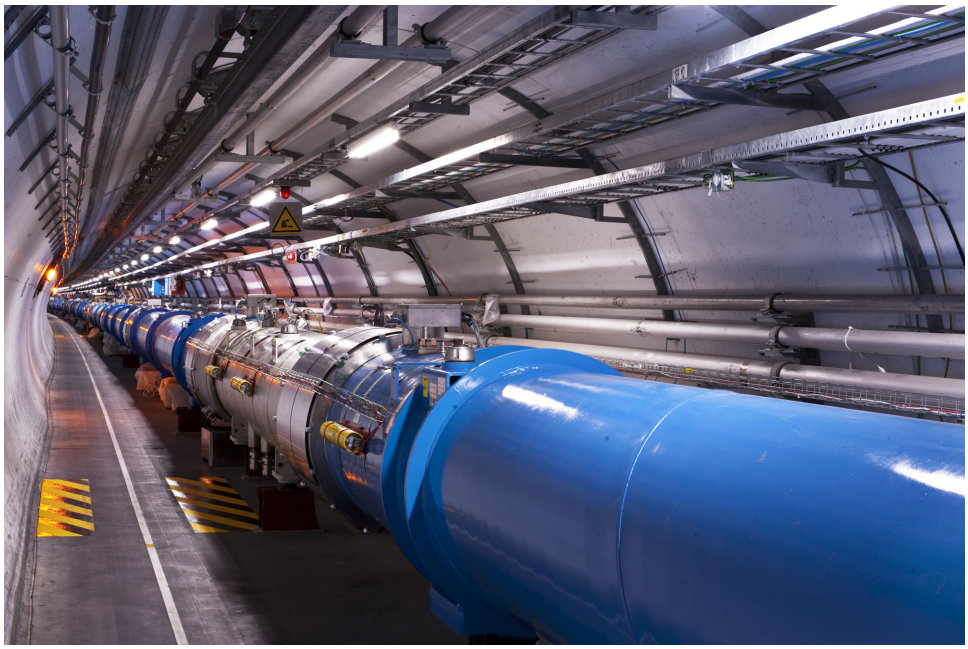

Figure 3.2: The LHC Tunnel.

magnetic field maintained by superconducting electromagnets. These latter 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 cooling the magnets to $-271.3\,^{\circ}\mathrm{C}$, a temperature very close to the absolute zero. For this reason, much of the accelerator is connected to a distribution system of liquid helium, which cools the magnets.

3.1.1 Brief history of the LHC project

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. After many years of work on the technical aspects and physics requirements of such a machine, their dreams came to fruition in December 1994 when CERN 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 Japan, the USA, India and other non-member States, the Council voted in 1995 to allow the project to proceed in a single phase. Between 1996 and 1998, four experiments (ALICE, ATLAS, CMS and LHCb) received official approval and construction work commenced

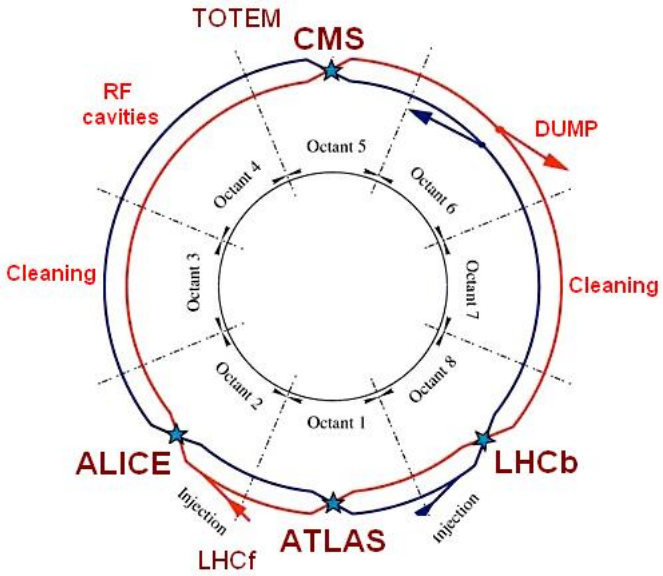

Figure 3.3: Structure of the LHC

on the four sites. In November 2000 the LEP accelerator stopped running and began to be dismantled to give up its place in the 27 kilometres tunnel to the LHC. In June 2003 the ATLAS detector cavern was completed and on the 1st of February 2005 a second detector cavern, the CMS one (53 metres long, 27 wide and 24 high), was inaugurated. Finally, more than three years later, at 10.28 AM on the September 10, 2008, a beam of protons was successfully steered around LHC for the first time.

3.1.2 Characteristics of the LHC

The LHC is not a perfect circle. It is made of eight arcs and eight 'insertions' (see Fig. 3.3).

The arcs contain the dipole 'bending' magnets, with 154 in each arc. An insertion consists of a long straight section plus two (one at each end) transition regions, 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 happened sector by sector, the hardware was commissioned sector by sector and all the dipoles of a sector were connected in series in the same continuous cryostat. Powering of each sector is essentially independent.

In the LHC particles circulate in a vacuum tube. The beam vacuum pressure is 10^{-13} atm (ultrahigh vacuum), because it is necessary to avoid collisions with gas molecules. Particles are manipulated using electromagnetic

devices:

- Thousands of **magnets** of different varieties and sizes are used to direct the beams around the accelerator. These include 1232 dipole magnets which keep the particles in their nearly circular orbits and 392 quadrupole magnets which focus the beams. Insertion quadrupoles are special magnets used to focus the beam down to the smallest possible size at the collision points, thereby maximizing the chance of two protons (or heavy ions) smashing head-on into each other. Others magnets (sextupoles, octupoles, decapoles, etc.) contributes to optimizing the particle trajectory.

  The dipoles of the LHC represented the most important technological challenge for the design of the accelerator. Indeed, the maximum energy that can be achieved, given a specific acceleration circumference, is directly proportional to the strength of the dipole field. No practical solution could have been designed using 'warm' magnets instead of superconducting ones. At the LHC the dipole magnets are then superconducting electromagnets which are able to provide a very high magnetic field (up to about 8 T) over their length. The LHC dipoles use niobium-titanium (NbTi) cables, which become superconducting (they conduct electricity without resistance) below a temperature of 10 K. In fact, the LHC operate at a temperature of 1.9 K, which is even lower than the temperature of outer space (2.7 K).

- **Accelerating cavities**, which are electromagnetic resonators, deliver radiofrequency power to the beam during the acceleration to the top energy. They are also used to keep the proton bunches tightly bunched to ensure high luminosity at the collision points and hence, maximize the number of collisions. Superconducting cavities with small energy losses and large stored energy are the best solution.

3.2 The CERN accelerator complex

The accelerator complex at CERN is a succession of machines with increasingly higher energies (see Fig. 3.4). Each machine injects the beam into the next one, which takes over to bring the beam to an even higher energy, and so on. The LHC is the last element of this chain. At the end of the acceleration process the particle beams reach, currently, the energy of 13 TeV. In addition, most of the other accelerators in the chain have their own experimental halls, where the beams are used for experiments at lower energies.

This is how the proton beams are accelerated through the CERN accelerator complex:

- Hydrogen atoms are taken from a bottle containing hydrogen.

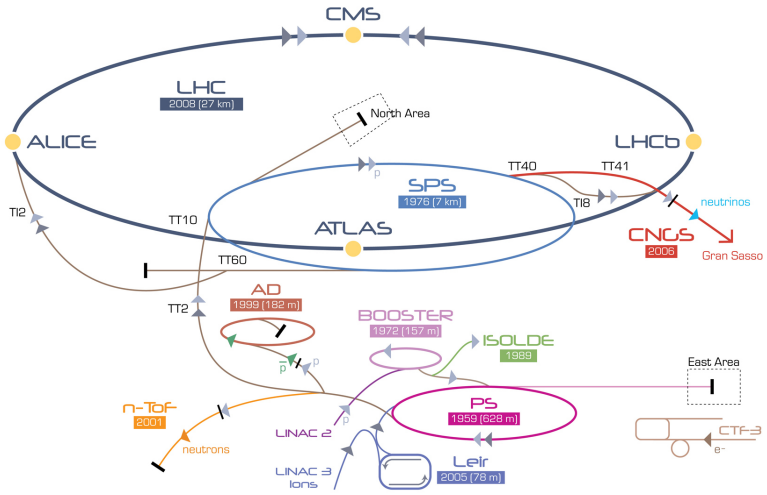

Figure 3.4: The CERN accelerator complex.

- Protons are obtained by stripping the orbiting electrons from hydrogen atoms.
- Protons are 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 are accelerated to their nominal energy of, currently, 6.5 TeV. Beams will circulate for many hours inside the LHC beam pipes under normal operating conditions.

In addition to accelerating protons, the complex can also accelerates lead ions.

3.3 The LHC Detectors

There are four main detectors installed at the LHC: ALICE, ATLAS, CMS and LHCb. They are installed in four huge underground caverns built around the four collision points of the LHC beams.

ALICE (A Large Ion Collider Experiment)

ALICE is a detector specialized in analysing lead-ion collisions. It studies the properties of quark-gluon plasma, a state of matter where quarks and gluons, under conditions of very high temperatures and densities, are no longer confined inside hadrons. Such a state of matter probably existed just after the Big Bang, before particles such as protons and neutrons were formed.

ATLAS (A Toroidal LHC ApparatuS)

ATLAS is a general-purpose detector designed to cover the widest possible range of physics at the LHC, from the search for the Higgs boson to super-symmetry (SUSY) and extra dimensions. The main feature of the ATLAS detector is its enormous doughnut-shaped magnet system. This consists of eight 25-meters long superconducting magnet coils, arranged to form a cylinder around the beam pipe through the centre of the detector. ATLAS is the largest-volume collider-detector ever constructed.

LHCb (Large Hadron Collider beauty)

LHCb specializes in the study of the slight asymmetry between matter and antimatter present in interactions of B-particles (particles containing the b quark). Understanding it should explain the strong matter-antimatter imbalance in the universe. Instead of surrounding the entire collision point with an enclosed detector, the LHCb experiment uses a series of sub-detectors to detect mainly forward particles.

3.4 The CMS detector

CMS (*Compact Muon Solenoid*) is a general purpose detector (see Fig. 3.5) with the same physics goals as ATLAS, but with different technical solutions and design.

The detector is 21 m long, 15 high m and 15 m wide and weighs 12500 tonnes. It is built around a huge superconducting solenoid and it is composed of many sub-detectors consisting of layers of material that exploit the different properties of particles produced in high-energy collisions in the LHC to catch and measure the energy and momentum of each one. Using the data produced by the various sub-detectors it is possible to build up a "picture" of events at the heart of the collision.

3.4.1 Detector structure

The CMS detector [7] is composed of four types of sub-detectors (see Fig. 3.6). Starting from the center we find:

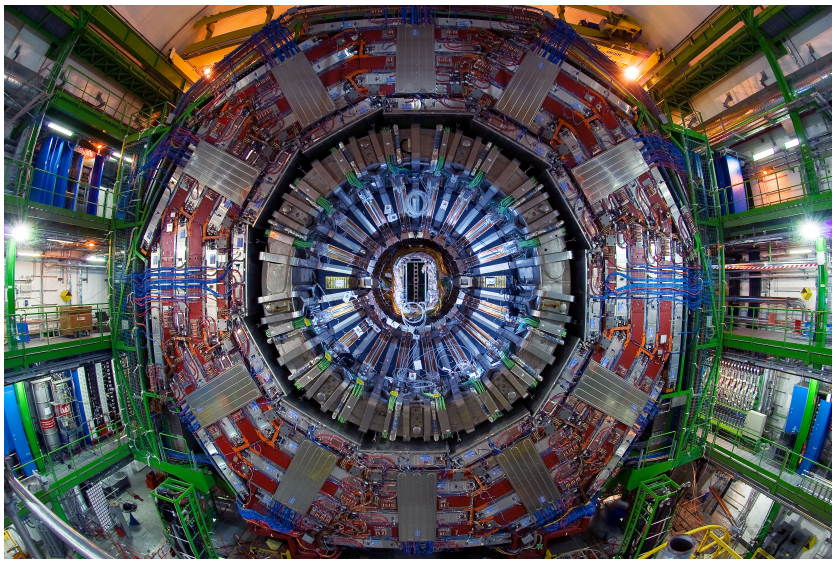

Figure 3.5: The CMS detector.

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

The other essential item is a very strong **magnet** (the Solenoid) that allows us to accurately measure very high-momentum particles such as muons. Indeed, the higher a charged particle momentum, the less its path is curved in the magnetic field, so when we know its path we can measure its momentum both inside the coil (with the tracking devices) and outside of the coil (with the muon chambers).

The Solenoid

The CMS solenoid is a coil of superconducting wire that creates a magnetic field when electricity flows through it. It has an overall length of 13m, a diameter of 7m and a weight of 12000 tonnes. It is a superconducting magnet, allowing electricity to flow without resistance and creating a powerful magnetic field of 4 T (about 100000 times than that of the Earth). For this reason the magnet needs to be cooled to 4.65 K. In fact at ordinary

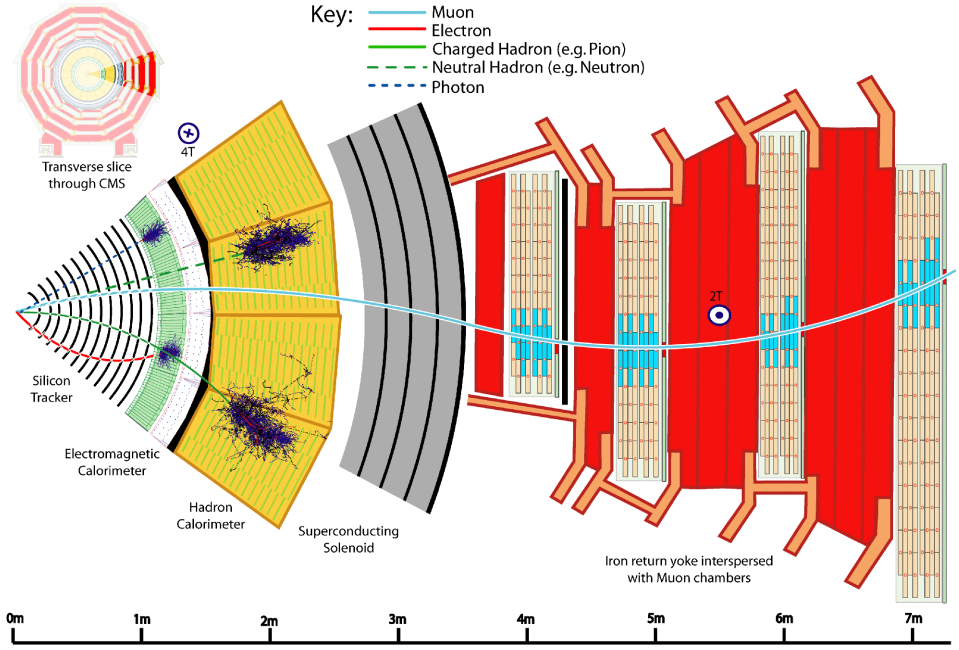

Figure 3.6: Section view of the CMS experiment. Different particles show different behaviours and trajectories within the various sub-detectors

Figure 3.7: The CMS magnet during CMS construction.

temperatures the strongest possible magnet has only half the strength of the CMS solenoid. It is the largest magnet of its type ever constructed and allows the tracker and calorimeter detectors to be placed inside the coil. This is why the detector is overall *compact*, compared to detectors of similar weight.

The Tracker Detector

To calculate the momentum of a particle we need to track its path through the magnetic field generated by the Solenoid. The CMS **tracker** records the paths taken by charged particles by finding their positions at a certain number of key points. The tracker can reconstruct the paths of muons, electrons and hadrons as well as see tracks coming from the decay of very short-lived particles containing b quarks. The tracker needs to record particle paths accurately but must be lightweight so as to disturb the particles as little as possible. It does this by taking position measurements so accurate (up to 10 μm) that tracks can be reliably reconstructed using just a few measurement points. It is also the innermost layer of the detector and so receives the highest flow of particles: the construction materials were therefore carefully chosen to resist radiation. The CMS tracker is made entirely of silicon:

- At the very core of the detector, dealing with the highest intensity of particles, we find the pixel detector: it has about the size of a shoebox and it contains 65 million pixels, allowing it to track the paths of particles emerging from the collision with extreme accuracy. It is very important for reconstructing the tracks of very short-lived particles. Because of the huge number of channels, the power for each pixel must be kept to a minimum. Even with each one generating only around 50 microwatts, the total power output is around the same as the energy produced by a hot plate. So as not to overheat the detector, the pixels are mounted on cooling tubes.

- 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. Due to the nature of their job, the tracker and its electronics are pummeled by radiation but they are designed to withstand it. To minimise disorder in the silicon, this part of the detector is kept at -20 °C, to "freeze" any damage and prevent it from perpetuating.

As particles travel through the tracker, the pixels and microstrips produce tiny electric signals that are amplified and detected, allowing us to reconstruct the particle trajectory in the magnetic field.

The Electromagnetic Calorimeter

In order to build up a "picture" of events occurring in the LHC, CMS must find the energies of emerging particles. Electrons, positrons and photons are measured using an **Electromagnetic Calorimeter** (ECAL). In order to detect them with the necessary precision in the very strict conditions of the LHC (high magnetic field, high levels of radiation and only 25 nanoseconds between collisions), very particular materials are required. CMS use lead tungstate crystals ($PbWO_4$). With a touch of oxygen in this crystalline form it is highly transparent and scintillates when electrons, positrons and photons pass through it. This means it produces light in proportion to the energy of the particles. These high-density crystals produce light in fast, short, well-defined photon bursts that allow for a precise, fast and fairly compact detector. Photodetectors, that have been especially designed to work within the high magnetic field, are also 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 sent for analysis.

The ECAL, made up of a barrel section and two endcaps, forms a layer between the tracker and the hadronic calorimeter. The cylindrical barrel consists of 61200 crystals divided 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 15000 additional crystals.

The Hadron Calorimeter

The **Hadron Calorimeter** (HCAL) measures the energy of hadrons, particles made of quarks and gluons (for example protons, neutrons, pions and kaons). Additionally, it provides indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos. The HCAL must be hermetic, to make sure it captures, to the extent possible, every particle emerging from the collisions. This way if we see particles shoot out one side of the detector, but not the other, with an imbalance in the momentum and energy (measured in the transverse direction relative to the beam line), we can deduce that invisible particles have been produced. The layers of the HCAL were built in a staggered fashion so that there are no gaps in direct lines that a particle might escape through.

The HCAL is a *sampling* calorimeter meaning it finds the position, energy and arrival time of the particles 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 tower), this total amount of light is a measure of the energy of a particle.

The HCAL is organised into barrel, endcap and forward sections. There are 36 barrel wedges, each weighing 26 tonnes. These form the last layer of detector inside the magnet coil, while a few additional layers, the outer barrel, sit outside the coil, ensuring that no energy leaks out the back of the barrel undetected. Similarly, 36 endcap wedges measure particle energies as they emerge through the ends of the solenoid magnet. Lastly, the two hadronic forward calorimeters are positioned at either end of CMS, to pick up the myriad of particles coming out of the collision region at very small angles relative to the beam line. These receive the bulk of the particle energy contained in the collision so must be very resistant to radiation and use different materials with respect to the other parts of the HCAL.

The Muon Detectors

As the name *Compact Muon Solenoid* suggests, detecting muons is one of the most important tasks of the CMS experiment. Muons are charged particles that are just like electrons and positrons, but are 200 times heavier. They are expected 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. Because muons can penetrate several metres of iron without interacting they are not stopped by the calorimeters. Therefore, chambers to detect muons are placed at the very edge of the experiment.

A muon is measured by tracking its position through the multiple layers of each of four muon station and, combining with tracker measurements, the detectors precisely trace the particle path. This gives a measurement of its momentum because we know that particles travelling with more momentum bend less in a magnetic field. The CMS magnet is very powerful, so even the paths of very high-energy muons can be bended.

In total there are 1400 muon chambers:

- 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) track the position of the particles and provide a trigger;
- 610 resistive plate chambers (RPCs) form a redundant trigger system, which quickly decides to keep the acquired muon data or not.

DTs and RPCs are arranged in concentric cylinders around the beam line (the barrel region) while CSCs and RPCs, make up the endcap disks that cover the ends of the barrel.

3.4.2 Trigger

When CMS is performing at its peak, about one billion proton-proton interactions will 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 would be unlikely to reveal new phenomena. It is therefore necessary to have a **trigger** whose purpose is to select the potentially interesting events and reduce the rate to just a few hundred events per second, allowing events to 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 arrives. 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.

3.5 Physics at the LHC

Our current understanding of the Universe is incomplete. The Standard Model of particles and forces summarizes our present knowledge of particle physics. The Standard Model has been tested by various experiments and it has proven particularly successful in anticipating the existence of previously undiscovered particles, for example W and Z bosons, gluons, or the top and charm quarks. However, it leaves many unsolved questions, which the LHC will help to answer:

- The Standard Model 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** is a theory that hypothesises the existence of more massive partners of the standard particles we know today. This theory 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. Physicists are searching for particles or phenomena responsible for dark matter (23% of the universe) and dark energy (73% of the Universe). A very popular idea is that dark matter is made of neutral, but still undiscovered, supersymmetric particles.

- The LHC will also help us to 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. The LHC could help to provide an answer to this imbalance between matter and antimatter.

- Heavy-ion collisions at the LHC will provide a window onto the state of matter that would have existed in the very 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.

One of the main results obtained with the LHC is the discovery of the Higgs boson. The Standard Model does not explain the origin of mass, nor why some particles are very heavy while others have no mass at all. According to the theory of the Higgs mechanism, the whole of 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. On July 4, 2012, the two main experiments at the LHC (ATLAS and CMS) both reported independently that they had found a new particle with a mass of about 125 GeV which was consistent with the Higgs boson [8] [9]. On October 8, 2013, the Nobel prize in physics was awarded jointly to François Englert and Peter Higgs *for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN Large Hadron Collider.*

Chapter 4

Data Analysis

This chapter describes the experimental procedure used to select events for the measure of the $t\bar{t}$ production cross section ($\sigma_{t\bar{t}}$).

4.1 Introduction

Quarks 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. 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.

The aim of this work is the calculation of the $t\bar{t}$ cross section in the fully-hadronic channel. This means that we are interested in events which follow the decay chain

$$t\bar{t} \rightarrow W^+b \;\; W^-\bar{b} \rightarrow qq'b \;\; qq'\bar{b},$$

in which the decay products are at least six jets arising from the quarks produced in the decay, two of which coming from b quarks. The fully-hadronic channel has the largest branching ratio among the possible decay channels (46%) so a large amount of signal events is produced. However this channel is overwhelmed by the QCD multijet background processes. In order to perform the wanted measurement is thus necessary a properly optimized event selection.

4.2 Monte Carlo simulations

As a result of pp collisions at LHC, a great amount of particles is produced and only few detected events are interesting for the analysis while the others

are uninteresting background events. Consequently the signal (S), in this case the $t\bar{t}$ pairs that decay in the fully-hadronic channel, is overwhelmed by a huge background (B) and at this level it is impossible to discriminate S from B. Monte Carlo (MC) simulations are used to devise a strategy for reducing the background while keeping as much signal as possible.

The CMS collaboration uses MC generation programs to model a number of physics processes relevant to $t\bar{t}$ production and decay. The next-to-leading-order (NLO) POWHEG [10] generator is used to generate $t\bar{t}$ signal events, assuming a top quark mass of $m_t = 172.5\ GeV$. An alternative sample can be obtained using the MG5_AMC@NLO [11] generator. In this work events generetated by the POWHEG generator have been used; these samples are inclusive of all possible $t\bar{t}$ decays, and contain about 20 millions events. On the other hand, as long as the background dominates over the signal, the data themselves can be considered as a good representation of the background. Comparing the data behavior ($S + B$) with that of the MC simulations (S) it is possible to notice signal and background differences in order to separate them in the clearest possible way.

4.3 The cross section

The *cross section* is a physical quantity that represents the occurrence probability of a given event. It is measured in cm^2 or alternatively in *barn* (b). The relation between cm^2 and b is:

$$1b = 10^{-24} cm^2. \tag{4.1}$$

In our specific case (pp collisions) the $t\bar{t}$ *production cross section* ($\sigma_{t\bar{t}}$) is the probability that a pp collision will produce $t\bar{t}$ pairs. This cross section can be evaluated from theoretical models: it depends on the top quark mass (see eq. 2.1) and on the available centre-of-mass energy ($\sqrt{s} = 13\ TeV$ at the LHC, for the 2015 run). The theoretical value for $\sigma_{t\bar{t}}$ is 832 pb [12]. From an experimental point of view the purpose of the cross section measurements is to check if there is an agreement with the theoretical predicted values. Any excess with respect to the theoretical cross section can be a hint of new physics.

The number of expected $t\bar{t}$ events at LHC can be determined using the following relation:

$$N_{t\bar{t}}^{exp} = \epsilon \times \sigma_{t\bar{t}} \times L, \tag{4.2}$$

where:

- ϵ is the Monte Carlo efficiency:

$$\epsilon = \frac{N_{cut}^{MC}}{N_{gen}^{MC}}, \tag{4.3}$$

where N_{gen}^{MC} is the number of $t\bar{t}$ events generated in the Monte Carlo sample while N_{cut}^{MC} is the number of Monte Carlo $t\bar{t}$ events that pass a certain selection;

- L is the *integrated luminosity*. It is measured in cm^{-2} or more conveniently in $barn^{-1}$ (or its submultiples) and it is defined as:

$$L = \int L_{inst}(t)dt, \tag{4.4}$$

where L_{inst} is the *instantaneous luminosity*. This latter quantity is measured in $cm^{-2}s^{-1}$ and depends only on the features of the accelerating machine and the particle bunches involved in the collisions. If two bunches containing n_1 and n_2 particles collide head-on with frequency ν the expression for the instantaneous luminosity is:

$$L_{inst} = \frac{n_1 n_2 N_b \nu}{4\pi\sigma_x\sigma_y}, \tag{4.5}$$

where σ_x and σ_y characterize the transverse bunch sizes. It is assumed that there are N_b circulating bunches in each direction, and that the bunches are identical in transverse profile, that the profiles are Gaussian and independent of position along the bunch, and the particle distributions are not altered during bunch crossing.

The instantaneous luminosity and the integrated luminosity are useful parameters to characterize the performance of a particle accelerator and a period of data collection. In particular, all collider experiments aim to maximize their integrated luminosities, as the higher the integrated luminosity, the more data is available to analyze.

Knowing the values of L, ϵ, and $N_{t\bar{t}}^{exp}$ in equation 4.2 it is possible to obtain a value for $\sigma_{t\bar{t}}$.

4.4 Data selection

The data used in this work were collected by the CMS experiment in 2015 and after pre-processing they have been saved in *root* files. The data selection was made by means of a ROOT macro written in C++. Broadly this macro:

- receives a parameter that indicates the samples that need to be processed (data or MC).

- reads the root files, extracts the events, acquires kinematic variables and order the jets with decreasing p_T (transverse momentum);
- requests a specific trigger;
- applies a selection on the number of jets (N_{jets});
- applies a selection on the number of b-tagged jets (N_{btag}). These are the jets coming from b quarks decay;
- applies a selection on the opening angle ΔR_{bb} between the b-tagged jets (see section 4.4.3).

These selections aim to improve the S/B ratio selecting the signal events and rejecting the background ones as much as possible. An estimate of the S/B ratio can be obtained this way:

- we count the number of MC events that pass a given selection;
- using equation 4.3 we calculate the MC efficiency;
- by putting in equation 4.2 the theoretical $t\bar{t}$ cross section ($\sigma_{t\bar{t}} = 832\ pb$), the data integrated luminosity ($L_{data} = 2.474\ fb^{-1}$) and the efficiency we obtain the number of expected events (i.e. the signal S);
- we count the number of data events that pass the same selection (N_{data});
- the background (B) is roughly given by: $N_{data} - S$.

4.4.1 Trigger selection, request on the number of jets and kinematic cuts

Events are selected by two different triggers:

- *HLT_PFHT450_SixJet40_PFBTagCSV*, which requires six jets with $p_T \geq 40\ GeV$, one b-tagged jet and p_T sum for all jets $\geq 450\ GeV$;
- *HLT_PFHT400_SixJet30_BTagCSV0p5_2PFBTagCSV*, which requires six jets with $p_T \geq 30\ GeV$, two b-tagged jet and p_T sum for all jets $\geq 400\ GeV$;

An event is selected if it passes at least one of these two triggers. Events satisfying the trigger requirements are then selected according to the number of jets and some kinematic cuts over jets variables.

Considering the kinematics of a fully-hadronic event the logical request on the number of jets is:

$$N_{jets} \geq 6. \tag{4.6}$$

To be selected, each jet must have $p_T \geq 40\ GeV$ and $|\eta| \leq 2.4$. The quantity η is called pseudorapidity and is associated to the polar angle (θ) between the jet direction and the beam collision axis:

$$\eta = -\log\left[\tan(\frac{\theta}{2})\right]. \tag{4.7}$$

4.4.2 b-tag selection

Jets that arise from bottom-quark hadronization (b-tagged 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. Accurately identifying 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 and, to a lesser extent, the charm hadrons can be used to identify the hadronic jets into which the b and c quarks fragment. These hadrons have relatively large masses, long lifetimes and daughter particles with hard momentum spectra and large impact parameter. 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-tagging).

A variety of reconstructed objects (tracks, vertices and identified leptons) can be used to build observables that discriminate between b jets 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. In this work the CSV b-tagging algorithm [13] has been used; in particular, in order to minimize the misidentification probability, and hence the multijet background, the **tight** operating point has been chosen.

The $t\bar{t}$ decay is characterised by the presence of two b-tagged jets in the final state so we require:

$$N_{btag} \geq 2. \tag{4.8}$$

4.4.3 ΔR_{bb} selection

Another quantity that has been taken into account is the ΔR_{bb} variable which is a function of the angles between the two b-tagged jets:

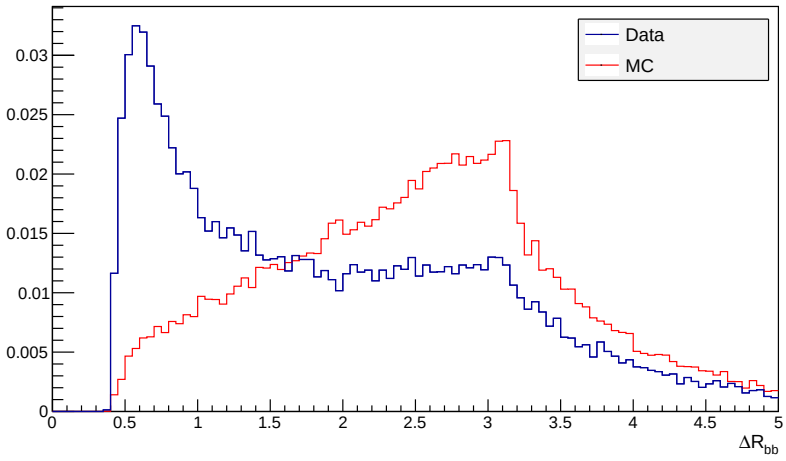

Figure 4.1: Behavior of ΔR_{bb} for MC events (red) and for data events (blue).

$$\Delta R_{bb} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{4.9}$$

where

$$\Delta\eta = \eta_1 - \eta_2, \tag{4.10}$$

$$\Delta\phi = \phi_1 - \phi_2. \tag{4.11}$$

Here η_i and ϕ_i are the pseudorapidity and the azimuthal angles of the i-th b-tagged jet.

In order to apply a selection on ΔR_{bb} we compared its behavior for MC events and for data events (see Fig. 4.1).

It can be noticed that the two behaviours are quite different. This is due to different production mechanisms of b quarks in background events. They can be divided into three topologies:

- *Direct production*: two b quarks are produced in a $2 \to 2$ Feynman diagram. In this case ΔR_{bb} has a value of about 3;
- *Gluon splitting*: two b quarks are produced from a gluon decay. The angle between these two quarks is usually small, so tipically $\Delta R_{bb} < 1$;
- *Flavor excitation* corresponds to a more elaborate diagram. ΔR_{bb} has an intermediate value between those of the two previous cases.

On the other hand, b-jets coming from $t\bar{t}$ decays are widely separeted ($1 \leqslant \Delta R_{bb} \leqslant 4$ approximately). In order to reject background events it was decided to select events with

Selection	N_{data}	S	S/B
Trigger	35149073	54691	$\sim 1/500$
$N_{jets} \geq 6$	1434017	34875	$\sim 1/40$
$N_{btag} \geq 2$	31006	6532	$\sim 1/4$
$\Delta R_{bb} \geq 1.5$	17799	5409	$\sim 1/2$

Table 4.1: S/B ratios for the applied selections.

$$\Delta R_{bb} \geqslant 1.5. \tag{4.12}$$

In this way we reject the entire background from gluon splitting and part of the one from flavor excitation without throwing away too much signal.

In table 4.1 the estimates of the S/B ratio for all the applied selections are reported.

4.5 Mass reconstruction with kinematic fit

The kinematic recostruction of the final state and of the top quark mass in the hypothesis

$$t\bar{t} \to W^+ b W^- \bar{b}$$

is essential for a data-based estimate of the background.

Once the preselection of data has been done, the macro reconstructs the top quark mass by using a kinematic fit carried out on the preselected events in order to improve the S/B ratio and provide a way to distinguish signal from background. The four-momenta $P^\mu = (E, \vec{p})$ of the six jets are evaluated and, going backwards, the W bosons four-momenta and masses are then reconstructed. Adding the b-jets four-momenta to the W bosons ones the four-momenta of the two top quarks are finally obtained:

$$P^\mu_{W^+} = P^\mu_{j_1} + P^\mu_{j_2} \tag{4.13}$$

$$P^\mu_{W^-} = P^\mu_{j_3} + P^\mu_{j_4} \tag{4.14}$$

$$P^\mu_t = P^\mu_{W^+} + P^\mu_b \tag{4.15}$$

$$P^\mu_{\bar{t}} = P^\mu_{W^-} + P^\mu_{\bar{b}} \tag{4.16}$$

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

Selection	N_{data}	S	ϵ	S/B
$\chi^2 < 10$	3000	2033	0.001	~ 2
$\chi^2 < 5$	1672	1409	0.0007	~ 5

Table 4.2: Efficiencies and S/B ratios for χ^2 cuts.

$$\chi^2 = \frac{(m_{jj}^{(1)} + m_W)^2}{\sigma_W^2} + \frac{(m_{jj}^{(2)} + m_W)^2}{\sigma_W^2} + \frac{(m_{jjb}^{(1)} + m_t^{rec})^2}{\sigma_t^2} + \frac{(m_{jjb}^{(2)} + m_t^{rec})^2}{\sigma_t^2} \tag{4.17}$$

where $m_{jj}^{(1,2)}$ are the masses of the two dijet systems coming from the W bosons, $m_{jjb}^{(1,2)}$ are the masses of the two three-jet systems coming from the two top quarks, $m_W = 80.4\, GeV$ is the W boson mass, σ_W and σ_W are the uncertainties associated to the mass distributions of W and t (including their natural widths) in MC $t\bar{t}$ events ($\sigma_W \sim 15\%\ m_{jj}$ and $\sigma_t \sim 11\%\ m_{jjb}$). Functions modeled on the χ^2 test allow checking the validity of a given hypothesis, typically the agreement between the data distribution of an experiment and the expected distribution. A small χ^2 value guarantees a good agreement.

The reconstructed top quark mass m_t^{rec} plays the role of a free parameter. By minimizing the χ^2 function using the MINUIT algorithm, the best value for the top quark mass, common to the two three-jet systems, is then found. Of course, it is impossible to know in advance which combination of the six jets is the one which correctly reconstructs the $t\bar{t}$ pair. This is why it is necessary to repeat the minimization process for every single permutation of the jets. Of these permutations only those where the two b partons are associated to two b-tagged jets with $\Delta R_{bb} \geq 1.5$ are considered. After doing this the values of χ^2 and m_t corresponding to the permutation with minimum (i.e. *best*) value of χ^2 are saved. This procedure has been performed both on the data and the MC samples: two files containing the *best* m_t^{rec} and corresponding χ^2 values have been created for the following analysis. As can be seen in Fig. 4.2 for $t\bar{t}$ signal events, the best m_t^{rec} quantity peaks indeed around the value of the top quark mass. Instead, if we consider all the permutations (see Fig. 4.3) the peak is overwhelmed.

The χ^2 distributions for data and signal are quite different (see Fig. 4.4 and 4.5). Cutting on the χ^2 value of the best permutation will help to reduce the background and increase the signal purity.

In table 4.2 there are two values of S/B ratio when, in addition to the previous selections, two possible χ^2 cuts are applied.

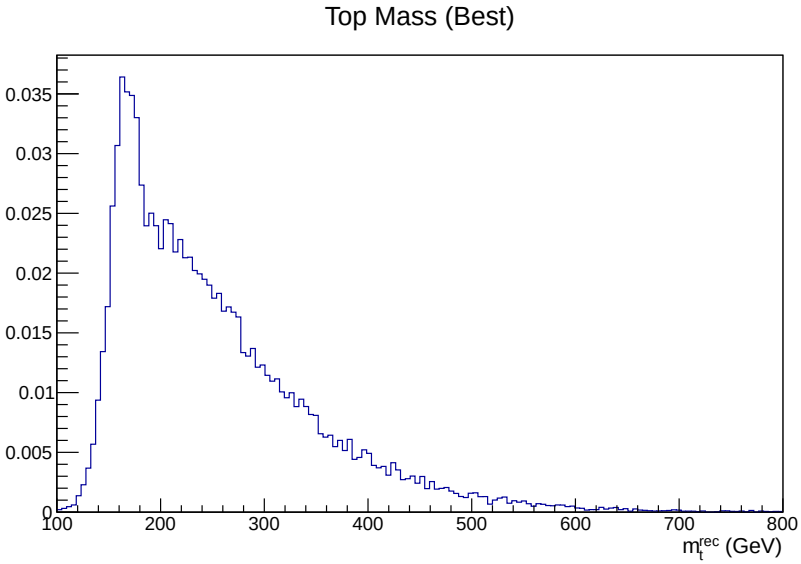

Figure 4.2: Top quark mass for the *best* permutations.

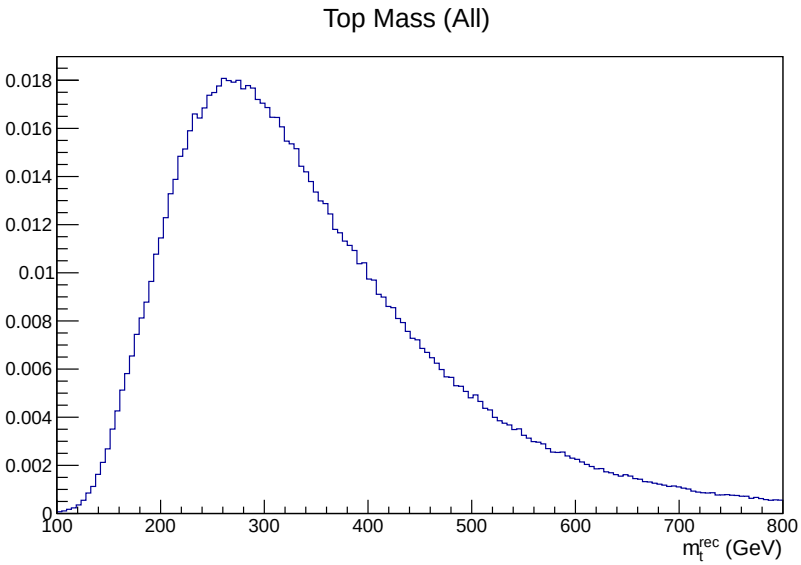

Figure 4.3: Top quark mass for all the permutations

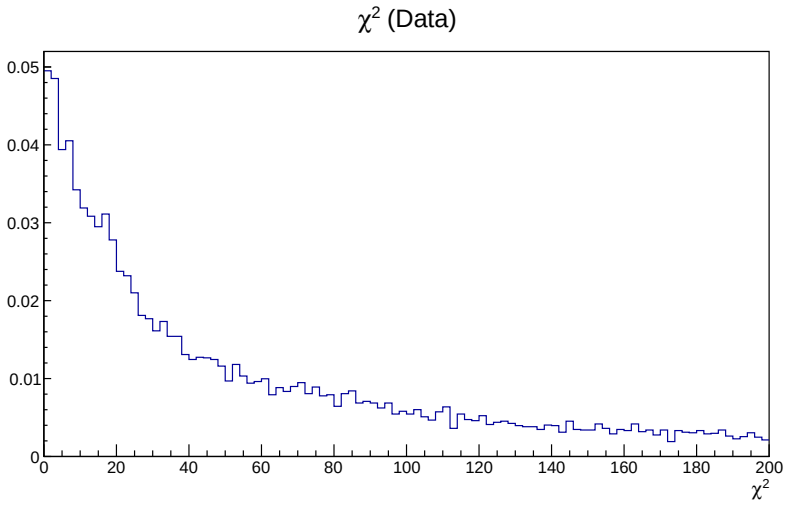

Figure 4.4: χ^2 behaviour for data.

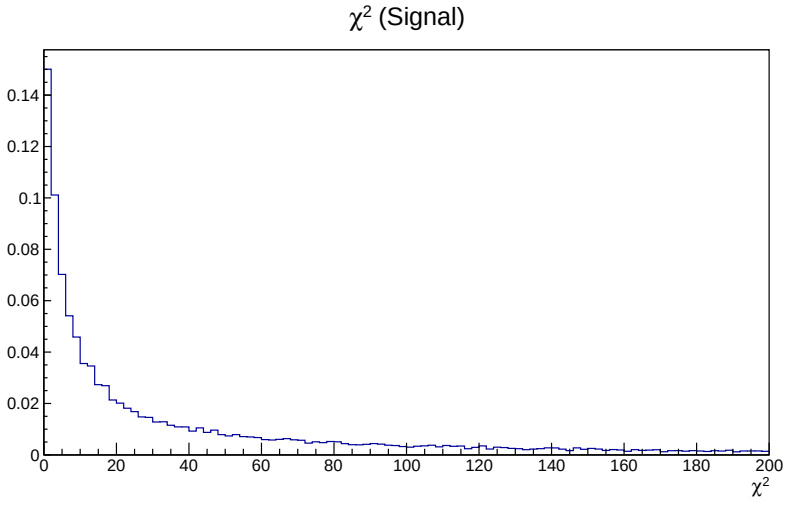

Figure 4.5: χ^2 behaviour for signal (MC).

4.6 Background estimate

The selected events can be modeled as the sum of two different contribution: the signal, (which is provided by the Monte Carlo simulations) and the background. The multijet background cannot be reliably estimated through simulations, so here we use a data-based method to estimate it. Such a method recurs to distributions for the reconstructed top quark mass. In order to obtain such distributions, it is necessary to create a sample of background events.

In this work the background has been estimated using the same data events and requiring the same triggers, kinematical cuts and number of jets per event. The difference is that in this case we require events without b-tagged jets (0-btag events) to ensure that we are not dealing with $t\bar{t}$ events (top quarks decay almost always in Wb as explained in section 2.2.3). At this point two randomly selected jets are arbitrarily "b-tagged"; then the analysis proceed exactly as in the previous case: we require the same ΔR_{bb} cut and reconstruct the top quark mass using the kinematic fit. Since these events have nothing to do with $t\bar{t}$ events they behave as pure background.

In order to ensure that this method reproduces the real background quite closely, another technique, called *event mixing*, can be used to validate the obtained results. This technique consists in mixing two events which satisfy the same conditions required previously: trigger, six jets, two b-tagged jets, kinematic cuts. Mixing is achieved by randomly choosing jets from two different events. Then the analysis goes on as in the previous case (ΔR_{bb} cut and kinematic fit).

The results for the two methods are reported in Fig. 4.6. A χ^2 cut ($\chi^2 > 20$) has been applied before filling the histograms to be sure to select a background-enriched sample.

The two distributions are quite similar but the one from 0-btag events is more smooth. Probably, mixed events still contain a small "contamination" from $t\bar{t}$ events. For this reason we prefer to use the 0-btag events to represent the background, while the other method will be used to assess systematic uncertainities.

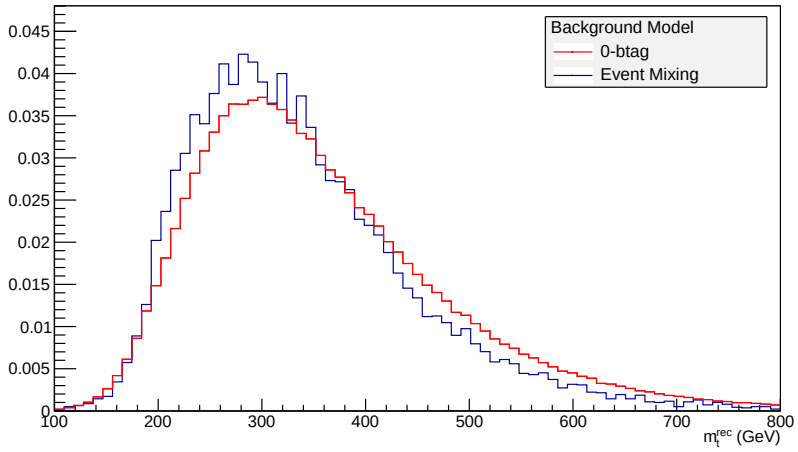

Figure 4.6: Reconstructed top quark mass for a background-enriched sample ($\chi^2 > 20$) using 0-btag events (red) and event mixing (blue).

Chapter 5

Cross section measurement

In order to measure the production cross section we need to extract the signal yield in the selected events. This is done by evaluating the fraction of signal and background events in the observed top quark mass distribution, comparing selected events to the expected distributions from signal and background.

5.1 Likelihood fit

5.1.1 Template histograms

Now that we have three files (data, MC and background) containing the best m_t^{rec} and χ^2 values, the $t\bar{t}$ cross section can be estimated. A ROOT macro called "fitXsec.C" reads the files and creates three histograms (one for each file) applying a χ^2 cut. Then the background and signal histograms are normalized to unit area. The obtained histograms are called *Templates* (see Fig. 5.1); they are the probability functions for the signal $P_s(i)$ and for the background $P_b(i)$ (i refers to the i-th bin of the distributions). The signal is characterized by a peak at a value close to the real top quark mass, while the background is uniformly distribuited (it increases and decreases without showing significant peaks).

5.1.2 Likelihood fit

Starting from these *template* distibutions, in order to obtain the $\sigma_{t\bar{t}}$ value, a likelihood fit is performed. In statistics, a likelihood function ($\mathcal{L}$) is a function of the parameters of a statistical model. Likelihood functions play a key role in statistical inference, especially methods of estimating a parameter from a set of statistics. Generally the likelihood of a set of parameter values (α), given outcomes x, is equal to the probability of those observed outcomes given those parameter values, that is:

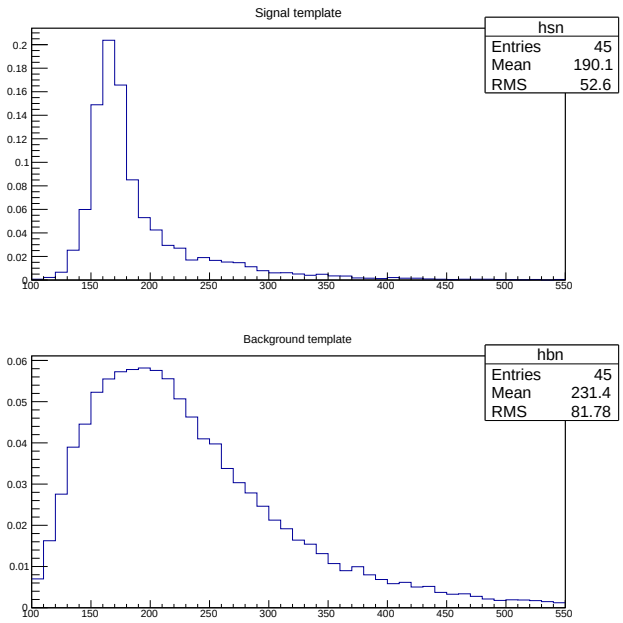

Figure 5.1: Template histograms (m_t^{rec}) for signal (top) and background (bottom).

$$\mathcal{L}(\alpha) = P(x|\alpha). \tag{5.1}$$

In our case the likelihood function is composed by two factors. The first is the normalization factor which has a Poissonian shape:

$$\mathcal{L}_{norm} = e^{-\mu}\frac{\mu^n}{n!}, \tag{5.2}$$

where n is the number of selected events, and μ is the number of expected signal + background events ($\mu = n_s + n_b$). The second factor is bounded to the shape of the distributions:

$$\mathcal{L}_{shape} = \prod_{i=1}^{N} \frac{n_s\, P_s(i) + n_b\, P_b(i)}{n_s + n_b}. \tag{5.3}$$

Therefore the total likelihood function is simply:

$$\mathcal{L} = \mathcal{L}_{norm} \times \mathcal{L}_{shape} = e^{-\mu}\frac{\mu^n}{n!} \times \prod_{i=1}^{N} \frac{n_s\, P_s(i) + n_b\, P_b(i)}{n_s + n_b}. \tag{5.4}$$

This function depends on two parameters: the background and signal yields (n_b and n_s). The latter is obtained applying equations 4.2 which, for simplicity, is reported below:

Parameter	$\chi^2 < 10$	$\chi^2 < 5$
L_{data}	2.474 fb^{-1}	2.474 fb^{-1}
n	2973	1658
$\epsilon'(\%)$	0.06	0.04

Table 5.1: Input parameters for the likelihood fit.

$$n_s = \epsilon(\chi^2) \times \sigma_{t\bar{t}} \times L_{data}. \tag{5.5}$$

The only unknown quantity in this equation is exactly $\sigma_{t\bar{t}}$, the others being reported in table 5.1.

Before performing the likelihood fit the efficiency ($\epsilon(\chi^2)$) has been corrected in order to take account of the difference of b-tagging efficiency as measured in the data with respect to Monte Carlo simulations: a scale factor ($SF_{b-tag} = 0.907$) has been introduced and the correct efficiency (ϵ') has been calculated:

$$\epsilon' = \epsilon(\chi^2) \times SF^2_{b-tag}. \tag{5.6}$$

The scale factor for the b-tagging algorithm is squared because we have required two b-tagged jets.

The target of the likelihood fit is to find the $\sigma_{t\bar{t}}$ and n_b values that maximize the likelihood function. For computational reasons, finding the local minima is easier than finding the maxima, therefore, using MINUIT algorithm, we apply a minimization on $-\log[\mathcal{L}(\sigma_{t\bar{t}}, n_b)]$. This is permitted because the *log* function is strictly monotone. The likelihood function then becomes:

$$-\log[\mathcal{L}(\sigma_{t\bar{t}}, n_b)] = \mu - n\log\mu + \log(n!) - \sum_{i=1}^{N} \log\left[\frac{n_s\, P_s(i) + n_b\, P_b(i)}{n_s + n_b}\right]. \tag{5.7}$$

The minimization provides the values of $\sigma_{t\bar{t}}$ and n_b. Table 5.2 shows the results for these two parameters when two differents χ^2 cuts are required.

Parameter	$\chi^2 < 10$	$\chi^2 < 5$
$\sigma_{t\bar{t}}$ (pb)	798 ± 53	823 ± 54
n_b	874 ± 78	879 ± 52

Table 5.2: Likelihood fit results for $\sigma_{t\bar{t}}$ and n_b. All the quoted uncertanties are statistical.

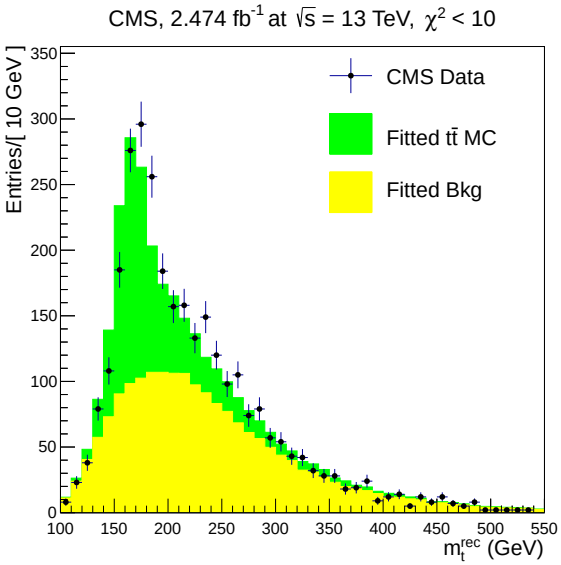

Figure 5.2: Reconstructed top quark mass for $\chi^2 < 10$. Comparison between data (black dots), signal (green) and expected background (yellow).

Now it is also possible to compare the data behaviour with the signal and background sum, as shown in Fig. 5.2 and 5.3. Since the $\chi^2 < 5$ request provide a smaller relative statistical uncertainty, we use the corresponding cross section value as our measurement.

We have used the reconstructed top quark mass as a way to separate signal from background and to evaluate the respective yields. For this reason we need to consider the fact that this reconstructed mass is affected by the uncertain knowledge of the so-called *jet energy scale* (JES). In the usual top quark mass measurements, the JES is calibrated in-situ recurring to a procedure which constrains the dijet invariant masses from W bosons to the world average value of 80.4 GeV. Applying such a method is beyond our goal but we implemented instead a simplified calibration by rescaling all jet energies/momenta by a factor which guarantees a better reconstruction of the W mass. Fig. 5.4 shows the dijet invariant masses from the kinematic fit for data and signal. A better agreement is seen in Fig. 5.5 if we rescale the energy/momentum of the simulated jets by a factor

$$K_{JES} = 1.04.$$

If we apply such a factor to the signal template the likelihood fit returns a cross section of

$$\sigma_{t\bar{t}} = 893 \pm 57\ pb,$$

while the reconstructed top quark mass fits better the data (see Fig. 5.6).

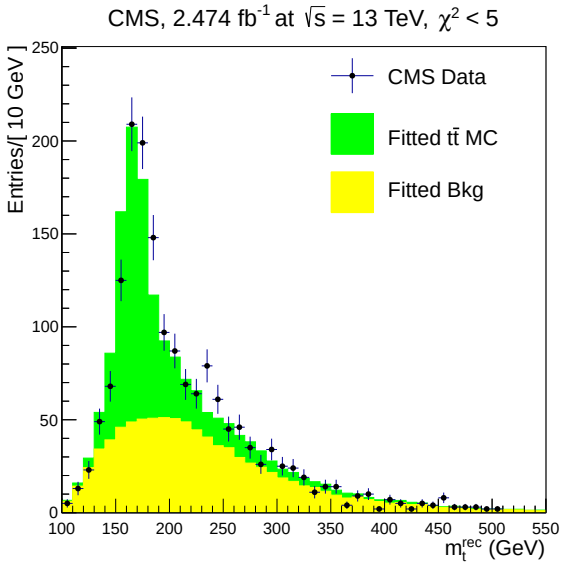

Figure 5.3: Reconstructed top quark mass for $\chi^2 < 5$. Comparison between data (black dots), signal (green) and expected background (yellow).

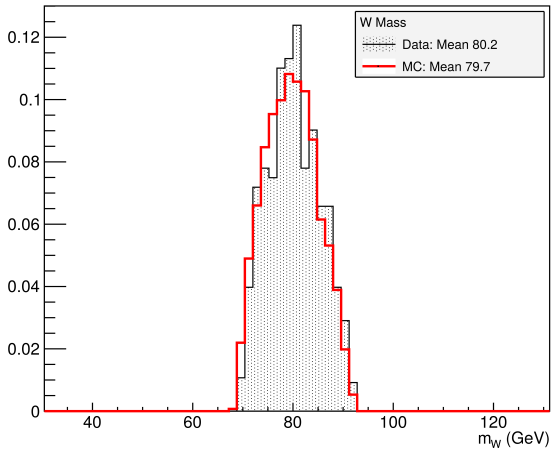

Figure 5.4: Reconstructed mass of the dijet system associated to the W bosons in the kinematic fit. Comparison between data and $t\bar{t}$ simulation.

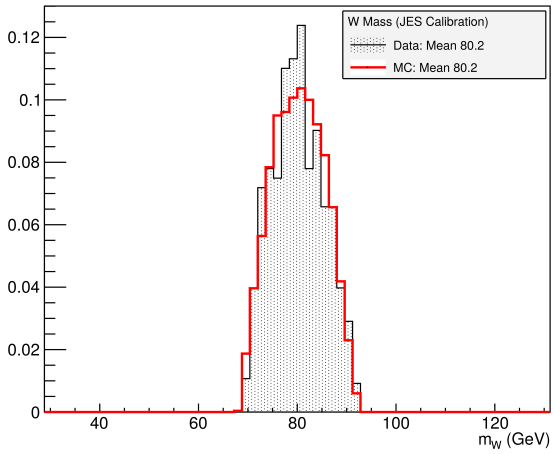

Figure 5.5: Reconstructed mass of the dijet system associated to the W bosons in the kinematic fit, after the JES calibration. Comparison between data and $t\bar{t}$ simulation.

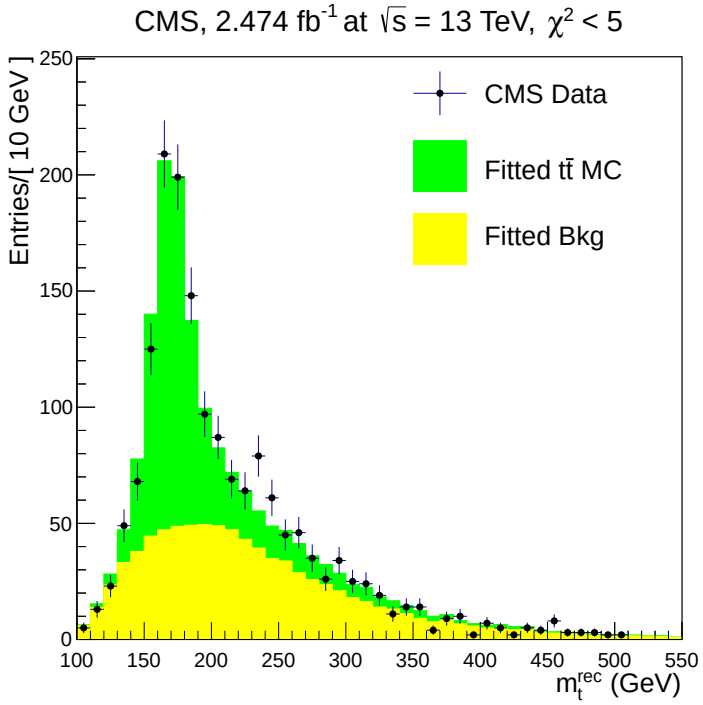

Figure 5.6: Reconstructed top quark mass for $\chi^2 < 5$ and after the JES calibration.

5.2 Systematic uncertainties

Dealing with systematic uncertainities in experiments such as CMS can be very complicated because there are lots of factors that can affect the measurements contributing to the total uncertainty.

The most important sources of systematic uncertainty that affect this measurement come from:

1. simulation of the trigger efficiency;
2. uncertain knowledge of the jet energy scale and resolution;
3. modeling of the signal (generator);
4. choice of the parton distribution functions and QCD scales;
5. modeling of the background;
6. uncertainty on the integrated luminosity value.
7. uncertainty on the b-tagging scale factor.

In the following we assess the uncertainties associated to items 2, 3, 5, 6 and 7.

5.2.1 Jet energy scale and resolution

The jet energies/momenta are known only with a certain resolution given by a factor called *JEC_Uncertainty_Total*. In order to estimate the related uncertainity, the energy components (P_x, P_y, P_z, E, E_T) of the jets in the MC samples have been multiplied by the factor:

$$(1 \pm JEC\_Uncertainty\_Total).$$

Redoing the analysis on these corrected samples two new cross section values have been obtained: one (σ_+) for the "plus" correction, the other (σ_-) for the "minus" one:

$$\sigma_+ = 787 \pm 51\ pb \qquad \sigma_- = 876 \pm 60\ pb.$$

Using these two values we can estimate the systematic uncertainity by taking half the difference with respect to the average:

$$\frac{|\sigma_+ - \sigma_-|}{\sigma_+ + \sigma_-} = 0.05, \tag{5.8}$$

that is a relative 5% uncertainty.

5.2.2 Modeling of the signal

This uncertainty is related to the choice of the Monte Carlo samples used to perform the analysis. Monte Carlo samples are simulations and although the MC generators are extremely sofisticated, there may be slight imperfections in the models they generate. It may be useful to compare the cross section value obtained using POWHEG MC default sample (σ_1) with the value that can be obtained performing the entire analysis using other MC samples (σ_2). In this case a sample produced with the MG5_AMC@NLO generator has been used obtaining for the cross section:

$$\sigma_2 = 837 \pm 53\ pb.$$

The systematic uncertainity is given by:

$$\frac{|\sigma_1 - \sigma_2|}{\sigma_1} = 0.06, \tag{5.9}$$

that is a relative 6% uncertainty.

5.2.3 Modeling of the background

As explained in section 4.6 we decided to use the 0-btag events to represent the background. However, another background estimation method has been tested, the *event mixing*. In order to see how the background model affects the cross section measurement the likelihood fit has been performed using the background events from event mixing. In this case the we obtain for the cross section:

$$\sigma_{mix} = 906 \pm 62\ pb$$

Now, the relative systematic uncertainity can be evaluated:

$$\frac{|\sigma_{0btag} - \sigma_{mix}|}{\sigma_{0btag}} = 0.015, \tag{5.10}$$

that is a relative 1.5% uncertainty. This means that the cross section measurement presented here is quite robust with respect to the modeling of the background.

5.2.4 Integrated luminosity

The value used for the integrated luminosity is known with a 2.7% uncertainty. This turns into an equal relative uncertainty on the cross section.

5.2.5 b-tagging scale factor

The scale factor for the b-tagging efficiency amount to 0.907 ± 0.037. If we vary it by adding or subtracting the uncertainty we obtain the two cross section values:

$$\sigma_{b+} = 824 \pm 53\ pb \qquad \sigma_{b-} = 970 \pm 62\ pb.$$

This corresponds to an uncertainty of

$$\frac{|\sigma_{b+} - \sigma_{b-}|}{\sigma_{b+} + \sigma_{b-}} = 0.08, \tag{5.11}$$

that is a relative 8% uncertainty.

5.2.6 Total systematic uncertainty

Since the sources of systematic uncertainty described above are expected to be quite uncorrelated, for the total uncertainty we consider the quadrature sum of the certainties described above. This amount to 11.6%.

5.2.7 Results

The likelihood fit of the reconstructed top quark mass templates, including the JES calibration, returns a best cross section of

$$893 \pm 57\ (stat) \pm 104\ (syst)\ pb.$$

This result is in good agreement with the theoretical value of 832 pb and with other CMS measurements in the same channel, as can be seen in Fig. 5.7.

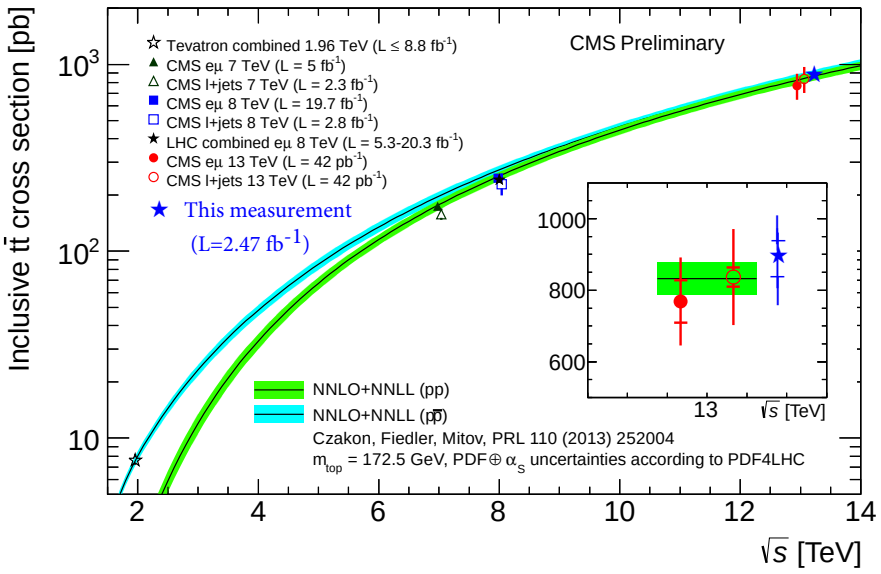

Figure 5.7: Behaviours of the theoretical cross section in function of the centre-of-mass energy displayed along with some experimental measurements. The error bar on this measurement (blue star) represent the statistical (inner bar) and the total (outer bar) uncertainties.

Chapter 6

Conclusions

By using data that have been collected in 2015 by the CMS experiment at LHC, we have developed a procedure that has led to a measurement of the $t\bar{t}$ production cross section in the fully-hadronic channel. First, events from the data samples have been selected in order to improve the signal-background (S/B) ratio and therefore separate $t\bar{t}$ signal events from QCD background events. This step has been optimized by using Monte Carlo simulated events. In order to be selected the events had to satisfy certain conditions and pass some specific cuts: multijet trigger, kinematic cuts, six or more jets, two of which coming from the hadronization of b quarks and a ΔR_{bb} cut.

After this selection a kinematic fit, based on the minimization of a χ^2-like function, was performed in order to reconstruct the top quark mass event per event thus providing a way to distinguish signal from background. The background has been estimated repeating the entire analysis using events whitout b-tagged jets.

Using the data, background and signal distributions (the latter obtained by the Monte Carlo simulations) a likelihood fit has been performed thus obtaining the cross section value. Finally, the effect of some significant sources of systematic uncertainty has been evaluated.

The final result

$$\boxed{\sigma_{t\bar{t}} = 893 \pm 57\ (stat) \pm 104\ (syst)\ pb}$$

is in good agreement with the theoretical value of 832 pb and with other CMS measurements in different channels.

Bibliography

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

[2] `http://www.slac.stanford.edu/pubs/beamline/25/3/25-3-carithers.pdf`.

[3] CDF Collaboration, *Observation of Top Quark Production in pp Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. 74 (1995) 2626.

[4] DØ Collaboration, *Observation of the Top Quark*, Phys. Rev. Lett. 74 (1995) 2632.

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

[6] `http://home.cern/topics/large-hadron-collider`.

[7] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST 3 (2008) S08004, 257.

[8] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012) 1.

[9] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (2012) 30.

[10] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP 06 (2010) 043.

[11] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP 07 (2014) 079.

[12] M. Czakon and A. Mitov, *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. 185 (2014) 2930.

[13] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, JINST 8 (2013) 04013.

Ringraziamenti

Desidero ringraziare infinitamente il Professor Andrea Castro per la cortesia, l'attenzione e la pazienza nei miei confronti, oltre che per la sua costante presenza in ogni aspetto del lavoro svolto durante questi mesi.
Ringrazio la mia famiglia, in particolare i miei genitori, per aver sempre creduto nelle mie capacità e avermi sempre sostenuto in ogni momento di questo percorso.
Ringrazio Jessica per avermi sempre incoraggiato a dare il meglio di me, soprattutto nei momenti più difficili e impegnativi.
Infine un grazie a chiunque mi sia stato vicino durante il cammino che ha portato a questo traguardo.